
Uniwersytet
Gdański

Wydział Biologii
Uniwersytetu Gdańskiego

mgr Wojciech Wesołowski

Charakterystyka bakteriofagów infekujących uropatogenne bakterie *Escherichia coli* oraz identyfikacja w ich genomach czynników o potencjalnym znaczeniu aplikacyjnym i poznawczym

Characterization of bacteriophages infecting uropathogenic *Escherichia coli* and identification of factors in their genomes with both applied and fundamental scientific significance

Praca przedstawiona
Radzie Dyscypliny Nauki Biologiczne Uniwersytetu Gdańskiego
celem uzyskania stopnia doktora
w dziedzinie nauk ścisłych i przyrodniczych
w dyscyplinie nauki biologiczne

Promotor: dr hab. Bożena Nejman-Faleńczyk, prof. UG
Katedra Biologii Molekularnej,
Wydział Biologii, Uniwersytet Gdański

Gdańsk 2026

Szczególne podziękowania składam Sylwii Bloch,
za ogromną wiedzę, którą mi przekazała.

Aleksandrze Łukasiak,
za lata solidnej współpracy.

Panom Marcinowi Górniakowi i Bartłomiejowi Tomiczkowi,
za poszerzenie mojej perspektywy w zakresie filogenetyki.

Natalii Lewandowskiej i Karolinie Zdrojewskiej,
za dyskusje, pomoc oraz cenne uwagi.

I wielu innym, których można pozdrowić Tuwimem,
lecz podniosła forma na to nie pozwala.

Spis treści

Streszczenie rozprawy doktorskiej 5
Lista publikacji wchodzących w skład rozprawy doktorskiej: 17
Abstract of doctoral dissertation 18
List of publications included in the doctoral dissertation: 29
Praca nr 1 30
Suplement do pracy nr 1 50
Oświadczenia współautorów 110
Praca nr 2 119
Suplement do pracy nr 2 148
Oświadczenia współautorów 185
Praca nr 3 205
Suplement do pracy nr 3 272
Oświadczenia współautorów 284
Praca nr 4 304
Suplement do pracy nr 4 335
Oświadczania współautorów 385
Działalność naukowa 395
Publikacje naukowe 396
Wystąpienia na konferencjach naukowych 398
Kursy, staże i szkolenia 399
Projekty naukowe 399

Streszczenie rozprawy doktorskiej

Zakażenia układu moczowego (ZUM) należą do najczęstszych infekcji bakteryjnych na świecie, stanowiąc drugą pod względem częstości grupę zakażeń po infekcjach dróg oddechowych. Dotykają one od około 150 do 250 milionów ludzi rocznie, stanowią około 30% wszystkich zakażeń szpitalnych i generują znaczne wydatki systemu opieki zdrowotnej oraz społeczno-ekonomiczne, przy czym w samych tylko Stanach Zjednoczonych roczny koszt leczenia niepowikłanych ZUM szacuje się na około 2 miliardy dolarów. Choć ZUM występują u obu płci, kobiety są nieproporcjonalnie częściej dotknięte tego typu infekcjami, gdyż ponad połowa kobiet na świecie doświadcza co najmniej jednego epizodu zakażenia w ciągu życia.

Głównym czynnikiem etiologicznym ZUM są uropatogenne szczepy *Escherichia coli* (*uropathogenic Escherichia coli*, UPEC), odpowiadające za około 75–85% niepowikłanych i 50–65% powikłanych zakażeń, które wraz z innymi przedstawicielami rzędu *Enterobacterales*, przede wszystkim *Klebsiella pneumoniae*, stanowią przyczynę większości przypadków infekcji. Dominacja bakterii *E. coli* wśród uropatogenów wynika przede wszystkim z ich zdolności do unikania odpowiedzi immunologicznej gospodarza oraz do zasiedlania dróg moczowych. Co istotne, bakterie UPEC nie są jednolitą grupą, albowiem obejmują szczepy istotnie różniące się profilami wirulencji i oporności. Genom *E. coli* można podzielić na zachowany ewolucyjnie genom rdzeniowy (*core genome*), obejmujący geny kodujące białka pełniące podstawowe funkcje życiowe komórki oraz sekwencje wysoce zmienne (*accessory genome*), obejmujące mobilne elementy genetyczne, takie jak plazmidy, profagi i wyspy genomowe, których nabywanie lub utrata, za pośrednictwem horyzontalnego transferu genów, kształtuje zdolność bakterii do kolonizacji układu moczowego, ich patogenność i oporność na antybiotyki.

Patogenność UPEC jest determinowana przez zróżnicowany zestaw czynników wirulencji, a nie przez jeden uniwersalny czynnik. Genomy tych bakterii zawierają geny związane z adhezją, umożliwiające przyleganie do komórek nabłonka i kolonizację, w tym fimbrie typu 1, P i S, adhezyny Dra/Afa oraz antygen 43. Niosą one również geny kodujące toksyny, które modulują odpowiedź immunologiczną i uszkadzają komórki gospodarza, jak na przykład α-hemolizyna, martwiczy czynnik cytotoksyczny 1 (*cytotoxic necrotizing factor 1*, CNF1) czy białko USP (*uropathogenic-specific protein*), a także geny zaangażowane w biosyntezę i transport sideroforów, umożliwiające efektywne pozyskiwanie żelaza. Dodatkowe geny i białka związane z wirulencją obejmują czynniki inwazyjne, białka autotransporterowe, elementy sekrecyjnego systemu typu VI oraz produkty uczestniczące w biosyntezie otoczki i

lipopolisacharydu, które ułatwiają ucieczkę przed odpowiedzią immunologiczną i hamują proces fagocytozy.

Tworzenie biofilmu jest kluczowym etapem patogenezy ZUM, umożliwiającym UPEC skuteczne przetrwanie i ekspansję w obrębie dróg moczowych. Rozwój biofilmu umożliwiają struktury komórkowe, takie jak fimbrie typu 1 i P, które rozpoznają receptory na powierzchni komórek nabłonka gospodarza i inicjują produkcję macierzy zewnątrzkomórkowej chroniącej bakterie przed antybiotykami i komórkami układu odpornościowego. W macierzy biofilmu występują dodatkowo komórki przetrwałe (*persister cells*), które potrafią przetrwać ataki układu odpornościowego oraz terapie antybiotykowe i umożliwiają ponowny wzrost populacji bakteryjnej po zakończeniu antybiotykoterapii, co sprawia, że eradykacja biofilmu jest skrajnie trudna, a nawroty zakażeń są częste. Biofilmy mogą również powstawać na powierzchniach abiotycznych, takich jak cewniki moczowe, co czyni leczenie odcewnikowych ZUM szczególnie trudnym zadaniem, albowiem wydłużające się cewnikowanie zwiększa ryzyko zakażenia o około 5% na dobę. W środowisku szpitalnym biofilm bakteryjny stanowi źródło przewlekłych zakażeń oraz miejsce intensywnego międzykomórkowego transferu genów, sprzyjającego selekcji i rozprzestrzenianiu się oporności na antybiotyki.

Oporność na środki przeciwdrobnoustrojowe wśród uropatogenów jest stale rosnącym problemem, napędzanym w dużej mierze niewłaściwym stosowaniem antybiotyków i generuje wysokie koszty leczenia. Odnotowane wskaźniki oporności różnią się znacznie w zależności od regionu geograficznego i grupy antybiotyków. Oporność na trimetoprim–sulfametoksazol wśród izolatów UPEC w Europie wynosi od 14,6 do 60%, podczas gdy oporność na amoksycylinę–kwas klawulanowy mieści się w zakresie od 5,3% w Niemczech do 37,6% we Francji, a w krajach rozwijających się odnotowuje się jeszcze wyższe wartości. Oporność na cefalosporyny II i III generacji również rośnie w Europie, a badania izolatów klinicznych UPEC wykazały oporność na cefalosporyny III generacji i amoksycylinę–kwas klawulanowy w niektórych regionach przekraczającą 70–85%. Rozległe stosowanie fluorochinolonów, w szczególności cyprofloksacyny, przyczyniło się do stałego wzrostu oporności na tę grupę leków, która jest wyraźnie bardziej rozpowszechniona w regionach o niskich i średnich dochodach (w zakresie 55,5–85,5%) niż w krajach wysoko rozwiniętych (w zakresie 5,1–32,0%). Z drugiej strony, izolaty UPEC na ogół pozostają wrażliwe na karbapenemy i nitrofurantoinę, przy czym ta ostatnia, nadal stanowi terapię pierwszego wyboru w zależności od lokalnych wzorców oporności i rodzaju zakażenia. Skalę problemu odzwierciedla włączenie *E. coli* do listy patogenów monitorowanych przez Europejskie Centrum ds. Zapobiegania i

Kontroli Chorób (ECDC) a także dane z globalnego systemu nadzoru WHO GLASS, wskazujące na istotne obciążenie związane z opornością na środki przeciwdrobnoustrojowe wśród *E. coli* i *K. pneumoniae*, będących częstymi przyczynami zakażeń krwi. W związku z rosnącą częstością oporności na środki przeciwdrobnoustrojowe wśród bakterii UPEC, pojawiła się pilna potrzeba opracowania terapii alternatywnych.

Wśród alternatywnych strategii walki z zakażeniami wywoływanymi przez wielolekooporne UPEC obiecującym podejściem jest terapia fagowa. Lityczne bakteriofagi posiadają kilka swoistych zalet w porównaniu z konwencjonalnymi środkami przeciwdrobnoustrojowymi: (I) mogą niszczyć biofilmy bakteryjne, które są oporne na leczenie antybiotykami, (II) namnażają się w miejscu zakażenia, zapewniając w ten sposób samoreplikujące się środki bakteriobójcze oraz (III) zabijają patogeny selektywnie, zachowując przy życiu pożyteczne bakterie mikrobiomu. Mimo to zastosowanie terapeutyczne fagów pozostaje wyzwaniem. Do terapii nadają się wyłącznie fagi o ściśle litycznym cyklu rozwojowym, a bioinformatyczna predykcja drogi rozwoju fagów często wymaga potwierdzenia eksperymentalnego. Rzadko badanym aspektem jest to, że aktywność fagów może być obniżona przez odpowiedź immunologiczną gospodarza. Ponadto, wiele bakteriofagów jest wysoce specyficznych wobec zaledwie kilku szczepów bakterii, co uniemożliwia ich szerokie, uniwersalne zastosowanie.

Co ważne, genomy fagów stanowią również cenne źródło informacji genetycznej, mogącej zawierać geny kodujące białka i inne czynniki o znaczeniu zarówno aplikacyjnym, jak i poznawczym. Z jednej strony produkty genów znajdujących się w genomach fagów mogą mieć użyteczne aktywności biologiczne, takie jak liza komórek bakteryjnych lub degradacja struktur na powierzchni bakterii, przez co mogą stanowić źródło nowych środków przeciwdrobnoustrojowych. Z drugiej strony, sekwencje genomów fagowych mogą zawierać informacje przydatne w badaniach różnorodności i ewolucji fagów, umożliwiając identyfikację cech genomowych przydatnych w określaniu pokrewieństwa oraz rekonstrukcji zależności filogenetycznych między fagami. Niewątpliwie systematyczna eksploracja genomów fagów może jednocześnie przyczyniać się do odkrywania nowych białek o znaczeniu aplikacyjnym oraz do lepszego zrozumienia molekularnej i ewolucyjnej charakterystyki bakteriofagów.

W niniejszej rozprawie skoncentrowaliśmy się na bakteriofagach infekujących bakterie UPEC. Badania obejmowały zarówno fagi już opublikowane, których sekwencje genomowe były dostępne w bazie GenBank, jak i poszukiwanie nowych fagów oraz ich kompleksową charakterystykę, uwzględniającą analizę morfologiczną i fizjologiczną, ocenę stabilności i

bezpieczeństwa wobec komórek ludzkich, a także sekwencjonowanie, składanie i szczegółową analizę ich genomów. Przeprowadziliśmy również analizę filogenetyczną badanych fagów w celu określenia ich pokrewieństwa z innymi wirusami infekującymi bakterie UPEC. Z kolei analizy genomowe ukierunkowaliśmy na identyfikację genów i cech genomowych o potencjalnym znaczeniu poznawczym i aplikacyjnym. Szczególną uwagę poświęciliśmy identyfikacji genów kodujących białka enzymatyczne, w tym białka zaangażowane w lizę komórek bakteryjnych i potencjalne depolimerazy, mogące stanowić źródło nowych narzędzi przeciwdrobnoustrojowych. Z drugiej strony, badania miały także na celu identyfikację genów o silnym sygnale filogenetycznym, odgrywających istotną rolę w rekonstrukcji relacji ewolucyjnych nawet dla blisko spokrewnionych fagów infekujących UPEC.

W **artykule nr 1** przeprowadziliśmy analizę 98 bakteriofagów i wykazaliśmy ich znaczny potencjał terapeutyczny w zwalczaniu UPEC. Stwierdziliśmy, że zdecydowana większość tych izolatów, a dokładnie 95%, została oceniona pod kątem zakresu gospodarza, co stanowi fundament zrozumienia spektrum ich aktywności wobec różnych szczepów bakteryjnych. W naszym badaniu wyróżniliśmy siedem fagów: vB_Ecos_Golestan, myPSH1131, Killian, 590b, FS2B, UGKSEcP1 i UGKSEcP2 jako dobrze scharakteryzowanych, potencjalnych kandydatów do dalszych badań nad terapią fagową. Te konkretne szczepy fagowe przeszły szczegółową analizę fizjologiczną i bioinformatyczną, obejmującą ocenę zakresu gospodarza, stabilności, kinetyki rozwoju i organizacji genomu. Co godne uwagi, analiza faga FS2B ujawniła jego silną aktywność lityczną, duży plon fagowy oraz wyjątkową stabilność w szerokim zakresie temperatur i pH, wraz z potwierdzonym litycznym cyklem rozwojowym i brakiem genów związanych z lizogenią. Ponadto zaobserwowaliśmy, że fag ten wykazywał synergistyczne działanie z badanymi antybiotykami, co sugeruje, że stanowi on przykład dobrze scharakteryzowanego faga, który prezentuje duży potencjał, aby stać się skuteczną, spersonalizowaną opcją terapeutyczną w nawracających zakażeniach układu moczowego.

Analizy bioinformatyczne wykazały ponadto znaczne zróżnicowanie genomowe fagów infekujących UPEC. Wielkości ich genomów mieściły się w zakresie od 39 091 do 169 905 pz, a liczba przewidzianych otwartych ramek odczytu (*open reading frames*, ORF) od 49 do 305. Adnotacje genomów oraz analizy porównawcze przeprowadziliśmy w celu określenia funkcji białek fagowych. Co istotne, osiem wcześniej zsekwencjonowanych, lecz nieopisanych genomów fagów zostało w tej pracy poddanych adnotacji po raz pierwszy. Spośród 792 unikatowych białek fagowych zgromadzonych w tabelach uzupełniających S2 i S3, aż 79 białek

fagowych o potencjalnym charakterze aplikacyjnym zidentyfikowaliśmy w nowo zadnotowanych genomach fagów anty-UPEC. Szczególny nacisk położyliśmy na identyfikację fagowych enzymów o potencjalnym zastosowaniu w strategiach przeciwdrobnoustrojowych, w tym białek uczestniczących w lizie komórek bakteryjnych oraz potencjalnych depolimeraz, zidentyfikowanych na podstawie strukturalnych analiz bioinformatycznych. Łącznie, u wszystkich badanych fagów zidentyfikowano 40 klasycznych endolizyn, 75 holin, 40 spanin i 16 potencjalnych depolimeraz. Aby dokładniej scharakteryzować potencjał rozwojowy fagów, wykonaliśmy predykcje cyklu życiowego, dostarczając nowych wyników i wskazując kilka fagów wymagających dalszej weryfikacji eksperymentalnej ze względu na potencjalnie lizogeniczny cykl rozwojowy. Ponadto opracowaliśmy wieloetapowe podejście do analizy filogenetycznej badanych fagów. W pierwszym etapie, 82 fagi pogrupowaliśmy w klady na podstawie ich wstępnego podobieństwa i pokrewieństwa ewolucyjnego, określonego na podstawie drzewa proteomicznego wygenerowanego z wykorzystaniem programu ViPTree. Następnie, za pomocą oprogramowania OMA, zidentyfikowaliśmy grupy genów ortologicznych. W kolejnym etapie oceniliśmy poszczególne matryce uszeregowań sekwencji białkowych pod kątem liczby miejsc parsymonicznie informatywnych. Na tej podstawie wytypowaliśmy geny dostarczające informacji filogenetycznej i wykorzystaliśmy je do skonstruowania drzew wielogenowych. Należały do nich między innymi geny kodujące białka ogonka, białka portalowe oraz białka związane z replikacją genomu. Z analizy wykluczyliśmy sekwencje o wysokim stopniu zakonserwowania.

Te nowo wykonane analizy w znacznym stopniu wzbogaciły dostępną charakterystykę bioinformatyczną genomów fagowych i dostarczyły nowych informacji o ich potencjale aplikacyjnym, cechach rozwojowych i relacjach ewolucyjnych tych fagów. Co ważne, analizy przeprowadzone z użyciem wielu narzędzi bioinformatycznych konsekwentnie wskazywały na brak czynników wirulencji i genów oporności na antybiotyki w analizowanych genomach fagów. Wyniki te demonstrują istotność i znaczenie kompleksowej charakterystyki fagów dla oceny ich stabilności, bezpieczeństwa wobec komórek ludzkich, ich cech rozwojowych i zależności ewolucyjnych, przed zastosowaniem terapeutycznym.

Mimo tych obiecujących wyników zidentyfikowaliśmy istotne luki metodologiczne i niespójności w opublikowanych pracach, które utrudniają ocenę przydatności badanych fagów w terapii fagowej. Nasza praca przeglądowa (**artykuł nr 1**) ujawniła, że kluczowe dane dotyczące bezpieczeństwa fagów w dużej mierze nie istnieją, a co ważne, żaden z 98 analizowanych fagów nie został przetestowany pod kątem cytotoksyczności wobec komórek

eukariotycznych. Ponadto tylko 42% przeanalizowanych fagów miało zbadaną wydajność infekcji (*efficiency of plating*, EOP). Brak wyników tej analizy, jest istotnym zaniedbaniem prowadzącym do nieprecyzyjnych ocen skuteczności działania opublikowanych fagów. Z kolei badania stabilności przeprowadzono jedynie dla 20% szczepów fagowych pod względem temperatury i 19% pod względem pH. Dodatkowo 15% genomów fagowych w ogóle nie zostało zsekwencjonowanych, a inne nie miały wykonanej prawidłowej adnotacji ani analizy filogenetycznej. W celu zaradzenia tym ograniczeniom doszliśmy do wniosku, że przyszłe badania powinny opierać się na ustandaryzowanych, kompleksowych protokołach charakterystyki fagów, a konkretnie, opowiedzieliśmy się za obowiązkowymi testami EOP, dokładnymi badaniami przesiewowymi pod kątem cytotoksyczności, badaniami stabilności w warunkach fizjologicznie istotnych oraz rygorystyczną weryfikacją bioinformatyczną genomów pod kątem obecności genów toksyn i oporności na antybiotyki, aby móc ocenić potencjał badanych fagów w zastosowaniu w terapii fagowej.

Mając na uwadze ograniczenia dotyczące charakterystyki bakteriofagów, zidentyfikowane **w pracy nr 1**, staraliśmy się im zaradzić we własnych badaniach eksperymentalnych, przedstawionych w **artykule nr 2**. W tej pracy wyizolowaliśmy i wszechstronnie scharakteryzowaliśmy kilka nowych litycznych bakteriofagów infekujących kliniczne szczepy UPEC oraz zaproponowaliśmy pięć nowych gatunków fagów: *Hanrivervirus vB_EcoS-149_4M*, *Warwickvirus vB_EcoS-149_6D*, *Nonavirus vB_EcoS-164_2*, *Kagunavirus vB_EcoS-179_3* i *Vectrevirus vB_EcoP-179_5*. Co ważne, wyszliśmy daleko poza opisowe raportowanie dominujące w literaturze: zamiast jedynie rejestrować zakres gospodarzy standardowo stosowanym testem nakropień (*spot test*), wykonaliśmy analizę EOP i zintegrowaliśmy charakterystykę fagów z porównawczą analizą genomową samych gospodarzy bakteryjnych. Ponadto wykonaliśmy szeroką analizę bioinformatyczną genomów fagowych, wykluczyliśmy występowanie genów toksyn i genów oporności na antybiotyki oraz udokumentowaliśmy obecność genów kodujących potencjalne białka lityczne, takie jak endolizyny, holiny i spaniny, u wszystkich nowo zidentyfikowanych fagów.

Aby ustalić pozycję filogenetyczną badanych fagów, zmodyfikowaliśmy nieco podejście zastosowane w **pracy przeglądowej nr 1**. Odległości genomowe pomiędzy analizowanymi fagami określiliśmy za pomocą programu VIRIDIC, a następnie przeprowadziliśmy rekonstrukcję relacji filogenetycznych na podstawie drzew wielogenowych. Do analiz wykorzystaliśmy grupy genów ortologicznych zidentyfikowane za pomocą oprogramowania OMA. Po wykonaniu uszeregowania sekwencji oraz ocenie ich

sygnału filogenetycznego wykluczyliśmy sekwencje charakteryzujące się niewystarczającą liczbą miejsc parsymonicznie informatywnych, a pozostałe wykorzystaliśmy do rekonstrukcji drzew. Wśród genów wykorzystanych do rekonstrukcji znajdowały się między innymi geny kodujące białka ogonka, małą i dużą podjednostkę terminazy, a także białka związane z lizą komórki, w tym holiny i spaniny.

Ponadto wykazaliśmy, że właściwości biologiczne i bezpieczeństwo badanych fagów istotnie różnicują ich potencjalną przydatność w terapii fagowej. Fag vB_EcoS-149_4M, charakteryzujący się najszerszym zakresem gospodarza, wysoką stabilnością w badanych warunkach oraz korzystnym profilem bezpieczeństwa, obejmującym brak genów kodujących toksyny i determinanty oporności na antybiotyki oraz brak wykrywalnego działania cytotoksycznego wobec badanych linii komórek ludzkich, stanowi obiecującego kandydata do dalszych zastosowań aplikacyjnych. Natomiast fagi vB_EcoS-179_3 i vB_EcoP-179_5 prezentują bardzo wąski zakres gospodarza, ograniczony do infekcji szczepu EC179, co wskazuje na ich potencjalną przydatność raczej w podejściu zindywidualizowanym niż w zastosowaniach o charakterze uniwersalnym.

Chociaż identyfikacja pięciu nowych gatunków bakteriofagów anty-UPEC poszerzyła dotychczasową pulę 98 opisanych fagów zdolnych do infekcji tych uropatogenów, to uzyskane wyniki wskazują jednak, że liczba opisanych fagów litycznych infekujących UPEC nadal pozostaje niewystarczająca w porównaniu z dużą liczbą i znacznym zróżnicowaniem genotypowym i fenotypowym przedstawicieli tej grupy bakterii. Szczególnie wymownym przykładem jest szczep EC256, dla którego w ramach przeprowadzonych badań nie udało się zidentyfikować infekującego go bakteriofaga.

Strategię wieloaspektowej charakterystyki pozyskanych ze środowiska bakteriofagów rozszerzyliśmy jeszcze bardziej w **pracy nr 3**, dotyczącej bakteriofaga vB-EcoS_57-3 i jego endolizyny 57_3Lys. W tej pracy wyizolowaliśmy faga ze ścieków miejskich, scharakteryzowaliśmy go jako faga o morfologii siphovirusów i oceniliśmy morfologię łysinek, która sugeruje działanie depolimerazy. W pełni zsekwencjonowaliśmy i opisaliśmy jego genom o długości 45 774 pz, obejmujący 91 ORF, oraz jednoznacznie ustaliliśmy jego pozycję taksonomiczną jako nowego przedstawiciela rodzaju *Veterinaerplatzvirus* w rodzinie *Drexlerviridae*. Ponadto eksperymentalnie zbadaliśmy cykl rozwojowy faga: określiliśmy czas latencji wynoszący 11 minut, plon fagowy na poziomie około 42 PFU na komórkę, umiarkowanie szybką adsorpcję osiągającą wydajność 80% w ciągu 10 minut oraz wysoce wydajny rozwój lityczny, prowadzący do istotnej redukcji liczby żywych komórek

bakteryjnych w ciągu 30 minut. Co istotne, przetestowaliśmy faga w warunkach zbliżonych do warunków zamierzonego zastosowania klinicznego, czyli w sztucznym moczu przy pH 4,5, 5,8 i 6,5, w którym fag pozostał stabilny i redukował liczbę bakterii do zaledwie 10% początkowej ich liczby po 24 godzinach. Dodatkowo wykluczyliśmy niekorzystne działanie lizatu fagowego na komórki ssacze, wykazując brak cytotoksyczności wobec ludzkich linii pęcherza moczowego (T24) i nerki (HEK293).

W dalszej części pracy zwróciliśmy uwagę na kodowane przez fagi białka, ponieważ enzymy lityczne pochodzenia fagowego stanowią ciekawą grupę czynników o właściwościach przeciwbakteryjnych. Jak podsumowaliśmy w **artykule nr 4**, endolizyny to enzymy degradujące peptydoglikan, produkowane w trakcie litycznego cyklu faga i mają wiele zalet, takich jak szybkie zabijanie bakterii, niskie prawdopodobieństwo rozwoju oporności, aktywność synergistyczna z innymi środkami przeciwdrobnoustrojowymi oraz skuteczność zarówno wobec komórek planktonicznych, jak i biofilmów bakteryjnych. Poprzez zastosowanie technik inżynierii genetycznej, ich spektrum działania można nawet celowo modyfikować, na przykład poprzez fuzję z komponentami uszkadzającymi błonę zewnętrzną, domenami wiążącymi ścianę komórkową dla bardziej specyficznego działania lub peptydami przeciwdrobnoustrojowymi. Niemniej jednak liczba doniesień na temat endolizyn wykazujących bezpośrednią aktywność wobec uropatogenów, w szczególności bakterii UPEC, pozostaje ograniczona. Istotnym ograniczeniem ich zastosowania wobec Gram-ujemnych bakterii UPEC jest obecność błony zewnętrznej, która stanowi barierę utrudniającą dostęp egzogennie aplikowanych enzymów do warstwy peptydoglikanu. Dodatkowe wyzwania obejmują ryzyko zaburzeń mikrobiomu wynikające ze zbyt szerokiego spektrum aktywności przeciwbakteryjnej, trudności w oczyszczaniu oraz fakt, że zbyt niskie stężenia endolizyn mogą zmieniać fenotyp bakterii w sposób potencjalnie niekorzystny dla pacjenta, co uwidacznia konieczność przeprowadzenia dokładnej oceny ich bezpieczeństwa.

Poszukując opisanych w **artykule nr 4** endolizyn, przeanalizowaliśmy genomy 75 litycznych fagów anty-UPEC dostępnych w bazie GenBank. Analiza ta pozwoliła nam zidentyfikować 39 potencjalnych endolizyn w 30 genomach fagowych. Zidentyfikowane enzymy obejmowały zarówno białka wcześniej opisane przez innych badaczy, jak i białka opisane po raz pierwszy przez nas. Przeprowadzona analiza *in silico* wykazała znaczną różnorodność potencjalnych endolizyn pod względem przewidywanej aktywności enzymatycznej. Wśród zidentyfikowanych białek wyróżniono pięć grup obejmujących enzymy z domenami lizozymowymi, amidazowymi i peptydazowymi, a także białka o przewidywanej

aktywności transglikozylazowej i hydrolazowej. Parametry fizykochemiczne zidentyfikowanych endolizyn wskazują, że masy cząsteczkowe tych białek mieściły się w zakresie od około 8,7 do 22,6 kDa, a ich punkty izoelektryczne wskazywały na zasadowy charakter. Wartości wskaźnika niestabilności sugerowały dobrą stabilność większości analizowanych białek, natomiast wysokie wartości wskaźnika alifatyczności wskazywały na ich potencjalną stabilność termiczną. Ujemne wartości współczynnika GRAVY wskazywały z kolei na hydrofilowy charakter białek i ich przewidywaną rozpuszczalność w wodzie. Dodatkowo w sekwencjach tych białek zidentyfikowaliśmy wiele sekwencji przeciwdrobnoustrojowych peptydów (*antimicrobial peptides*, AMP), co wskazuje na ich potencjalne właściwości lityczne.

Szczególnie interesujące było również nasze odkrycie, że w genomach siedmiu fagów występują duże białka o masie od 120 do 145 kDa z domenami przypominającymi endolizyny, które prawdopodobnie ułatwiają wstrzykiwanie materiału genetycznego faga do wnętrza komórki na początku infekcji, a nie ostateczną lizę komórki. Co więcej, przeprowadzona przez nas analiza filogenetyczna i funkcjonalna ujawniła problemy związane z prawidłową adnotacją niektórych, opisanych przez innych badaczy potencjalnych endolizyn. W przypadku kilku białek pierwotnie opisanych jako endolizyny analiza zidentyfikowanych domen wskazała na możliwość ich błędnej klasyfikacji. Przykładowo, w przypadku białek pochodzących z fagów Killian i ZCEC14 zidentyfikowano domeny charakterystyczne odpowiednio dla beztlenowej rybonukleozydo-trifosforanowej reduktazy oraz RNazy H T4, a także domenę inhibitora proteazy uczestniczącej w dojrzewaniu główki bakteriofaga. Obecność tych domen, a także przeprowadzona klasyfikacja filogenetyczna sugerowały, że analizowane białka mogą pełnić inne funkcje niż degradacja peptydoglikanu i prawdopodobnie nie są endolizynami. Wyniki te wskazują na konieczność szczegółowej weryfikacji białek adnotowanych jako endolizyny fagowe.

Kontynuacją prac bioinformatycznych dotyczących fagowych endolizyn były nasze kolejne badania eksperymentalne, opisane w **pracy nr 3**. W tej pracy dokonaliśmy walidacji funkcjonalnej endolizyny faga vB-EcoS_57-3, nazwanej 57_3Lys. *In silico* ustaliliśmy, że 57_3Lys to niewielkie białko o długości 161 aminokwasów i masie 17,55 kDa. Na podstawie analizy domen zidentyfikowanych w 57_3Lys udało się ustalić, że jest to endolizyna SAR klasy II, reprezentowanej przez endolizynę R21 lambdoidalnego faga 21. W obrębie sekwencji 57_3Lys zidentyfikowaliśmy N-końcowy, umiarkowanie hydrofobowy segment 20 reszt jako domenę „signal-anchor-release", katalityczną triadę $E(X)_8$-$D(X)_5$-T w pozycjach E26, D35 i

T41 oraz kationowy, amfipatyczny region podobny do peptydów przeciwdrobnoustrojowych AMP, między resztami 110 a 140. W ramach przeprowadzonych prac eksperymentalnych, sklonowaliśmy i nadprodukowaliśmy enzym, a następnie oczyszczaliśmy go metodą chromatografii powinowactwa, uzyskując białko o masie około 17,55 kDa, zgodnej z wartością przewidywaną *in silico*. Następnie, wykorzystując azydek sodu (NaN_3), będący inhibitorem translokazy Sec u *E. coli*, wykazaliśmy zależne od stężenia zahamowanie lizy komórek bakteryjnych indukowanej przez 57_3Lys. Przy stężeniu 5 mM NaN_3 efekt lityczny był wyraźnie ograniczony, natomiast przy 10 mM został niemal całkowicie zahamowany. W połączeniu z obecnością przewidywanej N-końcowej domeny SAR oraz charakterystycznego dla endolizyn SAR klasy II motywu $E(X)_8$-$D(X)_5$-T, wyniki te dostarczyły zbieżnych dowodów na udział szlaku Sec w translokacji 57_3Lys do przestrzeni peryplazmatycznej. Oprócz tego, wykazaliśmy, że oczyszczone białko 57_3Lys, zastosowane zewnętrznie wobec komórek klinicznego izolatu UPEC, EC57, wykazywało aktywność przeciwbakteryjną bez dodatku zewnętrznego permeabilizatora błony zewnętrznej, powodując istotne zmniejszenie zarówno mętności hodowli, jak i liczby żywych komórek wyrażonej w CFU/mL. Efekt ten rozwijał się powoli i był obserwowany po 24 godzinach inkubacji, podczas gdy w ciągu pierwszych 60 minut nie stwierdzono istotnego działania przeciwbakteryjnego. W przeciwieństwie do 57_3Lys, komercyjnie dostępny lizozym nie wykazywał istotnego działania przeciwbakteryjnego, co prawdopodobnie wynikało z ograniczonego dostępu enzymu do warstwy peptydoglikanu. Obserwowana aktywność 57_3Lys wobec komórek Gram-ujemnych może być związana z obecnością przewidywanego C-końcowego regionu AMP, obejmującego reszty 110–140 i wzbogaconego w reszty dodatnio naładowane oraz hydrofobowe, który może sprzyjać elektrostatycznym oddziaływaniom z lipopolisacharydem i destabilizacji błony zewnętrznej, ułatwiając dostęp domeny enzymatycznej do peptydoglikanu. Należy jednak podkreślić, że udział tego regionu w aktywności 57_3Lys pozostaje hipotezą wymagającą bezpośredniej weryfikacji eksperymentalnej**.**

Podsumowując, przedmiotem niniejszej rozprawy doktorskiej była kompleksowa analiza bakteriofagów infekujących uropatogenne szczepy *E. coli* oraz poszukiwanie w ich genomach czynników o potencjalnym znaczeniu aplikacyjnym i poznawczym. Podjęcie tej tematyki wynikało z narastającego problemu oporności na środki przeciwdrobnoustrojowe wśród uropatogenów, związanego między innymi z niewłaściwym stosowaniem antybiotyków. Skala tego zjawiska jest zróżnicowana w zależności od regionu geograficznego oraz stosowanej klasy antybiotyków, co wskazuje na potrzebę poszukiwania alternatywnych metod zwalczania

zakażeń wywoływanych przez UPEC. Jedną z takich możliwości stanowi terapia fagowa, której potencjał wynika między innymi ze zdolności fagów litycznych do eliminacji biofilmów, namnażania się w miejscu zakażenia oraz selektywnego oddziaływania na bakterie patogenne, przy ograniczonym wpływie na bakterie komensalne.

Przeprowadzony przez nas systematyczny przegląd 98 bakteriofagów anty-UPEC wykazał jednak istotne braki w sposobie ich dotychczasowej charakterystyki. Szczególnie widoczne były ograniczenia dotyczące oceny bezpieczeństwa, w tym brak badań cytotoksyczności wobec komórek eukariotycznych, a także niepełne oznaczenia EOP, ograniczony zakres badań stabilności oraz niekompletna adnotacja i analiza genomów. Wyniki te wskazały na potrzebę stosowania bardziej kompleksowego i ujednoliconego schematu oceny bakteriofagów. Wyniki uzyskane w części eksperymentalnej rozprawy stanowiły naszą bezpośrednią odpowiedź na zidentyfikowane wcześniej ograniczenia. Wyizolowaliśmy i kompleksowo scharakteryzowaliśmy nowe lityczne bakteriofagi infekujące kliniczne szczepy UPEC, co doprowadziło do zaproponowania pięciu nowych gatunków. Szczególną uwagę zwrócił fag vB_EcoS-149_4M, który ze względu na szeroki zakres gospodarza, wysoką stabilność oraz korzystny profil bezpieczeństwa został wskazany jako kandydat do dalszych badań w kierunku oceny jego przydatności w terapii fagowej. Podejście to rozszerzyliśmy dodatkowo poprzez pogłębioną charakterystykę bakteriofaga vB-EcoS_57-3 i jego endolizyny 57_3Lys. Istotnym elementem badań była ocena właściwości faga w warunkach związanych z potencjalnym zastosowaniem w zakażeniach układu moczowego. Wykazaliśmy jego stabilność i aktywność lityczną w sztucznym moczu o wartościach pH odpowiadających zakresowi występującemu w drogach moczowych, a jednocześnie nie stwierdziliśmy cytotoksycznego działania wobec ludzkich komórek. Równolegle przeprowadziliśmy analizę 75 genomów litycznych bakteriofagów anty-UPEC, koncentrując się na identyfikacji potencjalnych endolizyn. Zidentyfikowaliśmy 39 białek należących do pięciu grup funkcjonalnych, różniących się przewidywaną aktywnością enzymatyczną. Analiza ta pokazała również, że sama adnotacja genomu nie zawsze pozwala na wiarygodną identyfikację endolizyn. Szczegółowa analiza endolizyny 57_3Lys pozwoliła następnie na połączenie przewidywań bioinformatycznych z wynikami badań funkcjonalnych. Uzyskane wyniki wskazują, że zidentyfikowana przez nas endolizyna 57_3Lys wykazuje działanie antybakteryjne i w przyszłości może mieć praktyczne zastosowanie w leczeniu ZUM. Enzym ten był zdolny do znaczącego zmniejszenia zarówno mętności, jak i liczby żywych komórek klinicznego izolatu UPEC bez dodatku egzogennego permeabilizatora błony zewnętrznej, co podkreśla jego

aktywność wobec osłony bakterii Gram-ujemnych i umacnia potencjał terapeutyczny enzymów litycznych pochodzenia fagowego jako strategii komplementarnej wobec terapii fagowej.

Co istotne, uzyskane w ramach tej rozprawy wyniki mają znaczenie nie tylko w kontekście poszukiwania nowych czynników przeciwdrobnoustrojowych, ale również znaczenie poznawcze, istotne z punktu widzenia badań nad biologią i ewolucją bakteriofagów. Analiza genów wpływających na topologię drzew filogenetycznych pozwoliła wskazać sekwencje szczególnie przydatne do odtwarzania relacji ewolucyjnych pomiędzy fagami anty-UPEC. Wybrane geny, w tym geny kodujące białka ogona, białka portalowe, lityczne oraz białka związane z replikacją, charakteryzowały się silnym sygnałem filogenetycznym. Mogą one zatem stanowić użyteczne markery genomowe, pomocne w rozróżnianiu blisko spokrewnionych fagów oraz zwiększaniu wiarygodności ich klasyfikacji i rekonstrukcji zależności ewolucyjnych.

Ostatecznie wyniki uzyskane w ramach niniejszej rozprawy pokazują, że bakteriofagi są źródłem różnorodnych czynników o potencjalnym znaczeniu zarówno aplikacyjnym, jak i poznawczym. Ich kompleksowa charakterystyka pozwala nie tylko identyfikować i oceniać kandydatów do wykorzystania w nowych strategiach przeciwdrobnoustrojowych, lecz także dostarcza informacji przydatnych do poznania różnorodności, biologii i ewolucji tych wirusów. Szczególne znaczenie ma połączenie klasycznych badań biologicznych z analizą bioinformatyczną genomu i badaniem zależności filogenetycznych, co pozwala na pełniejszą ocenę zarówno możliwości wykorzystania fagów i ich produktów, jak i głębsze zrozumienie biologii i ewolucji bakteriofagów.

Lista publikacji wchodzących w skład rozprawy doktorskiej:

Praca nr 1

Łukasiak A[#], **Wesołowski W**[#], Neumann J, Lewandowska N, Węglińska E, Bloch S, Węgrzyn G, Nejman-Faleńczyk B. Groundwork for phage therapy: multi-faceted comparative analysis of lytic bacteriophages infecting uropathogenic *Escherichia coli*. Virology. 2026 Feb;615:110738. doi: 10.1016/j.virol.2025.110738.

[#] co-first authors

Praca nr 2

Wesołowski W[#], Czerwonka G[#], Zegadło K, Jagieła E, Łukasiak A, Grabowski Ł, Mantej J, Dziuba A, Bloch S, Topka-Bielecka G, Necel A, Kostovski D, Węglińska E, Narajczyk M, Węgrzyn G, Nejman-Faleńczyk B, Adamus-Białek W. Significant variability in virulence factors and antibiotic resistance of uropathogenic *Escherichia coli* strains may be counteracted by their susceptibility to lytic bacteriophages. Sci Rep. 2026 Jul 21;16(1):26677. doi: 10.1038/s41598-026-62793-9.

[#] co-first authors

Praca nr 3

Wesołowski W, Bloch S, Czerwonka G, Jagieła E, Grabowski Ł, Jeschke M, Jacewicz J, Piwnicka E, Morcinek-Orłowska J, Loika H, Czaplewska P, Węgrzyn G, Adamus-Białek W, Łukasiak A, Nejman-Faleńczyk B. Antibacterial activity of a newly identified phage-derived endolysin and its parental bacteriophage against clinical uropathogenic *Escherichia coli*. bioRxiv. doi: 10.64898/2026.09.09.750462.

Praca nr 4

Wesołowski W, Łukasiak A, Bloch S, Kuligowska K, Neumann J, Lewandowska N, Węglińska E, Węgrzyn G, Nejman-Faleńczyk B. Phage endolysins as promising and effective candidates for use against uropathogenic *Escherichia coli*. Viruses. 2025 Apr 13;17(4):560. doi: 10.3390/v17040560.

Abstract of doctoral dissertation

Urinary tract infections (UTIs) are among the most common bacterial infections worldwide, constituting the second most frequent group of infections after those affecting the respiratory tract. They affect approximately 150–250 million people annually, represent about 30% of all hospital-acquired infections, and impose a substantial medical and socioeconomic burden, with the annual cost of uncomplicated UTIs in the United States alone estimated at approximately $2 billion. Although UTIs occur in both sexes, women are disproportionately affected, with more than half of women worldwide experiencing at least one episode during their lifetime.

The predominant etiological agent of UTIs is uropathogenic *Escherichia coli* (UPEC), which accounts for approximately 75–85% of uncomplicated and 50–65% of complicated infections, and together with other members of the order *Enterobacterales*, notably *Klebsiella pneumoniae*, constitutes most cases. The predominance of *E. coli* among uropathogens stems primarily from its ability to evade the host immune response and to colonize the environment of the urinary tract. Importantly, UPEC does not constitute a monolithic group, as it comprises strains that differ substantially in virulence and resistance profiles. The genome of *E. coli* can be divided into a conserved core genome, encoding vital functions, and a highly flexible accessory portion comprising mobile genetic elements such as plasmids, prophages, and genomic islands, whose acquisition or loss, mediated by horizontal gene transfer, shapes colonization, pathogenicity, and antibiotic resistance.

The pathogenicity of UPEC is determined by a diverse repertoire of virulence determinants rather than by a single universal factor set. Genomes of these bacteria contain genes involved in adhesion, enabling attachment to epithelial cells and colonization, including type 1, P, and S fimbriae, Dra/Afa adhesins, and antigen 43. They also carry genes encoding toxins that modulate immune responses and damage host cells, such as α-hemolysin, cytotoxic necrotizing factor 1 (CNF1), and the uropathogenic-specific protein (USP), as well as genes involved in siderophore biosynthesis and transport, which facilitate efficient iron acquisition. Additional virulence-associated genes and proteins include invasion-associated factors, autotransporter proteins, type VI secretion system components, and products involved in capsule and lipopolysaccharide biosynthesis, which contribute to immune evasion and inhibition of phagocytosis.

Biofilm formation constitutes a crucial stage in the pathogenesis of UTIs, enabling UPEC to persist and expand effectively within the urinary tract. Biofilm development is

facilitated by structures such as type 1 and P fimbriae, which recognize receptors on the surface of host epithelial cells and initiate production of an extracellular matrix that protects the bacteria from antibiotics and immune cells. The biofilm matrix additionally harbours persister cells, which are capable of surviving immune system attacks and antibiotic therapies and permit immediate regrowth of the bacterial population after antibiotic treatment, rendering biofilm eradication extremely challenging and infection relapses frequent. Biofilms may also form on abiotic surfaces such as urinary catheters, which makes treatment of catheter-associated UTIs particularly difficult, as prolonged catheterization increases the risk of infection by approximately 5% per day. In the hospital environment, bacterial biofilm serves as a reservoir of chronic infection and a site of intensive intercellular gene transfer, promoting the selection and dissemination of antibiotic resistance.

Antimicrobial resistance among uropathogens is a continuously growing problem, driven largely by the incorrect use of antibiotics, and it generates high treatment costs. Reported resistance rates vary markedly by geographic region and antibiotic class. Resistance to trimethoprim–sulfamethoxazole among UPEC isolates in Europe spans approximately from 14.6 to 60%, whereas resistance to amoxicillin–clavulanate ranges from 5.3% in Germany to 37.6% in France, with even higher levels reported in developing countries. Resistance to second- and third-generation cephalosporins is also increasing in Europe, and studies of clinical UPEC isolates have documented resistance to third-generation cephalosporins and amoxicillin–clavulanate exceeding 70–85% in some regions. Extensive use of fluoroquinolones, particularly ciprofloxacin, has contributed to a steady rise in resistance to this class, which is markedly more prevalent in low- and middle-income regions (in the range of 55.5–85.5%) than in highly developed countries (in the range of 5.1–32.0%). On the other hand, UPEC isolates generally remain susceptible to carbapenems and nitrofurantoin, the latter still constituting first-line therapy depending on local resistance patterns and infection type. The scale of the problem is reflected in the inclusion of *E. coli* among the pathogens monitored by the European Centre for Disease Prevention and Control (ECDC), as well as in data from the WHO Global Antimicrobial Resistance and Use Surveillance System (GLASS), which indicate a substantial burden of antimicrobial resistance among *E. coli* and *K. pneumoniae*, both of which are frequent causes of bloodstream infections. Given the increasing prevalence of antimicrobial resistance among UPEC, there is an urgent need for alternative therapeutic solutions.

Among alternative strategies against infections caused by multidrug-resistant UPEC, bacteriophage therapy is a promising approach. Lytic bacteriophages offer several inherent

advantages over conventional antimicrobials: (I) they can destroy bacterial biofilms, which are resistant to antibiotic treatment, (II) they replicate at the site of infection, thereby providing self-amplifying antibacterial agents, and (III) they kill pathogens selectively, sparing the commensal microbiota. Nevertheless, the therapeutic application of phages remains challenging. Only phages with a strictly lytic life cycle are suitable for therapy, and bioinformatic predictions of the phage lifestyle often require experimental verification. Phage activity may be reduced by the host immune response, an aspect that remains rarely investigated. Also, the narrow strain specificity of many phages precludes their broad, universal application.

Importantly, phage genomes also represent a valuable source of genetic information that may reveal proteins and other genomic features with both applied and fundamental scientific significance. On the one hand, gene products encoded in phage genomes can possess useful biological activities such as the lysis of bacterial cells or the degradation of bacterial surface structure and thus serve as a source of new antimicrobial agents. On the other hand, phage genomic sequences provide a rich resource for studying phage diversity and evolution, allowing the identification of genomic features that can improve the reconstruction of evolutionary relationships. Thus, systematic exploration of phage genomes can simultaneously contribute to the discovery of novel application-oriented proteins and to a better understanding of the molecular and evolutionary characteristics of bacteriophages.

In this dissertation, we focused on bacteriophages infecting UPEC bacteria. The study encompassed both previously described phages, whose genome sequences were available in the GenBank database, and the isolation of novel phages followed by their comprehensive characterization, including morphological and physiological analyses, assessment of their stability and safety toward human cells, as well as genome sequencing, assembly, and detailed genomic analysis. We also performed phylogenetic analyses of the investigated phages to determine their relatedness to other viruses infecting UPEC. The genomic analyses were focused on identifying genes and genomic features of potential scientific and applied significance. Particular attention was given, on the one hand, to the identification of genes encoding enzymatic proteins, including proteins involved in bacterial cell lysis and putative depolymerases, which may represent a source of novel antimicrobial tools. On the other hand, the study aimed to identify genes with a strong phylogenetic signal to improve the quality and reliability of evolutionary relationship reconstructions, even for closely related phages infecting UPEC.

In **article no. 1,** we performed an analysis of 98 bacteriophages, and we demonstrated the considerable therapeutic potential of these viruses for combating UPEC. We found that most of these isolates, specifically 95%, were evaluated for host range, establishing a foundational understanding of their spectrum of activity against various bacterial strains. Our study identified seven phages: vB_Ecos_Golestan, myPSH1131, Killian, 590b, FS2B, UGKSEcP1, and UGKSEcP2, that stand out as well-characterized candidates for further research towards phage therapy development. These specific phage strains underwent detailed physiological and bioinformatic profiling, including assessments of host range, stability, development kinetics, and genome content. Notably, our examination of phage FS2B revealed its strong lytic activity, a high burst size, and exceptional stability across a wide temperature and pH range, alongside a confirmed lytic replication cycle and lack of lysogeny-related genes. Furthermore, we observed that this phage demonstrated synergistic interactions with tested antibiotics, suggesting that with continued refinement, such a well-documented candidate could become an effective, personalized treatment option for recurrent urinary tract infections.

Bioinformatic analyses further demonstrated substantial genomic diversity among UPEC-infecting phages, with genome sizes ranging from 39,091 to 169,905 bp and the number of predicted ORFs ranging from 49 to 305. Genome annotation and comparative analyses were performed to investigate phage protein functions. Notably, eight previously sequenced but unannotated phage genomes were annotated for the first time in this work. Among all 792 unique phage proteins gathered in Supplementary Tables S2 and S3, 79 interesting phage proteins were identified by us in the newly annotated anti-UPEC phage genomes. A particular focus was placed on the identification of phage-encoded enzymes with potential application in antimicrobial strategies, including proteins involved in bacterial cell lysis and putative depolymerases identified through structure-based bioinformatic analysis. A total of 40 classical endolysins, 75 holins, 40 spanins, and 16 putative depolymerases were identified in the analysed anti-UPEC phage genomes. To further characterize phage developmental potential, life-cycle predictions were performed, providing newly obtained results and revealing several phages requiring further experimental verification because of potentially temperate characteristics. Furthermore, we developed a multi-step approach for the phylogenetic analysis of the phages under investigation. In the first step, the 82 phages were grouped into clades based on their preliminary similarity and evolutionary relatedness, as determined using a proteomic tree generated with ViPTree. Subsequently, orthologous gene groups were identified using OMA. In the next step, individual multiple sequence alignments of the encoded proteins were

evaluated based on the number of parsimony-informative sites. On this basis, genes providing phylogenetic information were selected and used to construct multilocus phylogenetic trees. These included, among others, genes encoding tail proteins, portal proteins, and proteins involved in genome replication. Highly conserved sequences were excluded from the analysis.

Collectively, these newly performed analyses substantially expanded the available genomic characterization of anti-UPEC phages and provided new information on their functional potential, developmental characteristics, and evolutionary relationships. Importantly, screening performed using multiple bioinformatic tools consistently indicated the absence of virulence factors and antibiotic-resistance genes in the analysed phage genomes. These findings demonstrate the value of comprehensive genome analysis for assessing phage safety, developmental characteristics, and evolutionary relationships before therapeutic application.

Despite these promising findings, we identified substantial methodological gaps and inconsistencies that currently hinder the clinical translation of phage therapy. Our review (**article no. 1**) revealed that critical safety data is largely missing; notably, none of the 98 analysed phages were tested for cytotoxicity toward eukaryotic cells. Furthermore, only 42% of the phages had their efficiency of plating (EOP) values determined, a critical omission that leads to inaccurate efficacy assessments. In turn, stability studies were conducted on only 20% of isolates regarding temperature and 19% regarding pH. Additionally, 15% of the phages were not sequenced at all, and others lacked proper annotation or phylogenetic analysis. To address these deficiencies, we conclude that future research must adopt standardized, comprehensive characterization protocols. Specifically, we advocate for mandatory EOP assays, thorough cytotoxicity screening, stability testing under physiologically relevant conditions, and rigorous bioinformatic verification of genomes for toxin or antibiotic-resistance genes to assess the potential of the investigated phages for application in phage therapy.

Building directly upon the limitations we identified in **article no. 1.**, we sought to address these shortcomings through our own experimental work, as presented in **article no. 2**. In that study, we isolated and comprehensively characterized several novel lytic bacteriophages infecting clinical UPEC strains, and we formally proposed five novel phage species: *Hanrivervirus vB_EcoS-149_4M*, *Warwickvirus vB_EcoS-149_6D*, *Nonavirus vB_EcoS-164_2*, *Kagunavirus vB_EcoS-179_3*, and *Vectrevirus vB_EcoP-179_5*. Crucially, we went far beyond the descriptive reporting that dominates the literature: rather than merely recording host range by spot test, we performed EOP analysis and integrated phage characterization with a comparative genomic analysis of the bacterial hosts themselves. Moreover, we performed a

broad assessment of our candidates: we confirmed the absence of toxin and antibiotic-resistance genes in the phage genomes, documented the presence of putative lytic protein genes encoding endolysins, holins, and spanins in all newly identified phages.

To determine phylogenetic positions of the tested phages, we refined the approach used in the review **article no. 1**. The genomic distances among the analyzed phages were determined using VIRIDIC, followed by reconstruction of their phylogenetic relationships based on multilocus phylogenetic trees. Orthologous gene groups identified using OMA were used for these analyses. Following multiple sequence alignment and assessment of their phylogenetic signal, sequences with an insufficient number of parsimony-informative sites were excluded, while the remaining sequences were used for tree reconstruction. The genes selected for phylogenetic reconstruction included, among others, those encoding tail proteins, the small and large terminase subunits, and proteins involved in cell lysis, including holins and spanins.

Furthermore, we demonstrated that the biological properties and safety profiles of the investigated phages substantially influence their potential applicability in phage therapy. The phage vB_EcoS-149_4M, characterized by the broadest host range, high stability under the tested conditions, and a favorable safety profile, including the absence of genes encoding toxins and antibiotic resistance determinants, as well as no detectable cytotoxic activity against the human cell lines examined, represents a promising candidate for further application-oriented studies. In contrast, phages vB_EcoS-179_3 and vB_EcoP-179_5 exhibited a very narrow host range, restricted to infection of strain EC179, indicating that they may be more suitable for personalized approaches than for broad-spectrum applications.

Although the identification of five novel species of anti-UPEC bacteriophages expanded the previously described pool of 98 phages capable of infecting these pathogens, our results indicate that the number of characterized lytic phages infecting UPEC remains insufficient relative to the high abundance and substantial genotypic and phenotypic diversity of this bacterial group. A particularly illustrative example is strain EC256, for which no bacteriophage capable of infecting it was identified in the present study.

Next, we extended this multi-aspect characterization strategy even further in **manuscript no. 3** about bacteriophage vB-EcoS_57-3 and its phage-derived endolysin 57_3Lys. In this work, we isolated the phage from urban sewage, characterized its siphovirus-like morphology and plaque phenotype, suggesting depolymerase activity. We fully sequenced and annotated its 45,774 bp genome comprising 91 predicted ORFs and firmly established its taxonomic position as a new representative of the genus *Veterinaerplatzvirus* within the family

Drexlerviridae. Rather than stopping at descriptive features, we quantified the full phage development cycle: we determined a latent period of 11 minutes, a burst size of approximately 42 PFU per cell, moderately fast adsorption reaching 80% within 10 minutes, and highly efficient lytic development reducing viable cell counts dramatically within 30 minutes. Importantly, we tested our phage under conditions relevant to the intended clinical application: stability and activity in artificial urine at pH 4.5, 5.8, and 6.5, where the phage remained stable and reduced bacterial numbers to as little as 10% of the initial inoculum after 24 hours. Additionally, we excluded adverse effects of the phage lysate on mammalian cells by demonstrating no cytotoxicity against human bladder (T24) and kidney (HEK293) lines.

Following the characterization of bacteriophages, we turned our attention to the proteins they encode, since phage-derived lytic enzymes represent a complementary and versatile class of antibacterial agents. As we summarized in **article no. 4**, endolysins are peptidoglycan-degrading enzymes produced during the phage lytic cycle, and they offer several advantages such as rapid bacterial killing, a low probability of resistance development, synergistic activity with other antimicrobial agents, and effectiveness against both planktonic cells and biofilms. Through genetic engineering, their spectrum of activity can even be deliberately modified, for instance by fusing them with outer-membrane-disrupting components, cell wall-binding domains for targeting specificity, or antimicrobial peptides. Nevertheless, reports on endolysins directly tested against uropathogens, especially UPEC, are limited. The central obstacle is that the outer membrane of Gram-negative bacteria severely restricts access of externally applied enzymes to peptidoglycan. Additional challenges include the risk of microbiota disruption resulting from an overly broad antibacterial activity spectrum, purification difficulties, and the fact that low concentrations of endolysins may alter bacterial phenotypes in ways that could be unfavorable for the patient, underscoring the need for thorough safety evaluations.

In our search for the endolysins described in **article no. 4**, we analyzed the genomes of 75 lytic anti-UPEC phages available in the GenBank database. This analysis enabled us to identify 39 putative endolysins in 30 phage genomes. The identified enzymes included both proteins previously described by other researchers and proteins reported by us for the first time. The *in silico* analysis revealed considerable diversity among the putative endolysins with respect to their predicted enzymatic activities. The identified proteins were classified into five groups comprising enzymes with lysozyme, amidase, and peptidase domains, as well as proteins with predicted transglycosylase and hydrolase activities. The physicochemical properties of the identified endolysins indicated that their molecular masses ranged from

approximately 8.7 to 22.6 kDa, while their isoelectric points indicated a predominantly basic character. The instability index values suggested good stability for most of the analyzed proteins, whereas the high aliphatic index values indicated their potential thermal stability. Negative GRAVY values, in turn, indicated the hydrophilic nature of these proteins and their predicted solubility in water. In addition, multiple antimicrobial peptide (AMP) sequences were identified within the protein sequences, suggesting their potential lytic properties.

Particularly noteworthy was our finding that the genomes of seven phages contained large proteins ranging from 120 to 145 kDa with endolysin-like domains. These proteins are likely involved in facilitating the injection of phage genetic material into the host cell at the initial stage of infection rather than in the final lysis of the bacterial cell. Furthermore, our phylogenetic and functional analyses revealed issues concerning the correct annotation of several putative endolysins previously described by other researchers. For several proteins originally annotated as endolysins, analysis of the identified domains indicated that they may have been incorrectly classified. For example, proteins derived from phages Killian and ZCEC14 were found to contain domains characteristic of anaerobic ribonucleoside-triphosphate reductase and T4 RNase H, respectively, as well as a protease inhibitor domain associated with bacteriophage head maturation. The presence of these domains, together with the results of phylogenetic classification, suggested that the analyzed proteins may perform functions other than peptidoglycan degradation and are therefore unlikely to represent endolysins. These findings highlight the need for thorough verification of proteins annotated as phage endolysins.

The continuation of this bioinformatic groundwork came in our subsequent experimental study in **manuscript no. 3**. In this work, we performed functional validation of the vB-EcoS_57-3 phage endolysin, called 57_3Lys. *In silico*, we established that 57_3Lys is a small 161-amino-acid protein of 17.55 kDa. Based on 57_3Lys domain analysis, we found that it is a class II SAR endolysin, exemplified by the endolysin R21 of lambdoid phage 21. We identified its N-terminal, moderately hydrophobic 20-residue segment as a signal-anchor-release domain, its catalytic $E(X)_8$-$D(X)_5$-T triad at residues E26, D35, and T41, and a cationic, amphipathic antimicrobial peptide-like region between residues 110 and 140. Functionally, we cloned and overproduced the enzyme, purified it via affinity chromatography to the *in silico* predicted molecular weight. Next, using sodium azide (NaN_3), an inhibitor of the Sec translocase in *E. coli*, we demonstrated a concentration-dependent inhibition of 57_3Lys-induced bacterial cell lysis. At a concentration of 5 mM NaN_3, the lytic effect was markedly

reduced, whereas at 10 mM it was almost completely abolished. Combined with the presence of a predicted N-terminal SAR domain and the characteristic $E(X)_8$-$D(X)_5$-T motif of class II SAR endolysins, these results provided convergent evidence for the involvement of the Sec pathway in the translocation of 57_3Lys into the periplasmic space. Besides, we showed that purified 57_3Lys, applied externally, significantly reduced both the turbidity and the viable cell count of clinical EC57, a UPEC strain without any added outer-membrane permeabilizer. This activity developed slowly, becoming significant after 24 h of incubation, in contrast to commercially available hen egg-white lysozyme, which showed no such effect. This activity against the Gram-negative envelope offers a plausible functional role for the C-terminal AMP-like region, presumably facilitating electrostatic interaction with lipopolysaccharide and destabilization of the outer membrane. However, it should be emphasized that the involvement of this region in the activity of 57_3Lys remains a hypothesis that requires direct experimental validation.

In summary, the subject of this doctoral dissertation was a comprehensive analysis of bacteriophages infecting uropathogenic *E. coli* strains and the identification of factors within their genomes with potential application-oriented and fundamental research significance. This research was motivated by the growing problem of antimicrobial resistance among uropathogens, associated, among other factors, with the inappropriate use of antibiotics. The scale of this phenomenon varies depending on the geographical region and the class of antibiotics used, highlighting the need to explore alternative approaches to combating infections caused by UPEC. One such approach is phage therapy, whose potential is associated, among other factors, with the ability of lytic phages to eliminate biofilms, replicate at the site of infection, and selectively target pathogenic bacteria while exerting a limited effect on commensal bacteria.

However, our systematic review of 98 anti-UPEC bacteriophages revealed significant gaps in their characterization to date. The most notable limitations concerned safety assessment, including the lack of cytotoxicity studies using eukaryotic cells, as well as incomplete EOP determinations, limited assessment of phage stability, and incomplete genome annotation and analysis. These findings highlighted the need for a more comprehensive and standardized framework for bacteriophage characterization. The results obtained in the experimental part of this dissertation directly addressed the limitations identified in the systematic review. We isolated and comprehensively characterized novel lytic bacteriophages infecting clinical UPEC strains, leading to the proposal of five new phage species. Particular attention was given to

phage vB_EcoS-149_4M, which, owing to its broad host range, high stability, and favorable safety profile, was identified as a candidate for further studies aimed at assessing its suitability for phage therapy. We further extended this approach through an in-depth characterization of bacteriophage vB_EcoS_57-3 and its endolysin, 57_3Lys. An important component of this work was the assessment of phage properties under conditions relevant to its potential application in urinary tract infections. We demonstrated its stability and lytic activity in artificial urine at pH values corresponding to those occurring in the urinary tract, while no cytotoxic effects on human cells were observed. In parallel, we analyzed 75 genomes of lytic anti-UPEC bacteriophages, focusing on the identification of putative endolysins. We identified 39 proteins belonging to five functional groups that differed in their predicted enzymatic activities. This analysis also demonstrated that genome annotation alone does not always enable reliable identification of endolysins. Detailed characterization of 57_3Lys subsequently enabled us to combine bioinformatic predictions with functional experimental data. The results indicate that the endolysin 57_3Lys identified in this study exhibits antibacterial activity and may have potential future applications in the treatment of UTIs. The enzyme was capable of significantly reducing both the turbidity and the number of viable cells of a clinical UPEC isolate in the absence of an exogenous outer-membrane permeabilizer, highlighting its activity against the Gram-negative bacterial envelope and further supporting the therapeutic potential of phage-derived lytic enzymes as a complementary strategy to phage therapy.

Importantly, the findings of this dissertation are relevant not only to the identification of novel antimicrobial agents but also have fundamental research significance for the study of bacteriophage biology and evolution. Analysis of genes influencing phylogenetic tree topology enabled us to identify sequences particularly informative for reconstructing evolutionary relationships among anti-UPEC phages. The selected genes, including those encoding tail proteins, portal proteins, lysis-associated proteins, and proteins involved in genome replication, exhibited strong phylogenetic signals. They may therefore serve as useful genomic markers for distinguishing closely related phages and improving the reliability of their classification and the reconstruction of their evolutionary relationships.

Overall, the results obtained in this dissertation demonstrate that bacteriophages are a source of diverse factors with potential application-oriented and fundamental research significance. Their comprehensive characterization enables not only the identification and assessment of candidates for incorporation into novel antimicrobial strategies but also provides valuable information for understanding the diversity, biology, and evolution of these viruses. In

particular, the integration of classical biological approaches with genomic bioinformatics and phylogenetic analyses enables a more comprehensive assessment of both the potential applications of phages and their products and the underlying biology and evolutionary history of bacteriophages.

List of publications included in the doctoral dissertation:

Article no. 1

Łukasiak A[#], **Wesołowski W**[#], Neumann J, Lewandowska N, Węglińska E, Bloch S, Węgrzyn G, Nejman-Faleńczyk B. Groundwork for phage therapy: multi-faceted comparative analysis of lytic bacteriophages infecting uropathogenic *Escherichia coli*. Virology. 2026 Feb;615:110738. doi: 10.1016/j.virol.2025.110738.

[#] co-first authors

Article no. 2

Wesołowski W[#], Czerwonka G[#], Zegadło K, Jagieła E, Łukasiak A, Grabowski Ł, Mantej J, Dziuba A, Bloch S, Topka-Bielecka G, Necel A, Kostovski D, Węglińska E, Narajczyk M, Węgrzyn G, Nejman-Faleńczyk B, Adamus-Białek W. Significant variability in virulence factors and antibiotic resistance of uropathogenic *Escherichia coli* strains may be counteracted by their susceptibility to lytic bacteriophages. Sci Rep. 2026 Jul 21;16(1):26677. doi: 10.1038/s41598-026-62793-9.

[#] co-first authors

Article no. 3

Wesołowski W, Bloch S, Czerwonka G, Jagieła E, Grabowski Ł, Jeschke M, Jacewicz J, Piwnicka E, Morcinek-Orłowska J, Loika H, Czaplewska P, Węgrzyn G, Adamus-Białek W, Łukasiak A, Nejman-Faleńczyk B. Antibacterial activity of a newly identified phage-derived endolysin and its parental bacteriophage against clinical uropathogenic *Escherichia coli*. bioRxiv. doi: 10.64898/2026.09.09.750462.

Article no. 4

Wesołowski W, Łukasiak A, Bloch S, Kuligowska K, Neumann J, Lewandowska N, Węglińska E, Węgrzyn G, Nejman-Faleńczyk B. Phage endolysins as promising and effective candidates for use against uropathogenic *Escherichia coli*. Viruses. 2025 Apr 13;17(4):560. doi: 10.3390/v17040560.

Praca nr 1

Groundwork for phage therapy: multi-faceted comparative analysis of lytic bacteriophages infecting uropathogenic Escherichia coli.

Łukasiak A[#], **Wesołowski W**[#], Neumann J, Lewandowska N, Węglińska E, Bloch S, Węgrzyn G, Nejman-Faleńczyk B.

Virology. 2026 Feb;615:110738. doi: 10.1016/j.virol.2025.110738.

[#] co-first authors

Impact Factor: 2.6

Punktacja Ministerstwa Nauki i Szkolnictwa Wyższego: 100

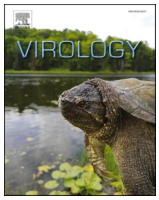
Virology 615 (2026) 110738



Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/virology



Groundwork for phage therapy: multi-faceted comparative analysis of lytic bacteriophages infecting uropathogenic *Escherichia coli*

Aleksandra Łukasiak [a,1], Wojciech Wesołowski [a,1], Julia Neumann [b], Natalia Lewandowska [a], Emilia Węglińska [a], Sylwia Bloch [a,c], Grzegorz Węgrzyn [a], Bożena Nejman-Faleńczyk [a,c,*]

[a] Department of Molecular Biology, Faculty of Biology, University of Gdańsk, Wita Stwosza 59, 80-308, Gdansk, Poland
[b] Laboratory of Environmental Chemoinformatics, Faculty of Chemistry, University of Gdansk, Wita Stwosza 56, Gdansk, 80-308, Poland
[c] BNF - New Bio Force Ltd., Kartuska 420a, 80-125, Gdańsk, Poland

ARTICLE INFO

Keywords:
Phages
UPEC bacteria
UTIs
Phage therapy

ABSTRACT

Phage therapy is a promising alternative for treating urinary tract infections (UTIs), particularly given the rapid emergence of antimicrobial resistance among uropathogenic *Escherichia coli* (UPEC). In this study, we review 98 reported lytic phages infecting UPEC, summarizing their morphology, host range, isolation sources, infection efficiency, progeny yield, adsorption rate, stability across pH and temperature ranges, biofilm-disrupting capacity, and other relevant characteristics. In addition, we conducted detailed bioinformatic and phylogenetic analyses of phage genomic sequences. This multi-faceted analysis highlights the considerable diversity and strong therapeutic potential of anti-UPEC phages, providing a solid scientific basis for their application in phage therapy. However, several limitations remain that hinder clinical translation. We discuss these challenges, drawing on examples from the literature, and outline that comprehensive physiological and genomic characterization, combined with standardized testing methodologies, can address current barriers to the effective use of anti-UPEC phages.

1. Introduction

Uropathogenic *Escherichia coli* (UPEC), responsible for urinary tract infections (UTIs), have developed resistance to common antibiotics like penicillins, cephalosporins, and fluoroquinolones (Carlet et al., 2011; Zowawi et al., 2015). UPECs are responsible for 50–65 % of complicated and 75–85 % of uncomplicated UTIs (Zalewska-Piątek and Piątek, 2020). Complicated infections, especially those related to catheterization, are challenging to treat due to bacterial biofilm formation, presence of persister cells in this structure, and more common antibiotic resistance in the population. Prolonged catheterization increases infection risk by 5 % per day, complicating recovery and reducing the quality of life (Hariati et al., 2018, 2019; Sabih and Leslie, 2023; Pallett and Hand, 2010). Antibiotic resistance is a growing crisis affecting multiple sectors, including medicine and industry. This problem also affects UPEC bacteria, and the main reason for the increasing number of multidrug-resistant strains is the incorrect use of antibiotics. In the treatment of complicated UTIs, the first-choice antibiotic is amoxicillin-clavulanic acid (Whelan et al., 2023), to which approximately 5.3 %, 13.9 %, 30 %, and 37.6 % of strains in Germany, Poland, England, and France, respectively, are estimated to be resistant (Kot, 2019; Abduzaimovic et al., 2016; Kot et al., 2016; Abernethy et al., 2017; Lavigne et al., 2016). In developing countries, these percentages are even higher and reach up to 43 % in Saudi Arabia (Alameer et al., 2025) In Europe, an increased number of strains resistant to second- and third-generation cephalosporins is also being reported. For the former, the percentage of drug-resistant strains ranged from 5.5 % in Belgium to 16.6 % in Spain. Ciontea et al., in 2015–2017 studied UPEC isolates from patients in Romania, where it turned out that as many as 87 % of isolates were resistant to third-generation cephalosporins and 71 % to amoxicillin-clavulanic acid (Ciontea et al., 2018). In 2025, Hejazi Dehaghani et al., showed that 50 % of UPEC clinical isolates were resistant to ceftazidime, 45 % to cefixime, and 20 % to cefepime Hejazi Dehaghani et al. (2025). In cases involving other groups of antibiotics,

* Corresponding author. Department of Molecular Biology, Faculty of Biology, University of Gdańsk, Wita Stwosza 59, 80-308, Gdansk, Poland.
E-mail addresses: aleksandra.lukasiak@phdstud.ug.edu.pl (A. Łukasiak), wojciech.wesolowski@phdstud.ug.edu.pl (W. Wesołowski), julia.brzoza@gmail.com (J. Neumann), natalia.lewandowska@phdstud.ug.edu.pl (N. Lewandowska), e.weglinska.166@studms.ug.edu.pl (E. Węglińska), sylwia.bloch@ug.edu.pl (S. Bloch), grzegorz.wegrzyn@ug.edu.pl (G. Węgrzyn), bozena.nejman-falenczyk@ug.edu.pl (B. Nejman-Faleńczyk).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.virol.2025.110738
Received 17 August 2025; Received in revised form 12 October 2025; Accepted 10 November 2025
Available online 11 November 2025
0042-6822/© 2025 Published by Elsevier Inc.

we can observe a similar trend, which has led to increased interest in alternative therapies against UPEC bacteria (Kot, 2019). In this light, phage therapy, using bacterial viruses called bacteriophages (shortly phages) alone (Necel et al., 2025; Topka-Bielecka et al., 2021a,b) or in combination with other agents (Devi et al., 2024; Necel et al., 2022), offers a promising alternative to antibiotics. Recent reports of our team and other scientists showed that combining phage and enzymatic therapies may be more effective than antibiotics (Wesołowski et al., 2025; Lu et al., 2023; Topka-Bielecka et al., 2021a,b). Although the number of studies describing phages has increased recently, the lack of standardized actions in this area and the still not enough number of well-characterized phages remain serious limitations. The available reports differ significantly in the scope of analyzed phage properties and methods used for this purpose, which makes their comparison difficult. This article reviews previous efforts to isolate phages targeting UPEC and compares their morphological, physiological, and sequence features.

2. Host spectrum and origin of phages targeting UPEC strains

Establishing the host range of a newly isolated phage is essential for evaluating its effectiveness in phage therapy. The phage should infect a specific pathogen without harming the natural microbiota. Additionally, it must efficiently eliminate the pathogenic bacterium. This allows to effectively constrain the bacterial cell resistance development against the phage, expediting the treatment process. Therefore, determining the efficiency of plating (EOP) has become more important alongside the classical spot test for host range assessment.

The analysis carried out in this study (Table 1) shows that the EOP assay is not typically performed when determining the range of UPEC bacteria sensitive to the tested phage. The host range for a newly isolated from Ganges river phage called myPSH1131 was examined. Out of the 53 carbapenem-resistant *E. coli* isolates that were tested and belonged to different pathotypes, such as enteropathogenic *E. coli* (EPEC), enterohemorrhagic *E. coli* (EHEC), enterotoxigenic *E. coli* (ETEC), enteroaggregative *E. coli* (EAEC), and one unknown pathotype; 31 of them were found to be infected by the phage using spot test (Manohar et al., 2018). However, using the full plate double-layer agar method, this number decreased to 29. In turn, among tested UPEC strains, infection occurred in 9 out of 12 strains (75 %), including 3 at a highly efficient level. What is interesting is that using the spot method, the phage myPSH1131 was able to infect UPEC strain PSH399, while no phage plaques were observed on double-layer agar plates obtained during the EOP assay (Manohar et al., 2018). Another study examined the ability of the phage vB_EcoS-Golestan to attack UPEC isolates obtained from patients suffering from UTI (Yazdi et al., 2020). It is worth noting that among the 52 UPEC strains, 41 were identified as MDR (multi-drug resistant) strains. Of the 52 UPEC clinical isolates, the phage infected 28 (53.8 %), including 23 (56 %) MDR strains, but the efficiency level of these infections was not determined. Additionally, the activity of vB_EcoS-Golestan phage against other bacterial species such as *Acinetobacter baumannii* ATCC19606, *Staphylococcus aureus* ATCC 43300, *Pseudomonas aeruginosa* ATCC 27853, *P. aeruginosa* ATCC 47085, *S. epidermidis* PTCC 1435: (*S. epidermidis* ATCC 14990), *Klebsiella pneumoniae* ATCC 700603, *K. oxytoca* PTCC 1402: (*K. oxytoca* ATCC 8724), *S. saprophyticus* ATCC 15305 was tested, but infection did not occur. An exception was the reference strain *E. coli* ATCC 25922 against which the phage vB_EcoS-Golestan turned out to be active (Yazdi et al., 2020).

A phage named vB_EcoP-EG1 was able to infect 10 out of the 21 (47.6 %) UPEC strains isolated from human urine and also 7 laboratory strains, such as *E. coli* MG1655, TOP10, S17-1, DH5α, BL21(DE3), JM110 and BW25113 (Gu et al., 2019). However, the efficiency of these infections was not determined (Gu et al., 2019). Phage Killian, which infected 6 out of 9 (66,7 %) UPEC strains, among which 3 strains were clinical isolates, was isolated from sewage, however, infection was determined as „inefficient" in four cases (Khunti et al., 2023). Furthermore, Killian efficiently infected *E. coli* laboratory strains ATCC 25922 and MC4100, as well as the probiotic strain *E. coli* Nissle 1917, which is important information from the perspective of developing a phage therapy safe for natural microbiota (Khunti et al., 2023). Recently, the effectiveness of four newly discovered phages from wastewater in fighting 53 clinical UPEC strains, including 16 MDR strains, was investigated (Malik et al., 2021). Phages named FS11, FS17, PS6, and PS8 infected 28 (52.8 %), 35 (66 %), 28 (52.8 %), and 26 (49 %) strains, respectively. It is important to emphasize that the cocktail composed of these phages exhibited a much broader spectrum of activity, lysing 86.7 % of UPEC strains, including 87.5 % of MDR strains (Malik et al., 2021). The broad host range was also demonstrated by phage KEP10, obtained from environmental water, infecting as many as 27 out of 41 (65.8 %) UPEC strains isolated from urine samples of hospitalized patients, but the effectiveness of these infections was not determined (Nishikawa et al., 2008).

The activity of newly isolated phages against *E. coli* strains obtained from stool samples, patients with acute UTI, and patients with recurrent UTI was investigated (González-Villalobos et al., 2021). Phages ϕEc1, ϕEc3, and ϕEc4, isolated from wastewater, exhibited lytic activity against strains from all tested groups. However, in the case of UPEC, infection occurred in 25 %, 18 %, and 12 % of all strains, respectively. The phage activity was also tested against other species such as reference strains of *Salmonella enterica* subsp. *enterica* serovar Typhi ATCC 6539, *S. enterica* subsp. *enterica* serovar Choleraesuis ATCC 10708, *S. enterica* subsp. *enterica* serovar Typhimurium ATCC 14028, *P. aeruginosa* ATCC 9027 and *Yersinia enterocolitica*, *K. pneumoniae*, *Shigella boydii*, and *S. aureus* clinical isolates, but no phage activity was observed (González-Villalobos et al., 2021). In turn, wastewater-isolated phages vKMB22 and vKMB26 were able to infect 18 (45 %) and 33 (82.5 %) out of 40 tested UPEC strains, respectively, with phage vKMB26 efficiently infecting 22 of them. Interestingly, a cocktail composed of these phages showed similar activity to that of vKMB26 phage alone. Additionally, even when diluted 100-fold, the phages from the cocktail still infected over half of the UPEC strains (57 %) (Slobodníková et al., 2021).

In another study (Vera-Mansilla et al., 2022), phages MLP1, MLP2, and MLP3, isolated from the Valdivia River, showed lytic activity against reference strains CFT073, CFT073 Δ*waaL*, UTI-89 and isolates from patients with UTI. Phage MLP1 has been found to infect only 2 out of the 8 (25 %) UPEC strains analyzed, indicating a relatively narrow range of activity. In contrast, phages MLP2 and MLP3, despite belonging to different families, demonstrated a broader and comparable range of activity, infecting 7 out of 8 (87.5 %) strains both. That study aimed to determine the plating efficiency of the examined phages. However, the results showed that phage MLP1 infected the clinical strain with significantly lower efficiency compared to the reference strain. Meanwhile, phages MLP2 and MLP3, despite having a similar host profile, exhibited differences in the degree of infection efficiency for these strains. Interestingly, the study included the *E. coli* strain Δ*waaL*, which lacks a gene for the O-antigen ligase, resulting in preventing LPS production. Interestingly, phage MLP1 was able to infect the CFT073 strain efficiently, however, the infection of the mutant strain decreased by about 4 orders of magnitude. This suggests that the O-antigen in the LPS may act as a receptor for phage MLP1. On the other hand, phage MLP2 infected the mutant with the same efficiency as the wild-type strain, indicating that phage MLP2 recognizes a different receptor (Vera-Mansilla et al., 2022).

The ability of sewage-isolated phages to combat biofilm formed by UPEC strains was tested (Chibeu et al., 2012). Phage vB_EcoM_ACG-C40, demonstrating the broadest range, infected as many as 28 out of 43 clinical isolates, accounting for 65 %, while the other two phages vB_EcoP_ACG-C91 and vB_EcoS_ACG-M12 exhibited significantly narrower ranges, at 39.5 % and 34.8 %, respectively. It's worth noting that 80.5 % of the tested strains were infected by at least one phage using the spot test, although the efficiency of these infections was not determined (Chibeu et al., 2012). In turn, 28 phages whose plaquing host range

Table 1
The list of reported anti-UPEC phages and their features.

Phage name	Source	Host range	Taxonomy	Length of genome (bp)	Number of ORFs	Proteomics	Virion size head/tail (nm)	Burst size	Adsorption	Cytotoxic effect	Model	Stability (^{0}C)*	Stability (pH)*	Stability (urine)	Accession number	Reference[a]
FU1	River	ND	ND (P)	ND	ND	+	47.2/12.5	ND	ND	ND	Liquid culture	ND	ND	ND	ND	67
FU3	River	ND	ND (P)	ND	ND	+	40/20	ND	ND	ND	Liquid culture	ND	ND	ND	ND	67
vB_EcoS-Golestan	Sewage	28/52	*Guernseyvirinae/ Kagunavirus* (S)	44,829	78	+	35/28	100	65.8 %/5	ND	Liquid culture	37–45	6–9	ND	MG099933	75
vB_EcoP-EG1	Sewage	13/35	*Autographiviridae/ Studiervirinae/ Teseptimavirus* (P)	39,919	51	+	50/10	152	ND	ND	Biofilm	4–45	ND	+	MG488277	24
IBEC 77	Urine	1/1	ND (P)	ND	ND	ND	ND	150	ND	ND	Biofilm *G. mellonella*	ND	ND	ND	ND	62
IBEC 78	Sewage	1/1	ND (P)	ND	ND	ND	ND	ND	ND	ND	Biofilm *G. mellonella*	ND	ND	ND	ND	62
IBEC 98	Urine	1/1	ND (P)	ND	ND	ND	ND	ND	ND	ND	Biofilm *G. mellonella*	ND	ND	ND	ND	62
AL505_P1	Sewage	1/1	ND (S)	ND	ND	ND	ND	ND	ND	ND	Biofilm Mice	ND	ND	ND	ND	20
AL505_P2	Sewage	1/1	ND (M)	ND	ND	ND	ND	ND	ND	ND	Biofilm Mice	ND	ND	ND	ND	20
AL505_P3	Sewage	1/1	ND (P)	ND	ND	ND	ND	ND	ND	ND	Biofilm Mice	ND	ND	ND	ND	20
myPSH1131	River	9/12	*Gordonclarkvirinae/ Kuravirus* (P)	76,163	97	ND	45/95	130	50 %/10	ND	*G. mellonella*	4–45	3–13	ND	MG983840	45
St11Ph5	Sewage	ND	*Schitoviridae/ Enquatrovirinae/ Gamaleyavirus* (P)	72,444	90	ND	ND	ND	ND	ND	ND	ND	ND	ND	MG208881	22
Killian	Sewage	6/9	*Straboviridae/ Tevenvirinae/ Tequatrovirus* (M)	169,905	276	ND	81.2/168	139	90 %/20	ND	Liquid culture	4–40	4–10	ND	OQ446694	32
FS11	Sewage	28/53	ND (M)	ND	ND	ND	72/63.6	121	96 %/25	ND	Liquid culture	0–50	4–12	ND	ND	44
FS17	Sewage	35/53	ND (M)	ND	ND	ND	63.4/109.7	98	96 %/20	ND	Liquid culture	0–50	4–12	ND	ND	44
PS6	Sewage	28/53	ND (S)	ND	ND	ND	46/113	78	96 %/32	ND	Liquid culture	0–50	4–12	ND	ND	44
PS8	Sewage	26/53	ND (S)	ND	ND	ND	60.2/105.3	87	96 %/35	ND	Liquid culture	0–50	4–12	ND	ND	44
590B	Sewage	4/50	*Guernseyvirinae/ Kagunavirus* (S)	44,390	75	ND	55/190	109	95 %/12	ND	Biofilm	−20–50	4–9	ND	MW722821	12
vB_EcoM_LNA1	Sewage	2/3	*Ounavirinae/ Felixounavirus* (M)	88,369	126	ND	ND	134	66.9 %/4	ND	Liquid culture	ND	ND	ND	MZ311863	54
vB_EcoM_LNA2	Sewage	3/3	*Straboviridae/ Tevenvirinae/Mosigvirus* (M)	168,414	267	ND	ND	68	79.7 %/4	ND	Liquid culture	ND	ND	ND	MZ311865	54
vB_EcoM_LNA3	Sewage	2/3	*Autographivirida/ Studiervirinae/ Kayfunavirus* (P)	39,663	53	ND	ND	12	95.7 %/4	ND	Liquid culture	ND	ND	ND	MZ311862	54
vB_EcoM_LNA4	Sewage	1/3	*Vequintavirinae/ Seunavirus* (M)	147,613	249	ND	ND	28	60.5 %/4	ND	Liquid culture	ND	ND	ND	MZ311864	54
vB_EcoM_LNA5	Sewage	2/3	*Autographiviridae/ Studiervirinae/ Kayfunavirus* (P)	39,091	58	ND	ND	100	99.4 %/4	ND	Liquid culture	ND	ND	ND	MZ311861	54

(*continued on next page*)

Table 1 (*continued*)

Phage name	Source	Host range	Taxonomy	Length of genome (bp)	Number of ORFs	Proteomics	Virion size head/tail (nm)	Burst size	Adsorption	Cytotoxic effect	Model	Stability (^{0}C)*	Stability (pH)*	Stability (urine)	Accession number	Reference[a]
vB_EcoM_ SHAK9454	Sewage	8/12	*Autographiviridae/ Studiervirinae/ Kayfunavirus*	39,471	56	ND	ND	ND	ND	ND	ND	ND	ND	ND	OR594184	7
vB_EcoM_ SHAK7704	Sewage	8/12	*Autographiviridae/ Studiervirinae/ Kayfunavirus*	44,446	64	ND	ND	ND	ND	ND	ND	ND	ND	ND	OR594181	7
vB_EcoM_ SHAK7854	Sewage	3/12	*Autographiviridae/ Studiervirinae/ Kayfunavirus*	39,734	57	ND	ND	ND	ND	ND	ND	ND	ND	ND	OR594182	7
vB_EcoM_ SHAK7858	Sewage	9/12	*Autographiviridae/ Studiervirinae/ Kayfunavirus* (S)	44,517	82	ND	ND	ND	ND	ND	ND	ND	ND	ND	OR594183	7
vB_EcoM_ SHAK7163	Sewage	7/12	*Autographiviridae/ Studiervirinae/ Kayfunavirus*	44,656	63	ND	ND	ND	ND	ND	ND	ND	ND	ND	OR594179	7
vB_EcoM_ SHAK7693	Sewage	8/12	*Autographiviridae/ Studiervirinae/ Kayfunavirus* (S)	45,233	83	ND	ND	ND	ND	ND	ND	ND	ND	ND	OR594180	7
KEP10	Environ–mental water	27/41	ND (M)	ND	ND	ND	83/104	ND	ND	ND	Mice	ND	ND	+	AB326953 (major capsid) AB326954 (tail fiber)	55
vB_EcoM- ϕEc1	Sewage	25/ 100	ND (M)	ND	ND	+	76/126	ND	ND	ND	Biofilm	ND	ND	ND	ND	23
vB_EcoM- ϕEc3	Sewage	18/ 100	*Guernseyvirinae/ Kagunavirus* (M)	45,288	79	+	ND	ND	ND	ND	Biofilm	ND	ND	ND	MZ576217	23
vB_EcoM- ϕEc4	Sewage	12/ 100	*Guernseyvirinae/ Kagunavirus* (M)	44,540	74	+	ND	ND	ND	ND	Biofilm	ND	ND	ND	MZ576218	23
MLP1	River	2/8	*Chaseviridae/ Cleopatravirinae*	51,940	72	ND	66/135	ND	ND	ND	Liquid culture	ND	ND	ND	OK148439	72
MLP2	River	7/8	*Straboviridae/ Tevenvirinae/ Tequatrovirus* (M)	168,787	269	ND	63/73	ND	ND	ND	Liquid culture	ND	ND	ND	OK148441	72
MLP3	River	7/8	*Gordonclarkvirinae/ Kuravirus* (P)	76,234	119	ND	48/11	ND	ND	ND	Liquid culture	ND	ND	ND	OK148440	72
vB_EcoM_ KMB22	Sewage	18/40	*Straboviridae/ Tequatrovirus* (M)	ND	ND	ND	ND	ND	ND	ND	Liquid culture	ND	ND	ND	ND	65
vB_EcoM_ KMB26	Sewage	33/40	*Straboviridae/ Tequatrovirus* (M)	168,222	266	ND	ND	ND	ND	ND	Liquid culture	ND	ND	ND	OW536688	65
vB_EcoP_ ACG-C91	Sewage	17/43	*Autographiviridae/ Molineuxvirinae/ Vectrevirus* (P)	43,731	55	65-68/12	ND	ND	ND	ND	Biofilm	ND	ND	ND	JN986844	15
vB_EcoM_ ACG-C40	Sewage	28/43	*Straboviridae/ Tevenvirinae/ Tequatrovirus* (M)	167,396	291	110/114	ND	ND	ND	ND	Biofilm	ND	ND	ND	JN986846	15
vB_EcoS_ ACG-M12	Sewage	15/43	*Drexlerviridae/ Braunvirinae/Guelphvirus* (S)	46,054	78	157/172	ND	ND	ND	ND	Biofilm	ND	ND	ND	JN986845	15
NTEC3	Sewage	9/45	*Guernseyvirinae/ Kagunavirus* (S)	44,240	72	ND	ND	ND	ND	ND	ND	ND	ND	ND	OK539620	14
CHD16UKE1	Sewage	33/79	*Straboviridae/ Tevenvirinae/ Tequatrovirus* (M)	168,543	300	ND	ND	ND	ND	ND	Liquid culture	ND	ND	ND	MZ234030	40

(*continued on next page*)

Table 1 (*continued*)

Phage name	Source	Host range	Taxonomy	Length of genome (bp)	Number of ORFs	Proteomics	Virion size head/tail (nm)	Burst size	Adsorption	Cytotoxic effect	Model	Stability (^{0}C)*	Stability (pH)*	Stability (urine)	Accession number	Reference[a]
CHD94UKE2	Sewage	23/79	*Straboviridae/ Tevenvirinae/ Tequatrovirus* (M)	167,922	300	ND	ND	ND	ND	ND	Liquid culture	ND	ND	ND	MZ234031	40
G3G7	Sewage	ND	*Straboviridae/ Tevenvirinae/ Tequatrovirus* (M)	168,649	300	ND	ND	ND	ND	ND	Liquid culture	ND	ND	ND	MZ234040	40
101117BS1	Sewage	18/79	*Straboviridae/ Tevenvirinae/ Tequatrovirus* (M)	167,080	300	ND	ND	ND	ND	ND	Liquid culture	ND	ND	ND	MZ234018	40
22664UKE3-2	Sewage	48/79	*Autographiviridae/ Studiervirinae/ Kayfunavirus*	40,482	58	ND	ND	ND	ND	ND	ND	ND	ND	ND	MZ234011	40
101120B2	Sewage	27/79	*Autographiviridae/ Studiervirinae/ Kayfunavirus*	39,899	59	ND	ND	ND	ND	ND	ND	ND	ND	ND	MZ234023	40
101118B1	Sewage	40/79	*Autographiviridae/ Molineuxvirinae/ Vectrevirus*	44,526	59	ND	ND	ND	ND	ND	ND	ND	ND	ND	MZ234020	40
101101UKE1	Sewage	15/79	*Autographiviridae/ Molineuxvirinae/ Vectrevirus*	44,450	61	ND	ND	ND	ND	ND	ND	ND	ND	ND	MZ234012	40
UTI89UKE2	Sewage	2/79	*Autographiviridae/ Molineuxvirinae/ Vectrevirus*	44,293	66	ND	ND	ND	ND	ND	ND	ND	ND	ND	MZ234049	40
UTI89UKE3	Sewage	3/79	*Autographiviridae/ Molineuxvirinae/ Vectrevirus*	44,294	66	ND	ND	ND	ND	ND	ND	ND	ND	ND	MZ234050	40
101117UKE2	Sewage	6/79	*Autographiviridae/ Molineuxvirinae/ Vectrevirus*	44,526	58	ND	ND	ND	ND	ND	ND	ND	ND	ND	MZ234019	40
101118UKE1	Sewage	32/79	*Autographiviridae/ Studiervirinae/ Kayfunavirus*	40,233	61	ND	ND	ND	ND	ND	ND	ND	ND	ND	MZ234021	40
101120B1-2	Sewage	15/79	*Autographiviridae/ Studiervirinae/ Kayfunavirus*	39,845	59	ND	ND	ND	ND	ND	ND	ND	ND	ND	MZ234022	40
101136BS1	Sewage	35/79	*Autographiviridae/ Studiervirinae/ Kayfunavirus*	39,375	59	ND	ND	ND	ND	ND	ND	ND	ND	ND	MZ234024	40
22664BS1	Sewage	33/79	*Autographiviridae/ Studiervirinae/ Kayfunavirus*	39,133	58	ND	ND	ND	ND	ND	ND	ND	ND	ND	MZ234009	40
172859UKE1	Sewage	3/79	*Straboviridae/ Tevenvirinae/Mosigvirus* (M)	168,667	280	ND	ND	ND	ND	ND	ND	ND	ND	ND	MZ234025	40
101112UKE3-1	Sewage	29/79	*Straboviridae/ Tevenvirinae/ Tequatrovirus* (M)	169,555	301	ND	ND	ND	ND	ND	ND	ND	ND	ND	MZ234013	40
CHD2BS1	Sewage	10/79	*Straboviridae/ Tevenvirinae/ Tequatrovirus* (M)	168,577	299	ND	ND	ND	ND	ND	ND	ND	ND	ND	MZ234027	40
101114BS3	Sewage	42/79	*Gordonclarkvirinae/ Kuravirus* (P)	75,747	143	ND	ND	ND	ND	ND	ND	ND	ND	ND	MZ234015	40

(*continued on next page*)

Table 1 (*continued*)

Phage name	Source	Host range	Taxonomy	Length of genome (bp)	Number of ORFs	Proteomics	Virion size head/tail (nm)	Burst size	Adsorption	Cytotoxic effect	Model	Stability (^{0}C)*	Stability (pH)*	Stability (urine)	Accession number	Reference[a]
101114UKE3	Sewage	25/79	*Gordonclarkvirinae/ Suseptimavirus* (P)	75,747	145	ND	ND	ND	ND	ND	ND	ND	ND	ND	MZ234017	40
CHD5UKE1	Sewage	1/79	*Gordonclarkvirinae/ Kuravirus* (P)	77,359	152	ND	ND	ND	ND	ND	ND	ND	ND	ND	MZ234028	40
101114B2	Sewage	30/79	*Dhillonvirus* (S)	44,971	69	ND	ND	ND	ND	ND	ND	ND	ND	ND	MZ234014	40
101114BS4	Sewage	21/79	*Dhillonvirus* (S)	45,251	70	ND	ND	ND	ND	ND	ND	ND	ND	ND	MZ234016	40
22664B1	Sewage	32/79	*Dhillonvirus* (S)	45,019	70	ND	ND	ND	ND	ND	ND	ND	ND	ND	MZ234008	40
CHD2B1	Sewage	12/79	*Dhillonvirus* (S)	45,144	68	ND	ND	ND	ND	ND	ND	ND	ND	ND	MZ234026	40
CHD5UKE2	Sewage	19/79	*Dhillonvirus* (S)	45,243	68	ND	ND	ND	ND	ND	ND	ND	ND	ND	MZ234029	40
UTI89UKE1	Sewage	1/79	*Guernseyvirinae/ Kagunavirus* (P)	41,265	73	ND	ND	ND	ND	ND	ND	ND	ND	ND	MZ234048	40
22664BS2	Sewage	34/79	*Guernseyvirinae/ Kagunavirus* (S)	45,176	85	ND	ND	ND	ND	ND	ND	ND	ND	ND	MZ234010	40
vB-EcoP_ KMB47	Sewage	17/77	*Autographiviridae/ Molineuxvirinae/ Vectrevirus*	44,488	54	ND	ND	211	98 %/5	ND	Liquid culture	ND	ND	ND	OR424385	46
vB-EcoP_ KMB14	Sewage	13/77	*Autographiviridae/ Molineuxvirinae/ Vectrevirus*	44,563	53	ND	ND	190	98 %/5	ND	Liquid culture	ND	ND	ND	OR424384	46
vB-EcoS_ KMB25	Sewage	4/77	*Drexlerviridae/ Tempevirinae/Tlsvirus*	50,475	87	ND	ND	ND	ND	ND	ND	ND	ND	ND	OR424386	46
vB-EcoS_ KMB46	Sewage	8/77	*Drexlerviridae/ Tempevirinae/ Warwickvirus*	50,953	96	ND	ND	ND	ND	ND	ND	ND	ND	ND	OR525699	46
vB-EcoM_ KMB37	Sewage	18/77	*Stephanstirmvirinae/ Justusliebigvirus*	148,640	305	ND	ND	ND	ND	ND	ND	ND	ND	ND	OR525700	46
vB-EcoM_ KMB43	Sewage	33/77	*Straboviridae/Krischvirus*	164,476	264	ND	ND	38	80 %/5	ND	Liquid culture	ND	ND	ND	OR525698	46
vB-EcoM_ KMB38	Sewage	13/77	*Straboviridae/ Tevenvirinae/Mosigvirus*	169,640	270	ND	ND	ND	ND	ND	ND	ND	ND	ND	OR525701	46
vB-EcoM_ KMB22	Sewage	26/77	*Straboviridae/ Tevenvirinae/ Tequatrovirus*	167,010	276	ND	ND	ND	ND	ND	ND	ND	ND	ND	OR539217	46
vB-EcoM_ KMB23	Sewage	25/77	*Straboviridae/ Tevenvirinae/ Tequatrovirus*	166,939	291	ND	ND	95	72 %/5	ND	Liquid culture	ND	ND	ND	OR525702	46
vB-EcoM_ KMB26	Sewage	45/77	*Straboviridae/ Tevenvirinae/ Tequatrovirus*	168,222	275	ND	ND	47	81 %/5	ND	Liquid culture	ND	ND	ND	OR539218	46
vB-EcoM_ KMB39	Sewage	21/77	*Straboviridae/ Tevenvirinae/ Tequatrovirus*	168,887	275	ND	ND	ND	ND	ND	ND	ND	ND	ND	OR539219	46
vB-EcoM_ KMB40	Sewage	28/77	*Straboviridae/ Tevenvirinae/ Tequatrovirus*	166,966	275	ND	ND	ND	ND	ND	ND	ND	ND	ND	OR539220	46
vB-EcoM_ KMB42	Sewage	29/77	*Straboviridae/ Tevenvirinae/ Tequatrovirus*	165,962	270	ND	ND	69	77 %/5	ND	Liquid culture	ND	ND	ND	OR539221	46
vB-EcoP_ KMB41	Sewage	4/77	*Murrayvirus*	41,822	62	ND	ND	ND	ND	ND	ND	ND	ND	ND	OR539216	46
vB-EcoS_ KMB35	Sewage	6/77	*Guernseyvirinae/ Kagunavirus*	43,891	82	ND	ND	ND	ND	ND	ND	ND	ND	ND	OR525704	46

(*continued on next page*)

Table 1 (*continued*)

Phage name	Source	Host range	Taxonomy	Length of genome (bp)	Number of ORFs	Proteomics	Virion size head/tail (nm)	Burst size	Adsorption	Cytotoxic effect	Model	Stability (^{0}C)*	Stability (pH)*	Stability (urine)	Accession number	Reference[a]
vB-Eco-KMB36	Sewage	2/77	*Guernseyvirinae/ Kagunavirus*	44,710	77	ND	ND	ND	ND	ND	ND	ND	ND	ND	OR525703	46
vB_Ec_ ZCEC14	Sewage	11/25	*Straboviridae/ Tevenvirinae/ Tequatrovirus*	165,162	265	ND	130/120	ND	ND	ND	ND	−20–60	7–11	ND	OP168898	28
Mimir124	River	1/8	*Schitoviridae/ Enquatrovirinae/ Gamaleyavirus* (P)	72,768	101	ND	70/ND	ND	100 %/3	ND	ND	ND	ND	ND	PQ310662	21
vB_EcoS_ UTEC10	Sewage	16/30	*Guernseyvirinae/ Kagunavirus*	44,772	73	ND	66.3/102.2	198	ND	ND	Biofilm Mice	5–50	6–10	ND	PP291582	59
FS2B	Sewage	37/53	*Gordonclarkvirinae/ Kuravirus*	77,407	118	ND	98.2/65.2	137	90 %/12	ND	Liquid culture	0–60	4–12	ND	MW929182 MW929183	43
vB_EcoA_ RDN8.1	Sewage	5/49	*Autographiviridae/ Studiervirinae/ Kayfunavirus*	39,508	49	ND	79/20	250	ND	ND	Liquid culture, Biofilm	−20–37	5–9	ND	MN414250	13
øEc_Makalu_001	Sewage	12/45	*Straboviridae/Krischvirus* (M)	163,752	284	ND	88.9/109	127	ND	ND	Liquid culture	25–37	6–9	ND	MN894885	18
øEc_Makalu_002	Sewage	14/45	*Straboviridae/Krischvirus* (M)	164,674	285	ND	88.5/9.4	74	ND	ND	Liquid culture	25–37	6–9	ND	MN709127	18
øEc_Makalu_003	Sewage	14/45	*Straboviridae/Krischvirus* (M)	162,966	284	ND	83.7/94.3	120	ND	ND	Liquid culture	25–37	6–9	ND	MN882349	18
Escherichia phage	Urine	6/21	ND (M)	ND	ND	ND	39/ 105.317	26	ND	ND	Liquid culture	37	7–8	ND	ND	17
UGKSEcP1	Sewage	9/11	*Straboviridae/ Tevenvirinae/ Tequatrovirus* (M)	167,433	277	ND	66/116	+	99.3 %/10	ND	Liquid culture	−20–35	7–10	ND	OV877085	66
UGKSEcP2	Sewage	9/11	*Straboviridae/ Tevenvirinae/ Tequatrovirus* (M)	167221	276	ND	66/116	+	99.3 %/12	ND	Liquid culture	−20–35	7–10	ND	OV876900	66
ES10	Sewage	ND	*Guernseyvirinae/ Kagunavirus*	48,315	74	ND	95/200	90	ND	ND	Liquid culture	0–35	5–7	–	OR466728	53

Please note: The host range was determined by spot test and provided as the number of strains sensitive to the phage/total UPEC strains tested. Underlined values indicate that an additional efficiency of plating (**EOP**) analysis was performed. Taxonomic units were provided based on the authors' data. Letters in brackets indicate *Myoviridae* (M), *Podoviridae* (P), and *Siphoviridae* (S) phage families, according to the previous nomenclature system. The adsorption results were provided as the percentage of phage particles adsorbed to bacterial cells/minutes. The "+" sign indicates that the experiment was conducted, however, the results were obtained with different methods and they are not provided in the Table. Each row of the table contains descriptions of phages from one publication. The abbreviation ND means that the data were not obtained by the authors and once written in a row it applies to all phages from the row unless otherwise stated. The table preserves the original names of phages given and used by the authors. As reported by the authors, phages vB_EcoM_KMB22 and vB_EcoM_KMB26 were analyzed twice (Slobodníková et al., 2021; Markusková et al., 2024). Due to different methodologies, we included results from both publications in the table. *Stability ranges are presented as proposed by the authors, however, due to different incubation times and interpretation approaches, it is difficult to compare them.

[a] The reference numbers correspond to the respective publications in the references section.

ranged from 61 % to 1 % were isolated. The infection efficiency was analyzed only for 7 phages (Loose et al., 2021). In another study, Ngiam et al. (2021) examined the host range for five phages - vB_EcoM_LNA1, vB_EcoM_LNA2, vB_EcoP_LNA3, vB_EcoM_LNA4, vB_EcoP_LNA5 using the spot test. It was found that each phage infects at least two out of three UPEC strains, with phage vB_EcoM_A2 infecting all tested strains. However, the results of experiments aimed at determining the infection efficiency indicated that phage vB_EcoM_LNA2, which formed lysis zones on the lawns of all tested strains in the spot test, did not efficiently infect any of the investigated UPEC isolates. These results suggested that all phages, except vB_EcoM_LNA2, demonstrated infectivity beyond their original hosts, and effectively infected some UPEC strains, implying that the creation of phage cocktails may be required in combating UTIs. These findings also suggest that the spot test is not always an adequate indicator of whether a particular strain falls within the host range of a tested phage (Ngiam et al., 2021).

Phage vB_Ec_ZCEC14, isolated from sewage in Egypt, produced a lysis zone against 11 out of 25 UPEC isolates (44 %), and 3 of them were infected with high efficiency. In turn, phage vB_EcoS_UTEC10, isolated from sewage outflow of the Zagazig University Hospital, was reported to show lytic activity against 53 % (n = 16) of the tested isolates in a spot test, and high efficiency of plating (≥0.5) against 46.7 % (n = 14). Mimir124 is another phage isolated on the UPEC host from river water (Golomidova et al., 2024). This phage was tested on 7 *E. coli* strains, however, no strain was found to be susceptible to Mimir124 other than its original host (Golomidova et al., 2024). Contrary to this, phage FS2B showed high lytic activity and in the spot test it exhibited a broad host range, infecting 69.8 % (n = 37) of the collected UPEC isolates, among them 64.8 % (n = 24) were identified as MDR (Malik et al., 2024). In turn, 6 phages isolated from sewage collected from Brisbane (named vB_EcoM_SHAK7704, vB_EcoM_SHAK7854, vB_EcoM_SHAK7858, vB_EcoM_SHAK7163, vB_EcoM_SHAK9454, and vB_EcoM_SHAK7693) infected from 2 to 6 UPEC strains out of 12 in total. Interestingly, when used in cocktails, these phages showed synergistic improvement against a mixed UPEC population (Asgharzadeh Kangachar et al., 2024). Similarly, six phages were characterized in detail and used in a cocktail (Markusková et al., 2024). These phages were able to infect from 2 to 44 out of 77 clinical strains (3–57 %). Among them, phages vKMB26 and vKMB43 showed the broadest host range, and lysed 57 % and 43 % of the strains, respectively (Markusková et al., 2024).

Reports describing UPEC-infecting phages FU1 and FU3, isolated from the Cisdane River, as well as the phage St11Ph5 were published, however, the host range for these phages has not been determined (Sufa et al., 2018; Golomidova et al., 2018). Similarly, the effectiveness of a cocktail composed of 25 different phages against UPEC bacteria from patients which caused UTI in mice was investigated, however, the host range of these phages remained unspecified (Mijbel Ali et al., 2021). According to another report, phage 590B has the narrowest host range among all the above-described phages, infecting only 8 % (4 out of 50 tested) of UPEC strains (Chaudhary et al., 2022a). The MDR *E. coli* 590 strain was used for the isolation of not only phage 590B but also phage NTEC3 which, compared to 590B, had a broader host spectrum and was able to infect 9 out of 45 UPEC strains tested in that study (Chaudhary et al., 2022c). During the next study, three phages (IBEC 77, IBEC 78, and IBEC 98) were isolated from sewage and urine, and tested against three different UPEC strains: CFT073, MG1655, and LF82 (Sanmukh et al., 2023). Nevertheless, none of these phages showed specificity against the other two strains. In addition, 18 more were isolated during that study, but none of them were found to attack the UPEC strain (Sanmukh et al., 2023). Using the clinical strains UPEC AL505 - AL505-P1, AL505-P2, and AL505-P3, newly isolated phages were obtained from wastewater (Galtier et al., 2016). However, that article did not include information about the host range for these phages. In another study, phages from sewage attacked 100 % of clinical strains gained from UTI hospitalized patients. However, individual phages from that mix were not isolated and described (Rahimzadeh et al., 2020).

As presented above, many phages infecting UPEC exhibit a broad host range among these bacteria (infecting from 8 % to 100 % of the tested strains). The average range was around 53.5 % using the spot test. Importantly, there are discrepancies in the results of spot tests and EOP, which may indicate that the spot test is not always a reliable indicator of whether a phage will be able to effectively kill pathogenic bacteria during therapy.

3. Stability of phages infecting UPEC strains

As phages are composed of nucleic acid and a protein capsid, they are highly susceptible to external factors. This makes evaluating their stability in various physical and chemical conditions crucial before using them in phage therapy. Not all phages are suitable for use in the human body's complex environment. Therefore, selecting phages that not only effectively eliminate bacteria but also remain stable is imperative. The stability of virions is also crucial for phage cocktail long-term storage, which requires low temperatures and optimal pH. The results obtained by the authors are described below according to their interpretations, although the large variation in the adopted methodology makes it difficult to compare them and draw more general conclusions.

The activity of phage vB_EcoS_Golestan was tested by incubating it at different temperatures (from 45 to 75 °C) for an hour (Yazdi et al., 2020). The phage was stable at 37 °C (control) and 45 °C. However, its activity gradually decreased at higher temperatures, and the phage became completely inactive at 75 °C (Yazdi et al., 2020). Phage vB_EcoP-EG1 was tested during a 1 h incubation at temperatures from 4 to 65 °C, and it was stable at a temperature range between 4 and 45 °C (Gu et al., 2019). When the temperature was increased to 50 °C, only 50 % of phage particles remained active (Gu et al., 2019). Both myPSH1131 and Killian phages remained stable at temperature ranges 4°C–45 °C and 4°C–40 °C, respectively, but both were inactivated rapidly at temperatures above 50 °C (Manohar et al., 2018; Khunti et al., 2023).

The viability of phages FS11, FS17, PS6, and PS8 did not fall below 80 % at the temperature range of 0–50 °C, and their stability at lower temperatures was also examined in the context of their storage (Malik et al., 2021). Phage lysates mixed with glycerol and stored at temperatures of −70 °C and −20 °C remained virtually unchanged for 3 months. However, the titers of lysates deprived of glycerol decreased rapidly after 3 months of storage at room temperature, while those of lysates stored at 4 °C did not change (Malik et al., 2021). Phage 590B also showed stability at both −20 °C and 45 °C. However, the survival rate of phage particles began to decline above 60 °C, as in the previous cases (Chaudhary et al., 2022a). Similarly, phage FS2B was quite stable from 0 to 60 °C and almost completely lost its viability above 80 °C (Malik et al., 2024). Phage vB_EcoS_UTEC10 was also stable over the temperature range 5–50 °C, and higher temperatures resulted in a gradual drop in phage titer (Rajab et al., 2024). Contrary to them, phage vB_Ec_ZCEC14 showed quite high thermal tolerance at 80 °C, as after an hour of incubation the number of phage particles decreased from 10^8 to 10^4 (Ismael et al., 2024).

The low survival rate of phages at high temperatures is due to their protein structure. Proteins are sensitive molecules and easily denature at high temperatures, however, this is not so important in the context of phage therapy. On the other hand, phage lysates must be stable at low temperatures, due to the need for their long-term storage. It is also important that phage virion survival is high at temperatures oscillating within the temperature range of the human body, ranging from 36 °C to 40 °C.

From all available studies examining the influence of pH value on the survival of phages, all tested viruses survived in solutions with pH values similar to human urine, ranging from 4.5 to 8.0 (Manohar et al., 2018; Malik et al., 2021; Chaudhary et al., 2022a; Khunti et al., 2023; Ismael et al., 2024), except phages vB_EcoS_Golestan and vB_EcoS_UTEC10, whose titers dropped at pH = 5 (Yazdi et al., 2020; Rajab et al., 2024).

All phages survived at slightly acidic pH but were completely inactivated at low pH (2.0). From all analyzed reports about phages infecting UPEC strains, where the stability of phage particles in solutions of different pH was tested, the phage myPSH1131 showed high survival in the widest pH spectrum, from 3 to 13 (Manohar et al., 2018). Our attention was also attracted by studies in which researchers presented the results of the survival of phages in human and mouse urine. Phages vB_EcoP-EG1 and KEP10 showed high stability after 24 h incubation in human urine (Nishikawa et al., 2008; Gu et al., 2019). Phage KEP10 was also incubated in mouse urine, and its titer did not change significantly (Nishikawa et al., 2008). Checking the stability of phages in urine (which is rich in various compounds such as urea, creatinine, uric acid, chloride, sodium, potassium, sulphate, ammonium, phosphate, and other ions and molecules) seems to be important from the perspective of using phage therapy against UTIs. Unfortunately, only a few such tests, mentioned above, have been reported so far.

4. Kinetics of development of phages targeting UPEC bacteria: pace and dynamics of viral infection

It is essential to ensure that phages in a phage cocktail act rapidly. This helps to minimize the risk of the development of bacterial cell resistance to the phages. It is also important because if the phages persist for too long time in the human body, the patient's immune system may develop anti-phage antibodies that could inactivate virions. Therefore, the phages used in phage therapy should quickly adsorb to the surface of the host cells and have a high burst size, thereby increasing the number of active phage particles against the pathogen. Moreover, they should have a short latency period to rapidly eliminate the bacterial cells and release progeny particles.

It was demonstrated that the phage vB_EcoS_Golestan lysed the culture of the uropathogenic strain *E. coli* 333 at various multiplicity of infection (MOI) values, reducing the bacterial cell count by several orders of magnitude (Yazdi et al., 2020). At MOI of 10, a rapid decrease in the quantity of bacterial cells from 10^8 to 10^7 CFU/ml was observed after 1 h. When testing MOI values of 0.1, 1, and 10, the cell count dropped to the order of 10^4 CFU/ml after 3 h from adding the lysate to the bacterial culture. The same cell count was achieved through phage infection at an MOI of 0.01 after 3 h of incubation. However, after 24 h of incubation with phage, the bacterial population renewed, although it did not reach the initial level. That study monitored both the number of bacterial cells (CFU/ml) and optical density measurements (OD_{600}), and these measurements were compatible. Furthermore, the phage vB_EcoS_Golestan titer corresponded to ~100 phage particles per cell, and the latent period lasted 40 min. Moreover, 65.8 % of phage particles adsorbed to bacterial cells within 5 min of introducing the lysate into the culture. Calcium and magnesium ions positively influenced the phage vB_EcoS_Golestan adsorption process. However, 15 min after the start of the infection, the quantity of adsorbed phages was nearly the same in both ion-depleted and ion-enriched cultures (Yazdi et al., 2020).

Phage vB_EcoP-EG1 releases about 150 phage particles from one infected *E. coli* MG1655 cell (Gu et al., 2019). Interestingly, it has a very short latent period, which took only 5 min. The culture of the *E. coli* MG1655 strain was also treated with phage lysate at various MOI values, including 0.1, 1, and 10. A rapid decrease in optical density values was observed after 1 h of incubation at each MOI variant. The study was conducted for 300 min, during which no regrowth of the bacterial culture was observed (Gu et al., 2019).

Compared to the use of phages individually, the greater effectiveness of the phage cocktail was confirmed. The liquid culture of the UPEC strain AL505 was lysed by a cocktail composed of three phages - AL505-P1, AL505-P2, and AL505-P3, and did not recover within 15 h, whereas the use of each phage separately resulted in the onset of bacterial population rebuilding already after 6 h (Galtier et al., 2016). It was emphasized that after 15, no mutants resistant to the applied phages could be identified. In the case of phage IBEC77, the burst size was approximately 150 phage particles released from UPEC cells (Sanmukh et al., 2023). The high burst sizes, 198 PFU/cell and 250 PFU/cell, and short latent period (~10 min), were observed for phages vB_EcoS_U-TEC10 and vB_EcoA_RDN8.1, respectively (Chaudhary et al., 2022b; Rajab et al., 2024), thus making them good candidates for phage therapy. For phage myPSH1131, the phage titer was 130 virions per cell, and the latent period was 20 min (Manohar et al., 2018). After the addition of the phage lysate (MOI = 0.001) into the culture of the primary host *E. coli* PSH131 strain (EPEC, enteropathogenic *E. coli*) for 5 min, 20 % of the phages were attached to the surface of bacterial cells. After 25 min, there were no free, unattached phage particles present in the mixture (Manohar et al., 2018). Similar to myPSH1131, the latent period of phage FS2B was approximately 20 min, and the burst size was 137 PFU/cell (Malik et al., 2024). In addition, approximately 90 % of phage FS2B particles adsorbed to the cells within 12 min (Malik et al., 2024). An even better result was observed for phage Mimir124, which adsorbed with an efficiency of above 90 % in 3 min (Golomidova et al., 2024). The study of the Killian phage on the UPEC CFT073 strain also revealed its high burst size, amounting to 139 phage particles per bacterial cell, and a latency period of 40 min (Khunti et al., 2023). Moreover, the Killian phage efficiently adsorbs to UPEC cells, as approximately 90 % of phage particles were bound to the host within 20 min of introducing the phage lysate into the bacterial culture. Interestingly, an experiment assessing the lysis efficiency of the *E. coli* CFT073 culture by the Killian phage showed that at high MOI values (1, 10, or 100), the phage causes a rapid decrease in optical density in the initial 8 h of incubation with bacteria. However, after this period, the bacterial population recovered rapidly. On the other hand, lower MOI values, especially 0.001 and 0.0001, resulted in less drastic lysis, but the population did not undergo recovery (Khunti et al., 2023).

Among four phages infecting UPEC strains, FS11 exhibited the highest phage titer, amounting to 121 phage particles per bacterial cell, and the shortest latency period of 24 min (Malik et al., 2021). The other phages, FS17, PS8, and PS6, also showed substantial burst sizes of 98, 87, and 78 phage particles released during bacterial cell lysis, with latency periods of 26, 42, and 35 min, respectively. The efficient infection process was further indicated by the high adsorption efficiency of these phages, reaching 96 % of bound phage virions within 20–35 min, depending on the phage strain. Additionally, the lytic effectiveness against the UPEC strain PB3 was examined for each phage individually and collectively in a cocktail with an MOI value of 0.1 of each phage used. The results indicated that each phage effectively lysed the host culture. However, when the cocktail was applied, the optical density measurements remained consistently low throughout the entire experiment (12 h), suggesting the superiority of phage cocktails over preparations composed of a single phage strain (Malik et al., 2021).

Among all phages infecting UPEC strains found in the literature, phage 590B possesses the most favourable kinetic parameters (Chaudhary et al., 2022a). Its burst size is 109 phage particles, with a latency period of only 15 min. Moreover, within 12 min of introducing the phage lysate into the UPEC590 strain culture, 95 % of the virion particles were adsorbed. Furthermore, phage 590B demonstrated lytic activity against UPEC strain at three different MOI values (0.01, 0.1, and 1). It is worth noting that at the lowest MOI, optical density measurements remained stable and at a low level for the entire 24 h duration of the experiment, whereas at the other two MOI values, bacterial population renewal began after 20 h (Chaudhary et al., 2022a).

It was demonstrated that the phage vB_EcoP_LNA3, isolated on the UPEC MS7163 strain, exhibits a latency period of only 2 min (Ngiam et al., 2021). However, it is characterized by a low phage yield, releasing only 12 phage particles from the cell after infection. On the other hand, another phage described in the same study, vB_EcoP_LNA5, also showed a very short 2-min latency period but had a significantly higher phage yield of up to 100 phage particles. Intracellular development of both phages appeared extremely rapid, however, such examples are known in the literature, e.g., phage vB_EcoS-95 with the eclipse and latent periods

of 3 min and 4 min, respectively (Topka et al., 2019). On the other hand, vB_EcoP_LNA5 efficiently adsorbed to bacterial cells, with 99.4 % of phage particles adsorbed within 4 min, compared to 66.9 % for phage vB_EcoP_LNA3. The phage yields for the remaining phages, vB_EcoM_LNA1, vB_EcoM_LNA2, and vB_EcoM_LNA4, were 134, 68, and 28 virion particles per cell, respectively, with latency periods of 15, 10, and 20 min, respectively. The efficiency of phage adsorption varied between 60 and 80 %. It is worth emphasizing that experiments for the phages vB_EcoM_LNA1 and vB_EcoM_LNA2 were conducted in the *E. coli* MG1655 strain, while for the remaining phages, they were performed in the primary UPEC hosts (Ngiam et al., 2021). In the case of phage FU3, which was able to reduce the population of the uropathogenic host by 2 orders of magnitude within 25 h at an MOI of 0.9, information of phage titer and adsorption efficiency was not provided (Sufa et al., 2018). The article also lacks data on the subsequently investigated phage FU4 (Sufa et al., 2018). It is important to note that many studies about UPEC phages do not include information about their development kinetics. However, understanding these aspects is crucial for phage therapy.

5. Application of phages destroying UPEC bacteria in experimental models

Phages used in phage therapy must not only be well-characterized but also tested on various experimental models in the laboratory. Such studies allow us to assess their effectiveness in more natural conditions. In various studies, the effectiveness of phages in eliminating biofilm generated by UPEC, in both living and non-living surfaces, such as urological catheters, has been frequently assessed.

The application of phage vB_EcoP_EG1 as a limiting and reducing factor for biofilm formation was investigated using *E. coli* MG1655 and UPEC strain 390G7 (Gu et al., 2019). Incubating both strains with the phage on 96-well plates for 24 h resulted in the formation of a biofilm, with biomass reduced by 50 % in the UPEC strain and 60 % in the MG1655 strain compared to cultures incubated without phage lysate (Gu et al., 2019). The biomass of the biofilm formed by the UPEC strain AL505 was effectively reduced by a cocktail composed of phages AL505-P1, AL505-P2, and AL505-P3; biomass reduction occurred by approximately 60–70 % depending on the use of different phage titers in the preparation (Galtier et al., 2016). In another study, 21 phages were isolated, including 3 specific to the UPEC strain CFT073 (Sanmukh et al., 2023). One of these phages, named IBEC77, and two non-infective phages were tested to examine their biofilm reduction abilities on the following strains: CFT073, characterized as strongly biofilm-forming, MG1655 with moderate biofilm-forming ability, and LF82, which produced the weakest biofilm. After 72 h of incubation with phages, it was found that phage IBEC77 significantly reduced the biofilm biomass of the CFT073 strain by 68 % relative to the control variant. This could be due to the cumulative action of the phage and its depolymerase. Subsequently, the UPEC strain CFT073, known for its high biofilm-forming capabilities, was selected to investigate the biofilm-eradicating potential of phages IBEC77 and IBEC40 (non-infective strain to UPEC) in a flow system. The results confirmed the findings obtained in the static system – phage IBEC77 significantly reduced the biofilm biomass and thickness. Interestingly, phage IBEC40, despite being incapable of infecting the UPEC strain, also reduced the biofilm biomass and thickness of the biofilm layer formed by this strain. This phage was deliberately chosen for the experiment due to the production of depolymerase, similar to phage IBEC77 (Sanmukh et al., 2023). Thus, biofilm could be reduced by a depolymerase associated with phage. Indeed, phages that secrete depolymerase could be a valuable tool in fighting against biofilms formed by strains not specifically targeted by the applied phage.

Phage 590B also proved to be effective in reducing the biofilm produced by the UPEC strain 590 in Foley's catheter (Chaudhary et al., 2022a). Additionally, it was investigated how the phage would reduce the number of bacterial cells in a previously disrupted biofilm. It was found that the action of the phage led to a reduction in the number of bacterial cells in the previously mechanically damaged biofilm by approximately 0.6–1 orders of magnitude (CFU/ml) for each of the three applied MOIs - 0.01, 0.1, and 1 in comparison to undisturbed biofilm. This confirmed the significant role that the intercellular matrix plays in the stability of the biofilm and how its disruption positively affects the action of the phage, allowing it to reach bacterial cells. Interestingly, both in the case of the disrupted and intact biofilm, the use of the phage at an MOI value of 0.01 resulted in the fewest viable bacterial cells (Chaudhary et al., 2022a). Similarly, the effect of three different phage titers of three different phages, vB_EcoM_ACG-C40, vB_EcoM_ACG-C91, and vB_EcoM_ACG-M12, on the biofilm formed in 96-well plates by the UPEC strain Can 91 was tested (Chibeu et al., 2012). It was suggested that lower phage titers in phage therapy may be equally effective as higher ones. The resurgence of bacterial populations after a 24-h incubation with the phage, which may be attributed to the emergence of phage-resistant mutants, was also assessed. It is worth noting that in the case of two out of three tested phages, this renewal occurred when the highest lysate titer was applied (Chibeu et al., 2012). Interestingly, phages vB_EcoM_ϕEc1 and vB_EcoM_ϕEc3 could reduce biofilm biomass in only a few of the *E. coli* biofilm-forming strains - 4 out of 9 and 3 out of 6, respectively (González-Villalobos et al., 2021). However, statistically significant differences were observed only in the case of the action of phage vB_EcoM_ϕEc1 on the biofilm formed by two strains. In contrast, phage vB_EcoM_ϕEc4 showed no activity against the biofilm. These surprising results were explained by the significantly shorter exposure time of the phage to the biofilm layer, which was only 1 h. Indeed, literature often reports much longer incubation times of phages with biofilm, ranging up to several dozen hours, as indicated in earlier sections of this article.

Phages infecting UPEC bacteria have also been tested on live organisms. Previously tested on biofilms, phages IBEC77 and IBEC40 were evaluated for their effective eradication of infection caused by *E. coli* CFT073 in *Galleria mellonella* larvae (Sanmukh et al., 2023). A combination of phages with the antibiotic ciprofloxacin was also used to assess the potential synergistic action of these components. When larvae were only exposed to bacteria, 50 % of them died within 22 h. However, when phages were administered separately, this time extended to 38 and 28.5 h, for phages IBEC77 and IBEC40, respectively. After administering a dose resembling phage IBEC77, this time extended to 50.5 h, whereas for phage IBEC40, it underwent a reduction, confirming earlier results obtained on biofilms. In the case of combined therapy involving the administration of phages together with ciprofloxacin, 100 % of larvae remained alive for 55 h. However, it was noted that the administration of ciprofloxacin at a dose of 20 mg/kg alone yielded the same effect in the same time (Sanmukh et al., 2023).

Combined therapy is a promising method for combating drug-resistant bacteria, nevertheless, it requires further research to assess different doses of phages, antibiotics, administration order, and timing. The significance of booster doses despite the short developmental cycle of phage was also confirmed in the study assessing the efficacy of phage myPSH1131 in treating *G. mellonella* larvae infected with the PSH1131 strain (EPEC), where only after the application of three doses of phage, larval survival reached 100 % (Manohar et al., 2018).

Phages infecting UPEC bacteria were tested for combating induced infections in mice. The application of a high MOI of 60 of phage KEP10 in mice infected with UPEC strain *E. coli* ECU5, resulted in 90 % survival of the individuals after one week, whereas untreated mice died within three days (Nishikawa et al., 2008). Despite the use of such a high titer of phage lysate, no behavioural abnormalities were observed in mice over a month (Nishikawa et al., 2008). The impact of administering a cocktail of three phages – vB_EcoM_AL505-P1, vB_EcoM_AL505-P2, and vB_EcoM_AL505-P3 into the multispecies bacterial culture occurring in the gut of mice was examined (Galtier et al., 2016). Before administration of the phage cocktail, mice were treated with streptomycin and kanamycin for a week to facilitate colonization of the colon by the antibiotic-resistant strain AL505. At the stage of administering

antibiotics and the phage cocktail, the composition of gut microbiota in the tested mice was assessed. Phages, as viruses that exhibit high specificity towards bacterial hosts, indeed led to lesser disruptions in the composition and species proportions of bacteria in the mouse gut compared to the used antibiotics. The major changes in species composition during phage cocktail treatment involved a drastic decrease in the abundance of *E. coli*, which is a consequence of using phages isolated with the UPEC strain AL505 (Galtier et al., 2016). Another study was conducted on the mouse model focusing on the action of a single phage named PEC80 and a phage cocktail composed of 25 phages targeting UPEC bacteria (Mijbel Ali et al., 2021). A higher efficacy of the phage cocktail in treating UPEC-induced urinary tract infections in mice compared to the individual phage strains against *E. coli* 8 and *E. coli* 302 was confirmed. However, the publication did not include basic information about the characteristics of the phages present in the cocktail or the PEC80 phage (Mijbel Ali et al., 2021). A mouse infection model was also used to assess the bacteriolytic activity of phage vB_EcoS_UTEC (Rajab et al., 2024). That work revealed that 90 % of mice survived the infection with *E. coli* in phage-treated variant, whereas only 45 % survived in the untreated one. In addition, this phage showed significant antibiofilm activity (Rajab et al., 2024). The least attention was paid to toxicity tests against eukaryotic cells because in none of the analyzed works was such an analysis performed. Nevertheless, it should be noted that such experiments are necessary due to demonstrated interactions between phages and the animal immune system (Podlacha et al., 2024).

6. Bioinformatic approaches applied in the analyses of the anti-UPEC phage genomes

6.1. Phage genome assembly and annotations

Lytic phages infecting UPEC identified so far vary considerably at the genomic sequence level. Genome sizes of the phages in question range from 39,091 base-pairs in phage LNA5 (Ngiam et al., 2021) to 169,905 base-pairs in Killian phage (Khunti et al., 2023) (Table 1). Most authors determined phage genome sequences using Illumina short reads. Exceptions include phages – ACG-C91, ACG-M12, and ACG-C40, which were sequenced by pyrosequencing (Chibeu et al., 2012) and phage St11Ph5, sequenced with the Ion Proton System (Golomidova et al., 2018). The following phages: vB_EcoM_LNA1, vB_EcoM_LNA2, vB_EcoP_LNA3, vB_EcoM_LNA4, vB_EcoP_LNA5, vB_EcoM_SHAK7704, vB_EcoM_SHAK7854, vB_EcoM_SHAK7858, vB_EcoM_SHAK7163, and vB_EcoM_SHAK9454 were sequenced using hybrid approach by Oxford Nanopore and Illumina platforms (Ngiam et al., 2021; Asgharzadeh Kangachar et al., 2024). Genome assemblies by different authors were generated with diverse algorithms. SPAdes was the most frequently applied (59 phages), followed by CLC Genomics Workbench (7 phages). Other programs were used just once (SKESA, IVA, Newbler, SOAPdenovo, Platanus).

In most publications, annotations for these genomes were typically performed with Glimmer (14 phages) and GeneMark (9 phages), allowing for coding sequence prediction. Other programs, such as Fgenes was used to annotate 2 phages: myPSH1131 and FS2B (Manohar et al., 2018; Malik et al., 2024), and Artemis and Prodigal were used once for annotations of EG1, and phiEC3/phiEC4 phages, respectively (Gu et al., 2019; González-Villalobos et al., 2021). The number of identified ORFs ranged from 49 in phage RDN8.1 (Chaudhary et al., 2022b) to 305 in KMB37 virus (Markusková et al., 2024), as shown in Table 1. Functional annotations employed different BLAST programmes and multiple databases, including UniProt, SwissProt, NCBI RefSeq and Conserved Domains, TIGRFAM, Pfam, or pVOGs. In addition, some authors used the PHAST, PHASTER, or RAST web servers (Aziz et al., 2008; Zhou et al., 2011; Arndt et al., 2016; Nawaz A. et al., 2024) alongside the programs described above (as in the case of the NTEC3 phage; Chaudhary et al., 2022c), while others relied solely on RAST (as in the case of the KMB26 phage; Slobodníková et al., 2021). Additional tools such as TMHMM, HMMER, and PhANNs were applied for annotation refinement. Notably, eight phages (vB_EcoM_LNA1, vB_EcoM_LNA2, vB_EcoP_LNA3, vB_EcoM_LNA4, vB_EcoP_LNA5, vB_EcoM_SHAK7704, vB_EcoM_SHAK7854, and vB_EcoM_SHAK9454) were sequenced but not annotated in the original publications. For the purposes of this study, these genomes were annotated using Pharokka v1.6.1 (Bouras et al., 2023). The resulting gene lists are provided in Supplementary Table S1.

Annotation revealed interesting phage proteins involved in host cell lysis, including 40 Rz-like or putative spanins, 75 holins or putative holins, and 45 endolysins. Proteins involved in DNA synthesis were also identified. These included 90 DNA polymerases, 18 DNA topoisomerases, 118 DNA helicases, and a variety of nucleases, like 287 endonucleases, 7 endoribonucleases, and 96 exonucleases (Supplementary Table S2). Following Latka et al. (2019) guidelines, 16 potential depolymerase genes in the genomes of 16 analyzed phages were identified using the Phyre2 software (Kelley et al., 2015) and are listed in Supplementary Table S3. Moreover, some authors employed programs for the identification of tRNAs (tRNAscan-SE, ARAGORN; Laslett and Canback, 2004; Lowe and Chan, 2016), rRNAs (Rnammer; Lagesen et al., 2007), promoters (PhagePromoter, PHIRE; Lavigne et al., 2004; Sampaio et al., 2019), and rho-independent terminators (ARNOLD; http://rssf.i2bc.paris-saclay.fr/toolbox/arnold/index.php). These searches identified one to twelve tRNA genes in 23 phages. No promoters, terminators, or rRNA were identified. Automated annotation pipelines were also employed in two studies. Prokka was applied to 31 phages, but no other curation was carried out (Golomidova et al., 2018; Loose et al., 2021; Ssekatawa et al., 2024). Pharokka-phold-phynteny pipeline was used once on phage Mimir124 (Golomidova et al., 2024), and pharokka pipeline was used once on phages øEc_Makalu_001, øEc_Makalu_002, øEc_Makalu_003 (Dhungana et al., 2024).

6.2. Phage genome content analyses

Because of the medical use of phages in therapy, a fundamental question is whether newly discovered phages carry toxin genes, virulence factors, or antibiotic resistance genes. Different strategies were applied by the authors to evaluate the safety of the identified viruses, with many approaches relying on annotation profiles. Apart from the databases described above, most authors relied on the Virulence Factor Database (VFDB), Comprehensive Antibiotic Resistance Database (CARD), or Pathogenicity Island Database (PAIDB). In addition, specialised programs such as VirulenceFinder (Joensen et al., 2014), ResFinder (Bortolaia et al., 2020), RGI (Alcock et al., 2023), or PhageLeads (Yukgehnaish et al., 2022) were frequently used to screen for undesired genes. Some authors relied solely on manual verification of information regarding the presence of virulence genes. However, not all publications provided explicit information on the presence or absence of toxins, virulence factors, or antibiotic resistance genes in UPEC phage genomes. To address this, we performed several analyses of the phage genomes compared in this study using the following programs: ABRicate 1.0.1 with all available databases (databases version 2023-Nov-4; https://github.com/tseemann/abricate), AMRFinderPlus 3.12.8 (database version 2024-01-31.1; Feldgarden et al., 2021), RGI 6.0.3 (database version 3.2.9; Alcock et al., 2023), VirulenceFinder 2.0.5 (Joensen et al., 2014), and ResFinder 4.4.2 software (Bortolaia et al., 2020). Importantly, all programs consistently confirmed that the analyzed phage genomes were free of virulence factors and antibiotic resistance genes.

Phage development cycle is another critical factor for the phage to be applied in phage therapy. Phages may develop by the lytic or lysogenic pathway, however, only lytic phages are suitable. Most publications classified phages by the absence of specific genomic markers, such as integrase, recombinase or excisionase. Currently, many tools are available for predicting the phage development cycle, yet only six of the analyzed studies made use of them. In two cases, the authors applied PHACTS (Loose et al., 2021; Ngiam et al., 2021), in one case, BACPHIP and PhageLeads were used together (Nawaz et al., 2024), and in another

case PHACTS and BACPHIP were used together (Vera-Mansilla et al., 2022). The PhageAI program was applied twice (Khunti et al., 2023, Dhungana et al., 2024), while PhaBOX and PhaTYP were used once (Asgharzadeh Kangachar et al., 2024). PhageLeads (Ismael et al., 2024) was also used once. To obtain comparable results across all phages tested in this study, we performed development cycle predictions using PHACTS 1.8 (McNair et al., 2012), BACPHLIP 0.9.6 (Hockenberry and Wilke, 2021), and PhaTYP 0.3.0 (Shang et al., 2023) (Supplementary Table S4). PHACTS evaluated 23 phages (LNA4, SHAK7858, UTEC10, Golestan, KMB35, CHD5UKE2, UTI89UKE1, 101118B1, CHD5UKE1, ES, MLP1, phiEc4, SHAK7693, St11Ph5, phiEc3, KMB36, 590B, UTI89UKE3, 22664BS2, ACG-C91, KMB41, 101117UKE2, MLP3) as temperate. However, in all cases the probability scores ranged between 0.50 and 0.54 – close to the limit value. BACPHLIP 0.9.6. produced more confident predictions, although in five cases (KMB36, 590B, KMB35, NTEC3, KMB37) the probability that the phage is lytic dropped below 0.7 value. The PhaTYP program classified 6 phages (FS2B, 101114BS3, 101114UKE3, CHD5UKE1, myPSH1131, MLP3) as temperate with scores over 0.999. Overall, deeper experimental verification is recommended for KMB35, KMB36, CHD5UKE1, MLP3, and 590B development cycles as they were independently classified as temperate by two programs. In addition, myPSH1131, 101114BS3 and 101114UKE3 warrant closer investigation due to their high PhaTYP score indicating lysogenic development, especially given PhaTYP's leading role among all such programs. Interestingly, when considering lysogeny-related gene carriage in these phages, only the integrase gene was identified in genomes of phages 101114BS3, 101114UKE3, and CHD5UKE1. No other genes typically associated with lysogenic development of phages, such as repressors of the lytic cycle or recombination-related proteins, were identified. This highlights cases where induction experiments in liquid culture of a potential lysogen strain should be performed to confirm or exclude a prophage state, since temperate phages - inherently capable of transferring toxic or antibiotic-resistance genes - are unsuitable for phage therapy.

6.3. Phylogenetic analyses

Phylogenetic relationships were investigated with VipTree webserver (Ngiam et al., 2021; Slobodníková et al., 2021; Chaudhary et al., 2022a; Vera-Mansilla et al., 2022; Khunti et al., 2023; Asgharzadeh Kangachar et al., 2024; Rajab et al., 2024; Ssekatawa et al., 2024; Nawaz et al., 2024) and MEGA software (Chaudhary et al., 2022a, 2022c; González-Villalobos et al., 2021; Gu et al., 2019; Rajab et al., 2024). Other tools such as phylogeny.fr website (Yazdi et al., 2020) and PhyML (Vera-Mansilla et al., 2022), IQ-TREE (Ssekatawa et al., 2024) were also utilised. Phylogenetic trees focused on single gene markers engaged genes coding for terminase large subunit (Ismael et al., 2024; Malik et al., 2024; Ngiam et al., 2021), major head protein (Ismael et al., 2024; Malik et al., 2024; Manohar et al., 2018; Ngiam et al., 2021; Yazdi et al., 2020; Ssekatawa et al., 2024), and DNA polymerase (Yazdi et al., 2020; Ngiam et al., 2021; Chaudhary et al., 2022c; Ismael et al., 2024; Malik et al., 2024). The most frequently used method was neighbour-joining with a bootstrap test of 1000 replicates. The only exception was phage 590B, for which the bootstrap was limited to 500 replicates (Chaudhary et al., 2022a). Maximum-likelihood method with a bootstrap test equal to 1000 replicates was used by the authors describing phages MLP1, MLP2, and MLP3, vB_Ec_ZCEC14, UGKSEcP1, and UGKSEcP2 (Vera-Mansilla et al., 2022; Ismael et al., 2024; Ssekatawa et al., 2024). The fast minimum evolution method was used in case of phages øEc_Makalu_001, øEc_Makalu_002, øEc_Makalu_003 (Dhungana et al., 2024). Sequence selection strategies for tree construction also varied. For phage NTEC3, the authors used the DNA polymerase genes with BLASTp values > 90 % (Chaudhary et al., 2022c). A similar approach was applied to phage 590B; however, instead of DNA polymerase, the authors used terminase large subunit sequences with BLASTp identity values above 99 % (Chaudhary et al., 2022a). For phages MLP1, MLP2, and MLP3, the authors first constructed VipTree-based phylogenetic tree, then selected closely related phages and built a maximum-likelihood tree based on terminase large subunit amino acid sequences (Vera-Mansilla et al., 2022). Today, it is well recognised that phage phylogeny is complex and modern approaches often require analysing multiple genes – or even the whole genome – when possible. The whole-genome approach was applied for EG1 using the neighbour-joining method with 1000 bootstrap replicates. Neighbour-joining without bootstrap was used for myPSH1131 (Manohar et al., 2018; Gu et al., 2019). In this study, we constructed a VipTree-based phylogenetic tree for all 82 analyzed phages (Fig. 1). Notably, 15 phages listed in Table 1 (FU1, FU3, IBEC 77, IBEC 78, IBEC 98, AL505_P1, AL505_P2, AL505_P3, FS11, FS17, PS6, PS8, KEP10, vB_EcoM- ϕEc1, Escherichia phage) were not deposited by their authors and, therefore, could not be included in this multigene tree. Conversely, phage FS2B was deposited twice, and both sequences were included in the analysis. Based on the results we concluded that phages are not grouped in monophyletic groups according to taxonomy deposited in NCBI. For instance, phages SHAK7858 and SHAK7693, classified under the *Studiervirinae* subfamily, are grouped within *Guernseyvirinae* subfamily. Similarly phages SHAK7704 and SHAK7163, although assigned to *Studiervirinae*, are closely related to *Molineuxvirinae* subfamily. All such discrepancies are illustrated in VipTree-based phylogenetic tree (Fig. 1).

The results further allowed the identification of 7 clades (I-VII). Clades were divided at 0.15 height of normalized maximum tree, selecting as many phages as possible with at least twenty orthologous genes available. Consequently, clades I-VII encompass 26, 4, 5, 12, 6, 10, and 12 phages, respectively, while phages LNA1, LNA4, KMB37, KMB41, MLP1, Mimir124, and St11Ph5 do not belong to any clade. Based on this grouping, 7 multigene trees (one for each clade) were made of 60, 24, 53, 20, 27, 28, and 14 orthologous genes, respectively, as shown in Figs. 2–8. All orthologous phage genes are listed in Supplementary Table S5. The multigene phylogenetic trees were generated as described in the Methodology Section. An effort was also made to reduce the number of genes required to reconstruct the same tree topology as the original multigene trees. For the trees presented in Figs. 2–8, reductions were made to 51, 1, 2, 6, 1, 18, 5 orthologous genes, respectively. Reductions were performed by removing from the analysis those genes whose alignments contained the fewest parsimony-informative sites, provided that the tree topology remained consistent with the original multigene tree. As reported in Supplementary Table S5, orthologous genes were gathered with the number of parsimony informative sites for each alignment, and analyzed using InterProScan 5.72 (Jones et al., 2014) to identify common protein features and address discrepancies in author-provided annotations. This analysis primarily utilised CDD, Hamap, Pfam, FunFam, SUPERFAMILY, and NCBIFam.

The most complex multigene tree was constructed for phages in clade 1 (Fig. 2). This clade comprised all 26 analyzed phages from the *Straboviridae* family, and gathered 60 orthologous genes. Interestingly, only 51 genes were sufficient to reconstruct the same tree topology; these are indicated by "+" in Supplementary Table S5. The most influential genes, contributing the greatest phylogenetic signal, encode proteins with features involved in tail fibres and baseplates production (801 and 502 parsimony informative sites, respectively), and in DNA replication (481 parsimony informative sites). Among the 51 genes significant for this clade, 13 encode capsid proteins, 16 are involved in DNA replication, 18 encode various enzymes, and 4 are not recognised by InterProScan.

Fig. 3 presents a tree constructed from 24 orthologous genes of four phages from clade 2. Interestingly, as indicated in Supplementary Table S5, only a single gene – a tail protein with 181 parsimony informative sites – is sufficient to reproduce the full tree topology. Fig. 4 shows all five phages belonging to the *Dhillonvirus* genus. Although they acquire 53 orthologous genes, only two – hypothetical proteins, collectively carrying 74 parsimony informative sites – are needed to reconstruct the same tree topology. Notably, within clade 3, 26 genes

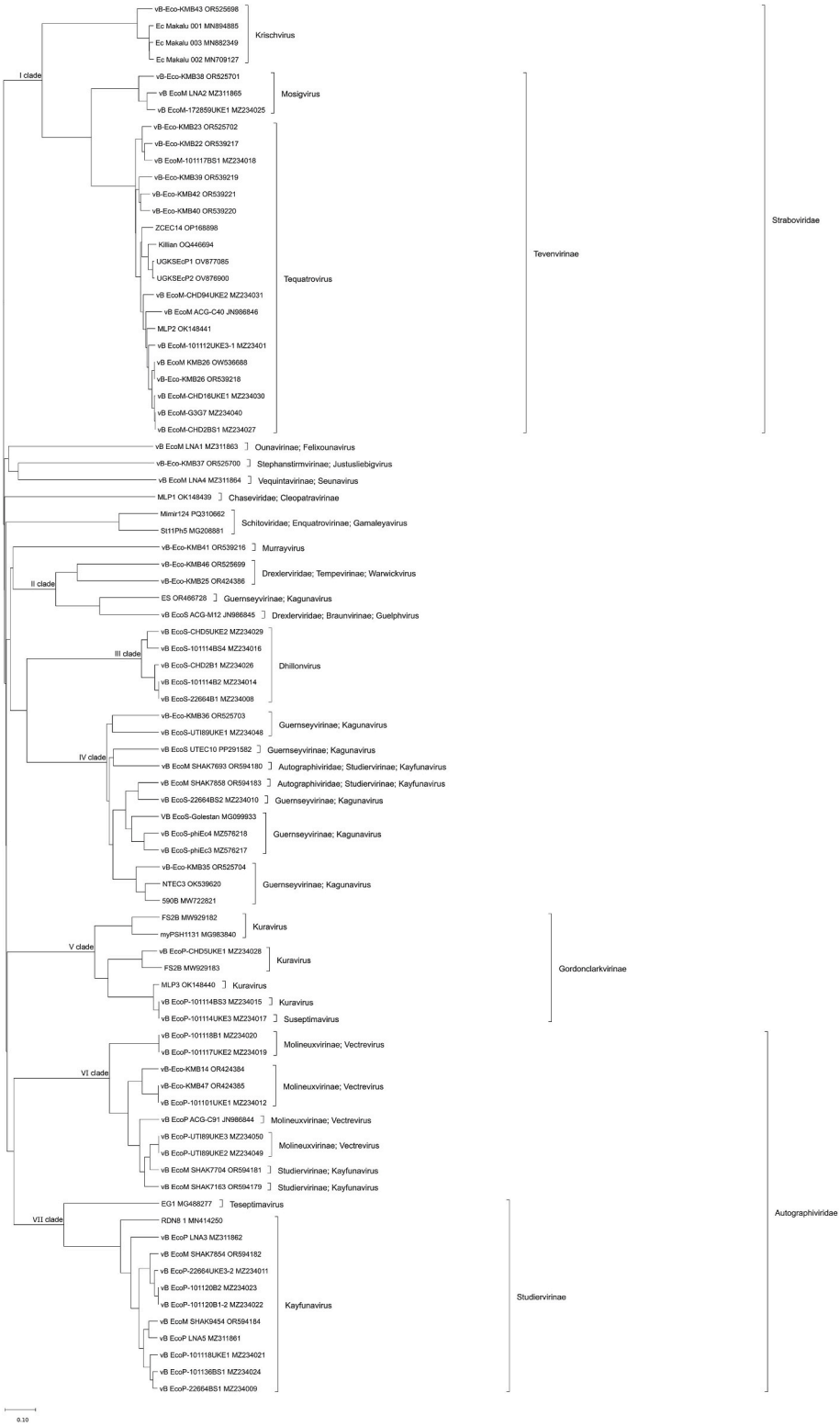


Fig. 1. The VipTree-based phylogenetic multigene tree for 82 phages infecting UPEC, whose sequences were available in the NCBI database. The analysis allowed to distinguish 7 clades, which are numbered I-VII. The following VipTree options were used to generate this tree: no reference, prodigal to get gene information, genetic code no. 11, no functional prediction of proteins. Bootstrap values were not generated.

show no phylogenetic signal, indicating they are conserved across all analyzed phages, and 29 genes lack any predicted domains according to InterProScan. Clade 4, shown in Fig. 5, includes phages from the *Guernseyvirinae* subfamily along with two phages from the *Studiervirinae* subfamily. In this case, it was possible to reduce the number of genes from 20 to 6, three of which encode phage tail proteins, while the other three do not have an explicit role assigned yet. Based on phylogenetic analysis, it is unlikely that SHAK7693 and SHAK7858 belong to the *Studiervirinae* subfamily, as they cluster closely with representatives of the *Guernseyvirinae* subfamily. Clade 5, visualised in Fig. 6, comprises

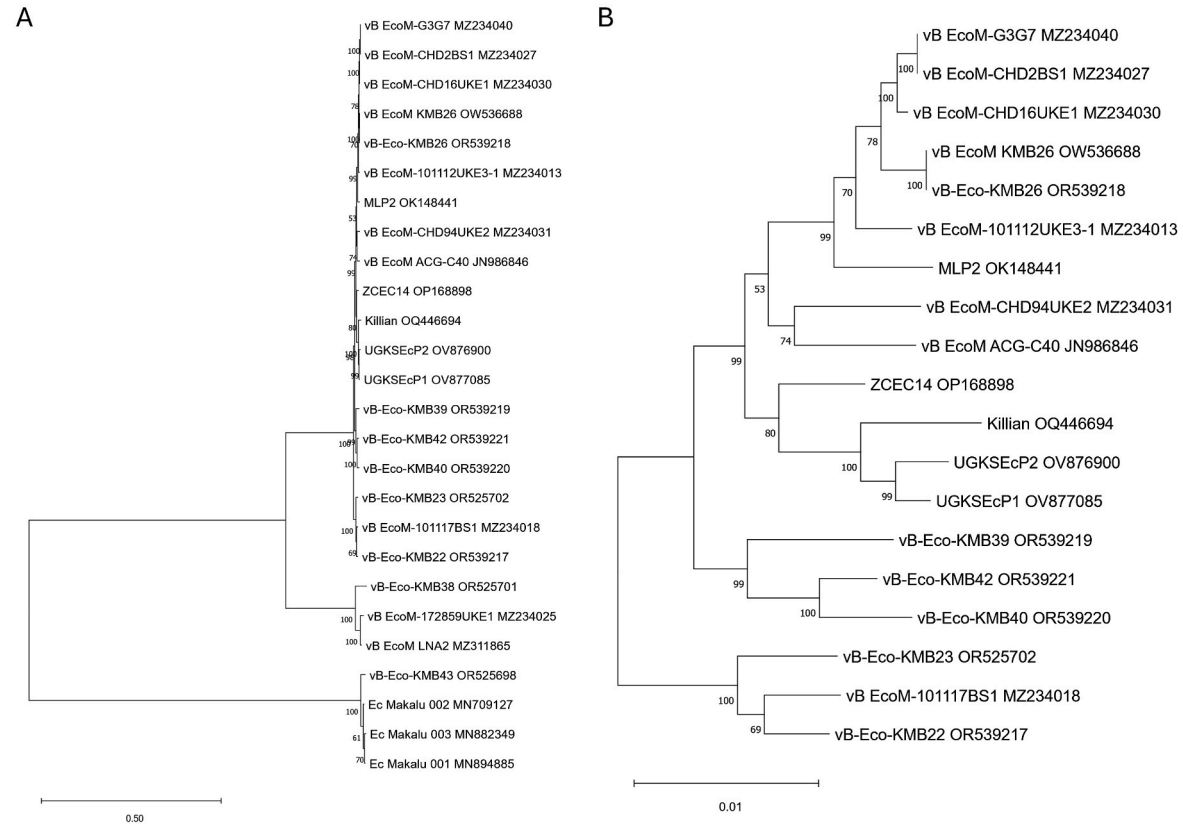


Fig. 2. The multigene maximum likelihood phylogenetic tree with 1000 bootstrap replications, made for 26 phages from the first clade and generated based on 60 orthologous phage genes presented in Supplementary Table S5. Panel **A** presents all analyzed phages. Panel **B** shows a modified tree, where phages KMB38, 172859UKE1, LNA2, KMB43, øEc_Makalu_002, øEc_Makalu_003, øEc_Makalu_001 were excluded to make the tree more transparent. Orthologous groups were chosen by OMA software, alignment was made with Clustal Omega, IQ-TREE was used for tree computation, and visualization was made with MEGA software.

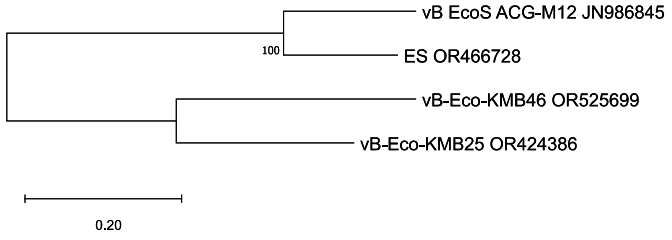


Fig. 3. The multigene maximum likelihood phylogenetic tree with 1000 bootstrap replications, made for 4 phages from the second clade and generated based on 24 orthologous phage genes presented in Supplementary Table S5. Orthologous groups were chosen by OMA software, alignment was made with Clustal Omega, IQ-TREE was used for tree computation, and visualization was made with MEGA software.

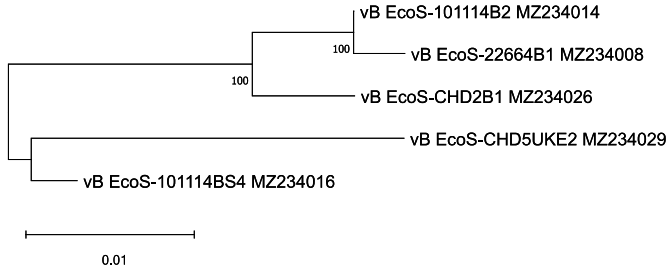


Fig. 4. The multigene maximum likelihood phylogenetic tree with 1000 bootstrap replications, made for 5 phages from the third clade and generated based on 53 orthologous phage genes presented in Supplementary Table S5. Orthologous groups were chosen by OMA software, alignment was made with Clustal Omega, IQ-TREE were used for tree computation, and visualization was made with MEGA software.

representatives of the *Gordonclarkvirinae* subfamily. In this case, only a single one out of 27 genes was required to reproduce the same tree topology as the multigene tree. The gene is glutathione synthetase with an ATP-binding domain carrying 151 parsimony informative sites.

Clades 6 and 7 include phages from the *Autographiviridae* family. Clade 6 comprises representatives of the *Molineuxvirinae* subfamily along with two phages from the *Studiervirinae* subfamily (Fig. 7). Interestingly, the reduction of shared orthologous genes was less effective compared to clades 2–5. Of the 28 available genes, 18 were required to preserve the tree topology. Among these 18 genes, 10 are unrecognised by InterProScan, 3 are involved in DNA replication, 1 is a tail tubular protein, and 4 are enzymes. The most influential is the orthologous group characterised as phage lysozyme with 551 parsimony informative sites. Notably, it is highly unlikely that SHAK7704 and SHAK7163 belong to the *Studiervirinae* subfamily, as they are too closely related to *Molineuxvirinae* phages. Finally, Fig. 8 presents phages from the *Studiervirinae* subfamily. Here, only 5 of the 14 orthologous genes are needed to maintain the same tree topology. Of these 5 genes, 3 are involved in DNA replication, 1 has a lysozyme-like function, and 1 has no identified domains.

7. Methodology

This anti-UPEC phages review was conducted by selecting articles published in the years 2008–2024, which were within the scope of this investigation. To search for such articles, the PubMed database was used

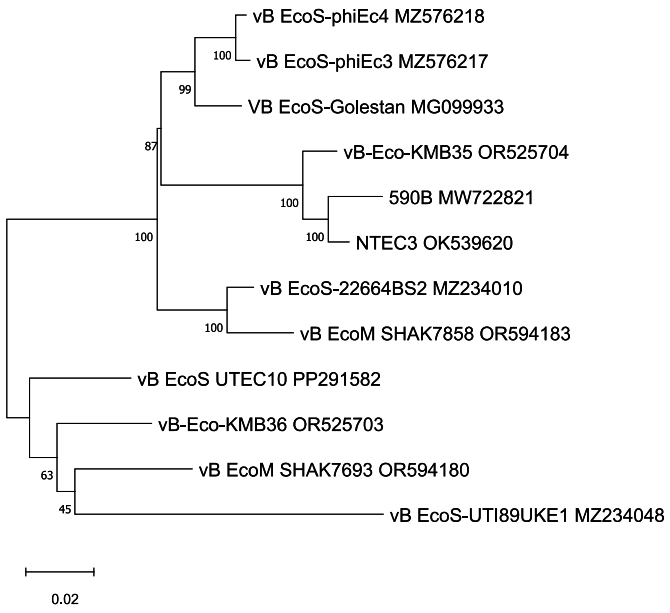


Fig. 5. The multigene maximum likelihood phylogenetic tree with 1000 bootstrap replications, made for 12 phages from the fourth clade and generated based on 20 orthologous phage genes presented in Supplementary Table S5. Orthologous groups were chosen by OMA software, alignment was made with Clustal Omega, IQ-TREE was used for tree computation, and visualization was made with MEGA software.

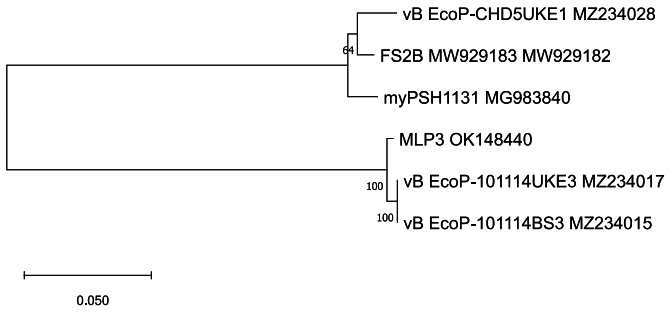


Fig. 6. The multigene maximum likelihood phylogenetic tree with 1000 bootstrap replications, made for 6 phages from the fifth clade and generated based on 27 orthologous phage genes presented in Supplementary Table S5. Orthologous groups were chosen by OMA software, alignment was made with Clustal Omega, IQ-TREE was used for tree computation, and visualization was made with MEGA software.

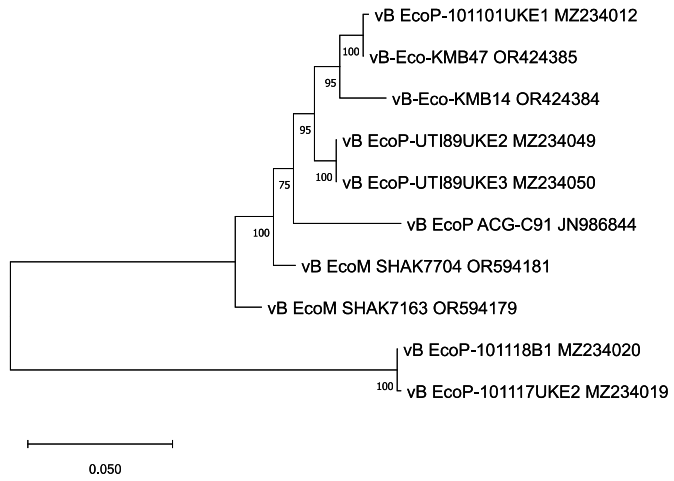


Fig. 7. The multigene maximum likelihood phylogenetic tree with 1000 bootstrap replications, made for 10 phages from the sixth clade and generated based on 28 orthologous phage genes presented in Supplementary Table S5. Orthologous groups were chosen by OMA software, alignment was made with Clustal Omega, IQ-TREE was used for tree computation, and visualization was made with MEGA software.

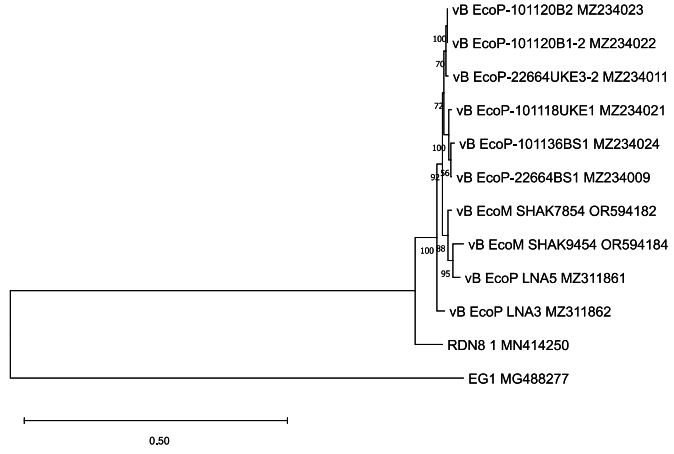


Fig. 8. The multigene maximum likelihood phylogenetic tree with 1000 bootstrap replications, made for 12 phages from the seventh clade and generated based on 14 orthologous phage genes presented in Supplementary Table S5. Orthologous groups were chosen by OMA software, alignment was made with Clustal Omega, IQ-TREE was used for tree computation, and visualization was made with MEGA software.

(https://pubmed.ncbi.nlm.nih.gov/; last accessed on October 9, 2025), and the following keywords in combinations were employed: UPEC, bacteria, phages, and phage therapy. In this way, a total number of 52 articles were identified. Non-English language articles were excluded, as were those that were not closely related to the phage therapy and phages infecting UPEC strains. After such an exclusion, 29 articles describing 98 phages were analyzed in detail. Interestingly, 10 out of the 29 articles were from 2024, 2 from 2023, 4 from 2022, 5 from 2021, 3 from 2018, and single articles from 2020, 2019, 2016, 2012, and 2008. In addition, 50 articles reviewing antibiotic resistance among UPEC phages, UTIs, and applied in silico methods were also included. In total, 79 articles were investigated in the groundwork presented here.

In silico analyses were performed using the following programs. Pharokka v1.6.1 (Bouras et al., 2023) was employed to annotate phage genomes, however, for depolymerases identification, the Phyre2 software (Kelley et al., 2015) was used according to the guidelines reported by Latka et al. (2019). To detect virulence factors, toxins, and antibiotic resistance genes, several programs were utilised, including ABRicate 1.0.1 with all available databases (database version 2023-Nov-4; https://github.com/tseemann/abricate), AMRFinderPlus 3.12.8 (database version 2024-01-31.1; Feldgarden et al., 2021), RGI 6.0.3 (database version 3.2.9; Alcock et al., 2023), VirulenceFinder 2.0.5 (Joensen et al., 2014), and ResFinder 4.4.2 (Bortolaia et al., 2020). To determine the phage life cycle, we employed PHACTS 1.8 (McNair et al., 2012), BACPHLIP 0.9.6 (Hockenberry and Wilke, 2021), and PhaTYP 0.3.0 (Shang et al., 2023) programmes.

To establish a common evolutionary origin, the VipTree server 4.0 (Nishimura et al., 2017) was used to cluster the analyzed phages into clades. Clades were defined at 0.15 of the normalized maximum tree height to include as many phages as possible with at least twenty orthologous genes. Orthologous gene groups were identified using OMA 2.6.0 (Altenhoff et al., 2019). Amino acid sequence alignments were performed with Clustal Omega 1.2.4 (Sievers et al., 2011), and maximum likelihood phylogenetic trees were constructed using IQ-TREE 3.0.1 (Minh et al., 2020) with 1000 bootstrap replicates. MEGA

software was used for tree visualization with midpoint rooting. Trees based on a reduced number of orthologous genes were generated by removing alignments with the smallest amount of parsimony-informative sites, the absence of which did not cause a change in the original multigene tree topology. Protein domain and feature analyses were conducted using InterProScan 5.72 (Jones et al., 2014).

8. Conclusions

Lytic phages have several advantages, particularly in combating multidrug-resistant bacteria like UPEC strains. They can destroy biofilms, self-replicate, and selectively kill pathogens, showing promise for treating bacterial infections. Moreover, they carry great value in the form of phage proteins e.g. endolysins and depolymerases, which can also serve as antimicrobial agents. Despite many advantages, phages have limitations, making their therapeutic use complicated. First, only phages with a lytic life cycle should be used in therapy, and as we have shown above, bioinformatic predictions often require experimental verification in this case. Second, some may show reduced activity due to an immune response that is investigated rarely. Additionally, there are phages highly specific to certain bacterial strains, which disqualifies them from broad use.

Unfortunately, significant discrepancies in host range testing methods reduce data comparability, and some tests, like the EOP assay, are often overlooked, leading to inaccurate effectiveness assessments. Considering 98 phages tested in this work, the EOP values were determined for only 42 % of them. There is also a lack of studies on the impact of UPEC-specific phages on eukaryotic cells, their stability, and insufficient bioinformatic analyses of their genomes for the presence of toxin or antibiotic-resistance genes. Among the phages analyzed here, none of them were tested for cytotoxicity towards eukaryotic cells (Table 1). Besides, in general, only 20 of 98 reported phages (20 %) were tested for stability at different temperatures (Supplementary Fig. S1, Panel A). It was found that all 20 phages were investigated at the range of temperatures optimal for bacterial growth (30–37 °C), however, more phages were analyzed at 30 °C than at 37 °C, which is more equal to the natural human body temperature. Besides, the temperature values at which the authors tested the phages varied significantly, and only 4 and 10 phages were tested at temperatures optimal for phage lysates storage (−20 °C and 4 °C, respectively). On the other hand, the highest number of phages (19 in total) were analyzed at 50 °C and 60 °C, the values which exceed the temperature of the human body or even values reached in fever, and are not very useful in relation to phage therapy (Supplementary Fig. S1, Panel A). Moreover, such a differentiated approach in the study of thermal phage stability and results interpretation makes it difficult to draw general conclusions. Our experience shows that the phage is stable when the phage count does not drop below 80 % during at least an hour of incubation, while at each larger drop, this stability is reduced. Unfortunately, such a universal approach was not always used by the authors. Similar difficulties occurred when assessing the stability of phages in response to different pH. In total, only 19 % of all 98 phages analyzed in this work were investigated in this context. As a rule, these were the same phages that were tested for stability at different temperatures. The only exception was the EG1 phage, whose stability at different pH conditions was not verified. Interestingly, phages were more frequently tested at pH values 4, 6, 8, and 10, while their number dropped significantly at extreme pH values (Supplementary Fig. S1, Panel B). Extremely low pH occurs in the human gastric fluid (1.5–2), although phages are usually not stable in such conditions and require additional protection. In turn, when analyzing the research on phage genomes, we found that 15 % of all tested phages were not even sequenced by the authors, and another 8 were sequenced but not annotated. Besides, the prediction of the cycle and the presence of toxin or antibiotic resistance genes was also not performed as standard, or the outdated software was often used. In addition, we also found that phylogenetic analysis requires more attention in terms of selecting genes that influence the topology of multigene trees and carry real informative value. In this light, some of these limitations can be easily overcome by performing a detailed physiological and bioinformatic characterization of the isolated phages and by paying more attention to the selection of appropriate and standardized methods.

However, based on the available results summarized in Table 1, we managed to divide the tested phages into three categories. The first category included phages that we considered well-characterized and analyzed in terms of host range, stability, kinetics of development, and genome content. Phages missing one of these results were determined as partially well-characterized, while phages lacking two or more were categorized as insufficiently characterized. Based on this division, only 7 phages: vB_Ecos_Golestan, myPSH1131, Killian, 590b, FS2B, UGKSEcP1, and UGKSEcP2 were identified as well characterized (Fig. 9), whereas the vast majority of the discussed phages have been insufficiently described. On the other hand, considering that basically all (95 %) of the phages gathered in Table 1 were tested for host range, we divided the phages according to the number of infected by them bacterial strains. The analysis revealed that phages 22664UKE3-2, vB_EcoM_KMB26, 101114BS3, 101118B1, and FS2B infected the highest numbers of isolates, respectively, 48, 45, 42, 40, and 37 (Supplementary Fig. S2). Interestingly, phage FS2B was indicated in both of the above-described analyses. Its detailed analysis confirmed that phage FS2B exhibited strong lytic activity, high burst size, and short adsorption and latent periods. This phage was also stable at 0-60°C, pH values between 4 and 12, and its titer remained almost unchanged in long-term storage at −20 °C and −70 °C in the presence of cryoprotectant. Besides, genome annotation confirmed the presence of genes associated with the lytic replication cycle, while lysogeny-related genes were absent. Moreover, combination assays indicated a positive synergistic interaction between phage FS2B and tested antibiotics (Malik et al., 2024). Undoubtedly, these preliminary findings suggest that phage FS2B holds strong potential as a promising candidate for the therapeutic management of infections caused by UPEC strains, however, some research (like cytotoxicity test or EOP assay) is still required to confirm it.

9. Future perspectives

Phage therapy offers a promising alternative for treating urinary tract infections caused by uropathogenic *Escherichia coli* bacteria, especially strains resistant to conventional antibiotics. Enhancements in phage isolation, genome sequencing, and its content verification, as well as phage cocktail design, may expand the therapeutic application of phages and improve the safety of phage therapy. Detailed phage characterization, research on phage–host interactions, including effects on human immunity and microbiota, will be crucial for reliable selection of optimal candidates for therapy. Standardization of testing methods and clinical protocols will be essential to ensure reproducibility and regulatory approval. With continued development and improvement, phage therapy could become an effective, personalized treatment option for recurrent and urinary tract infections caused by multidrug-resistant UPEC.

CRediT authorship contribution statement

Aleksandra Łukasiak: Writing – original draft, Visualization, Investigation, Funding acquisition, Data curation. **Wojciech Wesołowski:** Writing – original draft, Visualization, Methodology, Investigation, Funding acquisition, Data curation. **Julia Neumann:** Visualization, Data curation. **Natalia Lewandowska:** Writing – review & editing, Data curation. **Emilia Węglińska:** Writing – review & editing, Data curation. **Sylwia Bloch:** Writing – original draft, Supervision. **Grzegorz Węgrzyn:** Writing – review & editing, Formal analysis. **Bożena Nejman-Faleńczyk:** Writing – review & editing, Supervision, Formal analysis, Conceptualization.

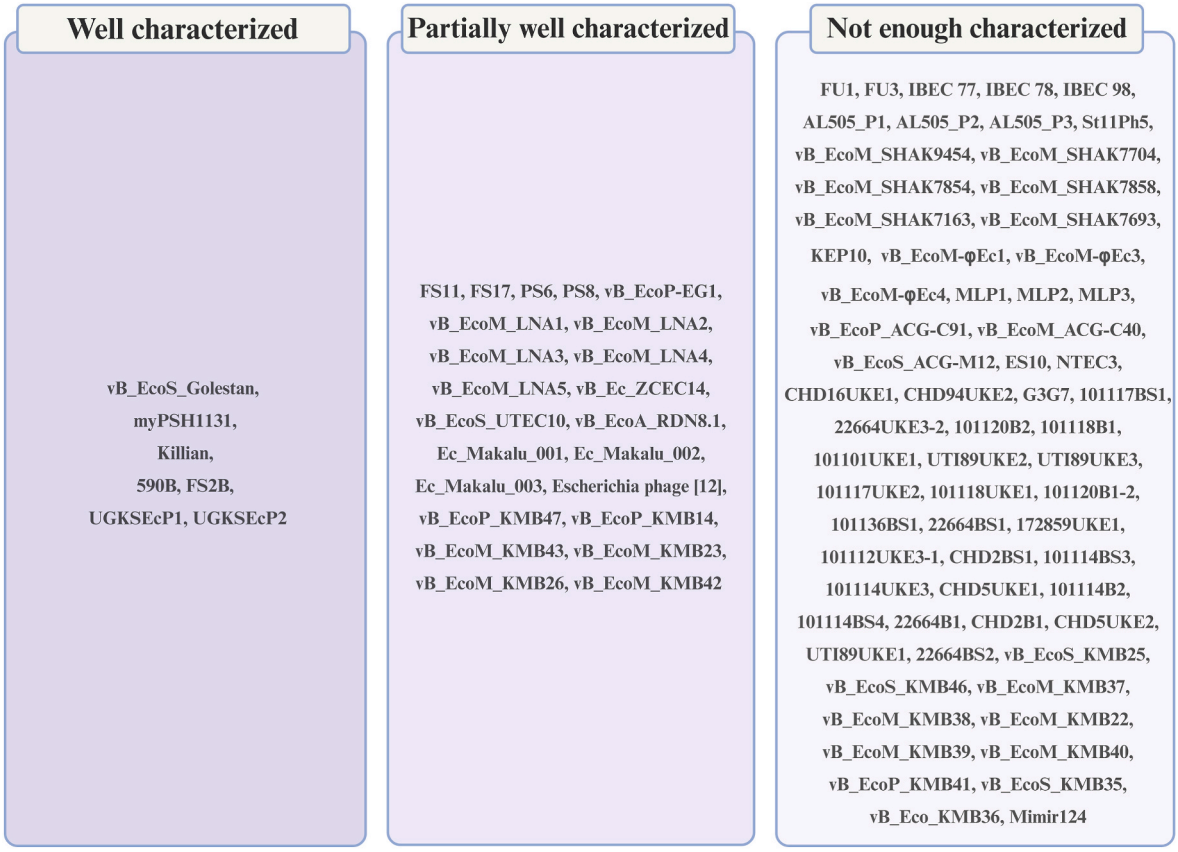


Fig. 9. Categorization of the reported anti-UPEC phages based on their level of characterization. Well-characterized phages were defined when four different parameters, such as host range, kinetics, temperature and pH stability, and genome content, were analyzed. Phages missing one of these results were considered as partially well-characterized. Phages lacking two or more of these experiments were deemed insufficiently characterized.

Funding

This research was funded by the National Science Centre, grant number 2023/49/N/NZ9/04058 to A.Ł., and by the UGrants-start grant number 533-BG10-GS0I-24 to W.W.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.virol.2025.110738.

Data availability

No data was used for the research described in the article.

References

Abduzaimovic, A., Aljicevic, M., Rebic, V., Vranic, S., Abduzaimovic, K., Sestic, S., 2016. Antibiotic resistance in urinary isolates of *Escherichia coli*. Mater. Sociomed. 28, 416–419. https://doi.org/10.5455/msm.2016.28.416-419.

Abernethy, J., Guy, R., Sheridan, E.A., Hopkins, S., Kiernan, M., Wilcox, M.H., Johnson, A.P., Hope, R., 2017. Epidemiology of *Escherichia coli* bacteraemia in England: results of an enhanced sentinel surveillance programme. J. Hosp. Infect. 95, 365–375. https://doi.org/10.1016/j.jhin.2016.12.008.

Alameer, K.M., Abuageelah, B.M., Alharbi, R.H., Alfaifi, M.H., Hurissi, E., Haddad, M., Dhayhi, N., Jafar, A.S., Mobarki, M., Awashi, H., Musawi, S., Alameer, A.M., Kariri, S.H., Alhazmi, A.H., 2025. Retrospective analysis of antibiotic resistance patterns of uropathogenic *Escherichia coli* with focus on extended-spectrum β-lactamase at a tertiary central hospital in Saudi Arabia. Health Sci. Rep. 8, e70378. https://doi.org/10.1002/hsr2.70378.

Alcock, B.P., Huynh, W., Chalil, R., Smith, K.W., Raphenya, A.R., Wlodarski, M.A., Edalatmand, A., Petkau, A., Syed, S.A., Tsang, K.K., Baker, S.J.C., Dave, M., McCarthy, M.C., Mukiri, K.M., Nasir, J.A., Golbon, B., Imtiaz, H., Jiang, X., Kaur, K., Kwong, M., Liang, Z.C., Niu, K.C., Shan, P., Yang, J.Y.J., Gray, K.L., Hoad, G.R., Jia, B., Bhando, T., Carfrae, L.A., Farha, M.A., French, S., Gordzevich, R., Rachwalski, K., Tu, M.M., Bordeleau, E., Dooley, D., Griffiths, E., Zubyk, H.L., Brown, E.D., Maguire, F., Beiko, R.G., Hsiao, W.W.L., Brinkman, F.S.L., Van Domselaar, G., McArthur, A.G., 2023. CARD 2023: expanded curation, support for machine learning, and resistome prediction at the comprehensive antibiotic resistance database. Nucleic Acids Res. 51, 690–699. https://doi.org/10.1093/nar/gkac920.

Altenhoff, A.M., Levy, J., Zarowiecki, M., Tomiczek, B., Warwick Vesztrocy, A., Dalquen, D.A., Müller, S., Telford, M.J., Glover, N.M., Dylus, D., Dessimoz, C., 2019. OMA standalone: orthology inference among public and custom genomes and transcriptomes. Genome Res. 29, 1152–1163. https://doi.org/10.1101/gr.243212.118.

Arndt, D., Grant, J.R., Marcu, A., Sajed, T., Pon, A., Liang, Y., Wishart, D.S., 2016. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res. 44, 16–21. https://doi.org/10.1093/nar/gkw387.

Asgharzadeh Kangachar, S., Logel, D.Y., Trofimova, E., Zhu, H.X., Zaugg, J., Schembri, M.A., Weynberg, K.D., Jaschke, P.R., 2024. Discovery and characterisation of new phage targeting uropathogenic *Escherichia coli*. Virology 597, 110148. https://doi.org/10.1016/j.virol.2024.110148.

Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., Meyer, F., Olsen, G.J., Olson, R., Osterman, A.L., Overbeek, R.A., McNeil, L.K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G.D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., Zagnitko, O., 2008. The RAST server: rapid annotations using subsystems technology. BMC Genom. 9, 75. https://doi.org/10.1186/1471-2164-9-75.

Bortolaia, V., Kaas, R.S., Ruppe, E., Roberts, M.C., Schwarz, S., Cattoir, V., Philippon, A., Allesoe, R.L., Rebelo, A.R., Florensa, A.F., Fagelhauer, L., Chakraborty, T., Neumann, B., Werner, G., Bender, J.K., Stingl, K., Nguyen, M., Coppens, J., Xavier, B. B., Malhotra-Kumar, S., Westh, H., Pinholt, M., Anjum, M.F., Duggett, N.A., Kempf, I., Nykäsenoja, S., Olkkola, S., Wieczorek, K., Amaro, A., Clemente, L., Mossong, J., Losch, S., Ragimbeau, C., Lund, O., Aarestrup, F.M., 2020. ResFinder 4.0 for predictions of phenotypes from genotypes. J. Antimicrob. Chemother. 75, 3491–3500. https://doi.org/10.1093/jac/dkaa345.

Bouras, G., Nepal, R., Houtak, G., Psaltis, A.J., Wormald, P.-J., Vreugde, S., 2023. Pharokka: a fast scalable bacteriophage annotation tool. Bioinformatics 39, 776. https://doi.org/10.1093/bioinformatics/btac776.

Carlet, J., Collignon, P., Goldmann, D., Goossens, H., Gyssens, I.C., Harbarth, S., Jarlier, V., Levy, S.B., N'Doye, B., Pittet, D., Richtmann, R., Seto, W.H., van der Meer, J.W., Voss, A., 2011. Society's failure to protect a precious resource: antibiotics. Lancet 378, 369–371. https://doi.org/10.1016/S0140-6736(11)60401-7.

Chaudhary, N., Maurya, R.K., Singh, D., Mohan, B., Taneja, N., 2022a. Genome analysis and antibiofilm activity of phage 590B against multidrug-resistant and extensively drug-resistant uropathogenic *Escherichia coli* isolates, India. Pathogens 11, 1448. https://doi.org/10.3390/pathogens11121448.

Chaudhary, N., Mohan, B., Mavuduru, R.S., Kumar, Y., Taneja, N., 2022b. Characterization, genome analysis and *in vitro* activity of a novel phage vB_EcoA_RDN8.1 active against multi-drug resistant and extensively drug-resistant biofilm-forming uropathogenic *Escherichia coli* isolates, India. J. Appl. Microbiol. 132, 3387–3404. https://doi.org/10.1111/jam.15439.

Chaudhary, N., Singh, D., Maurya, R.K., Mohan, B., Mavuduru, R.S., Taneja, N., 2022c. Whole genome sequencing and *in vitro* activity data of *Escherichia* phage NTEC3 against multidrug-resistant uropathogenic and extensively drug-resistant uropathogenic *E. coli* isolates. Data Brief 43, 108479. https://doi.org/10.1016/j.dib.2022.108479.

Chibeu, A., Lingohr, E.J., Masson, L., Manges, A., Harel, J., Ackermann, H.-W., Kropinski, A.M., Boerlin, P., 2012. Bacteriophages with the ability to degrade uropathogenic *Escherichia coli* biofilms. Viruses 4, 471–487. https://doi.org/10.3390/v4040471.

Ciontea, A.S., Cristea, D., Andrei, M.M., Popa, A., Usein, C.R., 2018. In vitro antimicrobial resistance of urinary *Escherichia coli* isolates from outpatients collected in a laboratory during two years, 2015–2017. Roum. Arch. Microbiol. Immunol. 77, 28–32.

Devi, B.A., Leela, K.V., Sugumar, S., 2024. Isolation and characterization of bacteriophages against *E. coli* urinary tract infection and evaluating their anti-biofilm activity and antibiotic synergy. Microb. Pathog. 193, 106789. https://doi.org/10.1016/j.micpath.2024.106789.

Dhungana, G., Nepal, R., Houtak, G., Bouras, G., Vreugde, S., Malla, R., 2024. Preclinical characterization and *in silico* safety assessment of three virulent bacteriophages targeting carbapenem-resistant uropathogenic *Escherichia coli*. Int. Microbiol. 27, 1747–1763. https://doi.org/10.1007/s10123-024-00508-8.

Feldgarden, M., Brover, V., Gonzalez-Escalona, N., Frye, J.G., Haendiges, J., Haft, D.H., Hoffmann, M., Pettengill, J.B., Prasad, A.B., Tillman, G.E., Tyson, G.H., Klimke, W., 2021. AMRFinderPlus and the reference gene catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. Sci. Rep. 11, 12728. https://doi.org/10.1038/s41598-021-91456-0.

Galtier, M., De Sordi, L., Maura, D., Arachchi, H., Volant, S., Dillies, M.-A., Debarbieux, L., 2016. Bacteriophages to reduce gut carriage of antibiotic resistant uropathogens with low impact on microbiota composition. Environ. Microbiol. 18, 2237–2245. https://doi.org/10.1111/1462-2920.13284.

Golomidova, A., Kupriyanov, Y., Gabdrakhmanov, R., Gurkova, M., Kulikov, E., Belalov, I., Uskevich, V., Bespiatykh, D., Letarova, M., Efimov, A., Kuznetsov, A., Shitikov, E., Pushkar, D., Letarov, A., Zurabov, F., 2024. Isolation, characterization, and unlocking the potential of Mimir124 phage for personalized treatment of difficult, multidrug-resistant uropathogenic *E. coli* strain. Int. J. Mol. Sci. 25, 12755. https://doi.org/10.3390/ijms252312755.

Golomidova, A.K., Kulikov, E.E., Babenko, V.V., Kostryukova, E.S., Letarov, A.V., 2018. Complete genome sequence of bacteriophage St11Ph5, which infects uropathogenic *Escherichia coli* strain up11. Genome Announc. 6, e01371. https://doi.org/10.1128/genomeA.01371-17.

González-Villalobos, E., Ribas-Aparicio, R.M., Montealegre, G.E.R., Belmont-Monroy, L., Ortega-García, Y., Aparicio-Ozores, G., Balcázar, J.L., Eslava-Campos, C.A., Hernández-Chiñas, U., Molina-López, J., 2021. |Isolation and characterization of novel bacteriophages as a potential therapeutic option for *Escherichia coli* urinary tract infections. Appl. Microbiol. Biotechnol. 105, 5617–5629. https://doi.org/10.1007/s00253-021-11432-6.

Gu, Y., Xu, Y., Xu, J., Yu, X., Huang, X., Liu, G., Liu, X., 2019. Identification of novel bacteriophage vB_EcoP-EG1 with lytic activity against planktonic and biofilm forms of uropathogenic *Escherichia coli*. Appl. Microbiol. Biotechnol. 103, 315–326. https://doi.org/10.1007/s00253-018-9471-x.

Hariati, H., Suza, D.E., Tarigan, R., 2019. Risk factors analysis for catheter-associated urinary tract infection in medan, Indonesia. Open Access Maced. J. Med. Sci. 7, 3189–3194. https://doi.org/10.3889/oamjms.2019.798.

Hejazi Dehaghani, S., Raheem, H.Q., Latifpour, M., Hallaji, M., 2025. Antibiotic resistance and virulence patterns of O25 and O16 serogroups in uropathogenic *Escherichia coli*. BMC Res. Notes 18, 157. https://doi.org/10.1186/s13104-025-07192-5.

Hockenberry, A.J., Wilke, C.O., 2021. BACPHLIP: predicting bacteriophage lifestyle from conserved protein domains. PeerJ 9, e11396. https://doi.org/10.7717/peerj.11396.

Ismael, N.M., Azzam, M., Abdelmoteleb, M., El-Shibiny, A., 2024. Phage vB_Ec_ZCEC14 to treat antibiotic-resistant *Escherichia coli* isolated from urinary tract infections. Virol. J. 21, 44. https://doi.org/10.1186/s12985-024-02306-0.

Joensen, K.G., Scheutz, F., Lund, O., Hasman, H., Kaas, R.S., Nielsen, E.M., Aarestrup, F. M., 2014. Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. J. Clin. Microbiol. 52, 1501–1510. https://doi.org/10.1128/JCM.03617-13.

Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.Y., Lopez, R., Hunter, S., 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 1 (30), 1236–1240. https://doi.org/10.1093/bioinformatics/btu031.

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., Sternberg, M.J.E., 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nat. Protoc. 10, 845–858. https://doi.org/10.1038/nprot.2015.053.

Khunti, P., Chantakorn, K., Tantibhadrasapa, A., Htoo, H.H., Thiennimitr, P., Nonejuie, P., Chaikeeratisak, V., 2023. A novel coli myophage and antibiotics synergistically inhibit the growth of the uropathogenic *E. coli* strain CFT073 in stoichiometric niches. Microbiol. Spectr. 11, e0088923. https://doi.org/10.1128/spectrum.00889-23.

Kot, B., Wicha, J., Gruzewska, A., Piechota, M., Wolska, K., Obrębska, M., 2016. Virulence factors, biofilm-forming ability, and antimicrobial resistance of urinary *Escherichia coli* strains isolated from hospitalized patients. Turk. J. Med. Sci. 46, 1908–1914. https://doi.org/10.3906/sag-1508-105.

Kot, B., 2019. Antibiotic resistance among uropathogenic *Escherichia coli*. Pol. J. Microbiol. 68, 403–415. https://doi.org/10.33073/pjm-2019-048.

Lagesen, K., Hallin, P., Rødland, E.A., Staerfeldt, H.-H., Rognes, T., Ussery, D.W., 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 35, 3100–3108. https://doi.org/10.1093/nar/gkm160.

Laslett, D., Canback, B., 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res. 32, 11–16. https://doi.org/10.1093/nar/gkh152.

Latka, A., Leiman, P.G., Drulis-Kawa, Z., Briers, Y., 2019. Modeling the architecture of depolymerase-containing receptor binding proteins in *Klebsiella* phages. Front. Microbiol. 10, 2649. https://doi.org/10.3389/fmicb.2019.02649.

Lavigne, R., Sun, W.D., Volckaert, G., 2004. PHIRE, a deterministic approach to reveal regulatory elements in bacteriophage genomes. Bioinformatics 20, 629–635. https://doi.org/10.1093/bioinformatics/btg456.

Lavigne, J.P., Bruyère, F., Bernard, L., Combescure, C., Ronco, E., Lanotte, P., Coloby, P., Thibault, M., Cariou, G., Des-places, N., Costa, P., Sotto, A., 2016. Resistance and virulence potential of uropathogenic *Escherichia coli* strains isolated from patients hospitalized in urology departments: a French prospective multicentre study. J. Med. Microbiol. 65, 530–537. https://doi.org/10.1099/jmm.0.000247.

Loose, M., Sáez Moreno, D., Mutti, M., Hitzenhammer, E., Visram, Z., Dippel, D., Schertler, S., Tišáková, L.P., Wittmann, J., Corsini, L., Wagenlehner, F., 2021. Natural bred ε2-phages have an improved host range and virulence against uropathogenic *Escherichia coli* over their ancestor phages. Antibiotics (Basel) 10, 1337. https://doi.org/10.3390/antibiotics10111337.

Lowe, T.M., Chan, P.P., 2016. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Res. 44, 54–57. https://doi.org/10.1093/nar/gkw413.

Lu, H., Li, Z., Elbaz, A., Ni, S.-Q., 2023. Synergistic action of phages and lytic proteins with antibiotics: a combination strategy to target bacteria and biofilms. BMC Microbiol. 23, 149. https://doi.org/10.1186/s12866-023-02881-2.

Malik, S., Nehra, K., Mann, A., Jagdish, R., Rana, J.S., 2024. Characterization and synergy studies of Caudoviricete *Escherichia* phage FS2B infecting multi-drug resistant uropathogenic *Escherichia coli* isolates. Int. Microbiol. 27, 155–166. https://doi.org/10.1007/s10123-023-00381-x.

Malik, S., Nehra, K., Rana, J.S., 2021. Bacteriophage cocktail and phage antibiotic synergism as promising alternatives to conventional antibiotics for the control of multi-drug-resistant uropathogenic *Escherichia coli*. Virus Res. 302, 198496. https://doi.org/10.1016/j.virusres.2021.198496.

Manohar, P., Tamhankar, A.J., Lundborg, C.S., Ramesh, N., 2018. Isolation, characterization and *in vivo* efficacy of *Escherichia* phage myPSH1131. PLoS One 13, e0206278. https://doi.org/10.1371/journal.pone.0206278.

Markusková, B., Elnwrani, S., Andrezál, M., Sedláčková, T., Szemes, T., Slobodníková, L., Kajsik, M., Drahovská, H., 2024. Characterization of bacteriophages infecting multidrug-resistant uropathogenic *Escherichia coli* strains. Arch. Virol. 169, 142. https://doi.org/10.1007/s00705-024-06063-x.

McNair, K., Bailey, B.A., Edwards, R.A., 2012. PHACTS, a computational approach to classifying the lifestyle of phages. Bioinformatics 28, 614–618. https://doi.org/10.1093/bioinformatics/bts014.

Mijbel Ali, B., Gatea Kaabi, S.A., Al-Bayati, M.A., Musafer, H.K., 2021. A novel phage cocktail therapy of the urinary tract infection in a mouse model. Arch Razi Inst. 76, 1229–1236. https://doi.org/10.22092/ari.2021.356004.1762.

Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., Lanfear, R., 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. 37, 1530–1534. https://doi.org/10.1093/molbev/msaa015.

Nishimura, Y., Yoshida, T., Kuronishi, M., Uehara, H., Ogata, H., Goto, S., 2017. ViPTree: the viral proteomic tree server. Bioinformatics 33, 2379–2380. https://doi.org/10.1093/bioinformatics/btx157.

Necel, A., Bloch, S., Topka-Bielecka, G., Janiszewska, A., Łukasiak, A., Nejman-Faleńczyk, B., Węgrzyn, G., 2022. Synergistic effects of bacteriophage vB_Eco4-M7 and selected antibiotics on the biofilm formed by Shiga toxin-producing *Escherichia coli*. Antibiotics 11, 712. https://doi.org/10.3390/antibiotics11060712.

Necel, A., Dydecka, A., Topka-Bielecka, G., Wesołowski, W., Lewandowska, N., Bloch, S., Nejman-Faleńczyk, B., 2025. What, how, and why? - anti-ehec phages and their application potential in medicine and food industry. J. Appl. Genet. 66, 219–240. https://doi.org/10.1007/s13353-024-00918-4.

Nawaz, A., Zafar, S., Alessa, A.H., Khalid, N.A., Shahzadi, M., Majid, A., Badshah, M., Shah, A.A., Khan, S., 2024. Characterization of ES10 lytic bacteriophage isolated from hospital waste against multidrug-resistant uropathogenic *E. coli*. Front. Microbiol. 15, 1320974. https://doi.org/10.3389/fmicb.2024.1320974.

Ngiam, L., Schembri, M.A., Weynberg, K., Guo, J., 2021. Bacteriophage isolated from non-target bacteria demonstrates broad host range infectivity against multidrug-

resistant bacteria. Environ. Microbiol. 23, 5569–5586. https://doi.org/10.1111/1462-2920.15714.
Nishikawa, H., Yasuda, M., Uchiyama, J., Rashel, M., Maeda, Y., Takemura, I., Sugihara, S., Ujihara, T., Shimizu, Y., Shuin, T., Matsuzaki, S., 2008. T-even-related bacteriophages as candidates for treatment of *Escherichia coli* urinary tract infections. Arch. Virol. 153, 507–515. https://doi.org/10.1007/s00705-007-0031-4.
Pallett, A., Hand, K., 2010. Complicated urinary tract infections: practical solutions for the treatment of multiresistant Gram-negative bacteria. J. Antimicrob. Chemother. 65 (Suppl. 3), iii25–iii33. https://doi.org/10.1093/jac/dkq298.
Podlacha, M., Gaffke, L., Grabowski, Ł., Mantej, J., Grabski, M., Pierzchalska, M., Pierzynowska, K., Węgrzyn, G., Węgrzyn, A., 2024. Bacteriophage DNA induces an interrupted immune response during phage therapy in a chicken model. Nat. Commun. 15, 2274. https://doi.org/10.1038/s41467-024-46555-7.
Rahimzadeh, G., Resch, G., Rezai, M.S., Nemati Hevelaee, E., 2020. Characterization and lytic activity of isolated *Escherichia coli* bacteriophages against *Escherichia coli in vitro*. Iran. J. Med. Sci. 45, 298–303. https://doi.org/10.30476/ijms.2019.45420.
Rajab, A.A.H., Fahmy, E.K., Esmaeel, S.E., Yousef, N., Askoura, M., 2024. *In vitro* and *in vivo* assessment of the competence of a novel lytic phage vB_EcoS_UTEC10 targeting multidrug resistant *Escherichia coli* with a robust biofilm eradication activity. Microb. Pathog. 197, 107058. https://doi.org/10.1016/j.micpath.2024.107058.
Sabih, A., Leslie, S.W., 2023. NICE guidance - complicated urinary tract infections: ceftolozane/tazobactam. BJU Int. 121, 825–834. https://doi.org/10.1111/bju.14364.
Sampaio, M., Rocha, M., Oliveira, H., Dias, O., 2019. Predicting promoters in phage genomes using PhagePromoter. Bioinformatics 35, 5301–5302. https://doi.org/10.1093/bioinformatics/btz580.
Sanmukh, S.G., Admella, J., Moya-Andérico, L., Fehér, T., Arévalo-Jaimes, B.V., Blanco-Cabra, N., Torrents, E., 2023. Accessing the *in vivo* efficiency of clinically isolated phages against uropathogenic and invasive biofilm-forming *Escherichia coli* strains for phage therapy. Cells 12, 344. https://doi.org/10.3390/cells12030344.
Shang, J., Tang, X., Sun, Y., 2023. PhaTYP: predicting the lifestyle for bacteriophages using BERT. Briefings Bioinf. 24, bbac487. https://doi.org/10.1093/bib/bbac487.
Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J.D., Higgins, D.G., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. Mol. Syst. Biol. 7, 539. https://doi.org/10.1038/msb.2011.75.
Slobodníková, L., Markusková, B., Kajsík, M., Andrezál, M., Straka, M., Liptáková, A., Drahovská, H., 2021. Characterization of anti-bacterial effect of the two new phages against uropathogenic *Escherichia coli*. Viruses 13, 1348. https://doi.org/10.3390/v13071348.
Ssekatawa, K., Ntulume, I., Byarugaba, D.K., Michniewski, S., Jameson, E., Wampande, E.M., Nakavuma, J., 2024. Isolation and characterization of novel lytic bacteriophages infecting carbapenem-resistant pathogenic diarrheagenic and uropathogenic *Escherichia coli*. Infect. Drug Resist. 17, 3367–3384. https://doi.org/10.2147/IDR.S466101.
Sufa, H.I., Budiarti, S., Rusmana, I., 2018. Diversity of uropathogenic *Escherichia coli* lytic phage from Cisadane river, West Java, Indonesia based on morphology and protein molecular weight characteristics. Biodivers. J. Biol. Divers. 19, 2359–2364. https://doi.org/10.13057/biodiv/d190646.
Topka, G., Bloch, S., Nejman-Faleńczyk, B., Gąsior, T., Jurczak-Kurek, A., Necel, A., Dydecka, A., Richert, M., Węgrzyn, G., Węgrzyn, A., 2019. Characterization of bacteriophage vB-EcoS-95, isolated from urban sewage and revealing extremely rapid lytic development. Front. Microbiol. 9, 3326. https://doi.org/10.3389/fmicb.2018.03326.
Topka-Bielecka, G., Dydecka, A., Necel, A., Bloch, S., Nejman-Faleńczyk, B., Węgrzyn, G., Węgrzyn, A., 2021a. Bacteriophage-derived depolymerases against bacterial biofilm. Antibiotics 10, 175. https://doi.org/10.3390/antibiotics10020175.
Topka-Bielecka, G., Nejman-Faleńczyk, B., Bloch, S., Dydecka, A., Necel, A., Węgrzyn, A., Węgrzyn, G., 2021b. Phage-bacteria interactions in potential applications of bacteriophage vB_EfaS-271 against *Enterococcus faecalis*. Viruses 13, 318. https://doi.org/10.3390/v13020318.
Vera-Mansilla, J., Sánchez, P., Silva-Valenzuela, C.A., Molina-Quiroz, R.C., 2022. Isolation and characterization of novel lytic phages infecting multidrug-resistant *Escherichia coli*. Microbiol. Spectr. 10, e0167821. https://doi.org/10.1128/spectrum.01678-21.
Wesołowski, W., Łukasiak, A., Bloch, S., Kuligowska, K., Neumann, J., Lewandowska, N., Węglińska, E., Węgrzyn, G., Nejman-Faleńczyk, B., 2025. Phage endolysins as promising and effective candidates for use against uropathogenic *Escherichia coli*. Viruses 17, 560. https://doi.org/10.3390/v17040560.
Whelan, S., Lucey, B., Finn, K., 2023. Uropathogenic *Escherichia coli* (UPEC)-associated urinary tract infections: the molecular basis for challenges to effective treatment. Microorganisms 11, 2169. https://doi.org/10.3390/microorganisms11092169.
Yazdi, M., Bouzari, M., Ghaemi, E.A., Shahin, K., 2020. Isolation, characterization and genomic analysis of a novel bacteriophage vB_EcoS-Golestan infecting multidrug-resistant *Escherichia coli* isolated from urinary tract infection. Sci. Rep. 10, 7690. https://doi.org/10.1038/s41598-020-63048-x.
Yukgehnaish, K., Rajandas, H., Parimannan, S., Manickam, R., Marimuthu, K., Petersen, B., Clokie, M.R.J., Millard, A., Sicheritz-Pontén, T., 2022. PhageLeads: rapid assessment of phage therapeutic suitability using an ensemble machine learning approach. Viruses 14, 342. https://doi.org/10.3390/v14020342.
Zalewska-Piątek, B., Piątek, R., 2020. Phage therapy as a novel strategy in the treatment of urinary tract infections caused by *E. coli*. Antibiotics 9, 304. https://doi.org/10.3390/antibiotics9060304.
Zhou, Y., Liang, Y., Lynch, K.H., Dennis, J.J., Wishart, D.S., 2011. PHAST: a fast phage search tool. Nucleic Acids Res. 39, 347–352. https://doi.org/10.1093/nar/gkr485.
Zowawi, H.M., Harris, P.N.A., Roberts, M.J., Tambyah, P.A., Schembri, M.A., Pezzani, M. D., Williamson, D.A., Paterson, D.L., 2015. The emerging threat of multidrug-resistant Gram-negative bacteria in urology. Nat. Rev. Urol. 12, 570–584. https://doi.org/10.1038/nrurol.2015.199.

Suplement do pracy nr 1

Supplementary Figure S1. The number of phages tested in different temperature ranges (Panel A) and pH values (Panel B). If a phage was tested in two temperature ranges/pH values, it was included in both bars to show the total number of viruses investigated under particular conditions.

Supplementary Figure S2. The number of bacterial strains infected by the analyzed phages. The analysis was performed based on the spot test results, since most phages were assayed using this method. As the EOP assays were performed for only 42% of the tested here phages, these results were not considered in this analysis.

Supplementary Table S1. The lists of newly annotated genes for phages: vB_EcoM_LNA1, vB_EcoM_LNA2, vB_EcoP_LNA3, vB_EcoM_LNA4, vB_EcoP_LNA5, vB_EcoM_SHAK7704, vB_EcoM_SHAK7854, and vB_EcoM_SHAK9454.

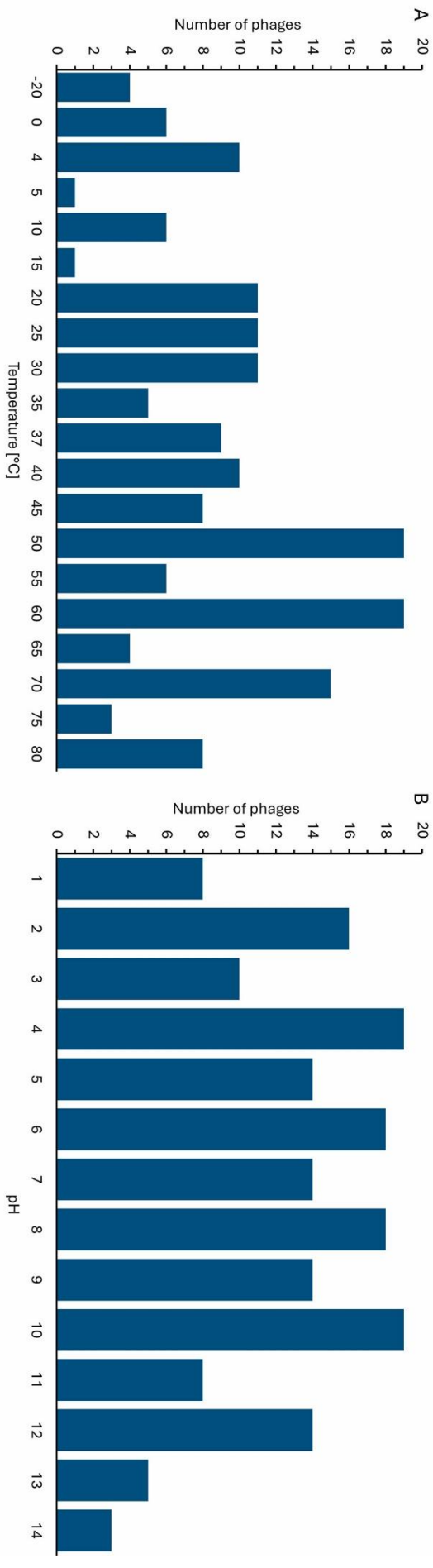
Supplementary Table S2. The list of annotated interesting phage proteins reported in genomes of 82 anti-UPEC phages, available in the NCBI database. Seventy-two of the predicted proteins were newly identified in the frame of this work.

Supplementary Table S3. The list of putative phage depolymerases newly identified in this work in genomes of the analyzed 82 anti-UPEC phages, according to the guidelines reported by Latka et al., 2019.

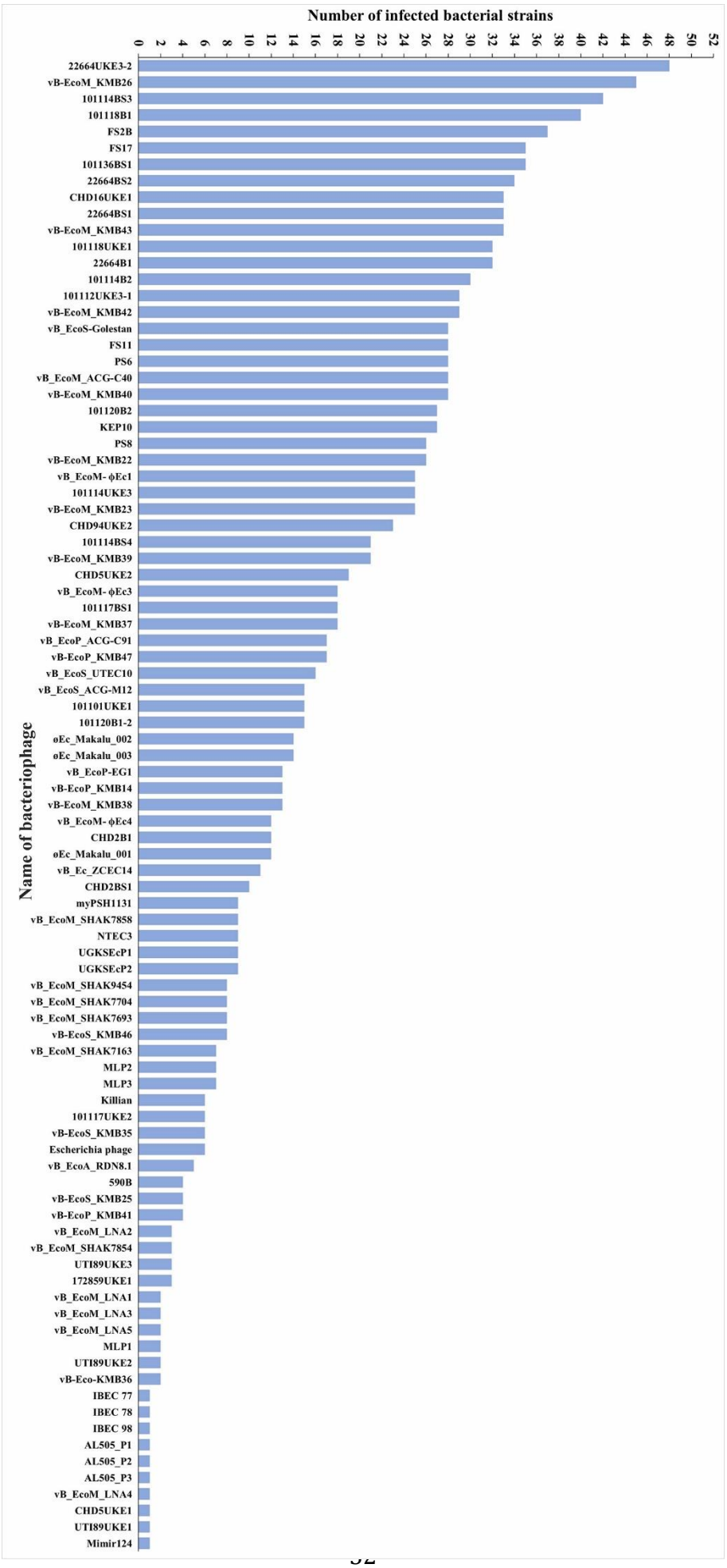
Supplementary Table S4. Results of phage development prediction, newly obtained in this work with PHACTS 1.8, BACPHLIP 0.9.6, and PhaTYP 0.3.0 tools.

Supplementary Table S5. The lists of selected genes localized in genomes of phages belonging to particular clades of the VipTree – based tree, presented in Figure 1. These phage genes were used by us to generate new multigene trees presented in Figures 2-8. Sequences of used genes were obtained from the NCBI database, and their names and positions in phage genomes were collected. Genes important for tree topology were identified in this work and marked by "+". The number of parsimony informative sites in particular alignments has been determined.

Supplementary Figure S1



Supplementary Figure S2



Supplementary Table S1

Annotation of phage vB_EcoM_LNA1.

Locus_tag	Start	End	Length_nt	Strand	Function_by_amino_acid_sequence	Protein_ID	Category
vB_EcoM_LNA1_MZ311863_CDS_0001	1	1602	1602	+	terminase large subunit	self_TKOZ1	protein
vB_EcoM_LNA1_MZ311863_CDS_0002	1619	3085	1467	+	portal protein	self_TKOZ2	protein
vB_EcoM_LNA1_MZ311863_CDS_0003	3085	3585	501	+	hypothetical protein	self_TKOZ3	protein
vB_EcoM_LNA1_MZ311863_CDS_0004	3585	3917	333	+	head protein	self_TKOZ4	protein
vB_EcoM_LNA1_MZ311863_CDS_0005	3929	5275	1347	+	head maturation protease	self_TKOZ5	protein
vB_EcoM_LNA1_MZ311863_CDS_0006	5287	5664	378	+	tail protease	self_TKOZ6	protein
vB_EcoM_LNA1_MZ311863_CDS_0007	5698	6804	1107	+	major head protein	self_TKOZ7	protein
vB_EcoM_LNA1_MZ311863_CDS_0008	6826	7275	450	+	hypothetical protein	self_TKOZ8	protein
vB_EcoM_LNA1_MZ311863_CDS_0009	7275	7757	483	+	tail completion or Neck1 protein	self_TKOZ9	protein
vB_EcoM_LNA1_MZ311863_CDS_0010	7754	8155	402	+	hypothetical protein	self_TKOZ10	protein
vB_EcoM_LNA1_MZ311863_CDS_0011	8130	8729	600	+	hypothetical protein	self_TKOZ11	protein
vB_EcoM_LNA1_MZ311863_CDS_0012	8730	10082	1353	+	tail sheath	self_TKOZ12	protein
vB_EcoM_LNA1_MZ311863_CDS_0013	10098	10544	447	+	virion structural protein	self_TKOZ13	protein
vB_EcoM_LNA1_MZ311863_CDS_0014	10618	11016	399	+	tail assembly chaperone	self_TKOZ14	protein
vB_EcoM_LNA1_MZ311863_CDS_0015	11019	11258	240	+	tail assembly chaperone	self_TKOZ15	protein
vB_EcoM_LNA1_MZ311863_CDS_0016	11258	13501	2244	+	tail length tape measure protein	self_TKOZ16	protein
vB_EcoM_LNA1_MZ311863_CDS_0017	13501	14298	798	+	hypothetical protein	self_TKOZ17	protein
vB_EcoM_LNA1_MZ311863_CDS_0018	14298	14639	342	+	virion structural protein	self_TKOZ18	protein
vB_EcoM_LNA1_MZ311863_CDS_0019	14639	15616	978	+	baseplate hub	self_TKOZ19	protein
vB_EcoM_LNA1_MZ311863_CDS_0020	15616	16239	624	+	baseplate spike	self_TKOZ20	protein
vB_EcoM_LNA1_MZ311863_CDS_0021	16239	16658	420	+	baseplate wedge subunit	self_TKOZ21	protein
vB_EcoM_LNA1_MZ311863_CDS_0022	16658	18127	1470	+	baseplate wedge subunit	self_TKOZ22	protein
vB_EcoM_LNA1_MZ311863_CDS_0023	18130	18987	858	+	baseplate protein	self_TKOZ23	protein
vB_EcoM_LNA1_MZ311863_CDS_0024	18987	19289	303	+	hypothetical protein	self_TKOZ24	protein
vB_EcoM_LNA1_MZ311863_CDS_0025	19292	20440	1149	+	tail fiber protein	self_TKOZ25	protein
vB_EcoM_LNA1_MZ311863_CDS_0026	20418	22838	2421	+	tail fiber protein	self_TKOZ26	protein
vB_EcoM_LNA1_MZ311863_CDS_0027	22918	23112	195	+	hypothetical protein	self_TKOZ27	protein
vB_EcoM_LNA1_MZ311863_CDS_0028	23113	23484	372	+	hypothetical protein	self_TKOZ28	protein
vB_EcoM_LNA1_MZ311863_CDS_0029	23521	24420	900	-	thymidylate synthase	self_TKOZ29	protein
vB_EcoM_LNA1_MZ311863_CDS_0030	24422	24967	546	-	dihydrofolate reductase	self_TKOZ30	protein
vB_EcoM_LNA1_MZ311863_CDS_0031	24964	25224	261	-	hypothetical protein	self_TKOZ31	protein
vB_EcoM_LNA1_MZ311863_CDS_0032	25225	25740	516	-	hypothetical protein	self_TKOZ32	protein
vB_EcoM_LNA1_MZ311863_CDS_0033	25754	26113	360	-	hypothetical protein	self_TKOZ33	protein
vB_EcoM_LNA1_MZ311863_CDS_0034	26115	26414	300	-	transcriptional regulator	self_TKOZ34	protein
vB_EcoM_LNA1_MZ311863_CDS_0035	26407	26622	216	-	hypothetical protein	self_TKOZ35	protein
vB_EcoM_LNA1_MZ311863_CDS_0036	26624	27007	384	-	hypothetical protein	self_TKOZ36	protein
vB_EcoM_LNA1_MZ311863_CDS_0037	27087	27194	108	-	hypothetical protein	self_TKOZ37	protein
vB_EcoM_LNA1_MZ311863_CDS_0038	27295	28395	1101	-	ATP-dependent DNA ligase	self_TKOZ38	protein
vB_EcoM_LNA1_MZ311863_CDS_0039	28392	28649	258	-	hypothetical protein	self_TKOZ39	protein
vB_EcoM_LNA1_MZ311863_CDS_0040	28609	28833	225	-	hypothetical protein	self_TKOZ40	protein
vB_EcoM_LNA1_MZ311863_CDS_0041	28856	29002	147	-	hypothetical protein	self_TKOZ41	protein
vB_EcoM_LNA1_MZ311863_CDS_0042	28999	29217	219	-	hypothetical protein	self_TKOZ42	protein
vB_EcoM_LNA1_MZ311863_CDS_0043	29214	29423	210	-	hypothetical protein	self_TKOZ43	protein
vB_EcoM_LNA1_MZ311863_CDS_0044	29475	30011	537	-	HNH endonuclease	self_TKOZ44	protein
vB_EcoM_LNA1_MZ311863_CDS_0045	29998	30465	468	-	DNA polymerase	self_TKOZ45	protein
vB_EcoM_LNA1_MZ311863_CDS_0046	30522	30710	189	+	hypothetical protein	self_TKOZ46	protein
vB_EcoM_LNA1_MZ311863_CDS_0047	30782	32965	2184	-	DNA polymerase	self_TKOZ47	protein
vB_EcoM_LNA1_MZ311863_CDS_0048	33224	33646	423	+	hypothetical protein	self_TKOZ48	protein
vB_EcoM_LNA1_MZ311863_CDS_0049	33648	34460	813	+	minor tail protein	self_TKOZ49	protein
vB_EcoM_LNA1_MZ311863_CDS_0050	34522	35265	744	+	deoxynucleoside monophosphate kinase	self_TKOZ50	protein
vB_EcoM_LNA1_MZ311863_CDS_0051	35274	35474	201	+	hypothetical protein	self_TKOZ51	protein
vB_EcoM_LNA1_MZ311863_CDS_0052	35467	37452	1986	+	DNA primase/helicase	self_TKOZ52	protein
vB_EcoM_LNA1_MZ311863_CDS_0053	37427	37708	282	+	hypothetical protein	self_TKOZ53	protein
vB_EcoM_LNA1_MZ311863_CDS_0054	37705	37854	150	+	hypothetical protein	self_TKOZ54	protein
vB_EcoM_LNA1_MZ311863_CDS_0055	37927	38784	858	+	hypothetical protein	self_TKOZ55	protein
vB_EcoM_LNA1_MZ311863_CDS_0056	38847	39887	1041	+	exonuclease	self_TKOZ56	protein
vB_EcoM_LNA1_MZ311863_CDS_0057	39877	40374	498	+	NAD synthetase	self_TKOZ57	protein
vB_EcoM_LNA1_MZ311863_CDS_0058	40396	40644	249	+	hypothetical protein	self_TKOZ58	protein
vB_EcoM_LNA1_MZ311863_CDS_0059	40620	41375	756	+	MazG-like pyrophosphatase	self_TKOZ59	protein
vB_EcoM_LNA1_MZ311863_CDS_0060	41356	41679	324	+	hypothetical protein	self_TKOZ60	protein
vB_EcoM_LNA1_MZ311863_CDS_0061	41771	42007	237	+	hypothetical protein	self_TKOZ61	protein
vB_EcoM_LNA1_MZ311863_CDS_0062	42054	44288	2235	+	ribonucleotide reductase large subunit	self_TKOZ62	protein
vB_EcoM_LNA1_MZ311863_CDS_0063	44260	44601	342	+	hypothetical protein	self_TKOZ63	protein
vB_EcoM_LNA1_MZ311863_CDS_0064	44598	45671	1074	+	ribonucleoside diphosphate reductase small subunit	self_TKOZ64	protein
vB_EcoM_LNA1_MZ311863_CDS_0065	45671	45913	243	+	thioredoxin domain	self_TKOZ65	protein
vB_EcoM_LNA1_MZ311863_CDS_0066	45906	46112	207	+	hypothetical protein	self_TKOZ66	protein
vB_EcoM_LNA1_MZ311863_CDS_0067	46161	47360	1200	+	NrdD-like anaerobic ribonucleotide reductase large subunit	self_TKOZ67	protein
vB_EcoM_LNA1_MZ311863_CDS_0068	47385	47495	111	+	hypothetical protein	self_TKOZ68	protein
vB_EcoM_LNA1_MZ311863_CDS_0069	47473	47937	465	+	HNH endonuclease	self_TKOZ69	protein

vB_EcoM_LNA1_MZ311863_CDS_0070	47978	48121	144	-	hypothetical protein	self_TKOZ70	protein
vB_EcoM_LNA1_MZ311863_CDS_0071	48193	49047	855	+	anaerobic ribonucleoside reductase large subunit	self_TKOZ71	protein
vB_EcoM_LNA1_MZ311863_CDS_0072	49110	49334	225	+	hypothetical protein	self_TKOZ72	protein
vB_EcoM_LNA1_MZ311863_CDS_0073	49348	49446	99	+	hypothetical protein	self_TKOZ73	protein
vB_EcoM_LNA1_MZ311863_CDS_0074	49449	49649	201	+	membrane protein	self_TKOZ74	protein
vB_EcoM_LNA1_MZ311863_CDS_0075	49646	50041	396	+	hypothetical protein	self_TKOZ75	protein
vB_EcoM_LNA1_MZ311863_CDS_0076	50038	50337	300	+	hypothetical protein	self_TKOZ76	protein
vB_EcoM_LNA1_MZ311863_CDS_0077	50347	50832	486	+	anaerobic ribonucleotide reductase small subunit	self_TKOZ77	protein
vB_EcoM_LNA1_MZ311863_CDS_0078	50795	51172	378	+	hypothetical protein	self_TKOZ78	protein
vB_EcoM_LNA1_MZ311863_CDS_0079	51139	51396	258	+	hypothetical protein	self_TKOZ79	protein
vB_EcoM_LNA1_MZ311863_CDS_0080	51399	51719	321	+	hypothetical protein	self_TKOZ80	protein
vB_EcoM_LNA1_MZ311863_CDS_0081	51771	52286	516	+	hypothetical protein	self_TKOZ81	protein
vB_EcoM_LNA1_MZ311863_CDS_0082	52279	52557	279	+	hypothetical protein	self_TKOZ82	protein
vB_EcoM_LNA1_MZ311863_CDS_0083	52569	53450	882	+	ribose-phosphate pyrophosphokinase	self_TKOZ83	protein
vB_EcoM_LNA1_MZ311863_CDS_0084	53496	55277	1782	+	nicotinamide phosphoribosyl transferase	self_TKOZ84	protein
vB_EcoM_LNA1_MZ311863_CDS_0085	55331	55666	336	+	hypothetical protein	self_TKOZ85	protein
vB_EcoM_LNA1_MZ311863_CDS_0086	55648	55824	177	+	hypothetical protein	self_TKOZ86	protein
vB_EcoM_LNA1_MZ311863_CDS_0087	55853	58219	2367	+	RIIA lysis inhibitor	self_TKOZ87	protein
vB_EcoM_LNA1_MZ311863_CDS_0088	58299	59408	1110	+	RIIB lysis inhibitor	self_TKOZ88	protein
vB_EcoM_LNA1_MZ311863_CDS_0089	59509	60057	549	+	unknown function	self_TKOZ89	protein
vB_EcoM_LNA1_MZ311863_CDS_0090	60035	60730	696	+	polynucleotide kinase	self_TKOZ90	protein
vB_EcoM_LNA1_MZ311863_CDS_0091	60741	61205	465	+	hypothetical protein	self_TKOZ91	protein
vB_EcoM_LNA1_MZ311863_CDS_0092	61256	62140	885	+	DNA methyltransferase	self_TKOZ92	protein
vB_EcoM_LNA1_MZ311863_CDS_0093	62152	62964	813	+	hypothetical protein	self_TKOZ93	protein
vB_EcoM_LNA1_MZ311863_CDS_0094	63011	63358	348	+	Rz-like spanin	self_TKOZ94	protein
vB_EcoM_LNA1_MZ311863_CDS_0095	63324	63521	198	+	hypothetical protein	self_TKOZ95	protein
vB_EcoM_LNA1_MZ311863_CDS_0096	63518	63787	270	+	hypothetical protein	self_TKOZ96	protein
vB_EcoM_LNA1_MZ311863_CDS_0097	63790	64113	324	+	hypothetical protein	self_TKOZ97	protein
vB_EcoM_LNA1_MZ311863_CDS_0098	64091	64318	228	+	hypothetical protein	self_TKOZ98	protein
vB_EcoM_LNA1_MZ311863_CDS_0099	64331	64531	201	+	hypothetical protein	self_TKOZ99	protein
vB_EcoM_LNA1_MZ311863_CDS_0100	64532	65317	786	+	phosphatase	self_TKOZ100	protein
vB_EcoM_LNA1_MZ311863_CDS_0101	65405	65557	153	+	hypothetical protein	self_TKOZ101	protein
vB_EcoM_LNA1_MZ311863_CDS_0102	65562	65657	96	+	hypothetical protein	self_TKOZ102	protein
vB_EcoM_LNA1_MZ311863_CDS_0103	65675	65827	153	+	hypothetical protein	self_TKOZ103	protein
vB_EcoM_LNA1_MZ311863_CDS_0104	65839	66027	189	+	hypothetical protein	self_TKOZ104	protein
vB_EcoM_LNA1_MZ311863_CDS_0105	66027	66233	207	+	hypothetical protein	self_TKOZ105	protein
vB_EcoM_LNA1_MZ311863_CDS_0106	66325	66705	381	+	hypothetical protein	self_TKOZ106	protein
vB_EcoM_LNA1_MZ311863_CDS_0107	66794	67090	297	+	hypothetical protein	self_TKOZ107	protein
vB_EcoM_LNA1_MZ311863_CDS_0108	67084	67416	333	+	hypothetical protein	self_TKOZ108	protein
vB_EcoM_LNA1_MZ311863_CDS_0109	67507	67779	273	+	hypothetical protein	self_TKOZ109	protein
vB_EcoM_LNA1_MZ311863_CDS_0110	67866	68270	405	+	hypothetical protein	self_TKOZ110	protein
vB_EcoM_LNA1_MZ311863_CDS_0111	68369	68572	204	+	hypothetical protein	self_TKOZ111	protein
vB_EcoM_LNA1_MZ311863_CDS_0112	68658	69200	543	+	hypothetical protein	self_TKOZ112	protein
vB_EcoM_LNA1_MZ311863_CDS_0113	69267	69500	234	+	hypothetical protein	self_TKOZ113	protein
vB_EcoM_LNA1_MZ311863_CDS_0114	69592	69933	342	+	hypothetical protein	self_TKOZ114	protein
vB_EcoM_LNA1_MZ311863_CDS_0115	70008	70124	117	-	hypothetical protein	self_TKOZ115	protein
vB_EcoM_LNA1_MZ311863_CDS_0116	70156	70686	531	+	hypothetical protein	self_TKOZ116	protein
vB_EcoM_LNA1_MZ311863_CDS_0117	70750	71013	264	+	hypothetical protein	self_TKOZ117	protein
vB_EcoM_LNA1_MZ311863_CDS_0118	71079	71327	249	+	hypothetical protein	self_TKOZ118	protein
vB_EcoM_LNA1_MZ311863_CDS_0119	71286	71414	129	-	hypothetical protein	self_TKOZ119	protein
vB_EcoM_LNA1_MZ311863_CDS_0120	72281	72412	132	-	hypothetical protein	self_TKOZ120	protein
vB_EcoM_LNA1_MZ311863_CDS_0121	72838	72975	138	-	hypothetical protein	self_TKOZ121	protein
vB_EcoM_LNA1_MZ311863_CDS_0122	73012	73560	549	-	hypothetical protein	self_TKOZ122	protein
vB_EcoM_LNA1_MZ311863_CDS_0123	73574	74038	465	-	hypothetical protein	self_TKOZ123	protein
vB_EcoM_LNA1_MZ311863_CDS_0124	74091	74315	225	-	hypothetical protein	self_TKOZ124	protein
vB_EcoM_LNA1_MZ311863_CDS_0125	74315	74410	96	-	hypothetical protein	self_TKOZ125	protein
vB_EcoM_LNA1_MZ311863_CDS_0126	74392	74895	504	-	hypothetical protein	self_TKOZ126	protein
vB_EcoM_LNA1_MZ311863_CDS_0127	74892	75110	219	-	hypothetical protein	self_TKOZ127	protein
vB_EcoM_LNA1_MZ311863_CDS_0128	75194	75775	582	-	hypothetical protein	self_TKOZ128	protein
vB_EcoM_LNA1_MZ311863_CDS_0129	75775	76119	345	-	hypothetical protein	self_TKOZ129	protein
vB_EcoM_LNA1_MZ311863_CDS_0130	76112	76405	294	-	hypothetical protein	self_TKOZ130	protein
vB_EcoM_LNA1_MZ311863_CDS_0131	76405	76800	396	-	hypothetical protein	self_TKOZ131	protein
vB_EcoM_LNA1_MZ311863_CDS_0132	76793	77191	399	-	hypothetical protein	self_TKOZ132	protein
vB_EcoM_LNA1_MZ311863_CDS_0133	77244	77708	465	-	endolysin	self_TKOZ133	protein
vB_EcoM_LNA1_MZ311863_CDS_0134	77708	79093	1386	-	tail protein	self_TKOZ134	protein
vB_EcoM_LNA1_MZ311863_CDS_0135	79090	79461	372	-	hypothetical protein	self_TKOZ135	protein
vB_EcoM_LNA1_MZ311863_CDS_0136	79510	80100	591	-	hypothetical protein	self_TKOZ136	protein
vB_EcoM_LNA1_MZ311863_CDS_0137	80094	80327	234	-	hypothetical protein	self_TKOZ137	protein
vB_EcoM_LNA1_MZ311863_CDS_0138	80386	80514	129	+	hypothetical protein	self_TKOZ138	protein
vB_EcoM_LNA1_MZ311863_CDS_0139	80720	81091	372	+	hypothetical protein	self_TKOZ139	protein
vB_EcoM_LNA1_MZ311863_CDS_0140	81173	82588	1416	+	RNA-binding protein	self_TKOZ140	protein

vB_EcoM_LNA1_MZ311863_CDS_0141	82620	82709	90	-	hypothetical protein	self_TKOZ141	protein
vB_EcoM_LNA1_MZ311863_tRNA_0001	82755	82831	77	+	tRNA-Pro(TGG)		tRNA
vB_EcoM_LNA1_MZ311863_CDS_0142	82834	83049	216	-	hypothetical protein	self_TKOZ142	protein
vB_EcoM_LNA1_MZ311863_tRNA_0002	82839	82916	78	+	tRNA-Glu(TTC)		tRNA
vB_EcoM_LNA1_MZ311863_tRNA_0003	83169	83245	77	+	tRNA-Asn(GTT)		tRNA
vB_EcoM_LNA1_MZ311863_tRNA_0004	83255	83342	88	+	tRNA-Tyr(GTA)		tRNA
vB_EcoM_LNA1_MZ311863_CDS_0143	83469	83588	120	+	hypothetical protein	self_TKOZ143	protein
vB_EcoM_LNA1_MZ311863_CDS_0144	83588	83689	102	+	hypothetical protein	self_TKOZ144	protein
vB_EcoM_LNA1_MZ311863_tRNA_0005	83771	83846	76	+	tRNA-Lys(TTT)		tRNA
vB_EcoM_LNA1_MZ311863_tRNA_0006	83851	83927	77	+	tRNA-Met(CAT)		tRNA
vB_EcoM_LNA1_MZ311863_tRNA_0007	83930	84005	76	+	tRNA-Ile(GAT)		tRNA
vB_EcoM_LNA1_MZ311863_CDS_0145	84075	84215	141	-	hypothetical protein	self_TKOZ145	protein
vB_EcoM_LNA1_MZ311863_CDS_0146	84265	84402	138	-	hypothetical protein	self_TKOZ146	protein
vB_EcoM_LNA1_MZ311863_CDS_0147	84570	84974	405	+	HNH endonuclease	self_TKOZ147	protein
vB_EcoM_LNA1_MZ311863_CDS_0148	85163	85309	147	+	hypothetical protein	self_TKOZ148	protein
vB_EcoM_LNA1_MZ311863_tRNA_0010	85332	85409	78	+	tRNA-Leu(TAG)		tRNA
vB_EcoM_LNA1_MZ311863_tRNA_0011	85417	85492	76	+	tRNA-Lys(CTT)		tRNA
vB_EcoM_LNA1_MZ311863_tRNA_0012	85499	85574	76	+	tRNA-Ala(TGC)		tRNA
vB_EcoM_LNA1_MZ311863_tRNA_0013	85581	85655	75	+	tRNA-Gly(TCC)		tRNA
vB_EcoM_LNA1_MZ311863_tRNA_0014	85662	85738	77	+	tRNA-Thr(TGT)		tRNA
vB_EcoM_LNA1_MZ311863_tRNA_0016	85910	85987	78	+	tRNA-Leu(CAA)		tRNA
vB_EcoM_LNA1_MZ311863_CDS_0149	86007	86564	558	+	hypothetical protein	self_TKOZ149	protein
vB_EcoM_LNA1_MZ311863_tRNA_0017	86574	86648	75	+	tRNA-Gln(TTG)		tRNA
vB_EcoM_LNA1_MZ311863_tRNA_0018	86651	86729	79	+	tRNA-Leu(TAA)		tRNA
vB_EcoM_LNA1_MZ311863_tRNA_0019	86735	86810	76	+	tRNA-Gln(CTG)		tRNA
vB_EcoM_LNA1_MZ311863_tRNA_0020	86842	86917	76	+	tRNA-His(GTG)		tRNA
vB_EcoM_LNA1_MZ311863_tRNA_0021	86932	86999	68	+	tRNA-Phe(GAA)		tRNA
vB_EcoM_LNA1_MZ311863_CDS_0150	87016	87750	735	+	hypothetical protein	self_TKOZ150	protein
vB_EcoM_LNA1_MZ311863_tRNA_0022	87850	87925	76	+	tRNA-Cys(GCA)		tRNA
vB_EcoM_LNA1_MZ311863_CDS_0151	87947	88123	177	-	hypothetical protein	self_TKOZ151	protein

Annotation of phage vB_EcoM_LNA2.

Locus_tag	Start	End	Length_nt	Strand	Function_by_amino_acid_sequence	Protein_ID	Category
vB_EcoM_LNA2_MZ311865_CDS_0001	2	382	381	-	unknown function	self_YRGQ1	protein
vB_EcoM_LNA2_MZ311865_CDS_0002	441	728	288	-	hypothetical protein	self_YRGQ2	protein
vB_EcoM_LNA2_MZ311865_CDS_0003	867	1847	981	-	unknown function	self_YRGQ3	protein
vB_EcoM_LNA2_MZ311865_CDS_0004	1917	2135	219	-	hypothetical protein	self_YRGQ4	protein
vB_EcoM_LNA2_MZ311865_CDS_0005	2244	2525	282	-	hypothetical protein	self_YRGQ5	protein
vB_EcoM_LNA2_MZ311865_CDS_0006	2557	3546	990	-	nucleotidyltransferase	self_YRGQ6	protein
vB_EcoM_LNA2_MZ311865_CDS_0007	3543	4016	474	-	membrane protein	self_YRGQ7	protein
vB_EcoM_LNA2_MZ311865_CDS_0008	4026	4556	531	-	hypothetical protein	self_YRGQ8	protein
vB_EcoM_LNA2_MZ311865_CDS_0009	4558	4737	180	-	helicase loader	self_YRGQ9	protein
vB_EcoM_LNA2_MZ311865_CDS_0010	4734	4844	111	-	hypothetical protein	self_YRGQ10	protein
vB_EcoM_LNA2_MZ311865_CDS_0011	4902	5111	210	-	hypothetical protein	self_YRGQ11	protein
vB_EcoM_LNA2_MZ311865_CDS_0012	5119	5298	180	-	hypothetical protein	self_YRGQ12	protein
vB_EcoM_LNA2_MZ311865_CDS_0013	5400	5702	303	-	lysis inhibition	self_YRGQ13	protein
vB_EcoM_LNA2_MZ311865_CDS_0014	5715	5927	213	-	hypothetical protein	self_YRGQ14	protein
vB_EcoM_LNA2_MZ311865_CDS_0015	5955	6536	582	-	thymidine kinase	self_YRGQ15	protein
vB_EcoM_LNA2_MZ311865_CDS_0016	6546	6728	183	-	hypothetical protein	self_YRGQ16	protein
vB_EcoM_LNA2_MZ311865_CDS_0017	6725	6937	213	-	hypothetical protein	self_YRGQ17	protein
vB_EcoM_LNA2_MZ311865_CDS_0018	6928	7140	213	-	hypothetical protein	self_YRGQ18	protein
vB_EcoM_LNA2_MZ311865_CDS_0019	7137	7517	381	-	phosphatase	self_YRGQ19	protein
vB_EcoM_LNA2_MZ311865_CDS_0020	7597	7935	339	-	valyl tRNA synthetase modifier	self_YRGQ20	protein
vB_EcoM_LNA2_MZ311865_CDS_0021	7941	8483	543	-	unknown function	self_YRGQ21	protein
vB_EcoM_LNA2_MZ311865_CDS_0022	8492	8950	459	-	endoribonuclease	self_YRGQ22	protein
vB_EcoM_LNA2_MZ311865_CDS_0023	9007	9306	300	-	hypothetical protein	self_YRGQ23	protein
vB_EcoM_LNA2_MZ311865_CDS_0024	9306	9572	267	-	hypothetical protein	self_YRGQ24	protein
vB_EcoM_LNA2_MZ311865_CDS_0025	9562	9786	225	-	hypothetical protein	self_YRGQ25	protein
vB_EcoM_LNA2_MZ311865_CDS_0026	9786	10148	363	-	hypothetical protein	self_YRGQ26	protein
vB_EcoM_LNA2_MZ311865_CDS_0027	10155	10463	309	-	unknown function	self_YRGQ27	protein
vB_EcoM_LNA2_MZ311865_CDS_0028	10460	10999	540	-	unknown function	self_YRGQ28	protein
vB_EcoM_LNA2_MZ311865_CDS_0029	11010	11114	105	-	hypothetical protein	self_YRGQ29	protein
vB_EcoM_LNA2_MZ311865_CDS_0030	11173	11502	330	-	hypothetical protein	self_YRGQ30	protein
vB_EcoM_LNA2_MZ311865_CDS_0031	11528	11842	315	-	hypothetical protein	self_YRGQ31	protein
vB_EcoM_LNA2_MZ311865_CDS_0032	11907	12161	255	-	internal head protein	self_YRGQ32	protein
vB_EcoM_LNA2_MZ311865_CDS_0033	12174	12587	414	-	endonuclease V N-glycosylase UV repair enzyme	self_YRGQ33	protein
vB_EcoM_LNA2_MZ311865_CDS_0034	12646	12927	282	-	internal head protein	self_YRGQ34	protein
vB_EcoM_LNA2_MZ311865_CDS_0035	12924	13412	489	-	hypothetical protein	self_YRGQ35	protein
vB_EcoM_LNA2_MZ311865_CDS_0036	13447	13905	459	-	nudix hydrolase	self_YRGQ36	protein
vB_EcoM_LNA2_MZ311865_CDS_0037	13895	14131	237	-	hypothetical protein	self_YRGQ37	protein
vB_EcoM_LNA2_MZ311865_CDS_0038	14128	14619	492	-	membrane protein	self_YRGQ38	protein
vB_EcoM_LNA2_MZ311865_CDS_0039	14619	15305	687	-	hypothetical protein	self_YRGQ39	protein
vB_EcoM_LNA2_MZ311865_CDS_0040	15307	15612	306	-	hypothetical protein	self_YRGQ40	protein
vB_EcoM_LNA2_MZ311865_CDS_0041	15681	16637	957	-	hypothetical protein	self_YRGQ41	protein
vB_EcoM_LNA2_MZ311865_CDS_0042	16666	16905	240	-	hypothetical protein	self_YRGQ42	protein
vB_EcoM_LNA2_MZ311865_CDS_0043	16968	17165	198	-	hypothetical protein	self_YRGQ43	protein
vB_EcoM_LNA2_MZ311865_CDS_0044	17175	17447	273	-	hypothetical protein	self_YRGQ44	protein
vB_EcoM_LNA2_MZ311865_CDS_0045	17457	17762	306	-	hypothetical protein	self_YRGQ45	protein
vB_EcoM_LNA2_MZ311865_CDS_0046	17759	18118	360	-	unknown function	self_YRGQ46	protein
vB_EcoM_LNA2_MZ311865_CDS_0047	18097	18492	396	-	unknown function	self_YRGQ47	protein
vB_EcoM_LNA2_MZ311865_CDS_0048	18461	19057	597	-	hypothetical protein	self_YRGQ48	protein
vB_EcoM_LNA2_MZ311865_CDS_0049	19107	19700	594	-	hypothetical protein	self_YRGQ49	protein
vB_EcoM_LNA2_MZ311865_CDS_0050	19771	19998	228	-	unknown function	self_YRGQ50	protein
vB_EcoM_LNA2_MZ311865_CDS_0051	20045	20209	165	-	hypothetical protein	self_YRGQ51	protein
vB_EcoM_LNA2_MZ311865_CDS_0052	20277	20582	306	-	hypothetical protein	self_YRGQ52	protein
vB_EcoM_LNA2_MZ311865_CDS_0053	20692	21318	627	-	hypothetical protein	self_YRGQ53	protein
vB_EcoM_LNA2_MZ311865_CDS_0054	21344	21433	90	+	hypothetical protein	self_YRGQ54	protein
vB_EcoM_LNA2_MZ311865_CDS_0055	21497	21625	129	-	hypothetical protein	self_YRGQ55	protein
vB_EcoM_LNA2_MZ311865_CDS_0056	21702	22046	345	-	hypothetical protein	self_YRGQ56	protein
vB_EcoM_LNA2_MZ311865_tRNA_0001	22068	22141	74	-	tRNA-Arg(TCT)		tRNA
vB_EcoM_LNA2_MZ311865_tRNA_0002	22148	22219	72	-	tRNA-Met(CAT)		tRNA
vB_EcoM_LNA2_MZ311865_CDS_0057	22220	22312	93	-	hypothetical protein	self_YRGQ57	protein
vB_EcoM_LNA2_MZ311865_CDS_0058	22299	22814	516	-	hypothetical protein	self_YRGQ58	protein
vB_EcoM_LNA2_MZ311865_CDS_0059	22819	23109	291	-	hypothetical protein	self_YRGQ59	protein
vB_EcoM_LNA2_MZ311865_CDS_0060	23106	23468	363	-	hypothetical protein	self_YRGQ60	protein
vB_EcoM_LNA2_MZ311865_CDS_0061	23470	23676	207	-	unknown function	self_YRGQ61	protein
vB_EcoM_LNA2_MZ311865_CDS_0062	23726	24013	288	-	internal virion protein	self_YRGQ62	protein
vB_EcoM_LNA2_MZ311865_CDS_0063	24091	24546	456	-	RNA ligase	self_YRGQ63	protein
vB_EcoM_LNA2_MZ311865_CDS_0064	24546	24776	231	-	tail fiber chaperone	self_YRGQ64	protein
vB_EcoM_LNA2_MZ311865_CDS_0065	24781	25515	735	-	deoxynucleoside monophosphate kinase	self_YRGQ65	protein
vB_EcoM_LNA2_MZ311865_CDS_0066	25569	26153	585	-	tail tube	self_YRGQ66	protein
vB_EcoM_LNA2_MZ311865_CDS_0067	26256	27077	822	-	DNA end protector	self_YRGQ67	protein

vB_EcoM_LNA2_MZ311865_CDS_0068	27080	27529	450	-	head closure	self_YRGQ68	protein
vB_EcoM_LNA2_MZ311865_CDS_0069	27663	28166	504	+	baseplate wedge subunit	self_YRGQ69	protein
vB_EcoM_LNA2_MZ311865_CDS_0070	28166	29899	1734	+	baseplate hub subunit and tail lysozyme	self_YRGQ70	protein
vB_EcoM_LNA2_MZ311865_CDS_0071	29903	30418	516	+	hypothetical protein	self_YRGQ71	protein
vB_EcoM_LNA2_MZ311865_CDS_0072	30421	30714	294	+	PAAR motif of membran proteins	self_YRGQ72	protein
vB_EcoM_LNA2_MZ311865_CDS_0073	30723	32696	1974	+	baseplate wedge subunit	self_YRGQ73	protein
vB_EcoM_LNA2_MZ311865_CDS_0074	32693	35791	3099	+	baseplate wedge subunit	self_YRGQ74	protein
vB_EcoM_LNA2_MZ311865_CDS_0075	35784	36788	1005	+	baseplate wedge subunit	self_YRGQ75	protein
vB_EcoM_LNA2_MZ311865_CDS_0076	36861	37733	873	+	baseplate wedge tail fiber protein connector	self_YRGQ76	protein
vB_EcoM_LNA2_MZ311865_CDS_0077	37733	39538	1806	+	baseplate wedge subunit	self_YRGQ77	protein
vB_EcoM_LNA2_MZ311865_CDS_0078	39538	40197	660	+	baseplate wedge subunit	self_YRGQ78	protein
vB_EcoM_LNA2_MZ311865_CDS_0079	40194	41777	1584	+	tail collar fiber protein	self_YRGQ79	protein
vB_EcoM_LNA2_MZ311865_CDS_0080	41777	43225	1449	+	fibritin neck whisker	self_YRGQ80	protein
vB_EcoM_LNA2_MZ311865_CDS_0081	43258	44184	927	+	head-tail adaptor Ad2	self_YRGQ81	protein
vB_EcoM_LNA2_MZ311865_CDS_0082	44186	44950	765	+	head closure Hc2	self_YRGQ82	protein
vB_EcoM_LNA2_MZ311865_CDS_0083	45003	45824	822	+	tail sheath stabilizer	self_YRGQ83	protein
vB_EcoM_LNA2_MZ311865_CDS_0084	45834	46328	495	+	terminase small subunit	self_YRGQ84	protein
vB_EcoM_LNA2_MZ311865_CDS_0085	46312	48147	1836	+	terminase large subunit	self_YRGQ85	protein
vB_EcoM_LNA2_MZ311865_CDS_0086	48178	50160	1983	+	tail sheath	self_YRGQ86	protein
vB_EcoM_LNA2_MZ311865_CDS_0087	50273	50764	492	+	tail tube	self_YRGQ87	protein
vB_EcoM_LNA2_MZ311865_CDS_0088	50748	50852	105	+	hypothetical protein	self_YRGQ88	protein
vB_EcoM_LNA2_MZ311865_CDS_0089	50849	52420	1572	+	portal protein	self_YRGQ89	protein
vB_EcoM_LNA2_MZ311865_CDS_0090	52420	52659	240	+	prohead	self_YRGQ90	protein
vB_EcoM_LNA2_MZ311865_CDS_0091	52659	53084	426	+	head scaffolding protein	self_YRGQ91	protein
vB_EcoM_LNA2_MZ311865_CDS_0092	53084	53725	642	+	head maturation protease	self_YRGQ92	protein
vB_EcoM_LNA2_MZ311865_CDS_0093	53759	54571	813	+	head scaffolding protein	self_YRGQ93	protein
vB_EcoM_LNA2_MZ311865_CDS_0094	54589	56157	1569	+	major head protein	self_YRGQ94	protein
vB_EcoM_LNA2_MZ311865_CDS_0095	56210	56479	270	-	hypothetical protein	self_YRGQ95	protein
vB_EcoM_LNA2_MZ311865_CDS_0096	56580	57863	1284	+	major head protein	self_YRGQ96	protein
vB_EcoM_LNA2_MZ311865_CDS_0097	57896	58894	999	-	RNA ligase	self_YRGQ97	protein
vB_EcoM_LNA2_MZ311865_CDS_0098	59014	59265	252	-	hypothetical protein	self_YRGQ98	protein
vB_EcoM_LNA2_MZ311865_CDS_0099	59262	59468	207	-	hypothetical protein	self_YRGQ99	protein
vB_EcoM_LNA2_MZ311865_CDS_0100	59557	60972	1416	-	Hoc-like head decoration	self_YRGQ100	protein
vB_EcoM_LNA2_MZ311865_CDS_0101	60982	61650	669	-	minor head protein inhibitor of protease	self_YRGQ101	protein
vB_EcoM_LNA2_MZ311865_CDS_0102	61701	63215	1515	+	DNA helicase	self_YRGQ102	protein
vB_EcoM_LNA2_MZ311865_CDS_0103	63241	63474	234	+	DNA helicase	self_YRGQ103	protein
vB_EcoM_LNA2_MZ311865_CDS_0104	63534	63701	168	-	unknown function	self_YRGQ104	protein
vB_EcoM_LNA2_MZ311865_CDS_0105	63734	63958	225	-	hypothetical protein	self_YRGQ105	protein
vB_EcoM_LNA2_MZ311865_CDS_0106	63958	64452	495	-	UvsY-like recombination mediator	self_YRGQ106	protein
vB_EcoM_LNA2_MZ311865_CDS_0107	64452	64850	399	-	baseplate wedge subunit	self_YRGQ107	protein
vB_EcoM_LNA2_MZ311865_CDS_0108	64850	65476	627	-	baseplate hub	self_YRGQ108	protein
vB_EcoM_LNA2_MZ311865_CDS_0109	65524	66276	753	+	baseplate hub	self_YRGQ109	protein
vB_EcoM_LNA2_MZ311865_CDS_0110	66273	67445	1173	+	baseplate hub	self_YRGQ110	protein
vB_EcoM_LNA2_MZ311865_CDS_0111	67393	67926	534	+	baseplate hub distal subunit	self_YRGQ111	protein
vB_EcoM_LNA2_MZ311865_CDS_0112	67923	69695	1773	+	baseplate hub subunit and tail length	self_YRGQ112	protein
vB_EcoM_LNA2_MZ311865_CDS_0113	69704	70813	1110	+	baseplate tail tube cap	self_YRGQ113	protein
vB_EcoM_LNA2_MZ311865_CDS_0114	70813	71775	963	+	tail tube	self_YRGQ114	protein
vB_EcoM_LNA2_MZ311865_CDS_0115	71808	72101	294	-	hypothetical protein	self_YRGQ115	protein
vB_EcoM_LNA2_MZ311865_CDS_0116	72160	74247	2088	-	Alt-like RNA polymerase ADP-ribosyltransferase	self_YRGQ116	protein
vB_EcoM_LNA2_MZ311865_CDS_0117	74303	74491	189	-	hypothetical protein	self_YRGQ117	protein
vB_EcoM_LNA2_MZ311865_CDS_0118	74491	75984	1494	-	ATP-dependent DNA ligase	self_YRGQ118	protein
vB_EcoM_LNA2_MZ311865_CDS_0119	75981	76250	270	-	hypothetical protein	self_YRGQ119	protein
vB_EcoM_LNA2_MZ311865_CDS_0120	76260	77075	816	-	unknown function	self_YRGQ120	protein
vB_EcoM_LNA2_MZ311865_CDS_0121	77072	77530	459	-	hypothetical protein	self_YRGQ121	protein
vB_EcoM_LNA2_MZ311865_CDS_0122	77523	77741	219	-	hypothetical protein	self_YRGQ122	protein
vB_EcoM_LNA2_MZ311865_CDS_0123	77731	77928	198	-	hypothetical protein	self_YRGQ123	protein
vB_EcoM_LNA2_MZ311865_CDS_0124	77983	78597	615	-	hypothetical protein	self_YRGQ124	protein
vB_EcoM_LNA2_MZ311865_CDS_0125	78633	78998	366	-	hypothetical protein	self_YRGQ125	protein
vB_EcoM_LNA2_MZ311865_CDS_0126	79073	79441	369	-	hypothetical protein	self_YRGQ126	protein
vB_EcoM_LNA2_MZ311865_CDS_0127	79574	79753	180	-	hypothetical protein	self_YRGQ127	protein
vB_EcoM_LNA2_MZ311865_CDS_0128	79810	79908	99	-	hypothetical protein	self_YRGQ128	protein
vB_EcoM_LNA2_MZ311865_CDS_0129	79889	79999	111	-	hypothetical protein	self_YRGQ129	protein
vB_EcoM_LNA2_MZ311865_CDS_0130	80022	80270	249	-	lysis inhibition	self_YRGQ130	protein
vB_EcoM_LNA2_MZ311865_CDS_0131	80395	80727	333	-	head morphogenesis	self_YRGQ131	protein
vB_EcoM_LNA2_MZ311865_CDS_0132	80785	81081	297	-	hypothetical protein	self_YRGQ132	protein
vB_EcoM_LNA2_MZ311865_CDS_0133	81083	81664	582	-	dCMP deaminase	self_YRGQ133	protein
vB_EcoM_LNA2_MZ311865_CDS_0134	81664	82656	993	-	phospho-2-dehydro-3-deoxyheptonate aldolase	self_YRGQ134	protein
vB_EcoM_LNA2_MZ311865_CDS_0135	82665	82790	126	-	hypothetical protein	self_YRGQ135	protein
vB_EcoM_LNA2_MZ311865_CDS_0136	82777	83013	237	-	hypothetical protein	self_YRGQ136	protein
vB_EcoM_LNA2_MZ311865_CDS_0137	83074	83349	276	-	hypothetical protein	self_YRGQ137	protein
vB_EcoM_LNA2_MZ311865_CDS_0138	83357	83581	225	-	hypothetical protein	self_YRGQ138	protein

vB_EcoM_LNA2_MZ311865_CDS_0139	83562	83753	192	-	hypothetical protein	self_YRGQ139	protein
vB_EcoM_LNA2_MZ311865_CDS_0140	83753	84529	777	-	polynucleotide kinase	self_YRGQ140	protein
vB_EcoM_LNA2_MZ311865_CDS_0141	84649	84840	192	-	hypothetical protein	self_YRGQ141	protein
vB_EcoM_LNA2_MZ311865_CDS_0142	84840	85139	300	-	hypothetical protein	self_YRGQ142	protein
vB_EcoM_LNA2_MZ311865_CDS_0143	85197	85415	219	-	hypothetical protein	self_YRGQ143	protein
vB_EcoM_LNA2_MZ311865_CDS_0144	85412	85702	291	-	Rz-like spanin	self_YRGQ144	protein
vB_EcoM_LNA2_MZ311865_CDS_0145	85699	86055	357	-	Rz-like spanin	self_YRGQ145	protein
vB_EcoM_LNA2_MZ311865_CDS_0146	86043	86606	564	-	inhibitor of host transcription	self_YRGQ146	protein
vB_EcoM_LNA2_MZ311865_CDS_0147	86606	87730	1125	-	RNA ligase and tail fiber protein attachment catalyst	self_YRGQ147	protein
vB_EcoM_LNA2_MZ311865_CDS_0148	87787	88197	411	-	endonuclease	self_YRGQ148	protein
vB_EcoM_LNA2_MZ311865_CDS_0149	88224	89402	1179	-	ribonucleotide reductase class Ia beta subunit	self_YRGQ149	protein
vB_EcoM_LNA2_MZ311865_CDS_0150	89456	91711	2256	-	ribonucleotide reductase large subunit	self_YRGQ150	protein
vB_EcoM_LNA2_MZ311865_CDS_0151	91799	92053	255	-	hypothetical protein	self_YRGQ151	protein
vB_EcoM_LNA2_MZ311865_CDS_0152	92055	92915	861	-	thymidylate synthase	self_YRGQ152	protein
vB_EcoM_LNA2_MZ311865_CDS_0153	92912	93499	588	-	dihydrofolate reductase	self_YRGQ153	protein
vB_EcoM_LNA2_MZ311865_CDS_0154	93492	93740	249	-	hypothetical protein	self_YRGQ154	protein
vB_EcoM_LNA2_MZ311865_CDS_0155	93743	93991	249	-	hypothetical protein	self_YRGQ155	protein
vB_EcoM_LNA2_MZ311865_CDS_0156	94054	94446	393	-	unknown function	self_YRGQ156	protein
vB_EcoM_LNA2_MZ311865_CDS_0157	94560	95423	864	-	single strand DNA binding protein	self_YRGQ157	protein
vB_EcoM_LNA2_MZ311865_CDS_0158	95461	95565	105	+	hypothetical protein	self_YRGQ158	protein
vB_EcoM_LNA2_MZ311865_CDS_0159	95589	96242	654	-	DNA helicase loader	self_YRGQ159	protein
vB_EcoM_LNA2_MZ311865_CDS_0160	96239	96577	339	-	late promoter transcriptional regulator	self_YRGQ160	protein
vB_EcoM_LNA2_MZ311865_CDS_0161	96555	96824	270	-	transcriptional regulator	self_YRGQ161	protein
vB_EcoM_LNA2_MZ311865_CDS_0162	96832	97749	918	-	exonuclease	self_YRGQ162	protein
vB_EcoM_LNA2_MZ311865_CDS_0163	97853	101731	3879	+	tail fiber protein proximal subunit	self_YRGQ163	protein
vB_EcoM_LNA2_MZ311865_CDS_0164	101740	102867	1128	+	long tail fiber protein proximal connector	self_YRGQ164	protein
vB_EcoM_LNA2_MZ311865_CDS_0165	102930	103595	666	+	hinge connector of long tail fiber protein distal connector	self_YRGQ165	protein
vB_EcoM_LNA2_MZ311865_CDS_0166	103604	106687	3084	+	long tail fiber protein distal subunit	self_YRGQ166	protein
vB_EcoM_LNA2_MZ311865_CDS_0167	106718	107269	552	+	tail fiber assembly	self_YRGQ167	protein
vB_EcoM_LNA2_MZ311865_CDS_0168	107279	107938	660	+	holin	self_YRGQ168	protein
vB_EcoM_LNA2_MZ311865_CDS_0169	107939	108211	273	-	anti-sigma factor	self_YRGQ169	protein
vB_EcoM_LNA2_MZ311865_CDS_0170	108258	108419	162	-	hypothetical protein	self_YRGQ170	protein
vB_EcoM_LNA2_MZ311865_CDS_0171	108409	108795	387	-	hypothetical protein	self_YRGQ171	protein
vB_EcoM_LNA2_MZ311865_CDS_0172	108792	109010	219	-	anti-restriction nuclease	self_YRGQ172	protein
vB_EcoM_LNA2_MZ311865_CDS_0173	108994	109275	282	-	unknown function	self_YRGQ173	protein
vB_EcoM_LNA2_MZ311865_CDS_0174	109275	109736	462	-	hypothetical protein	self_YRGQ174	protein
vB_EcoM_LNA2_MZ311865_CDS_0175	109733	110062	330	-	hypothetical protein	self_YRGQ175	protein
vB_EcoM_LNA2_MZ311865_CDS_0176	110073	110705	633	-	MotA-like activator of middle period transcription	self_YRGQ176	protein
vB_EcoM_LNA2_MZ311865_CDS_0177	110810	111025	216	-	hypothetical protein	self_YRGQ177	protein
vB_EcoM_LNA2_MZ311865_CDS_0178	111013	111141	129	-	hypothetical protein	self_YRGQ178	protein
vB_EcoM_LNA2_MZ311865_CDS_0179	111211	112536	1326	-	DNA topoisomerase II	self_YRGQ179	protein
vB_EcoM_LNA2_MZ311865_CDS_0180	112541	112681	141	-	hypothetical protein	self_YRGQ180	protein
vB_EcoM_LNA2_MZ311865_CDS_0181	112681	112827	147	-	hypothetical protein	self_YRGQ181	protein
vB_EcoM_LNA2_MZ311865_CDS_0182	112880	113326	447	-	Ndd-like nucleoid disruption protein	self_YRGQ182	protein
vB_EcoM_LNA2_MZ311865_CDS_0183	113311	113400	90	-	hypothetical protein	self_YRGQ183	protein
vB_EcoM_LNA2_MZ311865_CDS_0184	113403	113603	201	-	unknown function	self_YRGQ184	protein
vB_EcoM_LNA2_MZ311865_CDS_0185	113664	113774	111	-	hypothetical protein	self_YRGQ185	protein
vB_EcoM_LNA2_MZ311865_CDS_0186	113854	114117	264	-	hypothetical protein	self_YRGQ186	protein
vB_EcoM_LNA2_MZ311865_CDS_0187	114186	114743	558	-	DenB-like DNA endonuclease IV	self_YRGQ187	protein
vB_EcoM_LNA2_MZ311865_CDS_0188	114679	114966	288	-	hypothetical protein	self_YRGQ188	protein
vB_EcoM_LNA2_MZ311865_CDS_0189	115008	115955	948	-	RIIB lysis inhibitor	self_YRGQ189	protein
vB_EcoM_LNA2_MZ311865_CDS_0190	115965	118178	2214	-	RIIA lysis inhibitor	self_YRGQ190	protein
vB_EcoM_LNA2_MZ311865_CDS_0191	118191	118391	201	-	hypothetical protein	self_YRGQ191	protein
vB_EcoM_LNA2_MZ311865_CDS_0192	118484	119584	1101	-	hypothetical protein	self_YRGQ192	protein
vB_EcoM_LNA2_MZ311865_CDS_0193	119627	121444	1818	-	DNA topoisomerase II large subunit	self_YRGQ193	protein
vB_EcoM_LNA2_MZ311865_CDS_0194	121508	121951	444	-	unknown function	self_YRGQ194	protein
vB_EcoM_LNA2_MZ311865_CDS_0195	121954	122133	180	-	FmdB-like transcriptional regulator	self_YRGQ195	protein
vB_EcoM_LNA2_MZ311865_CDS_0196	122133	122546	414	-	unknown function	self_YRGQ196	protein
vB_EcoM_LNA2_MZ311865_CDS_0197	122546	122818	273	-	cef modifier of supressor tRNAs	self_YRGQ197	protein
vB_EcoM_LNA2_MZ311865_CDS_0198	122831	123355	525	-	unknown function	self_YRGQ198	protein
vB_EcoM_LNA2_MZ311865_CDS_0199	123415	123834	420	-	MotB-like transcriptional regulator	self_YRGQ199	protein
vB_EcoM_LNA2_MZ311865_CDS_0200	123844	124404	561	-	MotB-like transcriptional regulator	self_YRGQ200	protein
vB_EcoM_LNA2_MZ311865_CDS_0201	124404	124898	495	-	unknown function	self_YRGQ201	protein
vB_EcoM_LNA2_MZ311865_CDS_0202	124965	125642	678	-	exonuclease	self_YRGQ202	protein
vB_EcoM_LNA2_MZ311865_CDS_0203	125652	126965	1314	-	Dda-like helicase	self_YRGQ203	protein
vB_EcoM_LNA2_MZ311865_CDS_0204	126962	127273	312	-	hypothetical protein	self_YRGQ204	protein
vB_EcoM_LNA2_MZ311865_CDS_0205	127276	128022	747	-	Srd anti-sigma factor	self_YRGQ205	protein
vB_EcoM_LNA2_MZ311865_CDS_0206	128175	128783	609	-	RNA polymerase ADP-ribosylase	self_YRGQ206	protein
vB_EcoM_LNA2_MZ311865_CDS_0207	128841	129422	582	-	RNA polymerase ADP-ribosylase	self_YRGQ207	protein
vB_EcoM_LNA2_MZ311865_CDS_0208	129478	129642	165	-	hypothetical protein	self_YRGQ208	protein
vB_EcoM_LNA2_MZ311865_CDS_0209	129639	129818	180	-	hypothetical protein	self_YRGQ209	protein

vB_EcoM_LNA2_MZ311865_CDS_0210	129820	130284	465	-	Mrh transcription modulator under heat shock	self_YRGQ210	protein
vB_EcoM_LNA2_MZ311865_CDS_0211	130284	130463	180	-	hypothetical protein	self_YRGQ211	protein
vB_EcoM_LNA2_MZ311865_CDS_0212	130460	130708	249	-	hypothetical protein	self_YRGQ212	protein
vB_EcoM_LNA2_MZ311865_CDS_0213	130674	130901	228	-	hypothetical protein	self_YRGQ213	protein
vB_EcoM_LNA2_MZ311865_CDS_0214	131006	131242	237	-	virion structural protein	self_YRGQ214	protein
vB_EcoM_LNA2_MZ311865_CDS_0215	131289	131810	522	-	nucleoside triphosphate pyrophosphohydrolase	self_YRGQ215	protein
vB_EcoM_LNA2_MZ311865_CDS_0216	131901	132098	198	+	hypothetical protein	self_YRGQ216	protein
vB_EcoM_LNA2_MZ311865_CDS_0217	132095	133117	1023	-	DNA primase	self_YRGQ217	protein
vB_EcoM_LNA2_MZ311865_CDS_0218	133157	133624	468	-	hypothetical protein	self_YRGQ218	protein
vB_EcoM_LNA2_MZ311865_CDS_0219	133644	133841	198	-	hypothetical protein	self_YRGQ219	protein
vB_EcoM_LNA2_MZ311865_CDS_0220	133843	134493	651	-	hypothetical protein	self_YRGQ220	protein
vB_EcoM_LNA2_MZ311865_CDS_0221	134495	134788	294	-	spackle periplasmic	self_YRGQ221	protein
vB_EcoM_LNA2_MZ311865_CDS_0222	134852	135100	249	-	immunity to superinfection	self_YRGQ222	protein
vB_EcoM_LNA2_MZ311865_CDS_0223	135162	135530	369	-	hypothetical protein	self_YRGQ223	protein
vB_EcoM_LNA2_MZ311865_CDS_0224	135602	135790	189	-	Dmd discriminator of mRNA degradation	self_YRGQ224	protein
vB_EcoM_LNA2_MZ311865_CDS_0225	135787	136104	318	-	hypothetical protein	self_YRGQ225	protein
vB_EcoM_LNA2_MZ311865_CDS_0226	136160	136534	375	-	hypothetical protein	self_YRGQ226	protein
vB_EcoM_LNA2_MZ311865_CDS_0227	136623	137999	1377	-	DnaB-like replicative helicase	self_YRGQ227	protein
vB_EcoM_LNA2_MZ311865_CDS_0228	138075	138434	360	-	head vertex assembly chaperone	self_YRGQ228	protein
vB_EcoM_LNA2_MZ311865_CDS_0229	138409	139581	1173	-	DNA repair protein	self_YRGQ229	protein
vB_EcoM_LNA2_MZ311865_CDS_0230	139681	140229	549	-	thymidylate kinase	self_YRGQ230	protein
vB_EcoM_LNA2_MZ311865_CDS_0231	140231	141100	870	-	hypothetical protein	self_YRGQ231	protein
vB_EcoM_LNA2_MZ311865_CDS_0232	141130	141846	717	-	thymidylate synthase	self_YRGQ232	protein
vB_EcoM_LNA2_MZ311865_CDS_0233	141867	141968	102	-	hypothetical protein	self_YRGQ233	protein
vB_EcoM_LNA2_MZ311865_CDS_0234	141979	142128	150	-	hypothetical protein	self_YRGQ234	protein
vB_EcoM_LNA2_MZ311865_CDS_0235	142125	143285	1161	-	peptidase	self_YRGQ235	protein
vB_EcoM_LNA2_MZ311865_CDS_0236	143344	143745	402	-	phosphoheptose isomerase	self_YRGQ236	protein
vB_EcoM_LNA2_MZ311865_CDS_0237	143730	145415	1686	-	hypothetical protein	self_YRGQ237	protein
vB_EcoM_LNA2_MZ311865_CDS_0238	145457	145600	144	-	hypothetical protein	self_YRGQ238	protein
vB_EcoM_LNA2_MZ311865_CDS_0239	145597	146232	636	-	arabinose 5-phosphate isomerase	self_YRGQ239	protein
vB_EcoM_LNA2_MZ311865_CDS_0240	146273	148984	2712	-	DNA polymerase	self_YRGQ240	protein
vB_EcoM_LNA2_MZ311865_CDS_0241	148981	149076	96	-	hypothetical protein	self_YRGQ241	protein
vB_EcoM_LNA2_MZ311865_CDS_0242	149066	149443	378	-	translation repressor	self_YRGQ242	protein
vB_EcoM_LNA2_MZ311865_CDS_0243	149437	150000	564	-	clamp loader of DNA polymerase	self_YRGQ243	protein
vB_EcoM_LNA2_MZ311865_CDS_0244	150002	150964	963	-	clamp loader of DNA polymerase	self_YRGQ244	protein
vB_EcoM_LNA2_MZ311865_CDS_0245	151040	151726	687	-	DNA polymerase processivity factor	self_YRGQ245	protein
vB_EcoM_LNA2_MZ311865_CDS_0246	151769	152185	417	-	RNA polymerase binding	self_YRGQ246	protein
vB_EcoM_LNA2_MZ311865_CDS_0247	152198	152386	189	-	hypothetical protein	self_YRGQ247	protein
vB_EcoM_LNA2_MZ311865_CDS_0248	152441	154129	1689	-	SbcC-like subunit of palindrome specific endonuclease	self_YRGQ248	protein
vB_EcoM_LNA2_MZ311865_CDS_0249	154126	154365	240	-	hypothetical protein	self_YRGQ249	protein
vB_EcoM_LNA2_MZ311865_CDS_0250	154352	154609	258	-	hypothetical protein	self_YRGQ250	protein
vB_EcoM_LNA2_MZ311865_CDS_0251	154606	155625	1020	-	SbcD-like subunit of palindrome specific endonuclease	self_YRGQ251	protein
vB_EcoM_LNA2_MZ311865_CDS_0252	155696	155869	174	-	hypothetical protein	self_YRGQ252	protein
vB_EcoM_LNA2_MZ311865_CDS_0253	155909	156115	207	-	hypothetical protein	self_YRGQ253	protein
vB_EcoM_LNA2_MZ311865_CDS_0254	156090	156413	324	-	hypothetical protein	self_YRGQ254	protein
vB_EcoM_LNA2_MZ311865_CDS_0255	156415	156624	210	-	hypothetical protein	self_YRGQ255	protein
vB_EcoM_LNA2_MZ311865_CDS_0256	156614	157171	558	-	RNA polymerase sigma factor	self_YRGQ256	protein
vB_EcoM_LNA2_MZ311865_CDS_0257	157244	157513	270	-	hypothetical protein	self_YRGQ257	protein
vB_EcoM_LNA2_MZ311865_CDS_0258	157510	157737	228	-	hypothetical protein	self_YRGQ258	protein
vB_EcoM_LNA2_MZ311865_CDS_0259	157734	158066	333	-	hypothetical protein	self_YRGQ259	protein
vB_EcoM_LNA2_MZ311865_CDS_0260	158133	158372	240	-	hypothetical protein	self_YRGQ260	protein
vB_EcoM_LNA2_MZ311865_CDS_0261	158436	158636	201	-	hypothetical protein	self_YRGQ261	protein
vB_EcoM_LNA2_MZ311865_CDS_0262	158633	158767	135	-	hypothetical protein	self_YRGQ262	protein
vB_EcoM_LNA2_MZ311865_CDS_0263	158775	159068	294	-	hypothetical protein	self_YRGQ263	protein
vB_EcoM_LNA2_MZ311865_CDS_0264	159076	159324	249	-	hypothetical protein	self_YRGQ264	protein
vB_EcoM_LNA2_MZ311865_CDS_0265	159321	159665	345	-	hypothetical protein	self_YRGQ265	protein
vB_EcoM_LNA2_MZ311865_CDS_0266	159670	159978	309	-	thioredoxin domain	self_YRGQ266	protein
vB_EcoM_LNA2_MZ311865_CDS_0267	159981	160193	213	-	hypothetical protein	self_YRGQ267	protein
vB_EcoM_LNA2_MZ311865_CDS_0268	160210	160311	102	-	hypothetical protein	self_YRGQ268	protein
vB_EcoM_LNA2_MZ311865_CDS_0269	160304	160774	471	-	anaerobic ribonucleotide reductase small subunit	self_YRGQ269	protein
vB_EcoM_LNA2_MZ311865_CDS_0270	160771	162588	1818	-	anaerobic ribonucleoside reductase large subunit	self_YRGQ270	protein
vB_EcoM_LNA2_MZ311865_CDS_0271	162585	163058	474	-	endonuclease VII	self_YRGQ271	protein
vB_EcoM_LNA2_MZ311865_CDS_0272	163098	163277	180	-	hypothetical protein	self_YRGQ272	protein
vB_EcoM_LNA2_MZ311865_CDS_0273	163287	163700	414	-	inhibitor of host Lon protease	self_YRGQ273	protein
vB_EcoM_LNA2_MZ311865_CDS_0274	163687	163848	162	-	hypothetical protein	self_YRGQ274	protein
vB_EcoM_LNA2_MZ311865_CDS_0275	163820	163990	171	-	antitoxin from a toxin-antitoxin system	self_YRGQ275	protein
vB_EcoM_LNA2_MZ311865_CDS_0276	163993	164205	213	-	hypothetical protein	self_YRGQ276	protein
vB_EcoM_LNA2_MZ311865_CDS_0277	164205	164468	264	-	thioredoxin domain	self_YRGQ277	protein
vB_EcoM_LNA2_MZ311865_CDS_0278	164461	164874	414	-	hypothetical protein	self_YRGQ278	protein
vB_EcoM_LNA2_MZ311865_CDS_0279	164874	165179	306	-	unknown function	self_YRGQ279	protein
vB_EcoM_LNA2_MZ311865_CDS_0280	165179	166129	951	-	unknown function	self_YRGQ280	protein

vB_EcoM_LNA2_MZ311865_CDS_0281	166192	167151	960	-	hypothetical protein	self_YRGQ281	protein
vB_EcoM_LNA2_MZ311865_CDS_0282	167207	168202	996	-	unknown function	self_YRGQ282	protein

Annotation of phage vB_EcoM_LNA3.

Locus_tag	Start	End	Length_nt	Strand	Function_by_amino_acid_sequence	Protein_ID	Category
vB_EcoP_LNA3_MZ311862_CDS_0001	3	167	165	+	tail protein	self_ZLQQ1	protein
vB_EcoP_LNA3_MZ311862_CDS_0002	167	2524	2358	+	tail protein	self_ZLQQ2	protein
vB_EcoP_LNA3_MZ311862_CDS_0003	2601	3071	471	+	internal virion protein	self_ZLQQ3	protein
vB_EcoP_LNA3_MZ311862_CDS_0004	3056	3643	588	+	internal virion protein	self_ZLQQ4	protein
vB_EcoP_LNA3_MZ311862_CDS_0005	3655	5937	2283	+	internal virion protein	self_ZLQQ5	protein
vB_EcoP_LNA3_MZ311862_CDS_0006	5943	9830	3888	+	internal virion protein with endolysin domain	self_ZLQQ6	protein
vB_EcoP_LNA3_MZ311862_CDS_0007	9896	13090	3195	+	tail protein	self_ZLQQ7	protein
vB_EcoP_LNA3_MZ311862_CDS_0008	13223	13417	195	+	holin	self_ZLQQ8	protein
vB_EcoP_LNA3_MZ311862_CDS_0009	13414	13677	264	+	terminase small subunit	self_ZLQQ9	protein
vB_EcoP_LNA3_MZ311862_CDS_0010	13779	14228	450	+	Rz-like spanin	self_ZLQQ10	protein
vB_EcoP_LNA3_MZ311862_CDS_0011	14258	15988	1731	+	terminase large subunit	self_ZLQQ11	protein
vB_EcoP_LNA3_MZ311862_CDS_0012	16280	16438	159	+	hypothetical protein	self_ZLQQ12	protein
vB_EcoP_LNA3_MZ311862_CDS_0013	16652	16798	147	+	hypothetical protein	self_ZLQQ13	protein
vB_EcoP_LNA3_MZ311862_CDS_0014	17031	17246	216	+	hypothetical protein	self_ZLQQ14	protein
vB_EcoP_LNA3_MZ311862_CDS_0015	17457	17711	255	+	hypothetical protein	self_ZLQQ15	protein
vB_EcoP_LNA3_MZ311862_CDS_0016	17795	18298	504	+	ocr-like anti-restriction	self_ZLQQ16	protein
vB_EcoP_LNA3_MZ311862_CDS_0017	18298	18447	150	+	hypothetical protein	self_ZLQQ17	protein
vB_EcoP_LNA3_MZ311862_CDS_0018	18514	18681	168	+	hypothetical protein	self_ZLQQ18	protein
vB_EcoP_LNA3_MZ311862_CDS_0019	18685	18882	198	+	hypothetical protein	self_ZLQQ19	protein
vB_EcoP_LNA3_MZ311862_CDS_0020	18882	19142	261	+	hypothetical protein	self_ZLQQ20	protein
vB_EcoP_LNA3_MZ311862_CDS_0021	19240	21921	2682	+	N4-like RNA polymerase	self_ZLQQ21	protein
vB_EcoP_LNA3_MZ311862_CDS_0022	21934	22149	216	+	hypothetical protein	self_ZLQQ22	protein
vB_EcoP_LNA3_MZ311862_CDS_0023	22317	22496	180	+	hypothetical protein	self_ZLQQ23	protein
vB_EcoP_LNA3_MZ311862_CDS_0024	22568	22747	180	+	dGTPase inhibitor	self_ZLQQ24	protein
vB_EcoP_LNA3_MZ311862_CDS_0025	22747	23772	1026	+	DNA ligase	self_ZLQQ25	protein
vB_EcoP_LNA3_MZ311862_CDS_0026	23903	24163	261	+	hypothetical protein	self_ZLQQ26	protein
vB_EcoP_LNA3_MZ311862_CDS_0027	24156	24596	441	+	hypothetical protein	self_ZLQQ27	protein
vB_EcoP_LNA3_MZ311862_CDS_0028	24586	24813	228	+	hypothetical protein	self_ZLQQ28	protein
vB_EcoP_LNA3_MZ311862_CDS_0029	24806	24904	99	+	hypothetical protein	self_ZLQQ29	protein
vB_EcoP_LNA3_MZ311862_CDS_0030	24901	25059	159	+	RNA polymerase inhibitor	self_ZLQQ30	protein
vB_EcoP_LNA3_MZ311862_CDS_0031	25116	25829	714	+	Gp2.5-like ssDNA binding protein and ssDNA annealing protein	self_ZLQQ31	protein
vB_EcoP_LNA3_MZ311862_CDS_0032	25829	26281	453	+	endonuclease	self_ZLQQ32	protein
vB_EcoP_LNA3_MZ311862_CDS_0033	26281	26736	456	+	hypothetical protein	self_ZLQQ33	protein
vB_EcoP_LNA3_MZ311862_CDS_0034	26726	27184	459	+	amidase	self_ZLQQ34	protein
vB_EcoP_LNA3_MZ311862_CDS_0035	27209	27424	216	+	hypothetical protein	self_ZLQQ35	protein
vB_EcoP_LNA3_MZ311862_CDS_0036	27493	27672	180	+	hypothetical protein	self_ZLQQ36	protein
vB_EcoP_LNA3_MZ311862_CDS_0037	27683	29383	1701	+	DNA primase/helicase	self_ZLQQ37	protein
vB_EcoP_LNA3_MZ311862_CDS_0038	29435	29599	165	+	hypothetical protein	self_ZLQQ38	protein
vB_EcoP_LNA3_MZ311862_CDS_0039	29664	31835	2172	+	DNA polymerase I	self_ZLQQ39	protein
vB_EcoP_LNA3_MZ311862_CDS_0040	31835	31993	159	+	hypothetical protein	self_ZLQQ40	protein
vB_EcoP_LNA3_MZ311862_CDS_0041	31993	32277	285	+	hypothetical protein	self_ZLQQ41	protein
vB_EcoP_LNA3_MZ311862_CDS_0042	32274	32483	210	+	hypothetical protein	self_ZLQQ42	protein
vB_EcoP_LNA3_MZ311862_CDS_0043	32480	32935	456	+	hypothetical protein	self_ZLQQ43	protein
vB_EcoP_LNA3_MZ311862_CDS_0044	32925	33794	870	+	exonuclease	self_ZLQQ44	protein
vB_EcoP_LNA3_MZ311862_CDS_0045	33776	33868	93	+	hypothetical protein	self_ZLQQ45	protein
vB_EcoP_LNA3_MZ311862_CDS_0046	34005	34277	273	+	hypothetical protein	self_ZLQQ46	protein
vB_EcoP_LNA3_MZ311862_CDS_0047	34292	34516	225	+	hypothetical protein	self_ZLQQ47	protein
vB_EcoP_LNA3_MZ311862_CDS_0048	34521	34925	405	+	host range and adsorption protein	self_ZLQQ48	protein
vB_EcoP_LNA3_MZ311862_CDS_0049	34918	35166	249	+	host range and adsorption protein	self_ZLQQ49	protein
vB_EcoP_LNA3_MZ311862_CDS_0050	35178	36746	1569	+	head-tail adaptor	self_ZLQQ50	protein
vB_EcoP_LNA3_MZ311862_CDS_0051	36841	37725	885	+	head assembly	self_ZLQQ51	protein
vB_EcoP_LNA3_MZ311862_CDS_0052	37852	38895	1044	+	major head protein	self_ZLQQ52	protein
vB_EcoP_LNA3_MZ311862_CDS_0053	39068	39178	111	+	hypothetical protein	self_ZLQQ53	protein

Annotation of phage vB_EcoM_LNA4.

Locus_tag	Start	End	Length_nt	Strand	Function_by_amino_acid_sequence	Protein_ID	Category
vB_EcoM_LNA4_MZ311864_CDS_0001	1	123	123	+	hypothetical protein	self_AGKR1	protein
vB_EcoM_LNA4_MZ311864_CDS_0002	255	437	183	-	hypothetical protein	self_AGKR2	protein
vB_EcoM_LNA4_MZ311864_CDS_0003	434	550	117	-	hypothetical protein	self_AGKR3	protein
vB_EcoM_LNA4_MZ311864_CDS_0004	671	1099	429	-	hypothetical protein	self_AGKR4	protein
vB_EcoM_LNA4_MZ311864_CDS_0005	1112	1297	186	-	hypothetical protein	self_AGKR5	protein
vB_EcoM_LNA4_MZ311864_CDS_0006	1294	1524	231	-	hypothetical protein	self_AGKR6	protein
vB_EcoM_LNA4_MZ311864_CDS_0007	1521	1715	195	-	hypothetical protein	self_AGKR7	protein
vB_EcoM_LNA4_MZ311864_CDS_0008	1720	2040	321	-	hypothetical protein	self_AGKR8	protein
vB_EcoM_LNA4_MZ311864_CDS_0009	2064	2333	270	+	hypothetical protein	self_AGKR9	protein
vB_EcoM_LNA4_MZ311864_CDS_0010	2323	2649	327	+	hypothetical protein	self_AGKR10	protein
vB_EcoM_LNA4_MZ311864_CDS_0011	2657	2842	186	-	hypothetical protein	self_AGKR11	protein
vB_EcoM_LNA4_MZ311864_CDS_0012	2913	3158	246	-	hypothetical protein	self_AGKR12	protein
vB_EcoM_LNA4_MZ311864_CDS_0013	3255	3449	195	-	hypothetical protein	self_AGKR13	protein
vB_EcoM_LNA4_MZ311864_CDS_0014	3478	3615	138	-	hypothetical protein	self_AGKR14	protein
vB_EcoM_LNA4_MZ311864_CDS_0015	3637	3750	114	-	hypothetical protein	self_AGKR15	protein
vB_EcoM_LNA4_MZ311864_CDS_0016	3834	4985	1152	-	regulator of chromosome condensation	self_AGKR16	protein
vB_EcoM_LNA4_MZ311864_CDS_0017	5026	5274	249	-	hypothetical protein	self_AGKR17	protein
vB_EcoM_LNA4_MZ311864_CDS_0018	5261	5437	177	-	hypothetical protein	self_AGKR18	protein
vB_EcoM_LNA4_MZ311864_CDS_0019	5434	5628	195	-	hypothetical protein	self_AGKR19	protein
vB_EcoM_LNA4_MZ311864_CDS_0020	5625	5873	249	-	hypothetical protein	self_AGKR20	protein
vB_EcoM_LNA4_MZ311864_CDS_0021	5870	6127	258	-	hypothetical protein	self_AGKR21	protein
vB_EcoM_LNA4_MZ311864_CDS_0022	6121	6279	159	-	hypothetical protein	self_AGKR22	protein
vB_EcoM_LNA4_MZ311864_CDS_0023	6276	6485	210	-	hypothetical protein	self_AGKR23	protein
vB_EcoM_LNA4_MZ311864_CDS_0024	6475	7623	1149	-	hypothetical protein	self_AGKR24	protein
vB_EcoM_LNA4_MZ311864_CDS_0025	7659	7754	96	-	hypothetical protein	self_AGKR25	protein
vB_EcoM_LNA4_MZ311864_CDS_0026	7743	7991	249	+	hypothetical protein	self_AGKR26	protein
vB_EcoM_LNA4_MZ311864_CDS_0027	7991	8302	312	+	hypothetical protein	self_AGKR27	protein
vB_EcoM_LNA4_MZ311864_CDS_0028	8326	8829	504	-	hypothetical protein	self_AGKR28	protein
vB_EcoM_LNA4_MZ311864_CDS_0029	8814	9209	396	-	endonuclease VII	self_AGKR29	protein
vB_EcoM_LNA4_MZ311864_CDS_0030	9221	9709	489	-	virion structural protein	self_AGKR30	protein
vB_EcoM_LNA4_MZ311864_CDS_0031	9789	10019	231	-	hypothetical protein	self_AGKR31	protein
vB_EcoM_LNA4_MZ311864_CDS_0032	10016	10174	159	-	hypothetical protein	self_AGKR32	protein
vB_EcoM_LNA4_MZ311864_CDS_0033	10171	10347	177	-	hypothetical protein	self_AGKR33	protein
vB_EcoM_LNA4_MZ311864_CDS_0034	10334	10546	213	-	hypothetical protein	self_AGKR34	protein
vB_EcoM_LNA4_MZ311864_CDS_0035	10539	11612	1074	-	nucleotidyltransferase	self_AGKR35	protein
vB_EcoM_LNA4_MZ311864_CDS_0036	11621	11995	375	-	hypothetical protein	self_AGKR36	protein
vB_EcoM_LNA4_MZ311864_CDS_0037	11997	12332	336	-	membrane protein	self_AGKR37	protein
vB_EcoM_LNA4_MZ311864_CDS_0038	12329	12514	186	-	hypothetical protein	self_AGKR38	protein
vB_EcoM_LNA4_MZ311864_CDS_0039	12516	13004	489	-	hydrolase	self_AGKR39	protein
vB_EcoM_LNA4_MZ311864_CDS_0040	13001	13219	219	-	membrane protein	self_AGKR40	protein
vB_EcoM_LNA4_MZ311864_CDS_0041	13219	13536	318	-	hypothetical protein	self_AGKR41	protein
vB_EcoM_LNA4_MZ311864_CDS_0042	13546	13752	207	-	hypothetical protein	self_AGKR42	protein
vB_EcoM_LNA4_MZ311864_CDS_0043	13739	14194	456	-	hypothetical protein	self_AGKR43	protein
vB_EcoM_LNA4_MZ311864_CDS_0044	14203	15456	1254	-	tRNA nucleotidyltransferase	self_AGKR44	protein
vB_EcoM_LNA4_MZ311864_CDS_0045	15466	15732	267	-	hypothetical protein	self_AGKR45	protein
vB_EcoM_LNA4_MZ311864_CDS_0046	15729	16040	312	-	hypothetical protein	self_AGKR46	protein
vB_EcoM_LNA4_MZ311864_CDS_0047	16030	16281	252	-	hypothetical protein	self_AGKR47	protein
vB_EcoM_LNA4_MZ311864_CDS_0048	16278	16436	159	-	hypothetical protein	self_AGKR48	protein
vB_EcoM_LNA4_MZ311864_CDS_0049	16420	16824	405	-	hypothetical protein	self_AGKR49	protein
vB_EcoM_LNA4_MZ311864_CDS_0050	16809	17081	273	-	hypothetical protein	self_AGKR50	protein
vB_EcoM_LNA4_MZ311864_CDS_0051	17169	17288	120	-	hypothetical protein	self_AGKR51	protein
vB_EcoM_LNA4_MZ311864_CDS_0052	17281	17961	681	-	DNA methyltransferase	self_AGKR52	protein
vB_EcoM_LNA4_MZ311864_CDS_0053	17961	18893	933	-	ATP-dependent protease	self_AGKR53	protein
vB_EcoM_LNA4_MZ311864_CDS_0054	18895	19377	483	-	hypothetical protein	self_AGKR54	protein
vB_EcoM_LNA4_MZ311864_CDS_0055	19385	19702	318	-	hypothetical protein	self_AGKR55	protein
vB_EcoM_LNA4_MZ311864_CDS_0056	19669	20190	522	-	HNH endonuclease	self_AGKR56	protein
vB_EcoM_LNA4_MZ311864_CDS_0057	20165	20794	630	-	DprA-like DNA recombination-mediator protein	self_AGKR57	protein
vB_EcoM_LNA4_MZ311864_CDS_0058	20805	21605	801	-	PhoH-like phosphate starvation-inducible	self_AGKR58	protein
vB_EcoM_LNA4_MZ311864_CDS_0059	21617	22171	555	-	endolysin	self_AGKR59	protein
vB_EcoM_LNA4_MZ311864_CDS_0060	22184	22615	432	-	hypothetical protein	self_AGKR60	protein
vB_EcoM_LNA4_MZ311864_CDS_0061	22624	22764	141	-	hypothetical protein	self_AGKR61	protein
vB_EcoM_LNA4_MZ311864_CDS_0062	22842	23108	267	-	thioredoxin domain	self_AGKR62	protein
vB_EcoM_LNA4_MZ311864_CDS_0063	23120	24208	1089	-	ribonucleoside diphosphate reductase small subunit	self_AGKR63	protein
vB_EcoM_LNA4_MZ311864_CDS_0064	24247	26562	2316	-	ribonucleotide reductase large subunit	self_AGKR64	protein
vB_EcoM_LNA4_MZ311864_CDS_0065	26572	26802	231	-	hypothetical protein	self_AGKR65	protein
vB_EcoM_LNA4_MZ311864_CDS_0066	26802	27278	477	-	hypothetical protein	self_AGKR66	protein
vB_EcoM_LNA4_MZ311864_CDS_0067	27360	28358	999	-	thymidylate synthase	self_AGKR67	protein
vB_EcoM_LNA4_MZ311864_CDS_0068	28358	28681	324	-	hypothetical protein	self_AGKR68	protein

vB_EcoM_LNA4_MZ311864_CDS_0069	28678	28872	195	-	hypothetical protein	self_AGKR69	protein
vB_EcoM_LNA4_MZ311864_CDS_0070	28872	29474	603	-	unknown function	self_AGKR70	protein
vB_EcoM_LNA4_MZ311864_CDS_0071	29467	30072	606	-	deoxynucleoside monophosphate kinase	self_AGKR71	protein
vB_EcoM_LNA4_MZ311864_CDS_0072	30072	30257	186	-	hypothetical protein	self_AGKR72	protein
vB_EcoM_LNA4_MZ311864_CDS_0073	30288	31262	975	-	DNA polymerase exonuclease subunit	self_AGKR73	protein
vB_EcoM_LNA4_MZ311864_CDS_0074	31262	31387	126	-	hypothetical protein	self_AGKR74	protein
vB_EcoM_LNA4_MZ311864_CDS_0075	31464	31949	486	-	endonuclease VII	self_AGKR75	protein
vB_EcoM_LNA4_MZ311864_CDS_0076	31994	32647	654	-	RuvC-like Holliday junction resolvase	self_AGKR76	protein
vB_EcoM_LNA4_MZ311864_CDS_0077	32622	32927	306	-	hypothetical protein	self_AGKR77	protein
vB_EcoM_LNA4_MZ311864_CDS_0078	32924	34045	1122	-	exonuclease	self_AGKR78	protein
vB_EcoM_LNA4_MZ311864_CDS_0079	34054	34521	468	-	HNH endonuclease	self_AGKR79	protein
vB_EcoM_LNA4_MZ311864_CDS_0080	34493	34999	507	-	hypothetical protein	self_AGKR80	protein
vB_EcoM_LNA4_MZ311864_CDS_0081	35011	35205	195	-	hypothetical protein	self_AGKR81	protein
vB_EcoM_LNA4_MZ311864_CDS_0082	35269	36159	891	-	lipoprotein	self_AGKR82	protein
vB_EcoM_LNA4_MZ311864_CDS_0083	36152	36316	165	-	hypothetical protein	self_AGKR83	protein
vB_EcoM_LNA4_MZ311864_CDS_0084	36300	36503	204	-	membrane protein	self_AGKR84	protein
vB_EcoM_LNA4_MZ311864_CDS_0085	36503	36838	336	-	transcriptional regulator	self_AGKR85	protein
vB_EcoM_LNA4_MZ311864_CDS_0086	36907	37152	246	-	hypothetical protein	self_AGKR86	protein
vB_EcoM_LNA4_MZ311864_CDS_0087	37207	37896	690	-	nucleotide pyrophosphohydrolase	self_AGKR87	protein
vB_EcoM_LNA4_MZ311864_CDS_0088	37906	38277	372	-	hypothetical protein	self_AGKR88	protein
vB_EcoM_LNA4_MZ311864_CDS_0089	38277	38723	447	-	hypothetical protein	self_AGKR89	protein
vB_EcoM_LNA4_MZ311864_CDS_0090	38723	40027	1305	-	ATP-dependent DNA ligase	self_AGKR90	protein
vB_EcoM_LNA4_MZ311864_CDS_0091	40027	40803	777	-	Sir2 (NAD-dependent deacetylase)	self_AGKR91	protein
vB_EcoM_LNA4_MZ311864_CDS_0092	40800	41018	219	-	hypothetical protein	self_AGKR92	protein
vB_EcoM_LNA4_MZ311864_CDS_0093	41025	41909	885	-	ATPase	self_AGKR93	protein
vB_EcoM_LNA4_MZ311864_CDS_0094	41919	42095	177	-	hypothetical protein	self_AGKR94	protein
vB_EcoM_LNA4_MZ311864_CDS_0095	42114	42470	357	-	hypothetical protein	self_AGKR95	protein
vB_EcoM_LNA4_MZ311864_CDS_0096	42475	43071	597	-	phosphoesterase	self_AGKR96	protein
vB_EcoM_LNA4_MZ311864_CDS_0097	43068	43352	285	-	hypothetical protein	self_AGKR97	protein
vB_EcoM_LNA4_MZ311864_CDS_0098	43349	43618	270	-	hypothetical protein	self_AGKR98	protein
vB_EcoM_LNA4_MZ311864_CDS_0099	43736	44080	345	-	hypothetical protein	self_AGKR99	protein
vB_EcoM_LNA4_MZ311864_CDS_0100	44092	44214	123	-	hypothetical protein	self_AGKR100	protein
vB_EcoM_LNA4_MZ311864_CDS_0101	44241	44336	96	+	hypothetical protein	self_AGKR101	protein
vB_EcoM_LNA4_MZ311864_CDS_0102	44359	44676	318	+	replication initiation protein	self_AGKR102	protein
vB_EcoM_LNA4_MZ311864_CDS_0103	44678	45490	813	+	ribose-phosphate pyrophosphokinase	self_AGKR103	protein
vB_EcoM_LNA4_MZ311864_CDS_0104	45490	46023	534	+	hypothetical protein	self_AGKR104	protein
vB_EcoM_LNA4_MZ311864_CDS_0105	46077	47762	1686	+	nicotinamide phosphoribosyl transferase	self_AGKR105	protein
vB_EcoM_LNA4_MZ311864_CDS_0106	47824	47913	90	+	hypothetical protein	self_AGKR106	protein
vB_EcoM_LNA4_MZ311864_CDS_0107	47898	48113	216	+	hypothetical protein	self_AGKR107	protein
vB_EcoM_LNA4_MZ311864_CDS_0108	48106	48315	210	+	hypothetical protein	self_AGKR108	protein
vB_EcoM_LNA4_MZ311864_CDS_0109	48395	48640	246	+	hypothetical protein	self_AGKR109	protein
vB_EcoM_LNA4_MZ311864_CDS_0110	48704	48823	120	+	hypothetical protein	self_AGKR110	protein
vB_EcoM_LNA4_MZ311864_CDS_0111	48826	49005	180	+	hypothetical protein	self_AGKR111	protein
vB_EcoM_LNA4_MZ311864_tRNA_0001	49029	49105	77	+	tRNA-Met(CAT)		tRNA
vB_EcoM_LNA4_MZ311864_CDS_0112	49120	49287	168	+	hypothetical protein	self_AGKR112	protein
vB_EcoM_LNA4_MZ311864_CDS_0113	49357	49686	330	+	hypothetical protein	self_AGKR113	protein
vB_EcoM_LNA4_MZ311864_CDS_0114	49705	49830	126	+	hypothetical protein	self_AGKR114	protein
vB_EcoM_LNA4_MZ311864_tRNA_0002	49895	49970	76	+	tRNA-Lys(CTT)		tRNA
vB_EcoM_LNA4_MZ311864_CDS_0115	50145	50756	612	+	hypothetical protein	self_AGKR115	protein
vB_EcoM_LNA4_MZ311864_CDS_0116	50834	51481	648	+	tail fiber protein	self_AGKR116	protein
vB_EcoM_LNA4_MZ311864_CDS_0117	51629	51772	144	+	hypothetical protein	self_AGKR117	protein
vB_EcoM_LNA4_MZ311864_tRNA_0003	51869	51943	75	+	tRNA-Thr(TGT)		tRNA
vB_EcoM_LNA4_MZ311864_CDS_0118	52108	52212	105	+	hypothetical protein	self_AGKR118	protein
vB_EcoM_LNA4_MZ311864_tRNA_0005	52326	52403	78	+	tRNA-Leu(TAA)		tRNA
vB_EcoM_LNA4_MZ311864_tRNA_0006	52407	52481	75	+	tRNA-Arg(TCT)		tRNA
vB_EcoM_LNA4_MZ311864_tRNA_0007	52487	52560	74	+	tRNA-Cys(GCA)		tRNA
vB_EcoM_LNA4_MZ311864_CDS_0119	52647	53012	366	+	hypothetical protein	self_AGKR119	protein
vB_EcoM_LNA4_MZ311864_tRNA_0008	53015	53102	88	+	tRNA-Tyr(GTA)		tRNA
vB_EcoM_LNA4_MZ311864_tRNA_0009	53109	53183	75	+	tRNA-Ile(GAT)		tRNA
vB_EcoM_LNA4_MZ311864_tRNA_0011	53271	53346	76	+	tRNA-Glu(TTC)		tRNA
vB_EcoM_LNA4_MZ311864_tRNA_0012	53354	53430	77	+	tRNA-Trp(CCA)		tRNA
vB_EcoM_LNA4_MZ311864_tRNA_0013	53436	53511	76	+	tRNA-Gln(TTG)		tRNA
vB_EcoM_LNA4_MZ311864_tRNA_0014	53626	53718	93	+	tRNA-Ser(GGA)		tRNA
vB_EcoM_LNA4_MZ311864_CDS_0120	53731	53820	90	-	hypothetical protein	self_AGKR120	protein
vB_EcoM_LNA4_MZ311864_tRNA_0015	54066	54152	87	+	tRNA-Lys(TTT)		tRNA
vB_EcoM_LNA4_MZ311864_tRNA_0016	54322	54410	89	+	tRNA-Asn(GTT)		tRNA
vB_EcoM_LNA4_MZ311864_tRNA_0017	54454	54530	77	+	tRNA-Asp(GTC)		tRNA
vB_EcoM_LNA4_MZ311864_tRNA_0018	54714	54791	78	+	tRNA-Pro(TGG)		tRNA
vB_EcoM_LNA4_MZ311864_tRNA_0019	54797	54872	76	+	tRNA-Arg(ACG)		tRNA
vB_EcoM_LNA4_MZ311864_tRNA_0020	54908	54984	77	+	tRNA-Leu(TAG)		tRNA

vB_EcoM_LNA4_MZ311864_tRNA_0021	54992	55067	76	+	tRNA-Phe(GAA)		tRNA
vB_EcoM_LNA4_MZ311864_CDS_0121	55164	55628	465	+	endonuclease VII	self_AGKR121	protein
vB_EcoM_LNA4_MZ311864_CDS_0122	55675	55821	147	+	hypothetical protein	self_AGKR122	protein
vB_EcoM_LNA4_MZ311864_CDS_0123	55838	57262	1425	+	hypothetical protein	self_AGKR123	protein
vB_EcoM_LNA4_MZ311864_CDS_0124	57259	57663	405	+	Rz-like spanin	self_AGKR124	protein
vB_EcoM_LNA4_MZ311864_CDS_0125	57660	57950	291	+	Rz-like spanin	self_AGKR125	protein
vB_EcoM_LNA4_MZ311864_CDS_0126	58089	59549	1461	+	terminase large subunit	self_AGKR126	protein
vB_EcoM_LNA4_MZ311864_CDS_0127	59694	61115	1422	+	hypothetical protein	self_AGKR127	protein
vB_EcoM_LNA4_MZ311864_CDS_0128	61182	61811	630	+	hypothetical protein	self_AGKR128	protein
vB_EcoM_LNA4_MZ311864_CDS_0129	61808	62809	1002	+	hypothetical protein	self_AGKR129	protein
vB_EcoM_LNA4_MZ311864_CDS_0130	62830	63249	420	+	head decoration	self_AGKR130	protein
vB_EcoM_LNA4_MZ311864_CDS_0131	63271	64284	1014	+	major head protein	self_AGKR131	protein
vB_EcoM_LNA4_MZ311864_CDS_0132	64418	64942	525	+	hypothetical protein	self_AGKR132	protein
vB_EcoM_LNA4_MZ311864_CDS_0133	64952	67573	2622	+	tail fiber protein	self_AGKR133	protein
vB_EcoM_LNA4_MZ311864_CDS_0134	67609	68253	645	+	hypothetical protein	self_AGKR134	protein
vB_EcoM_LNA4_MZ311864_CDS_0135	68263	68814	552	+	tail needle knob	self_AGKR135	protein
vB_EcoM_LNA4_MZ311864_CDS_0136	68835	69134	300	+	head fiber protein	self_AGKR136	protein
vB_EcoM_LNA4_MZ311864_CDS_0137	69157	69561	405	-	membrane protein	self_AGKR137	protein
vB_EcoM_LNA4_MZ311864_CDS_0138	69696	70214	519	+	hypothetical protein	self_AGKR138	protein
vB_EcoM_LNA4_MZ311864_CDS_0139	70280	70756	477	+	tail completion or Neck1 protein	self_AGKR139	protein
vB_EcoM_LNA4_MZ311864_CDS_0140	70756	71190	435	+	minor head protein	self_AGKR140	protein
vB_EcoM_LNA4_MZ311864_CDS_0141	71190	71720	531	+	hypothetical protein	self_AGKR141	protein
vB_EcoM_LNA4_MZ311864_CDS_0142	71851	73263	1413	+	tail sheath	self_AGKR142	protein
vB_EcoM_LNA4_MZ311864_CDS_0143	73267	73740	474	+	virion structural protein	self_AGKR143	protein
vB_EcoM_LNA4_MZ311864_CDS_0144	73724	73819	96	+	hypothetical protein	self_AGKR144	protein
vB_EcoM_LNA4_MZ311864_CDS_0145	73816	74289	474	+	tail assembly chaperone	self_AGKR145	protein
vB_EcoM_LNA4_MZ311864_CDS_0146	74325	74567	243	+	hypothetical protein	self_AGKR146	protein
vB_EcoM_LNA4_MZ311864_CDS_0147	74621	77014	2394	+	tail length tape measure protein	self_AGKR147	protein
vB_EcoM_LNA4_MZ311864_CDS_0148	77096	77968	873	+	hypothetical protein	self_AGKR148	protein
vB_EcoM_LNA4_MZ311864_CDS_0149	77968	78318	351	+	virion structural protein	self_AGKR149	protein
vB_EcoM_LNA4_MZ311864_CDS_0150	78321	79310	990	+	baseplate hub	self_AGKR150	protein
vB_EcoM_LNA4_MZ311864_CDS_0151	79310	80011	702	+	baseplate spike	self_AGKR151	protein
vB_EcoM_LNA4_MZ311864_CDS_0152	80021	80638	618	+	hypothetical protein	self_AGKR152	protein
vB_EcoM_LNA4_MZ311864_CDS_0153	80649	81323	675	+	tail fiber protein	self_AGKR153	protein
vB_EcoM_LNA4_MZ311864_CDS_0154	81323	84112	2790	+	tail protein	self_AGKR154	protein
vB_EcoM_LNA4_MZ311864_CDS_0155	84122	86260	2139	+	colanic acid degradation	self_AGKR155	protein
vB_EcoM_LNA4_MZ311864_CDS_0156	86369	87862	1494	+	baseplate wedge subunit	self_AGKR156	protein
vB_EcoM_LNA4_MZ311864_CDS_0157	87873	88505	633	+	baseplate protein	self_AGKR157	protein
vB_EcoM_LNA4_MZ311864_CDS_0158	88517	89629	1113	+	tail fiber protein	self_AGKR158	protein
vB_EcoM_LNA4_MZ311864_CDS_0159	89641	90156	516	+	tail fiber assembly protein	self_AGKR159	protein
vB_EcoM_LNA4_MZ311864_CDS_0160	90166	90480	315	+	hypothetical protein	self_AGKR160	protein
vB_EcoM_LNA4_MZ311864_CDS_0161	90477	90950	474	+	hypothetical protein	self_AGKR161	protein
vB_EcoM_LNA4_MZ311864_CDS_0162	90966	91235	270	+	membrane protein	self_AGKR162	protein
vB_EcoM_LNA4_MZ311864_CDS_0163	91255	93054	1800	+	tail fiber protein	self_AGKR163	protein
vB_EcoM_LNA4_MZ311864_CDS_0164	93099	94802	1704	+	tail protein	self_AGKR164	protein
vB_EcoM_LNA4_MZ311864_CDS_0165	94834	95265	432	-	endonuclease	self_AGKR165	protein
vB_EcoM_LNA4_MZ311864_CDS_0166	95271	96407	1137	-	hypothetical protein	self_AGKR166	protein
vB_EcoM_LNA4_MZ311864_CDS_0167	96421	96768	348	-	hypothetical protein	self_AGKR167	protein
vB_EcoM_LNA4_MZ311864_CDS_0168	96831	97037	207	-	hypothetical protein	self_AGKR168	protein
vB_EcoM_LNA4_MZ311864_CDS_0169	97034	97219	186	-	hypothetical protein	self_AGKR169	protein
vB_EcoM_LNA4_MZ311864_CDS_0170	97280	97786	507	-	hypothetical protein	self_AGKR170	protein
vB_EcoM_LNA4_MZ311864_CDS_0171	97779	98057	279	-	hypothetical protein	self_AGKR171	protein
vB_EcoM_LNA4_MZ311864_CDS_0172	98048	98419	372	-	hypothetical protein	self_AGKR172	protein
vB_EcoM_LNA4_MZ311864_CDS_0173	98423	98992	570	-	hypothetical protein	self_AGKR173	protein
vB_EcoM_LNA4_MZ311864_CDS_0174	99007	99186	180	-	hypothetical protein	self_AGKR174	protein
vB_EcoM_LNA4_MZ311864_CDS_0175	99183	99890	708	-	PnuC-like nicotinamide mononucleotide transport	self_AGKR175	protein
vB_EcoM_LNA4_MZ311864_CDS_0176	99915	100142	228	-	hypothetical protein	self_AGKR176	protein
vB_EcoM_LNA4_MZ311864_CDS_0177	100139	100387	249	-	hypothetical protein	self_AGKR177	protein
vB_EcoM_LNA4_MZ311864_CDS_0178	100380	100718	339	-	ribonucleotide reductase NrdA-like	self_AGKR178	protein
vB_EcoM_LNA4_MZ311864_CDS_0179	100718	101143	426	-	hypothetical protein	self_AGKR179	protein
vB_EcoM_LNA4_MZ311864_CDS_0180	101130	101495	366	-	hypothetical protein	self_AGKR180	protein
vB_EcoM_LNA4_MZ311864_CDS_0181	101497	102564	1068	-	nicotinamide-nucleotide adenylyltransferase	self_AGKR181	protein
vB_EcoM_LNA4_MZ311864_CDS_0182	102572	102892	321	-	hypothetical protein	self_AGKR182	protein
vB_EcoM_LNA4_MZ311864_CDS_0183	102926	103132	207	-	hypothetical protein	self_AGKR183	protein
vB_EcoM_LNA4_MZ311864_CDS_0184	103143	103337	195	-	hypothetical protein	self_AGKR184	protein
vB_EcoM_LNA4_MZ311864_CDS_0185	103404	103802	399	-	hypothetical protein	self_AGKR185	protein
vB_EcoM_LNA4_MZ311864_CDS_0186	103799	103999	201	-	hypothetical protein	self_AGKR186	protein
vB_EcoM_LNA4_MZ311864_CDS_0187	104002	104475	474	-	HNH endonuclease	self_AGKR187	protein
vB_EcoM_LNA4_MZ311864_CDS_0188	104502	106916	2415	-	DNA polymerase	self_AGKR188	protein
vB_EcoM_LNA4_MZ311864_CDS_0189	106974	107105	132	+	hypothetical protein	self_AGKR189	protein

vB_EcoM_LNA4_MZ311864_CDS_0190	107243	107746	504	-	HNH endonuclease	self_AGKR190	protein
vB_EcoM_LNA4_MZ311864_CDS_0191	107878	109890	2013	-	DNA primase/helicase	self_AGKR191	protein
vB_EcoM_LNA4_MZ311864_CDS_0192	109900	110046	147	-	hypothetical protein	self_AGKR192	protein
vB_EcoM_LNA4_MZ311864_CDS_0193	110055	110705	651	-	DNA methyltransferase	self_AGKR193	protein
vB_EcoM_LNA4_MZ311864_CDS_0194	110758	111021	264	-	DNA ligase	self_AGKR194	protein
vB_EcoM_LNA4_MZ311864_CDS_0195	111021	111302	282	-	hypothetical protein	self_AGKR195	protein
vB_EcoM_LNA4_MZ311864_CDS_0196	111299	111790	492	-	hypothetical protein	self_AGKR196	protein
vB_EcoM_LNA4_MZ311864_CDS_0197	111783	112106	324	-	DNA ligase	self_AGKR197	protein
vB_EcoM_LNA4_MZ311864_CDS_0198	112087	112266	180	-	hypothetical protein	self_AGKR198	protein
vB_EcoM_LNA4_MZ311864_CDS_0199	112268	112777	510	-	hypothetical protein	self_AGKR199	protein
vB_EcoM_LNA4_MZ311864_CDS_0200	112789	114096	1308	-	helicase	self_AGKR200	protein
vB_EcoM_LNA4_MZ311864_CDS_0201	114096	115376	1281	-	RIIB lysis inhibitor	self_AGKR201	protein
vB_EcoM_LNA4_MZ311864_CDS_0202	115373	117367	1995	-	RIIA lysis inhibitor	self_AGKR202	protein
vB_EcoM_LNA4_MZ311864_CDS_0203	117379	117552	174	-	hypothetical protein	self_AGKR203	protein
vB_EcoM_LNA4_MZ311864_CDS_0204	117557	118129	573	-	endolysin	self_AGKR204	protein
vB_EcoM_LNA4_MZ311864_CDS_0205	118130	118486	357	-	hypothetical protein	self_AGKR205	protein
vB_EcoM_LNA4_MZ311864_CDS_0206	118515	118736	222	-	hypothetical protein	self_AGKR206	protein
vB_EcoM_LNA4_MZ311864_CDS_0207	118848	119894	1047	-	RNA ligase	self_AGKR207	protein
vB_EcoM_LNA4_MZ311864_CDS_0208	119887	120276	390	-	methyltransferase	self_AGKR208	protein
vB_EcoM_LNA4_MZ311864_CDS_0209	120263	120535	273	-	hypothetical protein	self_AGKR209	protein
vB_EcoM_LNA4_MZ311864_CDS_0210	120525	120767	243	-	hypothetical protein	self_AGKR210	protein
vB_EcoM_LNA4_MZ311864_CDS_0211	120767	120967	201	-	hypothetical protein	self_AGKR211	protein
vB_EcoM_LNA4_MZ311864_CDS_0212	120964	121146	183	-	membrane protein	self_AGKR212	protein
vB_EcoM_LNA4_MZ311864_CDS_0213	121146	121637	492	-	phosphatase	self_AGKR213	protein
vB_EcoM_LNA4_MZ311864_CDS_0214	121637	122140	504	-	HNH endonuclease	self_AGKR214	protein
vB_EcoM_LNA4_MZ311864_CDS_0215	122140	122352	213	-	hypothetical protein	self_AGKR215	protein
vB_EcoM_LNA4_MZ311864_CDS_0216	122349	122540	192	-	baseplate wedge subunit	self_AGKR216	protein
vB_EcoM_LNA4_MZ311864_CDS_0217	122525	122722	198	-	hypothetical protein	self_AGKR217	protein
vB_EcoM_LNA4_MZ311864_CDS_0218	122716	123009	294	-	hypothetical protein	self_AGKR218	protein
vB_EcoM_LNA4_MZ311864_CDS_0219	123006	123158	153	-	hypothetical protein	self_AGKR219	protein
vB_EcoM_LNA4_MZ311864_CDS_0220	123155	123310	156	-	hypothetical protein	self_AGKR220	protein
vB_EcoM_LNA4_MZ311864_CDS_0221	123339	123680	342	-	hypothetical protein	self_AGKR221	protein
vB_EcoM_LNA4_MZ311864_CDS_0222	123739	124428	690	-	hypothetical protein	self_AGKR222	protein
vB_EcoM_LNA4_MZ311864_CDS_0223	124438	125646	1209	-	ATPase	self_AGKR223	protein
vB_EcoM_LNA4_MZ311864_CDS_0224	125633	125875	243	-	hypothetical protein	self_AGKR224	protein
vB_EcoM_LNA4_MZ311864_CDS_0225	125860	126231	372	-	hypothetical protein	self_AGKR225	protein
vB_EcoM_LNA4_MZ311864_CDS_0226	126224	126712	489	-	hypothetical protein	self_AGKR226	protein
vB_EcoM_LNA4_MZ311864_CDS_0227	126709	126822	114	-	hypothetical protein	self_AGKR227	protein
vB_EcoM_LNA4_MZ311864_CDS_0228	126786	127142	357	-	hypothetical protein	self_AGKR228	protein
vB_EcoM_LNA4_MZ311864_CDS_0229	127139	127648	510	-	metal-dependent phosphohydrolase	self_AGKR229	protein
vB_EcoM_LNA4_MZ311864_CDS_0230	127612	128676	1065	-	thymidylate kinase	self_AGKR230	protein
vB_EcoM_LNA4_MZ311864_CDS_0231	128691	128783	93	+	hypothetical protein	self_AGKR231	protein
vB_EcoM_LNA4_MZ311864_CDS_0232	128771	129676	906	+	HNH endonuclease	self_AGKR232	protein
vB_EcoM_LNA4_MZ311864_CDS_0233	129676	129936	261	+	tail protein	self_AGKR233	protein
vB_EcoM_LNA4_MZ311864_CDS_0234	129968	131431	1464	+	HNH endonuclease	self_AGKR234	protein
vB_EcoM_LNA4_MZ311864_CDS_0235	131415	131801	387	+	tail protein	self_AGKR235	protein
vB_EcoM_LNA4_MZ311864_CDS_0236	131798	132094	297	+	replication initiation protein	self_AGKR236	protein
vB_EcoM_LNA4_MZ311864_CDS_0237	132234	132329	96	+	hypothetical protein	self_AGKR237	protein
vB_EcoM_LNA4_MZ311864_CDS_0238	132330	132758	429	+	hypothetical protein	self_AGKR238	protein
vB_EcoM_LNA4_MZ311864_CDS_0239	132758	133399	642	+	hypothetical protein	self_AGKR239	protein
vB_EcoM_LNA4_MZ311864_CDS_0240	133396	133734	339	+	MazG-like pyrophosphatase	self_AGKR240	protein
vB_EcoM_LNA4_MZ311864_CDS_0241	133731	134075	345	+	hypothetical protein	self_AGKR241	protein
vB_EcoM_LNA4_MZ311864_CDS_0242	134077	134223	147	+	hypothetical protein	self_AGKR242	protein
vB_EcoM_LNA4_MZ311864_CDS_0243	134348	135172	825	+	hypothetical protein	self_AGKR243	protein
vB_EcoM_LNA4_MZ311864_CDS_0244	135230	135361	132	+	hypothetical protein	self_AGKR244	protein
vB_EcoM_LNA4_MZ311864_CDS_0245	135348	135506	159	+	hypothetical protein	self_AGKR245	protein
vB_EcoM_LNA4_MZ311864_CDS_0246	135503	135676	174	-	hypothetical protein	self_AGKR246	protein
vB_EcoM_LNA4_MZ311864_CDS_0247	135730	135825	96	+	hypothetical protein	self_AGKR247	protein
vB_EcoM_LNA4_MZ311864_CDS_0248	135943	136134	192	+	hypothetical protein	self_AGKR248	protein
vB_EcoM_LNA4_MZ311864_CDS_0249	136180	136296	117	+	hypothetical protein	self_AGKR249	protein
vB_EcoM_LNA4_MZ311864_CDS_0250	136396	136926	531	+	hypothetical protein	self_AGKR250	protein
vB_EcoM_LNA4_MZ311864_CDS_0251	137000	137296	297	+	DNA ligase	self_AGKR251	protein
vB_EcoM_LNA4_MZ311864_CDS_0252	137387	137716	330	+	hypothetical protein	self_AGKR252	protein
vB_EcoM_LNA4_MZ311864_CDS_0253	137716	137817	102	+	hypothetical protein	self_AGKR253	protein
vB_EcoM_LNA4_MZ311864_CDS_0254	137878	138075	198	+	hypothetical protein	self_AGKR254	protein
vB_EcoM_LNA4_MZ311864_CDS_0255	138156	138449	294	+	hypothetical protein	self_AGKR255	protein
vB_EcoM_LNA4_MZ311864_CDS_0256	138453	138542	90	+	hypothetical protein	self_AGKR256	protein
vB_EcoM_LNA4_MZ311864_CDS_0257	138529	138927	399	+	hypothetical protein	self_AGKR257	protein
vB_EcoM_LNA4_MZ311864_CDS_0258	138929	139072	144	+	hypothetical protein	self_AGKR258	protein
vB_EcoM_LNA4_MZ311864_CDS_0259	139085	139270	186	+	hypothetical protein	self_AGKR259	protein

vB_EcoM_LNA4_MZ311864_CDS_0260	139367	139933	567	+	hypothetical protein	self_AGKR260	protein
vB_EcoM_LNA4_MZ311864_CDS_0261	139937	140284	348	+	hypothetical protein	self_AGKR261	protein
vB_EcoM_LNA4_MZ311864_CDS_0262	140299	140502	204	+	hypothetical protein	self_AGKR262	protein
vB_EcoM_LNA4_MZ311864_CDS_0263	140536	140706	171	+	hypothetical protein	self_AGKR263	protein
vB_EcoM_LNA4_MZ311864_CDS_0264	140832	141122	291	+	hypothetical protein	self_AGKR264	protein
vB_EcoM_LNA4_MZ311864_CDS_0265	141199	141300	102	+	hypothetical protein	self_AGKR265	protein
vB_EcoM_LNA4_MZ311864_CDS_0266	141302	141520	219	+	hypothetical protein	self_AGKR266	protein
vB_EcoM_LNA4_MZ311864_CDS_0267	141530	141715	186	+	hypothetical protein	self_AGKR267	protein
vB_EcoM_LNA4_MZ311864_CDS_0268	141800	141916	117	+	membrane protein	self_AGKR268	protein
vB_EcoM_LNA4_MZ311864_CDS_0269	141932	142273	342	+	hypothetical protein	self_AGKR269	protein
vB_EcoM_LNA4_MZ311864_CDS_0270	142327	142560	234	+	hypothetical protein	self_AGKR270	protein
vB_EcoM_LNA4_MZ311864_CDS_0271	142563	142895	333	+	hypothetical protein	self_AGKR271	protein
vB_EcoM_LNA4_MZ311864_CDS_0272	142982	143371	390	+	hypothetical protein	self_AGKR272	protein
vB_EcoM_LNA4_MZ311864_CDS_0273	143422	143646	225	+	hypothetical protein	self_AGKR273	protein
vB_EcoM_LNA4_MZ311864_CDS_0274	143643	143870	228	+	hypothetical protein	self_AGKR274	protein
vB_EcoM_LNA4_MZ311864_CDS_0275	143867	144016	150	+	hypothetical protein	self_AGKR275	protein
vB_EcoM_LNA4_MZ311864_CDS_0276	144013	144351	339	+	hypothetical protein	self_AGKR276	protein
vB_EcoM_LNA4_MZ311864_CDS_0277	144444	144809	366	+	hypothetical protein	self_AGKR277	protein
vB_EcoM_LNA4_MZ311864_CDS_0278	144886	145209	324	+	hypothetical protein	self_AGKR278	protein
vB_EcoM_LNA4_MZ311864_CDS_0279	145259	145624	366	+	hypothetical protein	self_AGKR279	protein
vB_EcoM_LNA4_MZ311864_CDS_0280	145631	145894	264	+	hypothetical protein	self_AGKR280	protein
vB_EcoM_LNA4_MZ311864_CDS_0281	145920	146060	141	+	hypothetical protein	self_AGKR281	protein
vB_EcoM_LNA4_MZ311864_CDS_0282	146111	146218	108	+	hypothetical protein	self_AGKR282	protein
vB_EcoM_LNA4_MZ311864_CDS_0283	146265	146489	225	+	hypothetical protein	self_AGKR283	protein
vB_EcoM_LNA4_MZ311864_CDS_0284	146741	146839	99	+	hypothetical protein	self_AGKR284	protein
vB_EcoM_LNA4_MZ311864_CDS_0285	146859	147047	189	+	hypothetical protein	self_AGKR285	protein
vB_EcoM_LNA4_MZ311864_CDS_0286	147139	147261	123	+	hypothetical protein	self_AGKR286	protein
vB_EcoM_LNA4_MZ311864_CDS_0287	147267	147371	105	+	hypothetical protein	self_AGKR287	protein

Annotation of phage vB_EcoM_LNA5.

Locus_tag	Start	End	Length_nt	Strand	Function_by_amino_acid_sequence	Protein_ID	Category
vB_EcoP_LNA5_MZ311861_CDS_0001	2	229	228	-	host range and adsorption protein	self_UTCO1	protein
vB_EcoP_LNA5_MZ311861_CDS_0002	234	458	225	-	hypothetical protein	self_UTCO2	protein
vB_EcoP_LNA5_MZ311861_CDS_0003	473	745	273	-	hypothetical protein	self_UTCO3	protein
vB_EcoP_LNA5_MZ311861_CDS_0004	882	974	93	-	hypothetical protein	self_UTCO4	protein
vB_EcoP_LNA5_MZ311861_CDS_0005	956	1828	873	-	exonuclease	self_UTCO5	protein
vB_EcoP_LNA5_MZ311861_CDS_0006	1819	2028	210	-	hypothetical protein	self_UTCO6	protein
vB_EcoP_LNA5_MZ311861_CDS_0007	2025	2309	285	-	hypothetical protein	self_UTCO7	protein
vB_EcoP_LNA5_MZ311861_CDS_0008	2309	4480	2172	-	DNA polymerase I	self_UTCO8	protein
vB_EcoP_LNA5_MZ311861_CDS_0009	4544	4672	129	-	hypothetical protein	self_UTCO9	protein
vB_EcoP_LNA5_MZ311861_CDS_0010	4692	4781	90	-	hypothetical protein	self_UTCO10	protein
vB_EcoP_LNA5_MZ311861_CDS_0011	4760	6460	1701	-	DNA primase/helicase	self_UTCO11	protein
vB_EcoP_LNA5_MZ311861_CDS_0012	6617	7075	459	-	amidase	self_UTCO12	protein
vB_EcoP_LNA5_MZ311861_CDS_0013	7065	7556	492	-	hypothetical protein	self_UTCO13	protein
vB_EcoP_LNA5_MZ311861_CDS_0014	7556	8008	453	-	endonuclease	self_UTCO14	protein
vB_EcoP_LNA5_MZ311861_CDS_0015	8008	8709	702	-	Gp2.5-like ssDNA binding protein and ssDNA annealing protein	self_UTCO15	protein
vB_EcoP_LNA5_MZ311861_CDS_0016	8768	8926	159	-	RNA polymerase inhibitor	self_UTCO16	protein
vB_EcoP_LNA5_MZ311861_CDS_0017	8923	9018	96	-	hypothetical protein	self_UTCO17	protein
vB_EcoP_LNA5_MZ311861_CDS_0018	9005	9478	474	-	nucleotide kinase	self_UTCO18	protein
vB_EcoP_LNA5_MZ311861_CDS_0019	9549	9803	255	-	hypothetical protein	self_UTCO19	protein
vB_EcoP_LNA5_MZ311861_CDS_0020	9936	11012	1077	-	DNA ligase	self_UTCO20	protein
vB_EcoP_LNA5_MZ311861_CDS_0021	11012	11263	252	-	dGTPase inhibitor	self_UTCO21	protein
vB_EcoP_LNA5_MZ311861_CDS_0022	11275	11448	174	-	hypothetical protein	self_UTCO22	protein
vB_EcoP_LNA5_MZ311861_CDS_0023	11616	11831	216	-	hypothetical protein	self_UTCO23	protein
vB_EcoP_LNA5_MZ311861_CDS_0024	11842	12072	231	-	hypothetical protein	self_UTCO24	protein
vB_EcoP_LNA5_MZ311861_CDS_0025	12088	14769	2682	-	N4-like RNA polymerase	self_UTCO25	protein
vB_EcoP_LNA5_MZ311861_CDS_0026	14867	15193	327	-	hypothetical protein	self_UTCO26	protein
vB_EcoP_LNA5_MZ311861_CDS_0027	15193	15396	204	-	hypothetical protein	self_UTCO27	protein
vB_EcoP_LNA5_MZ311861_CDS_0028	15393	15497	105	-	hypothetical protein	self_UTCO28	protein
vB_EcoP_LNA5_MZ311861_CDS_0029	15630	15857	228	-	hypothetical protein	self_UTCO29	protein
vB_EcoP_LNA5_MZ311861_CDS_0030	15921	16070	150	-	hypothetical protein	self_UTCO30	protein
vB_EcoP_LNA5_MZ311861_CDS_0031	16070	16573	504	-	ocr-like anti-restriction	self_UTCO31	protein
vB_EcoP_LNA5_MZ311861_CDS_0032	16657	16902	246	-	hypothetical protein	self_UTCO32	protein
vB_EcoP_LNA5_MZ311861_CDS_0033	17115	17333	219	-	hypothetical protein	self_UTCO33	protein
vB_EcoP_LNA5_MZ311861_CDS_0034	17486	17620	135	+	hypothetical protein	self_UTCO34	protein
vB_EcoP_LNA5_MZ311861_CDS_0035	17681	17776	96	+	hypothetical protein	self_UTCO35	protein
vB_EcoP_LNA5_MZ311861_CDS_0036	17790	17879	90	+	hypothetical protein	self_UTCO36	protein
vB_EcoP_LNA5_MZ311861_CDS_0037	17919	18077	159	-	hypothetical protein	self_UTCO37	protein
vB_EcoP_LNA5_MZ311861_CDS_0038	18052	18144	93	-	hypothetical protein	self_UTCO38	protein
vB_EcoP_LNA5_MZ311861_CDS_0039	18229	18345	117	-	hypothetical protein	self_UTCO39	protein
vB_EcoP_LNA5_MZ311861_CDS_0040	18375	20105	1731	-	terminase large subunit	self_UTCO40	protein
vB_EcoP_LNA5_MZ311861_CDS_0041	20135	20584	450	-	Rz-like spanin	self_UTCO41	protein
vB_EcoP_LNA5_MZ311861_CDS_0042	20686	20949	264	-	terminase small subunit	self_UTCO42	protein
vB_EcoP_LNA5_MZ311861_CDS_0043	20946	21140	195	-	holin	self_UTCO43	protein
vB_EcoP_LNA5_MZ311861_CDS_0044	21285	23465	2181	-	tail fiber protein	self_UTCO44	protein
vB_EcoP_LNA5_MZ311861_CDS_0045	23530	27417	3888	-	internal virion protein with endolysin domain	self_UTCO45	protein
vB_EcoP_LNA5_MZ311861_CDS_0046	27399	27848	450	-	HNH endonuclease	self_UTCO46	protein
vB_EcoP_LNA5_MZ311861_CDS_0047	27851	30133	2283	-	internal virion protein	self_UTCO47	protein
vB_EcoP_LNA5_MZ311861_CDS_0048	30145	30732	588	-	internal virion protein	self_UTCO48	protein
vB_EcoP_LNA5_MZ311861_CDS_0049	30711	31187	477	-	internal virion protein	self_UTCO49	protein
vB_EcoP_LNA5_MZ311861_CDS_0050	31264	33621	2358	-	tail protein	self_UTCO50	protein
vB_EcoP_LNA5_MZ311861_CDS_0051	33631	33858	228	-	RNA-binding protein	self_UTCO51	protein
vB_EcoP_LNA5_MZ311861_CDS_0052	33858	34424	567	-	tail protein	self_UTCO52	protein
vB_EcoP_LNA5_MZ311861_CDS_0053	34436	34525	90	-	hypothetical protein	self_UTCO53	protein
vB_EcoP_LNA5_MZ311861_CDS_0054	34584	34790	207	-	hypothetical protein	self_UTCO54	protein
vB_EcoP_LNA5_MZ311861_CDS_0055	34847	35896	1050	-	major head protein	self_UTCO55	protein
vB_EcoP_LNA5_MZ311861_CDS_0056	36023	36910	888	-	head assembly	self_UTCO56	protein
vB_EcoP_LNA5_MZ311861_CDS_0057	37000	38568	1569	-	head-tail adaptor	self_UTCO57	protein
vB_EcoP_LNA5_MZ311861_CDS_0058	38580	38828	249	-	host range and adsorption protein	self_UTCO58	protein

Annotation of phage vB_EcoM_SHAK7704.

Locus_tag	Start	End	Length_nt	Strand	Function_by_amino_acid_sequence	Protein_ID	Category
vB_EcoM_SHAK7704_OR594181_CDS_0001	1	1896	1896	+	terminase large subunit	self_GXWA1	protein
vB_EcoM_SHAK7704_OR594181_CDS_0002	2054	2335	282	+	acetyl-CoA acetyltransferase	self_GXWA2	protein
vB_EcoM_SHAK7704_OR594181_CDS_0003	2280	2636	357	+	acetyl-CoA acetyltransferase	self_GXWA3	protein
vB_EcoM_SHAK7704_OR594181_CDS_0004	2638	3327	690	+	hypothetical protein	self_GXWA4	protein
vB_EcoM_SHAK7704_OR594181_CDS_0005	3327	3671	345	+	endolysin	self_GXWA5	protein
vB_EcoM_SHAK7704_OR594181_CDS_0006	3665	3802	138	+	hypothetical protein	self_GXWA6	protein
vB_EcoM_SHAK7704_OR594181_CDS_0007	3780	3929	150	+	holin	self_GXWA7	protein
vB_EcoM_SHAK7704_OR594181_CDS_0008	4019	4216	198	+	Rz-like spanin	self_GXWA8	protein
vB_EcoM_SHAK7704_OR594181_CDS_0009	4348	4683	336	+	hypothetical protein	self_GXWA9	protein
vB_EcoM_SHAK7704_OR594181_CDS_0010	4751	7186	2436	+	tail spike protein	self_GXWA10	protein
vB_EcoM_SHAK7704_OR594181_CDS_0011	7164	7469	306	+	hypothetical protein	self_GXWA11	protein
vB_EcoM_SHAK7704_OR594181_CDS_0012	7456	7554	99	+	hypothetical protein	self_GXWA12	protein
vB_EcoM_SHAK7704_OR594181_CDS_0013	7578	7775	198	+	hypothetical protein	self_GXWA13	protein
vB_EcoM_SHAK7704_OR594181_CDS_0014	7953	8042	90	-	hypothetical protein	self_GXWA14	protein
vB_EcoM_SHAK7704_OR594181_CDS_0015	8134	8238	105	-	hypothetical protein	self_GXWA15	protein
vB_EcoM_SHAK7704_OR594181_CDS_0016	8281	8550	270	+	hypothetical protein	self_GXWA16	protein
vB_EcoM_SHAK7704_OR594181_CDS_0017	8537	8716	180	+	hypothetical protein	self_GXWA17	protein
vB_EcoM_SHAK7704_OR594181_CDS_0018	8716	8940	225	+	hypothetical protein	self_GXWA18	protein
vB_EcoM_SHAK7704_OR594181_CDS_0019	9016	9150	135	-	hypothetical protein	self_GXWA19	protein
vB_EcoM_SHAK7704_OR594181_CDS_0020	9225	9437	213	+	hypothetical protein	self_GXWA20	protein
vB_EcoM_SHAK7704_OR594181_CDS_0021	9500	9595	96	+	hypothetical protein	self_GXWA21	protein
vB_EcoM_SHAK7704_OR594181_CDS_0022	9646	9798	153	+	hypothetical protein	self_GXWA22	protein
vB_EcoM_SHAK7704_OR594181_CDS_0023	9862	10218	357	+	ocr-like anti-restriction	self_GXWA23	protein
vB_EcoM_SHAK7704_OR594181_CDS_0024	10221	10334	114	+	hypothetical protein	self_GXWA24	protein
vB_EcoM_SHAK7704_OR594181_CDS_0025	10411	10638	228	+	hypothetical protein	self_GXWA25	protein
vB_EcoM_SHAK7704_OR594181_CDS_0026	10703	11587	885	+	hypothetical protein	self_GXWA26	protein
vB_EcoM_SHAK7704_OR594181_CDS_0027	11651	14278	2628	+	RNA polymerase	self_GXWA27	protein
vB_EcoM_SHAK7704_OR594181_CDS_0028	14612	14740	129	+	hypothetical protein	self_GXWA28	protein
vB_EcoM_SHAK7704_OR594181_CDS_0029	14744	14962	219	+	hypothetical protein	self_GXWA29	protein
vB_EcoM_SHAK7704_OR594181_CDS_0030	14962	16950	1989	+	DNA primase/helicase	self_GXWA30	protein
vB_EcoM_SHAK7704_OR594181_CDS_0031	16950	17111	162	+	hypothetical protein	self_GXWA31	protein
vB_EcoM_SHAK7704_OR594181_CDS_0032	17184	17867	684	+	hypothetical protein	self_GXWA32	protein
vB_EcoM_SHAK7704_OR594181_CDS_0033	17878	18093	216	+	hypothetical protein	self_GXWA33	protein
vB_EcoM_SHAK7704_OR594181_CDS_0034	18083	18292	210	+	hypothetical protein	self_GXWA34	protein
vB_EcoM_SHAK7704_OR594181_CDS_0035	18351	18503	153	+	hypothetical protein	self_GXWA35	protein
vB_EcoM_SHAK7704_OR594181_CDS_0036	18490	21036	2547	+	DNA polymerase	self_GXWA36	protein
vB_EcoM_SHAK7704_OR594181_CDS_0037	21036	21152	117	+	hypothetical protein	self_GXWA37	protein
vB_EcoM_SHAK7704_OR594181_CDS_0038	21152	21250	99	+	hypothetical protein	self_GXWA38	protein
vB_EcoM_SHAK7704_OR594181_CDS_0039	21407	21784	378	+	polysaccharide chain length determinant protein	self_GXWA39	protein
vB_EcoM_SHAK7704_OR594181_CDS_0040	21790	22089	300	+	hypothetical protein	self_GXWA40	protein
vB_EcoM_SHAK7704_OR594181_CDS_0041	22167	22973	807	+	hypothetical protein	self_GXWA41	protein
vB_EcoM_SHAK7704_OR594181_CDS_0042	22983	23201	219	+	hypothetical protein	self_GXWA42	protein
vB_EcoM_SHAK7704_OR594181_CDS_0043	23307	23678	372	+	hypothetical protein	self_GXWA43	protein
vB_EcoM_SHAK7704_OR594181_CDS_0044	23782	23898	117	+	hypothetical protein	self_GXWA44	protein
vB_EcoM_SHAK7704_OR594181_CDS_0045	23981	25009	1029	+	exonuclease	self_GXWA45	protein
vB_EcoM_SHAK7704_OR594181_CDS_0046	24994	25407	414	+	endonuclease VII	self_GXWA46	protein
vB_EcoM_SHAK7704_OR594181_CDS_0047	25373	26407	1035	+	metallo-phosphoesterase	self_GXWA47	protein
vB_EcoM_SHAK7704_OR594181_CDS_0048	26476	26946	471	+	nucleotide kinase	self_GXWA48	protein
vB_EcoM_SHAK7704_OR594181_CDS_0049	26948	27064	117	+	hypothetical protein	self_GXWA49	protein
vB_EcoM_SHAK7704_OR594181_CDS_0050	27061	28008	948	+	ATP-dependent DNA ligase	self_GXWA50	protein
vB_EcoM_SHAK7704_OR594181_CDS_0051	27980	28195	216	+	hypothetical protein	self_GXWA51	protein
vB_EcoM_SHAK7704_OR594181_CDS_0052	28200	28328	129	+	hypothetical protein	self_GXWA52	protein
vB_EcoM_SHAK7704_OR594181_CDS_0053	28325	28786	462	+	acetyltransferase	self_GXWA53	protein
vB_EcoM_SHAK7704_OR594181_CDS_0054	28796	29005	210	+	hypothetical protein	self_GXWA54	protein
vB_EcoM_SHAK7704_OR594181_CDS_0055	29007	30557	1551	+	head-tail adaptor	self_GXWA55	protein
vB_EcoM_SHAK7704_OR594181_CDS_0056	30557	31477	921	+	head scaffolding protein	self_GXWA56	protein
vB_EcoM_SHAK7704_OR594181_CDS_0057	31554	32762	1209	+	major head protein	self_GXWA57	protein
vB_EcoM_SHAK7704_OR594181_CDS_0058	32816	33553	738	+	tail protein	self_GXWA58	protein
vB_EcoM_SHAK7704_OR594181_CDS_0059	33553	35955	2403	+	tail protein	self_GXWA59	protein
vB_EcoM_SHAK7704_OR594181_CDS_0060	35955	36677	723	+	internal virion protein	self_GXWA60	protein
vB_EcoM_SHAK7704_OR594181_CDS_0061	36677	39625	2949	+	lysozyme domain-containing protein	self_GXWA61	protein
vB_EcoM_SHAK7704_OR594181_CDS_0062	39690	42998	3309	+	internal virion protein with endolysin domain	self_GXWA62	protein
vB_EcoM_SHAK7704_OR594181_CDS_0063	42998	43960	963	+	tail fiber protein	self_GXWA63	protein
vB_EcoM_SHAK7704_OR594181_CDS_0064	43970	44164	195	+	holin	self_GXWA64	protein

Annotation of phage vB_EcoM_SHAK7854.

Locus_tag	Start	End	Length_nt	Strand	Function_by_amino_acid_sequence	Protein_ID	Category
vB_EcoM_SHAK7854_OR594182_CDS_0001	1	1731	1731	+	terminase large subunit	self_NJZI1	protein
vB_EcoM_SHAK7854_OR594182_CDS_0002	1761	1877	117	+	hypothetical protein	self_NJZI2	protein
vB_EcoM_SHAK7854_OR594182_CDS_0003	1962	2054	93	+	hypothetical protein	self_NJZI3	protein
vB_EcoM_SHAK7854_OR594182_CDS_0004	2090	2224	135	+	hypothetical protein	self_NJZI4	protein
vB_EcoM_SHAK7854_OR594182_CDS_0005	2401	2571	171	+	hypothetical protein	self_NJZI5	protein
vB_EcoM_SHAK7854_OR594182_CDS_0006	3044	3184	141	+	hypothetical protein	self_NJZI6	protein
vB_EcoM_SHAK7854_OR594182_CDS_0007	3398	3652	255	+	hypothetical protein	self_NJZI7	protein
vB_EcoM_SHAK7854_OR594182_CDS_0008	3726	4004	279	+	hypothetical protein	self_NJZI8	protein
vB_EcoM_SHAK7854_OR594182_CDS_0009	4085	4588	504	+	ocr-like anti-restriction	self_NJZI9	protein
vB_EcoM_SHAK7854_OR594182_CDS_0010	4588	4737	150	+	hypothetical protein	self_NJZI10	protein
vB_EcoM_SHAK7854_OR594182_CDS_0011	4801	5028	228	+	hypothetical protein	self_NJZI11	protein
vB_EcoM_SHAK7854_OR594182_CDS_0012	5161	5265	105	+	hypothetical protein	self_NJZI12	protein
vB_EcoM_SHAK7854_OR594182_CDS_0013	5253	5465	213	+	hypothetical protein	self_NJZI13	protein
vB_EcoM_SHAK7854_OR594182_CDS_0014	5465	5791	327	+	hypothetical protein	self_NJZI14	protein
vB_EcoM_SHAK7854_OR594182_CDS_0015	5889	8570	2682	+	N4-like RNA polymerase	self_NJZI15	protein
vB_EcoM_SHAK7854_OR594182_CDS_0016	8586	8816	231	+	hypothetical protein	self_NJZI16	protein
vB_EcoM_SHAK7854_OR594182_CDS_0017	8827	9042	216	+	hypothetical protein	self_NJZI17	protein
vB_EcoM_SHAK7854_OR594182_CDS_0018	9065	9181	117	+	hypothetical protein	self_NJZI18	protein
vB_EcoM_SHAK7854_OR594182_CDS_0019	9214	9393	180	+	hypothetical protein	self_NJZI19	protein
vB_EcoM_SHAK7854_OR594182_CDS_0020	9397	9648	252	+	dGTPase inhibitor	self_NJZI20	protein
vB_EcoM_SHAK7854_OR594182_CDS_0021	9648	10712	1065	+	DNA ligase	self_NJZI21	protein
vB_EcoM_SHAK7854_OR594182_CDS_0022	10845	11099	255	+	hypothetical protein	self_NJZI22	protein
vB_EcoM_SHAK7854_OR594182_CDS_0023	11099	11635	537	+	nucleotide kinase	self_NJZI23	protein
vB_EcoM_SHAK7854_OR594182_CDS_0024	11716	11814	99	+	hypothetical protein	self_NJZI24	protein
vB_EcoM_SHAK7854_OR594182_CDS_0025	11811	11969	159	+	RNA polymerase inhibitor	self_NJZI25	protein
vB_EcoM_SHAK7854_OR594182_CDS_0026	12028	12729	702	+	Gp2.5-like ssDNA binding protein and ssDNA annealing protein	self_NJZI26	protein
vB_EcoM_SHAK7854_OR594182_CDS_0027	12729	13181	453	+	endonuclease	self_NJZI27	protein
vB_EcoM_SHAK7854_OR594182_CDS_0028	13181	13672	492	+	hypothetical protein	self_NJZI28	protein
vB_EcoM_SHAK7854_OR594182_CDS_0029	13662	14120	459	+	amidase	self_NJZI29	protein
vB_EcoM_SHAK7854_OR594182_CDS_0030	14145	14360	216	+	hypothetical protein	self_NJZI30	protein
vB_EcoM_SHAK7854_OR594182_CDS_0031	14429	16129	1701	+	DNA primase/helicase	self_NJZI31	protein
vB_EcoM_SHAK7854_OR594182_CDS_0032	16177	16305	129	+	hypothetical protein	self_NJZI32	protein
vB_EcoM_SHAK7854_OR594182_CDS_0033	16369	18540	2172	+	DNA polymerase I	self_NJZI33	protein
vB_EcoM_SHAK7854_OR594182_CDS_0034	18540	18824	285	+	hypothetical protein	self_NJZI34	protein
vB_EcoM_SHAK7854_OR594182_CDS_0035	18821	19030	210	+	hypothetical protein	self_NJZI35	protein
vB_EcoM_SHAK7854_OR594182_CDS_0036	19021	19893	873	+	exonuclease	self_NJZI36	protein
vB_EcoM_SHAK7854_OR594182_CDS_0037	19875	19967	93	+	hypothetical protein	self_NJZI37	protein
vB_EcoM_SHAK7854_OR594182_CDS_0038	20104	20376	273	+	hypothetical protein	self_NJZI38	protein
vB_EcoM_SHAK7854_OR594182_CDS_0039	20391	20615	225	+	hypothetical protein	self_NJZI39	protein
vB_EcoM_SHAK7854_OR594182_CDS_0040	20620	21042	423	+	host range and adsorption protein	self_NJZI40	protein
vB_EcoM_SHAK7854_OR594182_CDS_0041	21035	21283	249	+	host range and adsorption protein	self_NJZI41	protein
vB_EcoM_SHAK7854_OR594182_CDS_0042	21295	22863	1569	+	head-tail adaptor	self_NJZI42	protein
vB_EcoM_SHAK7854_OR594182_CDS_0043	22953	23837	885	+	head assembly	self_NJZI43	protein
vB_EcoM_SHAK7854_OR594182_CDS_0044	23964	25013	1050	+	major head protein	self_NJZI44	protein
vB_EcoM_SHAK7854_OR594182_CDS_0045	25070	25276	207	+	hypothetical protein	self_NJZI45	protein
vB_EcoM_SHAK7854_OR594182_CDS_0046	25341	25430	90	+	hypothetical protein	self_NJZI46	protein
vB_EcoM_SHAK7854_OR594182_CDS_0047	25442	26008	567	+	tail protein	self_NJZI47	protein
vB_EcoM_SHAK7854_OR594182_CDS_0048	26008	26235	228	+	RNA-binding protein	self_NJZI48	protein
vB_EcoM_SHAK7854_OR594182_CDS_0049	26245	28602	2358	+	tail protein	self_NJZI49	protein
vB_EcoM_SHAK7854_OR594182_CDS_0050	28689	29159	471	+	internal virion protein	self_NJZI50	protein
vB_EcoM_SHAK7854_OR594182_CDS_0051	29144	29731	588	+	internal virion protein	self_NJZI51	protein
vB_EcoM_SHAK7854_OR594182_CDS_0052	29743	32025	2283	+	internal virion protein	self_NJZI52	protein
vB_EcoM_SHAK7854_OR594182_CDS_0053	32030	35917	3888	+	internal virion protein with endolysin domain	self_NJZI53	protein
vB_EcoM_SHAK7854_OR594182_CDS_0054	35982	38513	2532	+	tail spike protein	self_NJZI54	protein
vB_EcoM_SHAK7854_OR594182_CDS_0055	38551	38703	153	+	hypothetical protein	self_NJZI55	protein
vB_EcoM_SHAK7854_OR594182_CDS_0056	38700	38894	195	+	holin	self_NJZI56	protein
vB_EcoM_SHAK7854_OR594182_CDS_0057	38891	39154	264	+	terminase small subunit	self_NJZI57	protein

Annotation of phage vB_EcoM_SHAK9454.

Locus_tag	Start	End	Length_nt	Strand	Function_by_amino_acid_sequence	Protein_ID	Category
vB_EcoM_SHAK9454_OR594184_CDS_0001	1	1731	1731	+	terminase large subunit	self_LVCG1	protein
vB_EcoM_SHAK9454_OR594184_CDS_0002	1962	2054	93	+	hypothetical protein	self_LVCG2	protein
vB_EcoM_SHAK9454_OR594184_CDS_0003	2090	2224	135	+	hypothetical protein	self_LVCG3	protein
vB_EcoM_SHAK9454_OR594184_CDS_0004	2369	2458	90	+	hypothetical protein	self_LVCG4	protein
vB_EcoM_SHAK9454_OR594184_CDS_0005	2584	2775	192	+	hypothetical protein	self_LVCG5	protein
vB_EcoM_SHAK9454_OR594184_CDS_0006	2988	3206	219	+	hypothetical protein	self_LVCG6	protein
vB_EcoM_SHAK9454_OR594184_CDS_0007	3419	3664	246	+	hypothetical protein	self_LVCG7	protein
vB_EcoM_SHAK9454_OR594184_CDS_0008	3748	4251	504	+	ocr-like anti-restriction	self_LVCG8	protein
vB_EcoM_SHAK9454_OR594184_CDS_0009	4251	4400	150	+	hypothetical protein	self_LVCG9	protein
vB_EcoM_SHAK9454_OR594184_CDS_0010	4464	4691	228	+	hypothetical protein	self_LVCG10	protein
vB_EcoM_SHAK9454_OR594184_CDS_0011	4824	4928	105	+	hypothetical protein	self_LVCG11	protein
vB_EcoM_SHAK9454_OR594184_CDS_0012	4916	5128	213	+	hypothetical protein	self_LVCG12	protein
vB_EcoM_SHAK9454_OR594184_CDS_0013	5128	5454	327	+	hypothetical protein	self_LVCG13	protein
vB_EcoM_SHAK9454_OR594184_CDS_0014	5552	8233	2682	+	N4-like RNA polymerase	self_LVCG14	protein
vB_EcoM_SHAK9454_OR594184_CDS_0015	8249	8479	231	+	hypothetical protein	self_LVCG15	protein
vB_EcoM_SHAK9454_OR594184_CDS_0016	8490	8705	216	+	hypothetical protein	self_LVCG16	protein
vB_EcoM_SHAK9454_OR594184_CDS_0017	8728	8844	117	+	hypothetical protein	self_LVCG17	protein
vB_EcoM_SHAK9454_OR594184_CDS_0018	8874	9053	180	+	hypothetical protein	self_LVCG18	protein
vB_EcoM_SHAK9454_OR594184_CDS_0019	9058	9315	258	+	dGTPase inhibitor	self_LVCG19	protein
vB_EcoM_SHAK9454_OR594184_CDS_0020	9320	10363	1044	+	DNA ligase	self_LVCG20	protein
vB_EcoM_SHAK9454_OR594184_CDS_0021	10496	10750	255	+	hypothetical protein	self_LVCG21	protein
vB_EcoM_SHAK9454_OR594184_CDS_0022	10821	11294	474	+	nucleotide kinase	self_LVCG22	protein
vB_EcoM_SHAK9454_OR594184_CDS_0023	11281	11376	96	+	hypothetical protein	self_LVCG23	protein
vB_EcoM_SHAK9454_OR594184_CDS_0024	11373	11531	159	+	RNA polymerase inhibitor	self_LVCG24	protein
vB_EcoM_SHAK9454_OR594184_CDS_0025	11590	12288	699	+	Gp2.5-like ssDNA binding protein and ssDNA annealing protein	self_LVCG25	protein
vB_EcoM_SHAK9454_OR594184_CDS_0026	12288	12734	447	+	endonuclease	self_LVCG26	protein
vB_EcoM_SHAK9454_OR594184_CDS_0027	12721	13173	453	+	endonuclease	self_LVCG27	protein
vB_EcoM_SHAK9454_OR594184_CDS_0028	13173	13625	453	+	hypothetical protein	self_LVCG28	protein
vB_EcoM_SHAK9454_OR594184_CDS_0029	13615	14073	459	+	amidase	self_LVCG29	protein
vB_EcoM_SHAK9454_OR594184_CDS_0030	14138	15838	1701	+	DNA primase/helicase	self_LVCG30	protein
vB_EcoM_SHAK9454_OR594184_CDS_0031	15886	16014	129	+	hypothetical protein	self_LVCG31	protein
vB_EcoM_SHAK9454_OR594184_CDS_0032	16083	16514	432	+	endonuclease VII	self_LVCG32	protein
vB_EcoM_SHAK9454_OR594184_CDS_0033	16507	18678	2172	+	DNA polymerase I	self_LVCG33	protein
vB_EcoM_SHAK9454_OR594184_CDS_0034	18678	18962	285	+	hypothetical protein	self_LVCG34	protein
vB_EcoM_SHAK9454_OR594184_CDS_0035	18959	19168	210	+	hypothetical protein	self_LVCG35	protein
vB_EcoM_SHAK9454_OR594184_CDS_0036	19159	20031	873	+	exonuclease	self_LVCG36	protein
vB_EcoM_SHAK9454_OR594184_CDS_0037	20013	20105	93	+	hypothetical protein	self_LVCG37	protein
vB_EcoM_SHAK9454_OR594184_CDS_0038	20242	20514	273	+	hypothetical protein	self_LVCG38	protein
vB_EcoM_SHAK9454_OR594184_CDS_0039	20529	20753	225	+	hypothetical protein	self_LVCG39	protein
vB_EcoM_SHAK9454_OR594184_CDS_0040	20758	21180	423	+	host range and adsorption protein	self_LVCG40	protein
vB_EcoM_SHAK9454_OR594184_CDS_0041	21173	21421	249	+	host range and adsorption protein	self_LVCG41	protein
vB_EcoM_SHAK9454_OR594184_CDS_0042	21433	23001	1569	+	head-tail adaptor	self_LVCG42	protein
vB_EcoM_SHAK9454_OR594184_CDS_0043	23098	23982	885	+	head assembly	self_LVCG43	protein
vB_EcoM_SHAK9454_OR594184_CDS_0044	24109	25158	1050	+	major head protein	self_LVCG44	protein
vB_EcoM_SHAK9454_OR594184_CDS_0045	25257	25421	165	+	hypothetical protein	self_LVCG45	protein
vB_EcoM_SHAK9454_OR594184_CDS_0046	25481	25570	90	+	hypothetical protein	self_LVCG46	protein
vB_EcoM_SHAK9454_OR594184_CDS_0047	25582	26148	567	+	tail protein	self_LVCG47	protein
vB_EcoM_SHAK9454_OR594184_CDS_0048	26148	26375	228	+	RNA-binding protein	self_LVCG48	protein
vB_EcoM_SHAK9454_OR594184_CDS_0049	26385	28742	2358	+	tail protein	self_LVCG49	protein
vB_EcoM_SHAK9454_OR594184_CDS_0050	28819	29295	477	+	internal virion protein	self_LVCG50	protein
vB_EcoM_SHAK9454_OR594184_CDS_0051	29274	29861	588	+	internal virion protein	self_LVCG51	protein
vB_EcoM_SHAK9454_OR594184_CDS_0052	29873	32155	2283	+	internal virion protein	self_LVCG52	protein
vB_EcoM_SHAK9454_OR594184_CDS_0053	32160	36047	3888	+	internal virion protein with endolysin domain	self_LVCG53	protein
vB_EcoM_SHAK9454_OR594184_CDS_0054	36112	38292	2181	+	tail fiber protein	self_LVCG54	protein
vB_EcoM_SHAK9454_OR594184_CDS_0055	38437	38631	195	+	holin	self_LVCG55	protein
vB_EcoM_SHAK9454_OR594184_CDS_0056	38628	38891	264	+	terminase small subunit	self_LVCG56	protein

Supplementary Table S2

The list of interesting proteins occuring in genomes of the analyzed anti-UPEC bacteriophages. Proteins identified in this work (72 in total) are marked in bold.

Phage_name	Locus_tag	Start	End	Length_nt	Strand	Function_by_amino_acid_sequence	Protein_ID	Category
vB_EcoP_ACG-C91_JN986844	ACG-C91_0013	6892	8880	1989	+	putative DNA helicase	AFH19838.1	protein
vB_EcoP_ACG-C91_JN986844	ACG-C91_0018	10416	12962	2547	+	putative DNA polymerase	AFH19843.1	protein
vB_EcoP_ACG-C91_JN986844	ACG-C91_0042	35770	35994	225	+	putative holin	AFH19868.1	protein
vB_EcoP_ACG-C91_JN986844	ACG-C91_0043	35958	36344	387	+	putative HNH endonuclease	AFH19869.1	protein
vB_EcoS_ACG-M12_JN986845	ACG-M12_0031	17224	17721	498	+	putative HNH endonuclease	AFH19912.1	protein
vB_EcoS_ACG-M12_JN986845	ACG-M12_0061	39194	39688	495	+	putative HNH endonuclease	AFH19943.1	protein
vB_EcoS_ACG-M12_JN986845	ACG-M12_0065	41170	41385	216	+	putative holin	AFH19947.1	protein
vB_EcoS_ACG-M12_JN986845	ACG-M12_0066	41386	41871	486	+	putative endolysin	AFH19948.1	protein
vB_EcoM_ACG-C40_JN986846	ACG-C40_0012	7634	8317	684	-	exonuclease A	AFH19971.1	protein
vB_EcoM_ACG-C40_JN986846	ACG-C40_0015	8819	10138	1320	-	DNA helicase	AFH19974.1	protein
vB_EcoM_ACG-C40_JN986846	ACG-C40_0048	27196	29892	2697	-	DNA polymerase	AFH20008.1	protein
vB_EcoM_ACG-C40_JN986846	ACG-C40_0077	42558	43292	735	-	homing endonuclease	AFH20038.1	protein
vB_EcoM_ACG-C40_JN986846	ACG-C40_0079	45195	45668	474	-	packaging and recombination endonuclease VII	AFH20040.1	protein
vB_EcoM_ACG-C40_JN986846	ACG-C40_0116	60119	60580	462	-	site-specific RNA endonuclease	AFH20078.1	protein
vB_EcoM_ACG-C40_JN986846	ACG-C40_0179	108998	109615	618	+	homing endonuclease	AFH20143.1	protein
vB_EcoM_ACG-C40_JN986846	ACG-C40_0180	109618	111126	1509	+	RNA-DNA and DNA-DNA helicase	AFH20144.1	protein
vB_EcoM_ACG-C40_JN986846	ACG-C40_0181	111123	111815	693	+	homing endonuclease	AFH20145.1	protein
vB_EcoM_ACG-C40_JN986846	ACG-C40_0182	111841	112071	231	+	RNA-DNA and DNA-DNA helicase	AFH20146.1	protein
vB_EcoM_ACG-C40_JN986846	ACG-C40_0225	135609	136019	411	-	endonuclease II	AFH20190.1	protein
vB_EcoM_ACG-C40_JN986846	ACG-C40_0231	140558	141295	738	-	homing endonuclease	AFH20196.1	protein
vB_EcoM_ACG-C40_JN986846	ACG-C40_0240	145167	145799	633	-	homing endonuclease	AFH20206.1	protein
vB_EcoM_ACG-C40_JN986846	ACG-C40_0250	157742	158398	657	+	holin lysis mediator t	AFH20216.1	protein
vB_EcoM_ACG-C40_JN986846	ACG-C40_0253	158833	159111	279	-	inhibitor of MrcBC restriction endonuclease	AFH20219.1	protein
vB_EcoM_ACG-C40_JN986846	ACG-C40_0270	165365	165841	477	-	DNA endonuclease IV	AFH20238.1	protein
St11Ph5_MG208881	St11Ph5_00047	22560	23870	1311	+	DNA helicase	ATS92510.1	protein
St11Ph5_MG208881	St11Ph5_00049	24421	27000	2580	+	DNA polymerase	ATS92512.1	protein
St11Ph5_MG208881	St11Ph5_00074	59780	60286	507	-	putative Rz/RzI spanin protein	ATS92535.1	protein
St11Ph5_MG208881	St11Ph5_00076	60892	61152	261	-	putative holin protein	ATS92537.1	protein
VB_EcoS-Golestan_MG099933	E1_7	5234	5620	387	+	putative spanin	ATS93231.1	protein
VB_EcoS-Golestan_MG099933	E1_8	5586	5729	144	+	putative spanin	ATS93232.1	protein
VB_EcoS-Golestan_MG099933	E1_39	27955	30207	2253	-	putative DNA polymerase	ATS93263.1	protein
VB_EcoS-Golestan_MG099933	E1_66	40866	41174	309	+	putative class II holin-like protein	ATS93290.1	protein
VB_EcoS-Golestan_MG099933	E1_67	41167	41439	273	+	putative class I holin-like protein	ATS93291.1	protein
VB_EcoS-Golestan_MG099933	E1_68	41444	41902	459	+	lysozyme/endolysin	ATS93292.1	protein
EG1_MG488277	20_CDS	10722	11177	456	+	endonuclease	ATW57728.1	protein
EG1_MG488277	26_CDS	14452	16608	2157	+	DNA polymerase	ATW57734.1	protein
EG1_MG488277	31_CDS	17521	18423	903	+	exonuclease	ATW57739.1	protein
St11Ph5_MG208881	St11Ph5_00073	59833	60120	288	-	putative Rz/RzI spanin protein	ATZ35296.1	protein
myPSH1131_MG983840	PSH1131_005	8701	8919	219	+	putative holin	AVQ10010.1	protein
myPSH1131_MG983840	PSH1131_020	27490	28182	693	-	5'-3' exonuclease	AVQ10025.1	protein
myPSH1131_MG983840	PSH1131_029	33286	33519	234	-	5'-3' exonuclease	AVQ10034.1	protein
myPSH1131_MG983840	PSH1131_038	37901	39709	1809	-	DNA polymerase	AVQ10043.1	protein
myPSH1131_MG983840	PSH1131_052	43591	43956	366	-	DNA polymerase	AVQ10057.1	protein
myPSH1131_MG983840	PSH1131_064	48916	49251	336	-	DNA polymerase	AVQ10069.1	protein
myPSH1131_MG983840	PSH1131_067	53266	53379	114	-	type I DNA polymerase	AVQ10072.1	protein
UGKSEcP2_OV876900	UGKSECP2_00011	3761	4237	477	-	DenB-like DNA endonuclease IV	CAH1615298.1	protein
UGKSEcP2_OV876900	UGKSECP2_00026	13418	14101	684	-	exonuclease	CAH1615313.1	protein
UGKSEcP2_OV876900	UGKSECP2_00060	30738	33434	2697	-	DNA polymerase	CAH1615347.1	protein
UGKSEcP2_OV876900	UGKSECP2_00087	47676	48149	474	-	endonuclease VII	CAH1615374.1	protein
UGKSEcP2_OV876900	UGKSECP2_00125	64985	65446	462	-	endoribonuclease	CAH1615412.1	protein
UGKSEcP2_OV876900	UGKSECP2_00132	67514	68065	552	-	endonuclease V N-glycosylase UV repair enzyme	CAH1615419.1	protein
UGKSEcP2_OV876900	UGKSECP2_00134	68497	68991	495	-	endolysin	CAH1615421.1	protein
UGKSEcP2_OV876900	UGKSECP2_00200	115556	117067	1512	+	DNA helicase	CAH1615507.1	protein
UGKSEcP2_OV876900	UGKSECP2_00201	117093	117323	231	+	DNA helicase	CAH1615509.1	protein
UGKSEcP2_OV876900	UGKSECP2_00240	140285	140584	300	-	Rz-like spanin	CAH1615586.1	protein
UGKSEcP2_OV876900	UGKSECP2_00241	140581	140934	354	-	Rz-like spanin	CAH1615588.1	protein
UGKSEcP2_OV876900	UGKSECP2_00244	142670	143080	411	-	endonuclease	CAH1615594.1	protein
UGKSEcP2_OV876900	UGKSECP2_00267	163376	164032	657	+	holin	CAH1615640.1	protein
UGKSEcP1_OV877085	UGKSECP1_00010	6852	7082	231	-	DNA helicase	CAH1615835.1	protein
UGKSEcP1_OV877085	UGKSECP1_00011	7108	8619	1512	-	DNA helicase	CAH1615836.1	protein
UGKSEcP1_OV877085	UGKSECP1_00077	55175	55669	495	+	endolysin	CAH1615891.1	protein
UGKSEcP1_OV877085	UGKSECP1_00079	56101	56652	552	+	endonuclease V N-glycosylase UV repair enzyme	CAH1615893.1	protein
UGKSEcP1_OV877085	UGKSECP1_00086	58718	59179	462	+	endoribonuclease	CAH1615900.1	protein
UGKSEcP1_OV877085	UGKSECP1_00123	76016	76489	474	+	endonuclease VII	CAH1615937.1	protein
UGKSEcP1_OV877085	UGKSECP1_00150	90740	93436	2697	+	DNA polymerase	CAH1615964.1	protein
UGKSEcP1_OV877085	UGKSECP1_00184	109856	110539	684	+	exonuclease	CAH1615998.1	protein
UGKSEcP1_OV877085	UGKSECP1_00199	119724	120200	477	+	DenB-like DNA endonuclease IV	CAH1616013.1	protein
UGKSEcP1_OV877085	UGKSECP1_00220	127242	127898	657	-	holin	CAH1616034.1	protein
UGKSEcP1_OV877085	UGKSECP1_00243	148194	148604	411	+	endonuclease	CAH1616057.1	protein
UGKSEcP1_OV877085	UGKSECP1_00246	150340	150693	354	+	Rz-like spanin	CAH1616060.1	protein
UGKSEcP1_OV877085	UGKSECP1_00247	150690	150989	300	+	Rz-like spanin	CAH1616061.1	protein
vB_EcoM_KMB26_OW536688	12_CDS	7630	8313	684	-	T4-like phage DexA exonuclease A	CAH2618977.1	protein
vB_EcoM_KMB26_OW536688	13_CDS	8313	8555	243	-	Phage exonuclease	CAH2618978.1	protein
vB_EcoM_KMB26_OW536688	14_CDS	8548	8793	246	-	Phage exonuclease	CAH2618979.1	protein
vB_EcoM_KMB26_OW536688	16_CDS	9047	10366	1320	-	DNA helicase (EC, phage-associated	CAH2618981.1	protein
vB_EcoM_KMB26_OW536688	33_CDS	17510	18538	1029	-	DNA primase (EC / DNA helicase (EC, phage-associated	CAH2618998.1	protein
vB_EcoM_KMB26_OW536688	49_CDS	27054	29750	2697	-	DNA polymerase (EC, phage-associated	CAH2619014.1	protein
vB_EcoM_KMB26_OW536688	57_CDS	33383	35023	1641	-	Phage recombination-related endonuclease Gp46	CAH2619022.1	protein
vB_EcoM_KMB26_OW536688	60_CDS	35509	36528	1020	-	Phage recombination-related endonuclease Gp47	CAH2619025.1	protein
vB_EcoM_KMB26_OW536688	79_CDS	44230	44703	474	-	Phage endonuclease #T4-like phage gp49, endonuclease VII #T4 GC1525	CAH2619044.1	protein
vB_EcoM_KMB26_OW536688	124_CDS	63720	64136	417	-	Phage endonuclease	CAH2619089.1	protein
vB_EcoM_KMB26_OW536688	177_CDS	109198	110709	1512	+	DNA helicase, phage-associated	CAH2619142.1	protein
vB_EcoM_KMB26_OW536688	178_CDS	110735	110965	231	+	DNA helicase, phage-associated	CAH2619143.1	protein
vB_EcoM_KMB26_OW536688	219_CDS	134482	134835	354	-	Phage spanin Rz	CAH2619184.1	protein
vB_EcoM_KMB26_OW536688	222_CDS	136571	136981	411	-	Phage endonuclease	CAH2619187.1	protein
vB_EcoM_KMB26_OW536688	227_CDS	141520	142257	738	-	Phage-associated homing endonuclease	CAH2619192.1	protein
vB_EcoM_KMB26_OW536688	237_CDS	146483	146944	462	-	Phage-associated homing endonuclease	CAH2619202.1	protein
vB_EcoM_KMB26_OW536688	247_CDS	158944	159600	657	+	Phage holin	CAH2619212.1	protein
vB_EcoM_KMB26_OW536688	257_CDS	162943	164271	1329	-	DNA topoisomerase, phage-associated	CAH2619222.1	protein
vB_EcoM_KMB26_OW536688	265_CDS	166406	166963	558	-	Phage endonuclease	CAH2619230.1	protein
Ec_Makalu_002_MN709127	Makalu002_008	5315	5983	669	-	exonuclease	QGS83613.1	protein
Ec_Makalu_002_MN709127	Makalu002_012	6584	7975	1392	-	DNA helicase	QGS83617.1	protein
Ec_Makalu_002_MN709127	Makalu002_027	15982	18660	2679	-	DNA polymerase	QGS83632.1	protein
Ec_Makalu_002_MN709127	Makalu002_061	38815	39288	474	-	recombination endonuclease VII	QGS83666.1	protein

Ec_Makalu_002_MN709127	Makalu002_094	51263	51685	423	-	site-specific RNA endonuclease	QGS83699.1	protein
Ec_Makalu_002_MN709127	Makalu002_102	54429	54818	390	-	putative endolysin	QGS83707.1	protein
Ec_Makalu_002_MN709127	Makalu002_131	63706	64728	1023	+	putative homing endonuclease, HNH family	QGS83736.1	protein
Ec_Makalu_002_MN709127	Makalu002_149	83563	84378	816	+	homing endonuclease	QGS83754.1	protein
Ec_Makalu_002_MN709127	Makalu002_172	103865	105367	1503	+	ATP-dependent DNA helicase	QGS83777.1	protein
Ec_Makalu_002_MN709127	Makalu002_219	129371	129823	453	-	putative o-spanin	QGS83824.1	protein
Ec_Makalu_002_MN709127	Makalu002_220	129820	130143	324	-	putative i-spanin	QGS83825.1	protein
Ec_Makalu_002_MN709127	Makalu002_222	131293	131754	462	-	endonuclease II	QGS83827.1	protein
Ec_Makalu_002_MN709127	Makalu002_240	141938	142534	597	-	putative HNH endonuclease	QGS83845.1	protein
Ec_Makalu_002_MN709127	Makalu002_252	155311	155967	657	+	holin lysis mediator	QGS83857.1	protein
Ec_Makalu_003_MN882349	Makalu003_008	5315	5983	669	-	exonuclease	QHJ72956.1	protein
Ec_Makalu_003_MN882349	Makalu003_012	6584	7975	1392	-	ATP-dependent DNA helicase	QHJ72960.1	protein
Ec_Makalu_003_MN882349	Makalu003_027	15980	18658	2679	-	DNA polymerase	QHJ72975.1	protein
Ec_Makalu_003_MN882349	Makalu003_061	38812	39285	474	-	recombination endonuclease VII	QHJ73009.1	protein
Ec_Makalu_003_MN882349	Makalu003_095	51473	51895	423	-	site-specific RNA endonuclease	QHJ73043.1	protein
Ec_Makalu_003_MN882349	Makalu003_103	54639	55028	390	-	putative endolysin	QHJ73051.1	protein
Ec_Makalu_003_MN882349	Makalu003_136	67270	67809	540	+	putative HNH endonuclease	QHJ73084.1	protein
Ec_Makalu_003_MN882349	Makalu003_172	102546	104048	1503	+	ATP-dependent DNA helicase	QHJ73120.1	protein
Ec_Makalu_003_MN882349	Makalu003_222	129986	130447	462	-	endonuclease II	QHJ73170.1	protein
Ec_Makalu_003_MN882349	Makalu003_252	153602	154258	657	+	holin lysis mediator	QHJ73200.1	protein
Ec_Makalu_001_MN894885	EcMakalu001_008	5327	5995	669	-	exonuclease	QHJ73229.1	protein
Ec_Makalu_001_MN894885	EcMakalu001_012	6596	7963	1368	-	DNA helicase	QHJ73233.1	protein
Ec_Makalu_001_MN894885	EcMakalu001_027	15992	18670	2679	-	DNA polymerase	QHJ73248.1	protein
Ec_Makalu_001_MN894885	EcMakalu001_061	38827	39315	489	-	recombination endonuclease VII	QHJ73282.1	protein
Ec_Makalu_001_MN894885	EcMakalu001_093	51273	51695	423	-	site-specific RNA endonuclease	QHJ73314.1	protein
Ec_Makalu_001_MN894885	EcMakalu001_101	54440	54829	390	-	putative endolysin	QHJ73322.1	protein
Ec_Makalu_001_MN894885	EcMakalu001_147	82442	83257	816	+	putative Seg-like homing endonuclease, GIY-YIG family	QHJ73368.1	protein
Ec_Makalu_001_MN894885	EcMakalu001_170	101765	103267	1503	+	RNA-DNA and DNA-DNA helicase ATPase	QHJ73391.1	protein
Ec_Makalu_001_MN894885	EcMakalu001_171	103311	103547	237	+	putative ATP-dependent DNA helicase	QHJ73392.1	protein
Ec_Makalu_001_MN894885	EcMakalu001_218	127678	128130	453	-	O-spanin	QHJ73439.1	protein
Ec_Makalu_001_MN894885	EcMakalu001_219	128127	128450	324	-	I-spanin	QHJ73440.1	protein
Ec_Makalu_001_MN894885	EcMakalu001_221	129600	130061	462	-	endonuclease II	QHJ73442.1	protein
Ec_Makalu_001_MN894885	EcMakalu001_251	154405	155061	657	+	holin lysis mediator	QHJ73472.1	protein
RDN8_1_MN414250	12_CDS	8283	8525	243	+	HNH homing endonuclease	QID21009.1	protein
RDN8_1_MN414250	17_CDS	10273	10692	420	+	endonuclease I	QID21014.1	protein
RDN8_1_MN414250	19_CDS	11131	11589	459	+	endolysin	QID21016.1	protein
RDN8_1_MN414250	24_CDS	14086	16257	2172	+	DNA-directed DNA polymerase	QID21021.1	protein
RDN8_1_MN414250	29_CDS	17291	18163	873	+	exonuclease	QID21026.1	protein
RDN8_1_MN414250	47_CDS	36601	36798	198	+	holin class II	QID21044.1	protein
RDN8_1_MN414250	49_CDS	37591	38031	441	+	HNH homing endonuclease	QID21046.1	protein
590B_MW722821	6_CDS	4272	4658	387	+	putative spanin protein	QVR48586.1	protein
590B_MW722821	62_CDS	39352	39660	309	+	putative holin-like class II	QVR48620.1	protein
590B_MW722821	63_CDS	39653	39925	273	+	class I holin-like protein	QVR48621.1	protein
590B_MW722821	64_CDS	39903	40391	489	+	endolysin	QVR48622.1	protein
vB_EcoS-phiEc4_MZ576218	AAKBPDAN_00007	5250	5636	387	+	putative spanin	QYU43856.1	protein
vB_EcoS-phiEc4_MZ576218	AAKBPDAN_00008	5599	5745	147	+	putative spanin	QYU43857.1	protein
vB_EcoS-phiEc4_MZ576218	AAKBPDAN_00039	27986	30238	2253	-	DNA polymerase I	QYU43888.1	protein
vB_EcoS-phiEc4_MZ576218	AAKBPDAN_00065	41164	41436	273	+	putative holin-like class I protein	QYU43914.1	protein
vB_EcoS-phiEc4_MZ576218	AAKBPDAN_00066	41414	41899	486	+	endolysin	QYU43915.1	protein
vB_EcoS-phiEc3_MZ576217	MONNJMKB_00006	4928	5314	387	+	putative spanin	QZI78273.1	protein
vB_EcoS-phiEc3_MZ576217	MONNJMKB_00007	5280	5423	144	+	putative spanin	QZI78274.1	protein
vB_EcoS-phiEc3_MZ576217	MONNJMKB_00039	27798	30050	2253	-	DNA polymerase I	QZI78306.1	protein
vB_EcoS-phiEc3_MZ576217	MONNJMKB_00066	40982	41290	309	+	putative holin-like class II protein	QZI78333.1	protein
vB_EcoS-phiEc3_MZ576217	MONNJMKB_00067	41283	41555	273	+	putative holin-like class I protein	QZI78334.1	protein
vB_EcoS-phiEc3_MZ576217	MONNJMKB_00068	41533	42021	489	+	endolysin	QZI78335.1	protein
vB_EcoS-22664B1_MZ234008	22664B1_052	32136	32459	324	-	HNH endonuclease	QZI78411.1	protein
vB_EcoS-22664B1_MZ234008	22664B1_054	33148	34572	1425	-	DNA helicase	QZI78413.1	protein
vB_EcoS-22664B1_MZ234008	22664B1_059	35523	37814	2292	-	DNA polymerase I	QZI78418.1	protein
vB_EcoP-22664BS1_MZ234009	22664BS1_003	5693	5887	195	+	class II holin	QZI78433.1	protein
vB_EcoP-22664BS1_MZ234009	22664BS1_033	18780	19232	453	+	endonuclease I	QZI78463.1	protein
vB_EcoP-22664BS1_MZ234009	22664BS1_035	19873	20226	354	+	endolysin	QZI78465.1	protein
vB_EcoP-22664BS1_MZ234009	22664BS1_039	22584	24755	2172	+	DNA-directed DNA polymerase	QZI78469.1	protein
vB_EcoP-22664BS1_MZ234009	22664BS1_043	25549	26421	873	+	exonuclease	QZI78473.1	protein
vB_EcoS-22664BS2_MZ234010	22664BS2_051	29577	31109	1533	-	DNA helicase	QZI78540.1	protein
vB_EcoS-22664BS2_MZ234010	22664BS2_058	32533	34713	2181	-	DNA polymerase I	QZI78547.1	protein
vB_EcoP-22664UKE3-2_MZ234011	22664UKE3-2_003	6083	6277	195	+	class II holin	QZI78578.1	protein
vB_EcoP-22664UKE3-2_MZ234011	22664UKE3-2_006	7082	7519	438	+	HNH homing endonuclease	QZI78581.1	protein
vB_EcoP-22664UKE3-2_MZ234011	22664UKE3-2_034	20125	20577	453	+	endonuclease I	QZI78609.1	protein
vB_EcoP-22664UKE3-2_MZ234011	22664UKE3-2_036	21058	21516	459	+	endolysin	QZI78611.1	protein
vB_EcoP-22664UKE3-2_MZ234011	22664UKE3-2_040	23841	26012	2172	+	DNA-directed DNA polymerase	QZI78615.1	protein
vB_EcoP-22664UKE3-2_MZ234011	22664UKE3-2_043	26493	27365	873	+	exonuclease	QZI78618.1	protein
vB_EcoP-101101UKE1_MZ234012	101101UKE1_010	7572	10118	2547	+	DNA-directed DNA polymerase	QZI78645.1	protein
vB_EcoP-101101UKE1_MZ234012	101101UKE1_019	13293	14321	1029	+	exonuclease	QZI78654.1	protein
vB_EcoP-101101UKE1_MZ234012	101101UKE1_038	33309	33503	195	+	class II holin	QZI78673.1	protein
vB_EcoM-101112UKE3-1_MZ234013	101112UKE3-1_017	8536	9009	474	+	endonuclease	QZI78714.1	protein
vB_EcoM-101112UKE3-1_MZ234013	101112UKE3-1_038	17010	18029	1020	+	recombination-related endonuclease	QZI78735.1	protein
vB_EcoM-101112UKE3-1_MZ234013	101112UKE3-1_042	19343	21025	1683	+	recombination-related endonuclease	QZI78739.1	protein
vB_EcoM-101112UKE3-1_MZ234013	101112UKE3-1_051	24658	27354	2697	+	DNA polymerase	QZI78747.1	protein
vB_EcoM-101112UKE3-1_MZ234013	101112UKE3-1_067	34855	35883	1029	+	DNA primase / DNA helicase	QZI78763.1	protein
vB_EcoM-101112UKE3-1_MZ234013	101112UKE3-1_086	43029	44348	1320	+	DNA helicase	QZI78782.1	protein
vB_EcoM-101112UKE3-1_MZ234013	101112UKE3-1_088	44602	44847	246	+	exonuclease	QZI78784.1	protein
vB_EcoM-101112UKE3-1_MZ234013	101112UKE3-1_089	44840	45082	243	+	exonuclease	QZI78785.1	protein
vB_EcoM-101112UKE3-1_MZ234013	101112UKE3-1_090	45082	45765	684	+	exonuclease	QZI78786.1	protein
vB_EcoM-101112UKE3-1_MZ234013	101112UKE3-1_104	54379	54573	195	+	endonuclease	QZI78800.1	protein
vB_EcoM-101112UKE3-1_MZ234013	101112UKE3-1_105	54660	55217	558	+	endonuclease	QZI78801.1	protein
vB_EcoM-101112UKE3-1_MZ234013	101112UKE3-1_116	57407	58735	1329	+	DNA topoisomerase	QZI78812.1	protein
vB_EcoM-101112UKE3-1_MZ234013	101112UKE3-1_128	62078	62734	657	-	holin	QZI78824.1	protein
vB_EcoM-101112UKE3-1_MZ234013	101112UKE3-1_154	81777	82508	732	+	homing endonuclease	QZI78850.1	protein
vB_EcoM-101112UKE3-1_MZ234013	101112UKE3-1_156	83714	84124	411	+	endonuclease	QZI78852.1	protein
vB_EcoM-101112UKE3-1_MZ234013	101112UKE3-1_202	109733	109963	231	-	DNA helicase	QZI78898.1	protein
vB_EcoM-101112UKE3-1_MZ234013	101112UKE3-1_203	109989	111500	1512	-	DNA helicase	QZI78899.1	protein
vB_EcoM-101112UKE3-1_MZ234013	101112UKE3-1_223	128939	129751	813	-	homing endonuclease	QZI78919.1	protein
vB_EcoM-101112UKE3-1_MZ234013	101112UKE3-1_249	152793	153458	666	+	homing endonuclease	QZI78945.1	protein
vB_EcoM-101112UKE3-1_MZ234013	101112UKE3-1_263	158657	159073	417	+	endonuclease	QZI78959.1	protein
vB_EcoM-101112UKE3-1_MZ234013	101112UKE3-1_271	161274	161735	462	+	endonuclease	QZI78967.1	protein
vB_EcoS-101114B2_MZ234014	101114B2_051	32051	32374	324	-	HNH endonuclease	QZI79041.1	protein

vB_EcoS-101114B2_MZ234014	101114B2_053	33063	34487	1425	-	DNA helicase	QZI79043.1	protein
vB_EcoS-101114B2_MZ234014	101114B2_058	35438	37729	2292	-	DNA polymerase I	QZI79048.1	protein
vB_EcoS-101114BS4_MZ234016	101114BS4_047	30650	30973	324	-	HNH endonuclease	QZI79107.1	protein
vB_EcoS-101114BS4_MZ234016	101114BS4_049	31662	33086	1425	-	DNA helicase	QZI79109.1	protein
vB_EcoS-101114BS4_MZ234016	101114BS4_054	34037	36328	2292	-	DNA polymerase I	QZI79114.1	protein
vB_EcoP-101114UKE3_MZ234017	101114UKE3_032	31360	32181	822	-	exonuclease	QZI79163.1	protein
vB_EcoP-101114UKE3_MZ234017	101114UKE3_053	38440	40284	1845	-	DNA-directed DNA polymerase	QZI79184.1	protein
vB_EcoP-101114UKE3_MZ234017	101114UKE3_076	48822	49370	549	-	DNA polymerase	QZI79207.1	protein
vB_EcoP-101114UKE3_MZ234017	101114UKE3_121	68264	68569	306	-	DNA helicase	QZI79252.1	protein
vB_EcoM-101117BS1_MZ234018	101117BS1_018	8583	9056	474	+	endonuclease	QZI79294.1	protein
vB_EcoM-101117BS1_MZ234018	101117BS1_040	17059	18078	1020	+	recombination-related endonuclease	QZI79316.1	protein
vB_EcoM-101117BS1_MZ234018	101117BS1_043	18564	20204	1641	+	recombination-related endonuclease	QZI79319.1	protein
vB_EcoM-101117BS1_MZ234018	101117BS1_052	23863	26559	2697	+	DNA polymerase	QZI79327.1	protein
vB_EcoM-101117BS1_MZ234018	101117BS1_069	34407	35435	1029	+	DNA primase / DNA helicase	QZI79344.1	protein
vB_EcoM-101117BS1_MZ234018	101117BS1_086	41507	42826	1320	+	DNA helicase	QZI79361.1	protein
vB_EcoM-101117BS1_MZ234018	101117BS1_088	43080	43325	246	+	exonuclease	QZI79363.1	protein
vB_EcoM-101117BS1_MZ234018	101117BS1_089	43318	43560	243	+	exonuclease	QZI79364.1	protein
vB_EcoM-101117BS1_MZ234018	101117BS1_090	43560	44243	684	+	exonuclease	QZI79365.1	protein
vB_EcoM-101117BS1_MZ234018	101117BS1_104	53047	53241	195	+	endonuclease	QZI79379.1	protein
vB_EcoM-101117BS1_MZ234018	101117BS1_105	53324	53881	558	+	endonuclease	QZI79380.1	protein
vB_EcoM-101117BS1_MZ234018	101117BS1_118	56354	57682	1329	+	DNA topoisomerase	QZI79393.1	protein
vB_EcoM-101117BS1_MZ234018	101117BS1_128	60684	61340	657	-	holin	QZI79403.1	protein
vB_EcoM-101117BS1_MZ234018	101117BS1_151	81515	81925	411	+	endonuclease	QZI79426.1	protein
vB_EcoM-101117BS1_MZ234018	101117BS1_197	107549	107779	231	-	DNA helicase	QZI79472.1	protein
vB_EcoM-101117BS1_MZ234018	101117BS1_198	107805	109316	1512	-	DNA helicase	QZI79473.1	protein
vB_EcoM-101117BS1_MZ234018	101117BS1_258	155180	155596	417	+	endonuclease	QZI79533.1	protein
vB_EcoM-101117BS1_MZ234018	101117BS1_267	158602	159063	462	+	endonuclease	QZI79542.1	protein
vB_EcoP-101117UKE2_MZ234019	101117UKE2_011	7620	10193	2574	+	DNA-directed DNA polymerase	QZI79577.1	protein
vB_EcoP-101117UKE2_MZ234019	101117UKE2_018	12780	13808	1029	+	exonuclease	QZI79584.1	protein
vB_EcoP-101117UKE2_MZ234019	101117UKE2_037	35435	35629	195	+	class II holin	QZI79603.1	protein
vB_EcoP-101118B1_MZ234020	101118B1_011	7565	10138	2574	+	DNA-directed DNA polymerase	QZI79636.1	protein
vB_EcoP-101118B1_MZ234020	101118B1_018	12725	13753	1029	+	exonuclease	QZI79643.1	protein
vB_EcoP-101118B1_MZ234020	101118B1_038	35435	35629	195	+	class II holin	QZI79663.1	protein
vB_EcoP-101118UKE1_MZ234021	101118UKE1_003	5693	5887	195	+	class II holin	QZI79688.1	protein
vB_EcoP-101118UKE1_MZ234021	101118UKE1_034	19194	19646	453	+	endonuclease I	QZI79719.1	protein
vB_EcoP-101118UKE1_MZ234021	101118UKE1_036	20127	20585	459	+	endolysin	QZI79721.1	protein
vB_EcoP-101118UKE1_MZ234021	101118UKE1_041	23258	25426	2169	+	DNA-directed DNA polymerase	QZI79726.1	protein
vB_EcoP-101118UKE1_MZ234021	101118UKE1_045	26220	27092	873	+	exonuclease	QZI79730.1	protein
vB_EcoP-101120B1-2_MZ234022	101120B1-2_004	6234	6428	195	+	class II holin	QZI79751.1	protein
vB_EcoP-101120B1-2_MZ234022	101120B1-2_007	7233	7670	438	+	HNH homing endonuclease	QZI79754.1	protein
vB_EcoP-101120B1-2_MZ234022	101120B1-2_035	19907	20362	456	+	endonuclease I	QZI79782.1	protein
vB_EcoP-101120B1-2_MZ234022	101120B1-2_037	20840	21298	459	+	endolysin	QZI79784.1	protein
vB_EcoP-101120B1-2_MZ234022	101120B1-2_041	23623	25794	2172	+	DNA-directed DNA polymerase	QZI79788.1	protein
vB_EcoP-101120B1-2_MZ234022	101120B1-2_044	26275	27147	873	+	exonuclease	QZI79791.1	protein
vB_EcoP-101120B2_MZ234023	101220B2_004	6234	6428	195	+	class II holin	QZI79811.1	protein
vB_EcoP-101120B2_MZ234023	101220B2_007	7233	7670	438	+	HNH homing endonuclease	QZI79814.1	protein
vB_EcoP-101120B2_MZ234023	101220B2_035	19962	20417	456	+	endonuclease I	QZI79842.1	protein
vB_EcoP-101120B2_MZ234023	101220B2_037	20895	21353	459	+	endolysin	QZI79844.1	protein
vB_EcoP-101120B2_MZ234023	101220B2_041	23678	25849	2172	+	DNA-directed DNA polymerase	QZI79848.1	protein
vB_EcoP-101120B2_MZ234023	101220B2_044	26330	27202	873	+	exonuclease	QZI79851.1	protein
vB_EcoP-101136BS1_MZ234024	101136BS1_003	5693	5887	195	+	class II holin	QZI79870.1	protein
vB_EcoP-101136BS1_MZ234024	101136BS1_034	18795	19247	453	+	endonuclease I	QZI79901.1	protein
vB_EcoP-101136BS1_MZ234024	101136BS1_036	19728	20186	459	+	endolysin	QZI79903.1	protein
vB_EcoP-101136BS1_MZ234024	101136BS1_040	22436	24607	2172	+	DNA-directed DNA polymerase	QZI79907.1	protein
vB_EcoP-101136BS1_MZ234024	101136BS1_044	25401	26273	873	+	exonuclease	QZI79911.1	protein
vB_EcoP-101136BS1_MZ234024	101136BS1_058	35511	35912	402	+	endonuclease VII	QZI79925.1	protein
vB_EcoM-172859UKE1_MZ234025	172859UKE1_014	7132	7605	474	+	recombination endonuclease VII	QZI79941.1	protein
vB_EcoM-172859UKE1_MZ234025	172859UKE1_035	14871	15890	1020	+	recombination-related endonuclease	QZI79962.1	protein
vB_EcoM-172859UKE1_MZ234025	172859UKE1_038	17020	18666	1647	+	recombination-related endonuclease	QZI79965.1	protein
vB_EcoM-172859UKE1_MZ234025	172859UKE1_070	37989	39011	1023	+	DNA primase / DNA helicase	QZI79996.1	protein
vB_EcoM-172859UKE1_MZ234025	172859UKE1_084	44127	45440	1314	+	DNA helicase	QZI80010.1	protein
vB_EcoM-172859UKE1_MZ234025	172859UKE1_085	45450	46127	678	+	exonuclease	QZI80011.1	protein
vB_EcoM-172859UKE1_MZ234025	172859UKE1_100	56287	56844	558	+	endonuclease	QZI80026.1	protein
vB_EcoM-172859UKE1_MZ234025	172859UKE1_108	58494	59819	1326	+	DNA topoisomerase	QZI80034.1	protein
vB_EcoM-172859UKE1_MZ234025	172859UKE1_118	62772	63431	660	-	holin	QZI80044.1	protein
vB_EcoM-172859UKE1_MZ234025	172859UKE1_137	82516	82926	411	+	endonuclease	QZI80063.1	protein
vB_EcoM-172859UKE1_MZ234025	172859UKE1_181	107171	107404	234	-	DNA helicase	QZI80107.1	protein
vB_EcoM-172859UKE1_MZ234025	172859UKE1_182	107430	108944	1515	-	DNA helicase	QZI80108.1	protein
vB_EcoM-172859UKE1_MZ234025	172859UKE1_252	158372	158785	414	+	endonuclease	QZI80178.1	protein
vB_EcoM-172859UKE1_MZ234025	172859UKE1_263	161843	162301	459	+	endonuclease	QZI80189.1	protein
vB_EcoM-CHD2BS1_MZ234027	CHD2BS1_017	8537	9010	474	+	endonuclease	QZI80223.1	protein
vB_EcoM-CHD2BS1_MZ234027	CHD2BS1_038	17012	18031	1020	+	recombination-related endonuclease	QZI80244.1	protein
vB_EcoM-CHD2BS1_MZ234027	CHD2BS1_041	18572	20212	1641	+	recombination-related endonuclease	QZI80247.1	protein
vB_EcoM-CHD2BS1_MZ234027	CHD2BS1_050	23845	26541	2697	+	DNA polymerase	QZI80255.1	protein
vB_EcoM-CHD2BS1_MZ234027	CHD2BS1_068	35010	36038	1029	+	DNA primase / DNA helicase	QZI80273.1	protein
vB_EcoM-CHD2BS1_MZ234027	CHD2BS1_088	43525	44844	1320	+	DNA helicase	QZI80293.1	protein
vB_EcoM-CHD2BS1_MZ234027	CHD2BS1_090	45098	45343	246	+	exonuclease	QZI80295.1	protein
vB_EcoM-CHD2BS1_MZ234027	CHD2BS1_091	45336	45578	243	+	exonuclease	QZI80296.1	protein
vB_EcoM-CHD2BS1_MZ234027	CHD2BS1_092	45578	46261	684	+	exonuclease	QZI80297.1	protein
vB_EcoM-CHD2BS1_MZ234027	CHD2BS1_106	54875	55069	195	+	endonuclease	QZI80311.1	protein
vB_EcoM-CHD2BS1_MZ234027	CHD2BS1_108	55370	55927	558	+	endonuclease	QZI80313.1	protein
vB_EcoM-CHD2BS1_MZ234027	CHD2BS1_118	58062	59390	1329	+	DNA topoisomerase	QZI80323.1	protein
vB_EcoM-CHD2BS1_MZ234027	CHD2BS1_130	62746	63402	657	-	holin	QZI80335.1	protein
vB_EcoM-CHD2BS1_MZ234027	CHD2BS1_140	75238	75870	633	+	homing endonuclease	QZI80345.1	protein
vB_EcoM-CHD2BS1_MZ234027	CHD2BS1_151	79735	80472	738	+	homing endonuclease	QZI80356.1	protein
vB_EcoM-CHD2BS1_MZ234027	CHD2BS1_157	85012	85422	411	+	endonuclease	QZI80362.1	protein
vB_EcoM-CHD2BS1_MZ234027	CHD2BS1_203	111199	111429	231	-	DNA helicase	QZI80408.1	protein
vB_EcoM-CHD2BS1_MZ234027	CHD2BS1_204	111455	112966	1512	-	DNA helicase	QZI80409.1	protein
vB_EcoM-CHD2BS1_MZ234027	CHD2BS1_260	157682	158098	417	+	endonuclease	QZI80465.1	protein
vB_EcoM-CHD2BS1_MZ234027	CHD2BS1_268	160296	160757	462	+	endonuclease	QZI80473.1	protein
vB_EcoP-CHD5UKE1_MZ234028	CHD5UKE1_0032	31423	32241	819	-	exonuclease	QZI80528.1	protein
vB_EcoP-CHD5UKE1_MZ234028	CHD5UKE1_0053	39164	40972	1809	-	DNA-directed DNA polymerase	QZI80549.1	protein
vB_EcoP-CHD5UKE1_MZ234028	CHD5UKE1_0076	49572	50129	558	-	DNA polymerase	QZI80572.1	protein
vB_EcoP-CHD5UKE1_MZ234028	CHD5UKE1_126	69617	69913	297	-	DNA helicase	QZI80622.1	protein
vB_EcoP-CHD5UKE1_MZ234028	CHD5UKE1_149	75658	75921	264	-	DNA polymerase B region	QZI80645.1	protein
vB_EcoM-CHD16UKE1_MZ234030	CHD16UKE1_017	8537	9010	474	+	endonuclease	QZI80665.1	protein

vB_EcoM-CHD16UKE1_MZ234030	CHD16UKE1_038	17012	18031	1020	+	recombination-related endonuclease	QZI80686.1	protein
vB_EcoM-CHD16UKE1_MZ234030	CHD16UKE1_041	18517	20157	1641	+	recombination-related endonuclease	QZI80689.1	protein
vB_EcoM-CHD16UKE1_MZ234030	CHD16UKE1_050	23790	26486	2697	+	DNA polymerase	QZI80697.1	protein
vB_EcoM-CHD16UKE1_MZ234030	CHD16UKE1_068	34955	35983	1029	+	DNA primase / DNA helicase	QZI80715.1	protein
vB_EcoM-CHD16UKE1_MZ234030	CHD16UKE1_088	43471	44790	1320	+	DNA helicase	QZI80735.1	protein
vB_EcoM-CHD16UKE1_MZ234030	CHD16UKE1_090	45044	45289	246	+	exonuclease	QZI80737.1	protein
vB_EcoM-CHD16UKE1_MZ234030	CHD16UKE1_091	45282	45524	243	+	exonuclease	QZI80738.1	protein
vB_EcoM-CHD16UKE1_MZ234030	CHD16UKE1_092	45524	46207	684	+	exonuclease	QZI80739.1	protein
vB_EcoM-CHD16UKE1_MZ234030	CHD16UKE1_106	54821	55015	195	+	endonuclease	QZI80753.1	protein
vB_EcoM-CHD16UKE1_MZ234030	CHD16UKE1_108	55316	55873	558	+	endonuclease	QZI80755.1	protein
vB_EcoM-CHD16UKE1_MZ234030	CHD16UKE1_118	58008	59336	1329	+	DNA topoisomerase	QZI80765.1	protein
vB_EcoM-CHD16UKE1_MZ234030	CHD16UKE1_130	62679	63335	657	-	holin	QZI80777.1	protein
vB_EcoM-CHD16UKE1_MZ234030	CHD16UKE1_141	75215	75847	633	+	homing endonuclease	QZI80788.1	protein
vB_EcoM-CHD16UKE1_MZ234030	CHD16UKE1_152	79713	80450	738	+	homing endonuclease	QZI80799.1	protein
vB_EcoM-CHD16UKE1_MZ234030	CHD16UKE1_158	84992	85402	411	+	endonuclease	QZI80805.1	protein
vB_EcoM-CHD16UKE1_MZ234030	CHD16UKE1_204	111168	111398	231	-	DNA helicase	QZI80851.1	protein
vB_EcoM-CHD16UKE1_MZ234030	CHD16UKE1_205	111424	112935	1512	-	DNA helicase	QZI80852.1	protein
vB_EcoM-CHD16UKE1_MZ234030	CHD16UKE1_261	157648	158064	417	+	endonuclease	QZI80908.1	protein
vB_EcoM-CHD16UKE1_MZ234030	CHD16UKE1_269	160262	160723	462	+	endonuclease	QZI80916.1	protein
vB_EcoM-CHD94UKE2_MZ234031	CHD94UKE2_017	8540	9013	474	+	endonuclease	QZI80956.1	protein
vB_EcoM-CHD94UKE2_MZ234031	CHD94UKE2_019	10820	11650	831	+	homing endonuclease	QZI80958.1	protein
vB_EcoM-CHD94UKE2_MZ234031	CHD94UKE2_041	17899	18918	1020	+	recombination-related endonuclease	QZI80980.1	protein
vB_EcoM-CHD94UKE2_MZ234031	CHD94UKE2_044	19404	21044	1641	+	recombination-related endonuclease	QZI80983.1	protein
vB_EcoM-CHD94UKE2_MZ234031	CHD94UKE2_053	24677	27373	2697	+	DNA polymerase	QZI80991.1	protein
vB_EcoM-CHD94UKE2_MZ234031	CHD94UKE2_068	34239	35267	1029	+	DNA primase / DNA helicase	QZI81006.1	protein
vB_EcoM-CHD94UKE2_MZ234031	CHD94UKE2_071	36043	36717	675	+	homing endonuclease	QZI81009.1	protein
vB_EcoM-CHD94UKE2_MZ234031	CHD94UKE2_085	41597	42916	1320	+	DNA helicase	QZI81023.1	protein
vB_EcoM-CHD94UKE2_MZ234031	CHD94UKE2_087	43170	43415	246	+	exonuclease	QZI81025.1	protein
vB_EcoM-CHD94UKE2_MZ234031	CHD94UKE2_088	43408	43650	243	+	exonuclease	QZI81026.1	protein
vB_EcoM-CHD94UKE2_MZ234031	CHD94UKE2_089	43650	44333	684	+	exonuclease	QZI81027.1	protein
vB_EcoM-CHD94UKE2_MZ234031	CHD94UKE2_103	52944	53138	195	+	endonuclease	QZI81041.1	protein
vB_EcoM-CHD94UKE2_MZ234031	CHD94UKE2_104	53202	53759	558	+	endonuclease	QZI81042.1	protein
vB_EcoM-CHD94UKE2_MZ234031	CHD94UKE2_114	55892	57220	1329	+	DNA topoisomerase	QZI81052.1	protein
vB_EcoM-CHD94UKE2_MZ234031	CHD94UKE2_126	60721	61377	657	-	holin	QZI81064.1	protein
vB_EcoM-CHD94UKE2_MZ234031	CHD94UKE2_136	73158	73790	633	+	homing endonuclease	QZI81074.1	protein
vB_EcoM-CHD94UKE2_MZ234031	CHD94UKE2_150	82072	82482	411	+	endonuclease	QZI81088.1	protein
vB_EcoM-CHD94UKE2_MZ234031	CHD94UKE2_196	108245	108475	231	-	DNA helicase	QZI81134.1	protein
vB_EcoM-CHD94UKE2_MZ234031	CHD94UKE2_197	108501	110012	1512	-	DNA helicase	QZI81135.1	protein
vB_EcoM-CHD94UKE2_MZ234031	CHD94UKE2_256	155671	156084	414	+	endonuclease	QZI81194.1	protein
vB_EcoM-CHD94UKE2_MZ234031	CHD94UKE2_266	159226	159687	462	+	endonuclease	QZI81204.1	protein
vB_EcoM-G3G7_MZ234040	G3G7_017	8537	9010	474	+	endonuclease	QZI83144.1	protein
vB_EcoM-G3G7_MZ234040	G3G7_038	17012	18031	1020	+	recombination-related endonuclease	QZI83165.1	protein
vB_EcoM-G3G7_MZ234040	G3G7_041	18517	20157	1641	+	recombination-related endonuclease	QZI83168.1	protein
vB_EcoM-G3G7_MZ234040	G3G7_050	23790	26486	2697	+	DNA polymerase	QZI83176.1	protein
vB_EcoM-G3G7_MZ234040	G3G7_068	34955	35983	1029	+	DNA primase / DNA helicase	QZI83194.1	protein
vB_EcoM-G3G7_MZ234040	G3G7_088	43470	44789	1320	+	DNA helicase	QZI83214.1	protein
vB_EcoM-G3G7_MZ234040	G3G7_090	45043	45288	246	+	exonuclease	QZI83216.1	protein
vB_EcoM-G3G7_MZ234040	G3G7_091	45281	45523	243	+	exonuclease	QZI83217.1	protein
vB_EcoM-G3G7_MZ234040	G3G7_092	45523	46206	684	+	exonuclease	QZI83218.1	protein
vB_EcoM-G3G7_MZ234040	G3G7_106	54820	55014	195	+	endonuclease	QZI83232.1	protein
vB_EcoM-G3G7_MZ234040	G3G7_108	55315	55872	558	+	endonuclease	QZI83234.1	protein
vB_EcoM-G3G7_MZ234040	G3G7_118	58007	59335	1329	+	DNA topoisomerase	QZI83244.1	protein
vB_EcoM-G3G7_MZ234040	G3G7_130	62691	63347	657	-	holin	QZI83256.1	protein
vB_EcoM-G3G7_MZ234040	G3G7_141	75310	75942	633	+	homing endonuclease	QZI83267.1	protein
vB_EcoM-G3G7_MZ234040	G3G7_152	79807	80544	738	+	homing endonuclease	QZI83278.1	protein
vB_EcoM-G3G7_MZ234040	G3G7_158	85084	85494	411	+	endonuclease	QZI83284.1	protein
vB_EcoM-G3G7_MZ234040	G3G7_204	111271	111501	231	-	DNA helicase	QZI83330.1	protein
vB_EcoM-G3G7_MZ234040	G3G7_205	111527	113038	1512	-	DNA helicase	QZI83331.1	protein
vB_EcoM-G3G7_MZ234040	G3G7_261	157754	158170	417	+	endonuclease	QZI83387.1	protein
vB_EcoM-G3G7_MZ234040	G3G7_269	160368	160829	462	+	endonuclease	QZI83395.1	protein
vB_EcoP-UTI89UKE2_MZ234049	UTI89UKE2_011	7556	10102	2547	+	DNA-directed DNA polymerase	QZI84545.1	protein
vB_EcoP-UTI89UKE2_MZ234049	UTI89UKE2_021	13134	14162	1029	+	exonuclease	QZI84555.1	protein
vB_EcoP-UTI89UKE2_MZ234049	UTI89UKE2_040	33143	33337	195	+	class II holin	QZI84574.1	protein
vB_EcoP-UTI89UKE3_MZ234050	UTI89UKE3_011	7557	10103	2547	+	DNA-directed DNA polymerase	QZI84612.1	protein
vB_EcoP-UTI89UKE3_MZ234050	UTI89UKE3_021	13135	14163	1029	+	exonuclease	QZI84622.1	protein
vB_EcoP-UTI89UKE3_MZ234050	UTI89UKE3_040	33144	33338	195	+	class II holin	QZI84641.1	protein
vB_EcoM_LNA4_MZ311864	**vB_EcoM_LNA4_MZ311864_CDS_0121**	**55164**	**55628**	**465**	**+**	**endonuclease VII**	**self_AGKR121**	**protein**
vB_EcoM_LNA4_MZ311864	**vB_EcoM_LNA4_MZ311864_CDS_0124**	**57259**	**57663**	**405**	**+**	**Rz-like spanin**	**self_AGKR124**	**protein**
vB_EcoM_LNA4_MZ311864	**vB_EcoM_LNA4_MZ311864_CDS_0125**	**57660**	**57950**	**291**	**+**	**Rz-like spanin**	**self_AGKR125**	**protein**
vB_EcoM_LNA4_MZ311864	**vB_EcoM_LNA4_MZ311864_CDS_0165**	**94834**	**95265**	**432**	**-**	**endonuclease**	**self_AGKR165**	**protein**
vB_EcoM_LNA4_MZ311864	**vB_EcoM_LNA4_MZ311864_CDS_0187**	**104002**	**104475**	**474**	**-**	**HNH endonuclease**	**self_AGKR187**	**protein**
vB_EcoM_LNA4_MZ311864	**vB_EcoM_LNA4_MZ311864_CDS_0188**	**104502**	**106916**	**2415**	**-**	**DNA polymerase**	**self_AGKR188**	**protein**
vB_EcoM_LNA4_MZ311864	**vB_EcoM_LNA4_MZ311864_CDS_0190**	**107243**	**107746**	**504**	**-**	**HNH endonuclease**	**self_AGKR190**	**protein**
vB_EcoM_LNA4_MZ311864	**vB_EcoM_LNA4_MZ311864_CDS_0204**	**117557**	**118129**	**573**	**-**	**endolysin**	**self_AGKR204**	**protein**
vB_EcoM_LNA4_MZ311864	**vB_EcoM_LNA4_MZ311864_CDS_0214**	**121637**	**122140**	**504**	**-**	**HNH endonuclease**	**self_AGKR214**	**protein**
vB_EcoM_LNA4_MZ311864	**vB_EcoM_LNA4_MZ311864_CDS_0232**	**128771**	**129676**	**906**	**+**	**HNH endonuclease**	**self_AGKR232**	**protein**
vB_EcoM_LNA4_MZ311864	**vB_EcoM_LNA4_MZ311864_CDS_0234**	**129968**	**131431**	**1464**	**+**	**HNH endonuclease**	**self_AGKR234**	**protein**
vB_EcoM_LNA4_MZ311864	**vB_EcoM_LNA4_MZ311864_CDS_0029**	**8814**	**9209**	**396**	**-**	**endonuclease VII**	**self_AGKR29**	**protein**
vB_EcoM_LNA4_MZ311864	**vB_EcoM_LNA4_MZ311864_CDS_0056**	**19669**	**20190**	**522**	**-**	**HNH endonuclease**	**self_AGKR56**	**protein**
vB_EcoM_LNA4_MZ311864	**vB_EcoM_LNA4_MZ311864_CDS_0059**	**21617**	**22171**	**555**	**-**	**endolysin**	**self_AGKR59**	**protein**
vB_EcoM_LNA4_MZ311864	**vB_EcoM_LNA4_MZ311864_CDS_0075**	**31464**	**31949**	**486**	**-**	**endonuclease VII**	**self_AGKR75**	**protein**
vB_EcoM_LNA4_MZ311864	**vB_EcoM_LNA4_MZ311864_CDS_0078**	**32924**	**34045**	**1122**	**-**	**exonuclease**	**self_AGKR78**	**protein**
vB_EcoM_LNA4_MZ311864	**vB_EcoM_LNA4_MZ311864_CDS_0079**	**34054**	**34521**	**468**	**-**	**HNH endonuclease**	**self_AGKR79**	**protein**
vB_EcoM_SHAK7704_OR594181	**vB_EcoM_SHAK7704_OR594181_CDS_0036**	**18490**	**21036**	**2547**	**+**	**DNA polymerase**	**self_GXWA36**	**protein**
vB_EcoM_SHAK7704_OR594181	**vB_EcoM_SHAK7704_OR594181_CDS_0045**	**23981**	**25009**	**1029**	**+**	**exonuclease**	**self_GXWA45**	**protein**
vB_EcoM_SHAK7704_OR594181	**vB_EcoM_SHAK7704_OR594181_CDS_0046**	**24994**	**25407**	**414**	**+**	**endonuclease VII**	**self_GXWA46**	**protein**
vB_EcoM_SHAK7704_OR594181	**vB_EcoM_SHAK7704_OR594181_CDS_0005**	**3327**	**3671**	**345**	**+**	**endolysin**	**self_GXWA5**	**protein**
vB_EcoM_SHAK7704_OR594181	**vB_EcoM_SHAK7704_OR594181_CDS_0062**	**39690**	**42998**	**3309**	**+**	**internal virion protein with endolysin domain**	**self_GXWA62**	**protein**
vB_EcoM_SHAK7704_OR594181	**vB_EcoM_SHAK7704_OR594181_CDS_0064**	**43970**	**44164**	**195**	**+**	**holin**	**self_GXWA64**	**protein**
vB_EcoM_SHAK7704_OR594181	**vB_EcoM_SHAK7704_OR594181_CDS_0007**	**3780**	**3929**	**150**	**+**	**holin**	**self_GXWA7**	**protein**
vB_EcoM_SHAK7704_OR594181	**vB_EcoM_SHAK7704_OR594181_CDS_0008**	**4019**	**4216**	**198**	**+**	**Rz-like spanin**	**self_GXWA8**	**protein**
vB_EcoM_SHAK9454_OR594184	**vB_EcoM_SHAK9454_OR594184_CDS_0026**	**12288**	**12734**	**447**	**+**	**endonuclease**	**self_LVCG26**	**protein**
vB_EcoM_SHAK9454_OR594184	**vB_EcoM_SHAK9454_OR594184_CDS_0027**	**12721**	**13173**	**453**	**+**	**endonuclease**	**self_LVCG27**	**protein**
vB_EcoM_SHAK9454_OR594184	**vB_EcoM_SHAK9454_OR594184_CDS_0032**	**16083**	**16514**	**432**	**+**	**endonuclease VII**	**self_LVCG32**	**protein**
vB_EcoM_SHAK9454_OR594184	**vB_EcoM_SHAK9454_OR594184_CDS_0033**	**16507**	**18678**	**2172**	**+**	**DNA polymerase I**	**self_LVCG33**	**protein**
vB_EcoM_SHAK9454_OR594184	**vB_EcoM_SHAK9454_OR594184_CDS_0036**	**19159**	**20031**	**873**	**+**	**exonuclease**	**self_LVCG36**	**protein**

vB_EcoM_SHAK9454_OR594184	**vB_EcoM_SHAK9454_OR594184_CDS_0053**	**32160**	**36047**	**3888**	**+**	**internal virion protein with endolysin domain**	**self_LVCG53**	**protein**
vB_EcoM_SHAK9454_OR594184	**vB_EcoM_SHAK9454_OR594184_CDS_0055**	**38437**	**38631**	**195**	**+**	**holin**	**self_LVCG55**	**protein**
vB_EcoM_SHAK7854_OR594182	**vB_EcoM_SHAK7854_OR594182_CDS_0027**	**12729**	**13181**	**453**	**+**	**endonuclease**	**self_NJZI27**	**protein**
vB_EcoM_SHAK7854_OR594182	**vB_EcoM_SHAK7854_OR594182_CDS_0033**	**16369**	**18540**	**2172**	**+**	**DNA polymerase I**	**self_NJZI33**	**protein**
vB_EcoM_SHAK7854_OR594182	**vB_EcoM_SHAK7854_OR594182_CDS_0036**	**19021**	**19893**	**873**	**+**	**exonuclease**	**self_NJZI36**	**protein**
vB_EcoM_SHAK7854_OR594182	**vB_EcoM_SHAK7854_OR594182_CDS_0053**	**32030**	**35917**	**3888**	**+**	**internal virion protein with endolysin domain**	**self_NJZI53**	**protein**
vB_EcoM_SHAK7854_OR594182	**vB_EcoM_SHAK7854_OR594182_CDS_0056**	**38700**	**38894**	**195**	**+**	**holin**	**self_NJZI56**	**protein**
vB_EcoM_LNA1_MZ311863	**vB_EcoM_LNA1_MZ311863_CDS_0133**	**77244**	**77708**	**465**	**-**	**endolysin**	**self_TKOZ133**	**protein**
vB_EcoM_LNA1_MZ311863	**vB_EcoM_LNA1_MZ311863_CDS_0147**	**84570**	**84974**	**405**	**+**	**HNH endonuclease**	**self_TKOZ147**	**protein**
vB_EcoM_LNA1_MZ311863	**vB_EcoM_LNA1_MZ311863_CDS_0044**	**29475**	**30011**	**537**	**-**	**HNH endonuclease**	**self_TKOZ44**	**protein**
vB_EcoM_LNA1_MZ311863	**vB_EcoM_LNA1_MZ311863_CDS_0045**	**29998**	**30465**	**468**	**-**	**DNA polymerase**	**self_TKOZ45**	**protein**
vB_EcoM_LNA1_MZ311863	**vB_EcoM_LNA1_MZ311863_CDS_0047**	**30782**	**32965**	**2184**	**-**	**DNA polymerase**	**self_TKOZ47**	**protein**
vB_EcoM_LNA1_MZ311863	**vB_EcoM_LNA1_MZ311863_CDS_0056**	**38847**	**39887**	**1041**	**+**	**exonuclease**	**self_TKOZ56**	**protein**
vB_EcoM_LNA1_MZ311863	**vB_EcoM_LNA1_MZ311863_CDS_0069**	**47473**	**47937**	**465**	**+**	**HNH endonuclease**	**self_TKOZ69**	**protein**
vB_EcoM_LNA1_MZ311863	**vB_EcoM_LNA1_MZ311863_CDS_0094**	**63011**	**63358**	**348**	**+**	**Rz-like spanin**	**self_TKOZ94**	**protein**
vB_EcoP_LNA5_MZ311861	**vB_EcoP_LNA5_MZ311861_CDS_0014**	**7556**	**8008**	**453**	**-**	**endonuclease**	**self_UTCO14**	**protein**
vB_EcoP_LNA5_MZ311861	**vB_EcoP_LNA5_MZ311861_CDS_0041**	**20135**	**20584**	**450**	**-**	**Rz-like spanin**	**self_UTCO41**	**protein**
vB_EcoP_LNA5_MZ311861	**vB_EcoP_LNA5_MZ311861_CDS_0043**	**20946**	**21140**	**195**	**-**	**holin**	**self_UTCO43**	**protein**
vB_EcoP_LNA5_MZ311861	**vB_EcoP_LNA5_MZ311861_CDS_0045**	**23530**	**27417**	**3888**	**-**	**internal virion protein with endolysin domain**	**self_UTCO45**	**protein**
vB_EcoP_LNA5_MZ311861	**vB_EcoP_LNA5_MZ311861_CDS_0046**	**27399**	**27848**	**450**	**-**	**HNH endonuclease**	**self_UTCO46**	**protein**
vB_EcoP_LNA5_MZ311861	**vB_EcoP_LNA5_MZ311861_CDS_0005**	**956**	**1828**	**873**	**-**	**exonuclease**	**self_UTCO5**	**protein**
vB_EcoP_LNA5_MZ311861	**vB_EcoP_LNA5_MZ311861_CDS_0008**	**2309**	**4480**	**2172**	**-**	**DNA polymerase I**	**self_UTCO8**	**protein**
vB_EcoM_LNA2_MZ311865	**vB_EcoM_LNA2_MZ311865_CDS_0102**	**61701**	**63215**	**1515**	**+**	**DNA helicase**	**self_YRGQ102**	**protein**
vB_EcoM_LNA2_MZ311865	**vB_EcoM_LNA2_MZ311865_CDS_0103**	**63241**	**63474**	**234**	**+**	**DNA helicase**	**self_YRGQ103**	**protein**
vB_EcoM_LNA2_MZ311865	**vB_EcoM_LNA2_MZ311865_CDS_0144**	**85412**	**85702**	**291**	**-**	**Rz-like spanin**	**self_YRGQ144**	**protein**
vB_EcoM_LNA2_MZ311865	**vB_EcoM_LNA2_MZ311865_CDS_0145**	**85699**	**86055**	**357**	**-**	**Rz-like spanin**	**self_YRGQ145**	**protein**
vB_EcoM_LNA2_MZ311865	**vB_EcoM_LNA2_MZ311865_CDS_0148**	**87787**	**88197**	**411**	**-**	**endonuclease**	**self_YRGQ148**	**protein**
vB_EcoM_LNA2_MZ311865	**vB_EcoM_LNA2_MZ311865_CDS_0162**	**96832**	**97749**	**918**	**-**	**exonuclease**	**self_YRGQ162**	**protein**
vB_EcoM_LNA2_MZ311865	**vB_EcoM_LNA2_MZ311865_CDS_0168**	**107279**	**107938**	**660**	**+**	**holin**	**self_YRGQ168**	**protein**
vB_EcoM_LNA2_MZ311865	**vB_EcoM_LNA2_MZ311865_CDS_0179**	**111211**	**112536**	**1326**	**-**	**DNA topoisomerase II**	**self_YRGQ179**	**protein**
vB_EcoM_LNA2_MZ311865	**vB_EcoM_LNA2_MZ311865_CDS_0187**	**114186**	**114743**	**558**	**-**	**DenB-like DNA endonuclease IV**	**self_YRGQ187**	**protein**
vB_EcoM_LNA2_MZ311865	**vB_EcoM_LNA2_MZ311865_CDS_0202**	**124965**	**125642**	**678**	**-**	**exonuclease**	**self_YRGQ202**	**protein**
vB_EcoM_LNA2_MZ311865	**vB_EcoM_LNA2_MZ311865_CDS_0022**	**8492**	**8950**	**459**	**-**	**endoribonuclease**	**self_YRGQ22**	**protein**
vB_EcoM_LNA2_MZ311865	**vB_EcoM_LNA2_MZ311865_CDS_0240**	**146273**	**148984**	**2712**	**-**	**DNA polymerase**	**self_YRGQ240**	**protein**
vB_EcoM_LNA2_MZ311865	**vB_EcoM_LNA2_MZ311865_CDS_0271**	**162585**	**163058**	**474**	**-**	**endonuclease VII**	**self_YRGQ271**	**protein**
vB_EcoM_LNA2_MZ311865	**vB_EcoM_LNA2_MZ311865_CDS_0033**	**12174**	**12587**	**414**	**-**	**endonuclease V N-glycosylase UV repair enzyme**	**self_YRGQ33**	**protein**
vB_EcoP_LNA3_MZ311862	**vB_EcoP_LNA3_MZ311862_CDS_0010**	**13779**	**14228**	**450**	**+**	**Rz-like spanin**	**self_ZLQQ10**	**protein**
vB_EcoP_LNA3_MZ311862	**vB_EcoP_LNA3_MZ311862_CDS_0032**	**25829**	**26281**	**453**	**+**	**endonuclease**	**self_ZLQQ32**	**protein**
vB_EcoP_LNA3_MZ311862	**vB_EcoP_LNA3_MZ311862_CDS_0039**	**29664**	**31835**	**2172**	**+**	**DNA polymerase I**	**self_ZLQQ39**	**protein**
vB_EcoP_LNA3_MZ311862	**vB_EcoP_LNA3_MZ311862_CDS_0044**	**32925**	**33794**	**870**	**+**	**exonuclease**	**self_ZLQQ44**	**protein**
vB_EcoP_LNA3_MZ311862	**vB_EcoP_LNA3_MZ311862_CDS_0006**	**5943**	**9830**	**3888**	**+**	**internal virion protein with endolysin domain**	**self_ZLQQ6**	**protein**
vB_EcoP_LNA3_MZ311862	**vB_EcoP_LNA3_MZ311862_CDS_0008**	**13223**	**13417**	**195**	**+**	**holin**	**self_ZLQQ8**	**protein**
FS2B_MW929182	3_CDS	3221	3439	219	+	putative holin	UAG58399.1	protein
FS2B_MW929182	17_CDS	19824	20003	180	-	HNH endonuclease	UAG58413.1	protein
FS2B_MW929182	18_CDS	20017	20844	828	-	5'-3' exonuclease	UAG58414.1	protein
FS2B_MW929182	35_CDS	27708	29516	1809	-	DNA polymerase-like protein	UAG58431.1	protein
FS2B_MW929182	57_CDS	38048	38383	336	-	type I DNA polymerase	UAG58453.1	protein
590B_MW722821	19_CDS	11388	11858	471	-	HNH homing endonuclease	UCR96557.1	protein
590B_MW722821	30_CDS	23820	25244	1425	-	DNA helicase	UCR96560.1	protein
590B_MW722821	36_CDS	26741	28924	2184	-	DNA polymerase I	UCR96566.1	protein
590B_MW722821	44_CDS	32573	34795	2223	-	replicative DNA helicase	UCR96573.1	protein
NTEC3_OK539620	6_CDS	3391	3777	387	-	putative spanin protein	UDG73245.1	protein
NTEC3_OK539620	22_CDS	11540	12028	489	-	endolysin	UDG73261.1	protein
NTEC3_OK539620	30_CDS	14275	14748	474	-	homing endonuclease	UDG73269.1	protein
NTEC3_OK539620	42_CDS	17660	19900	2241	+	DNA helicase	UDG73281.1	protein
NTEC3_OK539620	49_CDS	23272	25452	2181	+	DNA polymerase I	UDG73288.1	protein
NTEC3_OK539620	52_CDS	25805	26089	285	+	endonuclease protein	UDG73291.1	protein
NTEC3_OK539620	55_CDS	27095	28519	1425	+	DNA helicase	UDG73294.1	protein
NTEC3_OK539620	67_CDS	40481	40951	471	+	HNH homing endonuclease	UDG73306.1	protein
MLP1_OK148439	25_CDS	21344	21685	342	+	holin	UEN68359.1	protein
MLP1_OK148439	46_CDS	32790	33185	396	+	endolysin	UEN68380.1	protein
MLP1_OK148439	61_CDS	41981	43915	1935	+	DNA polymerase	UEN68395.1	protein
MLP1_OK148439	64_CDS	45068	46105	1038	+	exonuclease	UEN68398.1	protein
MLP3_OK148440	8_CDS	9002	9823	822	-	exonuclease	UEN68414.1	protein
MLP3_OK148440	28_CDS	16096	17940	1845	-	DNA polymerase	UEN68434.1	protein
MLP3_OK148440	48_CDS	26258	26806	549	-	DNA polymerase	UEN68454.1	protein
MLP3_OK148440	87_CDS	45567	45872	306	-	DNA helicase	UEN68493.1	protein
MLP3_OK148440	109_CDS	62375	62887	513	+	HNH homing endonuclease	UEN68515.1	protein
MLP2_OK148441	15_CDS	7128	7601	474	+	endonuclease VII	UEN68540.1	protein
MLP2_OK148441	16_CDS	7598	8419	822	+	homing endonuclease	UEN68541.1	protein
MLP2_OK148441	36_CDS	16416	17435	1020	+	recombination-related endonuclease	UEN68561.1	protein
MLP2_OK148441	41_CDS	18748	20430	1683	+	recombination-related endonuclease	UEN68566.1	protein
MLP2_OK148441	49_CDS	24063	26759	2697	+	DNA polymerase	UEN68574.1	protein
MLP2_OK148441	64_CDS	34617	35645	1029	+	DNA primase / DNA helicase	UEN68589.1	protein
MLP2_OK148441	81_CDS	42777	44096	1320	+	DNA helicase	UEN68606.1	protein
MLP2_OK148441	83_CDS	44350	44595	246	+	exonuclease	UEN68608.1	protein
MLP2_OK148441	84_CDS	44588	44830	243	+	exonuclease	UEN68609.1	protein
MLP2_OK148441	85_CDS	44830	45513	684	+	exonuclease A	UEN68610.1	protein
MLP2_OK148441	101_CDS	54930	55406	477	+	endonuclease	UEN68626.1	protein
MLP2_OK148441	107_CDS	57541	58869	1329	+	DNA topoisomerase	UEN68632.1	protein
MLP2_OK148441	118_CDS	62212	62868	657	-	holin	UEN68643.1	protein
MLP2_OK148441	141_CDS	83382	83792	411	+	endonuclease	UEN68666.1	protein
MLP2_OK148441	144_CDS	85528	85881	354	+	spanin Rz	UEN68669.1	protein
MLP2_OK148441	185_CDS	109544	109774	231	-	DNA helicase	UEN68710.1	protein
MLP2_OK148441	186_CDS	109800	111311	1512	-	DNA helicase	UEN68711.1	protein
MLP2_OK148441	241_CDS	156976	157392	417	+	endonuclease	UEN68766.1	protein
MLP2_OK148441	248_CDS	159593	160054	462	+	endoribonuclease	UEN68773.1	protein
vB_EcoS-CHD2B1_MZ234026	CHD2B1_015	7520	9811	2292	+	DNA polymerase I	USL86271.1	protein
vB_EcoS-CHD2B1_MZ234026	CHD2B1_019	10670	12094	1425	+	DNA helicase	USL86275.1	protein
vB_EcoS-CHD2B1_MZ234026	CHD2B1_021	12822	13106	285	+	HNH endonuclease	USL86277.1	protein
vB_EcoS-CHD5UKE2_MZ234029	CHD5UKE2_048	24603	26894	2292	+	DNA polymerase I	USL86373.1	protein
vB_EcoS-CHD5UKE2_MZ234029	CHD5UKE2_053	27845	29269	1425	+	DNA helicase	USL86378.1	protein
vB_EcoS-CHD5UKE2_MZ234029	CHD5UKE2_055	29958	30281	324	+	HNH endonuclease	USL86380.1	protein
vB_EcoP-101114BS3_MZ234015	101114BS3_009	9813	10634	822	-	exonuclease	USM81136.1	protein
vB_EcoP-101114BS3_MZ234015	101114BS3_030	16893	18737	1845	-	DNA-directed DNA polymerase	USM81157.1	protein
vB_EcoP-101114BS3_MZ234015	101114BS3_053	27275	27823	549	-	DNA polymerase	USM81180.1	protein

vB_EcoP-101114BS3_MZ234015	101114BS3_098	46717	47022	306	-	DNA helicase	USM81225.1	protein
vB_EcoS-UTI89UKE1_MZ234048	UTI89UKE1_045	26618	28042	1425	-	DNA helicase	USM81316.1	protein
vB_EcoS-UTI89UKE1_MZ234048	UTI89UKE1_052	29715	31895	2181	-	DNA polymerase I	USM81323.1	protein
ZCEC14_OP168898	1_CDS	53	202	150	+	flap endonuclease	UVD33312.1	protein
ZCEC14_OP168898	18_CDS	13508	14425	918	+	endolysin	UVD33329.1	protein
ZCEC14_OP168898	23_CDS	16821	17048	228	+	DNA polymerase	UVD33334.1	protein
ZCEC14_OP168898	34_CDS	23952	24362	411	+	single-stranded-DNA-specific exonuclease	UVD33345.1	protein
ZCEC14_OP168898	39_CDS	26744	26974	231	+	flap endonuclease	UVD33349.1	protein
ZCEC14_OP168898	41_CDS	27270	28178	909	+	HNH homing endonuclease	UVD33351.1	protein
ZCEC14_OP168898	42_CDS	28243	28518	276	+	holin	UVD33352.1	protein
ZCEC14_OP168898	67_CDS	43518	45290	1773	-	DNA helicase	UVD33377.1	protein
ZCEC14_OP168898	72_CDS	48367	48765	399	+	putative HNH homing endonuclease	UVD33382.1	protein
ZCEC14_OP168898	78_CDS	51539	52219	681	+	endolysin	UVD33388.1	protein
ZCEC14_OP168898	88_CDS	59689	60114	426	-	putative replicative DNA helicase	UVD33398.1	protein
ZCEC14_OP168898	92_CDS	62619	63236	618	-	holin	UVD33402.1	protein
ZCEC14_OP168898	139_CDS	99740	100102	363	+	HNH homing endonuclease	UVD33449.1	protein
ZCEC14_OP168898	152_CDS	103689	104270	582	+	DNA polymerase	UVD33462.1	protein
ZCEC14_OP168898	157_CDS	105498	105701	204	+	putative DNA polymerase	UVD33467.1	protein
ZCEC14_OP168898	165_CDS	108203	109213	1011	+	putative HNH homing endonuclease	UVD33475.1	protein
ZCEC14_OP168898	168_CDS	109806	110774	969	+	putative DNA polymerase	UVD33478.1	protein
ZCEC14_OP168898	194_CDS	122637	122837	201	+	DNA helicase	UVD33502.1	protein
ZCEC14_OP168898	195_CDS	122890	123216	327	+	VSR endonuclease	UVD33503.1	protein
ZCEC14_OP168898	204_CDS	126654	127673	1020	+	holin	UVD33511.1	protein
ZCEC14_OP168898	205_CDS	127670	127933	264	+	endolysin	UVD33512.1	protein
ZCEC14_OP168898	229_CDS	144467	144667	201	-	DNA polymerase	UVD33536.1	protein
ZCEC14_OP168898	252_CDS	155146	155322	177	+	flap endonuclease	UVD33558.1	protein
Killian_OQ446694	PLBFAGBN_00003	3826	4482	657	+	holin protein	WEU67754.1	protein
Killian_OQ446694	PLBFAGBN_00013	7837	7986	150	-	flap endonuclease	WEU67764.1	protein
Killian_OQ446694	PLBFAGBN_00014	7983	9311	1329	-	DNA topoisomerase	WEU67765.1	protein
Killian_OQ446694	PLBFAGBN_00024	11659	12135	477	-	endonuclease	WEU67775.1	protein
Killian_OQ446694	PLBFAGBN_00057	29501	30151	651	-	HNH endonuclease	WEU67808.1	protein
Killian_OQ446694	PLBFAGBN_00073	38949	41645	2697	-	DNA polymerase	WEU67824.1	protein
Killian_OQ446694	PLBFAGBN_00082	46919	47182	264	-	endolysin	WEU67833.1	protein
Killian_OQ446694	PLBFAGBN_00083	47179	48198	1020	-	recombination-related endonuclease	WEU67834.1	protein
Killian_OQ446694	PLBFAGBN_00084	48176	48799	624	-	homing endonuclease	WEU67835.1	protein
Killian_OQ446694	PLBFAGBN_00091	52017	52232	216	-	recombination endonuclease VII	WEU67842.1	protein
Killian_OQ446694	PLBFAGBN_00100	54687	56504	1818	-	endolysin	WEU67851.1	protein
Killian_OQ446694	PLBFAGBN_00101	56495	57316	822	-	putative homing endonuclease	WEU67852.1	protein
Killian_OQ446694	PLBFAGBN_00102	57313	57786	474	-	endonuclease domain-containing protein	WEU67853.1	protein
Killian_OQ446694	PLBFAGBN_00144	75830	76159	330	-	endoribonuclease	WEU67895.1	protein
Killian_OQ446694	PLBFAGBN_00145	76156	76698	543	-	putative endoribonuclease	WEU67896.1	protein
Killian_OQ446694	PLBFAGBN_00146	76709	77260	552	-	endonuclease V N-glycosylase UV repair enzyme	WEU67897.1	protein
Killian_OQ446694	PLBFAGBN_00148	77692	78186	495	-	endolysin	WEU67899.1	protein
Killian_OQ446694	PLBFAGBN_00166	84533	85141	609	-	HNH endonuclease	WEU67910.1	protein
Killian_OQ446694	PLBFAGBN_00210	124433	125944	1512	+	DNA helicase	WEU67954.1	protein
Killian_OQ446694	PLBFAGBN_00211	125941	126633	693	+	homing endonuclease	WEU67955.1	protein
Killian_OQ446694	PLBFAGBN_00212	126659	126889	231	+	DNA helicase	WEU67956.1	protein
Killian_OQ446694	PLBFAGBN_00249	149309	149629	321	-	DNA helicase	WEU67993.1	protein
Killian_OQ446694	PLBFAGBN_00250	149626	149856	231	-	flap endonuclease	WEU67994.1	protein
Killian_OQ446694	PLBFAGBN_00255	152238	152648	411	-	endonuclease	WEU67999.1	protein
Killian_OQ446694	PLBFAGBN_00257	153851	154492	642	-	putative homing endonuclease	WEU68001.1	protein
Killian_OQ446694	PLBFAGBN_00273	163120	164037	918	-	endolysin	WEU68017.1	protein
ES_OR466728	3_CDS	875	1351	477	+	HNH endonuclease signature motif containing protein	WNO47585.1	protein
ES_OR466728	11_CDS	8977	9222	246	+	exonuclease	WNO47593.1	protein
ES_OR466728	17_CDS	11641	11889	249	+	exonuclease	WNO47599.1	protein
ES_OR466728	55_CDS	42316	42519	204	-	DNA helicase	WNO47637.1	protein
ES_OR466728	56_CDS	42531	42716	186	-	DNA helicase	WNO47638.1	protein
ES_OR466728	66_CDS	46107	46241	135	+	u-spanin	WNO47648.1	protein
ES_OR466728	67_CDS	46260	46421	162	+	endolysin	WNO47649.1	protein
vB-Eco-KMB41_OR539216	19_CDS	8327	10360	2034	+	DNA helicase	WPJ70605.1	protein
vB-Eco-KMB41_OR539216	21_CDS	10869	11909	1041	+	exonuclease	WPJ70607.1	protein
vB-Eco-KMB41_OR539216	24_CDS	13229	13669	441	+	endonuclease	WPJ70610.1	protein
vB-Eco-KMB22_OR539217	13_CDS	7843	8526	684	-	exonuclease	WPJ70663.1	protein
vB-Eco-KMB22_OR539217	14_CDS	8526	8768	243	-	exonuclease	WPJ70664.1	protein
vB-Eco-KMB22_OR539217	15_CDS	8761	9006	246	-	exonuclease	WPJ70665.1	protein
vB-Eco-KMB22_OR539217	17_CDS	9260	10579	1320	-	DNA helicase	WPJ70667.1	protein
vB-Eco-KMB22_OR539217	34_CDS	16647	17675	1029	-	DNA primase / DNA helicase	WPJ70684.1	protein
vB-Eco-KMB22_OR539217	49_CDS	25527	28223	2697	-	DNA polymerase	WPJ70699.1	protein
vB-Eco-KMB22_OR539217	57_CDS	31882	33522	1641	-	recombination-related endonuclease	WPJ70707.1	protein
vB-Eco-KMB22_OR539217	60_CDS	34008	35027	1020	-	recombination-related endonuclease	WPJ70710.1	protein
vB-Eco-KMB22_OR539217	79_CDS	43328	43801	474	-	endonuclease VII	WPJ70729.1	protein
vB-Eco-KMB22_OR539217	94_CDS	51453	52172	720	-	homing endonuclease	WPJ70744.1	protein
vB-Eco-KMB22_OR539217	117_CDS	60900	61361	462	-	endonuclease	WPJ70767.1	protein
vB-Eco-KMB22_OR539217	177_CDS	109169	110680	1512	+	DNA helicase	WPJ70827.1	protein
vB-Eco-KMB22_OR539217	178_CDS	110706	110936	231	+	DNA helicase	WPJ70828.1	protein
vB-Eco-KMB22_OR539217	222_CDS	136577	136987	411	-	endonuclease II	WPJ70872.1	protein
vB-Eco-KMB22_OR539217	235_CDS	145554	146186	633	-	homing endonuclease	WPJ70885.1	protein
vB-Eco-KMB22_OR539217	245_CDS	158059	158715	657	+	holin	WPJ70895.1	protein
vB-Eco-KMB22_OR539217	255_CDS	161772	163100	1329	-	DNA topoisomerase	WPJ70905.1	protein
vB-Eco-KMB22_OR539217	264_CDS	165228	165785	558	-	endonuclease IV	WPJ70914.1	protein
vB-Eco-KMB22_OR539217	265_CDS	165849	166043	195	-	endonuclease	WPJ70915.1	protein
vB-Eco-KMB26_OR539218	12_CDS	7630	8313	684	-	exonuclease	WPJ70928.1	protein
vB-Eco-KMB26_OR539218	13_CDS	8313	8555	243	-	exonuclease	WPJ70929.1	protein
vB-Eco-KMB26_OR539218	14_CDS	8548	8793	246	-	exonuclease	WPJ70930.1	protein
vB-Eco-KMB26_OR539218	16_CDS	9047	10366	1320	-	DNA helicase	WPJ70932.1	protein
vB-Eco-KMB26_OR539218	33_CDS	17510	18538	1029	-	DNA primase / DNA helicase	WPJ70949.1	protein
vB-Eco-KMB26_OR539218	49_CDS	27054	29750	2697	-	DNA polymerase	WPJ70965.1	protein
vB-Eco-KMB26_OR539218	57_CDS	33383	35023	1641	-	recombination-related endonuclease	WPJ70973.1	protein
vB-Eco-KMB26_OR539218	60_CDS	35509	36528	1020	-	recombination-related endonuclease	WPJ70976.1	protein
vB-Eco-KMB26_OR539218	78_CDS	44230	44703	474	-	endonuclease VII	WPJ70994.1	protein
vB-Eco-KMB26_OR539218	115_CDS	61060	61521	462	-	endonuclease	WPJ71031.1	protein
vB-Eco-KMB26_OR539218	122_CDS	63720	64136	417	-	endonuclease	WPJ71038.1	protein
vB-Eco-KMB26_OR539218	174_CDS	109198	110709	1512	+	DNA helicase	WPJ71090.1	protein
vB-Eco-KMB26_OR539218	175_CDS	110735	110965	231	+	DNA helicase	WPJ71091.1	protein
vB-Eco-KMB26_OR539218	219_CDS	136571	136981	411	-	endonuclease	WPJ71135.1	protein
vB-Eco-KMB26_OR539218	224_CDS	141520	142257	738	-	homing endonuclease	WPJ71140.1	protein

vB-Eco-KMB26_OR539218	234_CDS	146483	147115	633	-	homing endonuclease	WPJ71150.1	protein
vB-Eco-KMB26_OR539218	244_CDS	158944	159600	657	+	holin	WPJ71160.1	protein
vB-Eco-KMB26_OR539218	255_CDS	162943	164271	1329	-	DNA topoisomerase	WPJ71171.1	protein
vB-Eco-KMB26_OR539218	264_CDS	166406	166963	558	-	endonuclease	WPJ71180.1	protein
vB-Eco-KMB26_OR539218	265_CDS	167050	167244	195	-	endonuclease	WPJ71181.1	protein
vB-Eco-KMB39_OR539219	11_CDS	6893	7576	684	-	exonuclease	WPJ71193.1	protein
vB-Eco-KMB39_OR539219	12_CDS	7576	7818	243	-	exonuclease	WPJ71194.1	protein
vB-Eco-KMB39_OR539219	13_CDS	7811	8056	246	-	exonuclease	WPJ71195.1	protein
vB-Eco-KMB39_OR539219	14_CDS	8078	9397	1320	-	DNA helicase	WPJ71196.1	protein
vB-Eco-KMB39_OR539219	29_CDS	15097	16125	1029	-	DNA primase / DNA helicase	WPJ71211.1	protein
vB-Eco-KMB39_OR539219	43_CDS	23374	26070	2697	-	DNA polymerase	WPJ71225.1	protein
vB-Eco-KMB39_OR539219	51_CDS	29702	30418	717	+	homing endonuclease	WPJ71233.1	protein
vB-Eco-KMB39_OR539219	52_CDS	30411	32057	1647	-	recombination-related endonuclease	WPJ71234.1	protein
vB-Eco-KMB39_OR539219	55_CDS	32543	33562	1020	-	recombination-related endonuclease	WPJ71237.1	protein
vB-Eco-KMB39_OR539219	74_CDS	41564	42037	474	-	endonuclease VII	WPJ71256.1	protein
vB-Eco-KMB39_OR539219	112_CDS	58451	58912	462	-	endonuclease	WPJ71294.1	protein
vB-Eco-KMB39_OR539219	130_CDS	67654	68262	609	-	homing endonuclease	WPJ71312.1	protein
vB-Eco-KMB39_OR539219	140_CDS	72508	73167	660	-	HNH endonuclease	WPJ71322.1	protein
vB-Eco-KMB39_OR539219	175_CDS	108802	109419	618	+	homing endonuclease	WPJ71357.1	protein
vB-Eco-KMB39_OR539219	176_CDS	109422	110930	1509	+	DNA helicase	WPJ71358.1	protein
vB-Eco-KMB39_OR539219	177_CDS	110927	111619	693	+	homing endonuclease	WPJ71359.1	protein
vB-Eco-KMB39_OR539219	178_CDS	111645	111875	231	+	DNA helicase	WPJ71360.1	protein
vB-Eco-KMB39_OR539219	221_CDS	136879	137289	411	-	endonuclease	WPJ71403.1	protein
vB-Eco-KMB39_OR539219	223_CDS	138495	139226	732	-	homing endonuclease	WPJ71405.1	protein
vB-Eco-KMB39_OR539219	228_CDS	142507	143244	738	-	homing endonuclease	WPJ71410.1	protein
vB-Eco-KMB39_OR539219	237_CDS	147240	147872	633	-	homing endonuclease	WPJ71419.1	protein
vB-Eco-KMB39_OR539219	247_CDS	159974	160630	657	+	holin	WPJ71429.1	protein
vB-Eco-KMB39_OR539219	257_CDS	163634	164962	1329	-	DNA topoisomerase	WPJ71439.1	protein
vB-Eco-KMB39_OR539219	265_CDS	167095	167652	558	-	endonuclease	WPJ71447.1	protein
vB-Eco-KMB39_OR539219	266_CDS	167716	167910	195	-	endonuclease	WPJ71448.1	protein
vB-Eco-KMB40_OR539220	11_CDS	7255	7938	684	-	exonuclease	WPJ71460.1	protein
vB-Eco-KMB40_OR539220	12_CDS	7938	8180	243	-	exonuclease	WPJ71461.1	protein
vB-Eco-KMB40_OR539220	13_CDS	8173	8418	246	-	exonuclease	WPJ71462.1	protein
vB-Eco-KMB40_OR539220	14_CDS	8440	9759	1320	-	DNA helicase	WPJ71463.1	protein
vB-Eco-KMB40_OR539220	30_CDS	15567	16595	1029	-	DNA primase / DNA helicase	WPJ71479.1	protein
vB-Eco-KMB40_OR539220	45_CDS	24071	26767	2697	-	DNA polymerase	WPJ71494.1	protein
vB-Eco-KMB40_OR539220	52_CDS	30177	31859	1683	-	recombination-related endonuclease	WPJ71501.1	protein
vB-Eco-KMB40_OR539220	53_CDS	31875	32729	855	-	putative endonuclease	WPJ71502.1	protein
vB-Eco-KMB40_OR539220	56_CDS	33173	34192	1020	-	recombination-related endonuclease	WPJ71505.1	protein
vB-Eco-KMB40_OR539220	74_CDS	41896	42369	474	-	endonuclease VII	WPJ71523.1	protein
vB-Eco-KMB40_OR539220	112_CDS	58572	59033	462	-	endonuclease	WPJ71561.1	protein
vB-Eco-KMB40_OR539220	120_CDS	61714	62130	417	-	endonuclease	WPJ71569.1	protein
vB-Eco-KMB40_OR539220	174_CDS	107971	108588	618	+	homing endonuclease	WPJ71623.1	protein
vB-Eco-KMB40_OR539220	175_CDS	108591	110099	1509	+	DNA helicase	WPJ71624.1	protein
vB-Eco-KMB40_OR539220	176_CDS	110125	110358	234	+	DNA helicase	WPJ71625.1	protein
vB-Eco-KMB40_OR539220	219_CDS	135587	135997	411	-	endonuclease	WPJ71668.1	protein
vB-Eco-KMB40_OR539220	234_CDS	145041	145673	633	-	homing endonuclease	WPJ71683.1	protein
vB-Eco-KMB40_OR539220	244_CDS	157812	158468	657	+	holin	WPJ71693.1	protein
vB-Eco-KMB40_OR539220	254_CDS	161471	162799	1329	-	DNA topoisomerase	WPJ71703.1	protein
vB-Eco-KMB40_OR539220	262_CDS	164938	165414	477	-	endonuclease	WPJ71711.1	protein
vB-Eco-KMB40_OR539220	264_CDS	165796	165990	195	-	endonuclease	WPJ71713.1	protein
vB-Eco-KMB42_OR539221	13_CDS	7866	8549	684	-	exonuclease	WPJ71727.1	protein
vB-Eco-KMB42_OR539221	14_CDS	8542	8790	249	-	exonuclease	WPJ71728.1	protein
vB-Eco-KMB42_OR539221	15_CDS	8783	9028	246	-	exonuclease	WPJ71729.1	protein
vB-Eco-KMB42_OR539221	17_CDS	9282	10601	1320	-	DNA helicase	WPJ71731.1	protein
vB-Eco-KMB42_OR539221	33_CDS	16417	17445	1029	-	DNA primase / DNA helicase	WPJ71747.1	protein
vB-Eco-KMB42_OR539221	48_CDS	25303	27999	2697	-	DNA polymerase	WPJ71762.1	protein
vB-Eco-KMB42_OR539221	56_CDS	31632	33272	1641	-	recombination-related endonuclease	WPJ71770.1	protein
vB-Eco-KMB42_OR539221	59_CDS	33758	34777	1020	-	recombination-related endonuclease	WPJ71773.1	protein
vB-Eco-KMB42_OR539221	77_CDS	42773	43246	474	-	endonuclease VII	WPJ71791.1	protein
vB-Eco-KMB42_OR539221	113_CDS	59633	60094	462	-	endonuclease	WPJ71827.1	protein
vB-Eco-KMB42_OR539221	173_CDS	108375	109886	1512	+	DNA helicase	WPJ71887.1	protein
vB-Eco-KMB42_OR539221	174_CDS	109912	110145	234	+	DNA helicase	WPJ71888.1	protein
vB-Eco-KMB42_OR539221	217_CDS	135330	135740	411	-	endonuclease	WPJ71931.1	protein
vB-Eco-KMB42_OR539221	240_CDS	156823	157479	657	+	holin	WPJ71954.1	protein
vB-Eco-KMB42_OR539221	250_CDS	160481	161809	1329	-	DNA topoisomerase	WPJ71964.1	protein
vB-Eco-KMB42_OR539221	258_CDS	163939	164415	477	-	endonuclease	WPJ71972.1	protein
vB-Eco-KMB42_OR539221	260_CDS	164790	164984	195	-	endonuclease	WPJ71974.1	protein
vB-Eco-KMB14_OR424384	15_CDS	9910	12456	2547	+	DNA polymerase	WPK27879.1	protein
vB-Eco-KMB14_OR424384	23_CDS	15599	16576	978	+	exonuclease	WPK27887.1	protein
vB-Eco-KMB14_OR424384	24_CDS	16561	16971	411	+	endonuclease	WPK27888.1	protein
vB-Eco-KMB14_OR424384	42_CDS	35654	35848	195	+	class II holin	WPK27906.1	protein
vB-Eco-KMB14_OR424384	47_CDS	38777	39121	345	+	endolysin	WPK27911.1	protein
vB-Eco-KMB14_OR424384	49_CDS	39230	39379	150	+	holin	WPK27913.1	protein
vB-Eco-KMB47_OR424385	15_CDS	9480	12026	2547	+	DNA-directed DNA polymerase	WPK27933.1	protein
vB-Eco-KMB47_OR424385	23_CDS	15201	16229	1029	+	exonuclease	WPK27941.1	protein
vB-Eco-KMB47_OR424385	24_CDS	16214	16624	411	+	HNH endonuclease motif protein	WPK27942.1	protein
vB-Eco-KMB47_OR424385	43_CDS	35216	35410	195	+	class II holin	WPK27961.1	protein
vB-Eco-KMB47_OR424385	45_CDS	35695	36138	444	+	HNH homing endonuclease	WPK27963.1	protein
vB-Eco-KMB25_OR424386	32_CDS	27673	29646	1974	+	DNA helicase	WPK28005.1	protein
vB-Eco-KMB25_OR424386	33_CDS	29649	30056	408	+	endonuclease	WPK28006.1	protein
vB-Eco-KMB25_OR424386	40_CDS	33151	33375	225	+	holin	WPK28013.1	protein
vB-Eco-KMB25_OR424386	42_CDS	33866	34255	390	+	unimolecular spanin	WPK28015.1	protein
vB-Eco-KMB25_OR424386	46_CDS	35083	36663	1581	-	DNA helicase	WPK28019.1	protein
vB-Eco-KMB43_OR525698	8_CDS	5303	5971	669	-	exonuclease A	WQN06336.1	protein
vB-Eco-KMB43_OR525698	11_CDS	6395	7786	1392	-	DNA helicase	WQN06339.1	protein
vB-Eco-KMB43_OR525698	15_CDS	10023	11051	1029	-	DNA primase / DNA helicase	WQN06343.1	protein
vB-Eco-KMB43_OR525698	16_CDS	11059	11622	564	-	HNH homing endonuclease	WQN06344.1	protein
vB-Eco-KMB43_OR525698	26_CDS	16370	19048	2679	-	DNA polymerase	WQN06354.1	protein
vB-Eco-KMB43_OR525698	34_CDS	22837	24519	1683	-	recombination-related endonuclease	WQN06362.1	protein
vB-Eco-KMB43_OR525698	39_CDS	25927	26952	1026	-	recombination-related endonuclease	WQN06367.1	protein
vB-Eco-KMB43_OR525698	59_CDS	39163	39435	273	-	endonuclease VII	WQN06387.1	protein
vB-Eco-KMB43_OR525698	61_CDS	39944	40699	756	-	homing endonuclease	WQN06389.1	protein
vB-Eco-KMB43_OR525698	92_CDS	51447	51869	423	-	endoribonuclease	WQN06420.1	protein
vB-Eco-KMB43_OR525698	100_CDS	54619	55194	576	-	homing endonuclease	WQN06428.1	protein
vB-Eco-KMB43_OR525698	101_CDS	55240	55629	390	-	endolysin	WQN06429.1	protein

vB-Eco-KMB43_OR525698	164_CDS	101846	103330	1485	+	DNA helicase	WQN06492.1	protein
vB-Eco-KMB43_OR525698	165_CDS	103447	103683	237	+	DNA helicase	WQN06493.1	protein
vB-Eco-KMB43_OR525698	213_CDS	128805	129071	267	-	spanin Rz	WQN06541.1	protein
vB-Eco-KMB43_OR525698	215_CDS	130278	130739	462	-	endonuclease	WQN06543.1	protein
vB-Eco-KMB43_OR525698	242_CDS	154886	155542	657	+	holin	WQN06570.1	protein
vB-Eco-KMB43_OR525698	260_CDS	160880	162226	1347	-	DNA topoisomerase	WQN06588.1	protein
vB-Eco-KMB46_OR525699	7_CDS	5592	6647	1056	-	exonuclease	WQN06600.1	protein
vB-Eco-KMB46_OR525699	87_CDS	44819	45034	216	-	holin	WQN06680.1	protein
vB-Eco-KMB46_OR525699	95_CDS	48364	50361	1998	-	DNA helicase	WQN06688.1	protein
vB-Eco-KMB37_OR525700	87_CDS	35984	36904	921	-	exonuclease	WQN06777.1	protein
vB-Eco-KMB37_OR525700	90_CDS	38358	39242	885	-	exonuclease	WQN06780.1	protein
vB-Eco-KMB37_OR525700	136_CDS	55278	57362	2085	+	large terminase with endonuclease	WQN06826.1	protein
vB-Eco-KMB37_OR525700	178_CDS	96354	98933	2580	-	DNA polymerase	WQN06868.1	protein
vB-Eco-KMB37_OR525700	218_CDS	124434	124949	516	-	HNH endonuclease	WQN06908.1	protein
vB-Eco-KMB38_OR525701	13_CDS	8969	9646	678	-	exonuclease	WQN06997.1	protein
vB-Eco-KMB38_OR525701	14_CDS	9656	10969	1314	-	DNA helicase	WQN06998.1	protein
vB-Eco-KMB38_OR525701	27_CDS	16084	17106	1023	-	DNA primase / DNA helicase	WQN07011.1	protein
vB-Eco-KMB38_OR525701	50_CDS	30469	33180	2712	-	DNA polymerase	WQN07034.1	protein
vB-Eco-KMB38_OR525701	57_CDS	36637	38325	1689	-	recombination-related endonuclease	WQN07041.1	protein
vB-Eco-KMB38_OR525701	60_CDS	38802	39821	1020	-	recombination-related endonuclease	WQN07044.1	protein
vB-Eco-KMB38_OR525701	80_CDS	47086	47559	474	-	recombination endonuclease VII	WQN07064.1	protein
vB-Eco-KMB38_OR525701	111_CDS	61692	62150	459	-	site-specific RNA endonuclease	WQN07095.1	protein
vB-Eco-KMB38_OR525701	121_CDS	65293	65706	414	-	endonuclease V	WQN07105.1	protein
vB-Eco-KMB38_OR525701	184_CDS	114122	114853	732	+	homing endonuclease	WQN07168.1	protein
vB-Eco-KMB38_OR525701	185_CDS	114859	116373	1515	+	DNA helicase	WQN07169.1	protein
vB-Eco-KMB38_OR525701	186_CDS	116399	116632	234	+	DNA helicase	WQN07170.1	protein
vB-Eco-KMB38_OR525701	226_CDS	138577	138933	357	-	spanin Rz	WQN07210.1	protein
vB-Eco-KMB38_OR525701	229_CDS	140664	141074	411	-	endonuclease II	WQN07213.1	protein
vB-Eco-KMB38_OR525701	250_CDS	160976	161635	660	+	holin	WQN07234.1	protein
vB-Eco-KMB38_OR525701	260_CDS	164907	166232	1326	-	DNA topoisomerase II	WQN07244.1	protein
vB-Eco-KMB38_OR525701	267_CDS	167882	168358	477	-	DNA endonuclease IV	WQN07251.1	protein
vB-Eco-KMB23_OR525702	13_CDS	7765	8448	684	-	exonuclease	WQN07266.1	protein
vB-Eco-KMB23_OR525702	14_CDS	8448	8690	243	-	exonuclease	WQN07267.1	protein
vB-Eco-KMB23_OR525702	15_CDS	8683	8928	246	-	exonuclease	WQN07268.1	protein
vB-Eco-KMB23_OR525702	17_CDS	9182	10501	1320	-	DNA helicase	WQN07270.1	protein
vB-Eco-KMB23_OR525702	34_CDS	16570	17598	1029	-	DNA primase / DNA helicase	WQN07287.1	protein
vB-Eco-KMB23_OR525702	50_CDS	25442	28138	2697	-	DNA polymerase	WQN07303.1	protein
vB-Eco-KMB23_OR525702	58_CDS	31797	33479	1683	-	recombination-related endonuclease	WQN07311.1	protein
vB-Eco-KMB23_OR525702	59_CDS	33495	34349	855	-	homing endonuclease	WQN07312.1	protein
vB-Eco-KMB23_OR525702	62_CDS	34793	35812	1020	-	recombination-related endonuclease	WQN07315.1	protein
vB-Eco-KMB23_OR525702	82_CDS	43508	43981	474	-	endonuclease VII	WQN07335.1	protein
vB-Eco-KMB23_OR525702	121_CDS	60163	60624	462	-	site-specific RNA endonuclease	WQN07374.1	protein
vB-Eco-KMB23_OR525702	144_CDS	69845	70495	651	-	homing endonuclease	WQN07397.1	protein
vB-Eco-KMB23_OR525702	188_CDS	109718	111229	1512	+	DNA helicase	WQN07441.1	protein
vB-Eco-KMB23_OR525702	189_CDS	111255	111485	231	+	DNA helicase	WQN07442.1	protein
vB-Eco-KMB23_OR525702	234_CDS	136641	137051	411	-	endonuclease	WQN07487.1	protein
vB-Eco-KMB23_OR525702	259_CDS	157793	158449	657	+	holin	WQN07512.1	protein
vB-Eco-KMB23_OR525702	279_CDS	164920	165396	477	-	endonuclease IV	WQN07532.1	protein
vB-Eco-KMB23_OR525702	281_CDS	165778	165972	195	-	endonuclease	WQN07534.1	protein
vB-Eco-KMB36_OR525703	5_CDS	5559	5945	387	+	spanin	WQN07540.1	protein
vB-Eco-KMB36_OR525703	19_CDS	12411	12950	540	-	HNH homing endonuclease	WQN07554.1	protein
vB-Eco-KMB36_OR525703	34_CDS	24635	26059	1425	-	DNA helicase	WQN07569.1	protein
vB-Eco-KMB36_OR525703	37_CDS	26564	27061	498	-	HNH endonuclease motif protein	WQN07572.1	protein
vB-Eco-KMB36_OR525703	40_CDS	27265	30360	3096	-	DNA polymerase I	WQN07575.1	protein
vB-Eco-KMB36_OR525703	41_CDS	30353	30688	336	-	HNH homing endonuclease	WQN07576.1	protein
vB-Eco-KMB36_OR525703	66_CDS	41007	41258	252	+	putative holin	WQN07601.1	protein
vB-Eco-KMB35_OR525704	7_CDS	4905	5291	387	+	spanin	WQN07620.1	protein
vB-Eco-KMB35_OR525704	35_CDS	24277	25701	1425	-	DNA helicase	WQN07648.1	protein
vB-Eco-KMB35_OR525704	39_CDS	26981	27079	99	-	HNH endonuclease	WQN07652.1	protein
vB-Eco-KMB35_OR525704	41_CDS	27197	27928	732	-	restriction endonuclease	WQN07654.1	protein
vB-Eco-KMB35_OR525704	42_CDS	27972	30152	2181	-	DNA polymerase I	WQN07655.1	protein
vB-Eco-KMB35_OR525704	50_CDS	33774	36008	2235	-	replicative DNA helicased	WQN07663.1	protein
vB-Eco-KMB35_OR525704	69_CDS	40124	40432	309	+	putative holine	WQN07682.1	protein
vB-Eco-KMB35_OR525704	70_CDS	40425	40697	273	+	putative holin	WQN07683.1	protein
vB_EcoM_SHAK7163_OR594179	SHAK7163_0020	10447	12993	2547	+	DNA polymerase	WVW77634.1	protein
vB_EcoM_SHAK7163_OR594179	SHAK7163_0029	15938	16966	1029	+	exonuclease	WVW77643.1	protein
vB_EcoM_SHAK7163_OR594179	SHAK7163_0046	31648	34968	3321	+	internal virion protein with endolysin domain	WVW77660.1	protein
vB_EcoM_SHAK7163_OR594179	SHAK7163_0048	35940	36134	195	+	holin	WVW77662.1	protein
vB_EcoM_SHAK7163_OR594179	SHAK7163_0054	39749	40093	345	+	endolysin	WVW77668.1	protein
vB_EcoM_SHAK7163_OR594179	SHAK7163_0056	40202	40351	150	+	holin	WVW77670.1	protein
vB_EcoM_SHAK7163_OR594179	SHAK7163_0057	40441	40638	198	+	Rz-like spanin	WVW77671.1	protein
vB_EcoM_SHAK7693_OR594180	SHAK7693_0008	5483	5869	387	+	Rz-like spanin	WVW77687.1	protein
vB_EcoM_SHAK7693_OR594180	SHAK7693_0025	12924	14054	1131	-	DNA repair exonuclease	WVW77704.1	protein
vB_EcoM_SHAK7693_OR594180	SHAK7693_0031	17626	18276	651	-	HNH endonuclease	WVW77710.1	protein
vB_EcoM_SHAK7693_OR594180	SHAK7693_0039	25778	27166	1389	-	DNA helicase	WVW77718.1	protein
vB_EcoM_SHAK7693_OR594180	SHAK7693_0046	29005	31206	2202	-	DNA polymerase	WVW77725.1	protein
vB_EcoM_SHAK7693_OR594180	SHAK7693_0049	32133	32624	492	-	HNH endonuclease	WVW77728.1	protein
vB_EcoM_SHAK7693_OR594180	SHAK7693_0050	32621	33862	1242	-	exonuclease	WVW77729.1	protein
vB_EcoM_SHAK7693_OR594180	SHAK7693_0072	41883	42191	309	+	holin	WVW77751.1	protein
vB_EcoM_SHAK7693_OR594180	SHAK7693_0073	42184	42456	273	+	holin	WVW77752.1	protein
vB_EcoM_SHAK7693_OR594180	SHAK7693_0074	42419	42658	240	+	endolysin	WVW77753.1	protein
vB_EcoM_SHAK7693_OR594180	SHAK7693_0075	42651	42953	303	+	endolysin	WVW77754.1	protein
vB_EcoM_SHAK7858_OR594183	SHAK7858_0007	4898	5284	387	+	Rz-like spanin	WVW77770.1	protein
vB_EcoM_SHAK7858_OR594183	SHAK7858_0008	5250	5393	144	+	Rz-like spanin	WVW77771.1	protein
vB_EcoM_SHAK7858_OR594183	SHAK7858_0022	12701	13831	1131	-	DNA repair exonuclease	WVW77785.1	protein
vB_EcoM_SHAK7858_OR594183	SHAK7858_0035	24873	26297	1425	-	DNA helicase	WVW77798.1	protein
vB_EcoM_SHAK7858_OR594183	SHAK7858_0041	27162	29342	2181	-	DNA polymerase	WVW77804.1	protein
vB_EcoM_SHAK7858_OR594183	SHAK7858_0044	30288	31538	1251	-	exonuclease	WVW77807.1	protein
vB_EcoM_SHAK7858_OR594183	SHAK7858_0068	39935	40243	309	+	holin	WVW77831.1	protein
vB_EcoM_SHAK7858_OR594183	SHAK7858_0069	40236	40508	273	+	holin	WVW77832.1	protein
vB_EcoM_SHAK7858_OR594183	SHAK7858_0070	40486	41001	516	+	HNH endonuclease	WVW77833.1	protein
vB_EcoM_SHAK7858_OR594183	SHAK7858_0071	40989	41474	486	+	endolysin	WVW77834.1	protein
vB_EcoS_UTEC10_PP291582	7_CDS	1847	2332	486	-	endolysin	WWY66481.1	protein
vB_EcoS_UTEC10_PP291582	8_CDS	2310	2582	273	-	holin	WWY66482.1	protein
vB_EcoS_UTEC10_PP291582	31_CDS	13241	15412	2172	+	DNA polymerase	WWY66505.1	protein
vB_EcoS_UTEC10_PP291582	64_CDS	37822	37968	147	-	Rz-like spanin	WWY66538.1	protein

Mimir124_PQ310662	21_CDS	6030	6536	507	+	HNH endonuclease	XLZ23656.1	protein
Mimir124_PQ310662	29_CDS	10848	11324	477	+	HNH endonuclease	XLZ23664.1	protein
Mimir124_PQ310662	30_CDS	11324	12499	1176	+	HNH endonuclease	XLZ23665.1	protein
Mimir124_PQ310662	45_CDS	21725	22204	480	+	HNH endonuclease	XLZ23680.1	protein
Mimir124_PQ310662	49_CDS	24554	27133	2580	+	DNA polymerase	XLZ23684.1	protein
Mimir124_PQ310662	52_CDS	27980	28477	498	+	HNH endonuclease	XLZ23687.1	protein
Mimir124_PQ310662	53_CDS	28474	29454	981	+	exonuclease	XLZ23688.1	protein
Mimir124_PQ310662	82_CDS	59644	60138	495	-	Rz-like spanin	XLZ23717.1	protein
Mimir124_PQ310662	83_CDS	60125	60760	636	-	endolysin	XLZ23718.1	protein
Mimir124_PQ310662	84_CDS	60750	61010	261	-	holin	XLZ23719.1	protein

Supplementary Table S3

The list of putative phage depolymerases identified accordingly to the criteria described below the list. Proteins marked in bold (7 in total) were found in genomes of phages annotated first time in this work (Supplementary Table S1). The remaining depolymerases were annotated by the authors but not indicated strictly as depolymerases.

Phage name	Phage acc. No.	Protein name indicated by annotations	Protein acc. No.	Number of amino acids	Gene position	Strand	Template	Confidence	% i.d.	Aligned Residues	PDB header	Chain	PDB Molecule	PDBTitle
vB_EcoP-22664UKE3-2	MZ234011	hypothetical_protein	QZI78577.1	843	3368:5899	+	c2x3hA_	100	89	485	lyase	A	K5 LYASE	coliphage k5a lyase
vB_EcoP-101101UKE1	MZ234012	hypothetical_protein	QZI78684.1	632	37884:39782	+	c2x3hA_	100	98	498	lyase	A	K5 LYASE	coliphage k5a lyase
vB_EcoP-101120B1-2	MZ234022	hypothetical_protein	QZI79749.1	843	3368:5899	+	c2x3hA_	100	91	479	lyase	A	K5 LYASE	coliphage k5a lyase
vB_EcoP-101120B2	MZ234023	hypothetical_protein	QZI79809.1	843	3368:5899	+	c2x3hA_	100	91	478	lyase	A	K5 LYASE	coliphage k5a lyase
vB_EcoP-UTI89UKE2	MZ234049	unclassified_head-tail_protein	QZI84586.1	811	38367:40802	+	c3ju4A_	100	82	666	hydrolase	A	Endo-N-acetylneuraminidase	crystal structure analysis of endosialidasenf at 0.98 a resolution
vB_EcoP-UTI89UKE2	MZ234049	unclassified_head-tail_protein	QZI84586.1	811	38367:40802	+	c3gvjA_	100	81	666	hydrolase	A	Endo-N-acetylneuraminidase	crystal structure of an endo-neuraminidasenf mutant
vB_EcoP-UTI89UKE2	MZ234049	unclassified_head-tail_protein	QZI84586.1	811	38367:40802	+	c4hizA_	100	52	646	hydrolase,viral protein	A	Endosialidase	phage phi92 endosialidase
vB_EcoP-UTI89UKE3	MZ234050	unclassified_head-tail_protein	QZI84653.1	811	38368:40803	+	c3ju4A_	100	82	666	hydrolase	A	Endo-N-acetylneuraminidase	crystal structure analysis of endosialidasenf at 0.98 a resolution
vB_EcoP-UTI89UKE3	MZ234050	unclassified_head-tail_protein	QZI84653.1	811	38368:40803	+	c3gvjA_	100	82	666	hydrolase	A	Endo-N-acetylneuraminidase	crystal structure of an endo-neuraminidasenf mutant
vB_EcoP-UTI89UKE3	MZ234050	unclassified_head-tail_protein	QZI84653.1	811	38368:40803	+	c4hizA_	100	52	646	hydrolase,viral protein	A	Endosialidase	phage phi92 endosialidase
vB_EcoM_SHAK7704	**OR594181**	**tail_spike_protein**	**self_GXWA10**	**811**	**4751:7186**	**+**	**c3ju4A_**	**100**	**81**	**666**	**hydrolase**	**A**	**Endo-N-acetylneuraminidase**	**crystal structure analysis of endosialidasenf at 0.98 a resolution**
vB_EcoM_SHAK7704	**OR594181**	**tail_spike_protein**	**self_GXWA10**	**811**	**4751:7186**	**+**	**c3gvjA_**	**100**	**81**	**666**	**hydrolase**	**A**	**Endo-N-acetylneuraminidase**	**crystal structure of an endo-neuraminidasenf mutant**
vB_EcoM_SHAK7704	**OR594181**	**tail_spike_protein**	**self_GXWA10**	**811**	**4751:7186**	**+**	**c4hizA_**	**100**	**52**	**646**	**hydrolase,viral protein**	**A**	**Endosialidase**	**phage phi92 endosialidase**
vB_EcoM_SHAK7854	**OR594182**	**tail_spike_protein**	**self_NJZI54**	**843**	**35982:38513**	**+**	**c2x3hA_**	**100**	**79**	**476**	**lyase**	**A**	**K5 LYASE**	**coliphage k5a lyase**
vB_EcoP_LNA3	**MZ311862**	**tail_protein**	**self_ZLQQ7**	**1064**	**9896:13090**	**+**	**c3gvjA_**	**100**	**97**	**668**	**hydrolase**	**A**	**Endo-N-acetylneuraminidase**	**crystal structure of an endo-neuraminidasenf mutant**
vB_EcoP_LNA3	**MZ311862**	**tail_protein**	**self_ZLQQ7**	**1064**	**9896:13090**	**+**	**c3ju4A_**	**100**	**97**	**667**	**hydrolase**	**A**	**Endo-N-acetylneuraminidase**	**crystal structure analysis of endosialidasenf at 0.98 a resolution**
vB_EcoP_LNA3	**MZ311862**	**tail_protein**	**self_ZLQQ7**	**1064**	**9896:13090**	**+**	**c4hizA_**	**100**	**55**	**642**	**hydrolase,viral protein**	**A**	**Endosialidase**	**phage phi92 endosialidase**
vB_EcoS-UTI89UKE1	MZ234048	tailspike	USM81314.1	984	23501:26455	+	c4hizA_	100	80	656	hydrolase,viral protein	A	Endosialidase	phage phi92 endosialidase
vB_EcoS-UTI89UKE1	MZ234048	tailspike	USM81314.1	984	23501:26455	+	c3gvjA_	100	52	669	hydrolase	A	Endo-N-acetylneuraminidase	crystal structure of an endo-neuraminidasenf mutant
vB_EcoS-UTI89UKE1	MZ234048	tailspike	USM81314.1	984	23501:26455	+	c3ju4A_	100	52	668	hydrolase	A	Endo-N-acetylneuraminidase	crystal structure analysis of endosialidasenf at 0.98 a resolution
vB-Eco-KMB14	OR424384	tail_spike_protein	WPK27915.1	632	39798:41696	+	c2x3hA_	100	98	498	lyase	A	K5 LYASE	coliphage k5a lyase
vB-Eco-KMB47	OR424385	hypothetical_protein	WPK27970.1	632	39792:41690	+	c2x3hA_	100	98	498	lyase	A	K5 LYASE	coliphage k5a lyase
vB-Eco-KMB37	OR525700	colanidase_tailspike	WQN06859.1	804	85229:87643	+	c6e0wA_	100	99	599	hydrolase	A	Bacteriophage Phi92 gp150	crystal structure of the colanidase tailspike protein gp150 of phage phi92 complexed with one repeating unit of colanic acid
vB-Eco-KMB36	OR525703	tailspike	WQN07566.1	778	21945:24281	+	c2x3hA_	100	59	472	lyase	A	K5 LYASE	coliphage k5a lyase
vB_EcoM_SHAK7163	OR594179	tail_spike_protein	WVW77673.1	811	41173:43608	+	c3ju4A_	100	83	666	hydrolase	A	Endo-N-acetylneuraminidase	crystal structure analysis of endosialidasenf at 0.98 a resolution
vB_EcoM_SHAK7163	OR594179	tail_spike_protein	WVW77673.1	811	41173:43608	+	c3gvjA_	100	83	666	hydrolase	A	Endo-N-acetylneuraminidase	crystal structure of an endo-neuraminidasenf mutant
vB_EcoM_SHAK7163	OR594179	tail_spike_protein	WVW77673.1	811	41173:43608	+	c4hizA_	100	53	644	hydrolase,viral protein	A	Endosialidase	phage phi92 endosialidase

Mimir124	PQ310662	tail_fiber_protein	XLZ23723.1	788	61722:64088	-	c7vulA_	100	44	603	lyase	A	Non-contractile tail fiber protein	structure of klebsiella phage p560 depolymerase

Criteria for predicting putative depolymerase activity were applied by us as indicated by Latka et al., 2019. First, we have selected proteins longer than 200 residues, and annotated by the authors as tail/tail spike/hypothetical proteins in the NCBI database. Next, we focused only on proteins that showed homology to domains annotated as lyase [hyaluronate lyases (hyaluronidases), pectin/pectate lyases, alginate lyases, K5 lyases] or hydrolase (sialidases, rhamnosidases, levanases, dextranases, and xylanases) with a confidence = 100% and i.d.% of at least 40% as analyzed by Phyre2 software (Kelley et al., 2015). Finally, we confirmed that the length of homology with one of these enzymatic domains spans at least 100 residues and predicted a typical β-helical structure by Phyre2. Importantly all selected proteins possess such a structure and in the Table above, we have included the results of the described analyses. Importantly, confidence - represents the probability (from 0 to 100) that the match between the sequence considered as depolymerase and its template protein/domain is a true homology whereas the % i.d. means percentage identity between the potential depolymerase sequence and the template. Besides,in the table, we also included basic information from PDB (PDB header, Chain, PDB Molecule, PDBTitle) about the template protein as shown by Phyre2.

Citations

Latka A, Leiman PG, Drulis-Kawa Z, Briers Y. Modeling the Architecture of Depolymerase-Containing Receptor Binding Proteins in *Klebsiella* Phages. Front Microbiol. 2019 Nov 15;10:2649. doi: 10.3389/fmicb.2019.02649.

Kelley, L., Mezulis, S., Yates, C. et al. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc 10, 845–858 (2015). https://doi.org/10.1038/nprot.2015.053

Supplementary Table S4

Development prediction provided by Phacts.

Phage name _ Acc. No.	Phage Development Prediction	Probability	Standard deviation
Killian_OQ446694	Lytic	0.8084915084915085	0.03981755315791634
ZCEC14_OP168898	Lytic	0.8066933066933066	0.04993856011707866
vB-Eco-KMB42_OR539221	Lytic	0.8001998001998002	0.052733628246767875
UGKSEcP2_OV876900	Lytic	0.7993006993006994	0.057939388136500354
vB-Eco-KMB23_OR525702	Lytic	0.7909090909090908	0.03151860025847168
vB_EcoM-CHD2BS1_MZ234027	Lytic	0.7841158841158842	0.048862179518650066
vB_EcoM-101112UKE3-1_MZ234013	Lytic	0.7812187812187812	0.07698791978793594
vB-Eco-KMB40_OR539220	Lytic	0.7809190809190809	0.05155861480091239
vB_EcoM_KMB26_OW536688	Lytic	0.7802197802197802	0.0549159854079656
vB-Eco-KMB26_OR539218	Lytic	0.7795204795204794	0.04469019552021623
UGKSEcP1_OV877085	Lytic	0.767032967032967	0.04781824488965274
vB_EcoM-101117BS1_MZ234018	Lytic	0.764035964035964	0.05967105805184602
MLP2_OK148441	Lytic	0.7585414585414585	0.056644413777406304
vB_EcoM-G3G7_MZ234040	Lytic	0.7584415584415585	0.054589349015637396
vB_EcoM_ACG-C40_JN986846	Lytic	0.7551448551448552	0.05331044130086957
vB-Eco-KMB22_OR539217	Lytic	0.7549450549450549	0.03715974872496407
vB_EcoM-CHD94UKE2_MZ234031	Lytic	0.752947052947053	0.05051439438630861
vB-Eco-KMB39_OR539219	Lytic	0.7492507492507492	0.0865310191655401
vB-Eco-KMB43_OR525698	Lytic	0.7398601398601399	0.052147852147852135
Ec_Makalu_002_MN709127	Lytic	0.7269730269730271	0.03217046917418424
Ec_Makalu_003_MN882349	Lytic	0.7230769230769231	0.06846590503393363
vB_EcoM-CHD16UKE1_MZ234030	Lytic	0.7149850149850149	0.08105838947652758
vB_EcoM_LNA2_MZ311865	Lytic	0.7107892107892108	0.0513398072902401
Ec_Makalu_001_MN894885	Lytic	0.7077922077922079	0.057932153206275656
vB_EcoM-172859UKE1_MZ234025	Lytic	0.700899100899101	0.06610216817343365
vB-Eco-KMB38_OR525701	Lytic	0.6780219780219781	0.05192316702049144
EG1_MG488277	Lytic	0.6559440559440559	0.04749238561170755
vB_EcoP_LNA3_MZ311862	Lytic	0.6224775224775224	0.040845460315579604
vB_EcoP-101120B1-2_MZ234022	Lytic	0.6087912087912087	0.05809549424877568
vB_EcoM_SHAK7854_OR594182	Lytic	0.6038961038961039	0.05752415695778564
vB_EcoM_SHAK9454_OR594184	Lytic	0.5988011988011988	0.03798884378792097
vB_EcoP-101136BS1_MZ234024	Lytic	0.5978021978021979	0.05380935060126192
vB_EcoP-101118UKE1_MZ234021	Lytic	0.5943056943056944	0.04083324167140766
vB_EcoP-22664UKE3-2_MZ234011	Lytic	0.5928071928071929	0.04052349996910999
vB_EcoP-22664BS1_MZ234009	Lytic	0.5921078921078922	0.05455323009792227
vB_EcoP_LNA5_MZ311861	Lytic	0.5872127872127872	0.03922602578799767
vB_EcoP-101120B2_MZ234023	Lytic	0.5832167832167834	0.038960526649466067
vB_EcoP-101114UKE3_MZ234017	Lytic	0.5485514485514486	0.042992922311362966
RDN8_1_MN414250	Lytic	0.5369630369630369	0.02838024520552686
vB_EcoP-101101UKE1_MZ234012	Lytic	0.532967032967033	0.040570388723319334
vB_EcoM_LNA1_MZ311863	Lytic	0.5268731268731268	0.02939576353934294
vB-Eco-KMB25_OR424386	Lytic	0.5236763236763238	0.04774305104154496
FS2B_MW929183_MW929182	Lytic	0.5234765234765234	0.04401730360260719
vB_EcoM_SHAK7163_OR594179	Lytic	0.5228771228771228	0.03179649281088874
vB_EcoM_SHAK7704_OR594181	Lytic	0.5210789210789211	0.021598762012036112
vB-Eco-KMB47_OR424385	Lytic	0.5177822177822178	0.04933901993235517
vB-Eco-KMB46_OR525699	Lytic	0.515084915084915	0.0520846586228672
vB-Eco-KMB37_OR525700	Lytic	0.5148851148851149	0.03553186893573369
vB_EcoP-UTI89UKE2_MZ234049	Lytic	0.5124875124875125	0.03250363012912384

vB_EcoS_ACG-M12_JN986845	Lytic	0.5115884115884116	0.05592166313649559
myPSH1131_MG983840	Lytic	0.5071928071928072	0.04243591151396859
vB_EcoS-101114BS4_MZ234016	Lytic	0.5065934065934066	0.03412857038622432
Mimir124_PQ310662	Lytic	0.5058941058941059	0.061253019381096764
vB-Eco-KMB14_OR424384	Lytic	0.5052947052947052	0.04677800176280204
vB_EcoS-CHD2B1_MZ234026	Lytic	0.5028971028971029	0.041311318585690045
vB_EcoP-101114BS3_MZ234015	Lytic	0.5022977022977024	0.030575959290554382
vB_EcoS-101114B2_MZ234014	Lytic	0.5017982017982019	0.033718466647800725
NTEC3_OK539620	Lytic	0.500999000999001	0.043380690974404924
vB_EcoS-22664B1_MZ234008	Lytic	0.500899100899101	0.035775396045280135
vB_EcoM_LNA4_MZ311864	Temperate	0.5408591408591409	0.03875351375692474
vB_EcoM_SHAK7858_OR594183	Temperate	0.5226773226773227	0.01638361638361638
vB_EcoS_UTEC10_PP291582	Temperate	0.5221778221778222	0.050332303594103824
VB_EcoS-Golestan_MG099933	Temperate	0.5204795204795205	0.020565837442392577
vB-Eco-KMB35_OR525704	Temperate	0.5183816183816183	0.06047620688691993
vB_EcoS-CHD5UKE2_MZ234029	Temperate	0.5183816183816183	0.030799218621140163
vB_EcoS-UTI89UKE1_MZ234048	Temperate	0.5181818181818183	0.03479923584480584
vB_EcoP-101118B1_MZ234020	Temperate	0.5141858141858141	0.028656008883703066
vB_EcoP-CHD5UKE1_MZ234028	Temperate	0.5132867132867134	0.021481076910454693
ES_OR466728	Temperate	0.5127872127872128	0.04599322917954495
MLP1_OK148439	Temperate	0.5121878121878122	0.029846732723432533
vB_EcoS-phiEc4_MZ576218	Temperate	0.5114885114885115	0.047697459363083304
vB_EcoM_SHAK7693_OR594180	Temperate	0.5095904095904096	0.04673670053957779
St11Ph5_MG208881	Temperate	0.5087912087912089	0.03446205950427083
vB_EcoS-phiEc3_MZ576217	Temperate	0.5064935064935066	0.05187884423949967
vB-Eco-KMB36_OR525703	Temperate	0.5017982017982018	0.04651966978136441
590B_MW722821	Temperate	0.5016983016983018	0.040334905032335215
vB_EcoP-UTI89UKE3_MZ234050	Temperate	0.5013986013986014	0.016191522689992813
vB_EcoS-22664BS2_MZ234010	Temperate	0.5012987012987012	0.025838422633889652
vB_EcoP_ACG-C91_JN986844	Temperate	0.500999000999001	0.038583301685973556
vB-Eco-KMB41_OR539216	Temperate	0.5007992007992008	0.026085406316677637
vB_EcoP-101117UKE2_MZ234019	Temperate	0.5007992007992008	0.03137260803137165
MLP3_OK148440	Temperate	0.5004995004995004	0.04883878745886241

Development prediction provided by Bacphlip.

Phage name_ Acc. No	Virulent	Temperate
vB_EcoP-101114BS3_|_MZ234015	1	0
vB_EcoM-CHD94UKE2_|_MZ234031	1	0
MLP3_|_OK148440	1	0
EG1_|_MG488277	1	0
vB_EcoP-101101UKE1_|_MZ234012	1	0
Ec_Makalu_003_|_MN882349	1	0
vB_EcoM-101117BS1_|_MZ234018	1	0
vB_EcoP-101120B2_|_MZ234023	1	0
UGKSEcP1_|_OV877085	1	0
vB_EcoP-101136BS1_|_MZ234024	1	0
ZCEC14_|_OP168898	1	0
RDN8_1_|_MN414250	1	0
myPSH1131_|_MG983840	1	0
vB_EcoP-101120B1-2_|_MZ234022	1	0
vB-Eco-KMB22_|_OR539217	1	0
FS2B_|_MW929182	1	0
MLP1_|_OK148439	1	0
vB-Eco-KMB47_|_OR424385	1	0
UGKSEcP2_|_OV876900	1	0
Ec_Makalu_001_|_MN894885	1	0
vB_EcoM_LNA2_|_MZ311865	1	0
vB-Eco-KMB39_|_OR539219	1	0
vB_EcoP-101114UKE3_|_MZ234017	1	0
MLP2_|_OK148441	1	0
vB_EcoP-CHD5UKE1_|_MZ234028	1	0
vB_EcoP_LNA5_|_MZ311861	0,998656518	0,001343482
vB_EcoM_SHAK7854_|_OR594182	0,998656518	0,001343482
vB_EcoP-22664UKE3-2_|_MZ234011	0,998088282	0,001911718
vB-Eco-KMB25_|_OR424386	0,998044612	0,001955388
vB_EcoP-101117UKE2_|_MZ234019	0,997298624	0,002701376
vB_EcoP-22664BS1_|_MZ234009	0,997298624	0,002701376
vB_EcoM_SHAK9454_|_OR594184	0,997298624	0,002701376
vB_EcoP_ACG-C91_|_JN986844	0,997298624	0,002701376
vB_EcoP-101118B1_|_MZ234020	0,997298624	0,002701376
St11Ph5_|_MG208881	0,993351672	0,006648328
vB_EcoM_SHAK7704_|_OR594181	0,9875	0,0125
vB-Eco-KMB14_|_OR424384	0,9875	0,0125
vB-Eco-KMB23_|_OR525702	0,9875	0,0125
vB-Eco-KMB26_|_OR539218	0,985022026	0,014977974
vB_EcoM-G3G7_|_MZ234040	0,985022026	0,014977974
vB_EcoM_KMB26_|_OW536688	0,985022026	0,014977974
vB_EcoM-CHD16UKE1_|_MZ234030	0,985022026	0,014977974
vB_EcoM-CHD2BS1_|_MZ234027	0,985022026	0,014977974
vB-Eco-KMB42_|_OR539221	0,985022026	0,014977974
vB_EcoM-101112UKE3-1_|_MZ234013	0,975	0,025
Ec_Makalu_002_|_MN709127	0,975	0,025
Killian_|_OQ446694	0,975	0,025
vB_EcoP-101118UKE1_|_MZ234021	0,974136829	0,025863171

vB_EcoS-101114BS4_|_MZ234016	0,973802342	0,026197658
Mimir124_|_PQ310662	0,963505857	0,036494143
vB_EcoP-UTI89UKE3_|_MZ234050	0,9625	0,0375
vB_EcoS-CHD2B1_|_MZ234026	0,9625	0,0375
vB_EcoM_LNA1_|_MZ311863	0,9625	0,0375
vB_EcoP-UTI89UKE2_|_MZ234049	0,9625	0,0375
vB_EcoS-101114B2_|_MZ234014	0,961302342	0,038697658
vB_EcoS-22664B1_|_MZ234008	0,95	0,05
vB_EcoM-172859UKE1_|_MZ234025	0,95	0,05
vB_EcoS_ACG-M12_|_JN986845	0,949446177	0,050553823
vB_EcoP_LNA3_|_MZ311862	0,949210342	0,050789658
FS2B_|_MW929183	0,913270925	0,086729075
vB-Eco-KMB43_|_OR525698	0,9125	0,0875
vB-Eco-KMB40_|_OR539220	0,912005277	0,087994723
vB_EcoS-CHD5UKE2_|_MZ234029	0,9	0,1
vB_EcoM_LNA4_|_MZ311864	0,9	0,1
vB_EcoM_ACG-C40_|_JN986846	0,8875	0,1125
vB_EcoM_SHAK7163_|_OR594179	0,8875	0,1125
vB-Eco-KMB38_|_OR525701	0,8875	0,1125
vB_EcoS_UTEC10_|_PP291582	0,875	0,125
vB-Eco-KMB41_|_OR539216	0,8625	0,1375
vB_EcoS-phiEc4_|_MZ576218	0,8625	0,1375
vB_EcoS-phiEc3_|_MZ576217	0,8625	0,1375
ES_|_OR466728	0,85	0,15
vB-Eco-KMB46_|_OR525699	0,848044612	0,151955388
vB_EcoM_SHAK7858_|_OR594183	0,825	0,175
VB_EcoS-Golestan_|_MG099933	0,8	0,2
vB_EcoM_SHAK7693_|_OR594180	0,7875	0,2125
vB_EcoS-UTI89UKE1_|_MZ234048	0,75	0,25
vB_EcoS-22664BS2_|_MZ234010	0,7375	0,2625
vB-Eco-KMB36_|_OR525703	0,6875	0,3125
590B_|_MW722821	0,6875	0,3125
vB-Eco-KMB35_|_OR525704	0,65	0,35
NTEC3_|_OK539620	0,6	0,4
vB-Eco-KMB37_|_OR525700	0,5	0,5

Development prediction provided by Phatyp.

Phage name_ Acc No.	Phage Development Prediction	Score
vB-Eco-KMB38_|_OR525701	virulent	0,9998679
vB_EcoP_LNA3_|_MZ311862	virulent	0,99986744
vB-Eco-KMB37_|_OR525700	virulent	0,99986744
vB-Eco-KMB23_|_OR525702	virulent	0,99986744
vB-Eco-KMB22_|_OR539217	virulent	0,99986744
vB_EcoP_LNA5_|_MZ311861	virulent	0,99986696
vB_EcoP-101118UKE1_|_MZ234021	virulent	0,9998665
vB-Eco-KMB40_|_OR539220	virulent	0,9998665
vB_EcoS-22664B1_|_MZ234008	virulent	0,99986607
vB_EcoM-101112UKE3-1_|_MZ234013	virulent	0,99986607
vB_EcoP-101117UKE2_|_MZ234019	virulent	0,99986607
vB_EcoP-101118B1_|_MZ234020	virulent	0,99986607
vB_EcoP-101120B1-2_|_MZ234022	virulent	0,99986607
vB_EcoP-101120B2_|_MZ234023	virulent	0,99986607
vB_EcoP-101136BS1_|_MZ234024	virulent	0,99986607
vB-Eco-KMB26_|_OR539218	virulent	0,99986607
UGKSEcP2_|_OV876900	virulent	0,99986607
vB_EcoM_KMB26_|_OW536688	virulent	0,99986607
vB_EcoM_ACG-C40_|_JN986846	virulent	0,9998656
RDN8_1_|_MN414250	virulent	0,9998656
vB_EcoM-CHD16UKE1_|_MZ234030	virulent	0,9998656
vB_EcoM-CHD94UKE2_|_MZ234031	virulent	0,9998656
MLP2_|_OK148441	virulent	0,9998656
vB-Eco-KMB39_|_OR539219	virulent	0,9998656
vB_EcoM_SHAK7854_|_OR594182	virulent	0,9998656
vB_EcoP-22664UKE3-2_|_MZ234011	virulent	0,9998651
vB_EcoM-CHD2BS1_|_MZ234027	virulent	0,9998651
vB_EcoM-G3G7_|_MZ234040	virulent	0,9998651
vB_EcoM_LNA2_|_MZ311865	virulent	0,9998651
ZCEC14_|_OP168898	virulent	0,9998651
vB_EcoS-101114B2_|_MZ234014	virulent	0,9998645
vB_EcoS-phiEc3_|_MZ576217	virulent	0,9998645
Killian_|_OQ446694	virulent	0,9998645
vB-Eco-KMB46_|_OR525699	virulent	0,9998645
vB-Eco-KMB42_|_OR539221	virulent	0,9998645
UGKSEcP1_|_OV877085	virulent	0,9998645
vB_EcoS_UTEC10_|_PP291582	virulent	0,9998645
EG1_|_MG488277	virulent	0,99986404
vB_EcoS-22664BS2_|_MZ234010	virulent	0,99986404
vB_EcoS-CHD2B1_|_MZ234026	virulent	0,99986404
ES_|_OR466728	virulent	0,99986404
vB-Eco-KMB36_|_OR525703	virulent	0,99986404
vB_EcoM_SHAK7704_|_OR594181	virulent	0,99986404
vB_EcoM_SHAK9454_|_OR594184	virulent	0,99986404
vB_EcoS_ACG-M12_|_JN986845	virulent	0,9998636
VB_EcoS-Golestan_|_MG099933	virulent	0,9998636
Ec_Makalu_003_|_MN882349	virulent	0,9998636
vB_EcoP-22664BS1_|_MZ234009	virulent	0,9998636

vB_EcoM-101117BS1_|_MZ234018	virulent	0,9998636
vB_EcoM-172859UKE1_|_MZ234025	virulent	0,9998636
vB_EcoP_ACG-C91_|_JN986844	virulent	0,99986315
Ec_Makalu_001_|_MN894885	virulent	0,99986315
vB_EcoS-UTI89UKE1_|_MZ234048	virulent	0,99986315
vB_EcoP-UTI89UKE2_|_MZ234049	virulent	0,99986315
vB_EcoP-UTI89UKE3_|_MZ234050	virulent	0,99986315
Mimir124_|_PQ310662	virulent	0,99986315
Ec_Makalu_002_|_MN709127	virulent	0,9998627
vB_EcoS-phiEc4_|_MZ576218	virulent	0,9998627
NTEC3_|_OK539620	virulent	0,9998627
vB-Eco-KMB35_|_OR525704	virulent	0,9998627
vB_EcoM_SHAK7693_|_OR594180	virulent	0,9998627
vB_EcoP-101101UKE1_|_MZ234012	virulent	0,9998622
MLP1_|_OK148439	virulent	0,9998622
vB-Eco-KMB14_|_OR424384	virulent	0,9998622
vB-Eco-KMB43_|_OR525698	virulent	0,9998622
vB_EcoM_SHAK7858_|_OR594183	virulent	0,9998622
St11Ph5_|_MG208881	virulent	0,99986124
vB_EcoS-101114BS4_|_MZ234016	virulent	0,99986124
vB_EcoM_LNA1_|_MZ311863	virulent	0,99986124
590B_|_MW722821	virulent	0,9998608
vB-Eco-KMB47_|_OR424385	virulent	0,99985975
vB-Eco-KMB25_|_OR424386	virulent	0,99985975
vB_EcoM_SHAK7163_|_OR594179	virulent	0,99985975
vB_EcoS-CHD5UKE2_|_MZ234029	virulent	0,9998593
vB_EcoM_LNA4_|_MZ311864	virulent	0,9998593
vB-Eco-KMB41_|_OR539216	virulent	0,99983126
FS2B_|_MW929183	temperate	0,9998593
vB_EcoP-101114BS3_|_MZ234015	temperate	0,99985695
vB_EcoP-101114UKE3_|_MZ234017	temperate	0,99985695
vB_EcoP-CHD5UKE1_|_MZ234028	temperate	0,99985695
myPSH1131_|_MG983840	temperate	0,9998565
FS2B_|_MW929182	temperate	0,9998565
MLP3_|_OK148440	temperate	0,9998517

Supplementary Table S5

Names of the encoded proteins and positions of their genes in genomes of phages from the first clade of the VipTree-based multigene tree presented in Figure 1. These lists were used to generate tree presented in Figure 2. Genes important for tree topology were identified in this work.

The importance of the gene for tree topology marked by '+'	No. of parsimony informative sites in alignment	Characteristic protein features identified by InterProScan
+	801	pyocin_knob; Long-tail fiber proximal subunit, C-terminal, second; Long-tail fiber proximal subunit, C-terminal, trimerization domain
+	502	Baseplate wedge protein gp7, domain VI; Fibronectin type III; Baseplate wedge protein gp7.; Baseplate wedge protein gp7, helical domain; Baseplate wedge protein gp7, domain V; FN3
+	481	DNA polymerase family B, exonuclease domain; DNA-directed DNA polymerase.; DNA-directed DNA polymerase; DNA/RNA polymerases; Ribonuclease H-like; DNA polymerase family B; POLBc; DEDDy_DNA_polB_exo
+	348	Baseplate wedge protein gp10.; Baseplate wedge protein gp10 domain 3; gp9; Bacteriophage T4 gp9/10-like protein, N-terminal; Bacteriophage T4 gp9/10-like protein, C-terminal; Baseplate structural protein Gp10, C-terminal domain
+	325	Pre-baseplate central spike protein Gp5 [5].; Gp5 N-terminal OB domain; Gp5 C-terminal repeat (3 copies); Phage fibre proteins; gp5 N-terminal domain-like; T4-like_lys; Phage lysozyme; Lysozyme-like; Baseplate hub subunit and tail lysozyme
+	322	ATP dependent DNA ligase domain; Nucleic acid-binding proteins; DNA ligase/mRNA capping enzyme, catalytic domain
+	301	AAA domain; Rad50 coiled-coil Zn hook; P-loop containing nucleoside triphosphate hydrolases; ABC_ATPase; RecF/RecN/SMC N terminal domain
+	293	Baseplate structural protein gp6, C-terminal domain I; Baseplate wedge protein gp6-like, helical domain; Baseplate structural protein gp6, C-terminal domain II; Baseplate wedge protein gp6.; Baseplate structural protein gp6, C-terminal domain III; Baseplate wedge protein gp6, domain II
+	292	PFL-like glycyl radical enzymes; Ribonucleotide reductase, barrel domain; Ribonucleoside-diphosphate reductase; R1 subunit of ribonucleotide reductase, N-terminal domain; ATP cone domain; Ribonucleotide reductase, all-alpha domain
+	278	ATPase domain of HSP90 chaperone/DNA topoisomerase II/histidine kinase; Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase; DNA topoisomerase II large subunit; Ribosomal protein S5 domain 2-like; DNA gyrase B; Type II DNA topoisomerase; DNA topoisomerase 2
+	269	Tail sheath protein Gp18 domain III N-terminal region; Phage tail sheath protein subtilisin-like domain; Phage tail sheath protein beta-sandwich domain; Tail sheath protein Gp18-like domain I; Phage tail sheath C-terminal domain
+	249	Dda helicase SH3 domain; P-loop containing nucleoside triphosphate hydrolases; SF1_C_RecD; AAA domain; DEXSc_RecD-like; UvrD-like helicase C-terminal domain
+	245	Phage tail proteins; Baseplate structural protein, domain 2; Baseplate structural protein, domain 1
+	241	RNA ligase 1.; RNA ligase; T4 RNA ligase 1, C-terminal domain; T4 RNA ligase 1
+	241	Portal protein.; Bacteriophage T4-like portal protein (Gp20)
+	240	DEXHc_UvsW; Helicase conserved C-terminal domain; P-loop containing nucleoside triphosphate hydrolases; Type III restriction enzyme, res subunit; DNA helicase UvsW, N-terminal
+	240	RNR_III; Anaerobic ribonucleoside-triphosphate reductase; PFL-like glycyl radical enzymes; A2L zinc ribbon domain
+	231	Capsid vertex protein; Major capsid protein Gp23; Capsid vertex protein.
+	209	DnaB-like replicative helicase [41].; P-loop containing nucleoside triphosphate hydrolases; DnaB-like helicase C terminal domain; DNA primase/helicase, helicase subunit
+	190	P-loop containing nucleoside triphosphate hydrolases; recA bacterial DNA recombination protein; Bacteriophage T4, Recombination and repair protein, C-terminal domain
+	190	DNA primase.; DNA primase core
+	180	T4 RNase H, C terminal; PIN_T4-like; PIN domain-like; 5'-3' exonuclease, N-terminal resolvase-like domain; 5' to 3' exonuclease, C-terminal subdomain; H3TH_T4-like; PIN_53EXO-like
+	179	P-loop containing nucleoside triphosphate hydrolases; HAD-like; HAD_PNKP-C; Polynucleotide kinase; AAA domain; PITM 1-3, LNS2 domain
+	179	Metallo-dependent phosphatases; Calcineurin-like phosphoesterase
+	172	Ribonucleoside-diphosphate reductase 1 subunit beta; Ribonucleotide reductase, small chain; Ferritin-like; RNRR2
+	167	Major capsid protein.; Major capsid protein Gp23
+	166	Sliding-clamp-loader large subunit, AAA+ ATPase lid domain; ATPase family associated with various cellular activities (AAA); AAA; Sliding-clamp-loader large subunit, C-terminal domain; P-loop containing nucleoside triphosphate hydrolases; Sliding-clamp-loader large subunit.
+	159	T4 bacteriophage base plate protein
+	159	DNA clamp; Sliding clamp.; gp45 sliding clamp, C terminal; DNA polymerase processivity factor; Gp45 sliding clamp DNA polymerase
+	146	-
+	142	-
+	141	gp32 DNA binding protein like; Single-stranded DNA-binding protein [32].; Single-stranded DNA binding protein; Nucleic acid-binding proteins
+	139	T4-like virus Myoviridae tail sheath stabiliser; Tail sheath stabilizer and completion
+	136	-
+	127	3'-5' exoribonuclease Rv2179c-like domain; Ribonuclease H-like
+	126	Bacteriophage T4, Gp8; Baseplate structural protein gp8; Baseplate wedge protein gp8
+	126	Bacteriophage T holin; Holin.
+	124	-
+	123	Bacteriophage clamp loader A subunit; Sliding-clamp-loader small subunit.
+	107	Virus neck protein
+	104	all-alpha NTP pyrophosphatases; dUTPase
+	103	Transglycosylase SLT domain; Lysozyme-like
+	101	Loader of gp41 DNA helicase; gene 59 helicase assembly protein; T4 gene Gp59 loader of gp41 DNA helicase C-term; T4 gene Gp59 loader of gp41 DNA helicase; DNA helicase assembly protein [59].
+	95	Base plate wedge protein 53
+	91	P-loop containing nucleoside triphosphate hydrolases; Thymidine kinase [tdk].; Glucocorticoid receptor-like (DNA-binding domain); Thymidine kinase
+	91	Cytidine and deoxycytidylate deaminase zinc-binding region; Cytidine deaminase-like; deoxycytidylate_deaminase
+	86	Prohead core protein serine protease
+	82	Terminase DNA packaging enzyme; Phage terminase, small subunit
+	80	TnsA-like endonuclease N terminal; Head completion nuclease.
+	78	Recombination, repair and ssDNA binding protein UvsY
+	77	RNA polymerase sigma-like factor.
-	73	GIY-YIG endonuclease; GIY-YIG_EndoII_Hpy188I_like; GIY-YIG catalytic domain
-	70	-
-	70	T4-like virus tail tube protein gp19
-	57	Baseplate wedge protein gp25; gpW/gp25-like
-	56	Gene product 67
-	49	Bacteriophage translational regulator; Translational regulator protein regA
-	42	Transcriptional regulator DsbA
-	30	PAAR motif; PAAR_1; Putative phospholipase
-	29	UvsW.1 domain

vB_EcoM-G3G7	Killian	vB_EcoM-CHD16UKE1
long_tail_fiber 69173:73042	tail_protein 164142:168011	long_tail_fiber 69078:72947
baseplate_wedge_initiator 138499:141597	baseplate_wedge_subunit 95868:98966	baseplate_wedge_initiator 138395:141493
DNA_polymerase 23790:26486	DNA_polymerase 38949:41645	DNA_polymerase 23790:26486
hypothetical_protein 134766:136571	baseplate_wedge_subunit_and_tail_pin 100896:102701	hypothetical_protein 134663:136468
baseplate_hub_structural_protein_/lysozyme_R 144408:146135	glycoside_hydrolase_family_protein 91330:93057	baseplate_hub_structural_protein_/lysozyme_R 144304:146031
DNA_ligase 96805:98268	DNA_ligase 139931:141391	DNA_ligase 96704:98167
recombination-related_endonuclease 18517:20157	AAA_family_ATPase 45053:46693	recombination-related_endonuclease 18517:20157
baseplate_wedge_subunit 141594:143576	baseplate_wedge_subunit 93889:95871	baseplate_wedge_subunit 141490:143472
ribonucleotide_reductase_of_class_Ia_(aerobic),_alpha_subunit 81562:83826	ribonucleoside-diphosphate_reductase_subunit_alpha 154492:156756	ribonucleotide_reductase_of_class_Ia_(aerobic),_alpha_subunit 81470:83734
DNA_topoisomerase_large_subunit 49578:51395	ATP-binding_protein 16134:17951	DNA_topoisomerase_large_subunit 49579:51396
tail_sheath 124170:126149	tail_protein 111318:113297	tail_sheath 124067:126046
DNA_helicase 43470:44789	AAA_family_ATPase 22738:24057	DNA_helicase 43471:44790
baseplate_hub_subunit 107319:108491	baseplate_protein 129669:130844	baseplate_hub_subunit 107217:108389
RNA_ligase 85547:86671	RnlA_family_RNA_ligase 151061:152185	RNA_ligase 85455:86579
portal_protein 121904:123478	portal_protein 113989:115563	portal_protein 121801:123375
DNA_helicase 111527:113038	DNA_helicase 124433:125944	DNA_helicase 111424:112935
ribonucleotide_reductase_of_class_III_(anaerobic),_large_subunit 9007:10824	endolysin 54687:56504	ribonucleotide_reductase_of_class_III_(anaerobic),_large_subunit 9007:10824
capsid_vertex_protein 116811:118094	capsid_vertex_protein 119373:120656	capsid_vertex_protein 116708:117991
DNA_primase/helicase 32136:33563	AAA_family_ATPase 33094:34521	DNA_primase/helicase 32136:33563
recombination_protein 30614:31789	putative_RecA-like_recombination_protein 34868:36049	recombination_protein 30614:31789
DNA_primase_/_DNA_helicase 34955:35983	DNA_primase 30419:31447	DNA_primase_/_DNA_helicase 34955:35983
ribonuclease_H 73147:74064	endolysin 163120:164037	ribonuclease_H 73052:73969
3'-phosphatase,_5'-polynucleotide_kinase 88420:89325	3'-phosphatase_5'-polynucleotide_kinase 148404:149312	3'-phosphatase,_5'-polynucleotide_kinase 88328:89233
recombination-related_endonuclease 17012:18031	recombination-related_endonuclease 47179:48198	recombination-related_endonuclease 17012:18031
ribonucleotide_reductase_of_class_Ia_(aerobic),_beta_subunit 83878:85056	ribonucleotide_reductase_of_class_Ia_aerobic_beta_subunit 152676:153854	ribonucleotide_reductase_of_class_Ia_(aerobic),_beta_subunit 83786:84964
major_capsid_protein 118178:119743	major_capsid_protein 117724:119289	major_capsid_protein 118075:119640
replication_factor_C_small_subunit_/_DNA_polymerase_clamp_loader_subunit 21594:22553	putative_replication_factor 42659:43618	replication_factor_C_small_subunit_/_DNA_polymerase_clamp_loader_subunit 21594:22553
baseplate 108488:109240	baseplate_protein 128920:129672	baseplate 108386:109138
sliding_clamp_DNA_polymerase_accessory_protein 20856:21542	sliding_clamp_DNA_polymerase_accessory_protein 43668:44354	sliding_clamp_DNA_polymerase_accessory_protein 20856:21542
hypothetical_protein 113089:113769	inhibitor_of_prohead_protease 123702:124382	hypothetical_protein 112986:113666
prohead_assembly_protein 119762:120571	prohead_core_protein 116896:117705	prohead_assembly_protein 119659:120468
single_stranded_DNA-binding_protein 75997:76902	ssDNA_binding_protein 160868:161776	single_stranded_DNA-binding_protein 75902:76807
tail_sheath_stabilizer_and_completion_protein 128500:129318	tail_sheath_stabilizer_and_completion_protein 108149:108967	tail_sheath_stabilizer_and_completion_protein 128397:129215
DNA_end_protector_protein 147209:148033	DNA_end_protector_during_packaging 89432:90256	DNA_end_protector_protein 147105:147929
exonuclease 45523:46206	3'-5'_exoribonuclease 21321:22004	exonuclease 45524:46207
baseplate_wedge_subunit 137502:138506	baseplate_wedge_subunit 98959:99963	baseplate_wedge_subunit 137398:138402
holin 62691:63347	holin_protein 3826:4482	holin 62679:63335
neck_protein 130132:131061	neck_protein 106406:107335	neck_protein 130029:130958
DNA_polymerase_clamp_loader_subunit 22555:23118	DNA_polymerase_clamp_loader_subunit_A 42094:42657	DNA_polymerase_clamp_loader_subunit 22555:23118
neck_protein 129360:130130	head_closure 107337:108107	neck_protein 129257:130027
hypothetical_protein 37680:38198	dCTP_pyrophosphatase 28981:29499	hypothetical_protein 37680:38198
hypothetical_protein 160837:161382	transglycosylase_SLT_domain-containing_protein 73627:74172	hypothetical_protein 160731:161276
DNA_helicase_loader 74670:75308	loader_of_DNA_helicase_and_ssDNA_binding 161876:162529	DNA_helicase_loader 74575:75213
baseplate_wedge_subunit 146119:146709	baseplate_wedge_protein 90756:91346	baseplate_wedge_subunit 146015:146605
thymidine_kinase 162954:163535	thymidine_kinase 71481:72062	thymidine_kinase 162848:163429
dCMP_deaminase 91144:91725	dCMP_deaminase 146000:146581	dCMP_deaminase 91052:91633
prohead_assembly_protein 120602:121240	prohead_core_scaffolding_protein_and_protease 116227:116865	prohead_assembly_protein 120499:121137
terminase_small_subunit 127997:128491	small_terminase_protein 108976:109470	terminase_small_subunit 127894:128388
head_completion_protein 146757:147209	head_completion 90256:90708	head_completion_protein 146653:147105
single_stranded_DNA-binding_protein 110382:110795	recombination_mediator_protein 127365:127778	single_stranded_DNA-binding_protein 110279:110692
RNA_polymerase_sigma_factor_for_late_transcription 14140:14697	RNA_polymerase_sigma_factor 51115:51672	RNA_polymerase_sigma_factor_for_late_transcription 14140:14697
endonuclease 85084:85494	endonuclease 152238:152648	endonuclease 84992:85402
capsid_and_scaffold_protein 121240:121665	prohead_core_protein 115802:116227	capsid_and_scaffold_protein 121137:121562
tail_tube_protein 123562:124053	tail_protein 113414:113905	tail_tube_protein 123459:123950
baseplate_wedge_subunit 109917:110315	baseplate_wedge_subunit 127845:128243	baseplate_wedge_subunit 109814:110212
prohead_core_protein 121665:121904	putative_prohead_core_protein 115563:115802	prohead_core_protein 121562:121801
endoribonuclease_translational_repressor_of_early_genes 23120:23488	translational_repressor 41724:42092	endoribonuclease_translational_repressor_of_early_genes 23120:23488
double-stranded_DNA_binding_protein 74073:74342	double-stranded_DNA_binding_protein 162842:163111	double-stranded_DNA_binding_protein 73978:74247
phospholipase 143585:143878	PAAR_domain-containing_protein 93587:93880	phospholipase 143481:143774
DNA_helicase 111271:111501	DNA_helicase 126659:126889	DNA_helicase 111168:111398

UGKSEcP2	MLP2	vB-Eco-KMB42
tail_fiber_protein_proximal_subunit 153865:157734	long_tail_fiber_proximal_subunit 68564:72433	long_tail_fiber,_proximal_subunit 146932:150801
baseplate_wedge_subunit 86697:89795	baseplate_wedge_subunit 136481:139519	baseplate_wedge_initiator 79811:82909
DNA_polymerase 30738:33434	DNA_polymerase 24063:26759	DNA_polymerase 25303:27999
baseplate_wedge_subunit 91724:93529	baseplate_wedge_subunit_and_tail_pin 132749:134554	baseplate_wedge_subunit_and_tail_pin 84836:86641
baseplate_hub_subunit_and_tail_lysozyme 82159:83886	baseplate_hub_+_tail_lysozyme 142390:144117	baseplate_hub_structural_protein_/_lysozyme_R 75273:77000
ATP-dependent_DNA_ligase 130365:131825	DNA_ligase 95079:96542	DNA_ligase 123187:124650
SbcC-like_subunit_of_palindrome_specific_endonuclease 36844:38484	recombination-related_endonuclease 18748:20430	recombination-related_endonuclease 31632:33272
baseplate_wedge_subunit 84718:86700	baseplate_wedge_subunit 139576:141558	baseplate_wedge_subunit 77832:79814
Ribonucleoside-diphosphate_reductase_1_subunit_alpha 144338:146602	ribonucleotide_reductase_of_class_Ia_(aerobic),_alpha_subunit 79860:82124	ribonucleotide_reductase_of_class_Ia_(aerobic),_alpha_subunit 136998:139262
DNA_gyrase_subunit_B 8231:10048	DNA_gyrase_subunit_B 49112:50929	DNA_topoisomerase_large_subunit 2447:4264
tail_sheath 102146:104125	tail_sheath_monomer 122154:124133	tail_sheath 95258:97237
Dda-like_helicase 14835:16154	DNA_helicase 42777:44096	DNA_helicase 9282:10601
baseplate_hub 120103:121278	baseplate_hub_subunit 105590:106765	baseplate_hub_subunit 112924:114096
RNA_ligase_and_tail_fiber_protein_attachment_catalyst 141493:142617	RNA_ligase 83845:84969	RNA_ligase 134153:135277
portal_protein 104817:106391	portal_vertex_of_the_head 119888:121462	portal_vertex_of_the_head 97929:99503
DNA_helicase 115556:117067	DNA_helicase 109800:111311	DNA_helicase 108375:109886
Anaerobic_ribonucleoside-triphosphate_reductase 45862:47679	ribonucleotide_reductase_of_class_III_(anaerobic),_large_subunit 8410:10227	ribonucleotide_reductase_of_class_III_(anaerobic),_large_subunit 40959:42776
major_head_protein 110201:111484	capsid_vertex 114795:116078	capsid_vertex_protein 103319:104602
DnaB-like_replicative_helicase 24883:26310	DNA_primase/helicase 31798:33225	DNA_primase/helicase 19466:20893
recombinase 26657:27838	recombination_protein 30270:31451	recombination_protein 21240:22421
DNA_primase 22206:23234	DNA_primase_/_DNA_helicase 34617:35645	DNA_primase_/_DNA_helicase 16417:17445
RNaseH 152843:153760	ribonuclease_H 72538:73455	ribonuclease_H 145910:146827
polynucleotide_kinase 138835:139743	3'-phosphatase,_5'-polynucleotide_kinase 86719:87624	3'-phosphatase,_5'-polynucleotide_kinase 131498:132403
SbcD-like_subunit_of_palindrome_specific_endonuclease 38970:39989	recombination-related_endonuclease 16416:17435	recombination-related_endonuclease 33758:34777
Ribonucleoside-diphosphate_reductase_1_subunit_beta 143108:144286	ribonucleotide_reductase_of_class_Ia_(aerobic),_beta_subunit 82176:83354	ribonucleotide_reductase_of_class_Ia_(aerobic),_beta_subunit 135768:136946
major_head_protein 108552:110117	major_capsid_protein 116162:117727	major_capsid_protein 101670:103235
clamp_loader_of_DNA_polymerase 34448:35407	replication_factor_C_small_subunit_/_DNA_polymerase_clamp_loader_subunit 21867:22826	replication_factor_C_small_subunit_/_DNA_polymerase_clamp_loader_subunit 29236:30195
baseplate_hub 119354:120106	baseplate 106762:107514	baseplate 112175:112927
DNA_polymerase_processivity_factor 35459:36145	sliding_clamp_DNA_polymerase_accessory_protein 21129:21815	sliding_clamp_DNA_polymerase_accessory_protein 30247:30933
minor_head_protein_inhibitor_of_protease 114825:115505	inhibitor_of_prohead_protease 111362:112042	inhibitor_of_prohead_protease 107644:108324
head_scaffolding_protein 107724:108533	prohead_assembly_(scaffolding)_protein 117746:118555	prohead_assembly_(scaffolding)_protein 100842:101651
hypothetical_protein 150591:151499	single_stranded_DNA-binding_protein 74799:75707	single_stranded_DNA-binding_protein 143658:144566
tail_sheath_stabilizer 98977:99795	proximal_tail_sheath_stabilization_protein 126484:127302	tail_sheath_stabilizer_and_completion_protein 92089:92907
DNA_end_protector 80261:81085	DNA_end_protector_during_packaging 145191:146015	DNA_end_protector_protein 73375:74199
exonuclease 13418:14101	exonuclease_A 44830:45513	exonuclease 7866:8549
baseplate_wedge_subunit 89788:90792	baseplate_wedge_subunit 135484:136488	baseplate_wedge_subunit 82902:83906
holin 163376:164032	holin 62212:62868	holin 156823:157479
head-tail_adaptor_Ad2 97234:98163	head_completion,_neck_hetero-dimeric_protein 128116:129045	neck_protein 90346:91275
clamp_loader_of_DNA_polymerase 33883:34446	DNA_polymerase_clamp_loader_subunit 22828:23391	DNA_polymerase_clamp_loader_subunit 28671:29234
head_closure_Hc2 98165:98935	head_completion,_neck_hetero-dimeric_protein 127344:128114	neck_protein 91277:92047
nucleoside_triphosphate_pyrophosphohydrolase 21419:21937	dCTP_pyrophosphatase 37342:37860	dCTP_pyrophosphatase 15630:16148
hypothetical_protein 64432:64977	hypothetical_protein 160062:160607	hypothetical_protein 59080:59625
DNA_helicase_loader 151599:152252	DNA_helicase_loader 74046:74699	DNA_helicase_loader 144666:145319
baseplate_wedge_subunit 81585:82175	baseplate_wedge_subunit 144101:144691	baseplate_wedge_subunit 74699:75289
Thymidine_kinase 62286:62867	thymidine_kinase 162179:162760	thymidine_kinase 56934:57515
dCMP_deaminase 136434:137015	deoxycytidylate_deaminase 89444:90025	dCMP_deaminase 129616:130197
head_maturation_protease 107055:107693	prohead_assembly_(scaffolding)_protein 118586:119224	prohead_assembly_(scaffolding)_protein 100173:100811
terminase_small_subunit 99804:100298	terminase_small_subunit 125981:126475	terminase_small_subunit 92916:93410
head_closure 81085:81537	head_completion_protein 144739:145191	head_completion_protein 74199:74651
UvsY-like_recombination_mediator 117799:118212	single_stranded_DNA-binding_protein 108655:109068	single_stranded_DNA-binding_protein 110620:111033
late_sigma_transcription_factor 42305:42862	RNA_polymerase_sigma_factor_for_late_transcription 13543:14100	RNA_polymerase_sigma_factor_for_late_transcription 37093:37650
endonuclease 142670:143080	endonuclease 83382:83792	endonuclease 135330:135740
head_scaffolding_protein 106630:107055	capsid_and_scaffold_protein 119224:119649	capsid_and_scaffold 99748:100173
tail_tube 104242:104733	tail_fibers_protein 121546:122037	tail_fibers_protein 97354:97845
baseplate_wedge_subunit 118279:118677	baseplate_wedge_subunit 108190:108588	baseplate_wedge_subunit 111100:111498
prohead 106391:106630	prohead_core_protein 119649:119888	prohead_core_protein 99503:99748
translation_repressor 33513:33881	endoribonuclease_translational_repressor_of_early_genes 23393:23761	endoribonuclease_translational_repressor_of_early_genes 28301:28669
transcriptional_regulator 152565:152834	double-stranded_DNA_binding_protein 73464:73733	double-stranded_DNA_binding_protein,_late_transcriptional_regulation 145632:145901
PAAR_motif_of_membran_proteins 84416:84709	phospholipase 141567:141860	phospholipase 77530:77823
DNA_helicase 117093:117323	DNA_helicase 109544:109774	DNA_helicase 109912:110145

Ec_Makalu_003	vB-Eco-KMB38	vB-Eco-KMB43
long_tail_fiber,_proximal_subunit 143669:147409	long_tail_fiber 151123:154956	long_tail_fiber_proximal_subunit 144258:147998
baseplate_wedge_subunit 70016:73102	baseplate_wedge_initiator 85446:88544	baseplate_wedge_initiator 70264:73350
DNA_polymerase 15980:18658	DNA_polymerase 30469:33180	DNA_polymerase 16370:19048
baseplate_wedge_subunit_and_tail_pin 76540:78342	baseplate_wedge_subunit_and_tail_pin 90467:92281	baseplate_wedge_subunit_and_tail_pin 76789:78591
baseplate_hub_subunit_and_tail_lysozyme 64891:66690	baseplate_hub_structural_protein_/_lysozyme_R 80923:82656	baseplate_hub_+_tail_lysozyme 65699:67498
DNA_ligase 123584:125080	DNA_ligase 127646:129139	DNA_ligase 123875:125371
endonuclease_subunit 22447:24129	recombination-related_endonuclease 36637:38325	recombination-related_endonuclease 22837:24519
baseplate_wedge_subunit 68115:70019	baseplate_wedge_subunit 83479:85449	baseplate_wedge_subunit 68363:70267
ribonucleoside-diphosphate_reductase_subunit_alpha 134025:136268	ribonucleotide_reductase_of_class_Ia_(aerobic),_alpha_subunit 142334:144589	ribonucleotide_reductase_of_class_Ia_(aerobic),_alpha_subunit 135209:137452
DNA_topoisomerase_II_large_subunit 3217:5040	DNA_topoisomerase_large_subunit 3663:5480	DNA_topoisomerase_large_subunit 3205:5028
tail_sheath_protein 88146:90146	tail_sheath 100892:102874	tail_sheath_monomer 87402:89402
ATP-dependent_DNA_helicase 6584:7975	DNA_helicase 9656:10969	DNA_helicase 6395:7786
baseplate_central_spike_complex_protein 106888:107997	baseplate_hub_subunit 119428:120600	baseplate_hub_subunit 106243:107352
RNA_ligase_A 128881:130050	RNA_ligase 139484:140608	RNA_ligase 129173:130237
portal_vertex_protein 90773:92338	putative_portal_vertex_protein 103563:105134	portal_vertex_of_the_head 90029:91594
ATP-dependent_DNA_helicase 102546:104048	DNA_helicase 114859:116373	DNA_helicase 101846:103330
anaerobic_nucleoside_diphosphate_reductase 34202:36064	ribonucleotide_reductase_of_class_III_(anaerobic),_large_subunit 45272:47089	ribonucleotide_reductase_of_class_III_(anaerobic),_large_subunit 34555:36417
capsid_vertex_protein 96817:98058	capsid_vertex_protein 109286:110569	capsid_vertex 96176:97417
ATP-dependent_helicase 12866:14278	DNA_primase/helicase 20819:22261	DNA_primase/helicase 13256:14668
RecA-like_recombination_protein 14648:15718	RecA-like_recombination_protein 22605:23777	recombination_protein 15038:16108
DNA_primase 10210:11238	DNA_primase_/_DNA_helicase 16084:17106	DNA_primase_/_DNA_helicase 10023:11051
Ribonuclease_H 142461:143408	ribonuclease_H 150085:151002	ribonuclease_H 143048:143995
polynucleotide_kinase 127186:128064	3'-phosphatase,_5'-polynucleotide_kinase 136631:137530	3'-phosphatase,_5'-polynucleotide_kinase 127478:128356
endonuclease_subunit 25536:26561	recombination-related_endonuclease 38802:39821	recombination-related_endonuclease 25927:26952
aerobic_NDP_reductase,_small_subunit 132273:133433	ribonucleotide_reductase_of_class_Ia_(aerobic),_beta_subunit 141101:142279	ribonucleotide_reductase_of_class_Ia_(aerobic),_beta_subunit 132556:133719
major_capsid_protein 94591:96177	major_capsid_protein 107294:108862	major_capsid_protein 93841:95439
DNA_polymerase_accessory_protein 19683:20657	replication_factor_C_small_subunit_/_DNA_polymerase_clamp_loader_subunit 34198:35160	replication_factor_C_small_subunit_/_DNA_polymerase_clamp_loader_subunit 20073:21047
baseplate_hub_assembly_protein 106140:106910	baseplate_hub_assembly_protein 118679:119431	baseplate_assembly_protein 105495:106265
sliding_clmap_holder 20723:21409	sliding_clamp_DNA_polymerase_accessory_protein 35236:35922	sliding_clamp_DNA_polymerase_accessory_protein 21113:21799
inhibitor_of_prohead_protease 101759:102475	inhibitor_of_prohead_protease 113404:114072	protein_Inh 101050:101775
prohead_core_protein 93737:94531	prohead_assembly_(scaffolding)_protein 106464:107276	prohead_assembly_(scaffolding)_protein 92987:93781
single-stranded_DNA_binding_protein 139605:140573	single_stranded_DNA-binding_protein 147823:148725	single_stranded_DNA-binding_protein 140137:141105
tail_sheath_stabilizer_and_completion_protein 83954:84787	proximal_tail_sheath_stabilization_protein 97717:98538	proximal_tail_sheath_stabilization_protein 84203:85036
DNA_end_protector_protein 62963:63784	DNA_end_protector_protein 79013:79834	DNA_end_protector_during_packaging 63768:64589
exonuclease 5315:5983	exonuclease 8969:9646	exonuclease_A 5303:5971
baseplate_wedge_subunit 73095:74090	baseplate_wedge_subunit 88537:89541	baseplate_wedge_subunit 73343:74338
holin_lysis_mediator 153602:154258	holin 160976:161635	holin 154886:155542
neck_protein 82216:83148	neck_protein 95972:96898	head_completion,_neck_hetero-dimeric_protein 82467:83399
clamp_loader_small_subunit 19105:19683	DNA_polymerase_clamp_loader_subunit 33633:34196	DNA_polymerase_clamp_loader_subunit 19495:20073
neck_protein 83185:83925	neck_protein 96900:97664	head_completion,_neck_hetero-dimeric_protein 83436:84176
dCTP_pyrophosphatase 9677:10195	dCTP_pyrophosphatase 15280:15801	dCTP_pyrophosphatase 9490:10008
putative_20.7_kDa_protein 50797:51387	hypothetical_protein 61141:61683	hypothetical_protein 50770:51360
putative_DNA_helicase_loader_protein 140634:141314	DNA_helicase_loader 148842:149495	DNA_helicase_loader 141221:141901
baseplate_wedge_subunit 64325:64879	baseplate_wedge_subunit 80348:80923	baseplate_wedge_subunit 65133:65687
thymidine_kinase 48678:49274	thymidine_kinase 59155:59736	thymidine_kinase 48651:49247
deoxycytidylate_deaminase 126455:126961	dCMP_deaminase 133962:134543	deoxycytidylate_deaminase 126747:127253
prohead_core_scaffolding_protein_and_protease 93010:93705	prohead_assembly_(scaffolding)_protein 105789:106430	prohead_assembly_(scaffolding)_protein 92260:92955
small_terminase_protein 84801:85298	terminase_small_subunit 98548:99042	terminase_small_subunit 85050:85547
head_completion_protein 63796:64269	head_completion_protein 79837:80286	head_completion_protein 64601:65077
recombination,_repair_and_ssDNA_binding_protein 104590:104997	single_stranded_DNA-binding_protein 117113:117607	single_stranded_DNA-binding_protein 103945:104352
RNA_polymerase_sigma_factor 27158:27691	RNA_polymerase_sigma_factor_for_late_transcription 40814:41371	RNA_polymerase_sigma_factor_for_late_transcription 27550:28083
endonuclease_II 129986:130447	endonuclease_II 140664:141074	endonuclease 130278:130739
prohead_assembly_protein 92591:92998	prohead_core_protein 105364:105789	prohead_core_protein 91841:92248
tail_tube_protein 90220:90714	tail_inner_tube 102987:103478	tail_fibers_protein 89476:89970
baseplate_wedge_subunit 105045:105431	baseplate_wedge_subunit 117607:118005	baseplate_wedge_subunit 104400:104786
putative_prohead_core_protein_I 92339:92578	prohead_core_protein 105134:105364	prohead_core_protein 91595:91828
translation_repressor_protein 18741:19103	endoribonuclease_translational_repressor_of_early_genes 33262:33630	endoribonuclease_translational_repressor_of_early_genes 19131:19493
double-stranded_DNA-binding_protein 141571:141846	double-stranded_DNA_binding_protein 149808:150077	double-stranded_DNA_binding_protein 142158:142433
hypothetical_protein 67825:68118	putative_phospholipase 83177:83470	phospholipase 68073:68366
helicase 104092:104328	DNA_helicase 116399:116632	DNA_helicase 103447:103683

vB_EcoM-101117BS1	vB-Eco-KMB22	UGKSEcP1
long_tail_fiber 67076:70945	long_tail_fiber_proximal_subunit 148454:152323	tail_fiber_protein_proximal_subunit 133540:137409
baseplate_wedge_initiator 134813:137911	baseplate_wedge_initiator 80870:83968	baseplate_wedge_subunit 34373:37471
DNA_polymerase 23863:26559	DNA_polymerase 25527:28223	DNA_polymerase 90740:93436
hypothetical_protein 131078:132886	baseplate_wedge_subunit_and_tail_pin 85895:87703	baseplate_wedge_subunit 30640:32445
baseplate_hub_structural_protein_/lysozyme_R 140722:142449	baseplate_hub_structural_protein_/_lysozyme_R 76332:78059	baseplate_hub_subunit_and_tail_lysozyme 40282:42009
DNA_ligase 93041:94504	DNA_ligase 123978:125441	ATP-dependent_DNA_ligase 159653:161116
recombination-related_endonuclease 18564:20204	recombination-related_endonuclease 31882:33522	SbcC-like_subunit_of_palindrome_specific_endonuclease 85690:87330
baseplate_wedge_subunit 137908:139890	baseplate_wedge_subunit 78891:80873	baseplate_wedge_subunit 37468:39450
ribonucleotide_reductase_of_class_Ia_(aerobic),_alpha_subunit 77993:80257	ribonucleotide_reductase_of_class_Ia_(aerobic),_alpha_subunit 138245:140509	Ribonucleoside-diphosphate_reductase_1_subunit_alpha 144672:146936
DNA_topoisomerase_large_subunit 47805:49622	DNA_topoisomerase_large_subunit 2458:4275	DNA_gyrase_subunit_B 113908:115725
tail_sheath 120456:122435	tail_sheath 96346:98325	tail_sheath 20044:22023
DNA_helicase 41507:42826	DNA_helicase 9260:10579	Dda-like_helicase 108035:109354
baseplate_hub_subunit 103594:104769	baseplate_hub_subunit 113716:114891	baseplate_hub 2897:4072
RNA_ligase 81978:83102	RNA_ligase 135400:136524	RNA_ligase_and_tail_fiber_protein_attachment_catalyst 148657:149781
portal_protein 118190:119764	portal_vertex_of_the_head 99017:100591	portal_protein 17778:19352
DNA_helicase 107805:109316	DNA_helicase 109169:110680	DNA_helicase 7108:8619
ribonucleotide_reductase_of_class_III_(anaerobic),_large_subunit 9053:10870	ribonucleotide_reductase_of_class_III_(anaerobic),_large_subunit 41514:43331	Anaerobic_ribonucleoside-triphosphate_reductase 76486:78303
capsid_vertex_protein 113088:114371	capsid_vertex_protein 104401:105684	major_head_protein 12685:13968
DNA_primase/helicase 30965:32392	DNA_primase/helicase 19690:21117	DnaB-like_replicative_helicase 97865:99292
recombination_protein 29437:30618	recombination_protein 21464:22645	recombinase 96337:97518
DNA_primase_/_DNA_helicase 34407:35435	DNA_primase_/_DNA_helicase 16647:17675	DNA_primase 100939:101967
ribonuclease_H 71050:71967	ribonuclease_H 147432:148349	RNaseH 137514:138431
3'-phosphatase,_5'-polynucleotide_kinase 84852:85760	3'-phosphatase,_5'-polynucleotide_kinase 132746:133654	polynucleotide_kinase 151531:152439
recombination-related_endonuclease 17059:18078	recombination-related_endonuclease 34008:35027	SbcD-like_subunit_of_palindrome_specific_endonuclease 84185:85204
ribonucleotide_reductase_of_class_Ia_(aerobic),_beta_subunit 80309:81487	ribonucleotide_reductase_of_class_Ia_(aerobic),_beta_subunit 137015:138193	Ribonucleoside-diphosphate_reductase_1_subunit_beta 146988:148166
major_capsid_protein 114455:116020	major_capsid_protein 102752:104317	major_head_protein 14052:15617
replication_factor_C_small_subunit_/_DNA_polymerase_clamp_loader_subunit 21641:22600	replication_factor_C_small_subunit_/_DNA_polymerase_clamp_loader_subunit 29486:30445	clamp_loader_of_DNA_polymerase 88767:89726
baseplate 104769:105518	baseplate_hub_chaperone 112967:113716	baseplate_hub 4072:4821
sliding_clamp_DNA_polymerase_accessory_protein 20903:21589	sliding_clamp_DNA_polymerase_accessory_protein 30497:31183	DNA_polymerase_processivity_factor 88029:88715
hypothetical_protein 109367:110047	inhibitor_of_prohead_protease 108438:109118	minor_head_protein_inhibitor_of_protease 8670:9350
prohead_assembly_protein 116039:116848	prohead_core_protein 101924:102733	head_scaffolding_protein 15636:16445
single_stranded_DNA-binding_protein 73310:74218	single_stranded_DNA-binding_protein 144599:145504	hypothetical_protein 139775:140683
tail_sheath_stabilizer_and_completion_protein 124786:125604	tail_sheath_stabilizer_and_completion_protein 93177:93995	tail_sheath_stabilizer 24374:25192
DNA_end_protector_protein 143523:144347	DNA_end_protector_protein 74434:75258	DNA_end_protector 43083:43907
exonuclease 43560:44243	exonuclease 7843:8526	exonuclease 109856:110539
baseplate_wedge_subunit 133816:134820	baseplate_wedge_subunit 83961:84965	baseplate_wedge_subunit 33376:34380
holin 60684:61340	holin 158059:158715	holin 127242:127898
neck_protein 126418:127347	neck_protein 91434:92363	head-tail_adaptor_Ad2 26006:26935
DNA_polymerase_clamp_loader_subunit 22602:23165	DNA_polymerase_clamp_loader_subunit 28921:29484	clamp_loader_of_DNA_polymerase 89728:90291
neck_protein 125646:126416	neck_protein 92365:93135	head_closure_Hc2 25234:26004
dCTP_pyrophosphatase 35704:36222	dCTP_pyrophosphatase 15860:16378	nucleoside_triphosphate_pyrophosphohydrolase 102236:102754
hypothetical_protein 159071:159616	hypothetical_protein 60347:60892	hypothetical_protein 59187:59732
DNA_helicase_loader 72573:73211	DNA_helicase_loader 146188:146841	DNA_helicase_loader 139022:139675
baseplate_wedge_subunit 142433:143023	baseplate_wedge_subunit 75758:76348	baseplate_wedge_subunit 41993:42583
thymidine_kinase 161175:161756	thymidine_kinase 58017:58598	Thymidine_kinase 61297:61878
dCMP_deaminase 87579:88160	dCMP_deaminase 130345:130926	dCMP_deaminase 154259:154840
prohead_assembly_protein 116879:117517	prohead_core_scaffolding_protein_and_protease 101255:101893	head_maturation_protease 16476:17114
terminase_small_subunit 124283:124777	terminase_small_subunit 94004:94498	terminase_small_subunit 23871:24365
head_completion_protein 143071:143523	head_completion_protein 75258:75710	head_closure 42631:43083
single_stranded_DNA-binding_protein 106660:107073	recombination,_repair_and_ssDNA_binding_protein 111412:111825	UvsY-like_recombination_mediator 5963:6376
RNA_polymerase_sigma_factor_for_late_transcription 14186:14743	RNA_polymerase_sigma_factor_for_late_transcription 37943:38500	late_sigma_transcription_factor 81312:81869
endonuclease 81515:81925	endonuclease_II 136577:136987	endonuclease 148194:148604
capsid_and_scaffold_protein 117517:117942	capsid_and_scaffold 100830:101255	head_scaffolding_protein 17114:17539
tail_tube_protein 119848:120339	tail_fibers_protein 98442:98933	tail_tube 19436:19927
baseplate_wedge_subunit 106195:106593	baseplate_wedge_subunit 111892:112290	baseplate_wedge_subunit 5498:5896
prohead_core_protein 117942:118190	prohead_core_protein 100591:100830	prohead 17539:17778
endoribonuclease_translational_repressor_of_early_genes 23167:23535	endoribonuclease_translational_repressor_of_early_genes 28551:28919	translation_repressor 90293:90661
double-stranded_DNA_binding_protein 71976:72245	double-stranded_DNA_binding_protein,_late_transcriptional_regulation 147154:147423	transcriptional_regulator 138440:138709
phospholipase 139899:140192	phospholipase 78589:78882	PAAR_motif_of_membran_proteins 39459:39752
DNA_helicase 107549:107779	DNA_helicase 110706:110936	DNA_helicase 6852:7082

The name of gene product and its position in genome of particular phage:		
vB_EcoM-172859UKE1	**vB-Eco-KMB23**	**vB_EcoM_KMB26**
long_tail_fiber 68994:72869	long_tail_fiber_proximal_subunit 148185:152054	Phage_long_tail_fiber_proximal_subunit 149383:153252
baseplate_wedge_initiator 134850:137948	baseplate_wedge_initiator 80985:84083	Phage_baseplate_wedge_subunit 80401:83439
hypothetical_protein 22123:24834	DNA_polymerase 25442:28138	DNA_polymerase_(EC,_phage-associated 27054:29750
hypothetical_protein 131103:132908	baseplate_wedge_subunit_and_tail_pin 86010:87818	Phage_baseplate_wedge_subunit_and_tail_pin_(T4-like_gp10) 85366:87171
baseplate_hub_structural_protein_/lysozyme_R 140742:142475	baseplate_hub_structural_protein_/_lysozyme_R 76447:78174	T4-like_phage_baseplate_hub_+_tail_lysozyme 75809:77530
hypothetical_protein 94660:96153	DNA_ligase 124527:125990	DNA_ligase,_phage-associated 123968:125431
recombination-related_endonuclease 17020:18666	recombination-related_endonuclease 31797:33479	Phage_recombination-related_endonuclease_Gp46 33383:35023
baseplate_wedge_subunit 137945:139975	baseplate_wedge_subunit 79006:80988	Phage_baseplate_wedge_subunit_(T4-like_gp6) 78362:80344
ribonucleotide_reductase_of_class_Ia_(aerobic),_alpha_subunit 79003:81258	ribonucleotide_reductase_of_class_Ia_(aerobic),_alpha_subunit 138300:140564	Ribonucleotide_reductase_of_class_Ia_(aerobic),_alpha_subunit_(EC 138239:140503
DNA_topoisomerase_large_subunit 49586:51403	DNA_topoisomerase_large_subunit 2458:4275	DNA_gyrase_subunit_B_(EC 2447:4264
tail_sheath 120481:122463	tail_sheath 96462:97823	Phage_tail_sheath_monomer 95788:97767
DNA_helicase 44127:45440	DNA_helicase 9182:10501	DNA_helicase_(EC,_phage-associated 9047:10366
baseplate_hub_subunit 103202:104374	baseplate_hub_subunit 114271:115440	Phage_baseplate_hub_subunit 113745:114917
RNA_ligase 82981:84105	RNA_ligase 135464:136588	RNA_ligase,_phage-associated_#T4-like_RnlA_#T4_GC1653 135394:136518
putative_portal_vertex_protein 118221:119792	portal_(connector)_protein 99260:100834	Phage_portal_vertex_of_the_head 98459:100033
DNA_helicase 107430:108944	DNA_helicase 109718:111229	DNA_helicase,_phage-associated 109198:110709
ribonucleotide_reductase_of_class_III_(anaerobic),_large_subunit 7602:9419	ribonucleotide_reductase_of_class_III_(anaerobic),_large_subunit 41694:43511	Ribonucleotide_reductase_of_class_III_(anaerobic),_large_subunit_(EC 42416:44233
capsid_vertex_protein 112783:114063	capsid_vertex_protein 104662:105945	Phage_capsid_vertex 103843:105126
DNA_primase/helicase 33107:34483	DNA_primase/helicase 19605:21032	DNA_primase/helicase,_phage-associated 19931:21358
recombination_protein 31525:32697	recombination_protein 21379:22560	Phage_recombination_protein 21705:22880
DNA_primase_/_DNA_helicase 37989:39011	DNA_primase_/_DNA_helicase 16570:17598	DNA_primase_(EC_/_DNA_helicase_(EC,_phage-associated 17510:18538
ribonuclease_H 72973:73890	ribonuclease_H 147163:148080	Phage_ribonuclease_H_(EC 148361:149278
3'-phosphatase,_5'-polynucleotide_kinase 86182:86958	3'-phosphatase,_5'-polynucleotide_kinase 132824:133732	3'-phosphatase,_5'-polynucleotide_kinase,_phage-associated 132740:133645
recombination-related_endonuclease 14871:15890	recombination-related_endonuclease 34793:35812	Phage_recombination-related_endonuclease_Gp47 35509:36528
ribonucleotide_reductase_of_class_Ia_(aerobic),_beta_subunit 81314:82489	ribonucleotide_reductase_of_class_Ia_(aerobic),_beta_subunit 137079:138248	Ribonucleotide_reductase_of_class_Ia_(aerobic),_beta_subunit_(EC 137009:138187
major_capsid_protein 114487:116055	major_capsid_protein 103013:104578	Phage_major_capsid_protein 102194:103759
replication_factor_C_small_subunit_/_DNA_polymerase_clamp_loader_subunit 20143:21105	replication_factor_C_small_subunit_/_DNA_polymerase_clamp_loader_subunit 29401:30360	Replication_factor_C_small_subunit_/_Phage_DNA_polymerase_clamp_loader_subunit 30987:31946
baseplate 104371:105123	baseplate_hub_assembly_protein 113516:114265	Phage_baseplate 112996:113748
sliding_clamp_DNA_polymerase_accessory_protein 19381:20067	sliding_clamp_DNA_polymerase_accessory_protein 30412:31098	Sliding_clamp_DNA_polymerase_accessory_protein,_phage_associated 31998:32684
inhibitor_of_prohead_protease 108995:109663	inhibitor_of_prohead_protease 108987:109667	Inh_inhibitor_of_gp21_prohead_protease 108467:109147
prohead_assembly_protein 116073:116885	prohead_assembly_(scaffolding)_protein 102185:102994	Phage_prohead_assembly_(scaffolding)_protein 101366:102175
single_stranded_DNA-binding_protein 75252:76154	single_stranded_DNA-binding_protein 144911:145819	Single_stranded_DNA-binding_protein,_phage-associated 145518:146423
tail_completion_protein 124817:125638	tail_sheath_stabilizer_and_completion_protein 93293:94111	Proximal_tail_sheath_stabilization_protein 92667:93437
DNA_end_protector_protein 143564:144385	DNA_end_protector_protein 74549:75373	Phage_DNA_end_protector_during_packaging 73905:74729
exonuclease 45450:46127	exonuclease 7765:8448	T4-like_phage_DexA_exonuclease_A 7630:8313
baseplate_wedge_subunit 133853:134857	baseplate_wedge_subunit 84076:85080	Phage_baseplate_wedge_subunit_(T4-like_gp8) 83432:84436
holin 62772:63431	holin 157793:158449	Phage_holin 158944:159600
neck_protein 126457:127383	neck_protein 91550:92479	T4-like_phage_head_completion,_neck_hetero-dimeric_protein_(T4-like_gp13) 90876:91805
DNA_polymerase_clamp_loader_subunit 21107:21670	DNA_polymerase_clamp_loader_subunit 28836:29399	Phage_DNA_polymerase_clamp_loader_subunit_Gp62 30422:30985
neck_protein 125691:126455	neck_protein 92481:93251	T4-like_phage_head_completion,_neck_hetero-dimeric_protein_(T4-like_gp14) 91807:92577
dCTP_pyrophosphatase 39294:39815	dCTP_pyrophosphatase 15783:16301	dCTP_pyrophosphatase_(EC,_phage-associated 15295:15813
hypothetical_protein 162310:162852	hypothetical_protein 59610:60155	Phage_protein 60507:61052
DNA_helicase_loader 74480:75133	DNA_helicase_loader 145919:146557	Phage_DNA_helicase_loader 147117:147755
baseplate_wedge_subunit 142475:142978	baseplate_wedge_subunit 75873:76463	Phage_baseplate_wedge_subunit 75229:75819
thymidine_kinase 164257:164838	thymidine_kinase 57280:57861	Thymidine_kinase_(EC 58354:58935
dCMP_deaminase 89047:89628	dCMP_deaminase 130944:131525	dCMP_deaminase_(EC 130340:130921
prohead_assembly_protein 116919:117560	prohead_assembly_(scaffolding)_protein 101516:102154	Phage_prohead_assembly_(scaffolding)_protein 100697:101335
terminase_small_subunit 124313:124807	terminase_small_subunit 94120:94614	Phage_terminase,_small_subunit 93446:93940
head_completion_protein 142968:143561	head_completion_protein 75373:75825	Phage_head_completion_protein 74729:75181
single_stranded_DNA-binding_protein 106195:106689	single_stranded_DNA-binding_protein 111961:112374	Single_stranded_DNA-binding_protein,_phage-associated 111441:111854
RNA_polymerase_sigma_factor_for_late_transcription 13322:13879	RNA_polymerase_sigma_factor_for_late_transcription 38128:38685	T4-like_phage_RNA_polymerase_sigma_factor_for_late_transcription 38844:39401
endonuclease 82516:82926	endonuclease 136641:137051	Phage_endonuclease 136571:136981
capsid_and_scaffold_protein 117560:117985	capsid_and_scaffold 101091:101516	Phage_capsid_and_scaffold 100272:100697
tail_tube 119877:120368	tail_tube_protein 98685:99176	Phage_tail_fibers 97884:98375
baseplate_wedge_subunit 105797:106195	baseplate_wedge_subunit 112441:112839	Phage_baseplate_wedge_subunit_(T4-like_gp25) 111921:112319
prohead_core_protein 117985:118221	prohead_core_protein 100834:101091	Phage_prohead_core_protein 100033:100272
endoribonuclease_translational_repressor_of_early_genes 21664:22041	endoribonuclease_translational_repressor_of_early_genes 28466:28834	Phage_endoribonulcease_translational_repressor_of_early_genes,_regA 30052:30420
double-stranded_DNA_binding_protein 73898:74167	double-stranded_DNA_binding_protein 146885:147154	Phage_double-stranded_DNA_binding_protein_#T4-like_dsbA,_late_transcriptional_regulation_#T4_GC1668 148083:148352
putative_phospholipase 139927:140220	phospholipase 78704:78997	Phospholipase 78060:78353
DNA_helicase 107171:107404	DNA_helicase 111255:111485	DNA_helicase,_phage-associated 110735:110965

Ec_Makalu_002	vB-Eco-KMB26	vB_EcoM-101112UKE3-1
long_tail_fiber,_proximal_subunit 145552:149292	long_tail_fiber 149383:153252	long_tail_fiber 68407:72276
baseplate_wedge_protein 70377:73463	baseplate_wedge_initiator 80341:83439	baseplate_wedge_initiator 138121:141219
DNA_polymerase 15982:18660	DNA_polymerase 27054:29750	DNA_polymerase 24658:27354
baseplate_wedge_subunit_and_tail_pin 76901:78703	hypothetical_protein 85366:87171	hypothetical_protein 134388:136193
baseplate_hub_subunit_and_tail_lysozyme 65812:67611	baseplate_hub_structural_protein_/_lysozyme_R 75803:77530	baseplate_hub_structural_protein_/lysozyme_R 144030:145757
DNA_ligase 124891:126387	DNA_ligase 123968:125431	DNA_ligase 95267:96730
endonuclease_subunit 22448:24130	recombination-related_endonuclease 33383:35023	recombination-related_endonuclease 19343:21025
baseplate_wedge_protein 68476:70380	baseplate_wedge_subunit 78362:80344	baseplate_wedge_subunit 141216:143198
ribonucleoside-diphosphate_reductase_subunit_alpha 135323:137566	ribonucleotide_reductase_of_class_Ia_(aerobic),_alpha_subunit 138239:140503	ribonucleotide_reductase_of_class_Ia_(aerobic),_alpha_subunit 79513:81777
DNA_topoisomerase_II_large_subunit 3217:5040	DNA_topoisomerase_large_subunit 2447:4264	DNA_topoisomerase_large_subunit 49137:50954
tail_sheath_protein 89374:91374	tail_sheath 95788:97767	tail_sheath 122977:124956
DNA_helicase 6584:7975	DNA_helicase 9047:10366	DNA_helicase 43029:44348
baseplate_central_spike_complex_protein 108206:109315	baseplate_hub_subunit 113745:114917	baseplate_hub_subunit 105781:106953
RNA_ligase_A 130188:131357	RNA_ligase 135394:136518	RNA_ligase 84177:85301
portal_vertex_protein 92001:93566	portal_vertex_of_the_head 98459:100033	portal_protein 120071:121645
ATP-dependent_DNA_helicase 103865:105367	DNA_helicase 109198:110709	DNA_helicase 109989:111500
putative_anaerobic_ribonucleoside-triphosphate_reductase 34206:36068	ribonucleotide_reductase_of_class_III_(anaerobic),_large_subunit 42416:44233	ribonucleotide_reductase_of_class_III_(anaerobic),_large_subunit 9006:10823
capsid_vertex_protein 98128:99369	capsid_vertex_protein 103843:105126	capsid_vertex_protein 114990:116273
ATP-dependent_helicase 12868:14280	DNA_primase/helicase 19931:21358	DNA_primase/helicase 31764:33191
RecA-like_recombination_protein 14650:15720	recombination_protein 21705:22880	recombination_protein 30236:31417
DNA_primase 10212:11240	DNA_primase_/_DNA_helicase 17510:18538	DNA_primase_/_DNA_helicase 34855:35883
ribonuclease 144342:145289	ribonuclease_H 148361:149278	ribonuclease_H 72381:73298
polynucleotide_kinase 128493:129371	3'-phosphatase,_5'-polynucleotide_kinase 132740:133645	3'-phosphatase,_5'-polynucleotide_kinase 87050:87955
endonuclease_subunit 25537:26562	recombination-related_endonuclease 35509:36528	recombination-related_endonuclease 17010:18029
aerobic_NDP_reductase,_small_subunit 133571:134731	ribonucleotide_reductase_of_class_Ia_(aerobic),_beta_subunit 137009:138187	ribonucleotide_reductase_of_class_Ia_(aerobic),_beta_subunit 82508:83686
major_capsid_protein 95819:97405	major_capsid_protein 102194:103759	major_capsid_protein 116357:117922
clamp_loader,_small_subunit 19685:20659	replication_factor_C_small_subunit_/_DNA_polymerase_clamp_loader_subunit 30987:31946	replication_factor_C_small_subunit_/_DNA_polymerase_clamp_loader_subunit 22462:23421
baseplate_hub_assembly_protein 107458:108228	baseplate 112996:113748	baseplate 106950:107702
sliding_clamp 20725:21411	sliding_clamp_DNA_polymerase_accessory_protein 31998:32684	sliding_clamp_DNA_polymerase_accessory_protein 21724:22410
inhibitor_of_prohead_protease 103069:103794	hypothetical_protein 108467:109147	hypothetical_protein 111551:112231
prohead_core_scaffold_protein 94965:95759	prohead_assembly_(scaffolding)_protein 101366:102175	prohead_assembly_protein 117941:118750
single-stranded_DNA_binding_protein 140903:141877	single_stranded_DNA-binding_protein 145518:146423	single_stranded_DNA-binding_protein 74642:75550
tail_sheath_stabilizer_and_completion_protein 85182:86015	tail_sheath_stabilizer_and_completion_protein 92619:93437	tail_sheath_stabilizer_and_completion_protein 127307:128125
DNA_end_protector_protein 62758:63579	DNA_end_protector_protein 73905:74729	DNA_end_protector_protein 146831:147658
exonuclease 5315:5983	exonuclease 7630:8313	exonuclease 45082:45765
baseplate_wedge_protein 73456:74451	baseplate_wedge_subunit 83432:84436	baseplate_wedge_subunit 137124:138128
holin_lysis_mediator 155311:155967	holin 158944:159600	holin 62078:62734
neck_protein 82577:83509	neck_protein 90876:91805	neck_protein 129748:130677
Sliding-clamp-loader_subunit 19107:19685	DNA_polymerase_clamp_loader_subunit 30422:30985	DNA_polymerase_clamp_loader_subunit 23423:23986
neck_protein 84415:85155	neck_protein 91807:92577	neck_protein 128167:128937
dCTP_pyrophosphatase 9679:10197	hypothetical_protein 15295:15813	hypothetical_protein 37580:38098
hypothetical_protein 50587:51177	hypothetical_protein 60507:61052	hypothetical_protein 161743:162288
helicase_loader 142521:143198	DNA_helicase_loader 147117:147770	DNA_helicase_loader 73904:74542
baseplate_wedge_protein 65246:65800	baseplate_wedge_subunit 75229:75819	baseplate_wedge_subunit 145741:146331
thymidine_kinase 48468:49064	thymidine_kinase 58354:58935	thymidine_kinase 163860:164441
deoxycytidylate_deaminase 127762:128268	dCMP_deaminase 130340:130921	dCMP_deaminase 89774:90355
prohead_core_scaffolding_protein_and_protease 94238:94933	prohead_assembly_(scaffolding)_protein 100697:101335	prohead_assembly_protein 118781:119419
small_terminase_protein 86029:86526	terminase_small_subunit 93446:93940	terminase_small_subunit 126804:127298
head_completion_protein 64717:65190	head_completion_protein 74729:75181	head_completion_protein 146379:146831
recombination,_repair_and_ssDNA_binding_protein 105908:106315	single_stranded_DNA-binding_protein 111441:111854	single_stranded_DNA-binding_protein 108844:109257
RNA_polymerase_sigma_factor 27159:27692	RNA_polymerase_sigma_factor_for_late_transcription 38844:39401	RNA_polymerase_sigma_factor_for_late_transcription 14138:14695
endonuclease_II 131293:131754	endonuclease 136571:136981	endonuclease 83714:84124
prohead_core_protein 93819:94226	capsid_and_scaffold 100272:100697	capsid_and_scaffold_protein 119419:119844
tail_tube_protein 91448:91942	tail_fibers_protein 97884:98375	tail_tube_protein 121729:122220
baseplate_wedge_protein 106363:106749	baseplate_wedge_subunit 111921:112319	baseplate_wedge_subunit 108379:108777
putative_prohead_core_protein_I 93567:93806	prohead_core_protein 100033:100272	prohead_core_protein 119844:120071
translation_repressor_protein 18743:19105	endoribonuclease_translational_repressor_of_early_genes 30052:30420	endoribonuclease_translational_repressor_of_early_genes 23988:24356
double-stranded_DNA-binding_protein 143455:143730	double-stranded_DNA_binding_protein 148083:148352	double-stranded_DNA_binding_protein 73307:73576
hypothetical_protein 68186:68479	phospholipase 78060:78353	phospholipase 143207:143500
helicase 105411:105647	DNA_helicase 110735:110965	DNA_helicase 109733:109963

Ec_Makalu_001	vB-Eco-KMB40	ZCEC14
long_tail_fiber,_proximal_subunit 144472:148212	long_tail_fiber,_proximal_subunit 147941:151810	hypothetical_protein 9534:13403
baseplate_wedge_subunit 69256:72342	baseplate_wedge_initiator 79702:82800	hypothetical_protein 77606:80704
DNA_polymerase 15992:18670	DNA_polymerase 24071:26767	primase 133209:135905
baseplate_wedge_subunit_and_tail_pin 75780:77582	hypothetical_protein 84727:86532	hypothetical_protein 73872:75677
baseplate_hub_subunit_and_tail_lysozyme 64697:66490	baseplate_hub_structural_protein_/_lysozyme_R 75164:76891	nicotinamide_mononucleotide_transporter 83515:85242
DNA_ligase 123198:124694	DNA_ligase 123500:124963	hypothetical_protein 35218:36678
putative_recombination_endonuclease_subunit 22460:24142	recombination-related_endonuclease 30177:31859	tail_tip_protein 128159:129799
baseplate_wedge_subunit 67355:69259	baseplate_wedge_subunit 77723:79705	minor_tail_protein 80701:82683
ribonucleoside-diphosphate_reductase_subunit_alpha 133630:135873	ribonucleotide_reductase_of_class_Ia_(aerobic),_alpha_subunit 137246:139510	hypothetical_protein 20430:22694
DNA_topoisomerase_II_large_subunit 3229:5052	DNA_topoisomerase_large_subunit 2453:4270	hypothetical_protein 156032:157849
tail_sheath_protein 87260:89260	tail_sheath 95149:97128	hypothetical_protein 63276:65255
DNA_helicase 6596:7963	DNA_helicase 8440:9759	deoxi-UTP_pyrophosphatase 150173:151492
baseplate_central_spike_complex_protein 106106:107215	baseplate_hub_subunit 113237:114409	tail_fibers_protein 45765:46940
RNA_ligase_A 128495:129664	RNA_ligase 134410:135534	hypothetical_protein 24415:25539
portal_vertex_protein 89887:91452	portal_vertex_of_the_head 97820:99394	deoxynucleoside_kinase 60362:61936
RNA-DNA_and_DNA-DNA_helicase_ATPase 101765:103267	DNA_helicase 108591:110099	terminase_large_subunit 49977:51488
putative_anaerobic_ribonucleoside-triphosphate_reductase 34218:36080	ribonucleotide_reductase_of_class_III_(anaerobic),_large_subunit 40082:41899	hypothetical_protein 118646:120463
capsid_vertex_protein 96028:97269	capsid_vertex_protein 103204:104487	hypothetical_protein 55260:56543
DNA_primase-helicase_subunit 12878:14290	DNA_primase/helicase 18243:19670	hypothetical_protein 140323:141750
RecA-like_recombination_protein 14660:15730	recombination_protein 20017:21198	hypothetical_protein 138795:139976
DNA_primase 10222:11250	DNA_primase_/_DNA_helicase 15567:16595	hypothetical_protein 143442:144470
RnaseH 143264:144211	ribonuclease_H 146919:147836	endolysin 13508:14425
polynucleotide_kinase 126800:127678	3'-phosphatase,_5'-polynucleotide_kinase 131755:132660	HNH_homing_endonuclease 27270:28178
endonuclease_subunit 25549:26574	recombination-related_endonuclease 33173:34192	holin 126654:127673
aerobic_NDP_reductase,_small_subunit 131878:133038	ribonucleotide_reductase_of_class_Ia_(aerobic),_beta_subunit 136025:137194	hypothetical_protein 22746:23924
major_capsid_protein 93705:95291	major_capsid_protein 101555:103120	hypothetical_protein 56627:58192
clamp_loader,_small_subunit 19695:20669	replication_factor_C_small_subunit_/_DNA_polymerase_clamp_loader_subunit 27781:28740	putative_single-stranded_DNA_binding_protein 131236:132195
baseplate_hub_assembly_protein 105358:106128	baseplate 112488:113240	hypothetical_protein 46940:47689
sliding_clamp_DNA_polymerase_accessory_protein 20735:21421	sliding_clamp_DNA_polymerase_accessory_protein 28792:29478	ATP-dependent_Clp_protease 130498:131184
inhibitor_of_prohead_protease 100969:101694	hypothetical_protein 107241:107921	endolysin 51539:52219
prohead_core_protein 92851:93645	prohead_assembly_(scaffolding)_protein 100727:101536	baseplate_wedge 58211:59020
single-strand_DNA_binding_protein 140435:141376	single_stranded_DNA-binding_protein 144081:144986	hypothetical_protein 15768:16676
tail_sheath_stabilizer_and_completion_protein 84106:84894	tail_sheath_stabilizer_and_completion_protein 91980:92798	hypothetical_protein 67606:68424
DNA_end_protector_protein 62763:63584	DNA_end_protector_protein 73266:74090	hypothetical_protein 86316:87140
exonuclease 5327:5995	exonuclease 7255:7938	lysozyme 151994:152677
baseplate_wedge_subunit 72335:73330	baseplate_wedge_subunit 82793:83797	hypothetical_protein 76609:77613
holin_lysis_mediator 154405:155061	holin 157812:158468	DnaB-like_helicase 3216:3872
neck_protein 81456:82388	neck_protein 90237:91166	tail_fibers_protein 69238:70167
clamp_loader_small_subunit 19117:19695	DNA_polymerase_clamp_loader_subunit 27216:27779	D3_protein 132197:132760
neck_protein 83294:84034	neck_protein 91168:91938	hypothetical_protein 68466:69236
dCTP_pyrophosphatase 9689:10207	dCTP_pyrophosphatase 14780:15298	hypothetical_protein 144739:145257
hypothetical_protein 50597:51187	hypothetical_protein 58019:58564	hypothetical_protein 101389:101934
hypothetical_protein 141437:142117	DNA_helicase_loader 145675:146328	hypothetical_protein 15016:15669
baseplate_wedge_subunit 64125:64679	baseplate_wedge_subunit 74590:75180	hypothetical_protein 85226:85816
thymidine_kinase 48478:49074	thymidine_kinase 55879:56460	DNA_polymerase 103689:104270
deoxycytidylate_deaminase 126069:126575	dCMP_deaminase 129873:130454	hypothetical_protein 29673:30254
prohead_core_scaffolding_protein_and_protease 92124:92819	prohead_assembly_(scaffolding)_protein 100058:100696	hypothetical_protein 59051:59689
small_terminase_protein 84929:85405	terminase_small_subunit 92807:93301	hypothetical_protein 67103:67597
head_completion_protein 63596:64069	head_completion_protein 74090:74542	NrdH_family_redoxin 85864:86316
recombination,_repair_and_ssDNA_binding_protein 103808:104215	single_stranded_DNA-binding_protein 110933:111346	hypothetical_protein 48832:49245
RNA_polymerase_sigma_factor 27171:27704	RNA_polymerase_sigma_factor_for_late_transcription 36507:37064	hypothetical_protein 123781:124338
endonuclease_II 129600:130061	endonuclease 135587:135997	single-stranded-DNA-specific_exonuclease 23952:24362
prohead_core_protein 91705:92112	capsid_and_scaffold 99633:100058	putative_replicative_DNA_helicase 59689:60114
tail_tube_protein 89334:89828	tail_fibers_protein 97245:97736	hypothetical_protein 62020:62511
baseplate_wedge_protein 104263:104649	baseplate_wedge_subunit 111413:111811	putative_HNH_homing_endonuclease 48367:48765
putative_prohead_core_protein 91453:91692	prohead_core_protein 99394:99633	hypothetical_protein 60114:60362
hypothetical_protein 18753:19115	endoribonuclease_translational_repressor_of_early_genes 26846:27214	hypothetical_protein 132762:133130
double-stranded_DNA_binding_protein 142374:142649	double-stranded_DNA_binding_protein,_late_transcriptional_regulation 146641:146910	hypothetical_protein 14434:14703
hypothetical_protein 67065:67358	phospholipase 77421:77714	hypothetical_protein 82692:82985
putative_ATP-dependent_DNA_helicase 103311:103547	DNA_helicase 110125:110358	portal_protein 49721:49951

vB_EcoM-CHD94UKE2	vB_EcoM_ACG-C40	vB-Eco-KMB39
long_tail_fiber 67021:70890	long_tail_fiber_proximal_subunit 148067:151936	long_tail_fiber,_proximal_subunit 150140:154009
baseplate_wedge_initiator 135472:138570	baseplate_wedge_initiator 80721:83819	baseplate_wedge_initiator 79602:82700
DNA_polymerase 24677:27373	DNA_polymerase 27196:29892	DNA_polymerase 23374:26070
hypothetical_protein 131740:133545	baseplate_wedge_subunit_and_tail_pin 85746:87551	baseplate_wedge_subunit_and_tail_pin 84627:86432
baseplate_hub_structural_protein_/lysozyme_R 141381:143108	baseplate_hub_subunit_and_tail_lysozyme 75294:77024	baseplate_hub_structural_protein_/_lysozyme_R 75065:76792
DNA_ligase 93785:95248	DNA_ligase 123368:124831	DNA_ligase 124918:126381
recombination-related_endonuclease 19404:21044	recombination_endonuclease_subunit 33525:35207	recombination-related_endonuclease 30411:32057
baseplate_wedge_subunit 138567:140549	baseplate_wedge_subunit 78742:80724	baseplate_wedge_subunit 77623:79605
ribonucleotide_reductase_of_class_Ia_(aerobic),_alpha_subunit 78559:80823	ribonucleoside-diphosphate_reductase_subunit_alpha 137277:139541	ribonucleotide_reductase_of_class_Ia_(aerobic),_alpha_subunit 139226:141490
DNA_topoisomerase_large_subunit 47702:49519	topoisomerase_II_large_subunit 2447:4264	DNA_topoisomerase_large_subunit 2448:4265
tail_sheath 121144:123123	tail_sheath_monomer 96168:98147	tail_sheath 95051:97030
DNA_helicase 41597:42916	DNA_helicase 8819:10138	DNA_helicase 8078:9397
baseplate_hub_subunit 104293:105465	baseplate_hub_subunit 114851:116023	baseplate_hub_subunit 114655:115827
RNA_ligase 82535:83659	RNA_ligase_1_and_tail_fiber_attachment_catalyst 134432:135556	RNA_ligase 135702:136826
portal_protein 118878:120452	portal_vertex_protein_of_head 98839:100413	portal_vertex_of_the_head 98358:99932
DNA_helicase 108501:110012	RNA-DNA_and_DNA-DNA_helicase 109618:111126	DNA_helicase 109422:110930
ribonucleotide_reductase_of_class_III_(anaerobic),_large_subunit 9010:10827	anaerobic_NTP_reductase_large_subunit 43381:45198	ribonucleotide_reductase_of_class_III_(anaerobic),_large_subunit 39750:41567
capsid_vertex_protein 113785:115068	head_vertex_protein 104217:105500	capsid_vertex_protein 103748:105031
DNA_primase/helicase 31780:33207	DNA_primase-helicase_subunit 20728:22155	DNA_primase/helicase 17784:19211
recombination_protein 30252:31433	RecA-like_recombination_protein 22502:23683	recombination_protein 19558:20739
DNA_primase_/_DNA_helicase 34239:35267	DNA_primase_subunit 18050:19078	DNA_primase_/_DNA_helicase 15097:16125
ribonuclease_H 70995:71912	ribonuclease_RNaseH 147045:147962	ribonuclease_H 149118:150035
3'-phosphatase,_5'-polynucleotide_kinase 85409:86317	polynucleotide_5'-kinase_and_3'-phosphatase 131774:132682	3'-phosphatase,_5'-polynucleotide_kinase 133045:133953
recombination-related_endonuclease 17899:18918	recombination_endonuclease_subunit 35651:36670	recombination-related_endonuclease 32543:33562
ribonucleotide_reductase_of_class_Ia_(aerobic),_beta_subunit 80875:82044	aerobic_NDP_reductase_small_subunit 136047:137225	ribonucleotide_reductase_of_class_Ia_(aerobic),_beta_subunit 137317:138495
major_capsid_protein 115152:116717	major_capsid_protein 102574:104133	major_capsid_protein 102099:103664
replication_factor_C_small_subunit_/_DNA_polymerase_clamp_loader_subunit 22481:23440	clamp_loader_subunit 31129:32088	replication_factor_C_small_subunit_/_DNA_polymerase_clamp_loader_subunit 27307:28266
baseplate 105462:106214	baseplate_hub_assembly_protein 114102:114854	baseplate 113906:114655
sliding_clamp_DNA_polymerase_accessory_protein 21743:22429	sliding_clamp_DNA_polymerase_accessory_protein 32140:32826	sliding_clamp_DNA_polymerase_accessory_protein 28318:29004
hypothetical_protein 110063:110743	inhibitor_of_prohead_protease 108268:108948	inhibitor_of_prohead_protease 108072:108752
prohead_assembly_protein 116736:117545	prohead_core_scaffold_protein 101746:102555	prohead_assembly_(scaffolding)_protein 101271:102080
single_stranded_DNA-binding_protein 73845:74750	single-stranded_DNA_binding_protein 144202:145107	single_stranded_DNA-binding_protein 146280:147185
tail_sheath_stabilizer_and_completion_protein 125474:126292	tail_sheath_stabilizer_and_completion_protein 92999:93817	tail_sheath_stabilizer_and_completion_protein 91882:92700
DNA_end_protector_protein 144182:145006	DNA_end_protector_protein 73396:74220	DNA_end_protector_protein 73164:73991
exonuclease 43650:44333	exonuclease_A 7634:8317	exonuclease 6893:7576
baseplate_wedge_subunit 134475:135479	baseplate_wedge_subunit 83872:84816	baseplate_wedge_subunit 82693:83697
holin 60721:61377	holin_lysis_mediator_t 157742:158398	holin 159974:160630
neck_protein 127106:128035	neck_protein 91256:92185	neck_protein 90137:91066
DNA_polymerase_clamp_loader_subunit 23442:24005	clamp_loader_subunit 30564:31127	DNA_polymerase_clamp_loader_subunit 26742:27305
neck_protein 126334:127104	neck_protein 92187:92957	neck_protein 91069:91839
dCTP_pyrophosphatase 35538:36053	dCTPase 15835:16353	dCTP_pyrophosphatase 14310:14828
hypothetical_protein 159695:160240	hypothetical_protein 59578:60111	hypothetical_protein 57898:58443
DNA_helicase_loader 72518:73156	loader_of_DNA_helicase 145801:146454	DNA_helicase_loader 147874:148527
baseplate_wedge_subunit 143092:143682	baseplate_wedge_subunit 74720:75310	baseplate_wedge_subunit 74491:75081
thymidine_kinase 161799:162380	thymidine_kinase 57414:57995	thymidine_kinase 55752:56333
dCMP_deaminase 88137:88718	dCMP_deaminase 129894:130475	dCMP_deaminase 131165:131746
prohead_assembly_protein 117576:118214	prohead_core_scaffold_protein_and_protease 101077:101715	prohead_assembly_(scaffolding)_protein 100602:101240
terminase_small_subunit 124971:125465	terminase_DNA_packaging_enzyme_small_subunit 93826:94320	terminase_small_subunit 92709:93203
head_completion_protein 143730:144182	head_completion_protein 74220:74672	head_completion_protein 73991:74443
single_stranded_DNA-binding_protein 107356:107769	recombination,_repair_and_ssDNA_binding_protein 112547:112960	single_stranded_DNA-binding_protein 112351:112764
RNA_polymerase_sigma_factor_for_late_transcription 15026:15583	sigma_factor_recognition_late_T4_promoter 38986:39543	RNA_polymerase_sigma_factor_for_late_transcription 35878:36435
endonuclease 82072:82482	endonuclease_II 135609:136019	endonuclease 136879:137289
capsid_and_scaffold_protein 118214:118639	prohead_core_protein 100652:101077	capsid_and_scaffold 100177:100602
tail_tube_protein 120536:121027	tail_tube_protein 98264:98755	tail_fibers_protein 97783:98274
baseplate_wedge_subunit 106891:107289	baseplate_wedge_subunit 113027:113425	baseplate_wedge_subunit 112831:113229
prohead_core_protein 118639:118878	prohead_core_protein 100413:100652	prohead_core_protein 99932:100177
endoribonuclease_translational_repressor_of_early_genes 24007:24375	translational_repressor_protein 30194:30562	endoribonuclease_translational_repressor_of_early_genes 26372:26740
double-stranded_DNA_binding_protein 71921:72190	dsDNA_binding_protein 146767:147036	double-stranded_DNA_binding_protein,_late_transcriptional_regulation 148840:149109
phospholipase 140558:140851	hypothetical_protein 78440:78733	phospholipase 77321:77614
DNA_helicase 108245:108475	RNA-DNA_and_DNA-DNA_helicase 111841:112071	DNA_helicase 111645:111875

vB_EcoM-CHD2BS1	vB_EcoM_LNA2
long_tail_fiber 69101:72970	tail_fiber_protein_proximal_subunit 97853:101731
baseplate_wedge_initiator 138427:141525	baseplate_wedge_subunit 32693:35791
DNA_polymerase 23845:26541	DNA_polymerase 146273:148984
hypothetical_protein 134694:136499	baseplate_wedge_subunit 37733:39538
baseplate_hub_structural_protein_/lysozyme_R 144336:146063	baseplate_hub_subunit_and_tail_lysozyme 28166:29899
DNA_ligase 96733:98196	ATP-dependent_DNA_ligase 74491:75984
recombination-related_endonuclease 18572:20212	SbcC-like_subunit_of_palindrome_specific_endonuclease 152441:154129
baseplate_wedge_subunit 141522:143504	baseplate_wedge_subunit 30723:32696
ribonucleotide_reductase_of_class_Ia_(aerobic),_alpha_subunit 81490:83754	ribonucleotide_reductase_large_subunit 89456:91711
DNA_topoisomerase_large_subunit 49633:51450	DNA_topoisomerase_II_large_subunit 119627:121444
tail_sheath 124098:126077	tail_sheath 48178:50160
DNA_helicase 43525:44844	Dda-like_helicase 125652:126965
baseplate_hub_subunit 107247:108419	baseplate_hub 66273:67445
RNA_ligase 85475:86599	RNA_ligase_and_tail_fiber_protein_attachment_catalyst 86606:87730
portal_protein 121832:123406	portal_protein 50849:52420
DNA_helicase 111455:112966	DNA_helicase 61701:63215
ribonucleotide_reductase_of_class_III_(anaerobic),_large_subunit 9007:10824	anaerobic_ribonucleoside_reductase_large_subunit 160771:162588
capsid_vertex_protein 116739:118022	major_head_protein 56580:57863
DNA_primase/helicase 32191:33618	DnaB-like_replicative_helicase 136623:137999
recombination_protein 30669:31844	DNA_repair_protein 138409:139581
DNA_primase_/_DNA_helicase 35010:36038	DNA_primase 132095:133117
ribonuclease_H 73075:73992	exonuclease 96832:97749
3'-phosphatase,_5'-polynucleotide_kinase 88348:89253	polynucleotide_kinase 83753:84529
recombination-related_endonuclease 17012:18031	SbcD-like_subunit_of_palindrome_specific_endonuclease 154606:155625
ribonucleotide_reductase_of_class_Ia_(aerobic),_beta_subunit 83806:84984	ribonucleotide_reductase_class_Ia_beta_subunit 88224:89402
major_capsid_protein 118106:119671	major_head_protein 54589:56157
replication_factor_C_small_subunit_/_DNA_polymerase_clamp_loader_subunit 21649:22608	clamp_loader_of_DNA_polymerase 150002:150964
baseplate 108416:109168	baseplate_hub 65524:66276
sliding_clamp_DNA_polymerase_accessory_protein 20911:21597	DNA_polymerase_processivity_factor 151040:151726
hypothetical_protein 113017:113697	minor_head_protein_inhibitor_of_protease 60982:61650
prohead_assembly_protein 119690:120499	head_scaffolding_protein 53759:54571
single_stranded_DNA-binding_protein 75925:76830	single_strand_DNA_binding_protein 94560:95423
tail_sheath_stabilizer_and_completion_protein 128428:129246	tail_sheath_stabilizer 45003:45824
DNA_end_protector_protein 147137:147961	DNA_end_protector 26256:27077
exonuclease 45578:46261	exonuclease 124965:125642
baseplate_wedge_subunit 137430:138434	baseplate_wedge_subunit 35784:36788
holin 62746:63402	holin 107279:107938
neck_protein 130060:130989	head-tail_adaptor_Ad2 43258:44184
DNA_polymerase_clamp_loader_subunit 22610:23173	clamp_loader_of_DNA_polymerase 149437:150000
neck_protein 129288:130058	head_closure_Hc2 44186:44950
hypothetical_protein 37735:38253	nucleoside_triphosphate_pyrophosphohydrolase 131289:131810
hypothetical_protein 160765:161310	unknown_function 7941:8483
DNA_helicase_loader 74598:75236	DNA_helicase_loader 95589:96242
baseplate_wedge_subunit 146047:146637	baseplate_wedge_subunit 27663:28166
thymidine_kinase 162882:163463	thymidine_kinase 5955:6536
dCMP_deaminase 91072:91653	dCMP_deaminase 81083:81664
prohead_assembly_protein 120530:121168	head_maturation_protease 53084:53725
terminase_small_subunit 127925:128419	terminase_small_subunit 45834:46328
head_completion_protein 146685:147137	head_closure 27080:27529
single_stranded_DNA-binding_protein 110310:110723	UvsY-like_recombination_mediator 63958:64452
RNA_polymerase_sigma_factor_for_late_transcription 14140:14697	RNA_polymerase_sigma_factor 156614:157171
endonuclease 85012:85422	endonuclease 87787:88197
capsid_and_scaffold_protein 121168:121593	head_scaffolding_protein 52659:53084
tail_tube_protein 123490:123981	tail_tube 50273:50764
baseplate_wedge_subunit 109845:110243	baseplate_wedge_subunit 64452:64850
prohead_core_protein 121593:121832	prohead 52420:52659
endoribonuclease_translational_repressor_of_early_genes 23175:23543	translation_repressor 149066:149443
double-stranded_DNA_binding_protein 74001:74270	transcriptional_regulator 96555:96824
phospholipase 143513:143806	PAAR_motif_of_membran_proteins 30421:30714
DNA_helicase 111199:111429	DNA_helicase 63241:63474

Names of the encoded proteins and positions of their genes in genomes of phages from the second clade of the VipTree-based multigene tree presented in Figure 1. These lists were used to generate tree presented in Figure 3.

The importance of the gene for tree topology marked by '+'	No. of parsimony informative sites in alignment	Characteristic protein features identified by InterProScan	The name of gene product and its position in genome of particular phage:			
			vB_EcoS_ACG-M12	vB-Eco-KMB46	ES	vB-Eco-KMB25
+	181	Bacteriophage tail tip fiber protein; Fibronectin type III; FN3; Tip attachment protein J,FNIII-A domain; Putative phage tail protein	putative_tail_fiber_protein 20695:24111	host_specificity_protein_J 8341:11919	hypothetical_protein 18148:21534	tail_protein_with_putative_glycosyl_ hydrolase 16222:19680
-	135	Tape measure protein [H].; Lambda phage tail tape-measure protein (Tape_meas_lam_C); Prophage tail length tape measure protein	putative_tail_tape-measure_protein 14174:17152	tail_length_tape-measure_protein_H 14483:17164	tail_length_tape-measure_protein 12591:15563	tail_length_tape-measure_protein_1 10744:13656
-	82	Helicase conserved C-terminal domain; Type III restriction enzyme, res subunit; P-loop containing nucleoside triphosphate hydrolases	putative_ATP-dependent_helicase 33751:35745	DNA_helicase 48364:50361	putative_single-stranded_DNA_binding_protein 31974:33506	DNA_helicase 27673:29646
-	79	Terminase RNaseH-like domain	putative_terminase_large_subunit 4401:5969	terminase_large_subunit 26243:27811	terminase_large_subunit 1330:2892	terminase_large_subunit 537:2108
-	75	Major capsid protein	hypothetical_protein 9505:10449	hypothetical_protein 20588:21559	hypothetical_protein 7942:8886	major_capsid_protein 6362:7336
-	64	PDDEXK-like domain of unknown function (DUF3799)	putative_exodeoxyribonuclease_VIII 26618:27583	exonuclease 5592:6647	capsid_maturation_protease 23516:24487	exodeoxyribonuclease_VIII 21063:22118
-	56	Zinc-binding domain of primase-helicase; Zinc beta-ribbon	putative_DNA_primase 32162:33088	DNA_primase/helicase 28:957	hypothetical_protein 28820:29743	DNA_primase/helicase 25990:26916
-	53	Protein of unknown function (DUF3987)	hypothetical_protein 42590:44173	hypothetical_protein 41526:43106	hypothetical_protein 39556:41139	DNA_helicase 35083:36663
-	47	Phage portal protein	putative_portal_protein 6428:7684	portal_(connector)_protein 24880:26172	DUF1073_domain-containing_protein 3374:4645	portal_protein 2162:3457
-	43	endolysin_R21-like; Phage lysozyme; Lysozyme-like; Endolysin.; SAR-endolysin.	putative_endolysin 41386:41871	lysozyme_R 44328:44819	hypothetical_protein 38337:38822	lysozyme_or_muramidase 33375:33869
-	41	Metallo-dependent phosphatases	hypothetical_protein 37986:39122	putative_phosphoesterase_protein 45180:46295	hypothetical_protein 35526:36659	phosphoesterase 31924:33039
-	39	Uncharacterized protein conserved in bacteria (DUF2213)	putative_prohead_protease 7750:8859	capsid_and_scaffold 23002:24117	DUF2213_domain-containing_protein 4666:5781	capsid_and_scaffold 4187:5296
-	35	Phage minor tail protein L	putative_minor_tail_protein 18090:18845	tail_tip_assembly_protein_L 13306:14067	hypothetical_protein 15984:16739	minor_tail_protein 14071:14829
-	25	ERF superfamily	putative_recombination_protein 27656:28306	recombinase 4602:5255	head_morphogenesis_protein 25039:25689	recombinase 22532:23194
-	25	-	hypothetical_protein 11591:12028	hypothetical_protein 18994:19437	hypothetical_protein 10020:10463	tail_completion_or_Neck1_protein 8498:8938
-	24	Phage minor tail protein	putative_minor_tail_protein 17702:18052	minor_tail_protein 14130:14480	hypothetical_protein 15595:15945	minor_tail_protein 13656:14003
-	21	VRR-NUC domain	hypothetical_protein 35747:36166	hypothetical_protein 47945:48364	hypothetical_protein 33508:33927	endonuclease 29649:30056
-	21	Phage tail terminator protein	hypothetical_protein 12018:12419	hypothetical_protein 18603:19004	tail_fiber_protein 10463:10861	putative_tail_protein 8928:9320
-	20	-	hypothetical_protein 11227:11598	hypothetical_protein 19427:19801	putative_single-stranded_DNA_binding_protein 9656:10027	putative_head-tail_adaptor 8134:8505

-	18	Nucleic acid-binding proteins	putative_single-stranded_DNA_binding_protein 28354:28815	single-stranded_DNA-binding_protein 4106:4558	terminase_small_subunit 25730:26167	single-stranded_DNA-binding_protein 23232:23657
-	17	-	hypothetical_protein 41847:42236	lipoprotein 43936:44331	hypothetical_protein 38819:39205	unimolecular_spanin 33866:34255
-	15	Phage tail assembly chaperone	hypothetical_protein 13504:13818	hypothetical_protein 17521:17838	hypothetical_protein 11925:12239	tail_assembly_chaperone 10075:10389
-	14	DNA-packaging protein gp3	putative_terminase_small_subunit 3877:4383	terminase_small_subunit 27811:28338	hypothetical_protein 283:807	terminase_small_subunit 1:525
-	8	-	hypothetical_protein 42251:42586	hypothetical_protein 43103:43498	hypothetical_protein 39217:39552	hypothetical_protein 34688:35080

Names of the encoded proteins and positions of their genes in genomes of phages from the third clade of the VipTree-based multigene tree presented in Figure 1. These lists were used to generate tree presented in Figure 4.

The importance of the gene for tree topology marked by '+'	No. of parsimony informative sites in alignment	Characteristic protein features identified by InterProScan	The name of gene product and its position in genome of particular phage:				
			vB_EcoS-101114B2	vB_EcoS-CHD2B1	vB_EcoS-22664B1	vB_EcoS-CHD5UKE2	vB_EcoS-101114BS4
+	38	-	hypothetical_protein 31263:31727	hypothetical_protein 13430:13894	hypothetical_protein 31348:31812	hypothetical_protein 30362:30826	hypothetical_protein 30105:30569
+	36	Family of unknown function (DUF6453)	hypothetical_protein 25578:26543	hypothetical_protein 18613:19578	hypothetical_protein 25663:26628	hypothetical_protein 35166:36131	hypothetical_protein 24865:25830
-	21	Enterobacter phage Enc34, ssDNA-binding protein; Nucleic acid-binding proteins	hypothetical_protein 37900:38715	hypothetical_protein 6503:7294	hypothetical_protein 37985:38794	hypothetical_protein 23646:24377	hypothetical_protein 36553:37341
-	17	Bacteriophage tail tip fiber protein; Putative phage tail protein; Tip attachment protein J,FNIII-A domain	host_specificity_protein_J 22104:25550	host_specificity_protein_J 19606:23052	host_specificity_protein_J 22189:25635	host_specificity_protein_J 36159:39605	host_specificity_protein_J 21391:24837
-	13	Domain of unknown function (DUF6378)	hypothetical_protein 29535:30056	hypothetical_protein 15100:15621	hypothetical_protein 29620:30141	hypothetical_protein 32033:32626	hypothetical_protein 28307:28900
-	12	Nucleotide modification associated domain 5	hypothetical_protein 9913:10677	hypothetical_protein 34461:35225	hypothetical_protein 9998:10762	hypothetical_protein 6277:7041	hypothetical_protein 9206:9970
-	12	Domain of unknown function (DUF4055)	hypothetical_protein 6857:8377	hypothetical_protein 36761:38281	hypothetical_protein 6942:8462	hypothetical_protein 9154:10674	hypothetical_protein 5576:7096
-	10	Major capsid protein 13-like	major_capsid_protein 11629:12729	major_capsid_protein 32415:33509	major_capsid_protein 11714:12814	major_capsid_protein 3865:4965	major_capsid_protein 10922:12016
-	10	Protein of unknown function (DUF2800)	hypothetical_protein 39018:40448	hypothetical_protein 4748:6178	hypothetical_protein 39097:40527	hypothetical_protein 21921:23321	hypothetical_protein 37665:39095
-	7	-	hypothetical_protein 16041:16196	hypothetical_protein 28948:29103	hypothetical_protein 16126:16281	hypothetical_protein 399:554	hypothetical_protein 15328:15483
-	6	JAB1/MPN domain; Cysteine proteinases; MPN_NLPC_P60; NlpC/P60 family; Prokaryotic homologs of the JAB domain	tail_assembly_protein 20759:21499	tail_assembly_protein 23657:24397	tail_assembly_protein 20844:21584	tail_assembly_protein 40210:40950	tail_assembly_protein 20046:20786
-	6	AAA domain; P-loop containing nucleoside triphosphate hydrolases	hypothetical_protein 42329:44593	hypothetical_protein 608:2872	hypothetical_protein 42407:44671	hypothetical_protein 17928:20189	hypothetical_protein 41046:43238
-	5	-	hypothetical_protein 30586:30810	hypothetical_protein 14347:14571	hypothetical_protein 30671:30895	hypothetical_protein 31279:31503	hypothetical_protein 29428:29652
-	4	-	hypothetical_protein 417:686	hypothetical_protein 44560:44829	hypothetical_protein 447:716	hypothetical_protein 16842:17111	hypothetical_protein 44055:44324
-	4	DNA N-6-adenine-methyltransferase (Dam)	DNA_N-6-adenine_methyltransferase 30791:31276	DNA_N-6-adenine_methyltransferase 13881:14366	DNA_N-6-adenine_methyltransferase 30876:31361	DNA_N-6-adenine_methyltransferase 30813:31298	DNA_N-6-adenine_methyltransferase 29633:30118
-	4	-	assembly_protein_I 21496:22107	assembly_protein_I 23049:23666	assembly_protein_I 21581:22192	assembly_protein_I 39602:40213	assembly_protein_I 20783:21394
-	4	-	hypothetical_protein 1670:1957	hypothetical_protein 43181:43468	hypothetical_protein 1700:1987	hypothetical_protein 15575:15874	hypothetical_protein 174:461
-	4	-	minor_capsid_protein 8361:9464	minor_capsid_protein 35674:36777	minor_capsid_protein 8446:9549	minor_capsid_protein 8067:9170	minor_capsid_protein 7080:8183
-	3	-	hypothetical_protein 40636:40887	hypothetical_protein 4312:4563	hypothetical_protein 40715:40966	hypothetical_protein 21482:21733	hypothetical_protein 39280:39531
-	3	-	terminase_small_subunit 4889:5461	terminase_small_subunit 39677:40249	terminase_small_subunit 4974:5546	hypothetical_protein 12070:12642	hypothetical_protein 3608:4180
-	2	-	hypothetical_protein 10827:11531	hypothetical_protein 33607:34311	hypothetical_protein 10912:11616	hypothetical_protein 5063:5767	hypothetical_protein 10120:10824
-	1	-	hypothetical_protein 38783:39025	hypothetical_protein 6171:6413	hypothetical_protein 38862:39104	hypothetical_protein 23314:23556	hypothetical_protein 37430:37672
-	1	-	HNH_endonuclease 32051:32374	HNH_endonuclease 12822:13106	HNH_endonuclease 32136:32459	HNH_endonuclease 29958:30281	HNH_endonuclease 30650:30973
-	1	-	hypothetical_protein 9467:9610	hypothetical_protein 35528:35671	hypothetical_protein 9552:9695	hypothetical_protein 7921:8031	hypothetical_protein 8186:8329
-	1	Terminase RNaseH-like domain; Terminase large subunit, T4likevirus-type, N-terminal; P-loop containing nucleoside triphosphate hydrolases	terminase 5461:6846	terminase 38292:39677	terminase 5546:6931	terminase 10685:12070	terminase 4180:5565
-	1	-	hypothetical_protein 32371:33030	hypothetical_protein 12127:12786	hypothetical_protein 32456:33115	hypothetical_protein 29302:29961	hypothetical_protein 30970:31629
-	1	Phage minor tail protein	hypothetical_protein 19372:19971	hypothetical_protein 25185:25784	hypothetical_protein 19457:20056	hypothetical_protein 41738:42337	hypothetical_protein 18659:19258
-	0	-	hypothetical_protein 41857:41994	hypothetical_protein 3207:3344	hypothetical_protein 41935:42072	hypothetical_protein 20688:20831	hypothetical_protein 40502:40639
-	0	HAD-like	hypothetical_protein 30119:30589	hypothetical_protein 14568:15038	hypothetical_protein 30204:30674	hypothetical_protein 31500:31970	hypothetical_protein 28961:29431
-	0	Phage tail terminator protein	hypothetical_protein 14338:14760	hypothetical_protein 30384:30806	hypothetical_protein 14423:14845	hypothetical_protein 1834:2256	hypothetical_protein 13625:14047
-	0	-	hypothetical_protein 34753:34926	hypothetical_protein 10231:10404	hypothetical_protein 34838:35011	hypothetical_protein 27406:27579	hypothetical_protein 33352:33525

-	0	VRR-NUC-like; Restriction endonuclease-like	hypothetical_protein 34967:35245	hypothetical_protein 9912:10190	hypothetical_protein 35052:35330	hypothetical_protein 27087:27365	hypothetical_protein 33566:33844
-	0	DNA polymerase family A; Ribonuclease H-like; DNA/RNA polymerases	DNA_polymerase_I 35438:37729	DNA_polymerase_I 7520:9811	DNA_polymerase_I 35523:37814	DNA_polymerase_I 24603:26894	DNA_polymerase_I 34037:36328
-	0	-	hypothetical_protein 2010:2300	hypothetical_protein 42838:43128	hypothetical_protein 2040:2330	hypothetical_protein 15232:15522	hypothetical_protein 514:804
-	0	Putative DnaT-like ssDNA binding protein	hypothetical_protein 12788:13288	hypothetical_protein 31856:32356	hypothetical_protein 12873:13373	hypothetical_protein 3306:3806	hypothetical_protein 12075:12575
-	0	-	hypothetical_protein 13290:13403	hypothetical_protein 31741:31854	hypothetical_protein 13375:13488	hypothetical_protein 3191:3304	hypothetical_protein 12577:12690
-	0	P-loop containing nucleoside triphosphate hydrolases; SNF2-related domain	DNA_helicase 33063:34487	DNA_helicase 10670:12094	DNA_helicase 33148:34572	DNA_helicase 27845:29269	DNA_helicase 31662:33086
-	0	-	hypothetical_protein 4580:4792	hypothetical_protein 40346:40558	hypothetical_protein 4665:4877	hypothetical_protein 12739:12951	hypothetical_protein 3299:3511
-	0	-	hypothetical_protein 13390:13743	hypothetical_protein 31401:31754	hypothetical_protein 13475:13828	hypothetical_protein 2851:3204	hypothetical_protein 12677:13030
-	0	Phage tail assembly chaperone	hypothetical_protein 15617:15979	hypothetical_protein 29165:29527	hypothetical_protein 15702:16064	hypothetical_protein 616:978	hypothetical_protein 14904:15266
-	0	-	hypothetical_protein 2297:2536	hypothetical_protein 42602:42841	hypothetical_protein 2327:2566	hypothetical_protein 14990:15235	hypothetical_protein 801:1040
-	0	-	hypothetical_protein 28917:29492	hypothetical_protein 15664:16239	hypothetical_protein 29002:29577	hypothetical_protein 32669:33244	hypothetical_protein 27689:28264
-	0	-	hypothetical_protein 3034:3201	hypothetical_protein 41882:42088	hypothetical_protein 3064:3231	hypothetical_protein 14328:14495	hypothetical_protein 1538:1705
-	0	Phage minor tail protein L	minor_tail_protein 19968:20756	minor_tail_protein 24400:25188	minor_tail_protein 20053:20841	minor_tail_protein 40953:41741	minor_tail_protein 19255:20043
-	0	Phage lysozyme; Lysozyme-like; lyz_endolysin_autolysin	lysin 2526:3017	lysin 42121:42612	lysin 2556:3047	lysin 14512:15003	lysin 1030:1521
-	0	-	hypothetical_protein 16198:16293	hypothetical_protein 28851:28946	hypothetical_protein 16283:16378	hypothetical_protein 302:397	hypothetical_protein 15485:15580
-	0	-	hypothetical_protein 1545:1685	hypothetical_protein 43453:43593	hypothetical_protein 1575:1715	hypothetical_protein 15939:16085	hypothetical_protein 49:189
-	0	Phage tail tube protein, TTP	tail_tube_protein 14822:15547	tail_tube_protein 29597:30322	tail_tube_protein 14907:15632	tail_tube_protein 1048:1773	tail_tube_protein 14109:14834
-	0	-	hypothetical_protein 34492:34710	hypothetical_protein 10441:10665	hypothetical_protein 34577:34795	hypothetical_protein 27622:27840	hypothetical_protein 33091:33309
-	0	-	hypothetical_protein 37726:37896	hypothetical_protein 7353:7523	hypothetical_protein 37811:37981	hypothetical_protein 24436:24606	hypothetical_protein 36325:36501
-	0	Phage lipoprotein family	lipoprotein 26546:26767	lipoprotein 18389:18610	lipoprotein 26631:26852	lipoprotein 34942:35163	lipoprotein 25833:26054
-	0	-	hypothetical_protein 9668:9898	hypothetical_protein 35240:35470	hypothetical_protein 9753:9983	hypothetical_protein 7633:7863	hypothetical_protein 8387:8617
-	0	-	hypothetical_protein 13835:14338	hypothetical_protein 30806:31309	hypothetical_protein 13920:14423	hypothetical_protein 2256:2759	hypothetical_protein 13122:13625

Names of the encoded proteins and positions of their genes in genomes of phages from the fourth clade of the VipTree-based multigene tree presented in Figure 1. These lists were used to generate tree presented in Figure 5.

The importance of the gene for tree topology marked by '+'	No. of parsimony informative sites in alignment	Characteristic protein features identified by InterProScan	vB-Eco-KMB36	VB_EcoS-Golestan
+	158	-	tail_protein 19377:21932	putative_tail_protein 19660:22227
+	74	Phage tail assembly chaperone	tail_assembly_chaperone 14899:15315	hypothetical_protein 14257:14673
+	45	Phage tail tube protein	tail_tube_protein 11219:12385	putative_tail_protein 10484:11650
+	43	-	13.88_kDa_late_protein 33150:33650	hypothetical_protein 33347:33850
+	41	Phage tail assembly chaperone	tail_assembly_chaperone 15321:15680	hypothetical_protein 14679:15038
+	35	Protein of unknown function (DUF2800)	nuclease_protein 31596:32837	PD-(D/E)XK_nuclease_superfamily_protein 31450:32691
-	23	SNF2-related domain; P-loop containing nucleoside triphosphate hydrolases	DNA_helicase 24635:26059	helicase 24900:26324
-	20	VRR-NUC-like	hypothetical_protein 26287:26574	VRR-NUC_domain_protein 27465:27752
-	18	Putative DnaT-like ssDNA binding protein	putative_head-tail_connector 8828:9340	hypothetical_protein 8188:8700
-	15	-	spanin 5559:5945	putative_spanin 5234:5620
-	13	HTH_XRE; Helix-turn-helix domain; Putative DNA-binding domain	DNA-binding_transcriptional_activator 36672:36854	helix-turn-helix_domain_protein 37312:37500
-	11	-	hypothetical_protein 32834:33106	hypothetical_protein 33032:33304
-	9	Domain of unknown function (DUF1833)	putative_tail_protein 18509:19024	hypothetical_protein 18792:19307
-	9	Immunoglobulin	fibritin_protein 4757:5215	putative_tail_protein 3858:4319
-	6	Phage tail terminator protein	tail_completion_protein 10800:11216	hypothetical_protein 10065:10481
-	6	-	hypothetical_protein 27144:27275	hypothetical_protein 27834:27965
-	5	-	scaffold_protein 6137:6838	hypothetical_protein 5812:6513
-	4	-	DNA_binding_protein 43343:43579	hypothetical_protein 43675:43911
-	2	-	murein_DD-endopeptidase 19021:19386	hypothetical_protein 19304:19669
-	0	-	hypothetical_protein 24312:24449	exonuclease_subunit 24254:24391

			The name of gene product and its position in genome of particular phage:	
590B	**vB_EcoM_SHAK7693**	**vB_EcoS-phiEc4**	**vB_EcoS-22664BS2**	**vB_EcoS_UTEC10**
putative_tail_protein 18897:21464	tail_protein 20635:23124	tail_protein 19748:22243	hypothetical_protein 24409:26904	hypothetical_protein 20932:23418
hypothetical_protein 13542:13958	tail_assembly_chaperone 14463:14879	hypothetical_protein 14273:14689	hypothetical_protein 18955:19371	tail_assembly_chaperone 28447:28863
tail_protein 9949:11115	minor_tail_protein 11144:12310	putative_tail_protein 10500:11666	hypothetical_protein 15184:16350	tail_protein 31454:32620
13.88_kDa_late_protein 31715:32215	hypothetical_protein 34474:34944	hypothetical_protein 33371:33889	13.88_kDa_late_protein 37480:37980	hypothetical_protein 10262:10762
tail_assembly_chaperone 14057:14320	hypothetical_protein 14987:15244	hypothetical_protein 14695:15054	hypothetical_protein 19479:19736	tail_assembly_chaperone 28082:28441
nuclease_superfamily_protein 30161:31402	exonuclease 32621:33862	PD-(D/E)XK_nuclease_superfamily_protein 31473:32714	hypothetical_protein 35672:36910	nuclease 11075:12316
DNA_helicase 23820:25244	DNA_helicase 25778:27166	chromatin_remodeling_complex_ATPase 24916:26340	DNA_helicase 29577:31109	helicase 16985:18409
VRR-NUC_domain_protein 26203:26538	Holliday_junction_resolvase 28368:28652	putative_nuclease 27496:27783	hypothetical_protein 31896:32180	nuclease 15807:16109
hypothetical_protein 7548:8060	head-tail_adaptor_Ad1 8753:9265	hypothetical_protein 8204:8716	hypothetical_protein 12888:13400	head-to-tail_connector_complex_protein 34532:35044
putative_spanin_protein 4272:4658	Rz-like_spanin 5483:5869	putative_spanin 5250:5636	hypothetical_protein 9615:10001	hypothetical_protein 37931:38317
helix-turn-helix_domain_protein 35277:35459	excisionase_and_transcriptional_regulator 38163:38333	helix-turn-helix_domain_protein 37311:37481	hypothetical_protein 41561:41743	hypothetical_protein 7204:7374
hypothetical_protein 31399:31671	hypothetical_protein 33859:34131	hypothetical_protein 33055:33327	hypothetical_protein 37164:37436	hypothetical_protein 10806:11078
hypothetical_protein 18029:18544	minor_tail_protein 19701:20216	hypothetical_protein 18808:19323	hypothetical_protein 23469:23984	minor_tail_protein 23843:24358
putative_tail_protein 3202:3663	Hoc-like_head_decoration_protein 4428:4889	putative_tail_protein 3858:4319	fibritin_protein 8548:9006	fibritin_protein 39260:39721
tail_protein 9533:9946	tail_terminator 10728:11141	hypothetical_protein 10084:10497	hypothetical_protein 14765:15181	hypothetical_protein 32623:33036
hypothetical_protein 26620:26751	hypothetical_protein 28884:29015	hypothetical_protein 27865:27996	hypothetical_protein 32412:32543	hypothetical_protein 15402:15533
scaffold_protein 4850:5551	putative_head_scaffolding_protein 6061:6762	hypothetical_protein 5828:6532	hypothetical_protein 10193:10894	scaffold_protein 37038:37739
hypothetical_protein 42155:42391	hypothetical_protein 44176:44412	hypothetical_protein 43317:43553	hypothetical_protein 3120:3356	NinH_protein 152:388
hypothetical_protein 18541:18906	minor_tail_protein 20213:20578	peptidoglycan_endopeptidase 19320:19685	hypothetical_protein 23981:24346	hypothetical_protein 23481:23846
hypothetical_protein 23686:23823	hypothetical_protein 25414:25551	hypothetical_protein 24270:24407	hypothetical_protein 28931:29068	hypothetical_protein 18918:19055

vB_EcoS-phiEc3	vB_EcoM_SHAK7858	NTEC3	vB_EcoS-UTI89UKE1	vB-Eco-KMB35
tail_protein 19560:22055	tail_protein 19688:22189	tail_protein 30875:33442	hypothetical_protein 20999:23488	tail_protein 19238:21739
hypothetical_protein 14085:14501	tail_assembly_chaperone 14240:14656	hypothetical_protein 38381:38797	hypothetical_protein 16475:16891	tail_assembly_chaperone 13817:14233
putative_tail_protein 10497:11663	minor_tail_protein 10467:11633	putative_tail_protein 41224:42390	hypothetical_protein 12717:13883	tail_protein 10707:11873
hypothetical_protein 33183:33701	hypothetical_protein 31851:32369	hypothetical_protein 20258:20758	13.88_kDa_late_protein 34231:34740	13.88_kDa_late_protein 32667:33167
hypothetical_protein 14507:14866	hypothetical_protein 14764:15021	hypothetical_protein 38019:38291	hypothetical_protein 16990:17256	hypothetical_protein 14332:14595
PD-(D/E)XK_nuclease_superfamily_protein 31285:32526	exonuclease 30288:31538	nuclease_superfamily_protein 21071:22312	hypothetical_protein 32668:33918	hypothetical_protein 31113:32354
chromatin_remodeling_complex_ATPase 24728:26116	DNA_helicase 24873:26297	DNA_helicase 27095:28519	DNA_helicase 26618:28042	DNA_helicase 24277:25701
putative_nuclease 27308:27595	Holliday_junction_resolvase 26525:26809	endonuclease_protein 25805:26089	hypothetical_protein 29078:29362	replication_repair_nuclease_VRR-NUC 26707:26994
hypothetical_protein 8201:8713	head-tail_adaptor_Ad1 8171:8683	head-to-tail_connector_complex_protein 1:513	hypothetical_protein 10326:10838	head-tail_connector_protein 8178:8690
putative_spanin 4928:5314	Rz-like_spanin 4898:5284	putative_spanin_protein 3391:3777	hypothetical_protein 7921:8307	spanin 4905:5291
helix-turn-helix_domain_protein 37123:37293	excisionase_and_transcriptional_regulator 35938:36120	hypothetical_protein 17200:17370	hypothetical_protein 37763:37945	pyocin_activator_protein 36490:36672
hypothetical_protein 32867:33139	hypothetical_protein 31535:31807	hypothetical_protein 20802:21074	hypothetical_protein 33915:34187	hypothetical_protein 32351:32623
hypothetical_protein 18620:19135	minor_tail_protein 18754:19269	hypothetical_protein 33795:34310	hypothetical_protein 20065:20580	hypothetical_protein 18304:18819
putative_tail_protein 3858:4319	Hoc-like_head_decoration_protein 3825:4286	fibritin_protein 4386:4847	fibritin_protein 7117:7578	fibritin_protein 3835:4296
hypothetical_protein 10081:10494	tail_terminator 10048:10464	tail_protein 42393:42806	hypothetical_protein 12298:12714	tail_completion_protein 10291:10704
hypothetical_protein 27677:27808	hypothetical_protein 27041:27172	hypothetical_protein 25442:25573	hypothetical_protein 29594:29725	hypothetical_protein 27076:27207
hypothetical_protein 5506:6210	head_scaffolding_protein 5476:6180	scaffold_protein 2498:3199	hypothetical_protein 8496:9122	scaffold_protein 5483:6184
hypothetical_protein 44087:44323	NinH_family_protein 43289:43525	hypothetical_protein 9656:9892	hypothetical_protein 2096:2332	DNA_binding_protein 42893:43129
peptidoglycan_endopeptidase 19132:19497	minor_tail_protein 19266:19631	hypothetical_protein 33433:33798	hypothetical_protein 20577:20942	murein_endopeptidase_protein 18816:19181
hypothetical_protein 24082:24219	hypothetical_protein 24227:24364	hypothetical_protein 28516:28653	hypothetical_protein 26484:26621	hypothetical_protein 23961:24098

Names of the encoded proteins and positions of their genes in genomes of phages from the fifth clade of the VipTree-based multigene tree presented in Figure 1. These lists were used to generate tree presented in Figure 6.

The importance of the gene for tree topology marked by '+'	No. of parsimony informative sites in alignment	Characteristic protein features identified by InterProScan	The name of gene product and its position in genome of particular phage:					
			myPSH1131	vB_EcoP-101114UKE3	vB_EcoP-101114BS3	MLP3	vB_EcoP-CHD5UKE1	FS2B
+	151	Glutathione synthetase ATP-binding domain-like	ATP-grasp_enzyme 57073:58167	hypothetical_protein 52948:54183	hypothetical_protein 31401:32636	hypothetical_protein 30384:31619	hypothetical_protein 53254:54447	ATP-grasp_enzyme 41535:42728
-	104	-	hypothetical_protein 18536:19786	hypothetical_protein 24204:25871	hypothetical_protein 2657:4324	hypothetical_protein 1846:3513	hypothetical_protein 24124:25773	hypothetical_protein 12710:14365
-	80	-	hypothetical_protein 23310:24263	hypothetical_protein 22159:23211	hypothetical_protein 612:1664	hypothetical_protein 50:853	hypothetical_protein 22077:23129	hypothetical_protein 10663:11715
-	79	DNA/RNA polymerases; DNA polymerase family A	DNA_polymerase 37901:39709	DNA-directed_DNA_polymerase 38440:40284	DNA-directed_DNA_polymerase 16893:18737	DNA_polymerase 16096:17940	DNA-directed_DNA_polymerase 39164:40972	DNA_polymerase-like_protein 27708:29516
-	72	P-loop containing nucleoside triphosphate hydrolases	ATP-binding_protein 33971:34864	dNMP_kinase 32174:32770	dNMP_kinase 10627:11223	hypothetical_protein 9816:10412	dNMP_kinase 32228:33121	ATP-binding_protein 20822:21715
-	58	-	DNA_injection_protein 24709:25635	hypothetical_protein 23221:24189	hypothetical_protein 1674:2642	hypothetical_protein 863:1831	hypothetical_protein 23142:24110	putative_DNA_injection_protein 11728:12696
-	54	-	hypothetical_protein 65234:65596	hypothetical_protein 61464:61826	hypothetical_protein 39917:40279	hypothetical_protein 38900:39262	hypothetical_protein 61717:62079	hypothetical_protein 49998:50360
-	53	Phage phiEco32-like COOH.NH2 ligase-type 2	hypothetical_protein 63909:64508	hypothetical_protein 58477:59283	hypothetical_protein 36930:37736	hypothetical_protein 35913:36719	hypothetical_protein 58728:59534	hypothetical_protein 47009:47815
-	49	-	hypothetical_protein 74914:75279	hypothetical_protein 73547:73873	hypothetical_protein 52000:52326	terminase 50837:51163	terminase 74733:75098	hypothetical_protein 13612:13977
-	48	PIN domain-like; 5' to 3' exonuclease, C-terminal subdomain	5'-3'_exonuclease 27490:28182	exonuclease 31360:32181	exonuclease 9813:10634	exonuclease 9002:9823	exonuclease 31423:32241	5'-3'_exonuclease 20017:20844
-	42	Sigma2 domain of RNA polymerase sigma factors; Homeodomain-like	RNA_polymerase_ECF_sigma_factor 29574:30104	hypothetical_protein 32770:33420	hypothetical_protein 11223:11873	hypothetical_protein 10412:11062	hypothetical_protein 33507:34154	RNA_polymerase_ECF_sigma_factor 22101:22748
-	41	-	hypothetical_protein 30771:31517	hypothetical_protein 34218:34958	hypothetical_protein 12671:13411	hypothetical_protein 11859:12599	hypothetical_protein 35591:36337	hypothetical_protein 23761:24507
-	40	-	hypothetical_protein 61902:62309	hypothetical_protein 57841:58467	hypothetical_protein 36294:36920	hypothetical_protein 35277:35903	hypothetical_protein 58101:58718	hypothetical_protein 46382:46789
-	31	-	hypothetical_protein 36725:37117	hypothetical_protein 37073:37465	hypothetical_protein 15526:15918	hypothetical_protein 14729:15121	hypothetical_protein 37623:38015	hypothetical_protein 26167:26559
-	29	Terminase large subunit, T4likevirus-type, N-terminal; Terminase, large subunit.; Terminase RNaseH-like domain	terminase_large_subunit 151:1131	terminase_large_subunit 2119:3660	terminase_large_subunit 56264:57805	putative_DNA_packaging_protein 55443:56984	terminase_large_subunit 2030:3571	terminase_large_subunit 8195:9736
-	27	-	hypothetical_protein 42675:42842	hypothetical_protein 37749:37913	hypothetical_protein 16202:16366	hypothetical_protein 15405:15572	hypothetical_protein 38248:38415	hypothetical_protein 26792:26959
-	27	DNA ligase-like, N-terminal NAD+-binding domain; DNA ligase/mRNA capping enzyme, catalytic domain	NAD-dependent_DNA_ligase 44735:45064	hypothetical_protein 42503:42823	hypothetical_protein 20956:21276	hypothetical_protein 20159:20479	hypothetical_protein 43707:44036	NAD-dependent_DNA_ligase 32553:32876
-	25	-	hypothetical_protein 30363:30713	hypothetical_protein 33811:34158	hypothetical_protein 12264:12611	hypothetical_protein 11453:11800	hypothetical_protein 34759:35109	Appr-1-p_processing_enzyme_family 23353:23703
-	25	HNHc	terminase_small_subunit 26897:27265	terminase_small_subunit 30767:31171	terminase_small_subunit 9220:9624	terminase_small_subunit 8409:8813	terminase_small_subunit 30842:31243	hypothetical_protein 19436:19837
-	19	-	hypothetical_protein 55506:55748	hypothetical_protein 52679:52921	hypothetical_protein 31132:31374	hypothetical_protein 30115:30357	hypothetical_protein 52999:53247	hypothetical_protein 41280:41528
-	16	-	putative_holin 8701:8919	hypothetical_protein 14701:14919	hypothetical_protein 68846:69064	hypothetical_protein 68577:68795	hypothetical_protein 14635:14853	putative_holin 3221:3439
-	16	Cell Wall Hydrolase	hypothetical_protein 70851:71156	hypothetical_protein 68771:69076	hypothetical_protein 47224:47529	hypothetical_protein 46074:46379	hypothetical_protein 70113:70502	hypothetical_protein 57718:58107
-	13	-	hypothetical_protein 33093:33272	hypothetical_protein 31158:31349	hypothetical_protein 9611:9802	hypothetical_protein 8800:8991	hypothetical_protein 31230:31409	HNH_endonuclease 19824:20003
-	12	dUTPase-like; trimeric_dUTPase; dCTP deaminase-like	deoxycytidine_triphosphate_deaminase 46792:47325	deoxycytidine_triphosphate_deaminase 46841:47374	deoxycytidine_triphosphate_deaminase 25294:25827	deoxycytidine_triphosphate_deaminase 24289:24822	deoxycytidine_triphosphate_deaminase 47666:48199	dCTP_deaminase 36190:36723
-	10	-	hypothetical_protein 41262:41576	hypothetical_protein 41225:41539	hypothetical_protein 19678:19992	hypothetical_protein 18881:19198	hypothetical_protein 42060:42377	hypothetical_protein 31069:31383
-	6	-	hypothetical_protein 66442:66675	hypothetical_protein 62480:62713	hypothetical_protein 40933:41166	hypothetical_protein 39916:40149	hypothetical_protein 62925:63158	hypothetical_protein 50966:51199
-	4	SU10 adaptor protein	hypothetical_protein 4647:5399	hypothetical_protein 9019:9771	hypothetical_protein 63164:63916	hypothetical_protein 62877:63626	hypothetical_protein 8924:9676	hypothetical_protein 2091:2843

Names of the encoded proteins and positions of their genes in genomes of phages from the sixth clade of the VipTree-based multigene tree presented in Figure 1. These lists were used to generate tree presented in Figure 7.

The importance of the gene for tree topology marked by '+'	No. of parsimony informative sites in alignment	Characteristic protein features identified by InterProScan	vB_EcoP-101101UKE1	vB_EcoP-UTI89UKE2	vB_EcoP-101118B1
+	551	Phage lysozyme; Lysozyme-like	hypothetical_protein 26127:29075	hypothetical_protein 25850:28798	baseplate_hub_subunit_/_lysozyme 25484:28426
+	194	-	non-contractile_tail_tubular_protein 22979:25381	non-contractile_tail_tubular_protein 22702:25104	non-contractile_tail_tubular_protein 22353:24758
+	111	-	capsid_assembly_scaffolding_protein 19988:20908	capsid_assembly_scaffolding_protein 19710:20630	capsid_assembly_scaffolding_protein 19370:20278
+	65	-	major_capsid_protein 20985:22187	major_capsid_protein 20707:21909	major_capsid_protein 20358:21557
+	64	T7 virus internal virion protein gp14 family	internal_virion_protein 25381:26127	internal_virion_protein 25104:25850	internal_virion_protein 24758:25483
+	53	Tail tubular protein	non-contractile_tail_tubular_protein 22242:22979	non-contractile_tail_tubular_protein 21965:22702	non-contractile_tail_tubular_protein 21613:22353
+	50	-	hypothetical_protein 6137:6820	hypothetical_protein 6295:6978	hypothetical_protein 6295:7017
+	36	-	hypothetical_protein 11036:11413	hypothetical_protein 10472:10849	hypothetical_protein 10410:10787
+	35	-	hypothetical_protein 36273:36551	hypothetical_protein 35678:35953	hypothetical_protein 37966:38247
+	32	Terminase Bacteriophage T7, Ribonuclease H-like domain	terminase_large_subunit 34218:36116	terminase_large_subunit 33620:35518	terminase_large_subunit 35912:37807
+	31	-	hypothetical_protein 36570:36863	hypothetical_protein 35972:36259	hypothetical_protein 38265:38552
+	30	-	hypothetical_protein 1:924	hypothetical_protein 1:924	hypothetical_protein 1:924
+	27	DNA/RNA polymerases; DNA polymerase family A; Ribonuclease H-like	DNA-directed_DNA_polymerase 7572:10118	DNA-directed_DNA_polymerase 7556:10102	DNA-directed_DNA_polymerase 7565:10138
+	22	RecA-like_Gp4D_helicase; TOPRIM_primases; DnaB-like helicase C terminal domain; DNA primase core; P-loop containing nucleoside triphosphate hydrolases	primase/helicase_protein 4072:6060	primase/helicase_protein 4073:6061	primase/helicase_protein 4073:6061
+	19	NAT_SF; Acetyltransferase (GNAT) family; Acyl-CoA N-acyltransferases (Nat)	hypothetical_protein 17756:18217	hypothetical_protein 17478:17939	hypothetical_protein 17138:17599
+	17	-	hypothetical_protein 18227:18436	hypothetical_protein 17949:18158	hypothetical_protein 17609:17818
+	17	ATP dependent DNA ligase domain; DNA ligase OB-like domain; DNA ligase/mRNA capping enzyme, catalytic domain; Nucleic acid-binding proteins	ATP-dependent_DNA_ligase 16492:17439	ATP-dependent_DNA_ligase 16214:17161	ATP-dependent_DNA_ligase 15874:16821
+	16	-	hypothetical_protein 12630:13010	hypothetical_protein 12460:12831	hypothetical_protein 12089:12457
-	14	5' to 3' exonuclease, C-terminal subdomain	exonuclease 13293:14321	exonuclease 13134:14162	exonuclease 12725:13753
-	13	-	hypothetical_protein 10118:10234	hypothetical_protein 10102:10218	hypothetical_protein 10138:10254
-	10	Metallo-dependent phosphatases	phosphoesterase 14577:15716	phosphoesterase 14418:15560	phosphoesterase 14117:15148
-	7	-	hypothetical_protein 11491:12297	hypothetical_protein 11320:12126	hypothetical_protein 10950:11756
-	7	Hedgehog/DD-peptidase; Peptidase M15	hypothetical_protein 36863:37207	hypothetical_protein 36944:37288	hypothetical_protein 39216:39587
-	5	-	hypothetical_protein 12307:12525	hypothetical_protein 12136:12354	hypothetical_protein 11766:11984
-	2	-	hypothetical_protein 17411:17626	hypothetical_protein 17133:17348	hypothetical_protein 16793:17008
-	2	-	hypothetical_protein 43970:44083	hypothetical_protein 43855:43968	hypothetical_protein 44099:44212
-	1	-	hypothetical_protein 17646:17759	hypothetical_protein 17353:17481	hypothetical_protein 17013:17141
-	1	-	hypothetical_protein 37316:37465	hypothetical_protein 37397:37546	hypothetical_protein 39696:39845

	The name of gene product and its position in genome of particular phage:					
vB_EcoM_SHAK7704	**vB-Eco-KMB47**	**vB-Eco-KMB14**	**vB_EcoP_ACG-C91**	**vB_EcoP-101117UKE2**	**vB_EcoM_SHAK7163**	**vB_EcoP-UTI89UKE3**
lysozyme_domain-containing_protein 36677:39625	hypothetical_protein 28035:30983	baseplate_hub 28361:31309	hypothetical_protein 28598:31546	baseplate_hub_subunit_/_lysozyme 25539:28481	lysozyme_domain-containing_protein 28635:31583	hypothetical_protein 25851:28799
tail_protein 33553:35955	non-contractile_tail_tubular_protein 24887:27289	tail_tubular_protein_B 25237:27639	putative_tail_tubular_B_protein 25468:27870	non-contractile_tail_tubular_protein 22408:24813	tail_protein 25511:27913	non-contractile_tail_tubular_protein 22703:25105
head_scaffolding_protein 30557:31477	capsid_assembly_scaffolding_protein 21896:22816	capsid_assembly_scaffolding_protein 22243:23163	putative_scaffolding_protein 22474:23394	capsid_assembly_scaffolding_protein 19425:20333	head_scaffolding_protein 22511:23431	capsid_assembly_scaffolding_protein 19711:20631
major_head_protein 31554:32762	major_capsid_protein 22893:24095	major_capsid_protein 23240:24445	putative_major_capsid_protein 23471:24676	major_capsid_protein 20413:21612	major_head_protein 23511:24719	major_capsid_protein 20708:21910
internal_virion_protein 35955:36677	internal_virion_protein 27289:28035	internal_virion_protein 27639:28361	putative_internal_virion_protein 27870:28598	internal_virion_protein 24813:25538	internal_virion_protein 27913:28635	internal_virion_protein 25105:25851
tail_protein 32816:33553	non-contractile_tail_tubular_protein 24150:24887	tail_tubular_protein_A 24500:25237	putative_tail_tubular_A_protein 24731:25468	non-contractile_tail_tubular_protein 21668:22408	tail_protein 24774:25511	non-contractile_tail_tubular_protein 21966:22703
hypothetical_protein 17184:17867	hypothetical_protein 8045:8728	hypothetical_protein 8475:9158	hypothetical_protein 9114:9836	hypothetical_protein 6419:7072	hypothetical_protein 8940:9662	hypothetical_protein 6295:6978
polysaccharide_chain_length_determinant_protein 21407:21784	hypothetical_protein 12944:13321	hypothetical_protein 13170:13547	hypothetical_protein 13332:13709	hypothetical_protein 10465:10842	polysaccharide_chain_length_determinant_protein 13363:13740	hypothetical_protein 10473:10850
acetyl-CoA_acetyltransferase 2054:2335	hypothetical_protein 38181:38459	acetyl-CoA_acetyltransferase 38187:38465	hypothetical_protein 38739:39017	hypothetical_protein 37966:38247	acetyl-CoA_acetyltransferase 38470:38751	hypothetical_protein 35679:35954
terminase_large_subunit 1:1896	terminase_large_subunit 36125:38023	DNA_packaging_protein 36131:38029	putative_large_terminase_subunit 36684:38582	terminase_large_subunit 35912:37807	terminase_large_subunit 36417:38312	terminase_large_subunit 33621:35519
acetyl-CoA_acetyltransferase 2280:2636	hypothetical_protein 38478:38771	hypothetical_protein 38484:38777	hypothetical_protein 39030:39347	hypothetical_protein 38265:38552	acetyl-CoA_acetyltransferase 38696:39055	hypothetical_protein 35973:36260
hypothetical_protein 10703:11587	hypothetical_protein 1909:2832	hypothetical_protein 2186:3109	hypothetical_protein 2859:3743	hypothetical_protein 1:924	hypothetical_protein 2599:3570	hypothetical_protein 1:924
DNA_polymerase 18490:21036	DNA-directed_DNA_polymerase 9480:12026	DNA_polymerase 9910:12456	putative_DNA_polymerase 10416:12962	DNA-directed_DNA_polymerase 7620:10193	DNA_polymerase 10447:12993	DNA-directed_DNA_polymerase 7557:10103
DNA_primase/helicase 14962:16950	primase/helicase_protein 5980:7968	DNA_primase/helicase 6253:8241	putative_DNA_helicase 6892:8880	primase/helicase_protein 4073:6061	DNA_primase/helicase 6718:8706	primase/helicase_protein 4073:6061
acetyltransferase 28325:28786	hypothetical_protein 19664:20125	GNAT_family_N-acetyltransferase 20011:20472	putative_N-acetyltransferase 20242:20703	hypothetical_protein 17193:17654	acetyltransferase 20282:20743	hypothetical_protein 17479:17940
hypothetical_protein 28796:29005	hypothetical_protein 20135:20344	hypothetical_protein 20482:20691	hypothetical_protein 20713:20922	hypothetical_protein 17664:17873	hypothetical_protein 20753:20959	hypothetical_protein 17950:18159
ATP-dependent_DNA_ligase 27061:28008	ATP-dependent_DNA_ligase 18400:19347	DNA_ligase 18747:19694	hypothetical_protein 18978:19925	ATP-dependent_DNA_ligase 15929:16876	ATP-dependent_DNA_ligase 19018:19965	ATP-dependent_DNA_ligase 16215:17162
hypothetical_protein 23307:23678	hypothetical_protein 14538:14918	hypothetical_protein 14850:15221	hypothetical_protein 15233:15604	hypothetical_protein 12144:12512	hypothetical_protein 15264:15635	hypothetical_protein 12461:12832
exonuclease 23981:25009	exonuclease 15201:16229	exonuclease 15599:16576	exonucluease 15907:16932	exonuclease 12780:13808	exonuclease 15938:16966	exonuclease 13135:14163
hypothetical_protein 21036:21152	hypothetical_protein 12026:12142	hypothetical_protein 12456:12572	hypothetical_protein 12962:13078	hypothetical_protein 10193:10309	hypothetical_protein 12993:13109	hypothetical_protein 10103:10219
metallo-phosphoesterase 25373:26407	phosphoesterase 16617:17624	metallo-phosphoesterase 16964:17971	hypothetical_protein 17323:18330	phosphoesterase 14172:15203	metallo-phosphoesterase 17222:18364	phosphoesterase 14419:15561
hypothetical_protein 22167:22973	hypothetical_protein 13399:14205	hypothetical_protein 13710:14516	hypothetical_protein 14093:14899	hypothetical_protein 11005:11811	hypothetical_protein 14124:14930	hypothetical_protein 11321:12127
endolysin 3327:3671	hypothetical_protein 38771:39115	endolysin 38777:39121	hypothetical_protein 39340:39684	hypothetical_protein 39216:39587	endolysin 39749:40093	hypothetical_protein 36945:37289
hypothetical_protein 22983:23201	hypothetical_protein 14215:14433	hypothetical_protein 14526:14744	hypothetical_protein 14909:15127	hypothetical_protein 11821:12039	hypothetical_protein 14940:15158	hypothetical_protein 12137:12355
hypothetical_protein 27980:28195	hypothetical_protein 19319:19534	hypothetical_protein 19666:19881	hypothetical_protein 19897:20112	hypothetical_protein 16848:17063	hypothetical_protein 19937:20152	hypothetical_protein 17134:17349
hypothetical_protein 10221:10334	hypothetical_protein 1428:1541	hypothetical_protein 1747:1860	hypothetical_protein 2416:2535	hypothetical_protein 44099:44212	hypothetical_protein 2165:2278	hypothetical_protein 43856:43969
hypothetical_protein 28200:28328	hypothetical_protein 19545:19667	hypothetical_protein 19892:20014	hypothetical_protein 20132:20245	hypothetical_protein 17068:17196	hypothetical_protein 20157:20285	hypothetical_protein 17354:17482
holin 3780:3929	hypothetical_protein 39224:39373	holin 39230:39379	hypothetical_protein 39793:39942	hypothetical_protein 39696:39845	holin 40202:40351	hypothetical_protein 37398:37547

Names of the encoded proteins and positions of their genes in genomes of phages from the seventh clade of the VipTree-based multigene tree presented in Figure 1. These lists were used to generate tree presented in Figure 8.

The importance of the gene for tree topology marked by '+'	No. of parsimony informative sites in alignment	Characteristic protein features identified by InterProScan	RDN8_1	vB_EcoP-22664UKE3-2	vB_EcoM_SHAK7854
+	111	Nucleic acid-binding proteins; DNA ligase/mRNA capping enzyme, catalytic domain; DNA ligase C-terminal domain; ATP dependent DNA ligase domain	ATP-dependent_DNA_ligase 7067:8113	ATP-dependent_DNA_ligase 17061:18164	DNA_ligase 9648:10712
+	55	-	non-contractile_tail_tubular_protein 24360:26729	non-contractile_tail_tubular_protein 33694:36051	tail_protein 26245:28602
+	48	Transglycosylase SLT domain; LT-like; Lysozyme-like; Peptidoglycan transglycosylase gp16.	DNA_ejectosome_component 30137:34024	peptidoglycan_lytic_exo transglycosylase 1:3303	internal_virion_protein_with_endolysin_domain 32030:35917
+	36	DNA/RNA polymerases; DNA polymerase family A; Ribonuclease H-like; DNA-directed DNA polymerase.; DNA-directed DNA polymerase	DNA-directed_DNA_polymerase 14086:16257	DNA-directed_DNA_polymerase 23841:26012	DNA_polymerase_I 16369:18540
+	20	DNA-dependent RNA polymerase; DNA-directed RNA polymerase N-terminal; DNA/RNA polymerases; DNA-directed RNA polymerase	DNA_directed_RNA_polymerase 3518:6199	DNA-directed_RNA_polymerase 13304:15985	N4-like_RNA_polymerase 5889:8570
-	15	Single-stranded DNA-binding protein, Bacteriophage T7; Single-stranded DNA-binding protein.; Nucleic acid-binding proteins	single-stranded_DNA-binding_protein 9539:10237	single-stranded_DNA-binding_protein 19427:20125	Gp2.5-like_ssDNA_binding_protein_and_ssDNA_annealing_protein 12028:12729
-	11	Portal protein.; Bacteriophage head to tail connecting protein	collar/head-to-tail_connector_protein 19938:21506	head-to-tail_connector_protein 28749:30317	head-tail_adaptor 21295:22863
-	9	Endonuclease_I; Phage endonuclease I; Restriction endonuclease-like	endonuclease_I 10273:10692	endonuclease_I 20125:20577	endonuclease 12729:13181
-	8	Bacteriophage T7 DNA helicase/primase, N-terminal a+b fold; Toprim-like; DnaB-like helicase C terminal domain; DNA primase core; TOPRIM_primases; RecA-like_Gp4D_helicase; DNA helicase/primase [4].; P-loop containing nucleoside triphosphate hydrolases; Zinc beta-ribbon; Zinc-binding domain of primase-helicase	primase/helicase_protein 12080:13777	primase/helicase_protein 21900:23600	DNA_primase/helicase 14429:16129
-	7	Phage T7 capsid assembly protein	capsid_and_scaffold_protein 21651:22535	capsid_and_scaffold_protein 30407:31291	head_assembly 22953:23837
-	6	5'-3' exonuclease, N-terminal resolvase-like domain; PIN domain-like; Exonuclease	exonuclease 17291:18163	exonuclease 26493:27365	exonuclease 19021:19893
-	5	PGRP; N-acetylmuramoyl-L-alanine amidase; Endolysin.; N-acetylmuramoyl-L-alanine amidase-like; Endolysin	endolysin 11131:11589	endolysin 21058:21516	amidase 13662:14120
-	3	Holin.; Phage holin T7 family, holin superfamily II	holin_class_II 36601:36798	class_II_holin 6083:6277	holin 38700:38894
-	0	-	hypothetical_protein 16696:16905	hypothetical_protein 26293:26502	hypothetical_protein 18821:19030

The name of gene product and its position in genome of particular phage:								
vB_EcoP-101120B2	**vB_EcoP-101136BS1**	**vB_EcoP-22664BS1**	**vB_EcoM_SHAK9454**	**vB_EcoP-101118UKE1**	**vB_EcoP-101120B1-2**	**vB_EcoP_LNA5**	**EG1**	**vB_EcoP_LNA3**
ATP-dependent_DNA_ligase 16898:18001	ATP-dependent_DNA_ligase 15763:16929	ATP-dependent_DNA_ligase 15809:16912	DNA_ligase 9320:10363	ATP-dependent_DNA_ligase 16227:17330	ATP-dependent_DNA_ligase 16843:17946	DNA_ligase 9936:11012	DNA_ligase 7326:8399	DNA_ligase 22747:23772
non-contractile_tail_tubular_protein 33530:35887	non-contractile_tail_tubular_protein 32622:34979	non-contractile_tail_tubular_protein 32774:35131	tail_protein 26385:28742	non-contractile_tail_tubular_protein 33445:35802	non-contractile_tail_tubular_protein 33476:35833	tail_protein 31264:33621	tail_tubular_protein_B 24478:26862	tail_protein 167:2524
peptidoglycan_lytic_exotransglycosylase 1:3303	peptidoglycan_lytic_exotransglycosylase 1:3303	peptidoglycan_lytic_exotransglycosylase 1:3303	internal_virion_protein_with_endolysin_domain 32160:36047	peptidoglycan_lytic_exotransglycosylase 1:3303	peptidoglycan_lytic_exotransglycosylase 1:3303	internal_virion_protein_with_endolysin_domain 23530:27417	internal_virion_protein_D 30249:34223	internal_virion_protein_with_endolysin_domain 5943:9830
DNA-directed_DNA_polymerase 23678:25849	DNA-directed_DNA_polymerase 22436:24607	DNA-directed_DNA_polymerase 22584:24755	DNA_polymerase_I 16507:18678	DNA-directed_DNA_polymerase 23258:25426	DNA-directed_DNA_polymerase 23623:25794	DNA_polymerase_I 2309:4480	DNA_polymerase 14452:16608	DNA_polymerase_I 29664:31835
DNA-directed_RNA_polymerase 13141:15822	DNA-directed_RNA_polymerase 12070:14751	DNA-directed_RNA_polymerase 12052:14733	N4-like_RNA_polymerase 5552:8233	DNA-directed_RNA_polymerase 12470:15151	DNA-directed_RNA_polymerase 13086:15767	N4-like_RNA_polymerase 12088:14769	DNA-dependent_RNA_polymerase 3742:6393	N4-like_RNA_polymerase 19240:21921
single-stranded_DNA-binding_protein 19264:19962	single-stranded_DNA-binding_protein 18103:18795	single-stranded_DNA-binding_protein 18085:18780	Gp2.5-like_ssDNA_binding_protein_and_ssDNA_annealing_protein 11590:12288	single-stranded_DNA-binding_protein 18502:19194	single-stranded_DNA-binding_protein 19209:19907	Gp2.5-like_ssDNA_binding_protein_and_ssDNA_annealing_protein 8008:8709	single-stranded_DNA-binding_protein 10014:10721	Gp2.5-like_ssDNA_binding_protein_and_ssDNA_annealing_protein 25116:25829
head-to-tail_connector_protein 28586:30154	head-to-tail_connector_protein 27675:29243	head-to-tail_connector_protein 27823:29391	head-tail_adaptor 21433:23001	head-to-tail_connector_protein 28494:30062	head-to-tail_connector_protein 28531:30099	head-tail_adaptor 37000:38568	head-tail_connector_protein 19884:21494	head-tail_adaptor 35178:36746
endonuclease_I 19962:20417	endonuclease_I 18795:19247	endonuclease_I 18780:19232	endonuclease 12721:13173	endonuclease_I 19194:19646	endonuclease_I 19907:20362	endonuclease 7556:8008	endonuclease 10722:11177	endonuclease 25829:26281
primase/helicase_protein 21737:23437	primase/helicase_protein 20495:22195	primase/helicase_protein 20610:22310	DNA_primase/helicase 14138:15838	primase/helicase_protein 20893:22593	primase/helicase_protein 21682:23382	DNA_primase/helicase 4760:6460	primase/helicase_protein 11704:13404	DNA_primase/helicase 27683:29383
capsid_and_scaffold_protein 30244:31128	capsid_and_scaffold_protein 29333:30217	capsid_and_scaffold_protein 29486:30370	head_assembly 23098:23982	capsid_and_scaffold_protein 30157:31041	capsid_and_scaffold_protein 30189:31073	head_assembly 36023:36910	capsid_assembly_protein 21600:22514	head_assembly 36841:37725
exonuclease 26330:27202	exonuclease 25401:26273	exonuclease 25549:26421	exonuclease 19159:20031	exonuclease 26220:27092	exonuclease 26275:27147	exonuclease 956:1828	exonuclease 17521:18423	exonuclease 32925:33794
endolysin 20895:21353	endolysin 19728:20186	endolysin 19873:20226	amidase 13615:14073	endolysin 20127:20585	endolysin 20840:21298	amidase 6617:7075	lysozyme 11177:11632	amidase 26726:27184
class_II_holin 6234:6428	class_II_holin 5693:5887	class_II_holin 5693:5887	holin 38437:38631	class_II_holin 5693:5887	class_II_holin 6234:6428	holin 20946:21140	lysis_protein 36335:36538	holin 13223:13417
hypothetical_protein 26130:26339	hypothetical_protein 25201:25410	hypothetical_protein 25349:25558	hypothetical_protein 18959:19168	hypothetical_protein 26020:26229	hypothetical_protein 26075:26284	hypothetical_protein 1819:2028	hypothetical_protein 17167:17376	hypothetical_protein 32274:32483

Oświadczenia współautorów

Wojciech Wesołowski
.....
IMIĘ I NAZWISKO

Katedra Biologii Molekularnej,
Wydział Biologii,
Uniwersytet Gdański
.....
AFILIACJA

28.09.2026
.....
DATA

OŚWIADCZENIE

Niniejszym oświadczam, iż mój wkład w publikację:

Łukasiak A, Wesołowski W, Neumann J, Lewandowska N, Węglińska E, Bloch S, Węgrzyn G, Nejman-Faleńczyk B. Groundwork for phage therapy: multi-faceted comparative analysis of lytic bacteriophages infecting uropathogenic *Escherichia coli*. Virology. 2026 Feb;615:110738. doi: 10.1016/j.virol.2025.110738,

polegał na pozyskaniu funduszy, poszukiwaniu literatury, zebraniu i interpretacji danych, napisaniu rozdziałów 6 i 7, współudziale w przygotowaniu odpowiedzi na uwagi recenzentów, oraz na wykonaniu opisanych w rozdziale metodycznym analiz bioinformatycznych, w tym na:

- wykonaniu adnotacji ośmiu wcześniej zsekwencjonowanych, lecz niezaadnotowanych genomów fagów (Tabela S1),
- zidentyfikowaniu w analizowanych genomach fagowych potencjalnie interesujących białek o znaczeniu aplikacyjnym, w tym białek biorących udział w procesie lizy i depolimeryzacji (Tabela S2 oraz Tabela S3),
- przeprowadzeniu ujednoliconej metodycznie predykcji cyklu życiowego analizowanych fagów (Tabela S4),
- badanie zależności filogenetycznych między analizowanymi bakteriofagami za pomocą programu VipTree (Rycina 1),
- selekcji genów na podstawie siły sygnału filogenetycznego (Tabela S5) oraz wykonaniu na ich podstawie analizy filogenetycznej i konstrukcji drzew dla 7 zidentyfikowanych kladów (Ryciny 2-8).

Wesołowski
.....
PODPIS

Aleksandra Łukasiak
.....
IMIĘ I NAZWISKO

..... 27.08.2026 .....
DATA

Katedra Biologii Molekularnej,
Wydział Biologii,
Uniwersytet Gdański

.....
AFILIACJA

OŚWIADCZENIE

Niniejszym oświadczam, iż mój wkład w publikację:

Łukasiak A, Wesołowski W, Neumann J, Lewandowska N, Węglińska E, Bloch S, Węgrzyn G, Nejman-Faleńczyk B. Groundwork for phage therapy: multi-faceted comparative analysis of lytic bacteriophages infecting uropathogenic *Escherichia coli*. Virology. 2026 Feb;615:110738. doi: 10.1016/j.virol.2025.110738,

polegał na

(i) przeglądzie i zebraniu danych literaturowych;

(ii) współudziale w omawianiu koncepcji pracy;

(iii) przygotowaniu rozdziałów numer 1, 2, 3, 4 oraz 5;

(iv) przygotowaniu tabeli numer 1 oraz ryciny numer 9;

(v) przygotowaniu rycin numer 1 oraz 2 zawartych w suplemencie;

(vi) współudziale w redagowaniu ostatecznej wersji pracy, po uwagach recenzentów;

(vii) pozyskaniu funduszy.

.....
PODPIS

Julia Neumann
.....
IMIĘ I NAZWISKO

25/08/26r.
.....
DATA

Pracownia Chemoinformatyki Środowiska,
Wydział Chemii,
Uniwersytet Gdański

.....
AFILIACJA

OŚWIADCZENIE

Niniejszym oświadczam, iż mój wkład w publikację:

Łukasiak A#, **Wesołowski W#**, Neumann J, Lewandowska N, Węglińska E, Bloch S, Węgrzyn G, Nejman-Faleńczyk B. Groundwork for phage therapy: multi-faceted comparative analysis of lytic bacteriophages infecting uropathogenic *Escherichia coli*. Virology. 2026 Feb;615:110738. doi: 10.1016/j.virol.2025.110738.

polegał na współudziale w zbieraniu i walidacji danych do analiz filogenetycznych i wizualizacji drzew przedstawionych na rycinach nr 1 i 2.

Julia Neumann
.....
PODPIS

Natalia Lewandowska
.....
IMIĘ I NAZWISKO

24.08.2026
.....
DATA

Katedra Biologii Molekularnej,
Wydział Biologii,
Uniwersytet Gdański

.....
AFILIACJA

OŚWIADCZENIE

Niniejszym oświadczam, iż mój wkład w publikację:

Łukasiak A, Wesołowski W, Neumann J, Lewandowska N, Węglińska E, Bloch S, Węgrzyn G, Nejman-Faleńczyk B. Groundwork for phage therapy: multi-faceted comparative analysis of lytic bacteriophages infecting uropathogenic *Escherichia coli*. Virology. 2026 Feb;615:110738. doi: 10.1016/j.virol.2025.110738,

polegał na współudziale w zbieraniu danych oraz w przygotowaniu i edycji pierwotnej i ostatecznej wersji manuskryptu.

Natalia Lewandowska
.....
PODPIS

Emilia Piwnicka (Węglińska)
.....
IMIĘ I NAZWISKO

01.09.2026
.....
DATA

Katedra Biologii Molekularnej,
Wydział Biologii,
Uniwersytet Gdański

.....
AFILIACJA

OŚWIADCZENIE

Niniejszym oświadczam, iż mój wkład w publikację:

Łukasiak A, Wesołowski W, Neumann J, Lewandowska N, Węglińska E, Bloch S, Węgrzyn G, Nejman-Faleńczyk B. Groundwork for phage therapy: multi-faceted comparative analysis of lytic bacteriophages infecting uropathogenic *Escherichia coli*. Virology. 2026 Feb;615:110738. doi: 10.1016/j.virol.2025.110738,

polegał na współudziale w zbieraniu danych oraz w przygotowaniu i edycji pierwotnej i ostatecznej wersji manuskryptu.

Emilia Piwnicka
.....
PODPIS

Sylwia Bloch
.....
IMIĘ I NAZWISKO

Katedra Biologii Molekularnej,
Wydział Biologii,
Uniwersytet Gdański

.....
AFILIACJA

08.09.2026 r.
.....
DATA

OŚWIADCZENIE

Niniejszym oświadczam, iż mój wkład w publikację:

Łukasiak A, Wesołowski W, Neumann J, Lewandowska N, Węglińska E, Bloch S, Węgrzyn G, Nejman-Faleńczyk B. Groundwork for phage therapy: multi-faceted comparative analysis of lytic bacteriophages infecting uropathogenic *Escherichia coli*. Virology. 2026 Feb;615:110738. doi: 10.1016/j.virol.2025.110738,

polegał na:

(i) współudziale w nadzorowaniu prac związanych z powstawaniem rozdziałów 2-5 niniejszej pracy;

(ii) współudziale w przygotowaniu pierwotnej i ostatecznej wersji manuskryptu.

Bloch
.....
PODPIS

Grzegorz Węgrzyn
.....
IMIĘ I NAZWISKO

27 sierpnia 2026 r.
.....
DATA

Katedra Biologii Molekularnej,
Wydział Biologii,
Uniwersytet Gdański

.....
AFILIACJA

OŚWIADCZENIE

Niniejszym oświadczam, iż mój wkład w publikację:

Łukasiak A, Wesołowski W, Neumann J, Lewandowska N, Węglińska E, Bloch S, Węgrzyn G, Nejman-Faleńczyk B. Groundwork for phage therapy: multi-faceted comparative analysis of lytic bacteriophages infecting uropathogenic *Escherichia coli*. Virology. 2026 Feb;615:110738. doi: 10.1016/j.virol.2025.110738,

polegał na konsultacjach merytorycznych przy interpretacji wyników analiz oraz pomocy w redagowaniu tekstu manuskryptu.

KIEROWNIK
KATEDRY BIOLOGII MOLEKULARNEJ
prof. dr hab. Grzegorz Węgrzyn

.....
PODPIS

Bożena Nejman-Faleńczyk
.....
IMIĘ I NAZWISKO

09.03.2026 r.
.....
DATA

Katedra Biologii Molekularnej,
Wydział Biologii,
Uniwersytet Gdański

.....
AFILIACJA

OŚWIADCZENIE

Niniejszym oświadczam, iż mój wkład w publikację:

Łukasiak A, Wesołowski W, Neumann J, Lewandowska N, Węglińska E, Bloch S, Węgrzyn G, Nejman-Faleńczyk B. Groundwork for phage therapy: multi-faceted comparative analysis of lytic bacteriophages infecting uropathogenic *Escherichia coli*. Virology. 2026 Feb;615:110738. doi: 10.1016/j.virol.2025.110738,

polegał na:

- opracowaniu i konsultowaniu koncepcji badań oraz założeń naukowych niniejszej publikacji,
- współudziale w sprawowaniu merytorycznego nadzoru nad realizacją prac,
- interpretacji uzyskanych wyników oraz formułowaniu wniosków,
- koordynowaniu prac związanych z przygotowaniem manuskryptu i dbaniu o jego spójność merytoryczną i redakcyjną,
- koordynowaniu współpracy pomiędzy współautorami i określeniu zakresu ich pracy,
- przygotowaniu oraz edycji pierwotnej i ostatecznej wersji manuskryptu,
- komunikacji z redakcją czasopisma, w tym prowadzeniu korespondencji na etapie zgłoszenia, recenzowania i publikacji manuskryptu,
- koordynowaniu prac związanych z przygotowaniem odpowiedzi na uwagi recenzentów oraz wprowadzeniem wymaganych zmian do manuskryptu.

B. Nejman-Faleńczyk
.....
PODPIS

Praca nr 2

Significant variability in virulence factors and antibiotic resistance of uropathogenic *Escherichia coli* strains may be counteracted by their susceptibility to lytic bacteriophages.

Wesołowski W[#], Czerwonka G[#], Zegadło K, Jagieła E, Łukasiak A, Grabowski Ł, Mantej J, Dziuba A, Bloch S, Topka-Bielecka G, Necel A, Kostovski D, Węglińska E, Narajczyk M, Węgrzyn G, Nejman-Faleńczyk B, Adamus-Białek W.

Sci Rep. 2026 Jul 21;16(1):26677. doi: 10.1038/s41598-026-62793-9.

[#] co-first authors

Impact Factor: 4.9

Punktacja Ministerstwa Nauki i Szkolnictwa Wyższego: 140

scientific reports

Check for updates

OPEN

Significant variability in virulence factors and antibiotic resistance of uropathogenic *Escherichia coli* strains may be counteracted by their susceptibility to lytic bacteriophages

Wojciech Wesołowski[1,8], Grzegorz Czerwonka[2,8], Katarzyna Zegadło[2], Ernest Jagieła[2], Aleksandra Łukasiak[1], Łukasz Grabowski[1], Jagoda Mantej[3], Anna Dziuba[4], Sylwia Bloch[1], Gracja Topka-Bielecka[5], Agnieszka Necel[6], Daniel Kostovski[1], Emilia Węglińska[1], Magdalena Narajczyk[7], Grzegorz Węgrzyn[1], Bożena Nejman-Faleńczyk[1]✉ & Wioletta Adamus-Białek[4]✉

Uropathogenic *Escherichia coli* (UPEC) strains are the primary cause of urinary tract infections and pose a serious clinical problem due to their remarkable genetic variability, diverse virulence traits, and rapidly increasing antibiotic resistance. In this study, five clinical UPEC isolates (EC3, EC149, EC164, EC179, and EC256) were deeply investigated to better understand their pathogenic potential. The strains differed substantially in genome architecture, mobile genetic elements, plasmid composition, virulence factors composition, and antibiotic resistance profiles, illustrating the highly dynamic nature of UPEC populations. Of particular concern was the detection of plasmids linked to resistance against "last-resort" antibiotics, emphasizing the growing epidemiological risk associated with these pathogens. To address the limitations of conventional therapy, five lytic bacteriophages active against the analyzed UPEC strains were isolated and thoroughly characterized. Comparative genomic and phylogenetic analyses revealed that all phages represent previously undescribed species belonging to the genera *Hanrivervirus*, *Warwickvirus*, *Nonavirus*, *Kagunavirus*, and *Vectrevirus*. Functional assays demonstrated pronounced differences in host range and infectivity. Among them, phage vB_EcoS-149_4M emerged as the most promising candidate, combining broad antibacterial activity with high stability and a favourable safety profile, including the absence of toxin or antibiotic resistance genes and no detectable cytotoxicity toward human cell lines. Taken together, these findings highlight both the complexity and clinical threat posed by UPEC strains and the substantial potential of carefully selected bacteriophages as effective and precise antimicrobial tools. In effect, this multi-faceted approach provides a strong foundation for the development of precision phage therapy against multidrug-resistant UPEC infections.

[1]Department of Molecular Biology, Faculty of Biology, University of Gdansk, Wita Stwosza 59, Gdansk 80-308, Poland. [2]Division of Microbiology, Institute of Biology, Jan Kochanowski University, Uniwersytecka 7, Kielce 25-406, Poland. [3]Chair of Urology and Andrology, Faculty of Medicine, Ludwik Rydygier Collegium Medicum in Bydgoszcz, Nicolaus Copernicus University in Torun, M. Curie Skłodowskiej 9, Bydgoszcz 85-094, Poland. [4]Department of Medical Genetics and Laboratory Diagnostics, Faculty of Medicine, Jan Kochanowski University, IX Wieków Kielc 19a, Kielce 25-317, Poland. [5]Division of Environmental Toxicology, Faculty of Health Sciences with the Institute of Maritime and Tropical Medicine, Medical University of Gdansk, Dębowa 23 A, 80-204, Gdansk, Poland. [6]Department of Medical Microbiology, Faculty of Medicine, Medical University of Gdansk, Dębowa 25, Gdansk 80-204, Poland. [7]Laboratory of Bioimaging, Faculty of Biology, University of Gdansk, Wita Stwosza 59, Gdansk 80-308, Poland. [8]Wojciech Wesołowski and Grzegorz Czerwonka contributed equally to this work. ✉email: bozena.nejman-falenczyk@ug.edu.pl; wioletta.adamus-bialek@ujk.edu.pl

Keywords Uropathogenic *Escherichia coli* (UPEC), Bacterial and bacteriophage genomes, Virulence factors, Antibiotic-resistant genes, Anti-UPEC phages

Abbreviations

ARGs	Antibiotic resistance genes
EOP	Efficiency of plating
ESBLs	Extended-spectrum β-lactamases
HCA	Hierarchical cluster analysis
MFS	Major facilitator superfamily
RND	Resistance-nodulation-cell division
SNP	Single nucleotide polymorphism
UPEC	Uropathogenic Escherichia coli
UTI	Urinary tract infection
VAGs	Virulence associated genes

Urinary tract infections (UTIs) are among the most common bacterial infections affecting millions of individuals worldwide each year. They constitute about 30% of all hospital-acquired infections, and more than 50% of women worldwide will experience a UTI at least once in their lives[1,2]. The national economic burden of uncomplicated UTIs in the US costs approximately $2 billion annually, and several billions more in complicated, recurrent, or hospital-acquired infection[3].The primary causative agent of these infections is uropathogenic *Escherichia coli* (UPEC), a subset of strains that have evolved to thrive in the urinary tract[4,5]. The predominance of *E. coli* among other microorganisms causing UTIs stems primarily from its ability to evade the host's immune response and invade the extremely unfavorable urinary tract environment. Their remarkable adaptability is provided by horizontal gene transfer, in which pathogenic islands play a key role. These islands carry genes involved in adhesion, enabling attachment to epithelial cells and colonization, including genes encoding type 1 (*fim1*), P (*pap*), and S (*sfa*) fimbriae, Dra/Afa (*dra*/*afa*) and Iha (*iha*) adhesins, and Antigen 43 (*flu*). They also harbor genes encoding secreted toxins that modulate immune responses and damage host cells, such as hemolysin A (*hlyA*), cytotoxic necrotizing factor 1 (*cnf1*), and uropathogenic-specific protein (*usp*), as well as genes involved in siderophore biosynthesis and transport, which facilitate efficient iron acquisition (*iuc*, *aer*, *irp*). A feature that particularly favors the maintenance of urinary tract infection is the extraordinary ability of UPEC to cause intracellular bacterial communities (IBCs), which determines the escape from the components of the immune response and antibiotics, while at the same time giving time to activate subsequent mechanisms related to pathogenesis and antibiotic resistance[6,7]. These bacteria have been shown to lose certain virulence-associated genes (VAGs) when acquiring resistance, particularly to fluoroquinolones[8]. It should be noted that despite the highest resistance to penicillins and fluoroquinolones among all antibiotic classes, they are still used in empirical treatment, even despite evidence of cross-resistance[6,8].

All these features make UPEC seem to be the most specific in terms of its ability to cause urinary tract infection. This creates a huge inconvenience and even impairment in social and professional life and permanent loss of health, including increased risk of bladder cancer[9]. First-line drugs for UTI have still remained nitrofurantoin, trimethoprim-sulfamethoxazole and fosfomycin, however it depends on local resistance patterns, medical history, and type of infection. Unfortunately, in the case of chronic UTIs, no antibiotic has been proven to permanently eliminate the infection. Therefore, the search for new therapeutic solutions is extremely necessary[5,6].

Among these, bacteriophage therapy has garnered significant attention as a potential solution to combat UPEC infections[10–12]. Bacteriophages, or shortly phages, are viruses that specifically infect and lyse bacterial cells, making them a promising tool in the fight against antibiotic-resistant bacterial pathogens. UPEC is not a monolithic group but rather consists of various strains that differ in their virulence and resistance profiles[13,14]. This diversity poses a significant challenge for both traditional antibiotic therapies and phage therapy, as a phage that is effective against one UPEC strain may not necessarily be effective against another. Therefore, a comprehensive analysis of multiple UPEC genomes is essential to develop a broad-spectrum phage therapy or, alternatively, to tailor phage cocktails that are effective against a wide range of UPEC strains[15]. In addition to the genomic analysis, this research also considers the potential ecological implications of phage therapy. The introduction of bacteriophages into the urinary tract environment, which is already a complex microbial ecosystem, could have unforeseen consequences on both the bacterial community and the host. Therefore, understanding the interaction between UPEC, its resident phages, and the therapeutic phages is crucial for developing safe and effective treatments.

In the present study, we first selected five clinically relevant UPEC isolates representing diverse antibiotic resistance profiles, virulence gene repertoires, and phylogenetic backgrounds and characterized their genomes in detail. Because this diversity may constitute an important obstacle for phage-based interventions, we subsequently isolated lytic bacteriophages using these same strains as hosts and comprehensively characterized their biological properties. Finally, we evaluated the activity of the isolated phages against a broader collection of UPEC isolates to determine whether they display narrow or broad host ranges and to assess their potential utility as precision antimicrobial agents against multidrug-resistant UPEC populations.

This study employed a combination of bioinformatic tools and genomic sequencing technologies to dissect the genomes of the novel UPEC strains and the lytic phages infecting them, which represent previously undescribed species. Comparative genomics of bacterial strains was used to identify conserved and unique genetic elements across the strains, with particular attention given to genes associated with virulence, antibiotic resistance, and phage susceptibility. Furthermore, the study examined the presence of prophages in UPEC genomes, as these

endogenous phages could influence the outcome of bacteriophage therapy by conferring immunity in the host bacterium.

Materials and methods

Bacterial strains

All 37 *Escherichia coli* strains used in this study (EC1, EC3, EC13, EC22, EC40, EC53, EC57, EC65, EC113, EC124, EC130, EC131, EC134, EC149, EC162, EC164, EC179, EC189, EC192, EC217, EC223, EC227, EC229, EC234, EC255, EC256, EC259, EC263, EC294, EC295, EC296, EC313, EC319, EC341, EC344, EC348, EC378) were isolated from patients with UTI as described previously[6] and stored at -80 °C at the Department of Medical Genetics and Laboratory Diagnostics, Faculty of Medicine, Jan Kochanowski University in Kielce, Poland. The study was approved by the Bioethics Committee at the Collegium Medicum, Jan Kochanowski University in Kielce, Poland (Resolutions: No. 23/2017 of September 11, 2017; No. 22/2024 of March 13, 2024, and No. 35/2025 of June 12, 2025) and all experiments were performed in accordance with relevant named guidelines and regulations. Informed consents were obtained from all patients. Five isolates (EC3, EC149, EC164, EC179, and EC256) were selected for detailed analysis and lytic phages isolation due to their different antibiotic resistance profiles. The characteristics of the selected five *E. coli* strains are presented in Table 1.

Bacterial cultivation and DNA extraction

Bacterial strains used in this study were cultured at 37 °C in liquid Luria-Bertani broth (LB, BioShop) or plated on solid LB medium supplemented with 1.5% agar (LA, BTL). For the DNA isolation procedure, the liquid cultures were grown with aeration in a shaker incubator at 200 rpm and a temperature of 37 °C. DNA was isolated using the NucleoSpin Tissue spin column-based kit (Macherey-Nagel). Isolation was performed according to the manufacturer's protocol. The DNA concentration was determined spectrophotometrically at 260 nm.

Genome sequencing and de-novo assembly

Each bacterial isolate was sequenced commercially in the Genomed company (Warsaw, Poland) using two methods: short reads on the MiSeq platform and long reads on the MinION platform, to increase sequence reliability during genome assembly. The concentration of genomic DNA was measured using the PicoGreen (Invitrogen™) reagent. Genomic DNA was fragmented by sonication using Covaris E210, in accordance with the parameters recommended for preparing libraries for sequencing using MiSeq Illumina technology. Libraries were prepared using the NEBNext® Ultra™ kit II DNA Library Prep Kit for Illumina® according to the manufacturer's recommendations. Sequencing was performed using paired technology 2 × 300 bp ends and

	Strains name	EC3	EC149	EC164	EC179	EC256
Patient Data	Sex	F	F	F	M	F
	Age (years)	85	77	2.5	66	4
	Material	Urine	Urine	Urine	Urine	Urine
Phylogenetic group		A	E	B2	D	F
Antibiotic resistance						
Ampicillin		R	R	S	R	R
Ampicillin sulbactam		R	S	S	S	R
Amoxicillin clavulate		R	S	S	S	R
Piperacillin tazobactam		R	S	S	S	S
Cefuroxim		S	S	S	S	R
Cefotaxim		S	S	S	S	R
Ceftazidime		S	S	S	S	I
Cefepim		S	S	S	S	R
Ertapenem		S	S	S	S	S
Imipenem		S	S	S	S	S
Meropenem		S	S	S	S	S
Aztreonam		S	S	S	S	R
Amikacin		S	S	S	S	R
Gentamicin		S	S	S	S	R
Tobramycin		S	S	S	S	R
Norfloxacin		S	R	R	I	R
Ciprofloxacin		S	R	S	S	R
Nitrofurantoin		S	S	S	S	S
Co-trimoxazole		S	R	S	S	R

Table 1. The characteristics of clinical *Escherichia coli* strains selected for the isolation of lytic phages. Legend: F – female, M – male, S- susceptible, I – intermediate, R- resistance.

Reagent Kit v3 (Illumina). Libraries for the MinION platform were prepared with the Rapid Barcoding Kit (SQK-RBK004) and next sequenced with a SpotON Flow Cell Mk I (R9.4.1).

Raw sequences were assembled by hybrid assembly of Illumina and Oxford Nanopore Technologies reads via the Unicycler program (Galaxy Version 0.5.0 + galaxy1) on default settings, with the exception that the contigs shorter than 500 bp were excluded. Assembled sequences were deposited by the BankIt tool in the GenBank database under the numbers: GCF_036282145.1 for EC3, GCF_036281895.1 for EC149, GCF_036281865.1 for EC164, GCF_036281825.1 for EC179, and GCF_036281835.1 for EC256.

Genomes annotation

Assembled genomes were subjected to annotation in RAST server[16] using the ClassicRAST annotation scheme, FIGfams version 70, automatic error correction, and automatic frame shift correction[17] and additionally by Bakta Web server. Subsequently, according to the submission to NCBI, genomes were annotated using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP)[18].

Phylogenetic analysis of bacteria

Single nucleotide polymorphism (SNP) was the basis of the analysis by CSI phylogeny platform to generate phylogenetic tree. Tree was constructed by CSI Phylogeny algorithms at default settings: minimal depth at SNP positions x10, minimal relative depth at SNP positions 10%, minimal distance between SNPs (prune) 10 bp, minimal SNP quality 30, minimal read mapping quality 25, minimal Z-score 1.96, and without ignoring heterozygous SNPs[19]. The Newick file was uploaded into the iTOL online tool to generate the phylogeny tree based on the SNPs[20]. Genetic distances were calculated by the OrthoAni program[21]. Dendrograms were prepared with the Oat tool[21,22]. Genetic distances between strains were calculated by the online Average Nucleotide Identity (OrthoANI) calculator (EZbioCloud)[22,23].

Virulence-associated genes, CRISPR-cas analysis, and serotyping

The raw FASTA sequence(s) of a genome were screened against the VFDB database[24] using the *E. coli* CFT073 UPEC genome as the reference. The obtained data were tabulated and manually analyzed for virulence-associated genes (VAGs). The assembly was additionally searched for the presence of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) arrays using CRISPRCasFinder (online version; default settings)[25]. Serotypes of the strains were determined from whole-genome sequence data using the SerotypeFinder tool with default settings and the *Escherichia coli* reference database, based on sequence homology of O-antigen and flagellar antigen determinant genes. The analysis was performed using a minimum sequence identity threshold of 85% and a minimum target gene coverage of 60%[26].

Plasmid identification

Assembled circular plasmid contigs were screened through BLAST database for identification of similar plasmid sequences. The highest coverage of sequences pointed the most similar plasmid. Plasmid replicons were searched by PlasmidFinder service (2.0.1 Server 2020-07-01, Database version: 2023-01-18)[27] against the *Enterobacteriales* database, with 95% ientity threshold and 60% mnimum coverage.

Identification of antibiotic resistance genes

Whole-genome sequences of *E. coli* were screened for the presence of antibiotic resistance genes (ARGs) using the Resistance Gene Identifier (RGI) tool implemented in the Comprehensive Antibiotic Resistance Database (CARD)[28]. Genome assemblies were uploaded in FASTA format and analyzed using only Perfect and Strict hits. Results were used to assess the resistome composition of the analyzed *E. coli* genomes.

Prophage sequences identification

The materials used in this study consisted of assembled genome sequences of five *E. coli* strains: EC3, EC149, EC164, EC179, and EC256. The genomic sequences were analysed using the PHASTEST tool[29] to identify prophage regions. Identified prophages were labelled according to the strain name, sequence number in the bacterial genome, and region number in the specific sequence. Only prophages labelled as complete were selected for further analysis. The amino acid sequences of proteins identified as belonging to the holin or endolysin family (lytic cassette proteins) by PHASTEST were transferred to the Phyre2 server[30] to confirm the obtained protein annotations. Simultaneously, an analysis of antibiotic resistance genes within prophages was conducted using the CARD server[28]. The analysis results included only detected antibiotic resistance genes classified as perfect and strict matches.

Sequencing and in silico analyses of phage genomes

Phage genomes of vB_EcoS-149_4M, vB_EcoS-149_6D, vB_EcoS-164_2, vB_EcoS-179_3, and vB_EcoP-179_5 were sequenced using next-generation sequencing technology on an Illumina MiSeq platform. Assembly and annotation of the sequences were accomplished according to the Shen and Millard[31]. The quality of reads was controlled using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The raw data were filtered using a BBtools program. Reads subsampling was performed using Seqtk. *De novo* assembly was conducted using SPAdes[32]. Correction of the assembly was performed by Pilon[33] and prediction of phage termini by PhageTermVirome[34]. Lastly, Pharokka[35] was used to annotate phage genomes, PhaTYP[36] was applied for life cycle identification, and several programs, such as ABRicate (Seemann, 2016), AMRFinderPlus[37], RGI[28], VirulenceFinder[38], and Resfinder[39], were used for identification of antimicrobial resistance or virulence genes. Assembled and annotated phage genomes were deposited in the GenBank database under the following accession numbers: PX985795 for vB_EcoS-149_4M, PX985796 for vB_EcoS-149_6D, PX985797 for vB_EcoS-164_2,

PX985798 for vB_EcoS-179_3, and PX985799 for vB_EcoP-179_5. All genomic sequences are also available in the Zenodo repository under the following access link: https://doi.org/10.5281/zenodo.20098588. Visualization of phage genomes was made using Proksee[40], and genome comparison was performed using EasyFig software[41].

Taxonomic and phylogenetic analysis of phages

To establish the phylogeny relations of the analyzed phages, PhaGCN[42] was initially used to assign them to putative families or genera. For phage vB_EcoS-164_2, BLAST analysis was performed to identify closely related phages, since PhaGCN was unable to assign this phage to any genus or family. The reference phages selected for vB_EcoS-164_2 shared at least 50% squence similarity. Genomic distances among all analyzed phages were calculated using VIRIDIC[43]. In turn, OMA software was applied to establish their common evolutionary origin based on orthologous groups of amino acid sequences of gene products. Multiple sequence alignments were generated using Clustal Omega[44] and subsequently trimmed with trimAl[45] using the "gappyout" option. Each alignment was evaluated for the strength of its phylogenetic signal. Alignments were excluded if they contained fewer than 10 parsimony-informative sites or if the proportion of such sites was below 5%. The remainin orthologous groups were used to construct maximum likelihood phylogenetic trees with IQ-TREE[46]. Tree visualization was performed using MEGA, and tree rooting was determined based on hierarchical cluster analysis (HCA) using a distance matrix generated by VIRIDIC. What is more, in cases where analyzed phages formed species-level clusters, a single representative phage was selected using the medoid of the cluster method. Clustering with complete linkage was carried out using the SciPy library[47].

Isolation, propagation, and stability of phages

All bacteriophages tested in this work were isolated from urban sewage (Gdansk Wastewater Treatment Plant, Poland) and propagated according to the protocol described by Topka and colleagues[48]. Newly isolated phages were deposited in the collection of phage strains at the Department of Molecular Biology of the University of Gdansk, Gdansk, Poland. The double overlay plaque assay was employed to determine a titer of phage suspensions. Serial 10-fold dilutions of each phage suspension were prepared in TM buffer (10 mM Tris-HCl, 10 mM $MgSO_4$; pH 7.2). From each suspension, 5 µL was spotted onto a double agar layer. After overnight incubation of plates at 37 °C, plaques were counted, and the phage titer (expressed as PFU/mL) was calculated. Phage stability assays were performed as described previously[49].

Efficiency of plating (EOP) determination

All five phages were tested against its primary UPEC host (EC149, EC164, or EC179) and thirty-six other UPEC clinical isolates listed above, for a more thorough evaluation of productive infection as defined by the efficiency of plating (EOP) method proposed by[50].

Analysis of plaque and virion morphology

Plaque morphology was determined using the double-layer agar method and phage lysates serially diluted in TM buffer as described Topka[48].Transmission electron microscopy (TEM) photos of the phage virions' morphology were performed in the Bioimaging Laboratory, Faculty of Biology, University of Gdansk, Gdansk, Poland according the procedure presented by Topka-Bielecka and colleagues[49].

Assessment of phage cytotoxicity

Purified phage particles were first obtained using the caesium chloride density gradient centrifugation method according to the protocol provided by[51]. Next, the level of endotoxins in each phage suspension was measured using the Thermo Scientific Pierce chromogenic endotoxin quant kit (Thermo Scientific). Finally, the safety and the impact of the purified phage particles on the viability of the human bladder cancer cell line, T24, and the human embryonic kidney line, HEK293, were measured, using the Neutral Red Uptake (NRU) cytotoxicity test. The NRU cytotoxicity test was performed in accordance with the following protocol. Mammalian cells were grown in McCoy's 5 A Medium (for T24, Avantor) or Dulbecco's Modified Eagle's Medium (for HEK293, Gibco) containing 10% FS (Gibco), 1% pnicillin (Gibco), and 1% sreptomycin (Gibco). For the assay, mammalian cells were seeded into 96-well plates at a density of 1×10^4 cells per well and allowed to attach for 24 h at 37 °C and 5% CO_2. After 24 h incubation, the medium was removed from all wells, and cells were exposed to 200 µL of the mix solution containing culture medium and selected phage particles (in the highest available concentrations, 10^9 PFU/mL for phage vB_EcoS-149_4M; 10^{10} PFU/mL for phages vB_EcoS-149_6D and vB_EcoS-179_3; 10^{11} PFU/mL for phages vB_EcoS-164_2 and vB_EcoP-179_5). In addition, 1% Titon X-100 (Merck) and TM buffer were used as positive and negative controls, respectively. Plates were kept overnight in an incubator (37 °C, 5% C_2). Following 24 h incubation, the mixed solution was removed, and the mammalian cells were washed with 150 µL of PBS (Invitrogen) and treated with 100 uL of neutral red dye solution (0.33%, hermo Scientific), which efficiently penetrates membranes of viable cells. After 45 min. incubation in the dark at 37 °C, the solution was removed, and the cells were washed with 150 µL of PBS. Then, 100 µL solution containing 50% ehanol (Chempur) and 1% gacial acetic acid (Merck) in water was added to each well to extract the dye into the solution. The amount of released neutral red (proportional to the amount of viable cells) was measured as A540 in the Viroskanmicroplate reader. The cell viability was expressed as a percentage of the control values. The assay was performed in triplicate. The tested solution was considered cytotoxic if the cell viability was below 70%. ontrol and phage-treated T24 and HEK293 cells were also examined for morphological changes and cell growth inhibition under a light microscope at x10 magnification (OLYMPUS IX51 microscope). Cells were photographed at time zero and 24 h upon treatment with phage suspensions. Stability of phage particles during incubation with mammalian cells was quantified during this assay accordingly to the procedures described in Sect. 11 and expressed as PFU/mL.

Statistical analyses

Statistical analyses were performed using JASP software (version 0.97.1). Data are presented as means ± standard deviation (SD). Comparisons between two groups were performed using a Student's t-test. Experiments involving more than two groups or factors were analyzed using a two-way analysis of variance (two-way ANOVA), followed by Tukey's honestly significant difference (HSD) post hoc test when appropriate. The assumptions of ANOVA were evaluated by examining residual normality using Q–Q plots and homogeneity of variances using Levene's test. Differences were considered statistically significant at $p < 0.05$.

Results

Genomes' overview

Previously described urinary isolates of *E. coli*[6] were selected for this study based on their diverse drug resistance profiles: highly resistant (MDR strain), highly susceptible (resistance only to ampicillin), and mixed drug resistance (resistance to penicillins, fluoroquinolones, co-trimoxazole, and others). Additionally, a diverse profiles of virulence genes (presence of the 6 previously studied VAGs, or the absence of any, or a varying number of VAGs) and phylogenetic groups (A, B2, E, D, F) were also considered. These parameters were intended to aid in the search for the largest possible number of phages for a diverse group of strains. This diversity gave us the basis to hypothesize that we will find even more interesting features that may determine relationships with phages. The genomes of 5 strains (Table 1) maintaining the above-mentioned criteria, were sequenced in this work and compared with previously performed genetic analyses (CRISPR, VAGs, and phylogenetic groups by PCR) and phenotypic analyses (drug resistance testing by the automated VITEK 2 apparatus using AST Cards, BioMerieux). All studied *E. coli* strains revealed small genomic differences, including GC content, genome size, number of RNAs, and number of coding sequences, according to the Rapid Annotation Subsystem Technology (RAST) server annotation (Table 2). The genome sizes ranged from approximately 4.59 Mb to 5.51 Mb, with the largest genome being EC256 and the smallest EC3. The GC content across all genomes was relatively consistent, hovering around 50.5% o 50.8%.The number of subsystems varied slightly, with EC256 showing the highest count at 611 and EC3 the lowest at 585. The number of coding sequences (CDS) also varied, from 4,361 in EC3 to 5,469 in EC164. Lastly, the number of RNAs ranged from 79 in EC3 to 118 in EC164, indicating some variation in RNA content among these genomes.

Phylogenetic analysis

The phylogeny analysis of tested strains was performed by calculation of the OrthoANI values between genome sequences and the reference genome (K-12 substr. MG1655, RefSeq: GCF_000005845.2). The comparison revealed the similarity of genomes in range from 96.96% for EC164 to 99.38% for EC3, which indicated that all of them belong to the *E. coli* species (Table 3). Whole-genome relatedness among the investigated *Escherichia coli* strains was assessed using OrthoANI values and isolates shared OrthoANI values above 96.7%, confirming their assignment to the same species (Fig. 1). However, notable differences in genomic similarity were observed among the strains. The highest similarity was detected between EC3 and the reference strain *E. coli* K-12 MG1655 (99.43%), indicating a very close phylogenetic relationship. Similarly, EC149 also exhibited high similarity to MG1655 (99.01%). In contrast, the remaining UPEC isolates (EC164, EC179, and EC256) showed lower similarity to MG1655, with OrthoANI values ranging from 97.04% to 97.54%. Comparison with the well-characterized uropathogenic strain CFT073 revealed a different pattern. EC164 displayed the highest genomic similarity to CFT073 (98.86%), suggesting a closer relationship to the UPEC lineage. In contrast, EC3, EC149, EC179, and EC256 showed lower similarity to CFT073, with OrthoANI values ranging from 96.75% to 97.21%. Additionally, we have constructed a phylogenetic tree for the studied UPEC strains to investigate their relationships and assess genomic diversity through SNP analysis of five UPEC genomes. The SNP-based phylogenetic analysis revealed that the UPEC isolates formed distinct lineages relative to the reference strain. Isolate EC164 clustered most closely with the well-characterized UPEC strain CFT073, suggesting a high degree of genomic relatedness. This cluster was further associated with EC256, while EC179 and EC149 occupied progressively more distant branches. EC3 formed a separate lineage, showing the greatest divergence from the remaining UPEC isolates. The high bootstrap support values (0.955–1.000) for the major nodes indicate strong confidence in the inferred phylogenetic relationships. Overall, the SNP tree demonstrates substantial genomic diversity among the analyzed UPEC strains, with EC164 exhibiting the closest evolutionary relationship to the reference UPEC strain CFT073 (Fig. 2).

Category	EC3	EC149	EC164	EC179	EC256
GeneBank No.	GCF_036282145.1	GCF_036281895.1	GCF_036281865.1	GCF_036281825.1	GCF_036281835.1
Genome size (Mb)	4,589,678	4,860,575	5,463,518	5,243,574	5,506,402
GC content	50.8	50.7	50.6	50.6	50.5
Numer subsystems	585	593	602	610	611
Numer of coding sequences	4361	4712	5469	5022	5358
Number of RNAs	79	113	118	110	111

Table 2. General attributes of five UPEC genomes according to RAST server analysis.

Metric	Value				
	EC3	EC149	EC164	EC179	EC256
OrthoANIu value (%)	99.38	99.00	96.96	97.53	97.28
Genome MG1165 (bp)					
UPEC genome length (bp)	4,524,720	4,858,260	5,462,100	5,241,780	5,503,920
Average aligned length (bp)	2,974,079	3,052,734	2,892,033	3,068,755	2,989,802
MG1165 Genome coverage (%)	64.08	65.78	62.31	66.12	64.42
UPEC genome coverage (%)	65.73	62.84	52.95	58.54	54.32

Table 3. Overall genome relatedness expressed as calculated nucleotide identities of orthologous fragment pairs by OrthoANI algorithm for UPEC genomes compared to reference strain K-12 substr. MG1655 (RefSeq: GCF_000005845.2) (www.ezbiocloud.net/tools/ani).

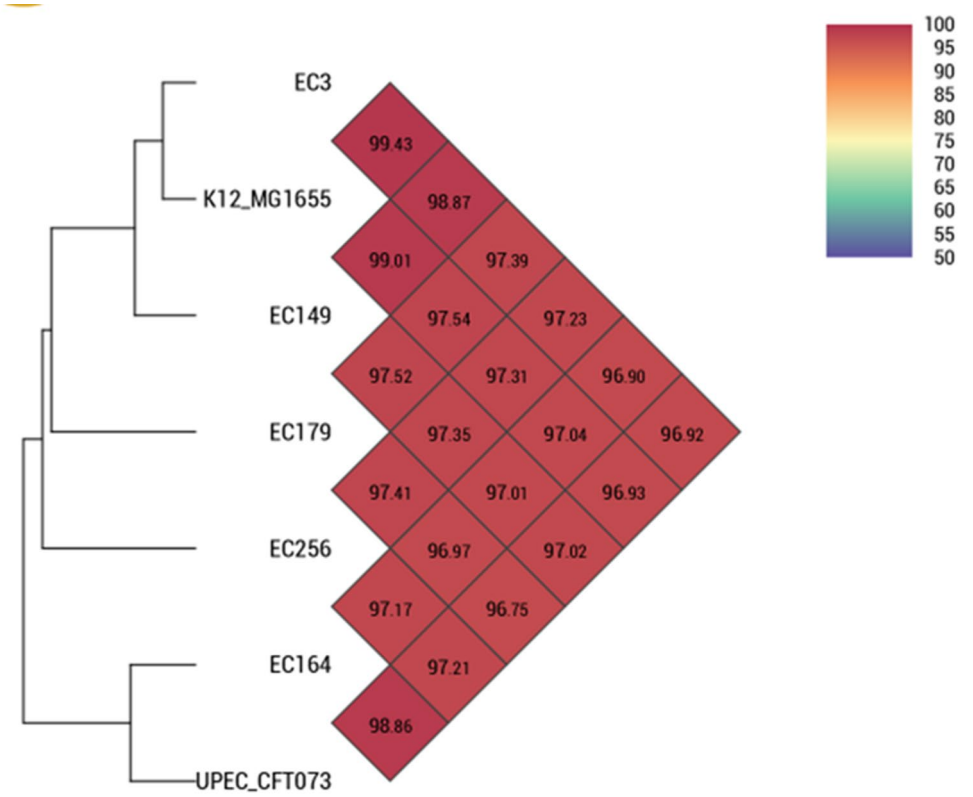


Fig. 1. UPEC genome relatedness based on orthologous average nucleotide identity (OrthoANI) values. The reference genomes belong to the MG1655 and CFT073 UPEC strains. The dendrogram was created using Oat 0.93 software[21].

Plasmid identification

Because plasmids can encode phage defense systems and influence susceptibility to bacteriophage infection, characterization of the plasmid content of UPEC strains is important for evaluating their suitability as targets for phage therapy and for identifying potential barriers to phage-mediated bacterial eradication. Genomic sequences were analyzed using the PlasmidFinder and BLAST tools, which led to the identification of 13 unique replicons among the tested UPEC *E. coli* strains. All tested UPEC strains carried at least one plasmid (Table 4). The plasmid pEC3-1 (41,901 bp) identified in host EC3 contains the IncI1-I(Alpha) replicon, exhibiting 100% identity and coverage with its closest match, plasmid R64 from *Salmonella enterica* subsp. *enterica* serovar Typhimurium. The pEC149-1 plasmid, with the pO111 replicon, showed 98.87% identity with the reference, suggesting a specific plasmid associated with serotype O111. Furthermore, the EC149 strain harboured three additional replicon sequences, where one was similar to the IncB/O/K/Z group (pEC149-2 plasmid), and the replicons of the two remaining plasmids were unidentified. The two plasmids identified in EC164 were pEC164-1 (185,871 bp) classified as IncFIB (98.51%), showing 99.93% similarity to pMTY17668_IncF, and pEC164-2 (98,778 bp) classified as IncY (99.87%), showing 99.41% similarity to pQ502. The EC179 strain also revealed two plasmids, one with replicon likely belonging to the IncFIA (99.72%) group, while the second remained

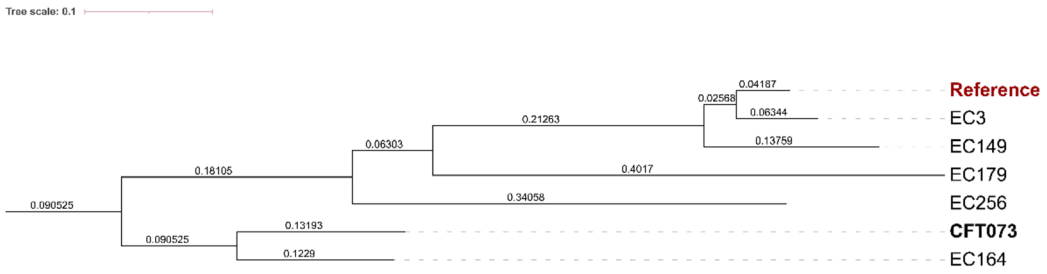


Fig. 2. Phylogenetic tree based on the SNP counts per substitution within genomes of the five UPEC isolates. The reference marked in red belongs to the MG1655 strain.

No	Name	Host name	Length	Plasmid replicon* (identity: %)	Closest similar plasmid	Coverage (%)	Identity (%)	Accession number
1.	pEC3-1	EC3	41,901	IncI1-I(Alpha) (100)	plasmid R64 (*Salmonella enterica* subsp. enterica serovar Typhimurium)	100	100	AP005147.1
2.	pEC149-1	EC149	99,554	pO111 (98.87)	pMTY10828_p0111	92	97.99	CP134391.1
3.	pEC149-2		91,896	IncB/O/K/Z (100)	pHUSEC41-1	100	99.73	HE603110.1
4.	pEC149-3		30,174	Unknown	pSCU-171-2	97	99.09	CP054365.1
5.	pEC149-4		6647	Unknown	pFF2-1 h-4	100	99.98	CP097202.1
6.	pEC164-1	EC164	185,871	IncFIB (98.51); IncFIC (FII – 95.79)	pMTY17668_IncF	98	99.93	CP134354.1
7.	pEC164-2		98,778	IncY (99.87)	pQ502	90	99.41	CP127257.1
8.	pEC179-1	EC179	163,705	IncFIA (99.72); IncFIB(98.34); IncFIC(FII – 95.79)	pMEI008_1	96	100	CP074681.1
9.	pEC179-2		1551	Unknown	pMUB-MIN6-6	100	100	CP069699.1
10.	pEC256-1	EC256	136,868	IncFII (100); IncFIB (98.51)	pCE0026f1	100	99.98	CP119001.1
11.	pEC256-2		90,451	IncM2 (99.83)	RCS40_p	95	99.99	LT985241.1
12.	pEC256-3		34,051	Unknown	p2-6712.08 (*Klebsiella pneumoniae*)	100	97.87	CP083002.1
13.	pEC256-4		2462	Unknown	pMB8775_4	100	98.26	CP134922.1

Table 4. Characteristics of identified plasmids. * according to PlasmidFinder tool.

unidentified. The EC256 strain carried four plasmids, with the identification results indicated that pEC256-1 and pEC256-2 belonged to the IncFIB and IncM2 groups, respectively, while pEC256-3 and pEC256-4 were classified as "unknown" with sequence similarities to *Klebsiella pneumoniae*.

Detection of virulence associated genes

Genes encoding adhesins

A comparative analysis of five UPEC genomes revealed the presence of several VAGs linked to adhesion, exhibiting varied distribution across strains (Table 5). Colonization factor antigen I (CFA/I) was identified exclusively in strain EC256, indicating its potential strain-specific role in adherence. Conversely, *E. coli* common pilus (*ECP*), a conserved adhesin, was prevalent across all genomes, where EC3 and EC149 displaying partial presence, reinforcing its ubiquitous role in UPEC adhesion. Similarly, *HCP* and *upaG*/*ehaG*, key contributors to host cell attachment, were consistently identified in all genomes, emphasizing their significance as core adhesion factors.

The presence of fimbrial adhesins varied among strains. A *fimH*, part of the type 1 fimbrial system, was partially represented in all genomes highlighting its fundamental role in urinary tract colonization. A *pap* fimbriae, linked to pyelonephritis, were detected in EC164 and EC256, albeit partially, suggesting their involvement in more severe infections. Neither *sfa* nor *foc* fimbriae genes were detected in the analyzed genomes. Other adhesion factors showed differential distribution. Gene encoding EaeH protein was common in most strains except EC3, while genes coding ELF pilus was identified in EC3, EC149, and EC256, pointing to strain-dependent roles in pathogenesis. Interestingly, *pilQRW* were found in EC3 and EC149. This diversity in adhesion-related VAGs underscores the adaptability of UPEC strains in urinary tract colonization and pathogenesis, driven by a repertoire of strain-specific and conserved adhesins.

Genes encoding autotransporters

We examined 10 virulence-associated autotransporter genes (Table 5). The genes analyzed included *aatA*, *agn43*, *cah*, *cdiA*, *ehaB*, *air/eaeX*, *pic*, *sat*, *upaH*, and *vat*, which are involved in adhesion, biofilm formation, tissue colonization, and host cell damage. The gene distribution varied across strains, with *cah* (biofilm formation) present in two genomes (EC3 and EC164) and *ehaB* (adhesion and biofilm development) found in three (EC164, EC179, and EC256).

VAG	Function	Escherichia coli				
		EC3	EC149	EC164	EC179	EC256
		Serotype				
		O107:37	O7:H4	O45:H7	O15:H6	O25:H17
cfaABCDE	Adhesion	-	-	-	-	+
ecpABCDER		(+)	(+)	+	+	+
elfACDG		+	+	-	-	+
eaeH		-	+	+	+	+
focACDFGHI		-	-	-	-	-
hcpABC		+	+	+	+	+
papABCDEFGHIJKX		-	-	(+)	-	(+)
sfaABCDEFGHS		-	-	-	-	-
fimABCDEFGHI		(+)	(+)	(+)	+	(+)
pilQRW		+	(+)	-	-	-
flgC		-	-	-	-	-
upaG/ehaG		+	+	+	+	+
aatA	Autotransporter	-	+	-	-	+
agn43		-	-	-	+	-
cah		+	-	+	-	+
cdiAB		-	-	-	-	-
ehaAB		-	-	(+)	(+)	(+)
air/eaeX		-	-	-	+	+
pic		-	-	-	-	-
sat		-	-	-	-	-
upaH		-	-	-	-	-
vat		-	-	+	+	-
ibeABC	Invasion	(+)	(+)	(+)	(+)	(+)
tia		-	-	+	-	+
iucABCD	Iron uptake	-	-	-	+	+
chuASTUWXY		-	-	+	+	+
ireA		-	-	+	-	+
sitABCD		-	+	+	+	+
iroBCDEN		-	-	+	+	+
fyuA		-	-	+	+	+
irp12				+	+	+
ybtAEPQSTUX				+	+	+
espL	Non-LEE encoded TTSS effectors	(+)	(+)	-	(+)	(+)
espR		(+)	(+)	-	(+)	(+)
espX		(+)	(+)	(+)	(+)	(+)
espY		-	-	-	(+)	(+)
aec (T6SS)	Secretion	(+)	-	(+)	(+)	(+)
sci (T6SS)		-	-	(+)	-	-
hlyABCD	Toxin	-	-	-	-	-
hlyE/clyA		+	-	+	+	+
usp		-	-	+	-	-
cnf1		-	-	-	-	-
senB		-	-	-	-	-
galE	Capsule (Immune evasion)	+	-	-	+	+
other	LOS	+	-	-	-	-
other	Capsule (Antiphagocytosis)	+	-	-	-	-

Table 5. Identified virulence associated genes (VAGs) in sequenced genomes of *E. coli* strains. Legend: +: presence of the genes; (+): lack of at least one of the genes in the operon; -: absence of the genes.

Invasion and iron uptake associated genes

A comparative analysis of VAGs related to invasion and iron uptake in five UPEC genomes revealed distinct gene distributions, underscoring their potential roles in UPEC pathogenesis (Table 5). The invasion-related gene *ibe* was consistently present across all strains, indicating its pivotal role in facilitating bacterial entry into host

cells. In contrast, *tia* was variably distributed, appearing in EC164, and EC256 but absent in others, suggesting strain-specific differences in invasion mechanisms. Genes associated with iron uptake, a critical function for bacterial survival in iron-limited environments like the urinary tract, exhibited considerable variability. The siderophore synthesis genes *iuc*, *chu*, *ireA*, *sit*, *iro*, and *fyuA/irp/ybt* were inconsistently distributed among strains. For instance, *chu* and *iro* were present in most strains except EC3 and EC149, whereas *iuc* was found only in EC179 and EC256. Similarly, *ireA* was uniquely present in EC164, highlighting its specific adaptation. The iron transporter *sit* was more widespread, detected in four of the five genomes (except EC3).

Secretion systems and toxins associated genes

A detailed analysis of VAGs linked to type III secretion system (TTSS) effectors, secretion systems, and toxins was conducted (Table 5). TTSS effector genes, including *espL*, *espR*, *espX*, and *espY*, were variably present among the strains. *espL* was partially identified in four strains, while *espR* and *espX* were found in four strains, with *espX* is present (partially) in all five strains. The *espY* gene was detected only in strains EC179 and EC256. These findings suggest variability in TTSS effector gene distribution among UPEC strains. Regarding secretion systems, genes associated with the T6SS systems, such as *aec ACE T6SS* (present in all genomes except EC149) and SCI-I T6SS (present only in genome EC164) were heterogeneously distributed. Toxin-related genes exhibited diverse profiles. The classical hemolysin operon *hlyABCD* was absent from all genomes, whereas *hlyE/clyA* was detected in four strains (except EC149). Other toxins, such as *usp*, *cnf1*, and *senB*, were found sporadically. These results highlight the heterogeneous distribution of VAGs among UPEC strains, reflecting their diverse pathogenic capabilities in urinary tract infections.

Capsule and LOS synthesis genes

Genes involved in capsule production, which play a critical role in immune evasion and antiphagocytic activity, were identified in selected strains. The *galE* gene, associated with capsule biosynthesis and immune evasion was present in EC3, EC179, and EC256, but was absent in the other strains, suggesting strain-specific utilization of this mechanism. Another capsule-related gene linked to antiphagocytic functions was identified only in EC3, highlighting its unique strategy among the analyzed genomes. In terms of LOS biosynthesis, a critical virulence factor for bacterial survival and host interaction, one gene was found in EC3. Its absence in the remaining strains points to alternative mechanisms compensating for LOS-related functions in these UPEC isolates.

Detection of phage associated regions

CRISPR/Cas sequences revealed multiple CRISPR loci with varying spacer numbers and repeat sequences. Evidence levels varied with some strains showing high-confidence CRISPR/Cas systems, notably Type IF Cas genes in EC164, and Type IE Cas genes in EC179 and EC256. The obtained CRISPR sequences are listed in Table 6. Importantly, the sequences of phages isolated in this work weren't found within the spacer regions of the identified CRISPRs.

The conducted research aimed to identify prophage regions in strains of bacteria from the species *E. coli* (EC3, EC149, EC164, EC179, EC256) and characterize the general parameters of these regions. Additionally, the analysis aimed to check the presence of genes encoding proteins associated with the lytic cassette and genes responsible for antibiotic resistance in the detected prophage regions. A description of the identified prophage areas in the analyzed *E. coli* strains was developed, along with the results of the search for sequences encoding potential antibiotic resistance proteins and those included in the phage lytic cassette. The analysis results were presented in Table 7. No prophage regions were detected in the genome of strain EC3. Names for individual prophage regions were created by combining the strain name from which they originated, the sequence number from the given strain with PHASTEST, and the region number for the specific sequence. Some of the proteins encoded by the analyzed sequences were labelled by the PHASTEST and Phyre2 services as putative. This fact has been considered in the presented results. Special attention should be paid to prophages labelled as EC.179.2(1) and EC.256.1(2). Within the first one, a gene conferring resistance to β-lactam antibiotics was identified. This sequence was marked by the RGI CARD service as a perfect match (100% agreement) and identified as the gene encoding TEM-1 β-lactamase. Region EC.256.1(2) contains a sequence encoding the efflux pump *emrE*, described in the genome of *E. coli* and marked as a close match (98.18% agreement) in the aforementioned server.

Detection of antibiotic resistance genes

Analyzed genomes showed a total of 81 ARGs were detected, with various mechanisms of antibiotic resistance identified across different UPEC strains (Supplementary Table 1). The majority of identified ARGs were associated with efflux pumps, particularly the resistance-nodulation-cell division (RND) and major facilitator superfamily (MFS) types. The identified genes contribute to multidrug resistance through different mechanisms. The efflux pump genes *acrB* and *mdtG* encode transporters involved in the extrusion of a variety of toxic compounds and antibiotics from bacterial cells. In contrast, *marA* and *hns* encode global transcriptional regulators that modulate the expression of multiple resistance-associated genes, including efflux systems and membrane permeability determinants. Together, these genes may contribute to reduced susceptibility to several classes of antimicrobials, including fluoroquinolones, tetracyclines, macrolides, and other structurally diverse compounds. Several identified genes encoded β-lactamases (e.g., *TEM-1*, *CTX-M-3*, *SHV-1*) which inactivate β-lactam antibiotics such as penicillins and cephalosporins. Genes for target modification and protection were also detected: *sul1*, *dfrA1*, *armA* modify antibiotic targets, rendering them less susceptible to drugs like sulfonamides and trimethoprim. Additionally, we identified the mutations in *gyrA* and *parC* which contributes to fluoroquinolone resistance. Many ARGs confer resistance to multiple antibiotic classes. For instance, *marA* and *acrB* provide resistance to fluoroquinolones, cephalosporins, and tetracyclines. Overall, the genetic analysis

Strain	CRISPR Id / Cas Type	Start	End	Spacer/Gene	Repeat consensus / cas genes	Direction	Evidence Level
EC3	EC3-1	239,143	239,843	11	CGGTTTATCCCCGCTGGCGCGGGGAACTC	-	4
	EC3-2	239,949	248,599	8	cas2_TypeIE, cas1_TypeIE, cas6_TypeIE, cas5_TypeIE, cas7_TypeIE, cse2_TypeIE, cse1_TypeIE, cas3_TypeI		
	EC3-3	265,390	266,641	20	CGGTTTATCCCCGCTGGCGCGGGGAACAC	-	4
	EC3-4	3,289,243	3,289,390	2	GTTCACTGCCGTACAGGCAGCTTAGAAA	+	1
EC149	EC149-1	1,415,474	1,415,618	1	CACAATGCCTGATGCGACGCTGGAGCGTCTTATCATGCCTACAAA	+	1
	EC149-2	1,485,190	1,485,462	4	CGGTTTATCCCCGCTGGCGCGGGGAACAC	-	3
	EC149-3	1,503,696	1,504,518	13	CGGTTTATCCCCGCTGGCGCGGGGAACAC	-	4
	EC149-4	3,694,018	3,694,164	1	TCCGGCCTACGGATGGCGCGAGAATTTGTAGGCCTGATAAGACGCG	ND	1
	EC149-5	3,722,732	3,722,876	1	ATGCCTGATGCGACGCTTGCCGCGTCTTATCAGGCCTACAAAA	ND	1
EC164	EC164-1	491,146	491,269	1	CGACCCCCACCATGTCAAGGTGGTGCTCTAACCAACTGAGCTA	ND	1
	EC164-2	912,967	913,049	1	TGGACACTGGCTGGACATCGGCTGGACA	+	1
	EC164-3	1,196,045	1,196,136	1	CCACCTTTTTTACCTGCTTCAGATGC	ND	1
	EC164-4	1,317,681	1,318,069	6	TTTCTAAGCTGCCTGTACGGCAGTGAAC	-	4
	CAS-TypeIF	1,318,610	1,326,957	6	cas6_TypeIF, csy3_TypeIF, csy2_TypeIF, csy1_TypeIF, cas3-cas2_TypeIF, cas1_TypeIF		
	EC164-5	1,327,130	1,327,337	3	TTTCTAAGCTGCCTGTACGGCAGTGAAC	-	1
	EC164-6	1,487,869	1,487,999	1	TTGCCTGATGCGACGCTTGCGCGTCTTATCAGGCCT	ND	1
	EC164-7	2,485,527	2,485,676	1	TGAACGCCTTATCCGACCTACACAGCACTGAACTCGTAGGCCTGATAAGACGCG	ND	1
EC179	EC179-1	411,426	411,575	1	CGCGTCTTATCAGGCCTACGAGTTCGGTGCTGTGTAGGTCGGATAAGGCGTTCA	ND	1
	EC179-2	898,813	898,954	1	GTTTGTAGGCATGATAAGACGCGTCAGCGTCGCATCAGGCA	ND	1
	EC179-3	1,410,874	1,411,015	1	GCCGGATGCGGCGTGAACGCCTTATCCGGCCTACAAAAGAAATGCAG	+	1
	EC179-4	3,084,207	3,084,333	1	TTTGTAGGCCTGATAAGACGCGCCAGCGTCGCATCAGGC	+	1
	EC179-5	3,655,982	3,657,047	17	CGGTTTATCCCCGCTGGCGCGGGGAACAC	-	4
	CAS-TypeIE	3,657,144	3,665,705	8	cas2_TypeIE, cas1_TypeIE, cas6_TypeIE, cas5_TypeIE, cas7_TypeIE, cse2_TypeIE, cse1_TypeIE, cas3_TypeI		
	EC179-6	3,682,750	3,683,327	9	CGGTTTATCCCCGCTGGCGCGGGGAACTC	-	4
	EC179-7	4,072,440	4,072,562	1	TTGCCGGATGCGGCGTAAACGCCTTATCCGGC		
EC256	EC256-1	531,399	531,522	1	CGACCCCCACCATGTCAAGGTGGTGCTCTAACCAACTGAGCTA	ND	1
	EC256-2	1,275,360	1,275,451	1	CCACCTTTTTTACCTGCTTCAGATGC	ND	1
	EC256-3	1,409,877	1,410,024	2	TTTCTAAGCTGCCTGTACGGCAGTGAAC	-	1
	EC256-4	4,432,516	4,433,767	20	GTGTTCCCCGCGCCAGCGGGGATAAACCG	+	4
	CAS-TypeIE	4,450,558	4,459,208	8	cas3_TypeI, cse1_TypeIE, cse2_TypeIE, cas7_TypeIE, cas5_TypeIE, cas6_TypeIE, cas1_TypeIE, cas2_TypeIE		
	EC256-5	4,459,314	4,460,014	11	GAGTTCCCCGCGCCAGCGGGGATAAACCG	+	4

Table 6. A list of CRISPR sequences detected by CRISPRCasFinder tool.

indicated a high level of multidrug resistance in studied UPEC strains, driven mainly by efflux mechanisms and β-lactamase production. However, these results were not fully consistent with phenotypic studies, except EC256 strain[6]. This is not surprising as drug resistance is multifactorial.

Susceptibility to lytic phages testing

The search for phages capable of infecting the five selected UPEC strains showing different antibiotic resistance profiles, led to the isolation of five new lytic phages from municipal sewage. The isolation procedure was based on the one-host enrichment method in which a raw urban sewage sample was mixed with a culture of the UPEC strain. With this approach, no phages infecting the EC3 and EC256 strains were acquired, however, two phages against EC149, two phages against EC179, and one phage against the EC164 strain were successfully isolated. All of them infected host UPEC bacteria with high efficiency, giving the titer of the order from 10^8 to 10^{11} PFU/mL. The obtained phages were named according to the nomenclature principles proposed by Kropinski and collaborators[52], taking into account the name of the original host strain. Thus, phages isolated using strain EC149 were named vB_EcoS-149_4M, and vB_EcoS-149_6D, whereas phages isolated on strains EC164 and EC179 were named vB_EcoS-164_2 and vB_EcoS-179_3 or vB_EcoP-179_5, respectively.

To evaluate the sensitivity of UPEC strains to the isolated phages, a more detailed analysis based on the measurement of the efficiency of plating (EOP) was performed. This method allowed for comparing the phage efficiency of infection between the primary host strain EC149, EC164, or EC179, and other UPEC clinical isolates (Fig. 3). Analysis revealed that phage vB_EcoS-149_4M presented the broadest strong activity, infecting 7 isolates with high efficiency and 3 with medium efficiency (Fig. 3). Interestingly, 3 of 7 strains infected by this phage with high efficiency (EC113, EC131, and EC134) showed EOP values above 1, which means that they are

Name of prophage region and length (kb)		Assigned bacteriophage (PHASTEST nomenclature)	Proteins	GC (%)	Holins	Endolysins	CARD - Perfect hits	CARD - Strict hits
Identified in *Escherichia coli* EC149								
EC149.1(1)	44.5	PHAGE_Entero_BP_4795_NC_004813(7)	52	46.11	1*	1		
EC149.1(2)	36.7	PHAGE_Entero_SfV_NC_003444(14)	22	47.32				
EC149.1(3)	20.0	PHAGE_Entero_SfI_NC_027339(6)	28	49.41	1*			
EC149.1(4)	13.8	PHAGE_Entero_HK106_NC_019768(3)	20	46.68				
EC149.2(1)	99.5	PHAGE_Entero_HK106_NC_019768(3)	119	48.07				
Identified in *Escherichia coli* EC164								
EC164.1(1)	38.3	PHAGE_Entero_fiAA91_ss_NC_022750(23)	36	49.72	2	1		
EC164.1(2)	46.5	PHAGE_Entero_lambda_NC_001416(23)	57	51.36	1*			
EC164.1(3)	57.3	PHAGE_Salmon_vB_SosS_Oslo_NC_018279(13)	78	52.15	2			
EC164.1(4)	44.3	PHAGE_Entero_DE3_NC_042057(25)	62	50.30	1*			
EC164.1(5)	53.9	PHAGE_Entero_mEp460_NC_019716(22)	59	50.45	1			
EC164.1(6)	54.1	PHAGE_Entero_mEp460_NC_019716(38)	60	49.44	1	1		
EC164.1(7)	46.7	PHAGE_Entero_mEp460_NC_019716(41)	55	50.58	1*	1		
EC164.1(8)	39.8	PHAGE_Burkho_BcepMu_NC_005882(31)	58	55.05		1		
EC164.1(9)	43.1	PHAGE_Shigel_SfII_NC_021857(34)	53	49.39	1			
EC164.1(10)	46.3	PHAGE_Escher_500465_1_NC_049342(37)	56	50.03	1			
EC164.1(11)	42.3	PHAGE_Entero_Sf101_NC_027398(19)	63	47.14	1			
EC164.3(1)	98.4	PHAGE_Entero_Sf101_NC_027398(19)	115	47.62				
Identified in *Escherichia coli* EC179								
EC.179.1(1)	22.4	PHAGE_Entero_fiAA91_ss_NC_022750(20)	29	54.04	2	1		
EC.179.1(2)	45.1	PHAGE_Entero_lambda_NC_001416(22)	46	50.92	1*			
EC.179.1(3)	49.2	PHAGE_Entero_lambda_NC_001416(28)	61	50.02	1*			
EC.179.1(4)	58.6	PHAGE_Entero_mEp460_NC_019716(24)	59	48.76	1	1		
EC.179.1(5)	36.9	PHAGE_Entero_fiAA91_ss_NC_022750(25)	47	50.11	2	1		
EC.179.1(6)	44.4	PHAGE_Escher_phiV10_NC_007804(41)	57	48.70	1*	1*		
EC.179.2(1)	43.8	PHAGE_Salmon_SJ46_NC_031129(6)	42	52.20			TEM-1	
Identified in *Escherichia coli* EC256								
EC.256.1(1)	16.5	PHAGE_Burkho_phiE255_NC_009237(16)	21	48.63		1		
EC.256.1(2)	46.9	PHAGE_Entero_DE3_NC_042057(23)	57	48.87	1*			emrE
EC.256.1(3)	56.8	PHAGE_Escher_HK639_NC_016158(8)	68	47.27	1*	1		
EC.256.1(4)	37.6	PHAGE_Entero_mEp460_NC_019716(21)	45	51.07	1*	1		
EC.256.1(5)	37.8	PHAGE_Entero_P88_NC_026014(29)	50	50.57	1			
EC.256.1(6)	45.3	PHAGE_Entero_fiAA91_ss_NC_022750(26)	46	50.50	2	1		
EC.256.4(1)	40.7	PHAGE_Escher_TL_2011b_NC_019445(37)	53	48.01	1*	1*		

Table 7. Results of bioinformatic analysis of prophage regions and protein-coding sequences within them. Legend: Proteins – number of encoded proteins; Holins - Consistent holins between PHASTEST and Phyre2; Endolysins - consistent endolysins between PHASTEST and Phyre2; TEM-1–100%, TEM β-lactamase; emrE – 98.18%, small multidrug resistance (SMR) antibiotic efflux pump, *Escherichia coli* emrE, * - putative.

even better bacterial strains for phage propagation than the primary strain EC149 (Fig. 3). In turn, phage vB_EcoS-149_6D infected 4 isolates with high efficiency but showed the highest number of low-efficiency infections, whereas phage vB_EcoS-164_2 had high efficiency against only 2 isolates, with most other infections being sporadic (10 UPEC isolates, Fig. 3). Phages vB_EcoS-179_3 and vB_EcoP-179_5 displayed very limited activity, each infecting only one isolate (the primary EC179 host) with high efficiency and showing no medium, low, or sporadic infections among other UPEC strains (Fig. 3). Surprisingly, none of the phages isolated on EC149 was able to infect strains EC164 and EC179, and vice versa (Fig. 3). Overall, the data indicate substantial variability in phage host range and infection strength, with some phages (particularly vB_EcoS-149_4M) demonstrating broader and stronger activity against UPEC isolates compared to other tested phages, which were more selective or had predominantly sporadic infections.

Based on these results, the five newly isolated phages were subjected to further studies of their morphology, stability, toxicity toward eukaryotic cells, and genome content to generate a background for phage therapy, as it may help to overcome the problems of developing resistance among UPEC strains.

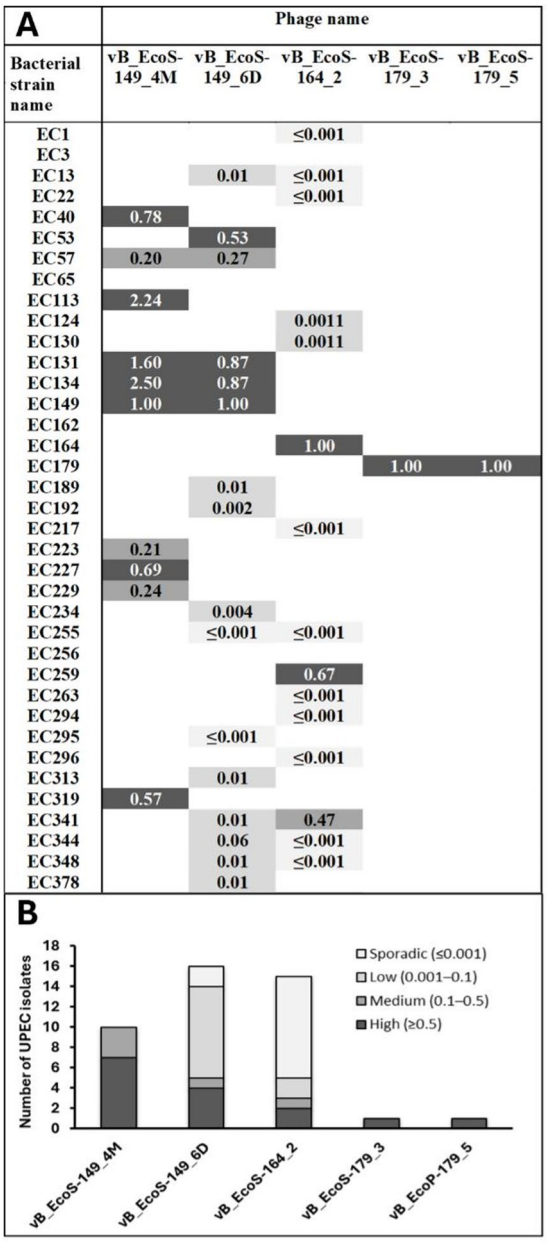


Fig. 3. **A**. EOP (efficiency of plating) values obtained on UPEC clinical isolates. According to the accepted classification[50], an EOP value ≤ 0.001 indicates sporadic production of phage particles in a given bacterial strain, while values in the range 0.001 < EOP < 0.1; 0.1 ≤ EOP < 0.5, and ≥ 0.5 indicate low □, medium □, and high □ efficiency of phage infection, respectively. Lack of infection is indicated by empty white. The obtained results are from three independent replicates. The SD values are not shown; however, they do not exceed 15%. The distribution of infection efficiencies for tested phages against the UPEC isolates.

Detailed comparative analysis of newly isolated anti-UPEC phages as background for phage therapy

Plaque and virion morphology

First, all tested phages were subjected to morphological analysis of their plaques and virions. As presented in Fig. 4, all of them formed clear plaques, typical for lytic phages, which differed in size. The smallest ones had a diameter < 1 mm and were formed by phage vB_EcoS-164_2, while the largest ones were about 2 mm and were generated by phage vB_EcoS-149_6D. In addition, for phages vB_EcoS-179_3 and vB_EcoP-179_5, barely visible haloes (turbid zones around the plaques) were observed. Their presence indicates the action of released (not associated with a tail) phage enzymes on bacterial cells. The plaques of the remaining phages did not show the presence of haloes. Although we cannot completely exclude the possibility that strain EC179 releases extracellular factors in response to phage infection, this explanation appears less likely. Plaque-associated halos are most commonly attributed to phage-encoded polysaccharide depolymerases that diffuse beyond the plaque

margins and enzymatically degrade bacterial surface polysaccharides, including capsules, exopolysaccharides, and biofilm matrix components. Supporting this hypothesis, bioinformatic analyses identified a tail spike protein (YEP44078.1) in phage vB_EcoS-179_3 and a tail fiber protein (YEP44190.1) in phage vB_EcoP-179_5 as candidate depolymerases. Notably, halo formation was restricted to phages infecting strain EC179, whereas phages infecting EC149 and EC164 formed clear plaques without detectable halos. This observation may indicate that strain EC179 possesses unique surface polysaccharide structures that serve as substrates for depolymerase activity and that both phages encode depolymerase-like proteins capable of interacting with these structures. Therefore, we speculate that the halo phenotype observed for EC179-associated phages results primarily from phage-mediated enzymatic degradation of host surface polysaccharides rather than from the release of host-derived factors.

Viron morphology of all phages was investigated under electron microscopy. This analysis revealed that they all belong to tailed phages and four present morphology characterized by flexible, long non-contractile tails and isometric heads, whereas phage vB_EcoP-179_5 has a short tail (Fig. 4).

The sensitivity to external agents

The stability of the tested phages following exposure to physical and chemical agents is summarized in Fig. 5. The values are expressed as percentages of surviving phages relative to the untreated control. A two-way ANOVA demonstrated significant effects of both phage type ($F(4,120) = 32.51$, $p < 0.001$) and experimental condition ($F(11,120) = 112.60$, $p < 0.001$) on phage survival. Moreover, a significant phage × condition interaction was observed ($F(44,120) = 14.94$, $p < 0.001$), indicating that the response to external agents depended on the particular phage tested and the tested external factors affected phage survival to significantly different extents.

Overall, phage vB_EcoS-149_4M exhibited the highest stability. It retained full infectivity under most tested conditions, including exposure to pH 4 and pH 10, temperatures of − 20 °C and 40 °C, osmotic shock, and treatment with DMSO, ethanol, sarkosyl, and chloroform. A reduction in titer was observed only after exposure to high temperature (62 °C), strong acidity (pH 2), and CTAB. In contrast, vB_EcoS-149_6D displayed moderate stability, remaining fully resistant to 40 °C, pH 10, chloroform, and sarkosyl, while maintaining relatively high survival at pH 4 (59%), DMSO (83%), CTAB (67%), and under osmotic shock (87%).

Phages vB_EcoS-179_3 and vB_EcoP-179_5 remained fully stable at − 20 °C and 40 °C, as well as at pH 4 and pH 10, but showed marked reductions in titer following exposure to detergents and ethanol. Phage vB_EcoS-164_2 exhibited the lowest overall stability, with its viability being strongly reduced under most conditions. Only exposure to pH 4, pH 10, DMSO, and chloroform had little or no effect on its infectivity (Fig. 5A).

Tukey's post hoc test further demonstrated that vB_EcoS-149_4M exhibited significantly higher overall survival than all remaining phages (all $p < 0.001$). Phage vB_EcoS-149_6D also showed significantly higher survival than vB_EcoP-179_5 ($p < 0.001$), vB_EcoS-164_2 ($p < 0.001$), and vB_EcoS-179_3 ($p = 0.029$). In contrast, no significant differences in overall survival were detected among vB_EcoP-179_5, vB_EcoS-164_2, and vB_EcoS-179_3 (all $p > 0.05$).

Taken together, these findings indicate that vB_EcoS-149_4M was the most resistant phage among those tested, whereas vB_EcoS-164_2, vB_EcoS-179_3, and vB_EcoP-179_5 formed a group of comparatively labile phages. Overall, a strong acid (pH = 2) completely inactivated all tested phages, and CTAB caused complete or nearly complete inactivation of most phages except vB_EcoS-149_6D. In contrast, weak acidic and alkaline conditions, as well as moderate heating (40 °C), were generally well tolerated. These differential stability profiles have practical implications for phage storage and handling.

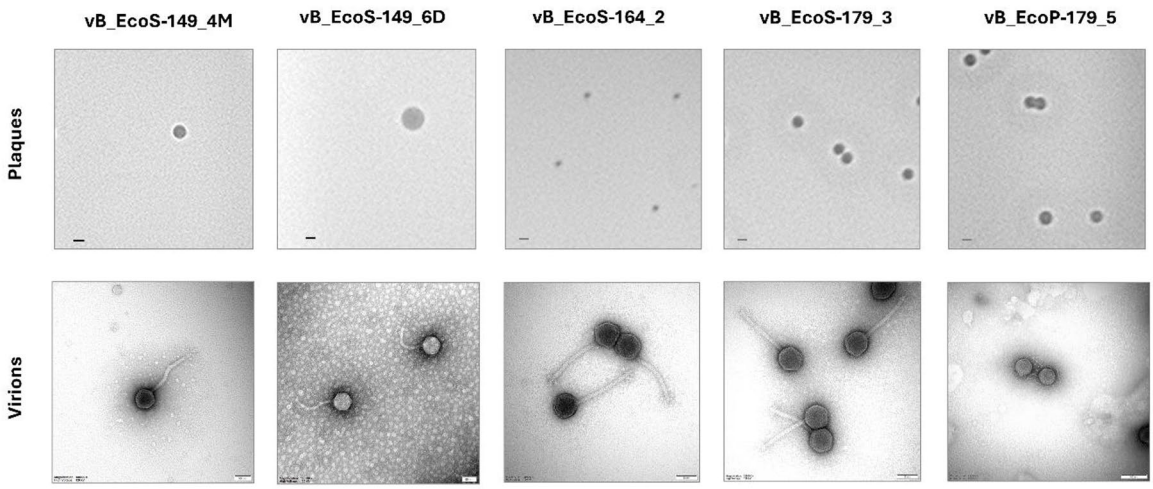


Fig. 4. Phage plaques (upper row) and virions (lower row) morphology. All photographs were made in the same magnification: In the upper row, the bar corresponds to 1 mm and plaque diameters are about: 1 mm for vB_EcoS-149_4M, 2 mm for vB_EcoS-149_6D, < 1 mm for vB_EcoS-164_2, 1 mm for vB_EcoS-179_3, and 1.5 mm for vB_EcoP-179_5. In the lower row, the bar corresponds to 50 nm. Phage particles were observed under a Philips CM 1000 electron microscope (Philips, Eindhoven, The Netherlands).

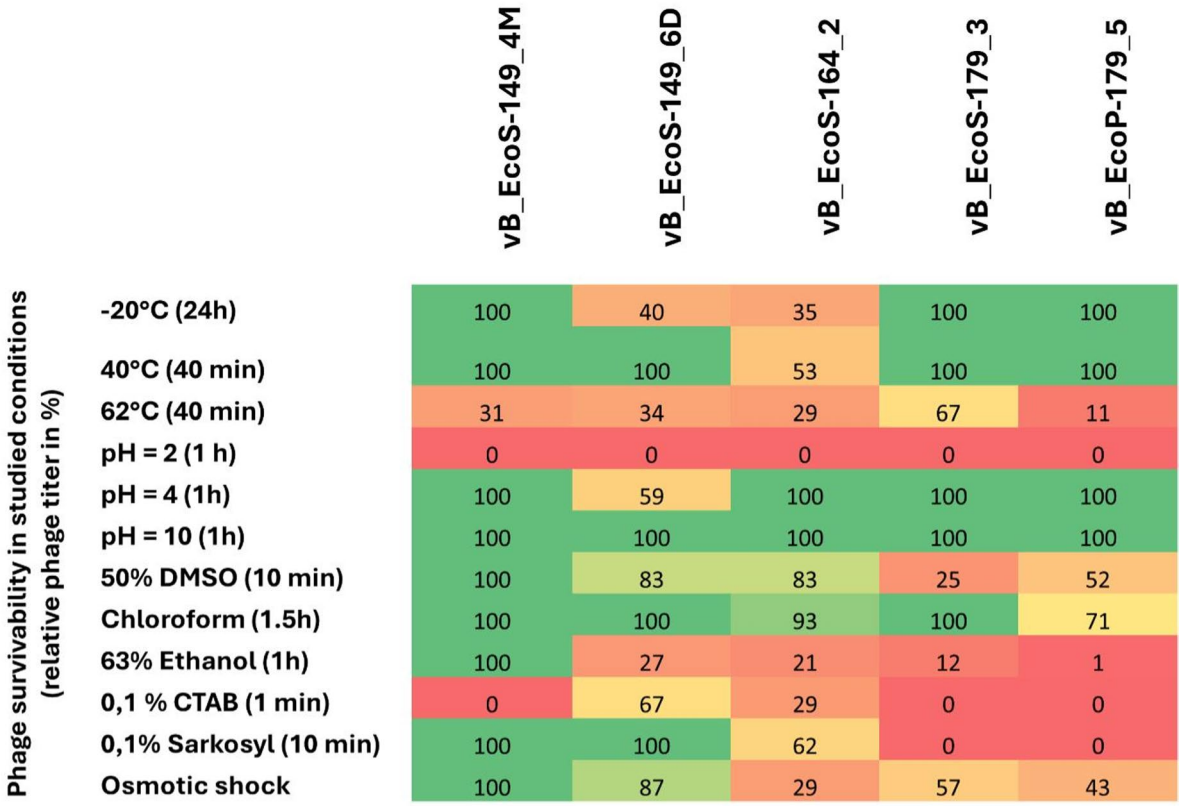


Fig. 5. Resistance of phage virions to physical and chemical agents. The phage titer (PFU/mL) was determined using the spot-test method on two-layer plates and expressed as a percentage of phages surviving the exposure to the external agent, relative to the untreated control variant. Conditional formatting was used to mark the values equal or close to 100% as green, middle values as yellow, and values equal or close to 0% as red. Phage survival, expressed as a percentage of the untreated control, was analyzed using a two-way analysis of variance (two-way ANOVA). The model included two fixed factors: phage identity (five levels) and experimental condition (twelve levels). Three independent replicates were analyzed for each phage–condition combination. Pairwise comparisons were performed using Tukey's honestly significant difference (HSD) post hoc test. Differences were considered statistically significant at $p < 0.05$.

The toxicity toward human cells

To test the safety of phages to mammalian cells, the viability of the human bladder cancer cell line, T24, and the human embryonic kidney line, HEK293, was assessed after treatment with all the tested phages (Fig. 6). The endotoxin levels were checked in all phage solutions and most didn't exceed the value of 1 EU/mL (Supplementary Figure S1). In addition, the stability of phage particles was tested after 24 h-long incubation with mammalian cells (Fig. 6).

The addition of phages (to a concentration of at least 10^8 PFU/mL) to T24 or HEK293 cultures resulted in no significant changes detected in the number of viable cells relative to the negative control, which includes non-treated cultures with viability at the level of 100% (Fig. 6A, B). The lack of a drop in cell survival below 70% in all tested cases (except positive control) suggested no deleterious effect of the tested phages on the viability of mammalian cells and was corroborated by observations of T24 and HEK293 human cells under a light microscope. Compared to untreated cells, no morphological changes could be observed after phage treatment (Fig. 6E, F). Importantly, all phage titers were comparable at time zero and after incubation with mammalian cells for 24 h, which indicated high virion stability under culturing conditions (Fig. 6C, D).

In summary, the obtained results show that all tested phages are safe for human cells and might be used in therapy.

Phage genome content

The complete genomes of all five phages were sequenced and subjected to detailed bioinformatic analysis. The results revealed that genome lengths ranged from 43,053 to 56,684 bp, with GC contents between 44% and 51%. Genome annotation identified multiple coding sequences (CDSs), ranging from 67 to 101 per phage genome; however, putative functions were assigned to less than half of them. Genes with assigned functions were used to generate phage genome maps, which are presented in Supplementary Figure S2A-E.

In addition, analysis using PhageTermVirome indicated that vB_EcoS-149_4M and vB_EcoS-149_6D possess headful packaging termini, whereas vB_EcoP-179_5 contains short direct terminal repeats. For vB_EcoS-164_2 and vB_EcoS-179_3, the phage termini were not determined with the applied tools (Table 8).

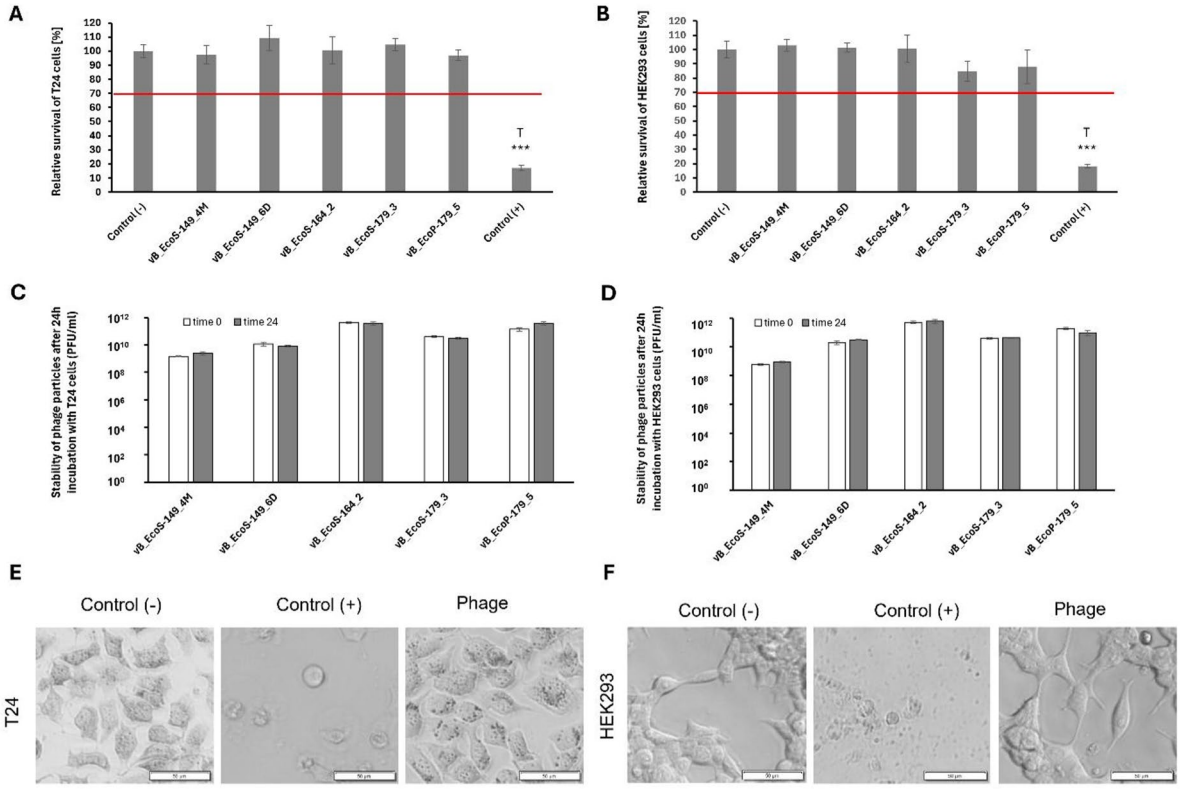


Fig. 6. Assessment of cytotoxicity of phage particles to mammalian cells. Panels (**A**) and (**B**) show results for the human bladder cancer cell line (T24) and the human embryonic kidney line (HEK293), respectively. The phage lysate was concentrated with PEG, purified by ultracentrifugation in a CsCl gradient, and then dialyzed and tested. Letter T means a drop in cell survival below 70%, which qualifies the agent as toxic. Panels (**C**) and (**D**) present the number of phage particles in 1 ml (PFU/mL) of suspension at time 0 and after 24 h of incubation of phage lysates with cells T24 and HEK293, respectively. Statistically significant differences were determined by pairwise comparisons between the phage lysate and the negative or positive control under standard culture conditions, using the Student's *t*-test and marked with the star symbols, where $p < 0.001$ (***). Importantly, no statistically significant differences were observed between any of the tested phage lysates and the negative control in either the T24 or HEK293 cell lines. Panels (**E**) and (**F**) present the morphology of T24 and HEK293 cells, respectively, after 24 h of incubation with phage vB_EcoS-164_2 (representative photos). Images were taken using light microscopy at magnification (x10) (OLYMPUS IX51 microscope). Triton X-100 and TM buffer were used as positive and negative controls, respectively. The bar on panels E and F corresponds to 50 µm.

Importantly, all analyzed phages were classified as virulent with a confidence of at least 0.98 (Table 8). Nevertheless, genes encoding proteins potentially involved in integration and excision processes were unexpectedly identified in phages vB_EcoS-164_2 and vB_EcoS-179_3, however, they were classified as moron genes acquired during earlier evolutionary events and retained in the phage genomes. Importantly, other sequences classified as morons were detected in all analyzed phages. In turn, screening for virulence factors and antimicrobial resistance genes did not reveal the presence of any such genes in the studied phage genomes.

Taxonomic and phylogenetic analyses

Performed classification revealed that phages vB_EcoS-149_4M and vB_EcoS-149_6D belong to the family *Drexlerviridae* but represent two distinct genera: phage vB_EcoS-149_4M clusters within the genus *Hanrivervirus* (Fig. 7), whereas phage vB_EcoS-149_6D groups within the genus *Warwickvirus* (Fig. 8).

Hierarchical cluster analysis (HCA) based on VIRIDIC genomic distances (Fig. 7A) showed that phage vB_EcoS-149_4M forms a separate branch within the *Hanrivervirus* cluster and exceeds the species demarcation threshold (red dashed line), indicating substantial genomic divergence from all currently described members of this genus.

This result was further supported by the maximum-likelihood phylogenetic tree (Fig. 7B), where we observe that vB_EcoS-149_4M forms a well-supported, isolated lineage relative to other *Hanrivervirus* representatives, including EC147, aalborv, haarsle, vojen, pSf-1, JakobBernoulli, KarlJaspers, JLBYU09, vB_EcoD_Fulton, Henu8, and aaroes. The long branch length separating vB_EcoS-149-4 M from its closest relatives indicates a considerable evolutionary distance. For phage vB_EcoS-149_4M, a total of 29 orthologous genes were identified.

Bacteriophage name [accession no]	Family	Genome length	Identified CDS	Functional CDS	GC content	Life cycle with score	Phage termini
vB-EcoS_149-4 M [PX985795]	*Drexlerviridae*	52,000	98	36	44	virulent 1.00	Headful (pac)
vB-EcoS_149-6D [PX985796]	*Drexlerviridae*	51,036	99	37	45	virulent 1.00	Headful (pac)
vB-EcoS_164-2 [PX985797]	-	56,684	101	43	44	virulent 0.98	-
vB-EcoS_179-3 [PX985798]	*Sarkviridae*	43,053	86	34	51	virulent 1.00	-
vB-EcoP_179-5 [PX985799]	*Autosignataviridae*	44,811	67	31	45	virulent 1.00	DTR (short)

Table 8. Phage genomes characteristic.

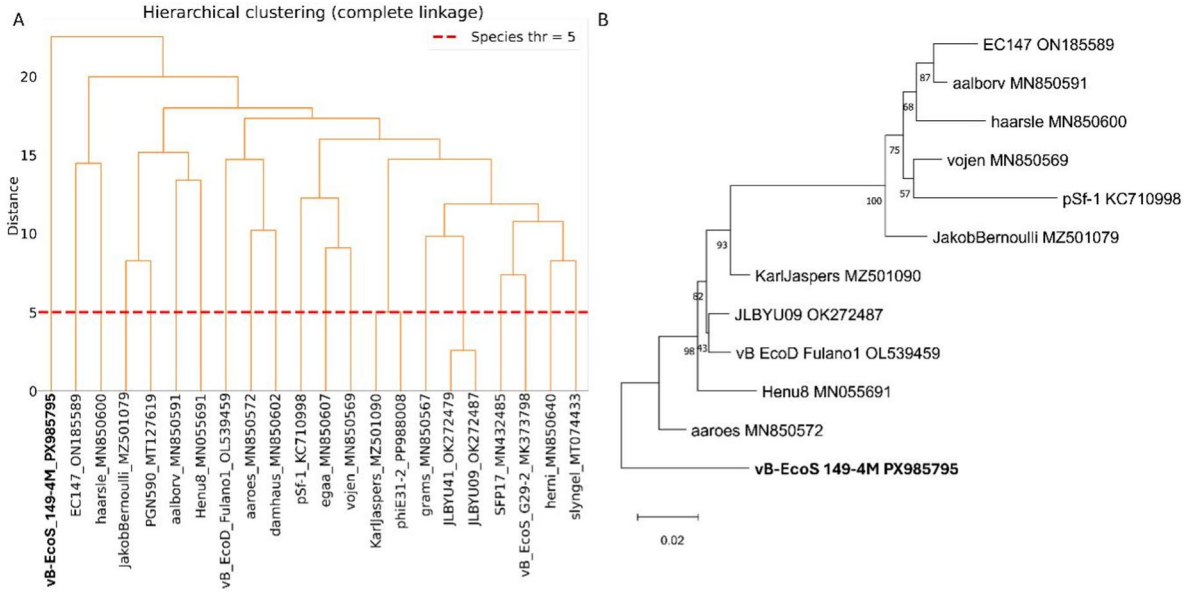


Fig. 7. Taxonomic and phylogenetic analysis of phage vB_EcoS-149_4M. Panel (**A**) presents the results of HCA based on the VIRIDIC intergenomic distance matrix, with the species demarcation threshold indicated by the red dashed line. This analysis illustrates the genomic relatedness of vB_EcoS-149_4M to representative members of the genus *Hanrivervirus* and supports its taxonomic assignment as a novel species within this genus. Panel (**B**) shows a maximum-likelihood phylogenetic tree reconstructed from three orthologous genes using IQ-TREE with 1,000 bootstrap replicates. Phage names and accession numbers are indicated on the tree.

Interestingly, only three orthologous groups contained a sufficient number of parsimony-informative sites and were therefore retained for phylogenetic reconstruction (Supplementary Table S2).

Further taxonomic classification revealed that phage vB_EcoS-149_6D belongs to the genus *Warwickvirus* within the family *Drexlerviridae* (Fig. 8). HCA based on VIRIDIC genomic distances (Fig. 8A) demonstrated that vB_EcoS-149_6D clusters below the species demarcation threshold (red dashed line) together with phage AV105, indicating that these two phages constitute a single species. This cluster is clearly separated from other *Warwickvirus* representatives, including Tuinin, MikeNSara, EC167, BEK12A, and PeterMerian, supporting its distinct taxonomic position at the species level.

The maximum-likelihood phylogenetic tree reconstructed from orthologous gene alignments (Fig. 8B) corroborated the clustering results. It shows that phage vB_EcoS-149_6D is separated from more distantly related *Warwickvirus* phages such as Tuinin, MikeNSara, and EC167, indicating a distinct evolutionary lineage within the genus. High bootstrap support values at the internal nodes further confirm the robustness of this phylogenetic placement. For phage vB_EcoS-149_6D, a total of 51 orthologous genes were identified, of which eight met the criteria for phylogenetic analysis (Supplementary Table S2).

Taken together, both the genomic distance clustering (Figs. 7A and 8A) and phylogenetic tree topology (Figs. 7B and 8B) consistently demonstrate that phage vB_EcoS-149_4M represents a distinct lineage within the genus *Hanrivervirus*, while phage vB_EcoS-149_6D belongs to the genus *Warwickvirus* and forms a separate species cluster with phage AV105. Based on these data, we propose to establish two novel species: *Hanrivervirus vB_EcoS-149_4M* and *Warwickvirus vB_EcoS-149_6D*.

Further analyses revealed that phage vB_EcoS-164_2 is related to phages currently assigned to two closely related genera that are not yet classified within any recognized family. As shown in Fig. 9, vB_EcoS-164_2 clusters

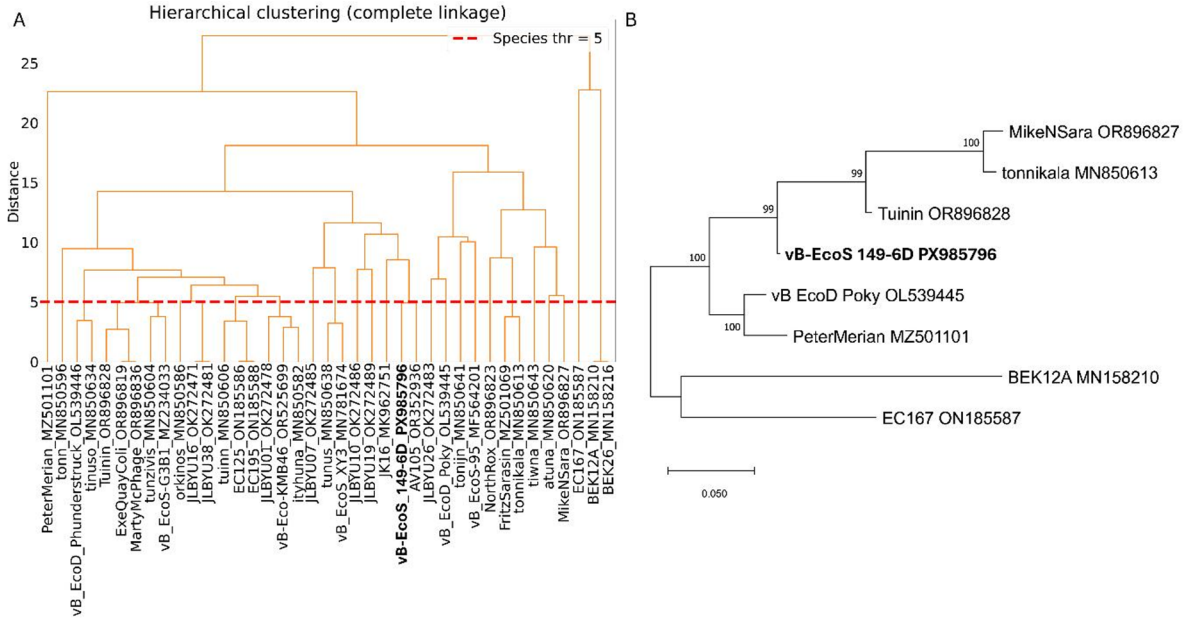


Fig. 8. Taxonomic and phylogenetic analysis of phage vB_EcoS-149_6D. Panel (**A**) presents the results of HCA based on the VIRIDIC intergenomic distance matrix, with the species demarcation threshold indicated by the red dashed line. This analysis illustrates the genomic relatedness of vB_EcoS-149_6D to representative members of the genus *Warwickvirus* and supports its taxonomic placement as a species cluster with phage AV105 within this genus. Panel (**B**) shows a maximum-likelihood phylogenetic tree reconstructed from eight orthologous genes using IQ-TREE with 1,000 bootstrap replicates. Phage names and accession numbers are indicated on the tree.

with phages of the genus *Nonavirus*, including several closely related *Sashaviruses*. Interestingly, phage 9NA, although the most distantly positioned, also clusters within the *Nonavirus* lineage (Fig. 9), further suggesting the genetic cohesion of this group. HCA showed that the analyzed phages form three subclusters that exceed the recommended by ICTV genus demarcation threshold (blue dashed line), suggesting the presence of three potential separate genera (Fig. 9A). However, the observed clustering remains compact and the overall genomic distances among these phages are relatively small, indicating a shared evolutionary background. In addition, the maximum-likelihood phylogenetic tree based on the concatenated orthologous sequence alignments (Fig. 9B) demonstrates that all analyzed phages form a coherent monophyletic lineage, with clear internal branching corresponding to the three subclusters observed in the HCA. Interestingly, the two outermost subclusters are formed by phages (vB_EcoS PJ16, vB_EcoS AVIO78A, KarinMuster_Bas82, vB_EcoS-164_2, and 9NA) belonging to the genus *Nonavirus*. Besides, the overall topology and short inter-branch distances support a close evolutionary relationship among all members of the group. In respect to this an taking into account the shared orthologous gene content (of which 13 exhibited sufficient phylogenetic signal to be retained for tree reconstruction as indicated in Supplementary Table S2), and the consistent phylogenetic placement of all analyzed phages, we propose to classify all these phages within a single genus, *Nonavirus* as indicated by a dashed black line in Fig. 9B. Our proposal results from the fact that despite the recommendation, ICTV accepts cases in which phages showing less than 70% of sequence similarity and exceeding the genus threshold (blue dashed line) are classified as belonging to the same genus due to their close evolutionary relationship. Furthermore, as phage vB_EcoS-164_2 apparently represents a distinct species within this genus, we propose to designate this species as *Nonavirus vB_EcoS-164_2*.

In turn, phage vB_EcoS-179_3 belongs to the family *Sarkviridae* and is most closely related to members of the genus *Kagunavirus* (Fig. 10). HCA showed that phage vB_EcoS-179_3 clusters within the *Kagunavirus* lineage but exceeds the recommended species demarcation threshold (red dashed line), indicating that it represents a distinct, previously uncharacterized species within this genus (Fig. 10A). As observed for the *Nonavirus* genus described above, although genome similarities among analyzed *Kagunavirus* phages are generally below 70%, and phages exceed the genus threshold (blue dashed line), they can still be formally classified within a single genus due to high evolutionary relationship.

The maximum-likelihood phylogenetic tree reconstructed from concatenated orthologous protein alignments further supported this classification (Fig. 10B). Phage vB_EcoS-179_3 forms a well-supported branch within the *Kagunavirus* clade, however, it is clearly separated from other *Kagunavirus* representatives. These results were obtained on 10 of 28 identified orthologous genes, which contained sufficient phylogenetic signal to be retained for tree reconstruction (Supplementary Table S2). Therefore, we suggest classifying phage vB_EcoS-179_3 as a novel species within the genus *Kagunavirus*. Accordingly, we propose the designation of this species as *Kagunavirus vB_EcoS-179_3*.

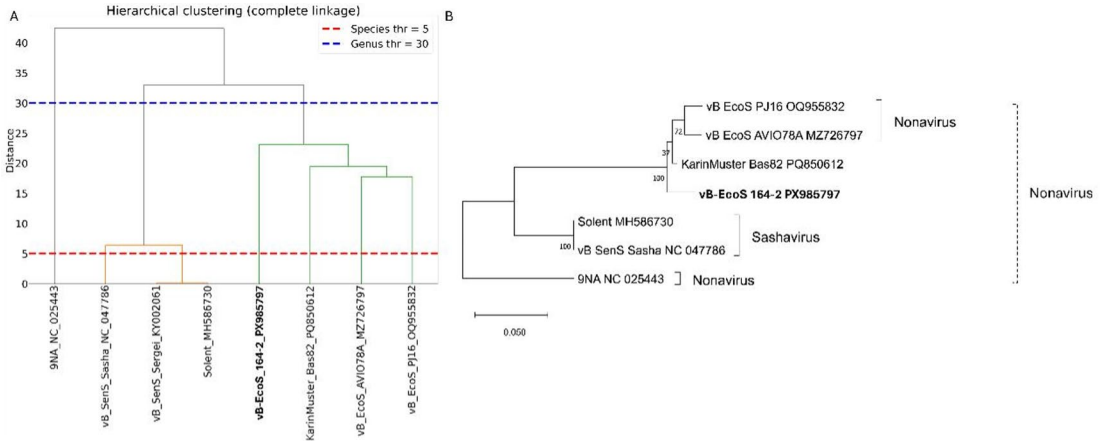


Fig. 9. Taxonomic and phylogenetic analysis of phage vB_EcoS-164_2. Panel (**A**) presents the results of HCA based on the VIRIDIC intergenomic distance matrix, with the species demarcation threshold indicated by the red dashed line and the genus demarcation threshold marked by the blue dashed line. This analysis illustrates the genomic relatedness of vB_EcoS-164_2 to closely related phages assigned to the genus *Nonavirus* and supports its taxonomic classification as a novel species within this genus. The orange cluster represents phages currently assigned to the genus *Sashavirus*, whereas the green cluster represents phages belonging to the genus *Nonavirus*, including phage vB_EcoS-164_2 and its closest relatives. Panel (**B**) shows a maximum-likelihood phylogenetic tree reconstructed from thirteen orthologous genes using IQ-TREE with 1,000 bootstrap replicates, depicting the evolutionary relationships among the analyzed phages and their placement within the *Nonavirus* lineage (black dashed line). Phage names and accession numbers are indicated on the tree.

Finally, phage vB_EcoP-179_5 is suggested to represent a novel species within the genus *Vectrevirus* of the family *Autosignataviridae* (Fig. 11). HCA showed that the analyzed phages segregate into two major clades. Phage vB_EcoP-179_5 is clearly separated from the main clade and forms an independent branch, indicating substantial genomic divergence from previously characterized *Vectrevirus* members (Fig. 11A). The maximum-likelihood phylogenetic tree (based on 10, the most informative genes as indicated in Supplementary Table S2) further supported the distinct phylogenetic position of phage vB_EcoP-179_5. This phage forms a well-supported, isolated lineage within the *Vectrevirus* clade and is clearly separated from its closest relatives. This is apparently consistent with its classification as a novel species, which we designated as *Vectrevirus vB_EcoP-179-5* (Fig. 11B).

The performed taxonomic and phylogenetic analyses clearly demonstrate the successful identification and characterization of novel bacteriophage species infecting uropathogenic *E. coli* bacteria. The integration of HCA based on VIRIDIC genomic distance matrices with maximum-likelihood phylogenetic reconstruction enabled accurate taxonomic placement of the analyzed phages and assessment of their evolutionary divergence. Supplementary Figure S3 presents a comprehensive hierarchical cluster analysis of all phages analyzed in this study (including the tested five), providing an integrated overview of their genomic relationships. The obtained dendrogram resolves the phages vB_EcoS-149_4M, vB_EcoS-149_6D, vB_EcoS-164_2, vB_EcoS-179_3, and vB_EcoP-179_5 into well-defined, distinct clusters, reflecting their pronounced genomic diversity and assignment to different phylogenetic lineages. Phages vB_EcoS-149_4M and vB_EcoS-149_6D cluster within lineages corresponding to the family *Drexlerviridae* but belong to separate genera *Hanrivervirus* and *Warwickvirus*, whereas vB_EcoS-164_2, vB_EcoS-179_3, and vB_EcoP-179_5 form independent lineages clearly separated from one another and from the remaining analyzed phages of *Nonavirus*, *Kagunavirus*, *and Vectrevirus* genus, respectively. This spatial separation on the dendrogram supports their distinct taxonomic status and confirms that they represent newly characterized species.

Genomic comparison between phages

As shown in Fig. 12, all analysed phages are similar to the selected reference phages from their respective genera. In the case of phages vB_EcoS-149_4M and vB_EcoS-149_6D, a small inversion is observed at the end of the genome, which is most likely the result of differences in genome assembly approaches. Nevertheless, gene regions responsible for DNA, RNA and nucleotide metabolism, as well as, connector, head and packaging, and lysis functions, are highly similar among all pairs of phages. The most pronounced differences are observed within the tail region of the vB_EcoS-164_2, vB_EcoS-179_3, and vB_EcoP-179_5 and their reference phage genomes, potentially reflecting variation in receptor-binding proteins and depolymerase-encoding genes.

As mentioned earlier, genes predicted to encode products participating in integration and excision processes were identified in the genomes of phages vB_EcoS-164_2 and vB_EcoS-179_3; however, both bioinformatic analyses and laboratory experiments indicate that all five phages analysed by us are lytic. To further investigate this discrepancy, the PhaTYP software was used to predict the lifecycles of all phages included in the phylogenetic analysis above. As shown in Supplementary Table S3, only phages Sasha, Sergei, KarinMuster Bas82, and Solent were classified as temperate, whereas all the rest, including the tested vB_EcoS-149_4M, vB_EcoS-149_6D,

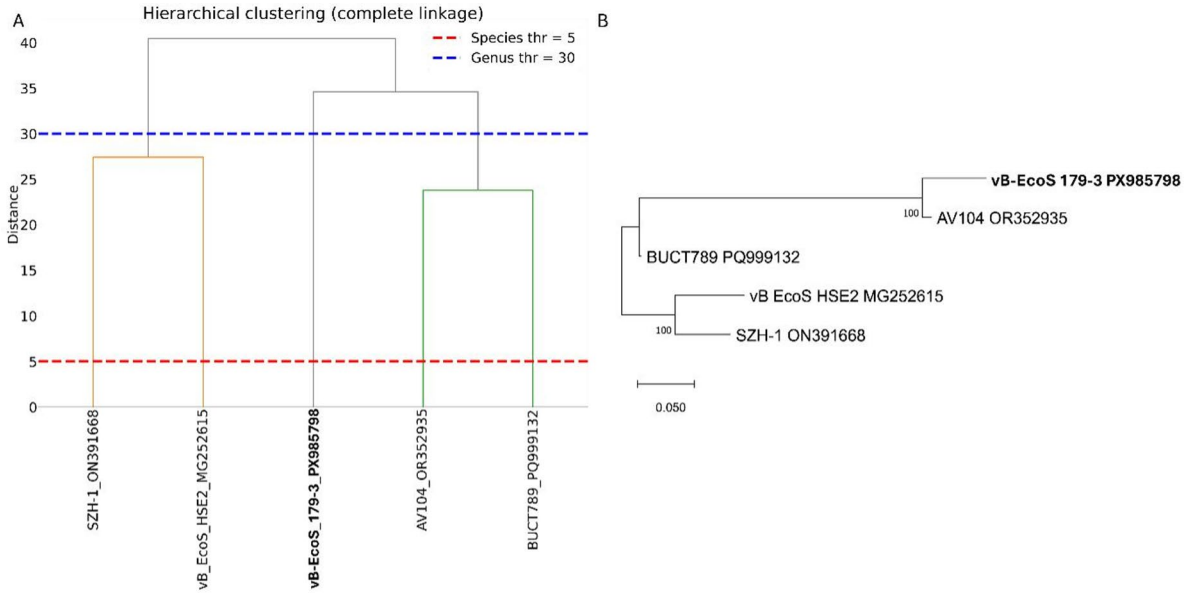


Fig. 10. Taxonomic and phylogenetic analysis of phage vB_EcoS-179_3. Panel (**A**) presents the results of HCA based on the VIRIDIC intergenomic distance matrix, with the species demarcation threshold indicated by the red dashed line and the genus demarcation threshold marked by the blue dashed line. This analysis illustrates the genomic relatedness of vB_EcoS-179_3 to members of the genus *Kagunavirus* and supports its classification as a novel species within this genus. The orange cluster and the green cluster correspond to the two major sublineages within the genus *Kagunavirus*, as identified by the HCA analysis. They do not represent different genera but rather distinct phylogenetic subclusters within the same genus. The analyzed *Kagunavirus* phages generally share less than 70% genome similarity and exceed the recommended genus threshold (blue dashed line). Nevertheless, their close evolutionary relationships support their classification within a single genus. Panel (**B**) shows a maximum-likelihood phylogenetic tree reconstructed from ten orthologous genes using IQ-TREE with 1,000 bootstrap replicates. Phage names and accession numbers are indicated on the tree.

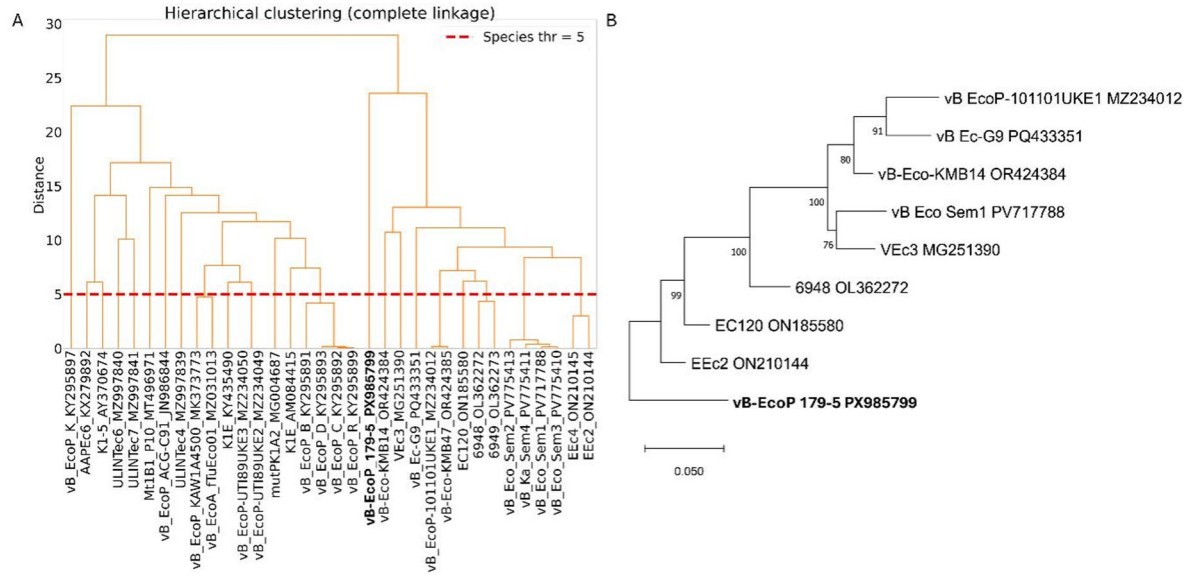


Fig. 11. Taxonomic and phylogenetic analysis of phage vB_EcoP-179_5. Panel (**A**) presents the results of HCA based on the VIRIDIC intergenomic distance matrix, with the species demarcation threshold indicated by the red dashed line. This analysis illustrates the genomic relatedness of vB_EcoP-179_5 to representative members of the genus *Vectrevirus* and supports its classification as a novel species. Panel (**B**) shows a maximum-likelihood phylogenetic tree reconstructed from ten orthologous genes using IQ-TREE with 1,000 bootstrap replicates. Phage names and accession numbers are indicated on the tree.

vB_EcoS-164_2, vB_EcoS-179_3, and vB_EcoP-179_5, were classified as lytic. Interestingly, Sasha, Sergei, and Solent were described as lytic phages by their authors[53]. This inconsistency cannot be fully explained based on the available information; therefore, experimental confirmation of the bioinformatic-based predictions is required.

Nonetheless, the presence in vB_EcoS-164_2 and vB_EcoS-179_3 genomes of genes annotated as products engaged in integration and excision processes may represent remnants of a temperate ancestor or so-called moron genes that were acquired in the past and have been retained in these genomes. As confirmed by phylogenetic analysis, both phages vB_EcoS-164_2 and vB_EcoS-179_3 are relatively unique, and the limited number of closely related phages currently available makes it difficult to draw definitive conclusions at this stage.

Discussion

Urinary tract infections, despite many years of intensive research, still pose a serious medical threat[54–59]. The main factor causing urinary tract infections is *E. coli* bacterium[5,60]. Although the concept of a distinct evolutionary specialization of UPEC remains debated and the boundaries between pathogenic and commensal *E. coli* are often blurred, several studies have suggested that certain phylogenetic lineages and combinations of virulence-associated traits may confer an increased propensity to cause urinary tract infections[57]. Accordingly, the ability of *E. coli* to cause UTIs does not rely on a universal set of virulence factors but rather on specific combinations of determinants that promote adaptation to the urinary tract environment[6]. These determinants include adhesion to epithelial cells, invasion of urinary tract cells, intracellular biofilm formation, antibiotic resistance, and the ability to transfer resistance determinants to bacterial cells inhabiting the same ecological niche, all of which are among the most important features enabling UPEC to cause persistent or recurrent infections[61]. Consequently, the risk, severity, and clinical outcome of UTIs are likely determined by complex interactions between host susceptibility and the urovirulence potential of genetically diverse *E. coli* strains[62–65] .

The use of bacteriophages and phage cocktails to combat infections caused by uropathogenic bacteria is one of the promising and constantly developed methods of biocontrol of pathogens[66–68]. This study aimed to provide preliminary insights into the relationship between the genomic features of a tested UPEC strains and their potential relevance to bacteriophage therapy. By elucidating the genetic underpinnings of UPEC strains, this research contributes to the development of targeted phage therapies that could offer a viable alternative to antibiotics in the treatment of UTIs. The findings from this study have the potential to inform future phage therapy protocols, not only for UTIs but also for other bacterial infections where antibiotic resistance is a growing concern. This study focused on the comparative analysis of five UPEC genomes with the aim of identifying potential targets for bacteriophage therapy. By examining the genomic structure, virulence factors, and resistance mechanisms present in these UPEC strains, this research sought to uncover the vulnerabilities that could be exploited by phages. A detailed understanding of the genomic features of UPEC is crucial for designing effective phage-based treatments, as it allows for the identification of specific bacterial receptors, metabolic pathways, and genetic determinants that are critical for phage infection and replication. The rationale behind selecting multiple UPEC genomes for this study lies in the genetic diversity observed among UPEC strains.

The basis for the large diversity of UPEC is the dynamic structure of their genome[69,70]. It can be conventionally divided into two parts. The first includes the core genome, which consists of housekeeping genes involved in replication, transcription and translation. The core genome determines the basic metabolic functions of bacteria and is characterized by stability. The second part is much more flexible, consisting of a set of genes that are often unique to individual strains and includes mobile elements, like plasmids, phages, and genomic islands[71–73]. These genes may be acquired or lost depending on adaptive needs and influences the colonization and pathogenicity of bacteria. The structure and size of the genome reflect the functioning and habitat of bacteria. Mobile elements increase the cell's adaptability to a changing environment, which therefore increases the genome size.

In our study, isolate EC3 had the smallest genome size. It also had the lowest number of subsystems, coding sequences and the number of RNAs. Isolate EC256 had the largest genome and the largest number of subsystems, isolate EC164 had the largest number of coding sequences and RNAs. The GC content was relatively consistent. Despite differences resulting from the need to adapt to the environment, comparisons of the genomes of all isolates confirmed that they belong to *E. coli* species and are relatively closely related to each other. The genome isolates and EC3 was closest to the UPEC_CFT073 reference genome. The most distinctive strain was EC256, but it should be remembered that it was the only multi-drug-resistant strain and seems to be the most virulent.

The most frequently identified plasmids were those from the IncF group. This is a family of plasmids commonly found in Gram-negative bacteria, including commensal ones. However, they may carry genes associated with virulence and antibiotic resistance, especially resistance to extended-spectrum β-lactamases (ESBLs), but also to aminoglycosides, tetracyclines and carbapenems especially resistance to extended-spectrum β-lactamases (ESBLs), but also to aminoglycosides, tetracyclines and carbapenems[74,75]. What is more, their remarkable intracellular versatility and the rapid evolution of the regulatory sequences of their replicons enable them to effectively spread among *Enterobacteriaceae*. In isolate EC149, we identified a pO111 plasmid that harbored the colistin resistance gene *mcr-1.1*, suggesting a potential role of this plasmid in the transfer of colistin resistance[76,77]. This is an important observation because colistin is often a "last resort" drug and the increase in resistance has serious consequences. We also detected the IncM2 plasmid in isolate EC256. In the literature, this plasmid is associated with the presence of the gene encoding NDM-1 carbapenemase. Moreover, plasmids from the IncM2 group encoding blaNDM-1 are considered emerging threats and require constant monitoring[78]. In addition to their role in the dissemination of antimicrobial resistance genes, plasmids may also influence bacteriophage susceptibility by carrying phage defense systems, such as restriction–modification systems, abortive infection systems, and other mobile genetic elements that can interfere with phage infection, thereby potentially affecting the efficacy of phage therapy[79].

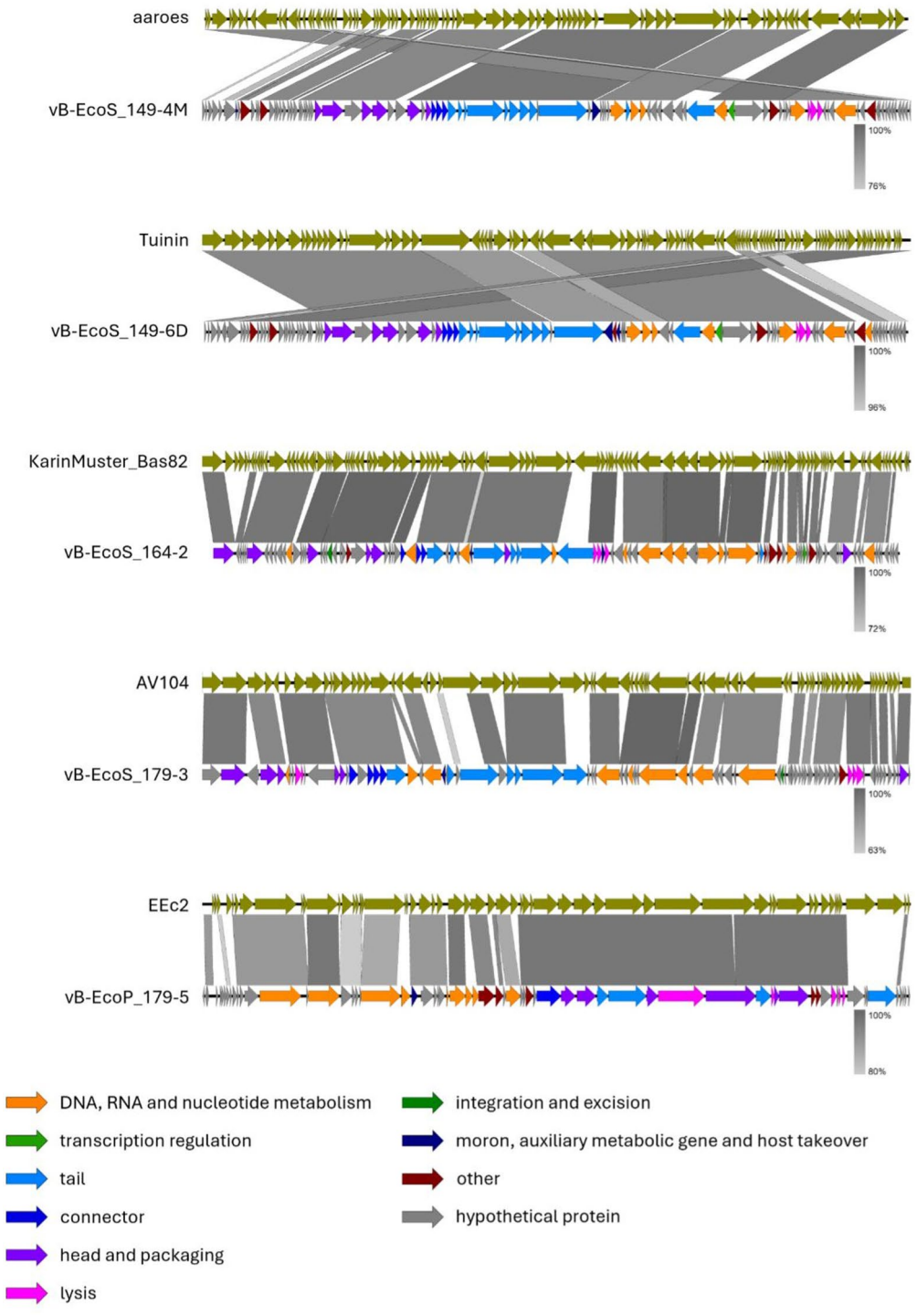


Fig. 12. Comparative genome organization of selected bacteriophages. Linear genome maps of phages vB_EcoS-149_4M, vB_EcoS-149_6D, vB_EcoS-164_2, vB_EcoS-179_3, and vB_EcoP-179_5 are shown in comparison with their closest related reference phages (selected in phylogenetic analysis). Predicted open reading frames (ORFs) are displayed as arrows indicating gene orientation and are color-coded according to functional categories, including: DNA, RNA and nucleotide metabolism (orange), transcription regulation (light green), tail (light blue), connector (dark blue), head and packaging (purple), lysis (pink), integration and excision (dark green), membrane and auxiliary metabolism (navy blue) and host takeover, other functions (dark red), and hypothetical proteins (grey). Grey-shaded regions between genomes represent nucleotide sequence similarity, illustrating conserved synteny and homologous genomic regions. The scale bar represents the percentage of nucleotide sequence identity. The comparison highlights both conserved genome architecture within related phages and structural divergence, supporting the taxonomic distinction of the newly characterized phages. Comparisons were made using Easyfig software.

The studied bacterial strains exhibit varied resistance profiles, with a concerning level of resistance to commonly used antibiotics like ampicillin. The presence of resistance in strains from different phylogenetic groups, especially to broad-spectrum antibiotics, suggests the potential complexity in treating infections caused by these strains, this is in line with our previous observations[6]. Globally, the high occurrence and increasing antibiotic resistance among uropathogenic *E. coli* constitute one of the most important health problems, which consequently generates high financial costs of treatment[80,81]. The consistent susceptibility to carbapenems and nitrofurantoin highlights these as potential treatment options for infections caused by these strains. However, the resistance observed in some strains to other commonly used antibiotics emphasizes the importance of accurate susceptibility testing before treatment. The alarming rate of drug resistance among *E. coli* strains, accompanied by the capacity for intracellular pathogenesis, has elevated to the ranks of the monitored species by European Centre for Disease Prevention and Control[82], . More than half of *E. coli* isolates were resistant to at least one antimicrobial group. Our previous study of 784 *E. coli* isolated from various clinical patients indicated similar phenomenon[6], especially in the case of penicillins and blood-derived isolates. Moreover, only 26% *E. coli* isolate were sensitive for all antibiotics. The Global Antibiotic Resistance Surveillance Report in 2025 confirmed that *E. coli* and *K. pneumoniae* are the leading drug-resistant Gram-negative bacteria found in bloodstream infections[83]. Among them, the *E. coli* uropathogens exhibit considerable variability in susceptibility to trimethoprim–sulfamethoxazole, with reported resistance levels spanning approximately 14.6% to 60% in Europe[84–88].Besides the extensive use of fluoroquinolones, particularly ciprofloxacin, has also contributed to a steady rise in resistance of UPEC bacteria to this class of drugs. Fluoroquinolone resistance in UPEC was markedly more prevalent in low- and middle-income regions (55.5–85.5%) compared with high-incomecountries, where reported values range from 5.1% to 32.0%[89]. In Poland, resstance toFQs was observed in about 30% of UPEC isolates[90]. It is worthadding that the presence of genetic determinants associated with fluoroquinolone resistance is accompanied by reduced susceptibility to these antibiotics, as observed in our studies. Although amoxicillin–clavulanate remains a recommended first-line option for the treatment of pyelonephritis and complicated urinary tract infections, its effectiveness varies geographically. Resistance rates among UPEC isolates to this medicament ranged from 5.3% in Germany[87] to 37.6% in France[91]. he continuing upward tren in antimicrobial resistance underscores the importance of prudent antibiotic use in the management of UTIs[89]. The aim of this work was therefore to thoroughly screen *E. coli* for its hidden potential for drug resistance and virulence, which could become a target for new tools to combat this pathogen. The studied strains were susceptible to most antibiotics, with the exception of one strain (EC256). Despite this, a large number of genes associated with drug resistance were detected. This observation further underscores how dangerous *E. coli* strains are. Here, most of the detected ARGs related to efflux system and the antibiotic inactivation. These data are in good agreement with literature data, where β-lactamases[89] and efflux pumps[92] are pointed as a major source of resistance.

Taken together, genomic analysis of several UPEC clinical isolates revealed substantial diversity among them, attributable to the dynamic nature of their genomes. All tested isolates carried stable core genes responsible for essential metabolic processes, virulence factors, and variable genome fragments in the form of mobile genetic elements, including prophages, plasmids, and determinants linked to multidrug resistance. The genomic variability observed among the analyzed isolates likely reflects adaptation to diverse ecological niches and poses challenges for treatment. It should be noted that each phylogenetic group was represented by a single isolate in the present study. Therefore, the observed genomic and phenotypic characteristics should be interpreted as isolate-specific observations rather than representative features of entire phylogenetic groups. Nevertheless, the findings illustrate the remarkable diversity that exists among UPEC strains belonging to different evolutionary lineages.

Because the genomic heterogeneity and multidrug resistance of UPEC may limit the effectiveness of conventional antimicrobial therapies, we next sought to isolate and characterize lytic bacteriophages using these same UPEC isolates as hosts and to determine their potential utility as precision antimicrobial agents against multidrug-resistant UPEC populations.

Five new lytic bacteriophages targeting UPEC were isolated from municipal sewage. Host range testing and EOP analyses revealed substantial variability in infectivity among these phages, with phage vB_EcoS-149_4M standing out as the most promising therapeutic candidate. This phage demonstrated the broadest and strongest activity, infecting 27% of all tested UPEC isolates with high or medium efficiency. Interestingly, several isolates allowed for phage propagation even better than the primary host, the EC149 strain. Furthermore, vB_EcoS-149_4M displayed satisfactory stability under various physical and chemical conditions, retained high infectivity during prolonged incubation with mammalian cell cultures, and exhibited no cytotoxic effects toward human bladder (T24) or kidney (HEK293) cells. An additional advantage was the lack of toxin and antibiotic resistance genes in its genome, as well as the presence of phage genes encoding lytic proteins. On the other hand, two other tested phages, vB_EcoS-179_3 and vB_EcoP-179_5, are highly specific and able to infect only the primary host, strain EC179.

Taxonomic and phylogenetic findings confirm that the investigated bacteriophages represent new species, thereby expanding current knowledge of anti-UPEC phage diversity. In total, five novel bacteriophage species were proposed in this study: *Hanrivervirus vB_EcoS-149_4M*, *Warwickvirus vB_EcoS-149_6D*, *Nonavirus vB_EcoS-164_2*, *Kagunavirus vB_EcoS-179_3*, and *Vectrevirus vB_EcoP-179_5*. All of them were clearly separated into distinct clusters, reflecting their pronounced genomic diversity and assignment to different phylogenetic lineages. Their identification enriches the pool of 98 anti-UPEC phages that we reviewed in our recent study[93] however, the total number of known lytic phages capable of infecting uropathogenic bacteria is still insufficient with respect to the number and diversity of identified UPEC strains. A good example is the highly virulent EC256 strain investigated in this study, for which no bacteriophage could be identified.

The marked differences in susceptibility to the isolated bacteriophages among the analyzed UPEC strains appear to reflect differences in their genomic architecture. Despite belonging to the same bacterial species and,

in some cases, forming closely related phylogenetic clusters, the strains exhibited distinct susceptibility patterns, as phages isolated on EC149 failed to infect EC164 and EC179, and vice versa (Figs. 2 and 3). This observation suggests that productive phage infection is primarily governed by strain-specific genomic determinants.

Comparative genomic analyses revealed considerable heterogeneity in genes associated with bacterial surface structures, including O-antigen-associated loci, surface polysaccharide biosynthesis genes, and fimbrial adhesins (Table 5). Since these structures determine the composition and accessibility of phage receptors, their differential distribution likely contributes to the observed variability in phage infectivity and host range. This interpretation is further supported by the broad host range of vB_EcoS-149_4M and the highly restricted activities of vB_EcoS-164_2, vB_EcoS-179_3, and vB_EcoP-179_5, as well as by the inability to isolate phages infecting EC3 and EC256 (Fig. 3).

In addition, differences in prophage content and CRISPR-Cas systems (Tables 6 and 7) may further modulate phage susceptibility through mechanisms such as superinfection exclusion, modifications of cell surface structures, or other defence functions encoded within mobile genetic elements. Collectively, our findings indicate that genomic heterogeneity among UPEC strains is accompanied by substantial variability in susceptibility to lytic bacteriophages, highlighting the importance of considering bacterial genomic diversity when developing precision phage therapy strategies against UPEC infections.

The combined bacterial and phage characterization highlights both the complexity of UPEC as a pathogen and the potential of well-selected bacteriophages as therapeutic tools. This broader study provides an essential groundwork for phage therapy research, as it not only identifies effective anti-UPEC phages but also includes their detailed assessments, highlighting their potential relevance for future applications. The efficacy of vB_EcoS-149_4M, in terms of host range, stability, and safety, positions it as a strong candidate for further preclinical development. In contrast, due to extremely high specificity toward one bacterial strain and sensitivity to most external factors, phages vB_EcoS-179_3 and vB_EcoP-179_5 do not seem to be appropriate for universal therapy, although they may be used in an individual approach. Such preclinical characterization is a critical step before moving toward in vivo testing and eventual clinical application to combat drug-resistant UPEC infections.

Overall, this research lays the groundwork for phage therapy against MDR UPEC infections by providing (i) a genomic insight into strain-specific determinants, (ii) a detailed preclinical assessment of candidates for therapeutic phages, (iii) evidence supporting the feasibility of integrating genomic data with phage selection to enhance therapeutic precision. Such an approach may have an impact not only on future UTI treatment protocols but also on broader phage therapy applications against other antibiotic-resistant bacterial pathogens.

Conclusions

Urinary tract infections caused by uropathogenic *Escherichia coli* remain a significant clinical challenge, largely due to the high genetic diversity, adaptive potential, and increasing antimicrobial resistance of these bacteria. The comparative genomic analysis performed in this study confirmed that UPEC strains possess a conserved core genome alongside highly dynamic mobile genetic elements, including plasmids and prophages. These elements not only contribute to virulence and environmental adaptability but also facilitate the dissemination of antibiotic resistance determinants, including those conferring resistance to last-resort drugs. The presence of numerous resistance-associated genes, even in phenotypically susceptible isolates, further emphasizes the hidden threat posed by these pathogens and highlights the limitations of conventional antibiotic-based therapies. Importantly, our findings demonstrate that integrating genomic insights with bacteriophage research provides a promising avenue for the development of alternative therapeutic strategies.

The combined genomic and phage-based approach presented here provides a solid foundation for future investigations and supports the potential implementation of precision phage therapy as an effective tool against multidrug-resistant UPEC pathogens.

Data availability

Data availabilityAssembled genome sequences were deposited by the BankIt tool in the GenBank database under the following accession numbers: GCF_036282145.1 for EC3, GCF_036281895.1 for EC149, GCF_036281865.1 for EC164, GCF_036281825.1 for EC179, GCF_036281835.1 for EC256, PX985795 for vB_EcoS-149_4M, PX985796 for vB_EcoS-149_6D, PX985797 for vB_EcoS-164_2, PX985798 for vB_EcoS-179_3, and PX985799 for vB_EcoP-179_5. These genomic sequences are also available in the Zenodo repository under the following access link: https://doi.org/10.5281/zenodo.20098588.

Received: 4 May 2026; Accepted: 14 July 2026
Published online: 21 July 2026

References

1. Mengistu, D. A. et al. Incidence of urinary tract infection among patients: systematic review and meta-analysis, **60**, 469580231168746. https://doi.org/10.1177/00469580231168746 (2023).
2. Cornelius, S. A., Basu, U., Zimmern, P. E. & De Nisco, N. J. Overcoming challenges in the management of recurrent urinary tract infections. *Expert Rev. Anti-Infective Therapy*. **22**, 1157–1169. https://doi.org/10.1080/14787210.2024.2412628 (2024).
3. Thompson, J., Marijam, A., Mitrani-Gold, F. S., Wright, J. & Joshi, A. V. Activity impairment, health-related quality of life, productivity, and self-reported resource use and associated costs of uncomplicated urinary tract infection among women in the United States. *PLoS One*. **18,** e0277728. https://doi.org/10.1371/journal.pone.0277728 (2023).
4. Torres-Puig, S. et al. Omics Technologies - What have they told us about uropathogenic *Escherichia coli* fitness and virulence during urinary tract infection? *Front. Cell. Infect. Microbiol.* **12**, 824039. https://doi.org/10.3389/fcimb.2022.824039 (2022).
5. Zhou, Y. et al. Urinary tract infections caused by uropathogenic *Escherichia coli*: mechanisms of infection and treatment options. *Int. J. Mol. Sci.* **24**, 10537. https://doi.org/10.3390/ijms241310537 (2023).

6. Dziuba, A. et al. Comparative study of virulence potential, phylogenetic origin, CRISPR-Cas regions and drug resistance of *Escherichia coli* isolates from urine and other clinical materials. *Front. Microbiol.* **14**, 1289683. https://doi.org/10.3389/fmicb.2023.1289683 (2023).
7. Dziuba, A., Białek, J., Wawszczak-Kasza, M., Janczura, J. & Dulębska, J. The Role of Intracellular Bacterial Communities of Uropathogenic *Escherichia Coli* in Chronic Urinary Tract Infection and New Therapeutic Ideas. *Curr. Clin. Microbiol. Rep.* **12,** 16. https://doi.org/10.1007/s40588-025-00253-0 (2025).
8. Adamus-Białek, W. et al. Ciprofloxacin, amoxicillin, and aminoglycosides stimulate genetic and phenotypic changes in uropathogenic *Escherichia coli* strains. *Virulence* **10**, 260–276. https://doi.org/10.1080/21505594.2019.1596507 (2019).
9. Bayne, C. E., Farah, D., Herbst, K. W. & Hsieh, M. H. Role of urinary tract infection in bladder cancer: a systematic review and meta-analysis. *World J. Urol.* **36**, 1181–1190. https://doi.org/10.1007/s00345-018-2257-z (2018).
10. Li, X. et al. Characterization of a novel phage vB_EcoP_P64441 and its potential role in controlling uropathogenic *Escherichia coli* (UPEC) and biofilms formation. *Virology.* **609**, 110570. https://doi.org/10.1016/j.virol.2025.110570 (2025).
11. Zulk, J. J. et al. Phage Resistance Accompanies Reduced Fitness of Uropathogenic *Escherichia coli* in the Urinary Environment. *mSphere* **7**, e0034522. https://doi.org/10.1128/msphere.00345-22 (2022).
12. Markusková, B. et al. Characterization of bacteriophages infecting multidrug-resistant uropathogenic *Escherichia coli* strains. *Arch. Virol.* **169**. 142. https://doi.org/10.1007/s00705-024-06063-x (2024).
13. Mirzahosseini, H. K. et al. Correlation of biofilm formation, virulence factors, and phylogenetic groups among *Escherichia coli* strains causing urinary tract infection: A global systematic review and meta-analysis. *J. Res. Med. Sci.* **28**, 66. https://doi.org/10.4103/jrms.jrms_637_22 (2023).
14. Halaji, M. et al. Phylogenetic Group Distribution of Uropathogenic *Escherichia coli* and Related Antimicrobial Resistance Pattern: A Meta-Analysis and Systematic Review. *Front. Cell. Infect. Microbiol.* **12**, 790184. https://doi.org/10.3389/fcimb.2022.790184 (2022).
15. Subhadra, B. et al. Complete genome sequence of uropathogenic *Escherichia coli* isolate UPEC 26 – 1. *Genes Genomics.* **40**, 643–655. https://doi.org/10.1007/s13258-018-0665-5 (2018).
16. Overbeek, R. et al. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res.* **42**, D206–D214. https://doi.org/10.1093/nar/gkt1226 (2014).
17. Aziz, R. K. et al. The RAST Server: Rapid annotations using subsystems technology. *BMC Genom.* **9**, 1–15. https://doi.org/10.1186/1471-2164-9-75 (2008).
18. Tatusova, T. et al. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* **44**, 6614–6624. https://doi.org/10.1093/nar/gkw569 (2016).
19. Kaas, R. S., Leekitcharoenphon, P., Aarestrup, F. M. & Lund, O. Solving the problem of comparing whole bacterial genomes across different sequencing platforms. *PLoS One.* **9**, e104984. https://doi.org/10.1371/journal.pone.0104984 (2014).
20. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296. https://doi.org/10.1093/nar/gkab301 (2021).
21. Lee, I., Kim, Y. O., Park, S. C. & Chun, J. OrthoANI: An improved algorithm and software for calculating average nucleotide identity. *Int. J. Syst. Evol. Microbiol.* **66**, 1100–1103. https://doi.org/10.1099/ijsem.0.000760 (2016).
22. Chalita, M. et al. EzBioCloud: a genome-driven database and platform for microbiome identification and discovery. *Int. J. Syst. Evol. Microbiol.* **74**, 006421. https://doi.org/10.1099/ijsem.0.006421 (2024).
23. Yoon, S. H., Ha, S., min, Lim, J., Kwon, S. & Chun, J. A large-scale evaluation of algorithms to calculate average nucleotide identity. Antonie van Leeuwenhoek. *Int. J. Gen. Mol. Microbiol.* **110**, 1281–1286. https://doi.org/10.1007/s10482-017-0844-4 (2017).
24. Zhou, S., Liu, B., Zheng, D., Chen, L. & Yang, J. VFDB 2025: an integrated resource for exploring anti-virulence compounds. *Nucleic Acids Res.* **53**, D871–D877. https://doi.org/10.1093/nar/gkae968 (2024).
25. Couvin, D. et al. CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res.* **46**, W246–W251. https://doi.org/10.1093/nar/gky425 (2018).
26. Joensen, K. G., Tetzschner, A. M. M., Iguchi, A., Aarestrup, F. M. & Scheutz, F. Rapid and easy in silico serotyping of Escherichia coli isolates by use of whole-genome sequencing data. *J. Clin. Microbiol.* **53**, 2410–2426. https://doi.org/10.1128/JCM.00008-15 (2015).
27. Carattoli, A. et al. In Silico detection and typing of plasmids using plasmidfinder and plasmid multilocus sequence typing. *Antimicrob. Agents Chemother.* **58**, 3895–3903. https://doi.org/10.1128/AAC.02412-14 (2014).
28. Alcock, B. P. et al. CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. *Nucleic Acids Res.* **51**, D690–D699. https://doi.org/10.1093/nar/gkac920 (2023).
29. Wishart, D. S. et al. PHASTEST: Faster than PHASTER, better than PHAST. *Nucleic Acids Res.* **51**, W443–W450. https://doi.org/10.1093/nar/gkad382 (2023).
30. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858. https://doi.org/10.1038/nprot.2015-053 (2015).
31. Shen, A., & Millard, A. Phage genome annotation: where to begin and end. *Appl. Res.* **2,** 183–193. https://doi.org/10.1089/phage.2021.0015. (2021).
32. Prjibelski, A., Antipov, D., Meleshko, D., Lapidus, A. & Korobeynikov, A. Using SPAdes de novo assembler. *Curr. Protoc. Bioinf.* **70**, e102. https://doi.org/10.1002/cpbi.102 (2020).
33. Walker, B. J. et al. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One.* **9**, e112963. https://doi.org/10.1371/journal.pone.0112963 (2014).
34. Garneau, J. R. et al. High-throughput identification of viral termini and packaging mechanisms in virome datasets using PhageTermVirome. *Sci. Rep.* **11**, 18319. https://doi.org/10.1038/s41598-021-97867-3 (2021).
35. Bouras, G. et al. Pharokka: a fast scalable bacteriophage annotation tool. *Bioinformatics* **39**, btac776. https://doi.org/10.1093/bioinformatics/btac776 (2023).
36. Shang, J., Tang, X. & Sun, Y. PhaTYP: predicting the lifestyle for bacteriophages using BERT. *Brief. Bioinform.* **24**, bbac487. https://doi.org/10.1093/bib/bbac487 (2023).
37. Feldgarden, M. et al. AMRFinderPlus and the Reference Gene Catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. *Sci. Rep.* **11**, 12728. https://doi.org/10.1038/s41598-021-91456-0 (2021).
38. Joensen, K. G. et al. Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic Escherichia coli. *J. Clin. Microbiol.* **52**, 1501–1510. https://doi.org/10.1128/JCM.03617-13 (2014).
39. Bortolaia, V. et al. ResFinder 4.0 for predictions of phenotypes from genotypes. *J. Antimicrob. Chemother.* **75**, 3491–3500. https://doi.org/10.1093/jac/dkaa345 (2020).
40. Grant, J. R. et al. Proksee: In-depth characterization and visualization of bacterial genomes. *Nucleic Acids Res.* **51**, W484–W492. https://doi.org/10.1093/nar/gkad326 (2023).
41. Sullivan, M. J., Petty, N. K., & Beatson, S. A. Easyfig: a genome comparison visualizer. *Bioinformatics* **27**, 1009–1010. https://doi.org/10.1093/bioinformatics/btr039 (2011).
42. Shang, J., Jiang, J. & Sun, Y. Bacteriophage classification for assembled contigs using graph convolutional network. *Bioinformatics* **37**, I25–33. https://doi.org/10.1093/bioinformatics/btab293 (2021).

43. Moraru, C., Varsani, A. & Kropinski, A. M. VIRIDIC - a novel tool to calculate the intergenomic similarities of prokaryote-infecting viruses. *Viruses* **12**, https://doi.org/10.3390/v12111268 (2020).
44. Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539. https://doi.org/10.1038/msb.2011.75 (2011).
45. Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973. https://doi.org/10.1093/bioinformatics/btp348 (2009).
46. Wong, T. K. et al. IQ-TREE 3: phylogenomic inference software using complex evolutionary models. *Mol Biol Evol.* **43**, msag117. https://doi.org/10.1093/molbev/msag117 (2026).
47. Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods*. **17**, 261–272. https://doi.org/10.1038/s41592-019-0686-2 (2020).
48. Topka, G. et al. Characterization of bacteriophage vB-EcoS-95, isolated from urban sewage and revealing extremely rapid lytic development. *Front. Microbiol.* **9**, 3326. https://doi.org/10.3389/fmicb.2018.03326 (2019).
49. Topka-Bielecka, G. et al. Characterization of the bacteriophage vB_EfaS-271 infecting *Enterococcus faecalis*. *Int. J. Mol. Sci.* **21**, 1–27. https://doi.org/10.3390/ijms21176345 (2020).
50. Mirzaei, M. K. & Nilsson, A. S. Isolation of phages for phage therapy: A comparison of spot tests and efficiency of plating analyses for determination of host range and efficacy. *PLoS One*. **10**, e0127606. https://doi.org/10.1371/journal.pone.0118557 (2015).
51. Sambrook, J. & Russell, D. W. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, 2001).
52. Kropinski, A. M., Prangishvili, D. & Lavigne, R. Position paper: The creation of a rational scheme for the nomenclature of viruses of Bacteria and Archaea. *Environ. Microbiol.* **11**, 2775–2777. https://doi.org/10.1111/j.1462-2920.2009.01970.x (2009).
53. Zeng, C. et al. DNA packaging and genomics of the *Salmonella* 9NA-Like phages. *J. Virol.* **93.** https://doi.org/10.1128/jvi.00848-19 (2019).
54. Timm Morgan, R., Russell Seongmi, K. & Hultgren Scott, J. Urinary tract infections: pathogenesis, host susceptibility and emerging therapeutics. *Nat. Rev. Microbiol.* **23**, 72–86. https://doi.org/10.1038/s41579-024-01092-4 (2025).
55. Bhimani, A. et al. Multidrug-resistant pathogens in UTIs: A concise review of epidemiology, diagnostics, and patent landscaping for precision healthcare. *Diagn. Microbiol. Infect. Dis.* **114**, 117125. https://doi.org/10.1016/j.diagmicrobio.2025.117125 (2026).
56. Godijk, N. G. et al. New methodology to assess the excess burden of antibiotic resistance using country-specific parameters: A case study regarding *E. coli* urinary tract infections. *BMJ Open.* **13**, e064335. https://doi.org/10.1136/bmjopen-2022-064335 (2023).
57. Klein, R. D. & Hultgren, S. J. Urinary tract infections: microbial pathogenesis, host–pathogen interactions and new treatment strategies. *Nat. Rev. Microbiol.* **18**, 211–226. https://doi.org/10.1038/s41579-020-0324-0 (2020).
58. Rosen, D. A., Hooton, T. M., Stamm, W. E., Humprey, P. A. & Hultgren, S. J. Detection of Intracellular Bacterial Communities in Human Urinary Tract Infection. *PLoS Med.* **4**, e329. https://doi.org/10.1371/journal (2007).
59. Zalewska-Piatek, B. & Piatek, R. Alternative treatment approaches of urinary tract infections caused by uropathogenic *Escherichia coli* strains. *Acta Biochim. Pol.* **66**, 129–138. https://doi.org/10.18388/abp.2018_2787 (2019).
60. Rozwadowski, M. & Gawel, D. Molecular Factors and Mechanisms Driving Multidrug Resistance in Uropathogenic *Escherichia coli*—An Update. *Genes* **13**, 1397. https://doi.org/10.3390/genes13081397 (2022).
61. Mahshouri, P., Alikhani, M. Y., Momtaz, H. E., Doosti-Irani, A. & Shokoohizadeh, L. Analysis of phylogroups, biofilm formation, virulence factors, antibiotic resistance and molecular typing of uropathogenic *Escherichia coli* strains isolated from patients with recurrent and non-recurrent urinary tract infections. *BMC Infect. Dis.* **25**, 267. https://doi.org/10.1186/s12879-025-10635-w (2025).
62. Flores, C. et al. A human urothelial microtissue model reveals shared colonization and survival strategies between uropathogens and commensals. *Sci. Adv.* **9**, eadi9834. https://doi.org/10.1126/sciadv.adi9834 (2023).
63. Ko, Y. H., Choi, J. Y. & Song, P. H. Host-Pathogen Interactions in Urinary Tract Infections. *Urogenit. Tract. Infect.* **14**, 71. https://doi.org/10.14777/uti.2019.14.3.71 (2019).
64. Schreiber, H. L., Spaulding, C. N., Dodson, K. W., Livny, J. & Hultgren, S. J. One size doesn't fit all: Unraveling the diversity of factors and interactions that drive E. coli urovirulence. *Annals Translational Med.* **5**, 28. https://doi.org/10.21037/atm.2016.12.73 (2017).
65. Schreiber, H. L. et al. Bacterial virulence phenotypes of *Escherichia coli* and host susceptibility determine risk for urinary tract infections. *Sci. Transl Med.* **9**, eaaf1283. https://doi.org/10.1126/scitranslmed.aaf1283 (2017).
66. Wesołowski, W. et al. Phage Endolysins as Promising and Effective Candidates for Use Against Uropathogenic *Escherichia coli*. *Viruses* **17**, 560. https://doi.org/10.3390/v17040560 (2025).
67. Hussein, A. H., Makky, S., Hager, R., Connerton, I. F. & El-Shibiny, A. Characterization and therapeutic potential of phage vB_Eco_ZCEC08 against multidrug-resistant uropathogenic *Escherichia coli*. *BMC Microbiol.* **25**, 221. https://doi.org/10.1186/s12866-025-03903-x (2025).
68. Rodea, G. E. et al. Genomic characterization of two bacteriophages (vB_EcoS-phiEc3 and vB_EcoS-phiEc4) belonging to the genus Kagunavirus with lytic activity against uropathogenic *Escherichia coli*. *Microb. Pathog.* **165**, 105494. https://doi.org/10.1016/j.micpath.2022.105494 (2022).
69. Lo, Y., Zhang, L., Foxman, B. & Zöllner, S. Whole-genome sequencing of uropathogenic *Escherichia coli* reveals long evolutionary history of diversity and virulence. *Infect. Genet. Evol.* **34**, 244–250. https://doi.org/10.1016/j.meegid.2015.06.023 (2015).
70. Nash, J. H. E. et al. Genome sequence of adherent-invasive *Escherichia coli* and comparative genomic analysis with other *E. coli* pathotypes. *BMC Genom.* **11**, 667. https://doi.org/10.1186/1471-2164-11-667 (2010).
71. Thänert, R. et al. Persisting uropathogenic *Escherichia coli* lineages show signatures of niche-specific within-host adaptation mediated by mobile genetic elements. *Cell. Host Microbe*. **30**, 1034–1047e6. https://doi.org/10.1016/j.chom.2022.04.008 (2022).
72. Morales, G. et al. The Role of Mobile Genetic Elements in Virulence Factor Carriage from Symptomatic and Asymptomatic Cases of *Escherichia coli* Bacteriuria. *Microbiol. Spectr.* **11**, e0471022. https://doi.org/10.1128/spectrum.04710-22 (2023).
73. Rijavec, M. et al. High prevalence of multidrug resistance and random distribution of mobile genetic elements among uropathogenic *Escherichia coli* (UPEC) of the four major phylogenetic groups. *Curr. Microbiol.* **53**, 158–162. https://doi.org/10.1007/s00284-005-0501-4 (2006).
74. Kopotsa, K., Osei Sekyere, J. & Mbelle, N. M. Plasmid evolution in carbapenemase-producing Enterobacteriaceae: a review. *Ann. N. Y. Acad. Sci.* **1457**, 61–91. https://doi.org/10.1111/nyas.14223 (2019).
75. Navon-Venezia, S., Kondratyeva, K. & Carattoli, A. *Klebsiella pneumoniae*: A major worldwide source and shuttle for antibiotic resistance. *FEMS Microbiol. Rev.* **41**, 252–275. https://doi.org/10.1093/femsre/fux013 (2017).
76. Lin, H. et al. Characteristics of the plasmid-mediated colistin-resistance gene mcr-1 in *Escherichia coli* isolated from a veterinary hospital in Shanghai. *Front. Microbiol.* **13**, 1002827. https://doi.org/10.3389/fmicb.2022.1002827 (2022).
77. Duggett, N. A. et al. Molecular epidemiology of isolates with multiple mcr plasmids from a pig farm in Great Britain: The effects of colistin withdrawal in the short and long term. *J. Antimicrob. Chemother.* **73**, 3025–3033. https://doi.org/10.1093/jac/dky292 (2018).
78. Getino, M. et al. A Broad-Host-Range Plasmid Outbreak: Dynamics of IncL/M Plasmids Transferring Carbapenemase Genes. *Antibiotics* **11**, 1641. https://doi.org/10.3390/antibiotics11111641 (2022).

79. Egido, J. E., Costa, A. R., Aparicio-Maldonado, C., Haas, P. J. & Brouns, S. J. J. Mechanisms and clinical importance of bacteriophage resistance. *FEMS Microbiol. Rev.* **46**, fuab048. https://doi.org/10.1093/femsre/fuab048 (2022).
80. Yang, X. et al. Disease burden and long-term trends of urinary tract infections: A worldwide report. *Front. Public. Health* **10**, 888205. https://doi.org/10.3389/fpubh.2022.888205 (2022).
81. Mancuso, G. et al. Urinary Tract Infections: The Current Scenario and Future Prospects. *Pathogens* **12**, 623. https://doi.org/10.3390/pathogens12040623 (2023).
82. ECDC. European Centre for Disease Prevention and Control. *Antimicrobial resistance in the EU/EEA (EARS-Net) - Annual Epidemiological Report 2020* (Stockholm, 2022).
83. WHO. Global antibiotic resistance surveillance report 2025. (2025).
84. Abduzaimovic, A. et al. Antibiotic Resistance in Urinary Isolates of *Escherichia coli*. *Materia Socio Med.* **28**, 416. https://doi.org/10.5455/msm.2016.28.416-419 (2016).
85. Erb, S. et al. Basic patient characteristics predict antimicrobial resistance in *E. coli* from urinary tract specimens: a retrospective cohort analysis of 5246 urine samples. *Swiss Med. Wkly.* **148**, w14660. https://doi.org/10.4414/smw.2018.14660 (2018).
86. Kot, B. et al. Virulence factors, biofilm-forming ability, and antimicrobial resistance of urinary *Escherichia coli* strains isolated from hospitalized patients. *Turk. J. Med. Sci.* **46**, 1908–1914. https://doi.org/10.3906/sag-1508-105 (2016).
87. Hitzenbichler, F. et al. Antibiotic resistance in *E. coli* isolates from patients with urinary tract infections presenting to the emergency department. *Infection* **46**, 325–331. https://doi.org/10.1007/s15010-018-1117-5 (2018).
88. Kresken, M. et al. Comparative in vitro activity of oral antimicrobial agents against Enterobacteriaceae from patients with community-acquired urinary tract infections in three European countries. *Clin. Microbiol. Infect.* **22**, 63e1. https://doi.org/10.1016/j.cmi.2015.08.019 (2016).
89. Kot, B. Antibiotic Resistance among Uropathogenic *Escherichia coli*. *Pol. J. Microbiol.* **68**, 403–415. https://doi.org/10.33073/PJM-2019-048 (2019).
90. Michno, M., Sydor, A., Wałaszek, M. & Sułowicz, W. Microbiology and drug resistance of pathogens in patients hospitalized at the nephrology department in the South of Poland. *Pol. J. Microbiol.* **67**, 517–524. https://doi.org/10.21307/pjm-2018-061 (2018).
91. Lavigne, J. P. et al. Resistance and virulence potential of uropathogenic *Escherichia coli* strains isolated from patients hospitalized in urology departments: A French prospective multicentre study. *J. Med. Microbiol.* **65**, 530–537. https://doi.org/10.1099/jmm.0.000247 (2016).
92. Sun, D. et al. (ed) Investigation of antibiotic resistance determinants and virulence factors of uropathogenic *Escherichia coli*. *J. Antibiot.* **73,** 314–319. https://doi.org/10.1038/s41429-020-0284-7 (2020).
93. Łukasiak, A. et al. Groundwork for phage therapy: multi-faceted comparative analysis of lytic bacteriophages infecting uropathogenic *Escherichia coli*. *Virology* **615**, 110738. https://doi.org/10.1016/j.virol.2025.110738 (2026).

Acknowledgements

We would like to thank Dawid Gniazdo for his contribution to the description of virulence-associated genes.

Author contributions

Conceptualization: WW, BN-F, WA-B• Data curation JM, GT-B, AN, DK, EW, WA-B, AD, EJ, KZ, GC,. Formal analysis WW, AŁ, ŁG, SB, WA-B, EJ, KZ, GC,. Funding acquisition WW, WA-B, GC. Investigation WW, AŁ, ŁG, SB, MN, WA-B, EJ, KZ, GC,. Methodology WW, AŁ, ŁG, JM, SB, BN-F, GC. Project administration WW, WA-B. Resources WA-B, AD, BN-F. Supervision BN-F, WA-B, GC. Validation WW, SB, BN-F. Visualization WW, BN-F, GC. Writing – original draft WW, BN-F, WA-B, EJ, KZ, GC,. Writing – review and editing AŁ, JM, SB, GW, WA-B, BN-F, GC.

Funding

This work was supported by the project no. 024/RID/2018/19 (WA-B) obtained within the program of the Minister of Science and Higher Education named "Regional Excellence Initiative in 2019–2022"; the project no.: RID/SP/0015/2024/01 (GC) financed by the program of the Minister of Science named "Regional Excellence Initiative", and the UGrants-start grant number 533-BG10-GS0I-24 (WW).

Declarations

Competing interests

The authors declare no competing interests.

Consent for publication

All authors have reviewed and approved the final version of the manuscript and consent to its publication.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-026-62793-9.

Correspondence and requests for materials should be addressed to B.N.-F. or W.A.-B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2026

Suplement do pracy nr 2

Supplementary Figure S1. Endotoxin levels in purified phage lysates. Endotoxin concentrations (EU/mL) were measured in solutions of phages vB_EcoS-149_4M, vB_EcoS-149_6D, vB_EcoS-164_2, vB_EcoS-179_3, and vB_EcoP-179_5 using the Thermo Scientific Pierce chromogenic endotoxin quant kit. Bars represent mean values, and error bars indicate standard deviation from three replicate measurements.

Supplementary Figure S2. Circular genome maps and functional annotation of newly characterized bacteriophages. Circular genome maps of phages (A) vB_EcoS-149_4M, (B) vB_EcoS-149_6D, (C) vB_EcoS-164_2, (D) vB_EcoS-179_3, and (E) vB_EcoP-179_5 are shown. Predicted open reading frames (ORFs) are displayed as arrows indicating transcriptional orientation and are color-coded according to functional categories, including DNA, RNA and nucleotide metabolism, transcription regulation, tail, connector, head and packaging, lysis, integration and excision, moron/auxiliary metabolism and host takeover, other functions, and hypothetical proteins (legend on the right side of each map). The inner rings represent GC content and GC skew across each genome. Annotated genes involved in genome replication, virion structure and assembly, host interaction, and lysis are indicated. Genome coordinates (kbp) are shown inside each map.

Supplementary Figure S3. Hierarchical cluster analysis of all bacteriophages analyzed in this study based on the VIRIDIC intergenomic distance matrix. Clustering was performed using the complete linkage method. The dendrogram illustrates genome-wide similarity relationships among all included phages and resolves them into well-defined clusters corresponding to major taxonomic lineages. The red dashed line indicates the species demarcation threshold, and the blue dashed line indicates the genus demarcation threshold. The positions of the newly characterized phages vB_EcoS-149_4M, vB_EcoS-149_6D, vB_EcoS-164_2, vB_EcoS-179_3, and vB_EcoP-179_5 highlight their separation into distinct clusters and support their classification into the genera *Hanrivervirus*, *Warwickvirus*, *Nonavirus*, *Kagunavirus*, and *Vectrevirus*, respectively.

Supplementary Table S1. ARG detected in UPEC genomes.

Supplementary Table S2. Identified orthologous genes and their annotated functions. Genes indicated by "+" contained a sufficient amount of parsimony-informative sites and were therefore retained for phylogenetic reconstruction.

Supplementary Table S3. Phage lifecycle classification predicted by the PhaTYP software.

Supplementary Table S1. ARG detected in UPEC genomes.

No	ARO Term	Detection criteria	AMR Gene Family	Drug Class	Resistance mechanism	Strain
1.	baeR	protein homolog model	resistance-nodulation-cell division (RND) antibiotic efflux pump	aminoglycoside antibiotic, aminocoumarin antibiotic	antibiotic efflux	EC149, EC164, EC179, EC256
2.	marA	protein homolog model	resistance-nodulation-cell division (RND) antibiotic efflux pump, General Bacterial Porin with reduced permeability to beta-lactams	fluoroquinolone antibiotic, monobactam, carbapenem, cephalosporin, glycylcycline, cephamycin, penam, tetracycline antibiotic, rifamycin antibiotic, phenicol antibiotic, penem, disinfecting agents and antiseptics	antibiotic efflux, reduced permeability to antibiotic	EC149, EC164, EC179, EC256
3.	H-NS	protein homolog model	major facilitator superfamily (MFS) antibiotic efflux pump, resistance-nodulation-cell division (RND) antibiotic efflux pump	macrolide antibiotic, fluoroquinolone antibiotic, cephalosporin, cephamycin, penam, tetracycline antibiotic	antibiotic efflux	EC149, EC164, EC179, EC256
4.	mdtG	protein homolog model	major facilitator superfamily (MFS) antibiotic efflux pump	phosphonic acid antibiotic	antibiotic efflux	EC149, EC164, EC179, EC256
5.	msbA	protein homolog model	ATP-binding cassette (ABC) antibiotic efflux pump	nitroimidazole antibiotic	antibiotic efflux	EC149, EC164, EC179, EC256
6.	kdpE	protein homolog model	kdpDE	aminoglycoside antibiotic	antibiotic efflux	EC149, EC164, EC179, EC256

7.	acrB	protein homolog model	resistance-nodulation-cell division (RND) antibiotic efflux pump	fluoroquinolone antibiotic, cephalosporin, glycylcycline, penam, tetracycline antibiotic, rifamycin antibiotic, phenicol antibiotic, disinfecting agents and antiseptics	antibiotic efflux	EC149, EC164, EC179, EC256
8.	sul2	protein homolog model	sulfonamide resistant sul	sulfonamide antibiotic	antibiotic target replacement	EC149, EC256
9.	cpxA	protein homolog model	resistance-nodulation-cell division (RND) antibiotic efflux pump	aminoglycoside antibiotic, aminocoumarin antibiotic	antibiotic efflux	EC149, EC164, EC256
10.	SAT-2	protein homolog model	streptothricin acetyltransferase (SAT)	nucleoside antibiotic	antibiotic inactivation	EC256
11.	mdtE	protein homolog model	resistance-nodulation-cell division (RND) antibiotic efflux pump	macrolide antibiotic, fluoroquinolone antibiotic, penam	antibiotic efflux	EC149, EC164, EC179, EC256
12.	TolC	protein homolog model	ATP-binding cassette (ABC) antibiotic efflux pump, major facilitator superfamily (MFS) antibiotic efflux pump, resistance-nodulation-cell division (RND) antibiotic efflux pump	macrolide antibiotic, fluoroquinolone antibiotic, aminoglycoside antibiotic, carbapenem, cephalosporin, glycylcycline, cephamycin, penam, tetracycline antibiotic, peptide antibiotic, aminocoumarin antibiotic, rifamycin antibiotic, phenicol antibiotic, penem, disinfecting agents and antiseptics	antibiotic efflux	EC149, EC164, EC179, EC256
13.	emrY	protein homolog model	major facilitator superfamily (MFS) antibiotic efflux pump	tetracycline antibiotic	antibiotic efflux	EC149, EC169, EC179

						, EC256
14.	TEM-1	protein homolog model	TEM beta-lactamase	monobactam, cephalosporin, penam, penem	antibiotic inactivation	EC149, EC179, EC256
15.	dfrA1	protein homolog model	trimethoprim resistant dihydrofolate reductase dfr	diaminopyrimidine antibiotic	antibiotic target replacement	EC256
16.	dfrA12	protein homolog model	trimethoprim resistant dihydrofolate reductase dfr	diaminopyrimidine antibiotic	antibiotic target replacement	EC256
17.	aadA2	protein homolog model	ANT(3")	aminoglycoside antibiotic	antibiotic inactivation	EC256
18.	qacEdelta1	protein homolog model	major facilitator superfamily (MFS) antibiotic efflux pump	disinfecting agents and antiseptics	antibiotic efflux	EC256
19.	sul1	protein homolog model	sulfonamide resistant sul	sulfonamide antibiotic	antibiotic target replacement	EC256
20.	armA	protein homolog model	16S rRNA methyltransferase (G1405)	aminoglycoside antibiotic	antibiotic target alteration	EC256
21.	msrE	protein homolog model	msr-type ABC-F protein	macrolide antibiotic, streptogramin antibiotic	antibiotic target protection	EC256
22.	mphE	protein homolog model	macrolide phosphotransferase (MPH)	macrolide antibiotic	antibiotic inactivation	EC3, EC256
23.	CTX-M-3	protein homolog model	CTX-M beta-lactamase	cephalosporin, cephamycin, penam	antibiotic inactivation	EC256
24.	baeS	protein homolog model	resistance-nodulation-cell division (RND) antibiotic efflux pump	aminoglycoside antibiotic, aminocoumarin antibiotic	antibiotic efflux	EC149, EC164, EC179, EC256
25.	mdtC	rotein homolog model	resistance-nodulation-cell division (RND) antibiotic efflux pump	aminocoumarin antibiotic	antibiotic efflux	EC149, EC164, EC179, EC256
26.	mdtB	protein homolog model	resistance-nodulation-cell division (RND) antibiotic efflux pump	aminocoumarin antibiotic	antibiotic efflux	EC149, EC164, EC179

						, EC256
27.	mdtA	protein homolog model	resistance-nodulation-cell division (RND) antibiotic efflux pump	aminocoumarin antibiotic	antibiotic efflux	EC149, EC164, EC179, EC256
28.	ugd	protein homolog model	pmr phosphoethanolamine transferase	peptide antibiotic	antibiotic target alteration	EC149, EC164, EC179, EC256
29.	Escherichia coli emrE	protein homolog model	small multidrug resistance (SMR) antibiotic efflux pump	macrolide antibiotic	antibiotic efflux	EC164, EC256
30.	Klebsiella pneumoniae KpnE	protein homolog model	small multidrug resistance (SMR) antibiotic efflux pump	macrolide antibiotic, aminoglycoside antibiotic, cephalosporin, tetracycline antibiotic, peptide antibiotic, rifamycin antibiotic, disinfecting agents and antiseptics	antibiotic efflux	EC149, EC164, EC179, EC256
31.	Klebsiella pneumoniae KpnF	protein homolog model	small multidrug resistance (SMR) antibiotic efflux pump	macrolide antibiotic, aminoglycoside antibiotic, cephalosporin, tetracycline antibiotic, peptide antibiotic, rifamycin antibiotic, disinfecting agents and antiseptics	antibiotic efflux	EC149, EC164, EC179, EC256
32.	mdtH	protein homolog model	major facilitator superfamily (MFS) antibiotic efflux pump	fluoroquinolone antibiotic	antibiotic efflux	EC149, EC164, EC179, EC256
33.	Escherichia coli mdfA	protein homolog model	major facilitator superfamily (MFS) antibiotic efflux pump	tetracycline antibiotic, disinfecting	antibiotic efflux	EC149, EC164

				agents and antiseptics		, EC179 , EC256
34.	Escherichia coli emrE	protein homolog model	small multidrug resistance (SMR) antibiotic efflux pump	macrolide antibiotic	antibiotic efflux	EC149 , EC179 , EC256
35.	Escherichia coli acrA	protein homolog model	resistance-nodulation-cell division (RND) antibiotic efflux pump	fluoroquinolone antibiotic, cephalosporin, glycylcycline, penam, tetracycline antibiotic, rifamycin antibiotic, phenicol antibiotic, disinfecting agents and antiseptics	antibiotic efflux	EC149 , EC164 , EC179 , EC256
36.	vanG	protein homolog model	glycopeptide resistance gene cluster, Van ligase	glycopeptide antibiotic	antibiotic target alteration	EC149 , EC164 , EC179 , EC256
37.	leuO	protein homolog model	major facilitator superfamily (MFS) antibiotic efflux pump	nucleoside antibiotic, disinfecting agents and antiseptics	antibiotic efflux	EC149 , EC164 , EC179 , EC256
38.	mdtM	protein homolog model	major facilitator superfamily (MFS) antibiotic efflux pump	fluoroquinolone antibiotic, lincosamide antibiotic, nucleoside antibiotic, phenicol antibiotic, disinfecting agents and antiseptics	antibiotic efflux	EC149 , EC164 , EC256
39.	tet(B)	protein homolog model	major facilitator superfamily (MFS) antibiotic efflux pump	tetracycline antibiotic	antibiotic efflux	EC256
40.	APH(3')-Ia	protein homolog model	APH(3')	aminoglycoside antibiotic	antibiotic inactivation	EC256

41.	APH(6)-Id	protein homolog model	APH(6)	aminoglycoside antibiotic	antibiotic inactivation	EC256
42.	APH(3")-Ib	protein homolog model	APH(3")	aminoglycoside antibiotic	antibiotic inactivation	EC149, EC256
43.	eptA	protein homolog model	pmr phosphoethanolamine transferase	peptide antibiotic	antibiotic target alteration	EC149, EC164, EC179, EC256
44.	mdtN	protein homolog model	major facilitator superfamily (MFS) antibiotic efflux pump	nucleoside antibiotic, disinfecting agents and antiseptics	antibiotic efflux	EC149, EC164, EC179, EC256
45.	mdtO	protein homolog model	major facilitator superfamily (MFS) antibiotic efflux pump	nucleoside antibiotic, disinfecting agents and antiseptics	antibiotic efflux	EC149, EC164, EC179, EC256
46.	mdtP	protein homolog model	major facilitator superfamily (MFS) antibiotic efflux pump	nucleoside antibiotic, disinfecting agents and antiseptics	antibiotic efflux	EC149, EC164, EC179, EC256
47.	aadA	protein homolog model	ANT(3")	aminoglycoside antibiotic	antibiotic inactivation	EC256
48.	gadX	protein homolog model	resistance-nodulation-cell division (RND) antibiotic efflux pump	macrolide antibiotic, fluoroquinolone antibiotic, penam	antibiotic efflux	EC149, EC164, EC179, EC256
49.	gadW	protein homolog model	resistance-nodulation-cell division (RND) antibiotic efflux pump	macrolide antibiotic, fluoroquinolone antibiotic, penam	antibiotic efflux	EC164, EC256
50.	mdtF	protein homolog model	resistance-nodulation-cell division (RND) antibiotic efflux pump	macrolide antibiotic, fluoroquinolone antibiotic, penam	antibiotic efflux	EC149, EC164,

						EC179, EC256
51.	CRP	protein homolog model	resistance-nodulation-cell division (RND) antibiotic efflux pump	macrolide antibiotic, fluoroquinolone antibiotic, penam	antibiotic efflux	EC149, EC164, EC179, EC256
52.	AcrF	protein homolog model	resistance-nodulation-cell division (RND) antibiotic efflux pump	fluoroquinolone antibiotic, cephalosporin, cephamycin, penam	antibiotic efflux	EC149, EC164, EC179, EC256
53.	AcrE	protein homolog model	resistance-nodulation-cell division (RND) antibiotic efflux pump	fluoroquinolone antibiotic, cephalosporin, cephamycin, penam	antibiotic efflux	EC149, EC164, EC179, EC256
54.	AcrS	protein homolog model	resistance-nodulation-cell division (RND) antibiotic efflux pump	fluoroquinolone antibiotic, cephalosporin, glycylcycline, cephamycin, penam, tetracycline antibiotic, rifamycin antibiotic, phenicol antibiotic, disinfecting agents and antiseptics	antibiotic efflux	EC149, EC164, EC179, EC256
55.	bacA	protein homolog model	undecaprenyl pyrophosphate related proteins	peptide antibiotic	antibiotic target alteration	EC149, EC164, EC179, EC256
56.	rsmA	protein homolog model	resistance-nodulation-cell division (RND) antibiotic efflux pump	fluoroquinolone antibiotic, diaminopyrimidine antibiotic, phenicol antibiotic	antibiotic efflux	EC149, EC164, EC179, EC256

57.	emrB	protein homolog model	major facilitator superfamily (MFS) antibiotic efflux pump	fluoroquinolone antibiotic	antibiotic efflux	EC149, EC164, EC179, EC256
58.	emrA	protein homolog model	major facilitator superfamily (MFS) antibiotic efflux pump	fluoroquinolone antibiotic	antibiotic efflux	EC149, EC164, EC179, EC256
59.	emrR	protein homolog model	major facilitator superfamily (MFS) antibiotic efflux pump	fluoroquinolone antibiotic	antibiotic efflux	EC149, EC164, EC179, EC256
60.	acrD	protein homolog model	resistance-nodulation-cell division (RND) antibiotic efflux pump	aminoglycoside antibiotic	antibiotic efflux	EC149, EC164, EC179, EC256
61.	evgS	protein homolog model	major facilitator superfamily (MFS) antibiotic efflux pump, resistance-nodulation-cell division (RND) antibiotic efflux pump	macrolide antibiotic, fluoroquinolone antibiotic, penam, tetracycline antibiotic	antibiotic efflux	EC149, EC164, EC179, EC256
62.	evgA	protein homolog model	major facilitator superfamily (MFS) antibiotic efflux pump, resistance-nodulation-cell division (RND) antibiotic efflux pump	macrolide antibiotic, fluoroquinolone antibiotic, penam, tetracycline antibiotic	antibiotic efflux	EC149, EC164, EC179, EC256
63.	emrK	protein homolog model	major facilitator superfamily (MFS) antibiotic efflux pump	tetracycline antibiotic	antibiotic efflux	EC149, EC164, EC179, EC256
64.	ArnT	protein homolog model	pmr phosphoethanolamine transferase	peptide antibiotic	antibiotic target alteration	EC149, EC164

						, EC179 , EC256
65.	PmrF	protein homolog model	pmr phosphoethanolamine transferase	peptide antibiotic	antibiotic target alteration	EC149 , EC164 , EC179 , EC256
66.	YojI	protein homolog model	ATP-binding cassette (ABC) antibiotic efflux pump	peptide antibiotic	antibiotic efflux	EC149 , EC164 , EC179 , EC256
67.	EC-5	protein homolog model	EC beta-lactamase	cephalosporin	antibiotic inactivation	EC164
68.	EC-8	protein homolog model	EC beta-lactamase	cephalosporin	antibiotic inactivation	EC179
69.	EC-15	protein homolog model	EC beta-lactamase	cephalosporin	antibiotic inactivation	EC149
70.	EC-19	protein homolog model	EC beta-lactamase	cephalosporin	antibiotic inactivation	EC256
71.	*Haemophilus influenzae* PBP3 conferring resistance to beta-lactam antibiotics	protein variant model	Penicillin-binding protein mutations conferring resistance to beta-lactam antibiotics	cephalosporin, cephamycin, penam	antibiotic target alteration	EC149 , EC164 , EC179 , EC256
72.	*Escherichia coli* EF-Tu mutants conferring resistance to Pulvomycin	protein variant model	elfamycin resistant EF-Tu	elfamycin antibiotic	antibiotic target alteration	EC149 , EC164 , EC179 , EC256
73.	*Escherichia coli* cyaA with mutation conferring resistance to fosfomycin	protein variant model	antibiotic-resistant cya adenylate cyclase	phosphonic acid antibiotic	antibiotic target alteration	EC149 , EC179 , EC256
74.	*Escherichia coli* parC conferring resistance to fluoroquinolones	protein variant model	fluoroquinolone resistant parC	fluoroquinolone antibiotic	antibiotic target alteration	EC149 , EC256

75.	*Escherichia coli* GlpT with mutation conferring resistance to fosfomycin	protein variant model	antibiotic-resistant GlpT	phosphonic acid antibiotic	antibiotic target alteration	EC149, EC164, EC179, EC256
76.	*Escherichia coli* gyrA conferring resistance to fluoroquinolones	protein variant model	fluoroquinolone resistant gyrA	fluoroquinolone antibiotic	antibiotic target alteration	EC149, EC164, EC179, EC256
77.	*Escherichia coli* AcrAB-TolC with MarR mutations conferring resistance to ciprofloxacin and tetracycline	overexpression model	resistance-nodulation-cell division (RND) antibiotic efflux pump	fluoroquinolone antibiotic, cephalosporin, glycylcycline, penam, tetracycline antibiotic, rifamycin antibiotic, phenicol antibiotic, disinfecting agents and antiseptics	antibiotic target alteration, antibiotic efflux	EC149, EC164, EC179, EC256
78.	*Escherichia coli* AcrAB-TolC with AcrR mutation conferring resistance to ciprofloxacin, tetracycline, and ceftazidime	overexpression model	resistance-nodulation-cell division (RND) antibiotic efflux pump	fluoroquinolone antibiotic, cephalosporin, glycylcycline, penam, tetracycline antibiotic, rifamycin antibiotic, phenicol antibiotic, disinfecting agents and antiseptics	antibiotic target alteration, antibiotic efflux	EC149, EC164, EC179, EC256
79.	tetR	overexpression model	major facilitator superfamily (MFS) antibiotic efflux pump	tetracycline antibiotic	antibiotic target alteration, antibiotic efflux	EC256
80.	*Escherichia coli* soxR with mutation conferring antibiotic resistance	overexpression model	ATP-binding cassette (ABC) antibiotic efflux pump, major facilitator superfamily (MFS) antibiotic efflux pump, resistance-nodulation-cell division (RND) antibiotic efflux pump	fluoroquinolone antibiotic, cephalosporin, glycylcycline, penam, tetracycline antibiotic, rifamycin antibiotic, phenicol antibiotic,	antibiotic target alteration, antibiotic efflux	EC149, EC164, EC179, EC256

				disinfecting agents and antiseptics		
81.	*Escherichia coli* soxS with mutation conferring antibiotic resistance	overexpression model	ATP-binding cassette (ABC) antibiotic efflux pump, major facilitator superfamily (MFS) antibiotic efflux pump, resistance-nodulation-cell division (RND) antibiotic efflux pump, General Bacterial Porin with reduced permeability to beta-lactams	fluoroquinolone antibiotic, monobactam, carbapenem, cephalosporin, glycylcycline, cephamycin, penam, tetracycline antibiotic, rifamycin antibiotic, phenicol antibiotic, penem, disinfecting agents and antiseptics		EC149, EC164, EC179, EC256

Supplementary Table S2

Identification of orthologous genes and their annotated functions. Genes indicated by "+" contained a sufficient amount of phylogenetically informative sites and were therefore retained for phylogenetic reconstruction of phage vB_EcoS-149_4M.

seq_length	parsimony	singleton	constant	gene in phylogeny	evolution model	Henu8_MN055691	haarsle_MN850600
1141	91	69	981	+	VT+G4	tail_fiber_protein 26964:30542	tail_fiber_protein 22167:25592
128	18	23	87	+	HIVb+G4	lipoprotein 45954:46340	RzlA 41634:42032
679	18	33	628	-, parsimony informative sites less than 5% of all alignment		ATP-dependent_helicase 39401:41488	ATP-dependent_helicase 35612:37657
310	14	11	285	-, parsimony informative sites less than 5% of all alignment		DNA_primase 37898:38932	DNA_primase 34061:35095
161	13	40	108	+	WAG+G4	hypothetical_protein 16256:16768	putative_zonula_occludens_toxin 11459:11971
890	12	50	828	-, parsimony informative sites less than 5% of all alignment		tail_tape-measure_protein 21727:24399	hypothetical_protein 16930:19602
324	11	20	293	-, parsimony informative sites less than 5% of all alignment		major_capsid_protein 17332:18306	MCP_protein 12535:13509
353	9	16	328	-, less than 10 parsimony informative sites		exonuclease 32295:33356	putative_exodeoxyribonuclease_VIII 27285:28346
370	8	26	336	-, less than 10 parsimony informative sites		major_head_subunit_precursor 15047:16243	major_capsid_protein 10334:11446
220	8	22	190	-, less than 10 parsimony informative sites		major_tail_protein 20297:20959	major_tail_protein 15500:16162
148	8	14	126	-, less than 10 parsimony informative sites		single-stranded_DNA_binding_protein 34463:34915	putative_single-stranded_DNA_binding_protein 29453:29893
243	8	17	218	-, less than 10 parsimony informative sites		tail_assembly_protein 25575:26306	minor_tail_protein 20778:21509
92	7	13	72	-, less than 10 parsimony informative sites		hypothetical_protein 33438:33716	hypothetical_protein 28428:28706
522	6	20	496	-, less than 10 parsimony informative sites		terminase_large_subunit 11445:13013	terminase_large_subunit 6649:8217
254	6	19	229	-, less than 10 parsimony informative sites		head_morphogenesis_protein 14362:15126	head_morphogenesis_protein 9566:10330
162	6	6	150	-, less than 10 parsimony informative sites		putative_lysozyme 45463:45951	endolysin 41143:41631
218	6	3	209	-, less than 10 parsimony informative sites		recombinase 33762:34418	putative_recombination_protein 28752:29408
75	5	25	45	-, less than 10 parsimony informative sites		hypothetical_protein 339:566	hypothetical_protein 46286:46513
370	5	14	351	-, less than 10 parsimony informative sites		phosphoesterase 43989:45101	putative_phosphoesterase 39666:40781
132	3	8	121	-, less than 10 parsimony informative sites		putative_tail_protein 19893:20291	minor_tail_protein 15096:15494
123	3	6	114	-, less than 10 parsimony informative sites		hypothetical_protein 19096:19467	ribonucleoside-triphosphate_reductase 14299:14670
138	3	6	129	-, less than 10 parsimony informative sites		hypothetical_protein 18683:19099	hypothetical_protein 13886:14302
253	2	6	245	-, less than 10 parsimony informative sites		minor_tail_protein 24784:25575	putative_minor_tail_protein 20017:20778
98	2	2	94	-, less than 10 parsimony informative sites		minor_tail_protein 24456:24752	putative_minor_tail_protein 19659:19955
76	1	3	72	-, less than 10 parsimony informative sites		hypothetical_protein 9854:10084	hypothetical_protein 5059:5289
527	1	8	518	-, less than 10 parsimony informative sites		helicase 46868:48451	putative_helicase 42857:44440
105	0	6	99	-, less than 10 parsimony informative sites		tape_measure_chaperone_protein 21059:21376	TfmS 16262:16579
71	0	4	67	-, less than 10 parsimony informative sites		holin 45248:45463	holin 40928:41143
90	0	0	90	-, less than 10 parsimony informative sites		hypothetical_protein 18363:18635	hypothetical_protein 13566:13838

aaroes_MN850572	EC147_ON185589	vojen_MN850569	pSf-1_KC710998	JLBYU09_OK272487
tail_fiber_protein 34543:38121	tail_tip_host_specificity_protein_J 16497:19922	tail_fiber_protein 26044:29469	hypothetical_protein 41091:44543	putative_tail_fiber_protein 26964:30542
hypothetical_protein 2537:2935	lipoprotein 34549:34947	RzlA 45541:45939	hypothetical_protein 8880:9296	putative_unimolecular_spanin 45441:45839
helicase 48199:50244	DNA_helicase 28458:30551	DNA_helicase 39520:41565	hypothetical_protein 2859:4541	putative_ATP-dependent_helicase 39356:41443
DNA_primase 46648:47682	DNA_primase/helicase 26955:27887	DNA_primase 37969:39003	hypothetical_protein 1308:2240	putative_DNA_primase 37852:38847
putative_zonula_occludens_toxin 23835:24347	hypothetical_protein 5789:6301	putative_zonula_occludens_toxin 15335:15847	hypothetical_protein 30383:30895	putative_zonula_occludens_toxin 16256:16768
hypothetical_protein 29306:31978	tail_length_tape-measure_protein_H 11260:13932	hypothetical_protein 20807:23479	hypothetical_protein 35854:38526	putative_tail_tape-measure_protein 21727:24399
MCP_protein 24911:25885	hypothetical_protein 6865:7839	MCP_protein 16412:17386	hypothetical_protein 31458:32432	putative_major_capsid_protein 17290:18306
putative_exodeoxyribonuclease_VIII 39873:40934	exonuclease 21705:22766	putative_exodeoxyribonuclease_VIII 31219:32280	hypothetical_protein 46260:47321	putative_exodeoxyribonuclease_VIII 32290:33351
prohead_protease 22710:23822	capsid_and_scaffold 4664:5776	capsid_protein 14210:15322	hypothetical_protein 29258:30370	putative_major_head_subunit_precursor 15131:16243
major_tail_protein 27876:28538	tail_fiber 9830:10492	putative_major_tail_protein 19377:20039	hypothetical_protein 34424:35086	putative_major_tail_protein 20297:20959
single-stranded_DNA_binding_protein 42041:42487	hypothetical_protein 23873:24325	putative_ssDNA-binding_protein 33386:33838	hypothetical_protein 48428:48880	putative_single_stranded_DNA-binding_protein 34458:34910
putative_minor_tail_protein 33154:33885	tail_tip_assembly_protein_K 15108:15839	putative_minor_tail_protein 24655:25386	hypothetical_protein 39702:40433	putative_minor_tail_protein 25575:26306
hypothetical_protein 41016:41282	hypothetical_protein 22848:23135	hypothetical_protein 32362:32640	hypothetical_protein 47403:47681	putative_tail_length_tape_measure_protein 33433:33711
terminase_large_subunit 19025:20593	terminase_large_subunit 979:2547	terminase_large_subunit 10525:12093	hypothetical_protein 25574:27142	putative_terminase_large_subunit 11446:13014
head_morphogenesis_protein 21942:22706	minor_capsid_protein 3896:4660	putative_minor_capsid_protein 13442:14206	hypothetical_protein 28490:29254	putative_head_morphogenesis_protein 14363:15127
endolysin 2046:2534	lysin 34058:34546	endolysin 45050:45538	hypothetical_protein 8389:8877	putative_endolysin 44950:45438
putative_recombination_protein 41340:41996	recombinase 23172:23828	putative_recombination_protein 32686:33342	hypothetical_protein 47727:48383	putative_recombinase 33757:34413
hypothetical_protein 7175:7402	hypothetical_protein 39202:39429	hypothetical_protein 49890:50117	hypothetical_protein 14046:14273	putative_polyribonucleotide_nucleotidyltransferase 1104:1331
putative_phosphoesterase 569:1699	hypothetical_protein 32581:33696	putative_phosphoesterase 43573:44688	hypothetical_protein 6912:8027	putative_phosphoesterase 43473:44588
minor_tail_protein 27472:27870	hypothetical_protein 9426:9824	minor_tail_protein 18973:19371	hypothetical_protein 34020:34418	putative_minor_tail_protein 19893:20291
ribonucleoside-triphosphate_reductase 26675:27046	hypothetical_protein 8629:9000	ribonucleoside-triphosphate_reductase 18176:18547	hypothetical_protein 33223:33594	putative_ribonucleoside-triphosphate_reductase 19096:19467
hypothetical_protein 26262:26678	hypothetical_protein 8216:8632	hypothetical_protein 17763:18179	hypothetical_protein 32810:33226	putative_head-tail_adaptor 18683:19099
putative_minor_tail_protein 32393:33154	tail_tip_assembly_protein_L 14347:15108	putative_minor_tail_protein 23894:24655	hypothetical_protein 38941:39702	putative_minor_tail_protein 24784:25575
putative_minor_tail_protein 32035:32331	minor_tail_protein 13935:14285	putative_minor_tail_protein 23536:23832	hypothetical_protein 38529:38879	putative_minor_tail_protein 24402:24752
hypothetical_protein 17442:17672	hypothetical_protein 45190:45420	hypothetical_protein 8940:9170	hypothetical_protein 23985:24215	hypothetical_protein 9838:10068
putative_helicase 3758:5341	hypothetical_protein 35773:37356	putative_helicase 46467:48050	hypothetical_protein 10620:12203	putative_helicase 46663:48246
TfmS 28638:28955	hypothetical_protein 10592:10909	tape_measure_chaperone 20139:20456	hypothetical_protein 35186:35503	putative_tape_measure_chaperone_protein 21059:21376
holin 1831:2046	hypothetical_protein 33843:34058	holin 44835:45050	hypothetical_protein 8174:8389	putative_holin 44735:44950
hypothetical_protein 25942:26214	hypothetical_protein 7896:8168	hypothetical_protein 17443:17715	hypothetical_protein 32490:32762	hypothetical_protein 18363:18635

JakobBernoulli_MZ501079	KarlJaspers_MZ501090	vB-EcoS_149-4M_04	vB_EcoD_Fulano1_OL539459	aalborv_MN850591
J-like_tail_tip_protein 16040:19465	J-like_tail_tip_protein 16041:19619	tail_protein 24705:28283	putative_tail_fiber_protein 16090:19668	tail_fiber_protein 10741:14166
u-spanin 35785:36183	u-spanin 36536:36934	Rz-like_spanin 45168:45566	putative_lipoprotein 35591:35989	RzlA 43942:44340
DNA_helicase 29769:31808	DNA_helicase 30499:32538	hypothetical_protein 39100:41193	putative_helicase 29548:31593	putative_ATP-dependent_helicase 1657:3702
DNA_primase 28218:29150	DNA_primase 28948:29880	DNA_primase 37596:38528	putative_DNA_primase 27997:29031	DNA_primase 4219:5253
capsid_decoration_protein 5333:5845	capsid_decoration_protein 5332:5844	hypothetical_protein 13627:14118	putative_zonula_occludens_toxin 5381:5893	putative_zonula_occludens_toxin 24363:24875
tail_tape_measure_protein 10803:13475	tail_tape_measure_protein 10804:13476	tail_length_tape_measure_protein 19467:22139	putative_tail_tape_measure_protein 10853:13525	hypothetical_protein 16731:19403
major_capsid_protein 6409:7383	major_capsid_protein 6408:7382	major_head_protein 15072:16046	putative_major_capsid_protein 6457:7431	MCP_protein 22825:23799
putative_exodeoxyribonuclease_VIII 21254:22315	putative_exodeoxyribonuclease_VIII 21316:22377	exonuclease_VIII 29991:31052	putative_exodeoxyribonuclease_VIII 21418:22479	putative_exodeoxyribonuclease_VIII 7930:8991
prohead_protease 4208:5320	prohead_protease 4207:5319	head_maturation_protease 12502:13614	hypothetical_protein 4256:5368	major_capsid_protein 24888:26000
major_tail_protein 9373:10035	major_tail_protein 9374:10036	major_tail_protein 18037:18699	putative_major_tail_protein 9423:10085	major_tail_protein 20171:20833
single-stranded_DNA_binding_protein 23422:23874	single-stranded_DNA_binding_protein 23487:23939	single_strand_DNA_binding_protein 32161:32607	putative_helix-destabilizing_protein 23586:24047	single-stranded_DNA_binding_protein 6371:6823
putative_minor_tail_protein 14651:15382	putative_minor_tail_protein 14652:15383	minor_tail_protein 23316:24047	putative_tail_assembly_protein 14701:15432	minor_tail_protein 14824:15555
hypothetical_protein 22397:22675	hypothetical_protein 22459:22740	tail_length_tape_measure_protein 31128:31415	hypothetical_protein 22561:22839	hypothetical_protein 7570:7848
terminase_large_subunit 522:2090	terminase_large_subunit 522:2090	terminase_large_subunit 8816:10384	putative_terminase_large_subunit 571:2139	terminase_large_subunit 28117:29685
minor_capsid_protein 3440:4204	minor_capsid_protein 3439:4203	head_morphogenesis 11734:12498	putative_head_morphogenesis_protein 3488:4252	putative_minor_capsid_protein 26004:26768
endolysin 35294:35782	endolysin 36045:36533	endolysin 44677:45165	putative_lysozyme 35100:35588	endolysin 44343:44831
putative_recombination_protein 22721:23377	putative_recombination_protein 22786:23442	Erf-like_ssDNA_annealing_protein 31461:32117	putative_recombination_protein 22885:23541	putative_recombination_protein 6868:7524
hypothetical_protein 40140:40367	hypothetical_protein 41189:41416	hypothetical_protein 49822:50049	putative_polyribonucleotide 41180:41470	hypothetical_protein 39461:39688
hypothetical_protein 33817:34932	hypothetical_protein 34568:35683	DNA_repair_exonuclease 43200:44315	putative_phosphoesterase 33623:34753	putative_phosphoesterase 45193:46308
minor_tail_protein 8969:9367	minor_tail_protein 8970:9368	tail_terminator 17633:18031	hypothetical_protein 9019:9417	minor_tail_protein 20839:21237
hypothetical_protein 8172:8543	hypothetical_protein 8173:8544	head_closure_Hc1 16836:17207	hypothetical_protein 8222:8593	ribonucleoside-triphosphate_reductase 21663:22034
head-tail_adaptor_protein 7759:8175	head-tail_adaptor_protein 7760:8176	virion_structural_protein 16423:16839	hypothetical_protein 7809:8225	hypothetical_protein 22031:22447
minor_tail_protein 13890:14651	minor_tail_protein 13891:14652	minor_tail_protein 22555:23316	putative_minor_tail_protein 13940:14701	putative_minor_tail_protein 15555:16316
minor_tail_protein 13532:13828	minor_tail_protein 13533:13829	minor_tail_protein 22197:22493	putative_tail_protein 13528:13878	putative_minor_tail_protein 16378:16674
hypothetical_protein 50259:50489	hypothetical_protein 50404:50634	hypothetical_protein 7225:7455	hypothetical_protein 50746:50976	hypothetical_protein 31035:31265
putative_helicase 36711:38294	putative_helicase 37759:39342	DNA_helicase 46393:47976	putative_helicase 37507:39090	putative_helicase 41535:43118
tape_measure_chaperone 10135:10452	tape_measure_chaperone 10136:10453	tail_length_tape_measure_protein 18799:19116	hypothetical_protein 10185:10502	TfmS 19754:20071
holin 35079:35294	holin 35830:36045	holin 44462:44677	hypothetical_protein 34885:35100	holin 44831:45046
hypothetical_protein 7439:7711	hypothetical_protein 7440:7712	hypothetical_protein 16103:16375	hypothetical_protein 7489:7761	hypothetical_protein 22495:22767

Identification of orthologous genes and their annotated functions. Genes indicated by "+" contained a sufficient amount of phylogenetically informative sites and were therefore retained for phylogenetic reconstruction of phage vB_EcoS-149_6D.

seq_length	parsimony	singleton	constant	gene in phylogeny	evolution model	BEK12A_MN158210	Tuinin_OR896828
235	127	4	104	+	PMB	hypothetical_protein 39324:40010	DNA_methyltransferase 37913:38686
1141	97	99	945	+	Q.pfam+G4	tail_fiber_protein 17349:20927	tail_protein 15893:19471
892	76	389	427	+	VT+F+G4	tail_length_tape-measure_protein_1 11754:14678	tail_length_tape_measure_protein 10648:13329
126	58	2	66	+	Blosum62+I	hypothetical_protein 49370:49750	hypothetical_protein 47513:47914
90	45	4	41	+	JTT	hypothetical_protein 44677:44949	hypothetical_protein 42720:42992
162	23	68	71	+	WAG+I	hypothetical_protein 6274:6786	hypothetical_protein 4823:5332
59	22	1	36	+	LG+I	hypothetical_protein 32989:33168	hypothetical_protein 31895:32074
219	17	45	157	+	WAG+G4	tail_fibers_protein 10318:10980	major_tail_protein 9212:9874
323	10	32	281	-, parsimony informative sites less than 5% of all alignment		hypothetical_protein 7350:8324	major_head_protein 6253:7224
426	9	68	349	-, less than 10 parsimony informative sites		portal_protein 3101:4381	hypothetical_protein 1640:2932
370	7	60	303	-, less than 10 parsimony informative sites		capsid_and_scaffold_protein 5149:6261	head_maturation_protease 3695:4810
95	7	41	47	-, less than 10 parsimony informative sites		hypothetical_protein 50300:50590	hypothetical_protein 48464:48754
87	6	8	73	-, less than 10 parsimony informative sites		hypothetical_protein 47869:48132	hypothetical_protein 46195:46458
69	6	10	53	-, less than 10 parsimony informative sites		hypothetical_protein 38332:38553	hypothetical_protein 37239:37460
196	5	16	175	-, less than 10 parsimony informative sites		hypothetical_protein 48709:49299	deoxynucleoside_monophosphate_kinase 46852:47442
131	5	10	116	-, less than 10 parsimony informative sites		hypothetical_protein 35459:35854	Rz-like_spanin 34365:34760
132	5	32	95	-, less than 10 parsimony informative sites		hypothetical_protein 9914:10315	tail_terminator 8808:9209
104	4	44	56	-, less than 10 parsimony informative sites		hypothetical_protein 11080:11394	tail_length_tape_measure_protein 9974:10291
371	4	3	364	-, less than 10 parsimony informative sites		hypothetical_protein 33495:34610	DNA_repair_exonuclease 32401:33516
148	4	60	84	-, less than 10 parsimony informative sites		single-stranded_DNA-binding_protein 24448:24900	single_strand_DNA_binding_protein 23253:23705
198	4	18	176	-, less than 10 parsimony informative sites		tail_assembly_protein 16567:17163	tail_protein 15218:15814
139	3	1	135	-, less than 10 parsimony informative sites		hypothetical_protein 30986:31405	hypothetical_protein 30332:30751
88	3	34	51	-, less than 10 parsimony informative sites		hypothetical_protein 8381:8653	hypothetical_protein 7234:7548
98	3	9	86	-, less than 10 parsimony informative sites		minor_tail_protein 14681:15031	minor_tail_protein 13332:13682
665	3	6	656	-, less than 10 parsimony informative sites		DNA_helicase 28989:30986	hypothetical_protein 28335:30332
147	3	25	119	-, less than 10 parsimony informative sites		hypothetical_protein 9481:9924	tail_completion_or_Neck1_protein 8375:8818
163	3	4	156	-, less than 10 parsimony informative sites		lysin 34971:35462	endolysin 33877:34368
54	3	4	47	-, less than 10 parsimony informative sites		hypothetical_protein 48150:48314	hypothetical_protein 46455:46640
251	3	48	200	-, less than 10 parsimony informative sites		hypothetical_protein 4381:5145	head_morphogenesis 2929:3684
351	3	76	272	-, less than 10 parsimony informative sites		exodeoxyribonuclease_VIII 22620:23675	exonuclease_VIII 21164:22219
85	2	2	81	-, less than 10 parsimony informative sites		hypothetical_protein 32656:32913	hypothetical_protein 31562:31819
74	2	4	68	-, less than 10 parsimony informative sites		hypothetical_protein 40536:40760	hypothetical_protein 38634:38858
253	2	45	206	-, less than 10 parsimony informative sites		minor_tail_protein 15094:15855	minor_tail_protein 13745:14506
522	2	40	480	-, less than 10 parsimony informative sites		terminase_large_subunit 985:2553	terminase_large_subunit 1:1569
123	2	4	117	-, less than 10 parsimony informative sites		hypothetical_protein 9117:9491	head_closure_Hc1 8011:8385
90	1	8	81	-, less than 10 parsimony informative sites		hypothetical_protein 43269:43541	hypothetical_protein 41293:41565
217	1	41	175	-, less than 10 parsimony informative sites		recombinase 23751:24404	Erf-like_ssDNA_annealing_protein 22556:23209
104	1	3	100	-, less than 10 parsimony informative sites		hypothetical_protein 51135:51449	hypothetical_protein 49269:49583
243	1	35	207	-, less than 10 parsimony informative sites		tail_assembly_protein 15855:16586	minor_tail_protein 14506:15237
526	1	25	500	-, less than 10 parsimony informative sites		hypothetical_protein 36683:38263	DNA_helicase 35590:37170
75	1	28	46	-, less than 10 parsimony informative sites		hypothetical_protein 40916:41143	hypothetical_protein 39013:39240
80	1	9	70	-, less than 10 parsimony informative sites		hypothetical_protein 33158:33400	hypothetical_protein 32064:32306
139	1	3	135	-, less than 10 parsimony informative sites		hypothetical_protein 8701:9120	virion_structural_protein 7532:8014
241	0	12	229	-, less than 10 parsimony informative sites		DNA_adenine_methyltransferase 31926:32651	DNA_methyltransferase 30805:31557
129	0	11	118	-, less than 10 parsimony informative sites		hypothetical_protein 36291:36680	hypothetical_protein 35198:35593
158	0	0	158	-, less than 10 parsimony informative sites		hypothetical_protein 28425:28901	transcriptional_regulator 27771:28247
71	0	1	70	-, less than 10 parsimony informative sites		holin 34756:34971	holin 33662:33877
50	0	0	50	-, less than 10 parsimony informative sites		hypothetical_protein 52232:52417	hypothetical_protein 50397:50549
76	0	0	76	-, less than 10 parsimony informative sites		hypothetical_protein 51649:51879	hypothetical_protein 49782:50012
81	0	40	41	-, less than 10 parsimony informative sites		hypothetical_protein 49850:50098	hypothetical_protein 48014:48262
70	0	1	69	-, less than 10 parsimony informative sites		hypothetical_protein 50666:50878	hypothetical_protein 48830:49042

vB_EcoD_Poky_OL539445	MikeNSara_OR896827	vB-EcoS_149-6D_03	PeterMerian_MZ501101
putative_DNA_cytosine_methyltransferase 37863:38573	DNA_methyltransferase 36699:37472	DNA_methyltransferase 47183:47923	Dcm-like_C5_cytosine_methyltransferase 38218:38958
putative_tail_fiber_protein 16482:20060	tail_protein 15898:19323	tail_protein 25361:28939	J-like_tail_tip_protein 16428:20006
putative_tail_tape_measure_protein 11237:13918	tail_length_tape_measure_protein 10653:13334	tail_length_tape_measure_protein 19901:22582	tail_tape_measure_protein 11183:13864
hypothetical_protein 48042:48467	hypothetical_protein 46887:47267	hypothetical_protein 5390:5791	hypothetical_protein 44404:44784
hypothetical_protein 43281:43541	hypothetical_protein 41760:42032	hypothetical_protein 939:1211	hypothetical_protein 42772:43032
putative_scaffolding_protein 5408:5917	hypothetical_protein 4823:5332	hypothetical_protein 14076:14585	capsid_decoration_protein 5350:5859
hypothetical_protein 31846:32025	hypothetical_protein 30682:30861	hypothetical_protein 41159:41344	hypothetical_protein 32195:32374
putative_major_tail_protein 9801:10463	major_tail_protein 9217:9879	major_tail_protein 18465:19127	major_tail_protein 9747:10409
putative_major_capsid_protein 6841:7812	major_head_protein 6256:7227	major_head_protein 15506:16477	major_capsid_protein 6784:7758
putative_portal_protein 2225:3517	hypothetical_protein 1640:2932	hypothetical_protein 10893:12185	portal_protein 2167:3459
putative_zonula_occludens_toxin 4280:5395	head_maturation_protease 3695:4810	head_maturation_protease 12948:14063	prohead_protease 4222:5337
hypothetical_protein 49017:49307	hypothetical_protein 47817:48107	hypothetical_protein 6341:6631	hypothetical_protein 45363:45650
hypothetical_protein 46769:47032	hypothetical_protein 45377:45640	hypothetical_protein 4072:4335	hypothetical_protein 43104:43367
hypothetical_protein 37189:37410	hypothetical_protein 36025:36246	hypothetical_protein 46509:46730	hypothetical_protein 37553:37762
hypothetical_protein 47426:48016	deoxynucleoside_monophosphate_kinase 46226:46816	deoxynucleoside_monophosphate_kinase 4729:5319	putative_P-loop_NTPase_protein 43743:44333
hypothetical_protein 34298:34711	Rz-like_spanin 33152:33547	Rz-like_spanin 43635:44030	u-spanin 34666:35061
putative_structural_protein 9397:9798	tail_terminator 8813:9214	tail_terminator 18061:18462	minor_tail_protein 9343:9744
putative_tape_measure_chaperone_protein 10563:10880	tail_length_tape_measure_protein 9979:10296	tail_length_tape_measure_protein 19227:19544	tape_measure_chaperone 10509:10826
putative_phosphoesterase 32352:33467	DNA_repair_exonuclease 31188:32303	DNA_repair_exonuclease 41671:42786	hypothetical_protein 32701:33816
putative_DNA_primase 23843:24295	single_strand_DNA_binding_protein 22847:23299	single_strand_DNA_binding_protein 32460:32912	single-stranded_DNA_binding_protein 23546:23992
putative_tail_assembly_protein 15807:16403	tail_protein 15223:15819	tail_protein 24472:25068	tail_assembly_protein 15753:16349
putative_VRR-NUC_domain-containing_protein 30283:30702	hypothetical_protein 29120:29539	hypothetical_protein 39602:40021	hypothetical_protein 30631:31050
hypothetical_protein 7871:8137	hypothetical_protein 7287:7553	hypothetical_protein 16487:16801	hypothetical_protein 7817:8083
putative_minor_tail_protein 13915:14271	minor_tail_protein 13337:13687	minor_tail_protein 22585:22935	minor_tail_protein 13867:14217
putative_ATP-dependent_helicase 28286:30283	hypothetical_protein 27123:29120	hypothetical_protein 37605:39602	DNA_helicase 28634:30631
hypothetical_protein 8964:9407	tail_completion_or_Neck1_protein 8380:8823	tail_completion_or_Neck1_protein 17628:18071	hypothetical_protein 8910:9353
putative_endolysin 33828:34319	endolysin 32664:33155	endolysin 43147:43638	endolysin 34178:34669
hypothetical_protein 47029:47214	hypothetical_protein 45637:45822	hypothetical_protein 4332:4517	hypothetical_protein 43364:43549
putative_head_morphogenesis_protein 3514:4269	head_morphogenesis 2929:3684	head_morphogenesis 12182:12937	minor_capsid_protein 3456:4211
putative_exodeoxyribonuclease 21754:22809	exonuclease_VIII 21018:22073	exonuclease_VIII 30632:31687	putative_exodeoxyribonuclease_VIII 21713:22774
hypothetical_protein 31513:31770	hypothetical_protein 30349:30606	hypothetical_protein 40832:41089	hypothetical_protein 31861:32118
hypothetical_protein 39075:39299	hypothetical_protein 37420:37644	hypothetical_protein 48395:48619	hypothetical_protein 39430:39654
putative_minor_tail_protein 14334:15095	minor_tail_protein 13750:14511	minor_tail_protein 22999:23760	minor_tail_protein 14280:15041
putative_large_terminase_subunit 586:2154	terminase_large_subunit 1:1569	terminase_large_subunit 9254:10822	terminase_large_subunit 528:2096
putative_ribonucleoside-triphosphate_reductase 8600:8971	head_closure_Hc1 8016:8387	head_closure_Hc1 17264:17638	hypothetical_protein 8546:8917
hypothetical_protein 41852:42124	hypothetical_protein 40352:40624	hypothetical_protein 50548:50820	hypothetical_protein 41363:41635
putative_recombination_protein 23146:23799	Erf-like_ssDNA_annealing_protein 22150:22803	Erf-like_ssDNA_annealing_protein 31763:32416	putative_recombination_protein 22851:23504
hypothetical_protein 49852:50166	hypothetical_protein 48571:48966	hypothetical_protein 7146:7460	hypothetical_protein 46195:46509
putative_tail_assembly_protein 15095:15826	minor_tail_protein 14511:15242	minor_tail_protein 23760:24491	putative_minor_tail_protein 15041:15772
putative_helicase 35540:37120	DNA_helicase 34376:35956	DNA_helicase 44860:46440	putative_helicase 35890:37473
hypothetical_protein 39455:39682	hypothetical_protein 37800:38027	hypothetical_protein 48774:49001	hypothetical_protein 39810:40037
hypothetical_protein 32015:32257	hypothetical_protein 30851:31093	hypothetical_protein 41334:41576	hypothetical_protein 32364:32606
putative_head-tail_adaptor 8184:8603	virion_structural_protein 7600:8019	virion_structural_protein 16785:17267	head-tail_adaptor_protein 8130:8549
putative_DNA_adenine_methyltransferase 30756:31508	DNA_methyltransferase 29619:30344	DNA_methyltransferase 40075:40827	Dam-like_N6-adenine_methyltransferase 31131:31856
hypothetical_protein 35148:35537	hypothetical_protein 33984:34379	hypothetical_protein 44468:44863	hypothetical_protein 35498:35893
putative_transcriptional_regulator 27722:28198	transcriptional_regulator 26559:27035	transcriptional_regulator 37041:37517	putative_transcriptional_regulator 28070:28546
putative_holin 33613:33828	holin 32449:32664	holin 42932:43147	holin 33963:34178
hypothetical_protein 50948:51133	hypothetical_protein 49781:49933	hypothetical_protein 8275:8427	hypothetical_protein 47291:47476
hypothetical_protein 50344:50595	hypothetical_protein 49165:49395	hypothetical_protein 7659:7889	hypothetical_protein 46708:46938
hypothetical_protein 48567:48815	hypothetical_protein 47367:47615	hypothetical_protein 5891:6139	hypothetical_protein 44884:45147
hypothetical_protein 49383:49595	hypothetical_protein 48183:48395	hypothetical_protein 6707:6919	hypothetical_protein 45726:45938

EC167_ON185587	tonnikala_MN850613
hypothetical_protein 38946:39632	putative_DNA_methylase 10944:11717
tail_tip_host_specificity_protein_J 16609:20187	putative_tail_fiber 40615:44040
tail_length_tape-measure_protein_H 11372:14044	tail_tape-measure_protein 35371:38052
hypothetical_protein 47644:48024	hypothetical_protein 20849:21229
hypothetical_protein 42916:43194	hypothetical_protein 16093:16365
hypothetical_protein 5350:5829	HtjA 29542:30051
hypothetical_protein 32129:32308	hypothetical_protein 4720:4899
tail_fiber 9942:10604	putative_major_tail_protein 33935:34597
hypothetical_protein 6984:7955	major_capsid_protein 30976:31947
portal_(connector)_protein 2167:3459	putative_portal_protein 26359:27651
capsid_and_scaffold 4222:5337	capsid_protein 28414:29529
hypothetical_protein 48574:48864	hypothetical_protein 21779:22069
hypothetical_protein 46143:46406	hypothetical_protein 19348:19611
hypothetical_protein 37954:38175	hypothetical_protein 10270:10491
dNMP_kinase 46983:47573	hypothetical_protein 20188:20778
lipoprotein 35081:35476	hypothetical_protein 7190:7585
hypothetical_protein 9538:9936	putative_structural_protein 33531:33932
hypothetical_protein 10704:11021	tape_measure_chaperone 34697:35014
hypothetical_protein 33036:34232	putative_phosphoesterase 5145:6341
single-stranded_DNA-binding_protein 23711:24163	single-stranded_DNA_binding_protein 47825:48271
tail_tip_assembly_protein_I 15932:16528	putative_tail_assembly_protein 39940:40536
hypothetical_protein 30567:30986	VRR-NUC_domain-containing_protein 3157:3576
hypothetical_protein 8012:8278	hypothetical_protein 32005:32271
hypothetical_protein 14101:14397	minor_tail_protein 38055:38405
DNA_helicase 28570:30567	ATP-dependent_helicase 1160:3157
hypothetical_protein 9105:9548	hypothetical_protein 33098:33541
lysozyme_R 34593:35084	endolysin 6702:7193
hypothetical_protein 46403:46588	hypothetical_protein 19608:19793
hypothetical_protein 3456:4211	head_morphogenesis_protein 27648:28403
exonuclease 21882:22937	exodeoxyribonuclease_VIII 45736:46791
hypothetical_protein 31796:32053	hypothetical_protein 4387:4644
hypothetical_protein 40158:40382	hypothetical_protein 11665:11889
tail_tip_assembly_protein_L 14459:15220	putative_minor_tail_protein 38467:39228
terminase_large_subunit 528:2096	terminase_large_subunit 24720:26288
hypothetical_protein 8741:9112	ribonucleoside-triphosphate_reductase 32734:33105
hypothetical_protein 41515:41787	hypothetical_protein 14664:14936
recombinase 23014:23667	putative_recombination_protein 47128:47781
hypothetical_protein 49408:49722	hypothetical_protein 22613:22927
tail_tip_assembly_protein_K 15220:15951	minor_tail_protein 39228:39959
hypothetical_protein 36305:37885	putative_helicase 8621:10201
hypothetical_protein 40539:40841	hypothetical_protein 12045:12272
hypothetical_protein 32298:32540	hypothetical_protein 4889:5131
hypothetical_protein 8325:8744	hypothetical_protein 32318:32737
DNA_adenine_methyltransferase 31066:31791	putative_DNA_N-6-adenine-methyltransferase 3657:4382
hypothetical_protein 35913:36308	YdbL 8229:8618
hypothetical_protein 28006:28482	putative_transcriptional_regulator 596:1072
holin 34378:34593	putative_holin 6487:6702
hypothetical_protein 50504:50689	hypothetical_protein 23708:23893
hypothetical_protein 49921:50151	hypothetical_protein 23126:23356
hypothetical_protein 48124:48372	hypothetical_protein 21329:21577
hypothetical_protein 48940:49161	hypothetical_protein 22144:22356

Identification of orthologous genes and their annotated functions. Genes indicated by "+" contained a sufficient amount of phylogenetically informative sites and were therefore retained for phylogenetic reconstruction of phage vB_EcoS-164_2.

seq_length	parsimony	singleton	constant	gene in phylogeny	evolution model	vB_EcoS_PJ16_OQ955832
114	38	4	72	+	Blosum62	hypothetical_protein 12497:12841
108	26	14	68	+	FLU	hypothetical_protein 13149:13478
62	25	3	34	+	LG	NinH_family_protein 14917:15111
107	25	1	81	+	Q.insect	holin 36064:36387
840	25	156	659	-, parsimony informative sites less than 5% of all alignment		tail_tip_protein 31473:34001
167	23	7	137	+	JTT+I	deoxyuridine_5'-_triphosphate_nucleotidohydrolase 13475:13978
319	23	13	283	+	LG	head_morphogenesis_protein 15284:16243
110	21	5	84	+	Q.yeast	hypothetical_protein 12149:12481
587	21	6	560	-, parsimony informative sites less than 5% of all alignment		terminase_large_subunit 4742:6505
451	19	12	420	-, parsimony informative sites less than 5% of all alignment		portal_protein 10761:12116
132	18	7	107	+	Q.bird+I	i-spanin_component 37001:37399
82	17	7	58	+	JTT	hypothetical_protein 12904:13152
253	16	59	178	+	JTT	helicase 43749:44501
115	14	0	101	+	LG+I	2'deoxynucleoside_5'-_phosphate_N-hydrolase 38642:38995
211	14	22	175	+	WAG+I	terminase_small_subunit 324:959
419	11	33	375	-, parsimony informative sites less than 5% of all alignment		hypothetical_protein 18711:19970
88	10	9	69	+	Q.insect+I	hypothetical_protein 16896:17171
618	9	256	353	-, less than 10 parsimony informative sites		DNA_polymerase 39788:41644
485	8	59	418	-, less than 10 parsimony informative sites		major_tail_tube_protein 24250:25707
122	8	4	110	-, less than 10 parsimony informative sites		hypothetical_protein 18301:18711
60	8	16	36	-, less than 10 parsimony informative sites		hypothetical_protein 14735:14917
127	7	3	117	-, less than 10 parsimony informative sites		L_alanyl_D_glutamate_peptidase_endolysin 36384:36773
163	7	28	128	-, less than 10 parsimony informative sites		nuclease 47397:47888
339	7	19	313	-, less than 10 parsimony informative sites		major_capsid_protein 20384:21403
142	6	15	121	-, less than 10 parsimony informative sites		type_1_neck_protein 23360:23791
65	5	8	52	-, less than 10 parsimony informative sites		hypothetical_protein 50823:51020
136	5	16	115	-, less than 10 parsimony informative sites		capsid_fiber_protein 19970:20380
338	5	68	265	-, less than 10 parsimony informative sites		hypothetical_protein 41725:42588
124	3	13	108	-, less than 10 parsimony informative sites		hypothetical_protein 22984:23358
381	3	78	300	-, less than 10 parsimony informative sites		hypothetical_protein 42604:43749
76	3	0	73	-, less than 10 parsimony informative sites		hypothetical_protein 36784:37014
784	3	64	717	-, less than 10 parsimony informative sites		DNA_helicase 48164:50518
140	3	17	120	-, less than 10 parsimony informative sites		tail_assembly_chaperone 25813:26235
70	2	16	52	-, less than 10 parsimony informative sites		hypothetical_protein 50609:50821
117	2	24	91	-, less than 10 parsimony informative sites		hypothetical_protein 21467:21820
88	2	15	71	-, less than 10 parsimony informative sites		hypothetical_protein 8748:9023
178	1	32	145	-, less than 10 parsimony informative sites		hypothetical_protein 39212:39748
178	1	29	148	-, less than 10 parsimony informative sites		hypothetical_protein 22574:22984
167	1	37	129	-, less than 10 parsimony informative sites		hypothetical_protein 30118:30621
70	1	28	41	-, less than 10 parsimony informative sites		hypothetical_protein 51013:51228
253	1	25	227	-, less than 10 parsimony informative sites		hypothetical_protein 52369:53136
97	1	17	79	-, less than 10 parsimony informative sites		N-acetyltransferase_tail_protein 31121:31480
140	1	15	124	-, less than 10 parsimony informative sites		structural_protein 23791:24213
168	0	26	142	-, less than 10 parsimony informative sites		enolase_like_protein 30618:31124

9NA_NC_025443	KarinMuster_Bas82_PQ850612	vB_EcoS_AVIO78A_MZ726797
hypothetical_protein 10840:11184	hypothetical_protein 7272:7616	hypothetical_protein 12911:13255
hypothetical_protein 11437:11763	hypothetical_protein 8406:8732	hypothetical_protein 14027:14356
NinH 13365:13568	hypothetical_protein 10415:10609	NinH 15819:16013
holin 34197:34523	holin 32645:32968	holin_or_antiholin 36526:36849
tail_protein 29533:32055	tail_protein 27592:30123	tail_tip_protein 31935:34463
putative_deoxyuridine_5'-triphosphate_nucleotidohydrolase 11763:12266	dUTPase 8732:9235	deoxyuridine_5'-triphosphate_nucleotidohydrolase 14353:14856
head_morphogenesis 13743:14702	head_morphogenesis_protein 10783:11742	head_morphogenesis_protein 16187:17146
hypothetical_protein 10480:10824	hypothetical_protein 6924:7256	hypothetical_protein 12563:12895
TerL_protein 4070:5833	terminase_large_subunit 1:1764	large_terminase_subunit 5131:6894
portal_protein 8779:10134	portal_protein 5535:6890	portal_protein 11174:12529
Rz-like_spanin 35137:35535	Rz-like_spanin 33582:33980	i-spanin_component 37463:37861
hypothetical_protein 11186:11440	hypothetical_protein 7679:7927	hypothetical_protein 13318:13575
hypothetical_protein 42326:43120	hypothetical_protein 40143:40913	helicase 44048:44800
nucleoside_2-deoxyribosyltransferase 36988:37338	hypothetical_protein 34776:35129	nucleoside_2-deoxyribosyltransferase 38657:39010
hypothetical_protein 296:931	terminase_small_subunit 53526:54278	terminase_small_subunit 276:911
hypothetical_protein 16609:17868	hypothetical_protein 14554:15813	hypothetical_protein 19798:21063
hypothetical_protein 15141:15407	hypothetical_protein 12137:12412	hypothetical_protein 17379:17654
DNA_polymerase 38237:40138	DNA_polymerase 36014:37870	DNA_polymerase 39919:41775
major_tail_protein 22692:24149	hypothetical_protein 19819:21276	major_tail_tube_protein 25066:26523
DNA_methyltransferase 16199:16609	DNA_methyltransferase 14144:14554	hypothetical_protein 19388:19798
hypothetical_protein 13177:13368	hypothetical_protein 10233:10415	hypothetical_protein 15637:15819
endolysin 34520:34909	endolysin 32965:33354	endolysin 36846:37235
hypothetical_protein 45585:46076	hypothetical_protein 43290:43781	hypothetical_protein 47177:47668
major_head_protein 18282:19301	major_head_protein 16227:17246	major_capsid_protein 21477:22496
tail_completion_or_Neck1_protein 20872:21300	tail_completion_or_Neck1_protein 18929:19360	hypothetical_protein 24176:24607
lipoprotein 49028:49225	lipoprotein 47082:47279	hypothetical_protein 50970:51167
major_virion_structural_protein 17868:18278	head_scaffolding_protein 15813:16223	major_virion_structural_protein 21063:21473
Gp2.5-like_ssDNA_binding_protein_and_ssDNA_annealing_protein 40218:41168	ssDNA_binding_and_annealing_protein 37951:38982	hypothetical_protein 41856:42887
hypothetical_protein 20498:20872	hypothetical_protein 18553:18927	hypothetical_protein 23800:24174
exonuclease 41181:42326	exonuclease 38998:40143	hypothetical_protein 42903:44048
membrane_protein 34920:35150	membrane_protein 33365:33595	rIIA_lysis_inhibitor 37246:37476
DNA_primase 46250:48604	DNA_primase 43955:46309	hypothetical_protein 47842:50196
tail_assembly_chaperone 24508:24933	tail_assembly_chaperone 21725:22147	tail_assembly_chaperone 26875:27297
hypothetical_protein 48695:48907	hypothetical_protein 46400:46612	hypothetical_protein 50287:50499
hypothetical_protein 19365:19748	hypothetical_protein 17309:17662	hypothetical_protein 22559:22912
hypothetical_protein 6973:7248	hypothetical_protein 4008:4274	hypothetical_protein 9051:9359
hypothetical_protein 37660:38196	hypothetical_protein 35438:35974	hypothetical_protein 39343:39879
hypothetical_protein 19953:20498	hypothetical_protein 18014:18553	hypothetical_protein 23282:23800
virion_structural_protein 28178:28681	virion_structural_protein 26237:26740	hypothetical_protein 30580:31083
hypothetical_protein 49218:49430	hypothetical_protein 47272:47487	hypothetical_protein 51160:51375
ParB-like_partition_protein 50488:51252	ParB-like_partition_protein 48409:49176	hypothetical_protein 52254:53021
hypothetical_protein 29181:29540	minor_tail_protein 27240:27599	hypothetical_protein 31583:31942
tail_terminator 21297:21719	tail_terminator 19360:19782	hypothetical_protein 24607:25029
minor_tail_protein 28678:29184	minor_tail_protein 26737:27243	enolase-like_protein 31080:31586

vB-EcoS_164-2_02	Solent_MH586730	vB_SenS_Sasha_NC_047786
hypothetical_protein 4547:4891	putative_transmembrane_protein 14163:14507	hypothetical_protein 11045:11389
hypothetical_protein 5663:5989	hypothetical_protein 14761:15087	hypothetical_protein 11643:11969
hypothetical_protein 7407:7601	NinH 16408:16596	NinH 13290:13478
holin 31407:31730	putative_holin_or_antiholin 36905:37231	holin 33787:34113
tail_protein 25429:27960	tail_tip_protein 32261:34789	tail_protein 29143:31671
dUTPase 5989:6492	deoxyuridine_5`-triphosphate_nucleotidohydrolase 15087:15590	deoxyuridine_5'-triphosphate_nucleotidohydrolase 11969:12472
head_morphogenesis 7775:8734	head_morphogenesis_protein 16780:17739	head_morphogenesis 13662:14621
hypothetical_protein 4199:4531	hypothetical_protein 13812:14147	hypothetical_protein 10694:11029
terminase_large_subunit 1:1764	large_terminase_subunit 6521:8284	TerL_protein 1805:3568
portal_protein 2810:4165	portal_protein 12424:13779	portal_protein 9306:10661
Rz-like_spanin 32344:32742	i-spanin 37839:38237	Rz-like_spanin 34721:35119
hypothetical_protein 4954:5202	hypothetical_protein 14513:14764	hypothetical_protein 11395:11646
hypothetical_protein 39126:39887	putative_TonB_receptor_protein 44863:45624	hypothetical_protein 41835:42596
nucleoside_2-deoxyribosyltransferase 33828:34178	putative_nucleoside_2-deoxyribosyltransferase 39690:40040	nucleoside_2-deoxyribosyltransferase 36572:36922
terminase_small_subunit 52003:52755	small_terminase_subunit 358:993	terminase_small_subunit 276:911
hypothetical_protein 11388:12647	hypothetical_protein 19662:20927	hypothetical_protein 16544:17809
hypothetical_protein 8967:9242	hypothetical_protein 18283:18558	hypothetical_protein 15165:15440
DNA_polymerase 35087:36943	DNA_polymerase 40944:42800	DNA_polymerase 37826:39682
minor_tail_protein 17659:19116	major_tail_protein 24933:26390	major_tail_tube_protein 21815:23272
DNA_methyltransferase 11020:11388	hypothetical_protein 19252:19662	DNA_methyltransferase 16134:16544
hypothetical_protein 7225:7407	hypothetical_protein 16229:16411	hypothetical_protein 13111:13293
endolysin 31727:32116	endolysin 37228:37611	endolysin 34110:34493
endonuclease 41761:42252	hypothetical_protein 48001:48492	hypothetical_protein 44973:45464
major_head_protein 13061:14080	major_capsid_protein 21341:22360	major_head_protein 18223:19242
tail_completion_or_Neck1_protein 16772:17200	hypothetical_protein 24043:24474	tail_completion_or_Neck1_protein 20925:21356
lipoprotein 45523:45720	putative_transmembrane_lipoprotein 51325:51522	lipoprotein 48297:48494
head_scaffolding_protein 12647:13057	major_virion_structural_protein 20927:21337	major_virion_structural_protein 17809:18219
Gp2.5-like_ssDNA_binding_protein_and_ssDNA_annealing_protein 37024:37965	hypothetical_protein 42881:43702	Gp2.5-like_ssDNA_binding_protein_and_ssDNA_annealing_protein 39763:40674
head_closure_Hc1 15478:15852	hypothetical_protein 23667:24041	hypothetical_protein 20549:20923
exonuclease 37981:39126	hypothetical_protein 43718:44863	exonuclease 40690:41835
membrane_protein 32097:32357	putative_holin_or_antiholin 37622:37852	membrane_protein 34504:34734
DNA_primase 42528:44882	primase 48666:51020	DNA_primase 45638:47992
tail_assembly_chaperone 19565:19987	tail_measure_chaperone 27201:27623	tail_assembly_chaperone 24083:24505
hypothetical_protein 44973:45185	hypothetical_protein 51111:51323	hypothetical_protein 48083:48295
hypothetical_protein 14143:14496	hypothetical_protein 22423:22776	hypothetical_protein 19305:19658
hypothetical_protein 1879:2145	putative_integral_membrane_protein 10518:10793	hypothetical_protein 7401:7676
hypothetical_protein 34511:35047	hypothetical_protein 40368:40904	hypothetical_protein 37250:37786
hypothetical_protein 14939:15478	hypothetical_protein 23257:23667	hypothetical_protein 20139:20549
virion_structural_protein 24074:24577	hypothetical_protein 30906:31409	virion_structural_protein 27788:28291
hypothetical_protein 45713:45928	hypothetical_protein 51515:51730	hypothetical_protein 48487:48702
ParB-like_partition_protein 47419:48186	putative_ParB-like_protein 53308:54075	ParB-like_partition_protein 50280:51047
minor_tail_protein 25077:25436	hypothetical_protein 31975:32268	hypothetical_protein 28857:29150
tail_terminator 17200:17622	hypothetical_protein 24474:24896	tail_terminator 21356:21778
minor_tail_protein 24574:25080	putative_enolase-like_protein 31406:31912	minor_tail_protein 28288:28794

Identification of orthologous genes and their annotated functions. Genes indicated by "+" contained a sufficient amount of phylogenetically informative sites and were therefore retained for phylogenetic reconstruction of phage vB_EcoP-179_3.

seq_length	parsimony	singleton	constant	gene in phylogeny	evolution model	BUCT789_PQ999132	vB_EcoS_HSE2_MG252615
407	221	27	159	+	Blosum62	terminase_large_subunit 10351:11601	terminase 21925:23175
490	78	17	395	+	VT+I	DUF4055_domain-containing_protein 11613:13106	putative_structural_protein 23187:24680
208	71	0	137	+	Q.pfam	ssDNA_binding_and_annealing_protein 37790:38422	helix-destabilizing_protein 8471:9097
747	71	54	622	+	WAG+I	replicative_helicase/primase 597:2855	putative_helicase-primase 12085:14322
138	70	3	65	+	VT+I	tail_assembly_chaperone 23000:23416	hypothetical_protein 35302:35718
388	35	17	336	+	JTT+I	tail_tube_protein 19726:20892	tail_protein 31869:33035
171	33	4	134	+	VT	hypothetical_protein 27479:27994	hypothetical_protein 38907:39422
131	30	5	96	+	VT+I	tail_protein 18915:19310	tail_protein 31058:31453
121	27	6	88	+	WAG	hypothetical_protein 27991:28356	peptidoglycan_endopeptidase 39419:39784
376	18	22	336	-, parsimony informative sites less than 5% of all alignment		calcineurin-like_phosphoesterase 21313:22443	calcineurin-like_phosphoesterase_superfamily_domain_protein 33774:34904
94	12	6	76	+	LG+I	VRR-NUC_domain_protein 34791:35105	VRR-NUC_domain_protein 5367:5651
474	11	19	444	-, parsimony informative sites less than 5% of all alignment		DNA_helicase 32874:34298	helicase 3715:5139
347	10	15	322	-, parsimony informative sites less than 5% of all alignment		minor_capsid_protein 13176:14219	head_protein 24717:25793
204	10	18	176	-, parsimony informative sites less than 5% of all alignment		prohead_core_protein_protease 17942:18559	hypothetical_protein 30088:30702
733	7	129	597	-, less than 10 parsimony informative sites		DNA_polymerase_I 35458:37731	DNA_polymerase 5932:8136
128	7	4	117	-, less than 10 parsimony informative sites		spanin 15023:15409	spanin 27168:27554
170	6	14	150	-, less than 10 parsimony informative sites		head-tail_adaptor_Ad1 17427:17939	hypothetical_protein 29573:30085
64	5	5	54	-, less than 10 parsimony informative sites		hypothetical_protein 9060:9296	transcriptional_regulator 20873:21109
93	5	10	78	-, less than 10 parsimony informative sites		hypothetical_protein 6083:6376	hypothetical_protein 18028:18309
119	4	9	106	-, less than 10 parsimony informative sites		minor_capsid_protein 18559:18918	ATP-binding_sugar_transporter_protein 30702:31061
413	4	68	341	-, less than 10 parsimony informative sites		nuclease_superfamily_protein 38513:39754	nuclease_superfamily_protein 9672:10913
79	4	7	68	-, less than 10 parsimony informative sites		hypothetical_protein 8369:8608	hypothetical_protein 20295:20534
60	2	2	56	-, less than 10 parsimony informative sites		hypothetical_protein 8595:8777	hypothetical_protein 20455:20703
137	2	4	131	-, less than 10 parsimony informative sites		tail_terminator 19310:19723	hypothetical_protein 31450:31866
153	2	18	133	-, less than 10 parsimony informative sites		tail_protein 14219:14680	tail_protein 25793:26254
67	1	2	64	-, less than 10 parsimony informative sites		hypothetical_protein 5880:6083	hypothetical_protein 17366:17584
161	0	16	145	-, less than 10 parsimony informative sites		endolysin 6988:7488	endolysin 18921:19421
76	0	2	74	-, less than 10 parsimony informative sites		hypothetical_protein 21086:21316	hypothetical_protein 33063:33293

SZH-1_ON391668	vB-EcoP_179-3_06	AV104_OR352935
large_terminase 42135:43385	hypothetical_protein 1:1131	terminase 1:1236
putative_structural_protein 40641:42122	portal_protein 1144:2616	62kDa_structural_protein 1249:2730
DNA_helix_destabilizing_protein 13324:13950	Gp2.5-like_ssDNA_binding_protein_and_ssDNA_annealing_protein 28850:29482	hypothetical_protein 29662:30294
replicative_DNA_helicase 8418:10661	replicative_helicase-primase 32574:34841	replicative_DNA_helicase 32974:35259
tail_assembly_chaperone 28881:29297	tail_assembly_chaperone 14914:15330	hypothetical_protein 13875:14291
tail_tube_protein 31409:32575	minor_tail_protein 11224:12390	hypothetical_protein 10282:11448
hypothetical_protein 25207:25722	minor_tail_protein 18519:19034	hypothetical_protein 18348:18863
tail_component 32991:33386	tail_completion_or_Neck1_protein 10413:10808	hypothetical_protein 9471:9866
hypothetical_protein 24845:25210	minor_tail_protein 19031:19396	hypothetical_protein 18860:19225
phosphatase 29854:30984	DNA_repair_exonuclease 13387:14517	hypothetical_protein 12189:13319
VRR-NUC_domain_protein 16788:17102	Holliday_junction_resolvase 25851:26165	hypothetical_protein 26510:26824
DNA_helicase 18262:19686	DNA_helicase 23943:25367	DNA_helicase 23926:25350
head_protein 39313:40356	head_morphogenesis 3530:4573	hypothetical_protein 2800:3843
hypothetical_protein 33852:34466	hypothetical_protein 9443:10057	hypothetical_protein 8501:9115
DNA_polymerase 14162:16435	DNA_polymerase 26518:28791	DNA_polymerase_I 27177:29450
spanin 37552:37938	Rz-like_spanin 5649:6035	hypothetical_protein 5032:5418
hypothetical_protein 34469:34981	head-tail_adaptor_Ad1 8928:9440	hypothetical_protein 7986:8498
hypothetical_protein 357:593	hypothetical_protein 42026:42226	hypothetical_protein 41842:42078
hypothetical_protein 4462:4755	hypothetical_protein 38469:38762	hypothetical_protein 38906:39199
head-to-tail_connector_complex_protein 33383:33742	tail_completion_or_Neck1_protein 10057:10416	hypothetical_protein 9115:9474
nuclease_superfamily_protein 11821:13062	exonuclease 29809:31050	hypothetical_protein 30562:31803
hypothetical_protein 1376:1615	hypothetical_protein 41210:41449	hypothetical_protein 41194:41433
hypothetical_protein 1207:1389	hypothetical_protein 41436:41618	hypothetical_protein 41420:41602
tail_completion_protein 32578:32991	tail_terminator 10808:11221	hypothetical_protein 9866:10279
tail_protein 38852:39313	Hoc-like_head_decoration 4573:5034	fibritin_protein 3843:4304
hypothetical_protein 5095:5298	hypothetical_protein 37860:38078	hypothetical_protein 38332:38535
lysozyme 2932:3432	endolysin 39795:40295	lysin 39826:40311
hypothetical_protein 30981:31211	hypothetical_protein 13160:13390	hypothetical_protein 11478:11708

Identification of orthologous genes and their annotated functions. Genes indicated by "+" contained a sufficient amount of phylogenetically informative sites and were therefore retained for phylogenetic reconstruction of phage vB_EcoP-179_5.

seq_length	parsimony	singleton	constant	gene in phylogeny	evolution model	vB_EcoP-101101UKE1_MZ234012	vB-Eco-KMB14_OR424384	6948_OL362272
398	62	14	322	+	WAG+G4	major_capsid_protein 20985:22187	major_capsid_protein 23240:24445	major_capsid_protein 31624:32832
190	40	22	128	+	VT+G4	hypothetical_protein 6137:6820	hypothetical_protein 8475:9158	hypothetical_protein 16838:17521
91	35	5	51	+	WAG+G4	hypothetical_protein 36273:36551	acetyl-CoA_acetyltransferase 38187:38465	hypothetical_protein 2363:2641
1061	35	20	1006	-, parsimony informative sites less than 5% of all alignment		peptidoglycan_lytic_exotransglycosylase 29140:32337	peptidoglycan_lytic_exotransglycosylase 31374:34682	peptidoglycan_lytic_exotransglycosylase 39761:42958
294	33	12	249	+	JTTDCMut+I	hypothetical_protein 1:924	hypothetical_protein 2186:3109	hypothetical_protein 10485:11408
315	32	12	271	+	LG+I	ATP-dependent_DNA_ligase 16492:17439	DNA_ligase 18747:19694	ATP-dependent_DNA_ligase 27000:27947
306	29	11	266	+	JTTDCMut+G4	capsid_assembly_scaffolding_protein 19988:20908	capsid_assembly_scaffolding_protein 22243:23163	capsid_assembly_scaffolding_protein 30624:31544
95	27	11	57	+	WAG+G4	hypothetical_protein 36570:36863	hypothetical_protein 38484:38777	hypothetical_protein 2660:2950
631	20	7	604	-, parsimony informative sites less than 5% of all alignment		terminase_large_subunit 34218:36116	DNA_packaging_protein 36131:38029	terminase_large_subunit 300:2198
114	19	5	90	+	WAG+G4	hypothetical_protein 36863:37207	endolysin 38777:39121	hypothetical_protein 2953:3297
240	14	4	222	+	HIVb+F+I	internal_virion_protein 25381:26127	internal_virion_protein 27639:28361	internal_virion_protein 36026:36748
59	10	15	34	+	cpREV+G4	hypothetical_protein 42821:43000	hypothetical_protein 595:774	hypothetical_protein 8853:9032
245	10	8	227	-, parsimony informative sites less than 5% of all alignment		non-contractile_tail_tubular_protein 22242:22979	tail_tubular_protein_A 24500:25237	non-contractile_tail_tubular_protein 32887:33624
123	8	4	111	-, less than 10 parsimony informative sites		hypothetical_protein 12630:13010	hypothetical_protein 14850:15221	hypothetical_protein 23332:23703
835	8	139	688	-, less than 10 parsimony informative sites		DNA-directed_DNA_polymerase 7572:10118	DNA_polymerase 9910:12456	DNA-directed_DNA_polymerase 18273:20819
786	6	24	756	-, less than 10 parsimony informative sites		DNA-directed_RNA_polymerase 988:3615	DNA-directed_RNA_polymerase 3173:5800	DNA-directed_RNA_polymerase 11472:14099
435	5	4	426	-, less than 10 parsimony informative sites		primase/helicase_protein 4072:6060	DNA_primase/helicase 6253:8241	primase/helicase_protein 14773:16761
320	4	17	299	-, less than 10 parsimony informative sites		non-contractile_tail_fiber_protein 32337:33299	non-contractile_tail_fiber_protein 34682:35644	non-contractile_tail_fiber_protein 42958:43920
982	4	18	960	-, less than 10 parsimony informative sites		hypothetical_protein 26127:29075	baseplate_hub 28361:31309	hypothetical_protein 36748:39696
123	3	33	87	-, less than 10 parsimony informative sites		hypothetical_protein 11036:11413	hypothetical_protein 13170:13547	hypothetical_protein 21737:22114
335	3	49	283	-, less than 10 parsimony informative sites		phosphoesterase 14577:15716	metallo-phosphoesterase 16964:17971	phosphoesterase 25387:26394
71	3	21	47	-, less than 10 parsimony informative sites		hypothetical_protein 6831:7049	hypothetical_protein 9169:9387	hypothetical_protein 17532:17750
153	3	6	144	-, less than 10 parsimony informative sites		hypothetical_protein 17756:18217	GNAT_family_N-acetyltransferase 20011:20472	hypothetical_protein 28392:28853
267	3	10	254	-, less than 10 parsimony informative sites		hypothetical_protein 11491:12297	hypothetical_protein 13710:14516	hypothetical_protein 22192:22998
99	2	4	93	-, less than 10 parsimony informative sites		terminase_small_subunit 33487:33786	terminase_small_subunit 35832:36131	terminase_small_subunit 1:300
69	1	1	67	-, less than 10 parsimony informative sites		hypothetical_protein 18227:18436	hypothetical_protein 20482:20691	hypothetical_protein 28863:29072
516	1	2	513	-, less than 10 parsimony informative sites		head-to-tail_connector_protein 18438:19988	collar 20693:22243	head-to-tail_connector_protein 29074:30624
72	1	5	66	-, less than 10 parsimony informative sites		hypothetical_protein 12307:12525	hypothetical_protein 14526:14744	hypothetical_protein 23008:23226
308	0	10	298	-, less than 10 parsimony informative sites		exonuclease 13293:14321	exonuclease 15599:16576	exonuclease 23971:24999

vB_Ec-G9_PQ433351	EC120_ON185580	vB-EcoP_179-5_05	EEc2_ON210144	vB_Eco_Sem1_PV717788	VEc3_MG251390
major_head_protein 16821:18026	major_capsid_protein 22709:23917	major_head_protein 23710:24918	major_capsid_protein 23554:24762	major_capsid_protein 40876:42081	major_capsid_protein 23474:24679
hypothetical_protein 2324:2905	hypothetical_protein 8214:8897	hypothetical_protein 8764:9453	hypothetical_protein 8861:9544	hypothetical_protein 26079:26753	hypothetical_protein 9172:9855
acetyl-CoA_acetyltransferase 31597:31938	hypothetical_protein 38526:38807	acetyl-CoA_acetyltransferase 38562:38843	hypothetical_protein 38491:38772	hypothetical_protein 11111:11389	Sema_domain-containing_protein 38861:39139
internal_virion_protein_with_endolysin_domain 24961:28158	hypothetical_protein 30848:34156	internal_virion_protein_with_endolysin_domain 31874:35059	peptidoglycan_lytic_exotransglycosylase 31691:34999	DNA_ejectosome_component 5306:7606	internal_virion_protein 31608:34916
hypothetical_protein 39806:40690	hypothetical_protein 2076:2999	hypothetical_protein 2656:3540	hypothetical_protein 2390:3274	hypothetical_protein 19570:20493	hypothetical_protein 2460:3383
ATP-dependent_DNA_ligase 12313:13275	ATP-dependent_DNA_ligase 18085:19032	ATP-dependent_DNA_ligase 19214:20161	ATP-dependent_DNA_ligase 18619:19566	ATP-dependent_DNA_ligase 36255:37202	DNA_ligase 18853:19800
head_scaffolding_protein 15824:16744	capsid_assembly_scaffolding_protein 21709:22629	head_scaffolding_protein 22710:23630	capsid_and_scaffold_protein 22554:23474	capsid_assembly_scaffolding_protein 39879:40799	scaffolding_protein 22477:23397
acetyl-CoA_acetyltransferase 31957:32247	hypothetical_protein 38825:39112	acetyl-CoA_acetyltransferase 38861:39148	hypothetical_protein 38790:39077	hypothetical_protein 11408:11701	hypothetical_protein 39158:39448
terminase_large_subunit 29607:31505	terminase_large_subunit 36475:38373	terminase_large_subunit 36508:38403	terminase_large_subunit 36448:38343	terminase_large_subunit 9055:10953	terminase_large_subunit 36797:38695
endolysin 32250:32594	hypothetical_protein 39115:39459	endolysin 39839:40183	hypothetical_protein 39769:40113	hypothetical_protein 11701:12045	peptidase 39451:39795
internal_virion_protein 21220:21948	internal_virion_protein 27113:27835	internal_virion_protein 28114:28860	internal_virion_protein 27956:28678	DNA_ejectosome_component_internal_virion_protein 564:1286	internal_virion_protein 27873:28595
hypothetical_protein 38419:38640	hypothetical_protein 548:727	hypothetical_protein 1184:1369	hypothetical_protein 783:965	hypothetical_protein 17967:18143	hypothetical_protein 961:1137
tail_protein 18081:18818	non-contractile_tail_tubular_protein 23974:24711	tail_protein 24975:25712	non-contractile_tail_tubular_protein 24817:25554	non-contractile_tail_tubular_protein 42136:42873	tail_tubular_protein_A 24734:25471
hypothetical_protein 8208:8585	hypothetical_protein 14244:14618	hypothetical_protein 14970:15338	hypothetical_protein 14892:15263	hypothetical_protein 32364:32735	hypothetical_protein 14997:15371
DNA_polymerase 3425:5998	DNA-directed_DNA_polymerase 9649:12195	DNA_polymerase 10030:12540	DNA-directed_DNA_polymerase 10296:12842	DNA-directed_DNA_polymerase 27756:30305	DNA_polymerase 10402:12948
RNA_polymerase 40754:43381	DNA-directed_RNA_polymerase 3063:5690	RNA_polymerase 3604:6231	DNA-directed_RNA_polymerase 3338:5965	DNA-directed_RNA_polymerase 20557:22917	RNA_polymerase 3447:6074
DNA_primase/helicase 72:1988	primase/helicase_protein 6149:8137	DNA_primase/helicase 6699:8687	primase/helicase_protein 6639:8627	primase/helicase_protein 24538:25845	DNA_helicase 6950:8938
tail_fiber_protein 28158:29120	non-contractile_tail_fiber_protein 34156:35118	tail_fiber_protein 35059:36021	non-contractile_tail_fiber_protein 34999:35961	non-contractile_tail_fiber_protein 7606:8568	tail_fiber_protein 34916:35878
lysozyme_domain-containing_protein 21948:24896	hypothetical_protein 27835:30783	lysozyme_domain-containing_protein 28860:31808	hypothetical_protein 28678:31626	baseplate_hub_subunit 1286:4234	lysozyme-like_domain_protein 28595:31543
polysaccharide_chain_length_determinant_protein 6623:6994	hypothetical_protein 12565:12942	polysaccharide_chain_length_determinant_protein 13244:13621	hypothetical_protein 13212:13589	hypothetical_protein 30675:31052	hypothetical_protein 13324:13695
metallo-phosphoesterase 10715:11722	phosphoesterase 16302:17309	metallo-phosphoesterase 17472:18503	phosphoesterase 17006:18013	phosphoesterase 34340:35479	calcineurin-like_phosphoesterase 17070:18077
hypothetical_protein 2916:3134	hypothetical_protein 8908:9126	hypothetical_protein 9464:9679	hypothetical_protein 9555:9773	hypothetical_protein 26772:26990	hypothetical_protein 9866:10084
acetyltransferase 13592:14053	hypothetical_protein 19477:19938	acetyltransferase 20478:20939	hypothetical_protein 20322:20783	hypothetical_protein 37647:38108	N-acetyltransferase 20245:20706
hypothetical_protein 7072:7875	hypothetical_protein 13105:13911	hypothetical_protein 13831:14637	hypothetical_protein 13752:14558	hypothetical_protein 31225:32031	hypothetical_protein 13858:14664
terminase_small_subunit 29308:29607	terminase_small_subunit 35744:36043	terminase_small_subunit 36209:36508	terminase_small_subunit 36149:36448	terminase_small_subunit 8756:9055	terminase_small_subunit 36066:36365
host_range_and_adsorption_protein 14063:14272	hypothetical_protein 19948:20157	hypothetical_protein 20949:21158	hypothetical_protein 20793:21002	hypothetical_protein 38118:38327	hypothetical_protein 20716:20925
head-tail_adaptor 14274:15824	head-to-tail_connector_protein 20159:21709	head-tail_adaptor 21160:22710	head-to-tail_connector_protein 21004:22554	collar_head-to-tail_connector_protein 38329:39879	head-tail_connector 20927:22477
hypothetical_protein 7885:8103	hypothetical_protein 13921:14139	hypothetical_protein 14647:14865	hypothetical_protein 14568:14786	hypothetical_protein 32041:32259	hypothetical_protein 14674:14892
exonuclease 8967:9896	exonuclease 14886:15914	exonuclease 15654:16679	exonuclease 15590:16618	exonuclease 33056:34084	exonuclease 15654:16682

Supplementary Table S3

Development prediction provided by Phatyp.

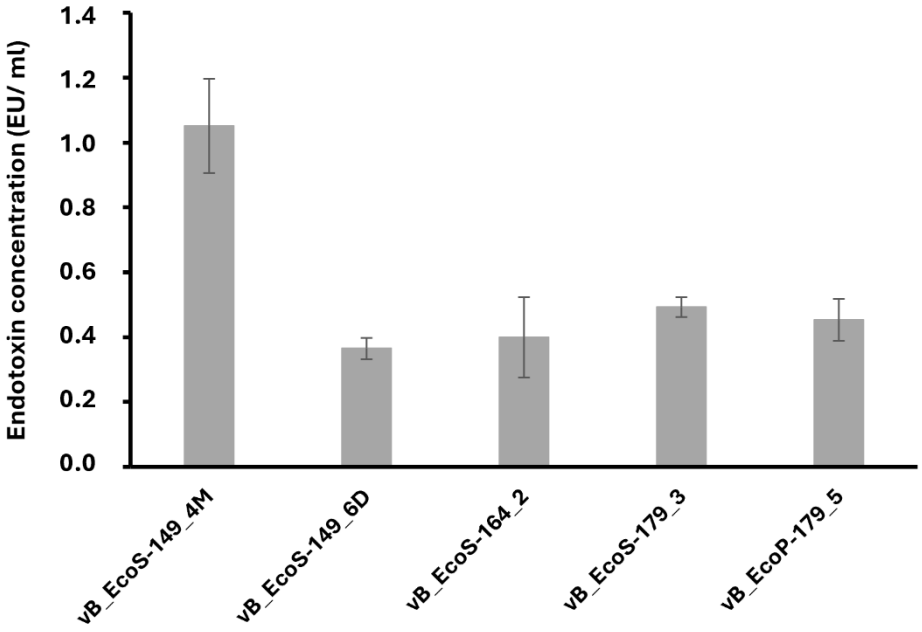
Phage Name_Accession No	Length	TYPE	PhaTYPScore
JK16_MK962751	51854	virulent	1.0
vB_EcoS_XY1_MN927225	45566	virulent	1.0
K1G_GU196277	43587	virulent	1.0
BUCT825_PV040721	44228	virulent	1.0
YSP2_MG241338	50316	virulent	1.0
vB_EcoP-101101UKE1_MZ234012	44450	virulent	1.0
vB-EcoS_164-2_PX985797	56684	virulent	0.98
Henu4_2_PQ159479	50260	virulent	1.0
K1ind1_GU196279	42292	virulent	1.0
vB_SenS_PHB07_MH102284	51818	virulent	1.0
vB_EcoA_fTuEco01_MZ031013	44320	virulent	1.0
MatMar_OR896829	50761	virulent	1.0
vB_EcoS-UTI89UKE1_MZ234048	41265	virulent	1.0
JLBYU07_OK272485	50290	virulent	1.0
vB_CfrD_SorkZaugg_OL539466	50023	virulent	1.0
vB-Eco-KMB14_OR424384	44563	virulent	1.0
JLBYU16_OK272471	49995	virulent	1.0
vB_Eco_Sem1_PV717788	44711	virulent	1.0
vB_EcoS_SCS92_ON548432	44240	virulent	1.0
damhaus_MN850602	51154	virulent	1.0
BEK12A_MN158210	52716	virulent	1.0
vB_EcoP_C_KY295892	44970	virulent	1.0
vB_EclS_AS5_OR753409	49578	virulent	1.0
JLBYU19_OK272489	49897	virulent	1.0
vB_EcoP_B_KY295891	44018	virulent	1.0
vB_EcoS_fTaEco03_MT711528	41972	virulent	1.0
tonijn_MN850641	51627	virulent	1.0
Tuinin_OR896828	51375	virulent	1.0
vB_CfrD_Brooksby_OL539443	50071	virulent	1.0
UTI-CM001_OM810255	41222	virulent	1.0
LL5_MH491968	49788	virulent	1.0
vB_CfrD_ElisaCorrea_OK499973	49523	virulent	1.0
NBSal001_MN994500	50922	virulent	1.0
vB_EcoS_PHB17_MN090155	48939	virulent	1.0
Jahat_MG145_MK552105	50984	virulent	1.0
vB_EcoP_ACG-C91_JN986844	43731	virulent	1.0
phiWAO78-1_MW331437	44551	virulent	1.0
EC167_ON185587	50988	virulent	1.0
vB_GS156_PV153712	48552	virulent	1.0
EEc2_ON210144	44883	virulent	1.0
phiE31-2_PP988008	51468	virulent	1.0
vB_EcoS-22664BS2_MZ234010	45176	virulent	1.0
JSSK01_OQ442786	44509	virulent	1.0
vB-Eco-KMB47_OR424385	44488	virulent	1.0
IsaakIselin_MZ501077	49911	virulent	1.0
EC120_ON185580	44544	virulent	1.0

Henu7_MN019128	49526	virulent	1.0
MBBL140_OR786094	49190	virulent	1.0
tinuso_MN850634	50856	virulent	1.0
Henu8_MN055691	49890	virulent	1.0
vB_Ka_Sem4_PV775411	45205	virulent	1.0
vB_EcoS_UTEC10_PP291582	44773	virulent	1.0
grams_MN850567	49530	virulent	1.0
tunus_MN850638	51111	virulent	1.0
NTEC3_OK539620	44240	virulent	1.0
vB-Eco-KMB25_OR424386	50475	virulent	1.0
WaterSpirit_OR896833	51324	virulent	1.0
9NA_NC_025443	52869	virulent	0.98
vB_EcoD_Poky_OL539445	51374	virulent	1.0
GSP032_ON855040	52100	virulent	1.0
K1E_KY435490	44246	virulent	1.0
BEBK14_MN158214	50853	virulent	1.0
ityhuna_MN850582	50768	virulent	1.0
vB-Eco-KMB36_OR525703	44710	virulent	1.0
KarlJaspers_MZ501090	51474	virulent	1.0
JakobBernoulli_MZ501079	51129	virulent	1.0
FritzSarasin_MZ501069	51777	virulent	1.0
26_MZ832314	42462	virulent	1.0
vB_EcoD_Phunderstruck_OL539446	51274	virulent	1.0
6949_OL362273	44371	virulent	1.0
tonn_MN850596	51012	virulent	1.0
36_KR296690	41085	virulent	1.0
tuinn_MN850606	50505	virulent	1.0
Stevie_KM236241	49816	virulent	1.0
vB_EcoS_fTaEco01_MT711527	41273	virulent	1.0
RP180_MK737937	44851	virulent	1.0
vB_CfrD_Supreme284_MW021750	49439	virulent	1.0
vB_EcoP-101118B1_MZ234020	44526	virulent	1.0
vB_VIPECOTH07_PQ417132	44205	virulent	1.0
MikeNSara_OR896827	50759	virulent	1.0
K1H_GU196278	41632	virulent	1.0
vB-Eco-KMB46_OR525699	50953	virulent	1.0
vB_EcoP-UTI89UKE3_MZ234050	44294	virulent	1.0
vB_SenS_Sasha_NC_047786	53263	temperate	1.0
N5822_MW032452	78128	virulent	1.0
vB-EcoP_179-3_PX985798	43053	virulent	1.0
vB_EcoS_fPoEco01_MT711526	44391	virulent	1.0
vB_EcoS-phiEc3_MZ576217	45288	virulent	1.0
JLBYU09_OK272487	48988	virulent	1.0
tiwna_MN850643	51014	virulent	1.0
K1-5_AY370674	44385	virulent	1.0
vB_EcoD_Fulano1_OL539459	51759	virulent	1.0
JLBYU38_OK272481	49995	virulent	1.0
BubbaBully_OR896831	51209	virulent	1.0

tonnikala_MN850613	51277	virulent	1.0
Motto_ON843697	49960	virulent	1.0
vB_EcoS_011D2_MT478992	50633	virulent	1.0
herni_MN850640	50971	virulent	1.0
PeterMerian_MZ501101	47775	virulent	1.0
vB_EcoP_K_KY295897	37775	virulent	1.0
vB_EcoS-Ro145clw_MG852086	42031	virulent	1.0
ULINTec2_MZ997838	40815	virulent	1.0
JLBYU41_OK272479	51277	virulent	1.0
VEc3_MG251390	44895	virulent	1.0
Pocis76_MW689258	48795	virulent	1.0
Tls_AY308796	49902	virulent	1.0
vB_EcoS-95_MF564201	50910	virulent	1.0
AV104_OR352935	43102	virulent	1.0
GJL01_KY657202	50407	virulent	1.0
aalborv_MN850591	46660	virulent	1.0
L_AB-2017_KY295896	41039	virulent	1.0
ZCEC5_MK542015	43372	virulent	1.0
vB_EcoS_W011D_MK778457	49847	virulent	1.0
JLBYU01_OK272478	51210	virulent	1.0
vB_CfrD_ZerotoHero_MW021748	49802	virulent	1.0
mutPK1A2_MG004687	44784	virulent	1.0
slyngel_MT074433	51048	virulent	1.0
ULINTec6_MZ997840	44728	virulent	1.0
vB_EcoD_Mishu_OK499984	49713	virulent	1.0
vB_SenS_UTK0006_OQ359885	49758	virulent	1.0
BEK26_MN158216	52716	virulent	1.0
vB_EcoS_XY2_MN927226	44138	virulent	1.0
vB_EcoS-G3B1_MZ234033	51299	virulent	1.0
vB_Eco_Sem2_PV775413	44244	virulent	1.0
Sephi301i_PP503414	48674	virulent	1.0
JohannLBurckhardt_MZ501085	49715	virulent	1.0
JLBYU26_OK272483	51253	virulent	1.0
SFP17_MN432485	51296	virulent	1.0
Shashou_MK931440	44155	virulent	1.0
vB_EcoS_G29-2_MK373798	51739	virulent	1.0
paulsarasin_MZ501099	51285	virulent	1.0
ULINTec7_MZ997841	46142	virulent	1.0
vojen_MN850569	50709	virulent	1.0
VB_EcoS-Golestan_MG099933	44829	virulent	1.0
vB_SenS_Sergei_KY002061	56051	temperate	1.0
F61_OP292674	46635	virulent	1.0
P-K7R_OM317910	52111	virulent	1.0
vB_EcoS_AVIO78A_MZ726797	57811	virulent	1.0
vB_EcoP-G3A1_MZ234032	44471	virulent	1.0
K1ind3_GU196281	43461	virulent	1.0
Sciku_MK931439	43130	virulent	1.0
6948_OL362272	44545	virulent	1.0

KarinMuster_Bas82_PQ850612	58508	temperate	1.0
EC147_ON185589	46060	virulent	1.0
590B_MW722821	44390	virulent	1.0
vB_EcoS-Ro145c2YLVW_MH051334	42145	virulent	1.0
pET128_PQ811887	49720	virulent	1.0
vB-EcoS_149-4M_PX985795	52000	virulent	1.0
vB_EcoP-UTI89UKE2_MZ234049	44293	virulent	1.0
haarsle_MN850600	48613	virulent	1.0
G_AB-2017_KY295895	41519	virulent	1.0
KMM3_OP902292	49164	virulent	1.0
Sazh_MH729819	49665	virulent	1.0
ExeQuayColi_OR896819	51258	virulent	1.0
pSf-1_KC710998	51821	virulent	1.0
BUCT789_PQ999132	40771	virulent	1.0
vB_EcoP_D_KY295893	44931	virulent	1.0
phage_vB_EcoD_Punny_OL539444	49765	virulent	1.0
AV105_OR352936	49722	virulent	1.0
vB_CfrD_Reinasaurus_OL539465	48103	virulent	1.0
vB_EcoS_fKuEco01_MT711525	40825	virulent	1.0
Snoke_MK931441	44454	virulent	1.0
K1ind2_GU196280	42765	virulent	1.0
fBC-Eco01_OM272052	43466	virulent	1.0
atuna_MN850620	50732	virulent	1.0
tunzivis_MN850604	50596	virulent	1.0
phSE-5_KX015771	49178	virulent	1.0
vB_EcoS_fFiEco02_MT711523	44723	virulent	1.0
fEg-Eco06_PQ141284	45491	virulent	1.0
vB_GS71_PV090992	48587	virulent	1.0
vB-EcoP_179-5_PX985799	44811	virulent	1.0
orkinos_MN850586	49798	virulent	1.0
K1E_AM084415	45251	virulent	1.0
vB_EcoP_R_KY295899	44941	virulent	1.0
P_AB-2017_KY295898	41184	virulent	1.0
EEc4_ON210145	44590	virulent	1.0
SZH-1_ON391668	43969	virulent	1.0
MartyMcPhage_OR896836	51258	virulent	1.0
vB_EcoS_fFiEco03_MT711524	43947	virulent	1.0
phage_vB_CfrD_Devorator_OL539461	49712	virulent	1.0
vB_EcoS_HSE2_MG252615	41317	virulent	1.0
U136B_MW598258	49223	virulent	1.0
NorthRox_OR896823	51138	virulent	1.0
Mt1B1_P10_MT496971	45359	virulent	1.0
aaroes_MN850572	51662	virulent	1.0
DK-2017_KY694971	50339	virulent	1.0
F115_OP292673	48162	virulent	1.0
phSE-2_KX015770	49167	virulent	1.0
JLBYU10_OK272486	50987	virulent	1.0
vB_SalS_TTS5_PV007924	51608	virulent	1.0

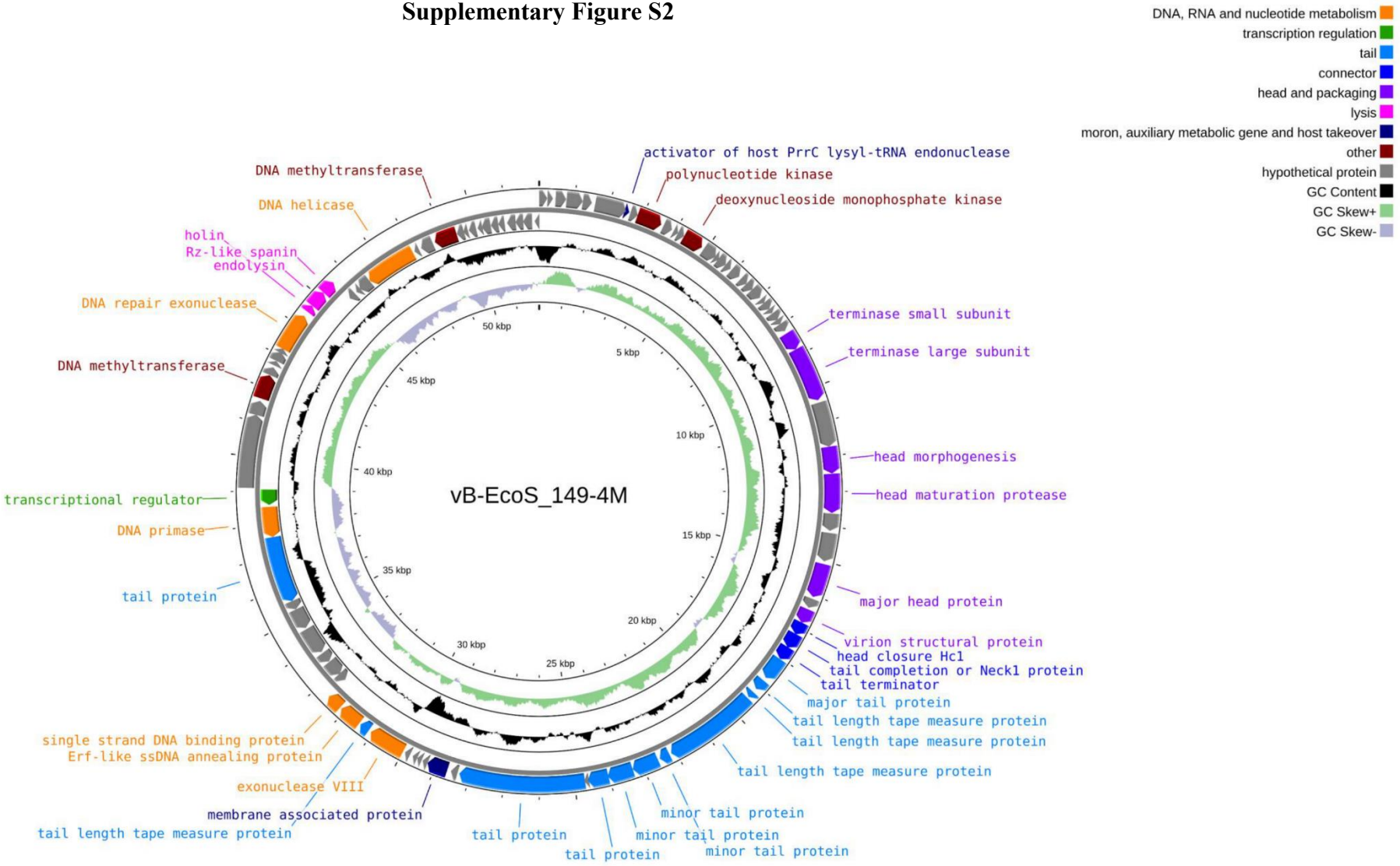
vB_Eco_Sem3_PV775410	45137	virulent	1.0
vB_EcoS_PJ16_OQ955832	57756	virulent	0.96
EC125_ON185586	50477	virulent	1.0
Solent_MH586730	55978	temperate	1.0
vB_VIPECOMC04_PQ423993	45658	virulent	1.0
vB_SenS_UTK0007_OQ359886	51921	virulent	1.0
Schulenburg_MK931438	44748	virulent	1.0
PGN590_MT127619	49043	virulent	1.0
ULINTec4_MZ997839	45159	virulent	1.0
BEK7_MN158209	51031	virulent	1.0
vB_EcoS_SCS44_ON001686	44829	virulent	1.0
vB-EcoS_149-6D_PX985796	51036	virulent	1.0
AAPEc6_KX279892	44559	virulent	1.0
BEC3_MN158217	51031	virulent	1.0
vB_CfrD_BlueShadow_OL539462	49619	virulent	1.0
vB-Eco-KMB35_OR525704	43891	virulent	1.0
DanielBernoulli_MZ501059	50731	virulent	1.0
vB_Ec-G9_PQ433351	44261	virulent	1.0
vB_EcoP_KAW1A4500_MK373773	44241	virulent	1.0
egaa_MN850607	51643	virulent	1.0
vB_Kpn_K50PH164C1_OY978835	49063	virulent	1.0
BEK1-23_MN158211	51031	virulent	1.0
vB_EcoS_XY3_MN781674	51345	virulent	1.0
EC195_ON185588	50477	virulent	1.0
vB_GS166_PV090993	48069	virulent	1.0
ST20_MF153391	44517	virulent	1.0
fEg-Eco19_OL539727	45805	virulent	1.0



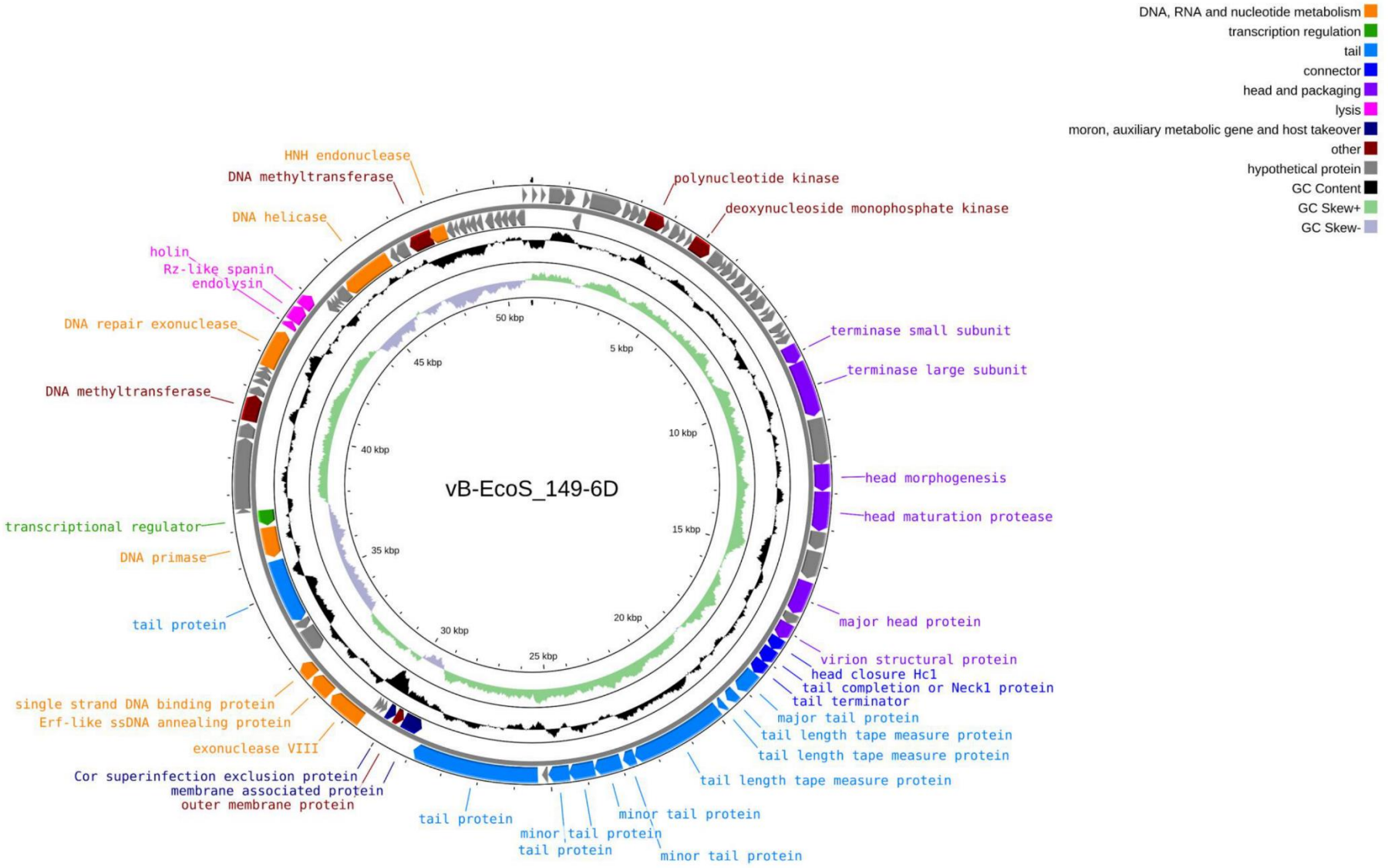
Supplementary Figure S1. Endotoxin levels in purified phage lysates. Endotoxin concentrations (EU/mL) were measured in solutions of phages vB_EcoS-149_4M, vB_EcoS-149_6D, vB_EcoS-164_2, vB_EcoS-179_3, and vB_EcoP-179_5 using the Thermo Scientific Pierce chromogenic endotoxin quant kit. Bars represent mean values, and error bars indicate standard deviation from three replicate measurements.

A

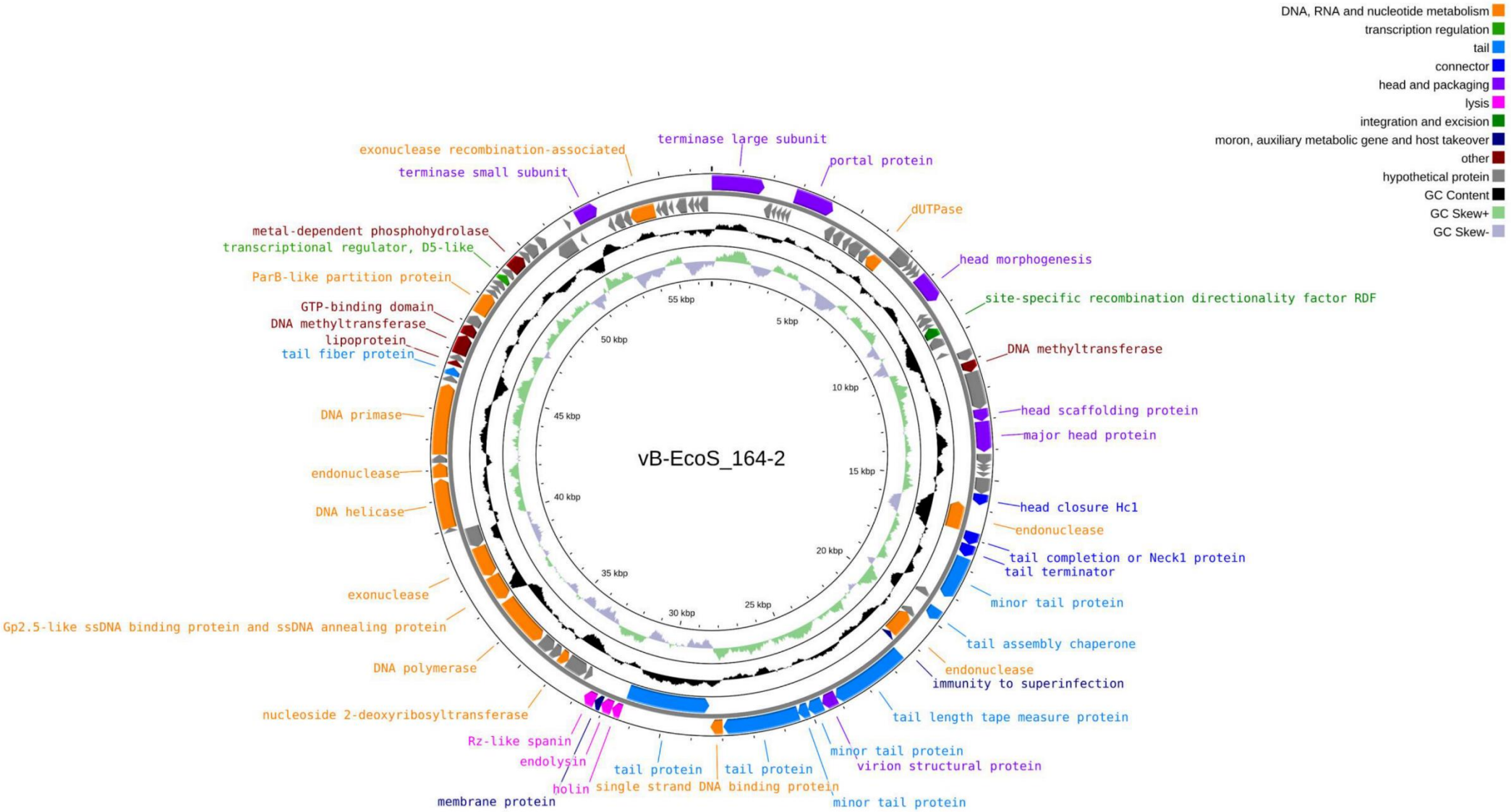
Supplementary Figure S2



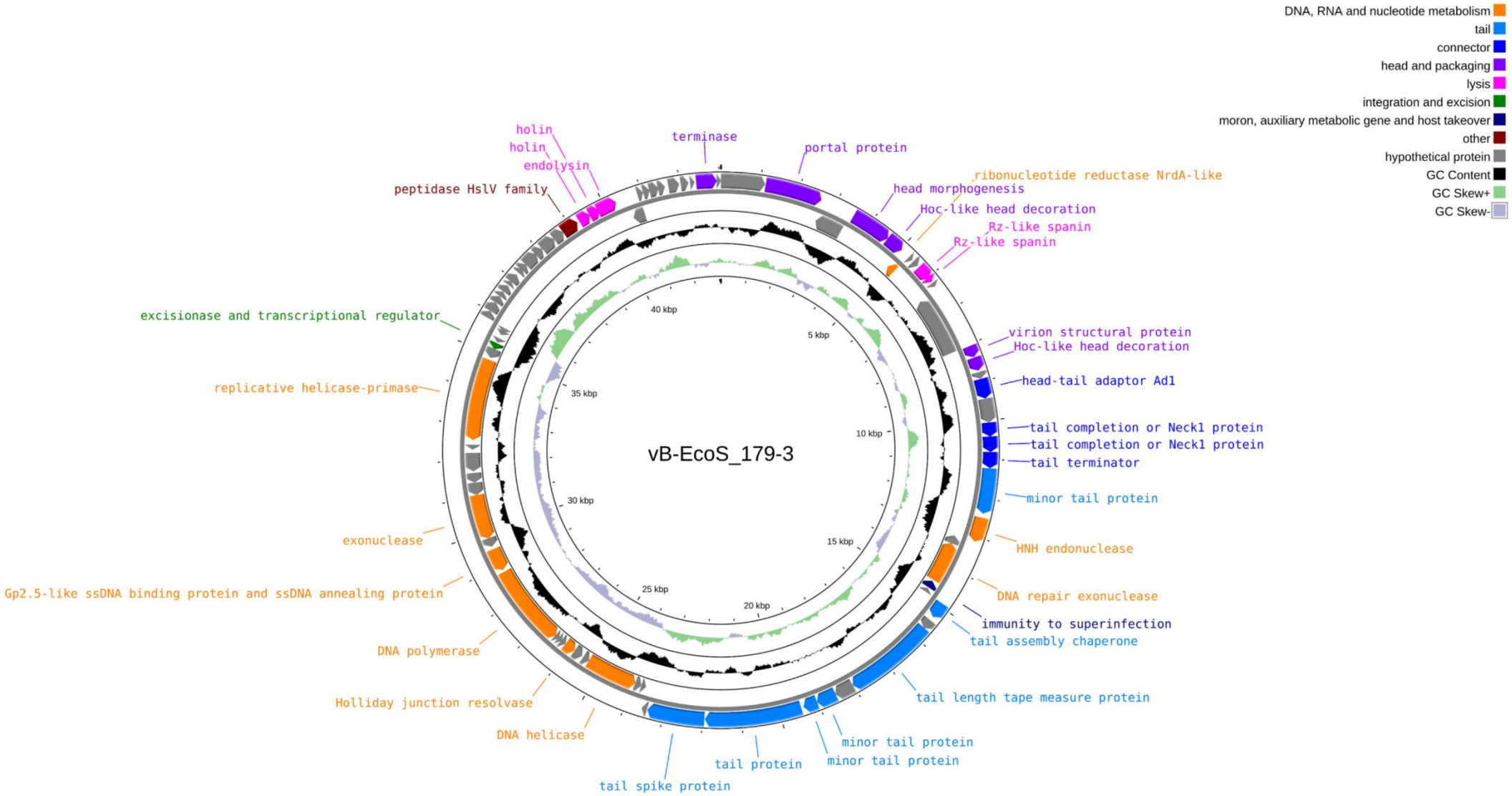
B


DNA, RNA and nucleotide metabolism
transcription regulation
tail
connector
head and packaging
lysis
moron, auxiliary metabolic gene and host takeover
other
hypothetical protein
GC Content
GC Skew+
GC Skew-
vB-EcoS_149-6D
5 kbp
10 kbp
15 kbp
20 kbp
25 kbp
30 kbp
35 kbp
40 kbp
45 kbp
50 kbp
HNH endonuclease
DNA methyltransferase
DNA helicase
holin
Rz-like spanin
endolysin
DNA repair exonuclease
DNA methyltransferase
transcriptional regulator
DNA primase
tail protein
single strand DNA binding protein
Erf-like ssDNA annealing protein
exonuclease VIII
Cor superinfection exclusion protein
membrane associated protein
outer membrane protein
tail protein
minor tail protein
tail protein
minor tail protein
tail length tape measure protein
tail length tape measure protein
tail length tape measure protein
major tail protein
tail terminator
tail completion or Neck1 protein
head closure Hc1
virion structural protein
major head protein
head maturation protease
head morphogenesis
terminase large subunit
terminase small subunit
polynucleotide kinase
deoxynucleoside monophosphate kinase

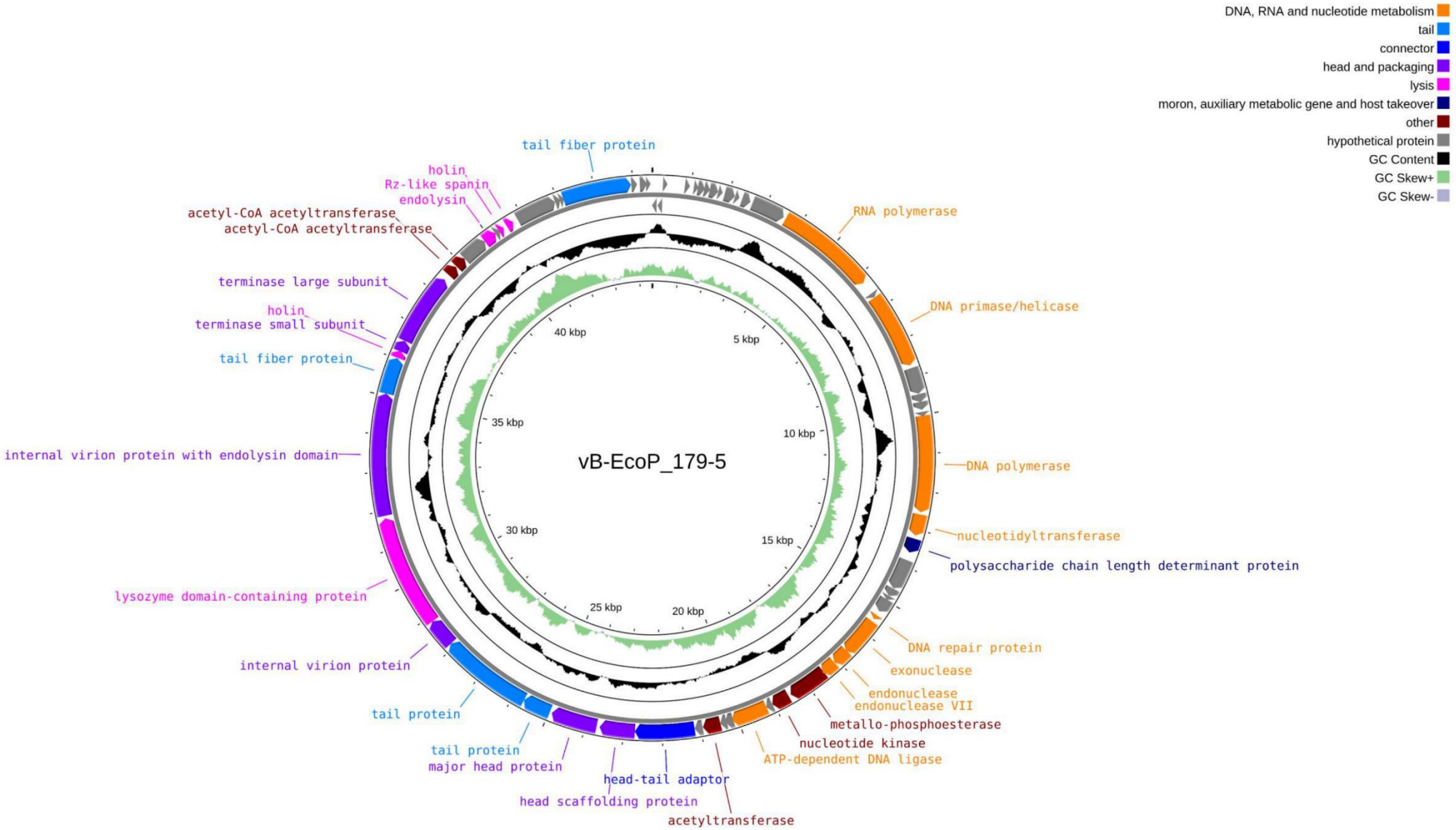
C


DNA, RNA and nucleotide metabolism
transcription regulation
tail
connector
head and packaging
lysis
integration and excision
moron, auxiliary metabolic gene and host takeover
other
hypothetical protein
GC Content
GC Skew+
GC Skew-
vB-EcoS_164-2
5 kbp
10 kbp
15 kbp
20 kbp
25 kbp
30 kbp
35 kbp
40 kbp
45 kbp
50 kbp
55 kbp
terminase large subunit
portal protein
dUTPase
head morphogenesis
site-specific recombination directionality factor RDF
DNA methyltransferase
head scaffolding protein
major head protein
head closure Hc1
endonuclease
tail completion or Neck1 protein
tail terminator
minor tail protein
tail assembly chaperone
endonuclease
immunity to superinfection
tail length tape measure protein
minor tail protein
virion structural protein
minor tail protein
tail protein
tail protein
single strand DNA binding protein
holin
endolysin
Rz-like spanin
membrane protein
nucleoside 2-deoxyribosyltransferase
DNA polymerase
Gp2.5-like ssDNA binding protein and ssDNA annealing protein
exonuclease
endonuclease
DNA helicase
DNA primase
tail fiber protein
lipoprotein
DNA methyltransferase
GTP-binding domain
ParB-like partition protein
transcriptional regulator, D5-like
metal-dependent phosphohydrolase
terminase small subunit
exonuclease recombination-associated

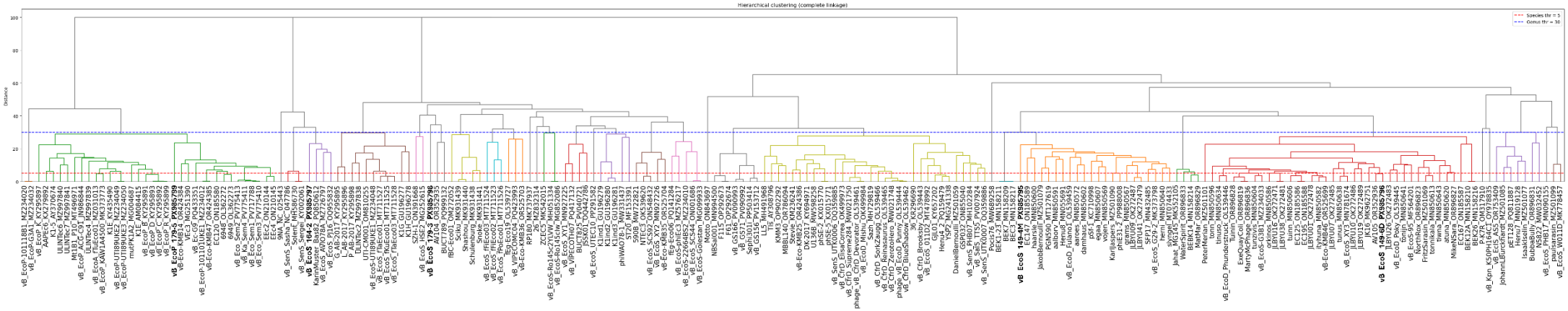
D


DNA, RNA and nucleotide metabolism
tail
connector
head and packaging
lysis
integration and excision
moron, auxiliary metabolic gene and host takeover
other
hypothetical protein
GC Content
GC Skew+
GC Skew-
vB-EcoS_179-3
5 kbp
10 kbp
15 kbp
20 kbp
25 kbp
30 kbp
35 kbp
40 kbp
holin
holin
endolysin
peptidase HslV family
terminase
portal protein
ribonucleotide reductase NrdA-like
head morphogenesis
Hoc-like head decoration
Rz-like spanin
Rz-like spanin
virion structural protein
Hoc-like head decoration
head-tail adaptor Ad1
tail completion or Neck1 protein
tail completion or Neck1 protein
tail terminator
minor tail protein
HNH endonuclease
DNA repair exonuclease
immunity to superinfection
tail assembly chaperone
tail length tape measure protein
minor tail protein
minor tail protein
tail protein
tail spike protein
DNA helicase
Holliday junction resolvase
DNA polymerase
Gp2.5-like ssDNA binding protein and ssDNA annealing protein
exonuclease
replicative helicase-primase
excisionase and transcriptional regulator

E


DNA, RNA and nucleotide metabolism
tail
connector
head and packaging
lysis
moron, auxiliary metabolic gene and host takeover
other
hypothetical protein
GC Content
GC Skew+
GC Skew-
vB-EcoP_179-5
5 kbp
10 kbp
15 kbp
20 kbp
25 kbp
30 kbp
35 kbp
40 kbp
tail fiber protein
holin
Rz-like spanin
endolysin
acetyl-CoA acetyltransferase
acetyl-CoA acetyltransferase
terminase large subunit
holin
terminase small subunit
tail fiber protein
internal virion protein with endolysin domain
lysozyme domain-containing protein
internal virion protein
tail protein
tail protein
major head protein
head scaffolding protein
head-tail adaptor
acetyltransferase
ATP-dependent DNA ligase
nucleotide kinase
metallo-phosphoesterase
endonuclease VII
endonuclease
exonuclease
DNA repair protein
polysaccharide chain length determinant protein
nucleotidyltransferase
DNA polymerase
DNA primase/helicase
RNA polymerase

Supplementary Figure S3



Oświadczenia współautorów

Wojciech Wesołowski
.....
IMIĘ I NAZWISKO

Katedra Biologii Molekularnej,
Wydział Biologii,
Uniwersytet Gdański
.....
AFILIACJA

28.09.2026
.....
DATA

OŚWIADCZENIE

Niniejszym oświadczam, iż mój wkład w publikację:

Wesołowski W, Czerwonka G, Zegadło K, Jagieła E, Łukasiak A, Grabowski Ł, Mantej J, Dziuba A, Bloch S, Topka-Bielecka G, Necel A, Kostovski D, Węglińska E, Narajczyk M, Węgrzyn G, Nejman-Faleńczyk B, Adamus-Białek W. Significant variability in virulence factors and antibiotic resistance of uropathogenic *Escherichia coli* strains may be counteracted by their susceptibility to lytic bacteriophages. Sci Rep. 2026 Jul 21. doi: 10.1038/s41598-026-62793-9,

polegał na:

- współudziale w zbadaniu wydajności infekcji badanymi fagami szczepów UPEC (Rycina nr 3),
- współudziale w przygotowaniu, zagęszczeniu i oczyszczeniu wirionów fagowych z wykorzystaniem ultrawirowania w gradiencie chlorku cezu celem wykonania ich zdjęć (Rycina nr 4) oraz przeprowadzeniu analizy na obecność zanieczyszczeń w postaci endotoksyn (Rycina S1),
- przygotowaniu, rozcieńczeniu i zmiareczkowaniu lizatów fagowych celem wykonania zdjęć łysinek (Rycina nr 4),
- współudziale w przeprowadzeniu analizy stabilności wirionów fagowych w obecności różnych czynników chemicznych i fizycznych (Rycina nr 5),
- złożeniu i wykonaniu funkcjonalnej adnotacji genomów fagowych, przygotowaniu na ich podstawie map kołowych genomów (Rycina S2) oraz zdeponowaniu fagowych sekwencji genomowych w bazie GenBank,
- wykonaniu predykcji cyklu życiowego badanych fagów oraz identyfikacji w ich genomach genów toksyn oraz genów oporności na antybiotyki z użyciem narzędzi bioinformatycznych (Tabela S3),s
- wykonaniu analizy taksonomicznej i filogenetycznej, selekcji genów ortologicznych na podstawie liczby miejsc parsymonicznie informatywnych oraz wygenerowaniu drzew dla każdego z badanych fagów (Rycina S3, Ryciny 7-11, Tabela S2),
- przeprowadzeniu analizy porównawczej organizacji genomów badanych fagów oraz fagów najbliżej z nimi spokrewnionych, wytypowanych na podstawie wyników analizy filogenetycznej (Rycina 12),
- współudziale w przygotowaniu odpowiedzi na uwagi recenzentów,
- a także na obróbce i interpretacji wyników oraz przygotowaniu rycin i tabel.

Wesołowski
.....
PODPIS

Grzegorz Czerwonka
.....
IMIĘ I NAZWISKO

24.08.2026
.....
DATA

Zakład Mikrobiologii,
Instytut Biologii,
Uniwersytet Jana Kochanowskiego w Kielcach
.....
AFILIACJA

OŚWIADCZENIE

Niniejszym oświadczam, iż mój wkład w publikację:

Wesołowski W, Czerwonka G, Zegadło K, Jagieła E, Łukasiak A, Grabowski Ł, Mantej J, Dziuba A, Bloch S, Topka-Bielecka G, Necel A, Kostovski D, Węglińska E, Narajczyk M, Węgrzyn G, Nejman-Faleńczyk B, Adamus-Białek W. Significant variability in virulence factors and antibiotic resistance of uropathogenic *Escherichia coli* strains may be counteracted by their susceptibility to lytic bacteriophages. Sci Rep. 2026 Jul 21. doi: 10.1038/s41598-026-62793-9.

polegał na:

gromadzeniu, opracowywaniu i weryfikacji danych badawczych (Data Curation), przeprowadzaniu analiz formalnych oraz interpretacji uzyskanych wyników (Formal Analysis), realizacji prac badawczych i eksperymentalnych (Investigation), a także współtworzeniu metodologii badań (Methodology). Ponadto uczestniczyłem w przygotowaniu pierwszej wersji manuskryptu (Writing – Original Draft) oraz w jego merytorycznej recenzji, redakcji i korekcie (Writing – Review and Editing).

.....

PODPIS

Katarzyna Zegadło
.....
IMIĘ I NAZWISKO

25.08.2026
DATA

Zakład Mikrobiologii,
Instytut Biologii,
Uniwersytet Jana Kochanowskiego w Kielcach
.....
AFILIACJA

OŚWIADCZENIE

Niniejszym oświadczam, iż mój wkład w publikację:

Wesołowski W, Czerwonka G, Zegadło K, Jagieła E, Łukasiak A, Grabowski Ł, Mantej J, Dziuba A, Bloch S, Topka-Bielecka G, Necel A, Kostovski D, Węglińska E, Narajczyk M, Węgrzyn G, Nejman-Faleńczyk B, Adamus-Białek W. Significant variability in virulence factors and antibiotic resistance of uropathogenic *Escherichia coli* strains may be counteracted by their susceptibility to lytic bacteriophages. Sci Rep. 2026 Jul 21. doi: 10.1038/s41598-026-62793-9.

polegał na:

Przeprowadzenie analiz plazmidów *Escherichia coli* (Data Curation, Formal Analysis, Investigation) oraz współudział w przygotowaniu sekcji manuskryptu dotyczącej identyfikacji i charakterystyki plazmidów.

Zegadło Katarzyna
.....
PODPIS

Ernest Jagieła
.....
IMIĘ I NAZWISKO

Kielce, 24.08.2026.
.....
DATA

Zakład Mikrobiologii,
Instytut Biologii,
Uniwersytet Jana Kochanowskiego w Kielcach
.....
AFILIACJA

OŚWIADCZENIE

Niniejszym oświadczam, iż mój wkład w publikację:

Wesołowski W, Czerwonka G, Zegadło K, Jagieła E, Łukasiak A, Grabowski Ł, Mantej J, Dziuba A, Bloch S, Topka-Bielecka G, Necel A, Kostovski D, Węglińska E, Narajczyk M, Węgrzyn G, Nejman-Faleńczyk B, Adamus-Białek W. Significant variability in virulence factors and antibiotic resistance of uropathogenic *Escherichia coli* strains may be counteracted by their susceptibility to lytic bacteriophages. Sci Rep. 2026 Jul 21. doi: 10.1038/s41598-026-62793-9.

polegał na:

Udziale w zbieraniu i przetwarzaniu danych dotyczących profagów oraz współudziale w przygotowaniu sekcji manuskryptu poświęconej profagom *E. coli*.

.....
PODPIS

Aleksandra Łukasiak
.....
IMIĘ I NAZWISKO

.... 27.08.2026 ....
DATA

Katedra Biologii Molekularnej,
Wydział Biologii,
Uniwersytet Gdański
.....
AFILIACJA

OŚWIADCZENIE

Niniejszym oświadczam, iż mój wkład w publikację:

Wesołowski W, Czerwonka G, Zegadło K, Jagieła E, Łukasiak A, Grabowski Ł, Mantej J, Dziuba A, Bloch S, Topka-Bielecka G, Necel A, Kostovski D, Węglińska E, Narajczyk M, Węgrzyn G, Nejman-Faleńczyk B, Adamus-Białek W. Significant variability in virulence factors and antibiotic resistance of uropathogenic *Escherichia coli* strains may be counteracted by their susceptibility to lytic bacteriophages. Sci Rep. 2026 Jul 21. doi: 10.1038/s41598-026-62793-9.

polegał na

(i) współudziale w określeniu spektrum gospodarza dla bakteriofagów;

(ii) współudziale w określeniu efektu cytotoksycznego wywoływanego przez lizaty fagowe na komórki eukariotyczne.

.....
PODPIS

Łukasz Grabowski
.....
IMIĘ I NAZWISKO

08.09.26
.....
DATA

Katedra Biologii Molekularnej,
Wydział Biologii,
Uniwersytet Gdański
.....
AFILIACJA

OŚWIADCZENIE

Niniejszym oświadczam, iż mój wkład w publikację:

Wesołowski W, Czerwonka G, Zegadło K, Jagieła E, Łukasiak A, Grabowski Ł, Mantej J, Dziuba A, Bloch S, Topka-Bielecka G, Necel A, Kostovski D, Węglińska E, Narajczyk M, Węgrzyn G, Nejman-Faleńczyk B, Adamus-Białek W. Significant variability in virulence factors and antibiotic resistance of uropathogenic *Escherichia coli* strains may be counteracted by their susceptibility to lytic bacteriophages. Sci Rep. 2026 Jul 21. doi: 10.1038/s41598-026-62793-9,

polegał na współudziale w przeprowadzeniu doświadczenia mającego na celu zbadania poziomu toksyczności lizatów fagowych wobec komórek eukariotycznych.

Ł. Grabowski
.....
PODPIS

Jagoda Mantej
.....
IMIĘ I NAZWISKO

27.08.2026
.....
DATA

Katedra Urologii i Andrologii,
Wydział Lekarski,
Collegium Medicum im. Ludwika Rydygiera w Bydgoszczy,
Uniwersytet Mikołaja Kopernika w Toruniu
.....
AFILIACJA

OŚWIADCZENIE

Niniejszym oświadczam, iż mój wkład w publikację:

Wesołowski W, Czerwonka G, Zegadło K, Jagieła E, Łukasiak A, Grabowski Ł, Mantej J, Dziuba A, Bloch S, Topka-Bielecka G, Necel A, Kostovski D, Węglińska E, Narajczyk M, Węgrzyn G, Nejman-Faleńczyk B, Adamus-Białek W. Significant variability in virulence factors and antibiotic resistance of uropathogenic *Escherichia coli* strains may be counteracted by their susceptibility to lytic bacteriophages. Sci Rep. 2026 Jul 21. doi: 10.1038/s41598-026-62793-9.

polegał na współudziale w analizie cytotoksyczności badanych bakteriofagów wobec komórek eukariotycznych.

Jagoda Mantej
.....
PODPIS

Anna Dziuba
.....
IMIĘ I NAZWISKO

24.08.2026
.....
DATA

Zakład Genetyki Medycznej i Diagnostyki Laboratoryjnej,
Wydział Lekarski,
Uniwersytet Jana Kochanowskiego w Kielcach

.....
AFILIACJA

OŚWIADCZENIE

Niniejszym oświadczam, iż mój wkład w publikację:

Wesołowski W, Czerwonka G, Zegadło K, Jagieła E, Łukasiak A, Grabowski Ł, Mantej J, Dziuba A, Bloch S, Topka-Bielecka G, Necel A, Kostovski D, Węglińska E, Narajczyk M, Węgrzyn G, Nejman-Faleńczyk B, Adamus-Białek W. Significant variability in virulence factors and antibiotic resistance of uropathogenic *Escherichia coli* strains may be counteracted by their susceptibility to lytic bacteriophages. Sci Rep. 2026 Jul 21. doi: 10.1038/s41598-026-62793-9.

polegał na: charakterystyce izolatów bakteryjnych wyselekcjonowanych do badań, gromadzeniu danych wykorzystanych do publikacji.

Anna Dziuba
.....
PODPIS

Sylwia Bloch
.....
IMIĘ I NAZWISKO

Katedra Biologii Molekularnej,
Wydział Biologii,
Uniwersytet Gdański
.....
AFILIACJA

08.09.2026 r.
.....
DATA

OŚWIADCZENIE

Niniejszym oświadczam, iż mój wkład w publikację:

Wesołowski W, Czerwonka G, Zegadło K, Jagieła E, Łukasiak A, Grabowski Ł, Mantej J, Dziuba A, Bloch S, Topka-Bielecka G, Necel A, Kostovski D, Węglińska E, Narajczyk M, Węgrzyn G, Nejman-Faleńczyk B, Adamus-Białek W. Significant variability in virulence factors and antibiotic resistance of uropathogenic *Escherichia coli* strains may be counteracted by their susceptibility to lytic bacteriophages. Sci Rep. 2026 Jul 21. doi: 10.1038/s41598-026-62793-9.

polegał na współudziale w:

(i) przygotowaniu i oczyszczaniu lizatów fagowych;

(ii) określeniu poziomu endotoksyn w oczyszczonych lizatach fagowych;

(iii) redagowaniu ostatecznej wersji manuskryptu.

.....
PODPIS

Gracja Topka-Bielecka
.....
IMIĘ I NAZWISKO

30.08.2026 r.
DATA

Zakład Toksykologii Środowiska
Wydział Nauk o Zdrowiu z Instytutem Medycyny Morskiej i Tropikalnej
Gdański Uniwersytet Medyczny
.....
AFILIACJA

OŚWIADCZENIE

Niniejszym oświadczam, iż mój wkład w publikację:

Wesołowski W, Czerwonka G, Zegadło K, Jagieła E, Łukasiak A, Grabowski Ł, Mantej J, Dziuba A, Bloch S, Topka-Bielecka G, Necel A, Kostovski D, Węglińska E, Narajczyk M, Węgrzyn G, Nejman-Faleńczyk B, Adamus-Białek W. Significant variability in virulence factors and antibiotic resistance of uropathogenic *Escherichia coli* strains may be counteracted by their susceptibility to lytic bacteriophages. Sci Rep. 2026 Jul 21. doi: 10.1038/s41598-026-62793-9,

polegał na współudziale w izolacji bakteriofagów z próbek ścieków miejskich z wykorzystaniem wybranych szczepów bakteryjnych.

Gracja Topka-Bielecka

PODPIS

Agnieszka Necel
.....
IMIĘ I NAZWISKO

31.08.2026
.....
DATA

Katedra i Zakład Mikrobiologii,
Wydział Lekarski,
Gdański Uniwersytet Medyczny
.....
AFILIACJA

OŚWIADCZENIE

Niniejszym oświadczam, iż mój wkład w publikację:

Wesołowski W, Czerwonka G, Zegadło K, Jagieła E, Łukasiak A, Grabowski Ł, Mantej J, Dziuba A, Bloch S, Topka-Bielecka G, Necel A, Kostovski D, Węglińska E, Narajczyk M, Węgrzyn G, Nejman-Faleńczyk B, Adamus-Białek W. Significant variability in virulence factors and antibiotic resistance of uropathogenic *Escherichia coli* strains may be counteracted by their susceptibility to lytic bacteriophages. Sci Rep. 2026 Jul 21. doi: 10.1038/s41598-026-62793-9.

polegał na współudziale w izolacji bakteriofagów z próbek ścieków miejskich z wykorzystaniem wybranych szczepów bakteryjnych.

Agnieszka Necel
.....
PODPIS

Daniel Kostovski
.....
IMIĘ I NAZWISKO

24.08.2026
DATA

Katedra Biologii Molekularnej,
Wydział Biologii,
Uniwersytet Gdański
.....
AFILIACJA

OŚWIADCZENIE

Niniejszym oświadczam, iż mój wkład w publikację:

Wesołowski W, Czerwonka G, Zegadło K, Jagieła E, Łukasiak A, Grabowski Ł, Mantej J, Dziuba A, Bloch S, Topka-Bielecka G, Necel A, Kostovski D, Węglińska E, Narajczyk M, Węgrzyn G, Nejman-Faleńczyk B, Adamus-Białek W. Significant variability in virulence factors and antibiotic resistance of uropathogenic *Escherichia coli* strains may be counteracted by their susceptibility to lytic bacteriophages. Sci Rep. 2026 Jul 21. doi: 10.1038/s41598-026-62793-9.

polegał na współudziale w przygotowaniu, zagęszczaniu i oczyszczaniu lizatów fagowych.

Kostovski Daniel
.....
PODPIS

Emilia Piwnicka (Węglińska)
.....
IMIĘ I NAZWISKO

01.09.2026
.....
DATA

Katedra Biologii Molekularnej,
Wydział Biologii,
Uniwersytet Gdański
.....
AFILIACJA

OŚWIADCZENIE

Niniejszym oświadczam, iż mój wkład w publikację:

Wesołowski W, Czerwonka G, Zegadło K, Jagieła E, Łukasiak A, Grabowski Ł, Mantej J, Dziuba A, Bloch S, Topka-Bielecka G, Necel A, Kostovski D, Węglińska E, Narajczyk M, Węgrzyn G, Nejman-Faleńczyk B, Adamus-Białek W. Significant variability in virulence factors and antibiotic resistance of uropathogenic *Escherichia coli* strains may be counteracted by their susceptibility to lytic bacteriophages. Sci Rep. 2026 Jul 21. doi: 10.1038/s41598-026-62793-9,

polegał na współudziale w zgromadzeniu i opracowaniu danych z analizy wydajności infekcji.

Emilia Piwnicka
.....
PODPIS

Magdalena Narajczyk
.....
IMIĘ I NAZWISKO

Laboratorium Bioobrazowania,
Wydział Biologii,
Uniwersytet Gdański
.....
AFILIACJA

26-08-2026.
DATA

OŚWIADCZENIE

Niniejszym oświadczam, iż mój wkład w publikację:

Wesołowski W, Czerwonka G, Zegadło K, Jagieła E, Łukasiak A, Grabowski Ł, Mantej J, Dziuba A, Bloch S, Topka-Bielecka G, Necel A, Kostovski D, Węglińska E, Narajczyk M, Węgrzyn G, Nejman-Faleńczyk B, Adamus-Białek W. Significant variability in virulence factors and antibiotic resistance of uropathogenic *Escherichia coli* strains may be counteracted by their susceptibility to lytic bacteriophages. Sci Rep. 2026 Jul 21. doi: 10.1038/s41598-026-62793-9,

polegał na przygotowaniu próbek i wykonaniu zdjęć wirionów bakteriofagów z użyciem transmisyjnego mikroskopu elektronowego.

Magdalena Narajczyk
.....
PODPIS

Grzegorz Węgrzyn
.....
IMIĘ I NAZWISKO

Katedra Biologii Molekularnej,
Wydział Biologii,
Uniwersytet Gdański
.....
AFILIACJA

27 sierpnia 2026 r.
.....
DATA

OŚWIADCZENIE

Niniejszym oświadczam, iż mój wkład w publikację:

Wesołowski W, Czerwonka G, Zegadło K, Jagieła E, Łukasiak A, Grabowski Ł, Mantej J, Dziuba A, Bloch S, Topka-Bielecka G, Necel A, Kostovski D, Węglińska E, Narajczyk M, Węgrzyn G, Nejman-Faleńczyk B, Adamus-Białek W. Significant variability in virulence factors and antibiotic resistance of uropathogenic *Escherichia coli* strains may be counteracted by their susceptibility to lytic bacteriophages. Sci Rep. 2026 Jul 21. doi: 10.1038/s41598-026-62793-9.

polegał na konsultacjach merytorycznych przy interpretacji wyników eksperymentów oraz pomocy w redagowaniu tekstu manuskryptu.

KIEROWNIK
KATEDRY BIOLOGII MOLEKULARNEJ
prof. dr hab. Grzegorz Węgrzyn

.....
PODPIS

Bożena Nejman-Faleńczyk
.....
IMIĘ I NAZWISKO

09.09.2026 r.
.....
DATA

Katedra Biologii Molekularnej,
Wydział Biologii,
Uniwersytet Gdański
.....
AFILIACJA

OŚWIADCZENIE

Niniejszym oświadczam, iż mój wkład w publikację:

Wesołowski W, Czerwonka G, Zegadło K, Jagieła E, Łukasiak A, Grabowski Ł, Mantej J, Dziuba A, Bloch S, Topka-Bielecka G, Necel A, Kostovski D, Węglińska E, Narajczyk M, Węgrzyn G, Nejman-Faleńczyk B, Adamus-Białek W. Significant variability in virulence factors and antibiotic resistance of uropathogenic *Escherichia coli* strains may be counteracted by their susceptibility to lytic bacteriophages. Sci Rep. 2026 Jul 21. doi: 10.1038/s41598-026-62793-9,

polegał na:

- opracowaniu i konsultowaniu koncepcji badań oraz założeń naukowych niniejszej publikacji,
- sprawowaniu merytorycznego nadzoru nad realizacją prac związanych z charakterystyką bakteriofagów,
- uczestnictwie w doborze metod badawczych, interpretacji uzyskanych wyników oraz formułowaniu wniosków dotyczących części pracy opisującej bakteriofagi,
- koordynowaniu prac związanych z przygotowaniem manuskryptu i dbaniu o jego spójność merytoryczną i redakcyjną,
- koordynowaniu współpracy pomiędzy współautorami i określeniu zakresu ich pracy,
- przygotowaniu oraz edycji pierwotnej i ostatecznej wersji manuskryptu,
- komunikacji z redakcją czasopisma, w tym prowadzeniu korespondencji na etapie zgłoszenia, recenzowania i publikacji manuskryptu,
- koordynowaniu prac związanych z przygotowaniem odpowiedzi na uwagi recenzentów oraz wprowadzeniem wymaganych zmian do manuskryptu.

B. Nejman-Faleńczyk
.....
PODPIS

Wioletta Adamus-Białek
.....
IMIĘ I NAZWISKO

23.08.2026
.....
DATA

Zakład Genetyki Medycznej i Diagnostyki Laboratoryjnej,
Wydział Lekarski,
Uniwersytet Jana Kochanowskiego w Kielcach

.....
AFILIACJA

OŚWIADCZENIE

Niniejszym oświadczam, iż mój wkład w publikację:

Wesołowski W, Czerwonka G, Zegadło K, Jagieła E, Łukasiak A, Grabowski Ł, Mantej J, Dziuba A, Bloch S, Topka-Bielecka G, Necel A, Kostovski D, Węglińska E, Narajczyk M, Węgrzyn G, Nejman-Faleńczyk B, Adamus-Białek W. Significant variability in virulence factors and antibiotic resistance of uropathogenic *Escherichia coli* strains may be counteracted by their susceptibility to lytic bacteriophages. Sci Rep. 2026 Jul 21. doi: 10.1038/s41598-026-62793-9.

polegał na udziale w opracowaniu koncepcji pracy i pozyskaniu finansowania badań związanego z sekwencjonowaniem, selekcji i dostarczeniu materiału do badań (kolekcja izolatów bakteryjnych) wraz z ich charakterystyką. Byłam zaangażowana w przeprowadzenie analiz i opracowanie wyników badań części dotyczącej bakterii, jak również zarządzanie projektem i nadzór nad nim, Brałam udział w opracowaniu wersji pierwotnej manuskryptu, a następnie jego redakcji i korekcie.

Signed by / Podpisano przez:
Wioletta Adamus-Białek
Date / Data: 2026-08-23 20:41

.....
PODPIS

KIR Szafir | elektronicznypodpis.pl

Raport z weryfikacji podpisu

INFORMACJE O DOKUMENCIE:	
Nazwa pliku	WAB praca nr 2-sig.pdf
Data weryfikacji podpisów	2026-08-24 15:54:05

SZCZEGÓŁY WERYFIKACJI:

Wioletta Adamus-Białek
.....
IMIĘ I NAZWISKO

23.08.2026
.....
DATA

Zakład Genetyki Medycznej i Diagnostyki Laboratoryjnej.
Wydział Lekarski.
Uniwersytet Jana Kochanowskiego w Kielcach

.....
AFILIACJA

OŚWIADCZENIE

Niniejszym oświadczam, iż mój wkład w publikację:

Wesołowski W. Czerwonka G, Zegadło K, Jagiełła E, Łukasiak A, Grabowski Ł, Mantej J, Dziuba A, Bloch S, Topka-Bielecka G, Necel A, Kostovski D, Węglińska E, Narajczyk M, Węgrzyn G, Nejman-Faleńczyk B, Adamus-Białek W. Significant variability in virulence factors and antibiotic resistance of uropathogenic *Escherichia coli* strains may be counteracted by their susceptibility to lytic bacteriophages. Sci Rep. 2026 Jul 21. doi: 10.1038/s41598-026-62793-9.

polegał na udziale w opracowaniu koncepcji pracy i pozyskaniu finansowania badań związanego z sekwencjonowaniem, selekcji i dostarczeniu materiału do badań (kolekcja izolatów bakteryjnych) wraz z ich charakterystyką. Byłam zaangażowana w przeprowadzenie analiz i opracowanie wyników badań części dotyczącej bakterii, jak również zarządzanie projektem i nadzór nad nim. Brałam udział w opracowaniu wersji pierwotnej manuskryptu, a następnie jego redakcji i korekcie.

Signed by / Podpisano przez: Wioletta Adamus-Białek
Date / Data: 2026-08-23 20:41

.....
PODPIS

Podpis: Pozytywny	
Rodzaj certyfikatu	Kwalifikowany
Format podpisu	PAdES-BES
Data złożenia podpisu	2026-08-23 20:41:18
Podpis zawiera znacznik czasu	Brak
Certyfikat podpisującego / składającego pieczęć	Nazwa powszechna: Wioletta Adamus-Białek Kraj: PL
Numer seryjny certyfikatu	183280257084507262196234348068648645422160390452
Wystawca certyfikatu	Nazwa powszechna: COPE SZAFIR - Kwalifikowany Nazwa organizacji: Krajowa Izba Rozliczeniowa S.A. Kraj: PL Identyfikator organizacji: VATPL-5260300517
Lista CRL wykorzystana do weryfikacji certyfikatu	-
Odpowiedź OCSP wykorzystana do weryfikacji certyfikatu	Status: GOOD ArchiveCutOff: 13 Mar 2017 14:07:47 GMT CN=COPE SZAFIR OCSP Responder OZK62 O=Krajowa Izba Rozliczeniowa S.A. L=Warszawa ST=mazowieckie C=PL - 24 Aug 2026 13:53:54 GMT
Zweryfikowano na dzień	2026-08-24 15:53:54 (Bieżący czas systemowy)
Status weryfikacji	**Pozytywny**
	Kwalifikowany podpis poprawnie zweryfikowany certyfikatem kwalifikowanym
Uwagi	Weryfikowany podpis został uznany za kwalifikowany. Zgodnie z Art. 25. Punkt 2 rozporządzenia Parlamentu Europejskiego i Rady (UE) nr 910/2014 z 23 lipca 2014 r. kwalifikowany podpis elektroniczny ma skutek prawny równoważny podpisowi własnoręcznemu.

Praca nr 3

Antibacterial activity of a newly identified phage-derived endolysin and its parental bacteriophage against clinical uropathogenic *Escherichia coli*.

Wesołowski W, Bloch S, Czerwonka G, Jagieła E, Grabowski Ł, Jeschke M, Jacewicz J, Piwnicka E, Morcinek-Orłowska J, Loika H, Czaplewska P, Węgrzyn G, Adamus-Białek W, Łukasiak A, Nejman-Faleńczyk B.

bioRxiv. doi: 10.64898/2026.09.09.750462.

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Antibacterial activity of a newly identified phage-derived endolysin and its parental bacteriophage against clinical uropathogenic *Escherichia coli*

Wojciech Wesołowski[1], Sylwia Bloch[1], Grzegorz Czerwonka[2], Ernest Jagieła[2], Łukasz Grabowski[1], Marta Jeschke[1], Jakub Jacewicz[1], Emilia Piwnicka[1], Joanna Morcinek-Orłowska[3], Hanna Loika[4], Paulina Czaplewska[5], Grzegorz Węgrzyn[1], Wioletta Adamus-Białek[6], Aleksandra Łukasiak[1], Bożena Nejman-Faleńczyk[1*]

[1]Department of Molecular Biology, Faculty of Biology, University of Gdańsk, Wita Stwosza 59, 80-308 Gdańsk, Poland

[2]Division of Microbiology, Institute of Biology, Jan Kochanowski University, Uniwersytecka 7, 25-406 Kielce, Poland

[3]Structural Biology Laboratory, Intercollegiate Faculty of Biotechnology, University of Gdańsk and Medical University of Gdańsk, Abrahama 58, 80-307 Gdańsk, Poland

[4]Department of Medical Microbiology, Faculty of Medicine, Medical University of Gdańsk, Dębowa 25, 80-204 Gdańsk, Poland

[5]Laboratory of Mass Spectrometry-Core Facility Laboratories, Intercollegiate Faculty of Biotechnology UG&MUG, University of Gdańsk, Abrahama 58, 80-307 Gdańsk, Poland

[6]Department of Medical Genetics and Laboratory Diagnostics, Faculty of Medicine, Jan Kochanowski University, IX Wieków Kielc 19a, 25-317 Kielce, Poland

*corresponding author: bozena.nejman-falenczyk@ug.edu.pl

Abstract

The increasing prevalence of antibiotic-resistant uropathogenic *Escherichia coli* highlights the need for antibacterial strategies that can complement or extend beyond conventional antibiotic treatment. Bacteriophages and phage-derived lytic enzymes represent promising alternatives because of their distinct mechanisms of bacterial killing and their potential activity against drug-resistant pathogens. In this study, we characterized the newly discovered UPEC-infecting bacteriophage vB-EcoS_57-3 and the endolysin

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

57_3Lys, encoded by this phage, combining genomic, structural, and functional approaches. The phage demonstrated lytic activity against clinical UPEC isolates and retained antibacterial potential under conditions relevant to the urinary tract. Genomic and sequence analyses revealed distinctive features of 57_3Lys associated with signal-anchor-release endolysins and suggested a less common mode of intracellular translocation and activation. Functional experiments supported the involvement of the bacterial secretion machinery in endolysin-mediated lysis. Notably, the purified enzyme also displayed antibacterial activity against intact clinical *E. coli* cells, despite the intrinsic barrier presented by the Gram-negative cell envelope. Although this activity developed slowly, it significantly reduced both bacterial culture turbidity and viable cell number. Together, these findings provide new insights into the antimicrobial strategies based on the phages and their phage lytic enzymes, supporting further exploration of vB-EcoS_57-3 and 57_3Lys as potential tools against antibiotic-resistant UPEC.

Keywords: uropathogenic *Escherichia coli* (UPEC), urinary tract infections (UTIs), bacteriophages, SAR-endolysins, phage therapy, antibacterial activity, antibiotic resistance

Highlights

- **vB-EcoS_57-3 reveals lytic activity against multiple clinical UPEC isolates.**
- **The phage retains antibacterial activity under artificial urinary conditions.**
- **Endolysin 57_3Lys exhibits distinctive SAR-associated features and activity mechanism.**
- **Purified 57_3Lys lyses intact UPEC cells in the absence of a permeabilizer.**

1. Introduction

Urinary tract infections (UTIs) are among the most common bacterial infections worldwide and constitute the second most frequent group of infections after those affecting the respiratory tract. They represent a substantial healthcare burden, affecting approximately 150–250 million people annually and generating considerable medical and socioeconomic costs. Although UTIs occur in both sexes, women are disproportionately

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

affected because of anatomical and physiological predispositions (Salman et al., 2026). Gram-negative bacteria are responsible for the majority of UTIs, with members of the order Enterobacterales, particularly *E. coli* and *Klebsiella pneumoniae*, accounting for most cases (Flores-Mireles et al., 2015; Cito et al., 2026).

Uropathogenic *E. coli* (UPEC) is the predominant etiological agent of UTIs and is responsible for approximately 85% of uncomplicated infections (Terlizzi et al., 2017). The pathogenicity of UPEC is determined by a diverse repertoire of virulence determinants, including adhesins, toxins such as α-hemolysin, biofilm-forming ability, and other factors that facilitate colonization, persistence, and dissemination within the urinary tract (Terlizzi et al., 2017; Fuad et al., 2026). The increasing prevalence of multidrug-resistant UPEC strains has substantially reduced the effectiveness of conventional antibiotic therapy (Khatun et al., 2026). Furthermore, recurrent infections requiring repeated antibiotic administration promote the accumulation and dissemination of antimicrobial-resistance determinants within the urogenital microbiome (Groleau et al., 2026). Consequently, there is an urgent need to develop alternative antibacterial strategies.

Among the most promising alternatives are bacteriophages, viruses that specifically infect bacteria (Łukasiak et al., 2026; Joshi et al., 2026). Phages are the most abundant biological entities on Earth, with an estimated global population of approximately 10^{31} particles, and represent an enormous reservoir of genetic diversity (Hatfull et al., 2015). Their narrow host specificity minimizes disruption of the commensal microbiota, while their ability to replicate at the site of infection enables self-amplifying antibacterial activity (Łukasiak et al., 2026). In parallel with the rapid increase in the number of newly described bacteriophages, both fundamental and translational phage research have expanded considerably over the past decade. Nevertheless, relatively few bacteriophages, particularly those infecting clinically relevant UPEC strains, have been comprehensively characterized, limiting both mechanistic understanding and therapeutic development (Hussein et al., 2025; Łukasiak et al., 2026; Joshi et al., 2026).

Besides whole phages, increasing attention has been directed toward phage-derived lytic enzymes, particularly endolysins, as next-generation antibacterial agents (Bai et al., 2020; Niaz et al., 2025; Wesołowski et al., 2025; Xia et al., 2026). During the lytic cycle, bacterial cell lysis is orchestrated by the coordinated activity of holins and endolysins.

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Holins permeabilize the cytoplasmic membrane, whereas endolysins hydrolyse glycosidic, peptide, or amide bonds within peptidoglycan, ultimately causing osmotic collapse and rapid bacterial lysis (Young, 2013; Young, 2014).

Currently, phage endolysins are classified into three major classes (Cernooka et al., 2025). Canonical endolysins, the most widespread class, accumulate in the cytoplasm during phage development and gain access to the peptidoglycan only after the formation of large membrane lesions by holins. A distinct group of double-stranded DNA bacteriophages encodes endolysins containing N-terminal signal-anchor-release (SAR) domains (Woźnica et al., 2015). Recently, a third major class of endolysins, termed C-terminal anchor (CTA) endolysins, has been proposed. CTA endolysins possess a C-terminal transmembrane anchor instead of an N-terminal SAR domain and appear to represent a distinct lysis strategy, although their activation mechanism remains to be fully elucidated (Cernooka et al., 2025).

Unlike two other classes, SAR endolysins are exported into the periplasm via the Sec secretion pathway independently of holin-mediated transport. Their SAR domain remains anchored in the inner membrane because it is not cleaved by signal peptidase, whereas the catalytic domain is already localized in the periplasm in an inactive conformation that prevents premature host cell lysis. Subsequent membrane depolarization, typically mediated by pinholins, releases the SAR domain from the membrane, induces a conformational rearrangement of the enzyme, thereby activating its lytic function (Xu et al., 2005; Woźnica et al., 2015). Structural diversity within both the catalytic and SAR domains underlies distinct activation mechanisms, leading to the classification of SAR endolysins into three functional classes (Xu et al., 2004). Despite their unique activation processes and considerable therapeutic potential, only a limited number of SAR endolysins have been experimentally characterized, particularly those encoded by newly isolated bacteriophages infecting clinically important pathogens.

In the present study, we describe a novel UPEC-infecting bacteriophage, vB-EcoS_57-3, together with its SAR endolysin, 57_3Lys. We demonstrate the antibacterial activity of both the phage and its endolysin against clinical multidrug-resistant UPEC isolates. Comparative genomic and functional analyses identify vB-EcoS_57-3 as a newly identified representative of the genus *Veterinaerplatzvirus* within the family *Drexlerviridae* and

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

reveal that 57_3Lys is a class II SAR endolysin whose activity depends on Sec-mediated translocation. These findings expand the current knowledge of SAR endolysins and support their potential application as antibacterial agents against antibiotic-resistant UPEC.

2. Materials and methods

2.1. Bacterial strains

The *E. coli* strain EC57 and other clinical UPEC isolates: EC02, EC10, EC28, EC47, EC55, EC70, EC106, EC116, EC131, EC149, EC164, EC179, EC296, EC302, and EC353 used in this study, were isolated from patients with urinary tract infections (UTIs), as described previously by Dziuba et al. (2023), and stored at −80°C at the Department of Medical Genetics and Laboratory Diagnostics, Faculty of Medicine, Jan Kochanowski University in Kielce, Poland. Laboratory *E. coli* strains: MG1655, C600, MC1061, BL21 and DH5α were obtained from the strain collection of the Department of Molecular Biology, University of Gdańsk.

2.2. Cultivation of bacteria and DNA extraction

All bacterial strains were cultured at 37°C in liquid Luria-Bertani broth (LB; BioShop, Burlington, ON, Canada) or on solid LB medium supplemented with 1.5% agar (LA; BTL, Łódź, Poland). For genomic DNA isolation, bacterial cultures were grown in LB broth with aeration at 200 rpm and 37°C in a shaker incubator. Genomic DNA was extracted using the NucleoSpin Tissue spin column kit (Macherey-Nagel, Düren, Germany) according to the manufacturer's instructions. DNA concentration and purity were determined spectrophotometrically by measuring absorbance at 260 nm.

2.3. Whole-genome sequencing of EC57

The bacterial isolate EC57 was sequenced by a commercial sequencing service using two complementary sequencing approaches: short-read sequencing on the Illumina MiSeq platform and long-read sequencing on the Oxford Nanopore MinION platform to

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

improve genome assembly quality and completeness (Genomed, Warszawa, Poland). Genomic DNA concentration was measured using the PicoGreen reagent (Invitrogen™, Thermo Fisher Scientific, Waltham, Massachusetts, USA). DNA fragmentation was performed by sonication using a Covaris E210 instrument with parameters recommended for Illumina MiSeq library preparation. Libraries for Illumina sequencing were prepared using the NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® (New England Biolabs, Ipswich, Massachusetts, USA) according to the manufacturer's instructions. Sequencing was performed using paired-end technology (2 × 300 bp) with the MiSeq Reagent Kit v3 (Illumina, San Diego, California, USA). Libraries for Oxford Nanopore sequencing were prepared using the Rapid Barcoding Kit (SQK-RBK004; Oxford Nanopore Technologies, Oxford, UK) and sequenced on a SpotON Flow Cell Mk I (R9.4.1).

2.4. De novo assembly and annotation

Raw sequencing reads generated from both sequencing platforms were assembled using a hybrid assembly approach implemented in Unicycler v0.5.0 (Galaxy version 0.5.0+galaxy1) with default parameters. Contigs shorter than 500 bp were discarded. The assembled genome was initially annotated using the Bakta Web server (Schwengers et al., 2021; Beyvers et al., 2025). After submission to NCBI, the genome was reannotated using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) (Tatusova et al., 2016). The genome sequence of strain EC57 was deposited in GenBank under the whole-genome shotgun (WGS) accession number JAWHTW010000001.1.

2.5. Phylogenetic analysis of EC57

A single nucleotide polymorphism (SNP)-based phylogenetic analysis was performed using CSI Phylogeny to reconstruct the phylogenetic tree. The analysis was conducted using the default parameters: minimum depth at SNP positions of ×10, minimum relative depth of 10%, minimum distance between SNPs (prune) of 10 bp, minimum SNP quality score of 30, minimum read mapping quality of 25, minimum Z-score of 1.96, and without exclusion of heterozygous SNPs (Kaas et al., 2014). The

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

resulting Newick-format file was subsequently imported into iTOL for phylogenetic tree visualization and annotation (Letunic and Bork, 2021).

2.6. *Identification of plasmid replicons, CRISPR-Cas systems, antimicrobial resistance genes, and virulence-associated genes in the EC57 genome*

Putative plasmid replicons were identified using PlasmidFinder v3.0.3 (Carattoli et al., 2014). The assembled genome sequence was analyzed using the default parameters, with a minimum nucleotide identity threshold of 95% and a minimum coverage threshold of 60%. Identified replicon sequences were assigned to incompatibility (Inc) groups based on the PlasmidFinder database, enabling characterization of the plasmid content of the EC57 genome.

CRISPR arrays and associated *cas* genes were identified using the CRISPRCasFinder web server (Couvin et al., 2018). The assembled genome sequence was analyzed to identify CRISPR arrays and classify CRISPR-Cas systems based on repeat sequences, spacer content, and the presence of adjacent *cas* genes. CRISPR arrays were assigned confidence levels according to the CRISPRCasFinder scoring system, and CRISPR-Cas subtypes were determined based on the composition and organization of the *cas* operon.

Putative virulence-associated genes (VAGs) were identified using the Virulence Factor Database (VFDB) (Liu et al., 2019). Genes were classified according to their functional categories, including adhesion, invasion, iron acquisition, secretion systems, toxins, autotransporters, and immune evasion.

Antimicrobial resistance genes (ARGs) were identified using the Resistance Gene Identifier (RGI) v6.0.4 against the Comprehensive Antibiotic Resistance Database (CARD) v3.2.9 (Alcock et al., 2023). Protein sequences predicted from the EC57 genome were analyzed using the Perfect and Strict hit criteria with default parameters. Identified ARGs were classified according to their resistance mechanism, AMR gene family, and associated antibiotic class.

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

2.7. Sequencing, assembly and annotation of phage vB-EcoS_57-3

The genome of phage vB-EcoS_57-3 was sequenced by Genomed (Warszawa, Poland) using next-generation sequencing (NGS) on an Illumina MiSeq platform with paired-end technology (2 x 300 bp). A double-stranded DNA (dsDNA) sequencing library was prepared using the NEBNext® Ultra™ II DNA Library Prep Kit (New England Biolabs, Ipswich, Massachusetts, USA). Raw sequencing reads were trimmed using Cutadapt v3.0 (Martin, 2011), and read quality was assessed with FastQC v0.12.1 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) by Genomed (Warszawa, Poland). Genome assembly and annotation were performed according to the workflow described by Shen and Millard (2021). *De novo* assembly was carried out using SPAdes v3.15.5 with the "*only-assembler*" option (Bankevich et al., 2012), based on 188,971 sequencing reads. Assembly artifacts were removed using apc.pl (available at: https://github.com/jfass/apc). Viral genome termini, DNA packaging strategy, and genome reordering were predicted using PhageTermVirome v4.1 with Illumina short reads and default parameters (Garneau et al., 2021). The assembled genome was subsequently polished using Pilon v1.24 (Walker et al., 2014) with filtered and sorted reads generated by SAMtools (Danecek et al., 2021). The final assembly consisted of a single contig representing the complete genome of phage vB-EcoS_57-3, with an average sequencing coverage of 1,710x.

Genome annotation was performed using Pharokka v1.5.1 (Bouras et al., 2023). Coding DNA sequences (CDSs) were predicted with PHANOTATE (McNair et al., 2019), whereas tRNA genes were identified using tRNAscan-SE 2.0 (Chan et al., 2021). Functional annotation was generated by comparing predicted CDSs against the PHROGs (Terzian et al., 2021), VFDB (Chen et al., 2005), and CARD (Alcock et al., 2020) databases using MMseqs2 (Steinegger et al., 2017) and PyHMMER (Larralde et al., 2023). Genome visualization, as well as GC content and GC skew analyses, were performed using the Proksee platform (Grant et al., 2023). The phage lifestyle was predicted using PhaTYP (Shang et al., 2023), implemented in PHAbox v2.1.13. To identify antimicrobial resistance determinants and virulence-associated genes within the phage genome, the following tools were used: ABRicate v1.0.1 (Seemann, 2023), AMRFinderPlus v3.11 (Feldgarden et al., 2021), RGI v6.0.3 (Alcock et al., 2023), VirulenceFinder (Joensen et al., 2014), and

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

ResFinder (Bortolaia et al., 2020). Comparative genomic analysis of the related phages CEB_EC3a, DTL, IME542, 12210I, Rtp, IME253, and ACG-M12 was performed using Easyfig v2.2.5 (Sullivan et al., 2011). The assembled and annotated genome sequence of phage vB-EcoS_57-3 was deposited in the GenBank database under accession number PZ349761.

2.8. Phylogenetic analysis of vB-EcoS_57-3

To determine the phylogenetic position and taxonomic classification of bacteriophage vB-EcoS_57-3, PhaGCN implemented in PHABOX v2.1.13 (Shang et al., 2021) was initially used to assign the phage to a putative taxonomic family. Subsequently, 211 complete reference genomes representing the identified family were retrieved from the NCBI Virus database. Pairwise intergenomic similarities among all 212 phages, including vB-EcoS_57-3, were calculated using VIRIDIC (Moraru et al., 2020), and the resulting distance matrix is provided in **Supplementary Table S1**. Hierarchical clustering based on complete linkage was performed using the SciPy v1.16.3 library (Virtanen et al., 2020). Orthologous protein groups were identified using OMA v2.6.0 (Altenhoff et al., 2019). Based on the VIRIDIC results, 19 representative phages were selected for multigene phylogenetic analysis. Phages sharing at least 30% genome-wide similarity with vB-EcoS_57-3 were included, provided they did not belong to the same species-level cluster. When multiple phages formed a species-level cluster, a single representative genome was selected using the medoid-based approach. Phage F61 was selected as the outgroup because it represented the most divergent phage while still sharing the orthologous protein set required for multigene phylogenetic analysis, including the endolysin.

Multiple sequence alignments of orthologous protein groups identified by OMA were generated using Clustal Omega v1.2.3 (Sievers and Higgins, 2018) and subsequently trimmed with trimAl v1.4.rev15 (Capella-Gutiérrez et al., 2009) using the *automated1* option. A maximum-likelihood phylogenetic tree was reconstructed from 11 orthologous protein groups (**Supplementary Table S2**) using IQ-TREE v3.1.1 (Wong et al., 2026). The optimal substitution models were selected with ModelFinder Plus (Kalyaanamoorthy et al., 2017), and phylogenetic inference was performed using the edge-linked partition model (Chernomor et al., 2016), with branch support assessed by 1,000 ultrafast bootstrap

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

replicates (Hoang et al., 2018) and 1,000 SH-like approximate likelihood ratio tests (Guindon et al., 2010). The resulting phylogenetic tree was visualized using Dendroscope v3.8.10 (Huson and Scornavacca, 2012).

2.9. Isolation of the vB-EcoS_57-3 bacteriophage from urban sewage

Bacteriophage vB-EcoS_57-3 against *E. coli* EC57 was isolated from urban sewage from the Gdańsk Wastewater Treatment Plant in Poland following the protocol described by Jurczak-Kurek et al. (2016), with some modifications. Briefly, 10 mL of the sewage sample was mixed with 1 mL of the host strain overnight culture and cultivated for a few hours at 37°C with shaking. After cell lysis, the sample was centrifuged at 10,000 x *g* for 30 min at 4°C. The obtained supernatant was treated with 4% chloroform (Chempur, Piekary Śląskie, Poland) for 15 min, centrifuged (2,000 x *g*; 10 min; 4°C), and filtered through a 0.22 µm syringe filter with surfactant-free cellulose acetate membrane (SFCA; Thermo Fisher Scientific, Waltham, Massachusetts, USA). In the next step, 10-fold dilutions of the phage lysate were prepared in TM buffer (10 mM Tris-HCl, 10 mM $MgSO_4$; pH 7.2). Then, 50 µL of each dilution was mixed with 1 mL of overnight host bacterial culture and 2 mL of top agar. The mixtures were poured onto Petri dishes (Alchem, Toruń, Poland) containing 25 mL of LA medium. After overnight incubation at 37°C, a single phage plaque was transferred to a new flask with EC57 bacterial culture and propagated three times by the procedure described above to obtain purified lysate with vB-EcoS_57-3 particles.

2.10. Lysate titration procedure

The titer of the phage lysate was determined using the standard double-overlay method described by Sambrook and Russell (2001), with some modifications. In the first step, the suspension of phage particles was serially 10-fold diluted in TM buffer (10 mM Tris-HCl, 10 mM $MgSO_4$; pH 7.2). Then, an appropriate volume of each dilution of phage lysate was mixed with 2 mL of top agar and 1 mL of the overnight bacterial cell culture. The mixtures were poured onto the standard Petri dishes filled with 25 mL of LA. Plates were incubated overnight at 37°C. The phage titer was calculated on the basis of counted plaques and presented as the number of plaque-forming units per milliliter (PFU/mL).

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

2.11. Preparation of the vB-EcoS_57-3 lysate

E. coli EC57 was grown in LB medium at 37°C to the early log phase (OD_{600} = 0.2). Then, host bacterial cells were infected with vB-EcoS_57-3 phage at an m.o.i. of 0.05 and cultivated until lysis was observed. The phage suspension was shaken with 4% chloroform (Chempur, Piekary Śląskie, Poland) for 15 min at 37°C, centrifuged at 2,000 x *g* for 20 min at 4°C to separate the host cell debris, and passed through a 0.22 µm syringe filter with SFCA membrane (Thermo Fisher Scientific, Waltham, Massachusetts, USA). The lysate was titrated on double-layer agar plates and incubated overnight at 37°C. The phage titer was determined on the basis of counted plaques. The phage stock was stored at 4°C until subsequent experiments.

2.12. Electron microscopy

The lysate with vB-EcoS_57-3 particles was concentrated in the presence of 10% polyethylene glycol 8000 (PEG8000; BioShop, Burlington, ON, Canada) for 18 h at 4°C. After centrifugation (8,000 x *g*; 20 min; 4°C), the supernatant was discarded, and pelleted phage particles were suspended in TM buffer (10 mM Tris-HCl, 10 mM $MgSO_4$; pH 7.2). PEG8000 was removed from the sample using 4% chloroform. Concentrated lysate was then subjected to ultracentrifugation (20,000 rpm; 2 h; 10°C) in the cesium chloride gradient as described by Sambrook and Russell (2001). Electron microscopic analysis of vB-EcoS_57-3 virions was performed in the Bioimaging Laboratory, Faculty of Biology, University of Gdańsk, Poland, by using the negative staining with uranyl acetate method as previously described by Topka-Bielecka et al. (2020).

2.13. Plaque morphology

The plaque morphology of phage vB-EcoS_57-3 was examined on an overnight culture of the EC57 isolate. To obtain visible plaques, 10-fold dilutions of vB-EcoS_57-3 stock were prepared in TM buffer (10 mM Tris-HCl, 10 mM $MgSO_4$; pH 7.2). In the next step, 1 mL of the overnight EC57 culture was mixed with 2 mL of top agar and 15 µL of an appropriate dilution of lysate. The mixture was poured onto a Petri dish with 25 mL of LA medium and incubated overnight at 37°C. The next day, plaque morphology and diameter were

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

339 determined. Visualization of phage vB-EcoS_57-3 plaques was performed using UVITEC
340 Alliance Q9 Mini software (UVITEC Ltd, Cambridge, UK).
341
342 ***2.14. Efficiency of plating (EOP)***
343 Efficiency of plaque formation on 21 *E. coli* strains was determined according to
344 the protocol described by Mirzaei and Nilsson (2015), with some modifications. Briefly,
345 10-fold serial dilutions of phage stock were prepared in TM buffer (10 mM Tris-HCl, 10
346 mM $MgSO_4$; pH 7.2). Then, 15 µL of the appropriate dilution was mixed with 1 mL of an
347 overnight bacterial culture and 2 mL of top agar. The mixture was poured onto an LA agar
348 plate and incubated overnight at 37°C. Efficiency of plaque formation of vB-EcoS_57-3
349 phage was calculated by dividing the phage titer on the tested strain by the phage titer on
350 the EC57 host. The EOP values for each virus-bacteria combination were classified as
351 follows: (i) EOP ≥ 0.05 - high infection efficiency, (ii) 0.5 > EOP ≥ 0.1 - medium infection
352 efficiency (iii) 0.1 > EOP ≥ 0.001 - low infection efficiency, and (iv) EOP ≤ 0.001-
353 inefficient infection.
354
355 ***2.15. Adsorption assay***
356 A phage adsorption assay was performed following a previously described method
357 by Czajkowski et al. (2014), with slight modifications. Briefly, host bacteria were
358 cultivated to an OD_{600} of 0.2 at 37°C. Then, 3 mL of EC57 cells were mixed with a phage
359 suspension to reach an m.o.i. of 0.05 and incubated at 37°C without shaking. Aliquots from
360 the prepared mixture were collected at 0 (control), 20 s, 1 min, 2 min, 3 min, 4 min, 6 min,
361 8 min, and 10 min post-infection intervals and centrifuged at 12,000 x *g* for 1 min at 4°C
362 to pellet the host bacteria with adsorbed phages. In the next step, the obtained supernatants
363 were filtered with 0.22 µm syringe filters with SFCA membranes (Thermo Fisher
364 Scientific, Waltham, Massachusetts, USA) and tested for free, non-adsorbed virus particles
365 by using the standard double-overlay method. The kinetics of phage adsorption to the host
366 cell surface was determined based on the formula: percentage of adsorbed phages =
367 ((control titer – residual titer)/control titer) x 100. Additionally, to estimate the number of
368 bacterial cells during the adsorption process, 100 µL of samples were collected at the times
369 indicated above, and their serial 10-fold dilutions were prepared in 0.85% NaCl

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

(STANLAB, Lublin, Poland). Then, 40 µL of each dilution were spread onto LA plates. After overnight incubation at 37°C, the number of bacterial cells per mL (CFU/mL) was calculated.

2.16. One-step growth curve

To determine the latent period and the burst size of phage vB-EcoS_57-3, a one-step growth experiment was conducted according to the procedure described by Necel et al. (2020), with some modifications. Host bacteria were grown in LB medium (BioShop, Burlington, ON, Canada) to an OD_{600} of 0.2. Then, 10 mL of bacterial culture was harvested by centrifugation (2,000 x *g*; 10 min; 4°C) and resuspended in 1 mL of LB medium supplemented with 3 mM NaN_3 (Avantor Performance Materials Poland S.A., Gliwice, Poland) to synchronize host growth. After 5 min incubation of the sample at 37°C, the phage particles were added to bacteria to an m.o.i. of 0.05. The phages were allowed to adsorb to the surface of bacterial cells for 1 min at 37°C. To remove unadsorbed virions, the sample was centrifuged (2,000 x *g*; 10 min; 4°C), and the bacterial pellet was suspended in 1 mL of LB medium supplemented with 3 mM NaN_3 (Avantor Performance Materials Poland S.A., Gliwice, Poland). In the next step, 25 µL of the suspension was added to 25 mL of LB medium (time 0) and cultivated at 37°C with shaking. The number of infective centers was estimated from the sample taken 1 min after infection by mixing 2 µL of the suspension with 1 mL of the overnight EC57 culture and 2 mL of top agar. Subsequently, the mixture was poured onto an LA plate. At indicated times, all samples were cleared by centrifugation (10,000 x *g*; 2 min; 4°C) and titrated to calculate the value of PFU/mL. After overnight incubation at 37°C, burst size was presented as the ratio of phage titer to the number of infection centers.

2.17. Lysis profile

EC57 cells were grown in LB medium to an OD_{600} of 0.2 at 37°C with shaking. Then, phage stock was added to the bacterial host to an m.o.i. of 0.005. A bacterial culture prepared in the same way but without virus supplementation was used as the control. Bacterial growth was monitored by measuring the optical density (OD_{600}) every 10 min for up to 70 min. Additionally, the number of bacterial cells per mL (CFU/mL) and phage titer

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

(PFU/mL) were also determined. After overnight incubation at 37°C, bacterial colonies and plaques were counted.

2.18. Stability of phage particles in the presence of physical and chemical factors

To determine the effects of various physical and chemical agents on the stability of the phage particles, the procedures described earlier by Jurczak-Kurek et al. (2016) were used, with some modifications. Briefly, 100 µL of phage lysate was mixed with 900 µL of TM buffer (10 mM Tris-HCl, 10 mM $MgSO_4$; pH 7.2). Then, the stability of phage suspension was tested in the presence of the following external factors: temperatures (-20°C, 40°C, and 62°C), pH values (2, 4, and 10), organic solvents (chloroform, DMSO, and ethanol), and detergents (CTAB and sarkosyl). Additionally, the effects of osmotic shock on the survival of virus particles were also analyzed. At appropriate times, samples were collected and assayed for phage presence by the standard double-overlay method described above. The percentage of residual infective virions under different conditions was calculated by comparing the initial and post-incubation phage titer (PFU/mL). Phage particles stored in TM buffer at 4°C were used as a control.

2.19. Stability and activity of phage virions in artificial urine

Synthetic urine was prepared according to the protocol previously described by Torzewska and Różalski (2015) and consisted of the following components (g/300 mL): 0.195 g of $CaCl_2$ x $2H_2O$ (BioShop, Burlington, ON, Canada), 0.195 g of $MgCl_2$ x $6H_2O$ (Chempur, Piekary Śląskie, Poland), 1.38 g of NaCl (STANLAB, Lublin, Poland), 0.69 g of Na_2SO_4 (Thermo Fischer Scientific, Waltham, Massachusetts, USA), 0.195 of g sodium citrate (BioShop, Burlington, ON, Canada), 0.006 g of sodium oxalate (Merck KGaA, Darmstadt, Germany), 1.26 g of KH_2PO_4 (Avantor Performance Materials Poland S.A., Gliwice, Poland), 0.48 g of KCl (Chempur, Piekary Śląskie, Poland), 0.3 g of NH_4Cl (Chempur, Piekary Śląskie, Poland), 7.5 g of urea (Avantor Performance Materials Poland S.A., Gliwice, Poland), and 0.33 g of creatine (Thermo Fischer Scientific, Waltham, Massachusetts, USA). The pH of artificial urine was adjusted to values of 4.5, 5.8, and 6.5 by using HCl (Merck KGaA, Darmstadt, Germany) or NaOH (Merck KGaA, Darmstadt,

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Germany). Then, prepared urines were sterilized by passing through 0.22 µm syringe filters with SFCA membrane (Thermo Fisher Scientific, Waltham, Massachusetts, USA).

In the first part of the experiment, the effect of artificial urine with different pH values on the stability of phage particles was analyzed. Briefly, 100 µL of phage lysate was mixed with 900 uL of synthetic urine with a pH of 4.5, 5.8, or 6.5 and incubated at 37°C. Samples were withdrawn at the indicated times (0, 2, 4, 6, and 8 h), and the concentration of phage particles (PFU/mL) was determined by using double-layer agar plates.

In the second part of the experiment, the activity of phage virions against host bacteria during incubation in the artificial urine with a pH of 4.5 was tested. EC57 bacteria were grown in fresh LB medium to an OD_{600} of 0.2 at 37°C with shaking. In the next step, 3.6 mL of bacterial culture was centrifuged (2,000 x *g*; 10 min; 4°C), and the obtained pellet was suspended in 0.9 mL of the artificial urine with a pH of 4.5. Then, host bacteria were treated with 100 µL of phage lysate with a titer of 3.6 x 10^9 PFU/mL, corresponding to an m.o.i. of 0.5. Additionally, to assess the stability of bacterial cells in synthetic urine at pH 4.5, 100 µL of TM buffer (10 mM Tris-HCl, 10 mM $MgSO_4$; pH 7.2) was mixed with 900 µL of the EC57 culture. The mixture was incubated at 37°C for 24 h. Bacterial viability was determined by enumerating colony-forming units (CFU/mL) at 0, 2, 4, 6, 8, and 24 h. At each time point, 100 µL samples were collected, serially diluted 10-fold in 0.85% NaCl (STANLAB, Lublin, Poland), and 40 µL aliquots of each dilution were spread onto LA plates. Following overnight incubation at 37°C, colonies were counted, and the CFU/mL values were calculated.

2.20. Cytotoxicity assay

The safety of phage vB-EcoS_57-3 obtained after purification in the cesium chloride gradient was tested with the Thermo Scientific Pierce chromogenic endotoxin quant kit (Thermo Fischer Scientific, Waltham, Massachusetts, USA) and Neutral Red Uptake (NRU) cytotoxicity assay according to the procedure described earlier by Wesołowski et al. (2026), with some modifications. The human bladder cancer cell line (T24) and the human embryonic kidney line (HEK293) were used in the experiment. The NRU cytotoxicity assay was conducted according to the following protocol. T24 and HEK293 cells were grown in McCoy's 5A Medium (Avantor Performance Materials

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Poland S.A., Gliwice, Poland) or Dulbecco's Modified Eagle's Medium (Gibco, Thermo Fischer Scientific, Waltham, Massachusetts, USA) supplemented with 10% FBS (Thermo Fischer Scientific, Waltham, Massachusetts, USA), 1% penicillin (Thermo Fischer Scientific, Waltham, Massachusetts, USA), and 1% streptomycin (Thermo Fischer Scientific, Waltham, Massachusetts, USA), respectively. Mammalian cells were seeded into 96-well plates at a density of 10^4 per well and incubated in the presence of 5% CO_2 at 37°C for 24 h. Then, the culture medium was discarded from each well and cells were treated with 200 µL of a mixture composed of appropriate medium and vB-EcoS_57-3 phage solution with a final titer of 10^8, 10^9 or 10^{10} PFU/mL. In addition, 1% Triton X-100 (Merck KGaA, Darmstadt, Germany) and TM buffer (10 mM Tris-HCl, 10 mM $MgSO_4$; pH 7.2) were used as positive and negative controls, respectively. Following 24 h incubation (37°C and 5% CO_2), the mixture was removed, and mammalian cells were washed with 150 µL of 1x PBS (Thermo Fisher Scientific, Waltham, Massachusetts, USA). In the next step, 0.33% NR dye solution (Thermo Fisher Scientific, Waltham, Massachusetts, USA) was added to each well, and the plate was incubated in the dark at 37°C for 45 min. After removal of the NR dye solution, the tested cells were washed with 150 µL of 1x PBS (Thermo Fisher Scientific, Waltham, Massachusetts, USA) and treated with 100 µL of NR desorb solution containing 50% ethanol (Chempur, Piekary Śląskie, Poland) and 1% glacial acetic acid (Merck KGaA, Darmstadt, Germany) in water. The amount of released NR proportional to the amount of viable mammalian cells was measured as A_{540} in the Viroskan microplate reader (Thermo Fisher Scientific, Waltham, Massachusetts, USA). The cell viability was estimated as a percentage of the control values. According to the PN-EN ISO 10993-5:2009 standard, the tested agent is considered cytotoxic if the mammalian cells' viability is below 70%. Additionally, the stability of vB-EcoS_57-3 particles in the presence of tested cells and media was determined by using the standard double-overlay method and expressed as PFU/mL.

2.21. Cloning of the gene encoding the 57_3Lys endolysin

The 57_3Lys gene was cloned from genomic DNA of phage vB-EcoS_57-3 by PCR using Walk DNA polymerase (A&A Biotechnology, Gdańsk, Poland) and specific primers: F_57_3Lys (5' GCGCCATGGCTCATATGAAGCAGAAGCTGCTAATTA 3') and

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

493 R_57_3Lys (5' TTTAAGCTTCTCGAGACGCTGCAATCCACTTAAACA 3'). The PCR
494 primers were designed by Primer3web (version 4.1.0) and synthesized by Genomed.
495 (Warszawa, Poland). The PCR product was cloned into a commercially available
496 pET26b(+) expression vector (Merck KGaA, Darmstadt, Germany), which carries a C-
497 terminal His-Tag sequence. Following digestion with *Nde*I and *Xho*I restriction
498 endonucleases (Thermo Fisher Scientific, Waltham, Massachusetts, USA), the PCR
499 product was ligated with the *Nde*I-*Xho*I fragment of plasmid pET26b(+) bearing a
500 kanamycin-resistant gene, using the T4 DNA ligase (Thermo Fisher Scientific, Waltham,
501 Massachusetts, USA). The construct pET26b_57_3Lys was introduced into *E. coli*
502 MC1061 by chemical transformation and then verified by sequencing (Genomed,
503 Warszawa, Poland).
504
505 ***2.22. In silico endolysin analysis***
506 The Prot-Param module from the Biopython library was used to predict molecular
507 weight, instability index, isoelectric point (pI), and grand average of hydropathy (GRAVY)
508 score (Cock et al., 2009). The aliphatic index was calculated using a previously described
509 formula (Ikai, 1980). A domain search was performed using InterProScan version 5.72-
510 103 (Jones et al., 2014). Additionally, to get deeper insight into the protein structure and
511 its properties, DeepTMHMM (Hallgren et al., 2022), SignalP6.0 (Nielsen, 2025), CD-
512 Search (Wang et al., 2023), and CAMPR3 (Waghu et al., 2015) were applied.
513
514 ***2.23. Overproduction and purification of the 57_3Lys endolysin***
515 The construct pET26b_57_3Lys was transformed into the BL21-AI™ *E. coli* strain
516 (Thermo Fisher Scientific, Waltham, Massachusetts, USA). Bacteria carrying the
517 overexpression plasmid were cultivated at 37°C in 200 mL of LB medium supplemented
518 with 30 µg/mL kanamycin (BioShop, Burlington, ON, Canada) and 12.5 µg/mL
519 tetracycline (Merck KGaA, Darmstadt, Germany) to an OD_{600} of 0.8. At that point,
520 overproduction of the 57_3Lys protein was induced by adding 0.5 mM isopropyl-β-D-
521 thiogalactopyranoside (IPTG; A&A Biotechnology, Gdańsk, Poland) and 0.2% L-(+)-
522 arabinose (BioShop, Burlington, ON, Canada), and the incubation was continued for 1 h at
523 37°C until complete lysis of bacterial cells occurred. Then, Pierce Protease Inhibitor

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Tablets (Thermo Fisher Scientific, Waltham, Massachusetts, USA) were added to the lysate and the mixture was shaken for 30 min at 37°C. Subsequently, the lysate was clarified by centrifugation (6,000 x *g*; 10 min; 4°C) and filtered through a 0.45 µm vacuum filter with a polyethersulfone membrane (PES; Avantor Performance Materials Poland S.A., Gliwice, Poland). In the next step, the lysate was loaded onto a 5 mL HisTrap™FF crude column (Cytiva, Marlborough, Massachusetts, USA). The protein was eluted with elution buffer (50 mM Tris-HCl, pH 7.5; 300 mM NaCl, 5% glycerol, 0.2 mM TCEP, and 200 mM imidazole) and dialyzed overnight against buffer containing 50 mM Tris-HCl (pH 7.5; BioShop, Burlington, ON, Canada), 200 mM NaCl (STANLAB, Lublin, Poland), 40% glycerol (STANLAB, Lublin, Poland), and 0.2 mM TCEP (Merck KGaA, Darmstadt, Germany). The purity of the 57_3Lys endolysin was assessed with the use of 20% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), and the concentration was determined fluorometrically by the Qubit™ Protein Assay Kit (Thermo Fisher Scientific, Waltham, Massachusetts, USA). Afterwards, the identity of the purified protein was further verified by LC-MS/MS analysis of excised and in-gel digested protein bands, as described previously (Fiołka et al., 2021), using an M5 MicroLC system with a C18 column coupled to a ZenoTOF 7600 mass spectrometer (SCIEX). The purified protein was stored at −20°C until further analysis.

2.24. Sec machinery assay

To test whether Sec machinery is responsible for endolysin secretion, the host cell lysis kinetics by 57_3Lys was determined in the presence of various concentrations of NaN_3 according to the protocol described by Bai et al. (2020), with some modifications. Briefly, *E. coli* BL21-AI™ cells bearing plasmid pET26b_57_3Lys were grown in LB medium at 37°C to an OD_{600} of 0.8 with aeration. Then, overproduction of 57_3Lys from the pET26b_57_3Lys plasmid in host cells was induced by 0.5 mM IPTG (A&A Biotechnology, Gdańsk, Poland) and 0.2% L-(+)-arabinose (BioShop, Burlington, ON, Canada). Additionally, some cultures were treated with NaN_3 added to a final concentration of 1 mM, 5 mM, or 10 mM. Controls were grown without addition of induction agents and NaN_3 or in the presence of 10 mM NaN_3. Bacteria were incubated at 37°C with shaking,

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

554 and their density was monitored by OD_{600} measurement at the same time intervals, every
555 10 min for 180 min.
556
557
558
559 ***2.25. Activity assay of the 57_3Lys endolysin***
560 Turbidity and cell number reduction assays of EC47, EC57, and EC131 cultures by
561 purified 57_3Lys endolysin were performed at 37°C in a standard 96-well titration plate.
562 Briefly, tested bacteria were grown in LB medium (BioShop, Burlington, ON, Canada) at
563 37°C to an OD_{600} of 0.8. Then, 50 mL of bacterial culture was spun down (2,000 x *g*; 20
564 min; RT) and subsequently suspended in 5 mL of 50 mM Tris-HCl (pH 7.5). Bacteria were
565 shaken for 45 min at RT. To measure the protein activity, 200 µL of 57_3Lys endolysin at
566 a concentration of 60 ng/µL or 200 µL of the reaction buffer (negative control; 50 mM
567 Tris-HCl, pH 7.5; 200 mM NaCl; 40% glycerol, and 0.2 mM TCEP) were mixed with 100
568 µL of tested bacterial culture. Additionally, the antibacterial activity of the chicken egg
569 white lysozyme (Merck KGaA, Darmstadt, Germany) against host bacteria was
570 determined. The changes in OD_{600} were recorded after 24 h of incubation at 37°C by using
571 the Viroskan microplate reader (Thermo Fisher Scientific, Waltham, Massachusetts, USA).
572 During this experiment, the number of survivors per mL was also examined. To estimate
573 CFU/mL, 20 µL of bacterial samples were collected and 10-fold diluted in 0.85% NaCl
574 (STANLAB, Lublin, Poland). In the next step, 10 µL of each dilution was spotted onto a
575 LA plate. After overnight incubation at 37°C, colonies were counted, and the CFU/mL was
576 calculated.
577
578 ***2.26. Statistical analysis***
579 All experiments were carried out in at least three or even more biological replicates.
580 The means ± SD were used in the statistical analysis of the data and the graphics. A
581 comparison of two averages in experiments was performed using Student's *t*-test by using
582 Microsoft® Excel 365 with Real Statistics Resource Pack. The significant differences were
583 marked by asterisks as follows: $p < 0.01$ (**) or $p < 0.001$ (***).
584

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

3. Results

3.1. Overview of the uropathogenic clinical EC57 isolate, a primary host for phage vB-EcoS_57-3

3.1.1. EC57 genomic content and pathogenicity determinants

E. coli EC57 was part of the clinical *E. coli* strain collection previously analyzed by Dziuba et al. (2023). This strain was isolated from the urine of a 14-year-old boy and belongs to the B2 phylogenetic group.

According to Bakta annotation, the genome of *E. coli* EC57 consists of one complete circular chromosome and three plasmids with a total size of 5,111,555 bp and was assembled into a single large chromosomal contig. Genome annotation performed using Bakta identified a total of 5,108 genomic features, including 4,684 protein-coding sequences (CDSs) and 65 small open reading frames (sORFs). In addition, 88 tRNA genes, 22 rRNA genes, one tmRNA, 189 ncRNAs, and 52 ncRNA-associated regions were identified. The relatively high number of regulatory ncRNAs suggests complex post-transcriptional regulation and adaptive potential under changing environmental conditions. The genome also contained two predicted chromosomal origins of replication (*oriC*), as well as one plasmid replication origin (*oriV*) and one origin of transfer (*oriT*), confirming the mobilizable nature of at least one plasmid present in the strain. Overall, the genome architecture of *E. coli* EC57 reflects a genetically versatile strain with a well-conserved core genome supplemented by plasmid-associated functions that may contribute to environmental adaptation, survival, and acquisition of accessory traits, including antimicrobial resistance or virulence-associated determinants.

PlasmidFinder tool identified a single plasmid, designated pEC57-1, in the *E. coli* EC57 genome. The plasmid DNA consists of 69,890 bp, and pEC57-1 was classified as a member of the IncFII incompatibility group, showing 100% identity to the IncFII replication region. Comparative analysis revealed that the replication region of plasmid pEC57-1 is identical to that of the *E. coli* plasmid pC15-1a (ENA accession number AY458016), showing 258/261 bp coverage and 100% nucleotide identity according to PlasmidFinder. However, comparison of the complete plasmid sequence using BLAST

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

indicated that pEC57-1 is most closely related to plasmid p1_A28403 from *Enterobacter hormaechei* (accession number CP182042.1), with approximately 90% query coverage and 99.74% nucleotide identity. These results suggest that although pEC57-1 carries an IncFII replication region identical to that of pC15-1a, its overall sequence differs substantially from previously described plasmids. Even after lowering the detection thresholds to 60% minimum identity and 30% minimum coverage, no additional plasmid replicons were identified. However, further examination of the genome assembly revealed two additional circular sequences potentially representing plasmids, they contained no recognizable plasmid replicon.

CRISPRCasFinder identified eight CRISPR loci in the *E. coli* EC57 genome (**Table 1**). Six loci (EC57-1, EC57-2, EC57-3, EC57-6, EC57-7, and EC57-8) consisted of single-spacer CRISPR arrays lacking adjacent *cas* genes and were classified as low-confidence structures (evidence level 1). A complete, high-confidence Type I-F CRISPR-Cas system (evidence level 4) was identified between positions 3,151,428 and 3,162,042 bp. This locus comprised two CRISPR arrays (EC57-4 and EC57-5) containing 15 and 10 spacers, respectively, flanking a complete *cas* operon encoding *cas6*, *csy3*, *csy2*, *csy1*, *cas3-cas2*, and *cas1*. Both arrays shared the repeat consensus sequence TTTCTAAGCTGCCTGTACGGCAGTGAAC and were oriented in the reverse direction (**Table 1**).

CARD analysis identified two acquired antimicrobial resistance genes, *blaTEM-1* and *tet(A)*, encoding a TEM-1 β-lactamase and a tetracycline efflux pump, respectively. In addition, several chromosomally encoded antimicrobial resistance determinants were detected, including multidrug efflux systems (*acrAB-tolC, acrDEF, mdtABC-EFGHMNOP*, and *emrABKYR*) and regulatory systems associated with antimicrobial resistance (*evgAS*, *baeRS*, *marA*, *soxRS*, and *crp*). The genome also contained genes implicated in resistance to peptide antibiotics (*arnT*, *pmrF*, *ugd*, and *eptA*) and β-lactam resistance (*ampC*). Furthermore, CARD identified resistance-associated protein variants, including PBP3 encoded by *ftsI* and GlpT encoded by *glpT*.

VFDB analysis identified a broad repertoire of virulence-associated genes (VAGs) in *E. coli* EC57. The genome contained multiple adhesin-associated genes, including the complete *E. coli* common pilus (*ecp*) and type 1 fimbrial (*fim*) operons, as well as genes

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

associated with F1C (*foc*), P (*pap*), and S (*sfa*) fimbriae, hemorrhagic *E. coli* pilus (*hcp*), the EaeH adhesin (*eaeH*), and UpaG (*upaG*). Several genes encoding autotransporter proteins were also detected, including *agn43*, *cah*, *cdiA*, *ehaB*, *pic*, *upaH*, and *vat*.

The genome also contained invasion-associated genes (*ibeABC* and *tia*) and genes involved in multiple iron acquisition systems, including the *chu* heme uptake system, the salmochelin *iro* operon, and the yersiniabactin-associated genes *fyuA*, *irp1*, *irp2*, and *ybtAEPQSTUX*. Furthermore, genes encoding several toxins were identified, including α-hemolysin (*hlyABCD*), cytotoxic necrotizing factor 1 (*cnf1*), uropathogenic-specific protein (*usp*), and hemolysin E/cytolysin A (*hlyE/clyA*). Genes encoding components of type VI secretion systems (T6SSs), including the ACE-T6SS and SCI-I T6SS, were also detected, together with *espL1*, which encodes a non-LEE-encoded type III secretion system effector.

3.1.2. EC57 phylogeny

SNP-based phylogenetic analysis demonstrated that *E. coli* EC57 clustered within the uropathogenic *E. coli* (UPEC) lineage and was clearly separated from the reference strain *E. coli* K-12 MG1655 (**Figure 1**). The K-12 reference strain grouped together with strains VR50, EcPF16, CI5, and EcPF15, forming a distinct branch distant from EC57. In contrast, EC57 formed a well-supported subclade (bootstrap support = 1.000) with strains 04-00955 and VG38-53E1, while strain A4 represented the closest sister lineage to this cluster (bootstrap support = 0.989). The short branch lengths separating EC57 from 04-00955 and VG38-53E1 indicate a high degree of genomic similarity and recent common ancestry. The EC57-containing clade was clearly separated from other well-characterized UPEC reference strains, including CFT073, UTI89, J96, and 536, which formed independent phylogenetic lineages. These results confirm that EC57 belongs to the UPEC phylogenetic group while representing a distinct evolutionary lineage within this pathotype.

3.2. Phage vB-EcoS_57-3 isolation and morphological characterization

The above-characterized UPEC strain EC57 was used to isolate vB-EcoS_57-3 bacteriophage from samples of urban sewage. The isolation procedure was based on the one-host enrichment method in which a raw urban sewage sample was mixed with a culture

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

of the EC57 strain to obtain the lysate of vB-EcoS_57-3 bacteriophage, as described in the Materials and Methods section. The bacteriophage vB-EcoS_57-3 was morphologically characterized using transmission electron microscopy. The virion exhibited a typical siphovirus-like morphology with an icosahedral capsid and a long, flexible, non-contractile tail (**Figure 2A**). Large plaques (diameter 4.5 ± 0.5 mm), with a halo, formed on the *E. coli* EC57 lawn, confirmed its lytic activity against the EC57 isolate and suggested the production of free-releasing polysaccharide depolymerase protein (**Figure 2B**).

3.3. Genome content

The vB-EcoS_57-3 phage possesses a 45,774 bp genome with 91 predicted ORFs, and a GC content of 44%. The whole genome of vB-EcoS_57-3 has been sequenced and deposited in GenBank (accession number PZ349761). Genome annotation revealed modular organization, including genes associated with, e.g., DNA replication, transcription, virion structure, host interaction, and lysis (**Figure 3**). Among all identified coding regions, only 31 ORFs were predicted to be functional genes. Phage lytic enzymes, holin and endolysin, have been identified automatically during annotation by Pharokka, while phage polysaccharide depolymerase was first recognized as a structural protein and subsequently classified by the Phyre2 program as Latka et al., (2019) proposed. Interestingly, only a single (arginine) tRNA gene was predicted in the phage genome (**Figure 3**). In addition, analysis using PhageTermVirome indicated that vB-EcoS_57-3 possesses Headful packing, whereas the PhaTYP classified it as virulent with a confidence equal 1.0. Finally, the comparative genomic analysis (**Figure 4**) performed by the EasyFig program showed conserved genome organization of vB-EcoS_57-3 with closely related phages CEB_EC3a, DTL, IME542, 12210I, Rtp, IME253, and ACG-M12.

3.4. Taxonomic and phylogenetic analyses

According to the PhaGCN analysis, phage vB-EcoS_57-3 was classified within the family *Drexlerviridae* and the subfamily *Braunvirinae*, with confidence scores of 0.99 and 0.61, respectively. To further investigate its taxonomic position, VIRIDIC was used to assess intergenomic relationships between vB-EcoS_57-3 and *Drexlerviridae* reference phages available in RefSeq (**Supplementary Figure S1 and Supplementary Table S1**).

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

VIRIDIC analysis showed that vB-EcoS_57-3 and vB_EcoS-12210I share 71.45% intergenomic similarity (**Supplementary Table S1**), supporting the assignment of vB-EcoS_57-3 to the genus *Veterinaerplatzvirus* (**Supplementary Figure S1**).

Phylogenetic analysis of 20 phages based on the concatenated alignments of 11 orthologous protein groups (**Supplementary Table S2**) placed vB-EcoS_57-3 within a well-supported clade comprising the *Braunvirinae* phages CEB_EC3a, DTL, IME542, vB_EcoS-12210I, Rtp, IME253, and ACG-M12 (**Figure 5**). Within this clade, vB-EcoS_57-3 was most closely related to vB_EcoS-12210I, with the two phages forming a strongly supported sister group (99.8/95), further supporting their close evolutionary relationship. The 11 orthologous protein groups used for phylogenetic reconstruction included DNA primase, major head protein, endolysin, a single-stranded DNA-binding protein, a virion structural protein, several tail-associated proteins, and one hypothetical protein (**Supplementary Table S2**).

3.5. Host range and specificity against uropathogens

Host range analysis (**Table 2**) indicated variable infectivity efficiency across tested *E. coli* strains. Interestingly, phage vB-EcoS_57-3 infects only UPEC bacteria and shows no lytic activity against non-pathogenic laboratory *E. coli* strains. Among tested uropathogens (Dziuba et al., 2023), high or medium efficiency of plating (EOP $\geq$ 0.5 or 0.1, respectively) was observed for six clinical isolates, including EC28, EC47, EC57, EC116, EC131, and EC353, whereas the other ten strains showed low, limited, or no susceptibility (**Table 2**). They were previously characterized based on antimicrobial susceptibility and virulence potential (Dziuba et al., 2023). Half of them were susceptible to all antibiotics, while the remaining showed variable resistance to penicillins, cephalosporins, aminoglycosides, or fluoroquinolones. All were susceptible to carbapenems and were ESBL-negative. Besides, most of the tested strains (72%) belonged to the B2 phylogenetic group, all revealed CRISPR-Cas regions, and all (except one) possessed virulence-associated genes (Dziuba et al., 2023) and the ability to form biofilms (data not shown). Therefore, we conclude that this phage is a moderate host-range virus, however, specific to uropathogenic *E. coli* bacteria.

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

3.6. Bacteriophage development

The parameters of phage vB-EcoS_57-3 development were determined in the EC57 host strain. The analyzed lysis profile of infection (**Figure 6**) revealed a rapid decrease in OD_{600} (**Figure 6A**) and viable cell counts up to 100 cells in 30 minutes after addition of phage (**Figure 6B**). These drops were accompanied by a significant increase in phage particles up to 10^{10} PFU/mL (**Figure 6C**), confirming efficient lytic development. One-step growth analysis (**Figure 6D**) demonstrated efficient phage replication, with a burst size of approximately 42 PFU per cell and an 11-minute-long latent period.

Adsorption of vB-EcoS_57-3 on cells of the EC57 host (**Figure 7A**) was moderately fast but efficient. Initially, a rapid rise in adsorption efficiency was observed, suggesting a fast initial interaction phase between phage particles and bacterial cells, however, this was followed by a transient decrease in the percentage of adsorbed phages, which may reflect short-term binding instability. From the second minute, the adsorption efficiency increased again and continued to rise more gradually, approaching a plateau efficiency equal to 80% at the 10^{th} minute, indicating a possible saturation of available binding sites on the bacterial surface. Importantly, bacterial cell numbers remained stable during all analyzed adsorption stages (**Figure 7B**), proving that the observed drop was not an effect of phage entry inside bacteria and its further lytic development leading to the decrease in bacteria number, but likely due to the reversible phage binding to the cell surface at the early stages of the infection.

3.7. Resistance of vB-EcoS_57-3 virions to external conditions and urine environment

The sensitivity of vB-EcoS_57-3 to various external conditions, including various temperatures, pH conditions, solvents, and detergents was tested (**Table 3**). The virions appeared relatively resistant to freezing, pH values, or osmotic shock, although they could not survive under conditions of 65°C and a pH of 2. Resistance to disinfectants, organic solvents, and detergents differed depending on the nature of the tested compound and was the highest for ethanol and chloroform and the lowest for 0.1% cetyltrimethylammonium bromide (CTAB), in which phages lost their activity completely.

In artificial urine (**Figure 8A**), vB-EcoS_57-3 remained relatively stable across all tested pH values (4.5, 5.8, and 6.5), with only minor fluctuations in PFU counts over time.

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

At pH 5.8 and 6.5, phage titers were consistently maintained at high levels throughout the incubation period, indicating favorable conditions for virion stability. In contrast, at pH 4.5, a slight reduction in PFU was observed at certain time points (4, 6, 8 h), suggesting that more acidic conditions may have a modest negative impact on phage stability. Nevertheless, even under acidic conditions, the decrease was not pronounced, demonstrating that vB-EcoS_57-3 retains antibacterial activity in artificial urine, significantly reducing bacterial counts up to 10% of the initial amount, after 24-hour of incubation (**Figure 8B**). In contrast, in the absence of phage particles, bacterial populations remained stable after 24 hours of incubation, with CFU values close to or only slightly reduced from the initial levels (**Figure 8C**). The survival of bacteria was correspondingly high, indicating that artificial urine at pH 4.5 alone does not significantly inhibit bacterial growth. This highlights that the observed reduction in bacterial counts in **Figure 8B** is specifically attributable to phage activity rather than environmental stress.

3.8. The safety of vB-EcoS_57-3 toward human cells

To test the safety of phage vB-EcoS_57-3 to mammalian cells, the viability of the human bladder cancer cell line, T24 (**Figure 9A**), and the human embryonic kidney line, HEK293 (**Figure 9B**), was assessed after treatment with the phage lysate at 3 concentrations (10^{10}, 10^{9}, and 10^{8} PFU/mL). The endotoxin level in the applied phage lysate was estimated at 0.57 ± 0.03 EU/mL, which is considered to be safe for mammalian cells. Besides, the stability of phage particles was tested after 24-hour incubation with mammalian cells T24 (**Figure 9C**) and HEK293 (**Figure 9D**). The addition of phages to T24 or HEK293 cultures resulted in no significant changes detected in the number of viable cells relative to the negative control, which included non-treated cultures with viability estimated at the level of 100% (**Figure 9A, B**). The lack of a drop in cell survival below 70% in all tested cases (except the positive control) suggested no deleterious effect of the tested phage on the viability of mammalian cells.

3.9. In silico analysis of the 57_3Lys endolysin properties

The vB-EcoS_57-3 genome analysis allowed for the identification of the endolysin (GenBank accession number YEN46648.1) in the predicted lysis cassette. *In silico* analysis

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

(**Table 4**) revealed that the analyzed endolysin gene is located between nucleotides 40,824 and 41,309 in the phage genome and has a length of 486 bp, encoding a relatively small protein of 161 amino acids. The compact size of both the gene and the encoded protein is characteristic of phage lytic enzymes, particularly endolysins. The predicted molecular weight of 17.55 kDa further supported the classification of this protein as a small enzymatic protein, consistent with known phage lysins. The isoelectric point (pI) of 9.7 indicates that the protein is basic in nature, suggesting a net positive charge at physiological pH. This property may facilitate interactions with negatively charged bacterial cell wall components, such as peptidoglycan.

The instability index of 32.544 classifies the protein as stable (values below 40 indicate stability), suggesting that it is likely to maintain its structural integrity under standard conditions. The aliphatic index of 101.18 is relatively high, indicating a substantial proportion of aliphatic amino acids, which is typically associated with enhanced thermal stability. The GRAVY score of -0.057 indicates that the protein is overall slightly hydrophilic. This suggests good solubility in aqueous environments, which is advantageous for enzymes that act on cell wall substrates in extracellular or periplasmic spaces. Domain prediction analysis identified multiple overlapping domains associated with lysozyme and endolysin activity, including SAR endolysin and lysozyme domains at positions 3-158 and 25-151, respectively (**Table 4**), thus confirming that the protein belongs to the phage endolysin/lysozyme family, responsible for degrading bacterial peptidoglycan during host cell lysis.

3.10. Identification of motifs characteristic for class II SAR endolysins within 57_3Lys

Searching for motifs characteristic for phage endolysins revealed three important domains within the 57_3Lys organization structure, typical for class II SAR endolysins (**Figure 10**), represented by R_{21}, the endolysin of the lambdoid phage 21 (Xu et al., 2004; Kuty et al., 2010). These enzymes have the canonical E-D-T catalytic triad and no C residue in the SAR domain.

The N-terminal region (residues 1-20) of 57_3Lys contains a clearly identifiable signal peptide with SAR-domain properties, as indicated by the sequence MKQKLLISAATAAAIYIAAP (**Figure 10** and **Supplementary Figure S2**). This

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

segment is moderately hydrophobic, has no cysteine residue, and is predicted to function as a transient membrane anchor. Its identification by SignalP 6.0 suggests that the 57_3Lys protein, as a typical class II SAR endolysin, is targeted to the Sec-dependent secretory pathway, where it is translocated across the cytoplasmic membrane and remains anchored within the membrane by the N-terminal signal peptide until membrane depolymerization occurs. Unlike classical signal peptides, the SAR domain of class II endolysins is not immediately released but instead retains the enzyme in an inactive, membrane-tethered state.

Downstream of the signal peptide, the 57_3Lys protein sequence forms the catalytic core of the endolysin as revealed by NCBI Conserved Domain software (**Figure 10** and **Supplementary Figure S2**). Within this region, three key residues, E26, D35, and T41, are highlighted as the putative catalytic motif. This arrangement corresponds to the conserved E-$(X)_8$-D-$(X)_5$-T motif, which is characteristic for class II SAR endolysins rather than the canonical cysteine-dependent catalytic triads of class I SAR endolysins. The positioning of these residues suggests a hydrolytic mechanism targeting peptidoglycan, although enzymatic activity is dependent on conformational rearrangements that reposition the catalytic glutamate upon release from the membrane, rather than solely on residue composition.

In addition, the CAMP analysis predicted an antimicrobial peptide (AMP) region between residues 110-140, located within the C-terminal part of the 57_3Lys protein (**Figure 10** and **Supplementary Figure S2**). This region shows strong similarity to cecropin-like α-helical antimicrobial peptides, particularly in its enrichment in positively charged and hydrophobic residues. Despite relatively low primary sequence identity (~20-25%) to Cecropin A from *Hyalophora cecropia*, the physicochemical properties, such as cationic charge distribution and amphipathic character, are conserved, suggesting a similar membrane-disruptive mode of action. This implies that the 57_3Lys protein may possess a dual function: enzymatic degradation of the peptidoglycan and direct disruption of bacterial membranes.

Phylogenetic analysis of 20 orthologous endolysins revealed that the endolysin of vB-EcoS_57-3 (YEN46648.1) clustered with endolysins from related phages (**Figure 11**). The vB-EcoS_57-3 endolysin of vB-EcoS_57-3 formed a well-supported sister group with

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

the endolysin of vB_EcoS_ACG-M12 (SH-aLRT/UFBoot: 85.6/92), while the endolysin of vB_EcoS-12210I was positioned as the next related lineage. The analyzed endolysins generally showed similar physicochemical properties and domain composition (**Supplementary Table S3**). Most were 160-161 amino acids long, with predicted molecular weights of approximately 17.3-17.6 kDa and basic isoelectric points.

Domain analysis further demonstrated a highly conserved architecture among the analyzed endolysins. The vB-EcoS_57-3 endolysin contained signatures corresponding to phage lysozyme (PF00959), endolysin_R21-like (cd16900), lysozyme-like (SSF53955), endolysin (MF_04110), and SAR-endolysin (MF_04136), a domain profile shared by most of the analyzed proteins (**Table 4**, **Supplementary Table S3**). Notable exceptions included the shorter putative phage-related lysozyme of phiJLA23 (131 aa), for which the SAR-endolysin signature was not detected. Overall, the phylogenetic and domain analyses indicate that the vB-EcoS_57-3 endolysin belongs to a group of closely related, structurally conserved phage lysozymes.

3.11. Involvement of the Sec machinery in 57_3Lys-mediated lysis of E. coli bacteria

To examine whether the identified SAR motif functions as a secretion sequence for Sec translocase and whether Sec machinery is responsible for 57_3Lys secretion, the endolysin was produced from the plasmid in the presence of NaN_3 (**Figure 12**), a well-known inhibitor of *E. coli* Sec translocase (Bai et al., 2020). In this experiment, the host cell lysis kinetics of 57_3Lys were determined by adding various concentrations of NaN_3. Increasing concentrations of NaN_3 inhibited the cell lysis process induced by the 57_3Lys produced from the plasmid after the addition of IPTG and arabinose (**Figure 12**). The addition of 5 mM NaN_3 reduced the lytic effect of the 57_3Lys endolysins by about half an OD_{600} unit, whereas the addition of 10 mM NaN_3 almost completely inhibited it and returned the culture turbidity to the level of the variant with non-produced 57_3Lys. This line of evidence implies that the Sec system might be involved in 57_3Lys-mediated cell lysis *in vitro*.

3.12. Functional analysis of the purified 57_3Lys lytic activity

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Functional assays showed that overexpression of 57_3Lys significantly reduced bacterial turbidity, indicating effective lysis (**Figure 12**), so we decided to purify this protein and analyze its lytic action against the EC57 clinical UPEC isolate. The Ni^{2+}-NTA affinity chromatography allowed us to obtain a fraction of the purified 57_3Lys endolysin corresponding to the predicted molecular weight of approximately 17.55 kDa, further verified by mass spectrometry (**Supplementary Figure S3**).

Purified 57_3Lys was used in turbidity and cell number reduction assays that demonstrated its antibacterial activity, significantly reducing the OD_{600} value of the EC57 bacterial culture and the amount of viable EC57 cells expressed as CFU/mL (**Figure 13 A**, **B**). Importantly, the antibacterial effect was observed after 24 h incubation of the bacterial cells with the 57_3Lys endolysin and was not detected during the first 60 minutes of the treatment (**Supplementary Figure S4**). Curiously, incubation with the commercially purchased lysozyme has no significant antibacterial effect, probably due to a lack of access of the lysozyme to the peptidoglycan layer. The observed antibacterial effect of the 57_3Lys in this case may be the result of the outer membrane destabilization by the internal AMP peptide action.

4. Discussion

In this study, we characterized the newly isolated UPEC-infecting bacteriophage vB-EcoS_57-3 and its endolysin 57_3Lys. The phage efficiently lysed EC57, showed activity against selected clinical UPEC isolates, including antibiotic-resistant strains, remained active under urinary conditions, and did not measurably affect the viability of tested human bladder and kidney cell lines. 57_3Lys was identified as a class II SAR endolysin, and functional data indicated that its lytic activity depends strongly on the host Sec pathway. Importantly, purified 57_3Lys also reduced the turbidity and the number of viable EC57 cells without an externally added outer-membrane permeabilizer, suggesting intrinsic activity against Gram-negative bacteria.

Sequence analysis identified an N-terminal, moderately hydrophobic segment of approximately 20 amino acids with properties expected for a SAR domain, together with the E(X)8-D(X)5-T catalytic motif characteristic of class II SAR endolysins. This architecture differs from canonical cytoplasmic phage endolysins, whose access to the

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

periplasm generally depends on holin-generated membrane lesions. In SAR systems, the N-terminal signal-anchor engages the host Sec machinery, allowing translocation of most of the endolysin through the cytoplasmic membrane while the SAR domain remains transiently tethered to it. Membrane depolarization subsequently releases the N-terminal signal and permits formation of the active periplasmic enzyme. The paradigm was established for the P1 endolysin Lyz, for which Sec-dependent export and non-proteolytic release of the N-terminal SAR sequence were demonstrated experimentally (Xu et al., 2004). Related pinholin-SAR systems have been described for phage 21 and phiKMV, illustrating that membrane depolarization rather than formation of large protein-conducting holes can control the timing of lysis (Park et al., 2007; Briers et al., 2011).

Our functional data obtained for 57_3Lys support this pathway. Increasing concentrations of sodium azide, an inhibitor of the *E. coli* SecA-dependent translocation machinery (Bai et al., 2020), progressively suppressed lysis following induction of 57_3Lys expression. At 5 mM NaN_3 the lytic effect was markedly reduced, whereas 10 mM NaN_3 almost completely abolished the decrease in culture turbidity, bringing the phenotype close to that of the variant without induced 57-3Lys production. Because sodium azide can exert cellular effects beyond a single protein-transport event, these experiments should not be interpreted as direct visualization of 57_3Lys translocation. Nevertheless, the dose-dependent inhibition, considered together with the predicted N-terminal SAR signal and phylogenetic placement of 57_3Lys among class II SAR endolysins, provides convergent functional evidence that Sec-mediated export is required for efficient 57_3Lys-mediated lysis.

The absence of cysteine in the SAR region of 57_3Lys further supports a class II SAR endolysin-type activation model. In contrast to class I SAR endolysin P1 Lyz, controlled by topological, conformational and disulfide-dependent changes after membrane release (Xu et al., 2005), class II-like enzymes lacking the cysteine residues within the SAR domain are expected to rely primarily on conformational repositioning of catalytic residues E, D, T after release from the membrane (Woźnica et al., 2015). Importantly, the described 57_3Lys features also argue against describing its N-terminal region as a conventional cleavable signal peptide. In the established class II SAR paradigm, the signal-anchor is

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

retained rather than removed by signal peptidase and is released from the membrane upon depolarization (Xu et al., 2004; Park et al., 2007).

A next particularly notable finding was the antibacterial activity of externally applied 57_3Lys against intact EC57 cells. Because the Gram-negative outer membrane normally restricts access of endolysins to peptidoglycan, this activity suggests that 57_3Lys possesses features facilitating interaction with or passage across the outer membrane. The effect developed slowly, becoming significant after 24 h rather than within the first 60 min, indicating that outer-membrane access may represent the rate-limiting step. Similar intrinsic activity has been reported for ABgp46, LysMK34, LysAm24, LysECD7 and LysSi3 (Oliveira et al., 2016; Antonova et al., 2019; Abdelkader et al., 2022). More recently, PA16cLys demonstrated activity against mature biofilms of uropathogenic *P. aeruginosa* without an additional permeabilizer, further supporting the potential of intrinsically active lysins in urinary infections (Zhang et al., 2026).

The predicted C-terminal AMP-like region of 57_3Lys, enriched in positively charged and hydrophobic residues, together with the high predicted pI (~9.7), provides a plausible explanation for this activity. Such cationic and amphipathic regions may interact electrostatically with lipopolysaccharide and destabilize the outer membrane. Similar principles underlie cecropin A-fused lysins and artilysins (Briers et al., 2014; Abdelkader et al., 2022). However, the contribution of this region to 57_3Lys activity remains to be experimentally demonstrated. The relatively slow killing compared with ABgp46 and engineered membrane-active lysins such as Art-175 (Briers et al., 2014; Oliveira et al., 2016) suggests that improving outer-membrane penetration could substantially enhance 57_3Lys antibacterial efficacy.

Finally, the present study demonstrates that both the parental phage vB-EcoS_57-3 and its endolysin 57_3Lys possess features relevant to this approach. The phage provides self-amplifying, receptor-dependent killing and showed stability and antibacterial activity in artificial urine, whereas the purified endolysin bypasses the need for phage adsorption and replication. This distinction may be particularly relevant when phage resistance emerges. A recent report on PA16cLys is illustrative, as the endolysin retained activity against phage-resistant *P. aeruginosa* mutants (Zhang et al., 2026). Overall, the findings support further investigation of both vB-EcoS_57-3 and 57_3Lys as potential alternative or

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

adjunctive strategies against antibiotic-resistant UPEC, while highlighting the need to improve and mechanistically define the extracellular activity of 57_3Lys. Besides, combined or sequential phage–endolysin treatment could be investigated as a potential strategy to improve antibacterial efficacy and address phage resistance.

CRediT authorship contribution statement

Wojciech Wesołowski: Writing – original draft, Visualization, Investigation, Methodology, Data curation

Sylwia Bloch: Methodology, Investigation, Writing – review & editing

Grzegorz Czerwonka: Methodology, Investigation, Writing – original draft

Ernest Jagieła: Investigation, Visualization

Łukasz Grabowski: Investigation, Visualization

Marta Jeschke: Investigation, Visualization

Jakub Jacewicz: Investigation, Visualization

Emilia Piwnicka: Writing – original draft, Data curation

Joanna Morcinek-Orłowska: Methodology, Data curation

Hanna Loika: Investigation, Visualization

Paulina Czaplewska: Investigation, Methodology

Grzegorz Węgrzyn: Writing – review & editing, Formal analysis.

Wioletta Adamus-Białek: Writing – review & editing, Data curation, Supervision

Aleksandra Łukasiak: Investigation, Funding acquisition, Data curation.

Bożena Nejman-Faleńczyk: Writing – original draft, Writing – review & editing, Supervision, Formal analysis, Conceptualization.

Funding

This research was funded by the National Science Centre (Poland), grant number 2023/49/N/NZ9/04058 to A.Ł.

Declaration of competing interests

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

The authors declare that they have no known competing financial interests or personal relationships

5. References

Abdelkader K, Gutiérrez D, Tamés-Caunedo H, Ruas-Madiedo P, Safaan A, Khairalla AS, Gaber Y, Dishisha T, Briers Y. Engineering a lysin with intrinsic antibacterial activity (LysMK34) by cecropin A fusion enhances its antibacterial properties against *Acinetobacter baumannii*. Appl Environ Microbiol. 2022 Jan 11;88(1):e0151521. doi: 10.1128/AEM.01515-21.

Alcock BP, Huynh W, Chalil R, Smith KW, Raphenya AR, Wlodarski MA, Edalatmand A, Petkau A, Syed SA, Tsang KK, Baker SJC, Dave M, McCarthy MC, Mukiri KM, Nasir JA, Golbon B, Imtiaz H, Jiang X, Kaur K, Kwong M, Liang ZC, Niu KC, Shan P, Yang JYJ, Gray KL, Hoad GR, Jia B, Bhando T, Carfrae LA, Farha MA, French S, Gordzevich R, Rachwalski K, Tu MM, Bordeleau E, Dooley D, Griffiths E, Zubyk HL, Brown ED, Maguire F, Beiko RG, Hsiao WWL, Brinkman FSL, Van Domselaar G, McArthur AG. CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. Nucleic Acids Res. 2023 Jan 6;51(D1):D690-D699. doi: 10.1093/nar/gkac920.

Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, Huynh W, Nguyen AV, Cheng AA, Liu S, Min SY, Miroshnichenko A, Tran HK, Werfalli RE, Nasir JA, Oloni M, Speicher DJ, Florescu A, Singh B, Faltyn M, Hernandez-Koutoucheva A, Sharma AN, Bordeleau E, Pawlowski AC, Zubyk HL, Dooley D, Griffiths E, Maguire F, Winsor GL, Beiko RG, Brinkman FSL, Hsiao WWL, Domselaar GV, McArthur AG. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Res. 2020 Jan 8;48(D1):D517-D525. doi: 10.1093/nar/gkz935.

Altenhoff AM, Levy J, Zarowiecki M, Tomiczek B, Vesztrocy AW, Dalquen DA, Müller S, Telford MJ, Glover NM, Dylus D, Dessimoz C. OMA standalone: orthology inference among public and custom genomes and transcriptomes. Genome Res. 2019 Jul;29(7):1152-1163. doi: 10.1101/gr.243212.118.

Antonova NP, Vasina DV, Lendel AM, Usachev EV, Makarov VV, Gintsburg AL, Tkachuk AP, Gushchin VA. Broad bactericidal activity of the *Myoviridae* bacteriophage lysins LysAm24, LysECD7, and LysSi3 against Gram-negative ESKAPE pathogens. Viruses. 2019 Mar 21;11(3):284. doi: 10.3390/v11030284.

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

1053 Bai J, Lee S, Ryu S. Identification and in vitro characterization of a novel phage endolysin
1054 that targets Gram-negative bacteria. Microorganisms. 2020 Mar 21;8(3):447. doi:
1055 10.3390/microorganisms8030447.

1056 Bankevich A., Nurk S., Antipov D., Gurevich A.A., Dvorkin M., Kulikov A.S., Lesin
1057 V.M., Nikolenko S.I., Pham S., Prjibelski A.D., Pyshkin A.V., Sirotkin A.V., Vyahhi
1058 N., Tesler G., Alekseyev M.A., Pevzner P.A. SPAdes: a new genome assembly
1059 algorithm, its applications to single-cell sequencing. J Comput Biol. 2012 19(5) 455-
1060 77. doi: 10.1089/cmb.2012.0021.

1061 Beyvers S, Jelonek L, Goesmann A, Schwengers O. Bakta Web—rapid and standardized
1062 genome annotation on scalable infrastructures. Nucleic Acids Res. 2025 Jul
1063 7;53(W1):W51-W56. doi: 10.1093/nar/gkaf335.

1064 Bortolaia V, Kaas RS, Ruppe E, Roberts MC, Schwarz S, Cattoir V, Philippon A, Allesoe
1065 RL, Rebelo AR, Florensa AF, Fagelhauer L, Chakraborty T, Neumann B, Werner G,
1066 Bender JK, Stingl K, Nguyen M, Coppens J, Xavier BB, Malhotra-Kumar S, Westh H,
1067 Pinholt M, Anjum MF, Duggett NA, Kempf I, Nykäsenoja S, Olkkola S, Wieczorek K,
1068 Amaro A, Clemente L, Mossong J, Losch S, Ragimbeau C, Lund O, Aarestrup FM.
1069 ResFinder 4.0 for predictions of phenotypes from genotypes. J Antimicrob Chemother.
1070 2020 Dec 1;75(12):3491-3500. doi: 10.1093/jac/dkaa345.

1071 Bouras G, Nepal R, Houtak G, Psaltis AJ, Wormald PJ, Vreugde S. Pharokka: a fast
1072 scalable bacteriophage annotation tool. Bioinformatics. 2023 Jan 1;39(1):btac776. doi:
1073 10.1093/bioinformatics/btac776.

1074 Briers Y, Peeters LM, Volckaert G, Lavigne R. The lysis cassette of bacteriophage ϕKMV
1075 encodes a signal-arrest-release endolysin and a pinholin. Bacteriophage. 2011;1(1):25-
1076 30. doi: 10.4161/bact.1.1.14868.

1077 Briers Y, Walmagh M, Grymonprez B, Biebl M, Pirnay JP, Defraine V, Michiels J, Cenens
1078 W, Aertsen A, Miller S, Lavigne R. Art-175 is a highly efficient antibacterial against
1079 multidrug-resistant strains and persisters of *Pseudomonas aeruginosa*. Antimicrob
1080 Agents Chemother. 2014 Jul;58(7):3774-3784. doi: 10.1128/AAC.02668-14.

1081 Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated
1082 alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009 Aug
1083 1;25(15):1972-1973. doi: 10.1093/bioinformatics/btp348.

1084 Carattoli A, Zankari E, García-Fernández A, Larsen MV, Lund O, Villa L, Aarestrup FM,
1085 Hasman H. In silico detection and typing of plasmids using PlasmidFinder and plasmid
1086 multilocus sequence typing. Antimicrob Agents Chemother. 2014 Jul;58(7):3895-
1087 3903. doi: 10.1128/AAC.02412-14.

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

1088 Cernooka E, Zrelovs N, Kazaks A. C-terminal anchor endolysins—proposing a third class
1089 of tailed bacteriophage endolysins. FEBS Lett. 2025 Jun;599(11):1499-1508. doi:
1090 10.1002/1873-3468.70042.

1091 Chan PP, Lin BY, Mak AJ, Lowe TM. tRNAscan-SE 2.0: improved detection and
1092 functional classification of transfer RNA genes. Nucleic Acids Res. 2021 Sep
1093 20;49(16):9077-9096. doi: 10.1093/nar/gkab688.

1094 Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, Jin Q. VFDB: a reference database for
1095 bacterial virulence factors. Nucleic Acids Res. 2005 Jan 1;33:D325-D328. doi:
1096 10.1093/nar/gki008

1097 Chernomor O, von Haeseler A, Minh BQ. Terrace aware data structure for phylogenomic
1098 inference from supermatrices. Syst Biol. 2016 Nov;65(6):997-1008. doi:
1099 10.1093/sysbio/syw037.

1100 Cito B, Corneille KA, Gates IM, Bruno KK, Ruhimbasa KC, Gloire MH, Bahaya J, Kawala
1101 ZL, Makeshe OM, Mwenebitu DL. Bacteriological profile and antibiotic resistance
1102 patterns of uropathogenic bacteria at HPGRB. Infect Drug Resist. 2026;19:603097.
1103 doi: 10.2147/IDR.S603097.

1104 Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck
1105 T, Kauff F, Wilczynski B, de Hoon MJL. Biopython: freely available Python tools for
1106 computational molecular biology and bioinformatics. Bioinformatics. 2009 Jun
1107 1;25(11):1422-1423. doi: 10.1093/bioinformatics/btp163.

1108 Couvin D, Bernheim A, Toffano-Nioche C, Touchon M, Michalik J, Néron B, Rocha EPC,
1109 Vergnaud G, Gautheret D, Pourcel C. CRISPRCasFinder, an update of CRISRFinder,
1110 includes a portable version, enhanced performance and integrates search for Cas
1111 proteins. Nucleic Acids Res. 2018 Jul 2;46(W1):W246-W251. doi:
1112 10.1093/nar/gky425.

1113 Czajkowski, R., Ozymko, Z., Lojkowska, E. Isolation, characterization of novel soilborne
1114 lytic bacteriophages infecting *Dickeya spp.* biovar 3 ('D. solani'). Plant Pathol, 2014
1115 63: 758-772. doi: 10.1111/ppa.12157.

1116 Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane
1117 T, McCarthy SA, Davies RM, Li H. Twelve years of SAMtools and BCFtools.
1118 Gigascience. 2021 Feb 16;10(2):giab008. doi: 10.1093/gigascience/giab008.

1119 Dziuba A, Dzierżak S, Sodo A, Wawszczak-Kasza M, Zegadło K, Białek J, Zych N,
1120 Kiebzak W, Matykiewicz J, Głuszek S, Adamus-Białek W. Comparative study of
1121 virulence potential, phylogenetic origin, CRISPR-Cas regions and drug resistance of
1122 *Escherichia coli* isolates from urine and other clinical materials. Front Microbiol. 2023
1123 Nov 29;14:1289683. doi: 10.3389/fmicb.2023.1289683.

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Feldgarden M, Brover V, Gonzalez-Escalona N, Frye JG, Haendiges J, Haft DH, Hoffmann M, Pettengill JG, Prasad AB, Tillman GE, Tyson GH, Klimke W. AMRFinderPlus and the Reference Gene Catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. Sci Rep. 2021 Jun 16;11(1):12728. doi: 10.1038/s41598-021-91456-0.

Fiołka MJ, Czaplewska P, Wójcik-Mieszawska S, Lewandowska A, Lewtak K, Sofińska-Chmiel W, Buchwald T. Metabolic, structural, and proteomic changes in *Candida albicans* cells induced by the protein-carbohydrate fraction of Dendrobaena veneta coelomic fluid. Sci Rep. 2021 Aug 18;11(1):16711. doi: 10.1038/s41598-021-96093-1.

Flores-Mireles AL, Walker JN, Caparon M, Hultgren SJ. Urinary tract infections: epidemiology, mechanisms of infection and treatment options. Nat Rev Microbiol. 2015 May;13(5):269-284. doi: 10.1038/nrmicro3432.

Fuad M, Mahmud Z, Mishu ID, Tamanna S, Ansari F, Manik MRK, Howlader ZH. Molecular characterization of extended-spectrum β-lactamase (ESBL) and virulent genes in multidrug-resistant *Escherichia coli* isolated from pharmaceutical and environmental wastewaters in Dhaka, Bangladesh. Sci Rep. 2026 Apr 1;16(1):15633. doi: 10.1038/s41598-026-46554-2.

Garneau J.R., Legrand V., Marbouty M., Press M.O., Vik D.R., Fortier L.C., Sullivan M.B., Bikard D., Monot M. 2021 High-throughput identification of viral termini, packaging mechanisms in virome datasets using PhageTermVirome. Scientific Reports, 11(1) 18319. doi: 10.1038/s41598-021-97867-3.

Grant JR, Enns E, Marinier E, Mandal A, Herman EK, Chen CY, Graham M, Van Domselaar G, Stothard P. Proksee: in-depth characterization and visualization of bacterial genomes. Nucleic Acids Res. 2023 Jul 5;51(W1):W484-W492. doi: 10.1093/nar/gkad326.

Groleau MC, Houle S, Quevedo AC, McKay G, Nguyen D, Dozois CM, Tufenkji N, Déziel E. Cranberry juice potentiates sensitivity of uropathogenic *Escherichia coli* (UPEC) strains to fosfomycin and decreases occurrence of spontaneous resistance. Appl Environ Microbiol. 2026 May 20;92(5):e0252125. doi: 10.1128/aem.02521-25.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 2010 May;59(3):307-321. doi: 10.1093/sysbio/syq010.

Hallgren J, Tsirigos KD, Pedersen MD, Almagro Armenteros JJ, Marcatili P, Nielsen H, Krogh A, Winther O. DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. bioRxiv. 2022 Apr 10. doi:10.1101/2022.04.08.487609.

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Hatfull GF. Dark matter of the biosphere: the amazing world of bacteriophage diversity. J Virol. 2015 Aug;89(16):8107-8110. doi: 10.1128/JVI.01340-15.

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol. 2018 Feb 1;35(2):518-522. doi: 10.1093/molbev/msx281.

Huson DH, Scornavacca C. Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. Syst Biol. 2012 Dec 1;61(6):1061-1067. doi: 10.1093/sysbio/sys062.

Hussein AH, Makky S, Hager R, Connerton IF, El-Shibiny A. Characterization, therapeutic potential of phage vB_Eco_ZCEC08 against multidrug-resistant uropathogenic *Escherichia coli*. BMC Microbiology. 2025; 25: 221. doi: 10.1186/s12866-025-03903-x.

Ikai A. Thermostability and aliphatic index of globular proteins. J Biochem. 1980 Dec;88(6):1895-1898.

Joensen KG, Scheutz F, Lund O, Hasman H, Kaas RS, Nielsen EM, Aarestrup FM. Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. J Clin Microbiol. 2014 May;52(5):1501-1510. doi: 10.1128/JCM.03617-13.

Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R, Hunter S. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014 May 1;30(9):1236-1240. doi: 10.1093/bioinformatics/btu031.

Joshi B, Zulk JJ, Serchejian C, Hameed ZA, Larson AB, Terwilliger AL, Kumar D, Mysorekar IU, Britton RA, Maresso AW, Patras KA. Bacteriophage-mediated reduction of uropathogenic *Escherichia coli* from the urogenital epithelium. Infect Immun. 2026 Apr 15;94(4):e0054325. doi: 10.1128/iai.00543-25.

Jurczak-Kurek A, Gąsior T, Nejman-Faleńczyk B, Bloch S, Dydecka A, Topka G, Necel A, Jakubowska-Deredas M, Narajczyk M, Richert M, Mieszkowska A, Wróbel B, Węgrzyn G, Węgrzyn A. Biodiversity of bacteriophages: morphological and biological properties of a large group of phages isolated from urban sewage. Sci Rep. 2016 Oct 4;6:34338. doi: 10.1038/srep34338.

Kaas RS, Leekitcharoenphon P, Aarestrup FM, Lund O. Solving the problem of comparing whole bacterial genomes across different sequencing platforms. PLoS One. 2014 Aug 11;9(8):e104984. doi: 10.1371/journal.pone.0104984.

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017 Jun;14(6):587-589. doi: 10.1038/nmeth.4285.

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Khatun MN, Chakraborty S, Akter T, Ifte JS, Begum JA, Islam A, Parvin R, Hossain MT, Chowdhury EH. Migratory birds facilitate the spread of multidrug-resistant pathogenic *Escherichia coli* in Tanguar Haor of Bangladesh. Environ Microbiol Rep. 2026 Apr;18(2):e70344. doi: 10.1111/1758-2229.70344.

Kuty GF, Xu M, Struck DK, Summer EJ, Young R. Regulation of a phage endolysin by disulfide caging. J Bacteriol. 2010 Nov;192(21):5682-5687. doi: 10.1128/JB.00674-10.

Larralde M, Zeller G. PyHMMER: a Python library binding to HMMER for efficient sequence analysis. Bioinformatics. 2023 May 4;39(5):btad214. doi: 10.1093/bioinformatics/btad214.

Latka A, Leiman PG, Drulis-Kawa Z, Briers Y. Modeling the architecture of depolymerase-containing receptor binding proteins in *Klebsiella* phages. Front Microbiol. 2019 Nov 15;10:2649. doi: 10.3389/fmicb.2019.02649.

Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021 Jul 2;49(W1):W293-W296. doi: 10.1093/nar/gkab301.

Liu B, Zheng D, Jin Q, Chen L, Yang J. VFDB 2019: a comparative pathogenomic platform with an interactive web interface. Nucleic Acids Res. 2019 Jan 8;47(D1):D687-D692. doi: 10.1093/nar/gky1080.

Łukasiak A, Wesołowski W, Neumann J, Lewandowska N, Węglińska E, Bloch S, Węgrzyn G, Nejman-Faleńczyk B. Groundwork for phage therapy: multi-faceted comparative analysis of lytic bacteriophages infecting uropathogenic *Escherichia coli*. Virology. 2026; 615: 110738. doi: 10.1016/j.virol.2025.110738.

Martin M., Cutadapt removes adapter sequences from high-throughput sequencing reads, EMBnet.journal 17 2011 10-12. doi: 10.14806/ej.17.1.200.

McNair K, Zhou C, Dinsdale EA, Souza B, Edwards RA. PHANOTATE: a novel approach to gene identification in phage genomes. Bioinformatics. 2019 Nov 1;35(22):4537-4542. doi: 10.1093/bioinformatics/btz265.

Mirzaei MK, Nilsson AS. Isolation of phages for phage therapy: a comparison of spot tests and efficiency of plating analyses for determination of host range and efficacy. PLoS One. 2015 Mar 11;10(3):e0118557. doi: 10.1371/journal.pone.0118557.

Moraru C, Varsani A, Kropinski AM. VIRIDIC—a novel tool to calculate the intergenomic similarities of prokaryote-infecting viruses. Viruses. 2020 Nov 10;12(11):1268. doi: 10.3390/v12111268.

Necel A, Bloch S, Nejman-Faleńczyk B, Grabski M, Topka G, Dydecka A, Kosznik-Kwaśnicka K, Grabowski Ł, Jurczak-Kurek A, Wołkowicz T, Węgrzyn G, Węgrzyn A. Characterization of a bacteriophage, vB_Eco4M-7, that effectively infects many

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Escherichia coli O157 strains. Sci Rep. 2020 Feb 28;10(1):3743. doi: 10.1038/s41598-020-60568-4.

Niaz H, Abbas A, Ayub A, Rahman A, Adnan F. Characterization and genome analysis of jumbo Escherichia phage UE-S1 and the antibacterial potential of its lysozyme against uropathogenic *Escherichia coli*. BMC Microbiol. 2025 Dec 11;26(1):27. doi: 10.1186/s12866-025-04513-3.

Nielsen H. Practical Applications of Language Models in Protein Sorting Prediction: SignalP 6.0, DeepLoc 2.1, and DeepLocPro 1.0. Methods Mol Biol. 2025 2941:153-175. doi: 10.1007/978-1-0716-4623-6_10.

Oliveira H, Vilas Boas D, Mesnage S, Kluskens LD, Lavigne R, Sillankorva S, Secundo F, Azeredo J. Structural and enzymatic characterization of ABgp46, a novel phage endolysin with broad anti-Gram-negative bacterial activity. Front Microbiol. 2016 Feb 26;7:208. doi: 10.3389/fmicb.2016.00208.

Park T, Struck DK, Dankenbring CA, Young R. The pinholin of lambdoid phage 21: control of lysis by membrane depolarization. J Bacteriol. 2007 Dec;189(24):9135-9139. doi: 10.1128/JB.00847-07.

Salman HS, Majeed HM, Shallal YF. Prevalence, etiology, and antibiotic resistance patterns of urinary tract infections in a Baghdad hospital: focus on uropathogenic *Escherichia coli* and virulence factors. Int J Microbiol. 2026 Apr 20:2026:9117996. doi: 10.1155/ijm/9117996.

Sambrook, J, Russell, D. W. 2001. Molecular Cloning: A Laboratory Manual, 3rd Edn. New York, NY: Cold Spring Harbor Laboratory Press.

Schwengers O, Jelonek L, Dieckmann MA, Beyvers S, Blom J, Goesmann A. Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. Microb Genom. 2021 Nov;7(11):000685. doi: 10.1099/mgen.0.000685.

Seemann T. 2023. ABRicate. GitHub repository. Retrieved from https://github.com/tseemann/abricate.

Shang J, Jiang J, Sun Y. Bacteriophage classification for assembled contigs using graph convolutional network. Bioinformatics. 2021 Jul 12;37(Suppl 1):i25-i33. doi: 10.1093/bioinformatics/btab293.

Shang J, Tang X, Sun Y. PhaTYP: predicting the lifestyle for bacteriophages using BERT. Brief Bioinform. 2023 Jan 19;24(1):bbac487. doi: 10.1093/bib/bbac487.

Shen A, Millard A. Phage Genome Annotation: Where to Begin, End. PHAGE: Therapy, Applications, Research. 2021;2:183–93. doi: 10.1089/phage.2021.0015.

Sievers F, Higgins DG. Clustal Omega for making accurate alignments of many protein sequences. Protein Sci. 2018 Jan;27(1):135-145. doi: 10.1002/pro.3290.

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

1270 Steinegger M, Söding J. MMseqs2 enables sensitive protein sequence searching for the
1271 analysis of massive data sets. Nat Biotechnol. 2017 Nov;35(11):1026-1028. doi:
1272 10.1038/nbt.3988.

1273 Sullivan MJ, Petty NK, Beatson SA. Easyfig: a genome comparison visualizer.
1274 Bioinformatics. 2011 Apr 1;27(7):1009-1010. doi: 10.1093/bioinformatics/btr039.

1275 Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L,
1276 Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. NCBI prokaryotic genome
1277 annotation pipeline. Nucleic Acids Res. 2016 Aug 19;44(14):6614-6624. doi:
1278 10.1093/nar/gkw569.

1279 Terlizzi ME, Gribaudo G, Maffei ME. Uropathogenic *Escherichia coli* (UPEC) infections:
1280 virulence factors, bladder responses, antibiotic, and non-antibiotic antimicrobial
1281 strategies. Front Microbiol. 2017 Aug 15;8:1566. doi: 10.3389/fmicb.2017.01566.

1282 Terzian P, Olo Ndela E, Galiez C, Lossouarn J, Pérez Bucio RE, Mom R, Toussaint A,
1283 Petit MA, Enault F. PHROG: families of prokaryotic virus proteins clustered using
1284 remote homology. NAR Genom Bioinform. 2021 Aug 5;3(3):lqab067. doi:
1285 10.1093/nargab/lqab067.

1286 Topka-Bielecka G, Bloch S, Nejman-Faleńczyk B, Grabski M, Jurczak-Kurek A, Górniak
1287 M, Dydecka A, Necel A, Węgrzyn G, Węgrzyn A. Characterization of the
1288 bacteriophage vB_EfaS-271 infecting *Enterococcus faecalis*. Int J Mol Sci. 2020 Sep
1289 1;21(17):6345. doi: 10.3390/ijms21176345.

1290 Torzewska A, Różalski A. Various intensity of *Proteus mirabilis*-induced crystallization
1291 resulting from the changes in the mineral composition of urine. Acta Biochim Pol. 2015
1292 62(1):127-132. doi: 10.18388/abp.2014_882.

1293 Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski
1294 E, Peterson P, Weckesser W, Bright J, van der Walt S, Brett M, Wilson J, Jarrod
1295 Millman K, Mayorov N, Nelson ARJ, Jones E, Kern R, Larson E, Carey CJ, Polat I,
1296 Feng Y, Moore EW, VanderPlas J, Laxalde D, Perktold J, Cimrman R, Henriksen I,
1297 Quintero EA, Harris CR, Archibald AM, Ribeiro AH, Pedregosa F, van Mulbregt M.
1298 SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods.
1299 2020 Mar;17(3):261-272. doi: 10.1038/s41592-019-0686-2.

1300 Waghu FH, Barai RS, Gurung P, Idicula-Thomas S. CAMPR3: a database on sequences,
1301 structures and signatures of antimicrobial peptides. Nucleic Acids Res. 2016 Jan
1302 4;44(D1):D1094-D1097. doi: 10.1093/nar/gkv1051.

1303 Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q,
1304 Wortman J, Young KY, Earl AM. Pilon: an integrated tool for comprehensive
1305 microbial variant detection and genome assembly improvement. PLoS One. 2014 Nov
1306 19;9(11):e112963. doi: 10.1371/journal.pone.0112963.

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Lu S, Marchler GH, Song JS, Thanki N, Yamashita RA, Yang R, Zhang D, Bryant SH. The conserved domain database in 2023. Nucleic Acids Res. 2023 Jan 6;51(D1):D384-D388. doi: 10.1093/nar/gkac1096.

Wesołowski W, Czerwonka G, Zegadło K, Jagieła E, Łukasiak A, Grabowski Ł, Mantej J, Dziuba A, Bloch S, Topka-Bielecka G, Necel A, Kostovski D, Węglińska E, Narajczyk M, Węgrzyn G, Nejman-Faleńczyk B, Adamus-Białek W. Significant variability in virulence factors and antibiotic resistance of uropathogenic *Escherichia coli* strains may be counteracted by their susceptibility to lytic bacteriophages. Sci Rep. 2026 Jul 21. doi: 10.1038/s41598-026-62793-9.

Wesołowski W, Łukasiak A, Bloch S, Kuligowska K, Neumann J, Lewandowska N, Węglińska E, Węgrzyn G, Nejman-Faleńczyk B. Phage endolysins as promising and effective candidates for use against uropathogenic *Escherichia coli*. Viruses. 2025 Apr 13;17(4):560. doi: 10.3390/v17040560.

Wong TKF, Ly-Trong N, Ren H, Demotte P, Baños H, Roger AJ, Susko E, Bielow C, De Maio N, Goldman N, Hahn MW, dos Reis M, Vinh LS, Huttley G, Lanfear R, Minh BQ. IQ-TREE 3: phylogenomic inference software using complex evolutionary models. Mol Biol Evol. 2026 May 1;43(5):msag117. doi: 10.1093/molbev/msag117.

Woźnica WM, Bigos J, Łobocka MB. Liza komórek bakteryjnych w procesie uwalniania bakteriofagów – kanoniczne i nowo poznane mechanizmy. Postępy Hig Med Dosw. 2015; 69:114-126. doi: 10.5604/01.3001.0009.6484.

Xia Y, Chen Z, Pan J, Liu X, Ding L, Jiang L. Phage and endolysin-based inhibition of *Klebsiella pneumoniae*: from mechanisms to combination therapies. Front Cell Infect Microbiol. 2026 16:1819965. doi: 10.3389/fcimb.2026.1819965.

Xu M, Arulandu A, Struck DK, Swanson S, Sacchettini JC, Young R. Disulfide isomerization after membrane release of its SAR domain activates P1 lysozyme. Science. 2005 Jan 7;307(5706):113-117. doi: 10.1126/science.1105143.

Xu M, Struck DK, Deaton J, Wang IN, Young R. A signal-arrest-release sequence mediates export and control of the phage P1 endolysin. Proc Natl Acad Sci U S A. 2004 Apr 27;101(17):6415-6420. doi: 10.1073/pnas.0400957101.

Young R. Phage lysis: do we have the whole story yet? Curr Opin Microbiol. 2013 Dec;16(6):790-797. doi: 10.1016/j.mib.2013.08.008.

Young R. Phage lysis: three steps, three choices, one outcome. J Microbiol. 2014 Mar;52(3):243-258. doi: 10.1007/s12275-014-4087-z.

Zhang S, Niu H, Hong Y, Yang R, Shahbaz Z, You J, Quan Z, Li J, Li X, Xu Y, Wang L. Characterization and therapeutic potential of a novel lytic phage-derived endolysin

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

PA16cLys against uropathogenic *Pseudomonas aeruginosa* biofilms. Biotechnol Bioeng. 2026 Jul 5. doi: 10.1002/bit.70301.

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Table 1. CRISPR arrays and *cas* gene sequences detected in EC57 genome

CRISPR Id / Cas Type	Start	End	No of spacers /genes	Repeat consensus sequences / *cas* genes	Direction	Evidence Level
EC57-1	1685647	1685760	1	TGCCGGATGCGCTTCGCTTATCCGGCCTACAAA	ND	1
EC57-2	2399780	2399903	1	CGACCCCCACCATGTCAAGGTGGTGCTCTAACCAA CTGAGCTA	ND	1
EC57-3	3029915	3030006	1	CCACCTTTTTTACCTGCTTCAGATGC	ND	1
EC57-4	3151428	3152355	15	TTTCTAAGCTGCCTGTACGGCAGTGAAC	-	4
CAS-Type I-F	3152894	3161241	6	*cas6, csy3, csy2, csy1, cas3-cas2, cas1*		
EC57-5	3161414	3162042	10	TTTCTAAGCTGCCTGTACGGCAGTGAAC	-	4
EC57-6	4011107	4011197	1	AACGCCTTATCCGGCCTACAAAACTGACAA	ND	1
EC57-7	4087523	4087640	1	TTTGTAGGCCGGATAAGCAAAGCGCATCCGGCA	ND	1
EC57-8	4815925	4816051	1	TGCCGGATGCGGCACGAGTGCCTTATCCGGCCTAC	ND	1

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Table 2. EOP (efficiency of plating) analysis of vB-EcoS_57-3 phage on *E. coli* laboratory strains and UPEC clinical isolates

	***E. coli* strain**	**EOP after vB-EcoS_57-3 infection***
Laboratory *E. coli* strains	MG1655	0
	C600	0
	MC1061	0
	BL21	0
	DH5α	0
Clinical UPEC isolates	EC02	0
	EC10	≤0.001
	EC28	1.48
	EC47	1.37
	EC55	0
	EC57	1
	EC70	0.007
	EC106	≤0.001
	EC116	0.72
	EC131	0.42
	EC149	0
	EC164	0
	EC179	≤0.001
	EC296	0
	EC302	0
	EC353	0.84

*According to the accepted classification proposed by Mirzaei et al. (2015) an EOP value ≤0.001 indicates sporadic production of phage particles in a given bacterial strain, while values in the range 0.001< EOP <0.1; 0.1≤ EOP <0.5, and ≥0.5 indicate low, medium, and high efficiency of phage infection, respectively. Lack of infection is indicated by 0. The obtained results are from three independent replicates. The SD values are not shown; however, they do not exceed 15%.

Table 3. Resistance of vB-EcoS_57-3 phage virion to physical and chemical agents. The values presented in the table represent the percentage of viable phage particles under the specified conditions.

Phage name	Temperature			pH (1h, 37°C)			Osmotic shock	0.1% CTAB (1 min)	0.1% Sarkosyl (10 min)	63% Ethanol (1h)	50% DMSO (10 min)	Chloroform (1.5 h)
	-20°C (24h)	40°C (40 min)	62°C (40 min)	pH 2 (1h)	pH 4 (1h)	pH 10 (1h)						
vB-EcoS_57-3	100	100	1.5	0	100	100	100	0	50	100	50	100

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Table 4. Characteristics of the 57_3Lys endolysin

Analyzed parameter	Result
Gene position in genome	40,824-41,309
Gene length (nt)	486
Protein accession number	YEN46648.1
Protein length (aa)	161
Molecular weight (kDa)	17.55
Isoelectric point (Ip)	9.7
Instability index	32.544
Aliphatic index	101.18
GRAVY	-0.057
Melting point (°C)	51
Signal peptide position in protein sequence (aa)	1-20
AMP position in protein sequence (aa)	110-140
Catalytic triad and its position in protein sequence	E (26) D (35) T (41)
Predicted domain and its position in protein sequence (aa)	26-158 endolysin_R21-like (cd16900) 3-158 SAR-endolysin (MF_04136) 25-151 Phage lysozyme (PF00959) 6-159 Lysozyme-like (SSF53955) 4-158 Endolysin (MF_04110)

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Figure legends

Figure 1. SNP-based phylogenetic tree of *E. coli* EC57 and representative *E. coli* strains. The phylogenetic tree was constructed using whole-genome single nucleotide polymorphism (SNP) analysis performed with CSI Phylogeny and visualized using iTOL. The laboratory strain *E. coli* K-12 MG1655 (reference) was included as a non-pathogenic reference genome. The grey-shaded sector indicates reference UPEC strains included for phylogenetic comparison.

Figure 2. Electron micrograph of phage vB-EcoS_57-3 (**A**) and its plaques formed in double-layer agar plates with *E. coli* EC57 strain (**B**). The bar corresponds to 50 nm on panel A and 1 mm on panel B.

Figure 3. Genome map of bacteriophage vB-EcoS_57-3. Predicted open reading frames (ORFs) are displayed as arrows indicating transcriptional orientation and are color-coded according to functional categories, including DNA, RNA and nucleotide metabolism, transcription regulation, tail, connector, head and packaging, lysis, integration and excision, moron/auxiliary metabolism and host takeover, other functions, and hypothetical proteins (legend on the right side). The inner rings represent GC content and GC skew across each genome. Annotated genes involved in genome replication, virion structure and assembly, host interaction, and lysis are indicated. Genome coordinates (kbp) are shown inside each map.

Figure 4. Comparative genome organization of vB-EcoS_57-3 linear genome bacteriophage with their closest related reference phages. Predicted open reading frames (ORFs) are displayed as arrows indicating gene orientation and are color-coded according to functional categories listed in the legend. Grey-shaded regions between genomes represent nucleotide sequence similarity, illustrating conserved synteny and homologous genomic regions. The scale bar represents the percentage of nucleotide sequence identity. The comparison highlights both conserved genome architecture within related phages and structural divergence, supporting the taxonomic distinction of the newly characterized phages. Comparisons were made using Easyfig software.

Figure 5. Maximum-likelihood phylogenetic tree of bacteriophage vB-EcoS_57-3 and related phages based on orthologous protein sequences. The phylogenetic tree was reconstructed from concatenated alignments of 11 orthologous protein groups shared among 20 phages (**Supplementary Table S2**) using IQ-TREE v3.1.1. Phylogenetic inference was performed using an edge-linked partition model, with the optimal substitution model for each partition selected by ModelFinder Plus. Branch support values are shown at the nodes as SH-aLRT support (%) / ultrafast bootstrap support (%), based on

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

1,000 replicates each. Phage vB-EcoS_57-3 is highlighted in bold. Phage F61 was used as the outgroup. Different name colors indicate the genus-level taxonomic assignment of the tested phages as follows: red *Rogunavirus*, orange *Eastlansingvirus*, dark yellow *Wilsonroadvirus*, bottle green *Lindendrivevirus*, green *Loudonvirus*, teal *Christensenvirus*, dark cyan *Veterinaerplatzvirus*, blue *Rtpvirus*, violet *Guelphvirus*, pink *Tlsvirus*. The scale bar represents the number of amino acid substitutions per site.

Figure 6. Kinetics of lytic development of bacteriophage vB-EcoS_57-3 in *E. coli* EC57 bacteria (**A-C**) and progeny production by vB-EcoS_57-3 phage in the single life cycle (**D**). Filled squares represent the bacterial strain EC57 following infection with phage vB-EcoS_57-3, whereas open squares represent the uninfected control EC57 without phage addition. Results are shown as (**A**) bacterial culture density measured at OD_{600}, (**B**) a number of surviving cells after the vB-EcoS_57-3 infection per 1 mL (CFU/mL), (**C**) a number of phages per 1 mL (PFU/mL), and (**D**) PFU per 1 cell. The average burst size in EC57 at the 16^{th} minute after infection has been estimated to be 42 PFU per cell, indicating efficient lytic development of the phage. Results are presented as mean values ± SD from three independent experiments. Please note that in some cases, the bars are smaller than the sizes of the symbols. The Student's *t*-test was used to calculate statistical significance. The significance of differences between infected cells and control EC57 culture is observed and marked by asterisks: $P < 0.001$ (***).

Figure 7. Kinetics of vB-EcoS_57-3 phage adsorption on host EC57 cells (**A**) and number of bacterial cells during this process (**B**). Filled squares represent the bacterial strain EC57 following infection with phage vB-EcoS_57-3, whereas open squares represent the uninfected control EC57 without phage addition. Results are presented as mean values ± SD from three independent experiments. Please note that the bars are smaller than the sizes of the symbols.

Figure 8. Stability and activity of phage vB-EcoS_57-3 in artificial urine. (**A**) Stability analysis of phage virions was performed at 2, 4, 6, and 8 hours of incubation in artificial urine with different pH (4.5; 5.8; 6.5). The number of phage particles was presented as the number of plaque-forming units (PFU) present in 1 mL of urine. (**B**, **C**) Activity of bacteriophage vB-EcoS_57-3 presented as the number of bacterial cells of the EC57 strain that survived 24-hour incubation in artificial urine at pH = 4.5, in the presence (**B**) and without the addition of phage particles (**C**). The number of bacterial cells was expressed as colony-forming units (CFU) per 1 mL of urine. The number of cells that survived the infection relative to the initial number of bacteria is expressed in percentages above the bars. The Student's *t*-test was used to calculate statistical significance. Statistically significant results are marked with asterisk symbols: $P < 0.01$ (**) and $P < 0.001$ (***).

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Figure 9. Assessment of cytotoxicity of vB-EcoS_57-3 phage particles to mammalian cells T24 (**A**) and HEK293 (**B**) after 24 h incubation. The phage lysate tested at 3 concentrations (10^{10}, 10^{9}, and 10^{8} PFU/mL). The letter T means a drop in cell survival below the red dotted threshold line at 70%, which qualifies the agent as toxic. The stability of phage vB-EcoS_57-3 during incubation with mammalian cells T24 (**C**) and HEK293 (**D**) shown as the number of phage particles in 1 mL (PFU/mL) of lysate measured at time 0 and after 24 hours of incubation. The Student's *t*-test was used to calculate statistical significance, which are marked with asterisk symbols: $P < 0.001$ (***).

Figure 10. Sequence of 57_3Lys endolysin with predicted motifs characteristic for class II SAR endolysins: a typical signal peptide with features of a SAR domain, including a moderately hydrophobic segment encompassing the first 20 amino acids; putative catalytic residues (E26, D35, and T41); and an AMP peptide at the C-terminus, between 110 and 140 amino acids.

Figure 11. Phylogenetic relationships of the endolysin from bacteriophage vB-EcoS_57-3 and orthologous endolysins from related phages. The maximum-likelihood phylogenetic tree was constructed based on amino acid sequences of 20 orthologous endolysins. Branch support values at the nodes are presented as SH-aLRT support (%) / ultrafast bootstrap support (%), based on 1,000 replicates each. The endolysin of vB-EcoS_57-3 (YEN46648.1) is highlighted in bold. Different protein name colors indicate the genus-level taxonomic assignment of the analyzed phages as follows: red *Rogunavirus*, orange *Eastlansingvirus*, dark yellow *Wilsonroadvirus*, bottle green *Lindendrivevirus*, green *Loudonvirus*, teal *Christensenvirus*, dark cyan *Veterinaerplatzvirus*, blue *Rtpvirus*, violet *Guelphvirus*, pink *Tlsvirus*. The scale bar represents the number of amino acid substitutions per site.

Figure 12. Kinetics of *E. coli* cell lysis by the 57_3Lys endolysin in the presence of the Sec translocase inhibitor, NaN_3. In all experiments, expression of 57_3Lys from plasmid in *E. coli* cells was induced by adding 0.5 mM of IPTG and 0.2% arabinose, except for the controls. Dark blue circles indicate samples without NaN_3, IPTG and arabinose (control 1), orange circles indicate samples treated with 10 mM NaN_3 in the absence of IPTG and arabinose (control 2), while dark green, light blue, dark pink, and light green circles represent samples treated with 0, 1, 5, and 10 mM NaN_3, respectively. All experiments for the cell lysis kinetics were performed in triplicate and the results of each treatment are represented by the mean ± SD. Statistical analyses were performed by the Student's *t*-test. The significance of differences between control 1 and particular variants are observed from the 10^{th} minute and marked by asterisks: $P < 0.001$ (***). Please note that the bars are smaller than the sizes of the symbols.

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Figure 13. Turbidity (**A**) and cell number (**B**) reduction assays of EC57 culture by purified 57_3Lys endolysin after 24 hours of incubation. Each experiment was performed in triplicate, and the results of each treatment are represented by the mean ± SD. Statistical analyses were performed by the Student's *t*-test. The significance of differences between endolysin variant and particular control variants are observed and marked by asterisks: $P < 0.01$ (**) and $P < 0.001$ (***).

bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Supplementary materials

Supplementary Figure S1. VIRIDIC-based clustering of phage vB-EcoS_57-3 and reference phages from the family *Drexlerviridae*. Hierarchical clustering based on pairwise intergenomic distances calculated using VIRIDIC for phage vB-EcoS_57-3 and reference phages from the family *Drexlerviridae*. Clustering was performed using complete linkage. Dashed lines indicate the distance thresholds corresponding to the species level (5%; red), genus level (30%; blue), and for limiting the number of the analyzed phages up to 20 cases (70%; green).

Supplementary Figure S2. Results of bioinformatic analysis of the 57_3Lys endolysin sequence confirming the presence of a 20 aa-long signal molecule at the N end (>SignalP 6.0 and >Deep TMHMM); a catalytic triad E-D-T encompassing 26th, 35th, and 41st residues, respectively (>NCBI Conserved Domain); and an AMP peptide occurring at a position between 110 and 140 amino acids (>CAMP).

Supplementary Figure S3. SDS-PAGE analysis of protein purification fractions obtained after Ni2+-NTA affinity chromatography of 57_3Lys endolysin. The 7th lane, corresponding to the 2nd elution fraction contains a prominent band at approximately 17.55 kDa, corresponding to the predicted molecular weight of 57_3Lys, indicating successful expression and purification. The remaining gel lanes contain: 1 - Lysate of *E. coli* BL21-AI [pET26b_57_3Lys]; 2 - Flow through; 3 - Wash 1; 4 - Wash 2; 5 - Wash 3; 6 - 1st elution fraction; 8 - 3rd elution fraction. The last, 9th lane contains SeeBlue® Pre-Stained Protein Standard marker (kDa). The gel was stained with BlueStain Sensitive Plus.

Supplementary Figure S4. Turbidity reduction assays of EC57 culture by purified 57_3Lys endolysin after 60 minutes of incubation. Turbidity reduction assays were performed in triplicate and the results of each treatment are represented by the mean ± SD.

Supplementary Table S1. Intergenomic distances among phage vB-EcoS_57-3 and reference phages calculated using VIRIDIC. Pairwise intergenomic distances between phage vB-EcoS_57-3 and reference phages from the family *Drexlerviridae*, calculated using VIRIDIC. Values are expressed as genomic distances (%), where lower values indicate higher intergenomic similarity. **(plik dostarczony na nośniku)**

Supplementary Table S2. Orthologous protein groups used for phylogenetic analysis of vB-EcoS_57-3. The first six columns provide information on the alignments and their use in the phylogenetic analysis. Alignment length - length of the protein sequence alignment (number of amino acid positions); Parsimony - number of parsimony-informative sites in the alignment; Singleton - number of singleton sites in the alignment; Constant - number

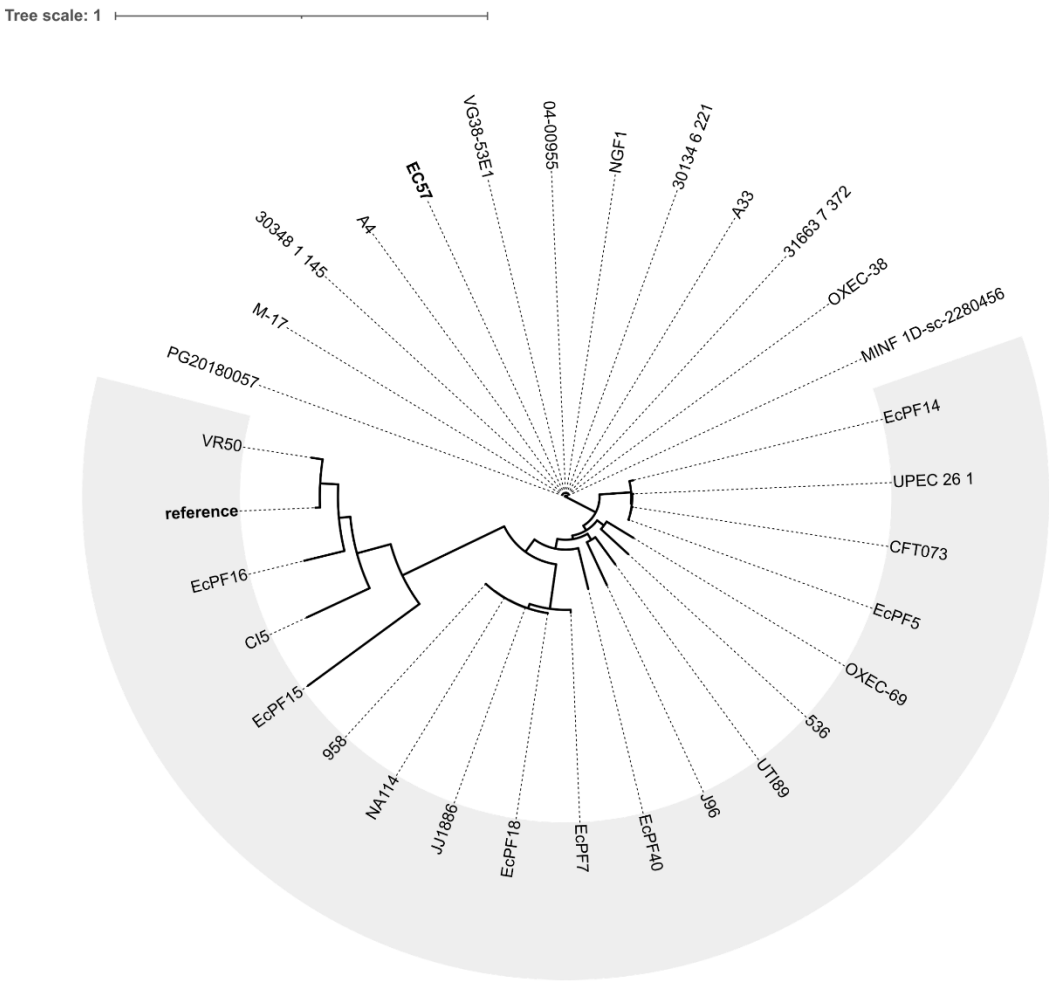
bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

of constant sites in the alignment; Included in phylogeny - indicates whether the corresponding orthologous protein group was included in the phylogenetic analysis, together with the justification for its inclusion (+) or exclusion (-), where applicable; Evolution model - best-fit amino acid substitution model selected for the corresponding alignment partition by ModelFinder Plus implemented in IQ-TREE. The remaining columns correspond to individual phages, with each column header providing the phage name and its GenBank accession number. Cells in these columns contain the annotated protein assigned to the corresponding orthologous group and its protein accession number.

Supplementary Table S3. Physicochemical properties and functional annotation of the endolysin from bacteriophage vB-EcoS_57-3 and orthologous endolysins included in the phylogenetic analysis. The table summarizes the predicted physicochemical properties and functional annotations of the 20 orthologous endolysins included in the phylogenetic analysis (**Figure 11**). For each protein, the phage name, protein accession number, protein length, predicted molecular weight (MW), theoretical isoelectric point (pI), instability index, aliphatic index, and GRAVY value are provided. Functional annotations and sequence signatures identified using InterPro are also presented, including phage lysozyme, endolysin_R21-like, lysozyme-like, endolysin, and SAR-endolysin signatures, if detected.

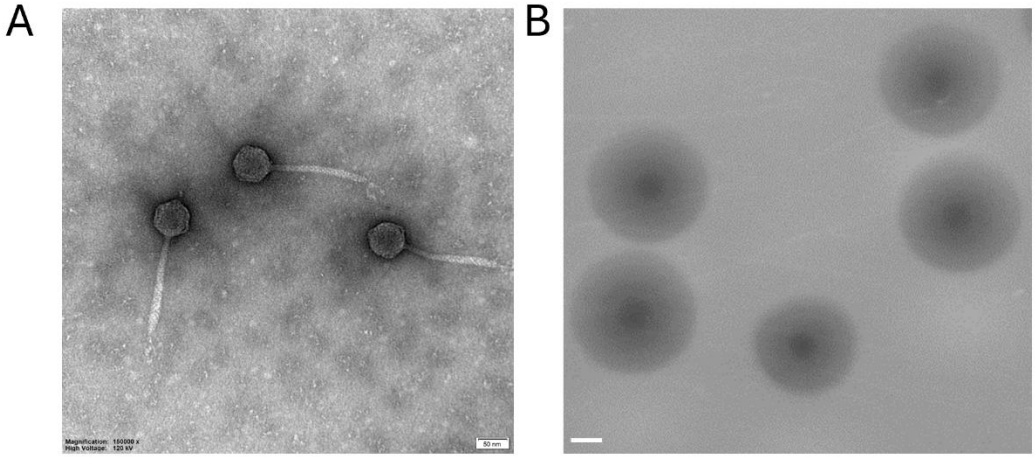
bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Figure 1



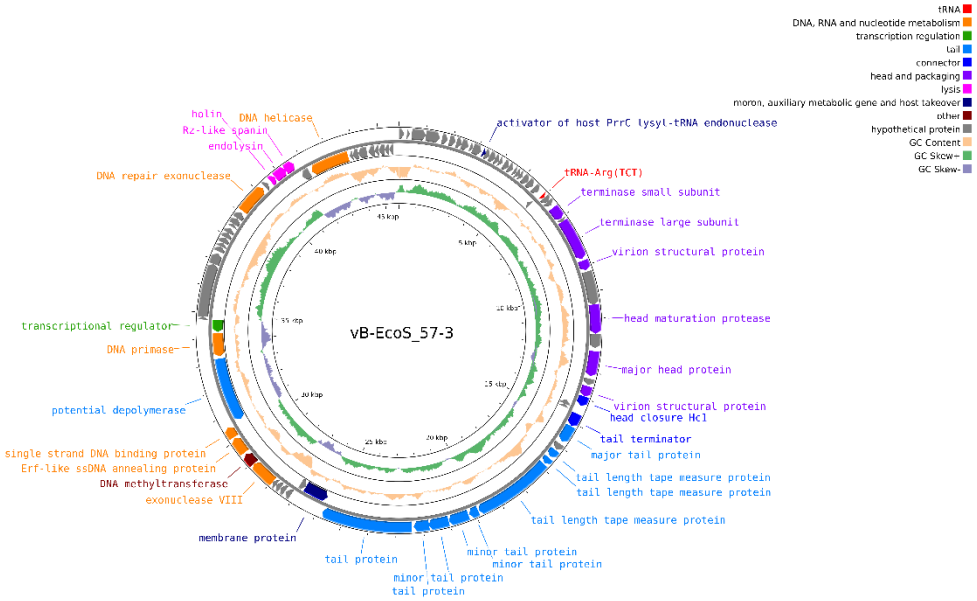
bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Figure 2



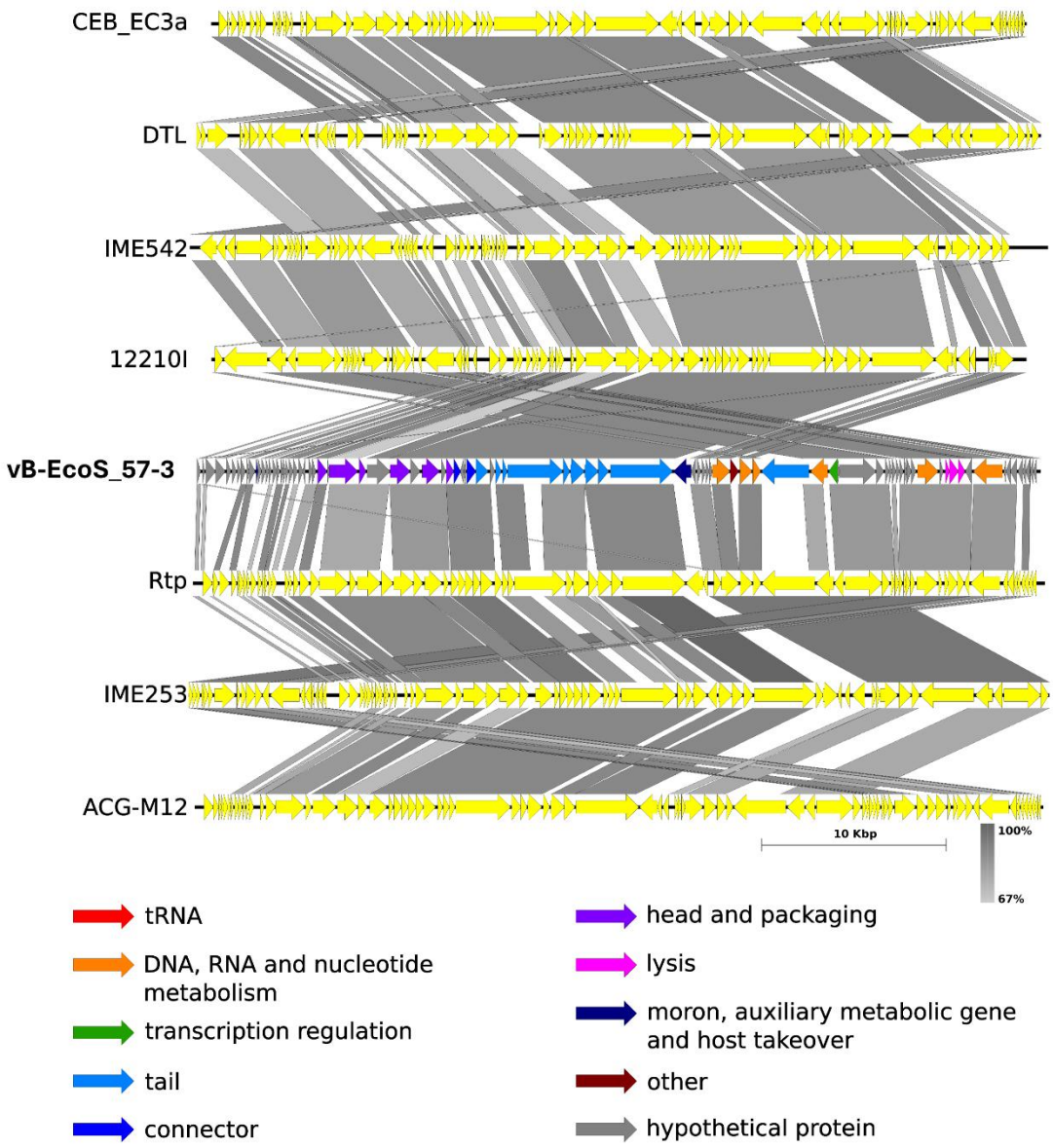
bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Figure 3



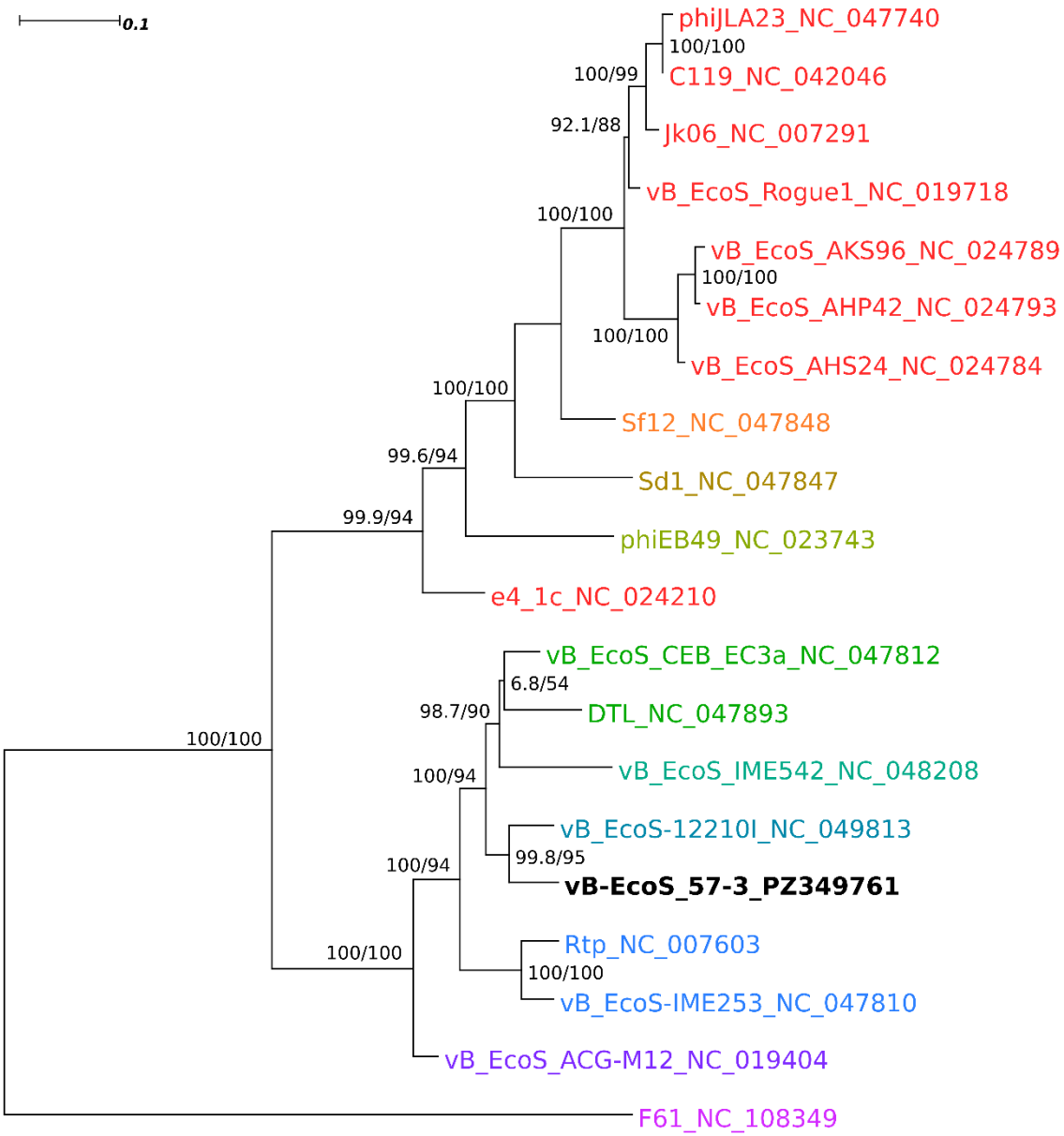
bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Figure 4



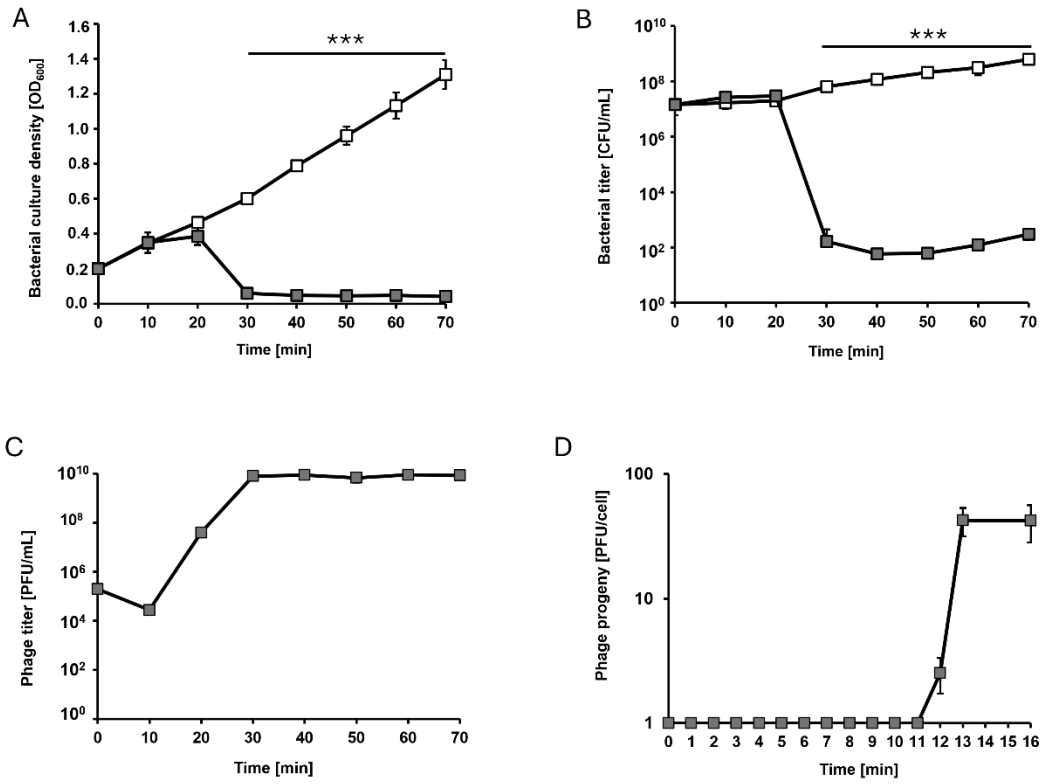
bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Figure 5



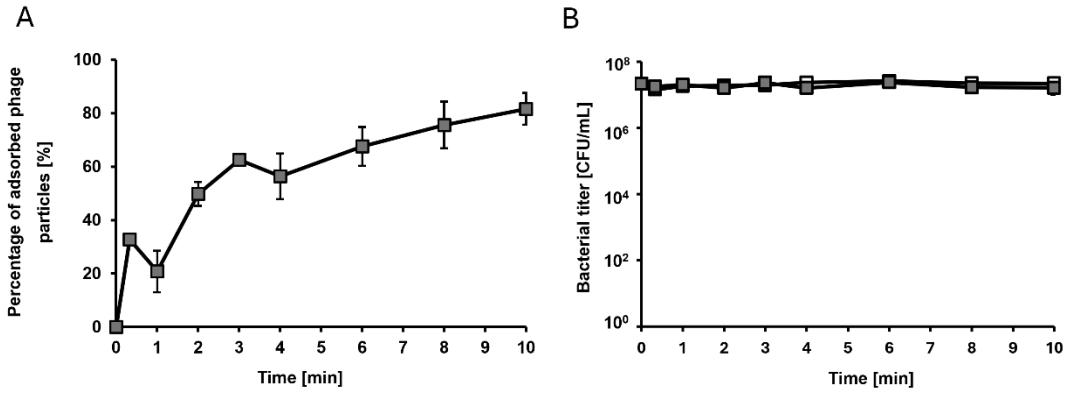
bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Figure 6



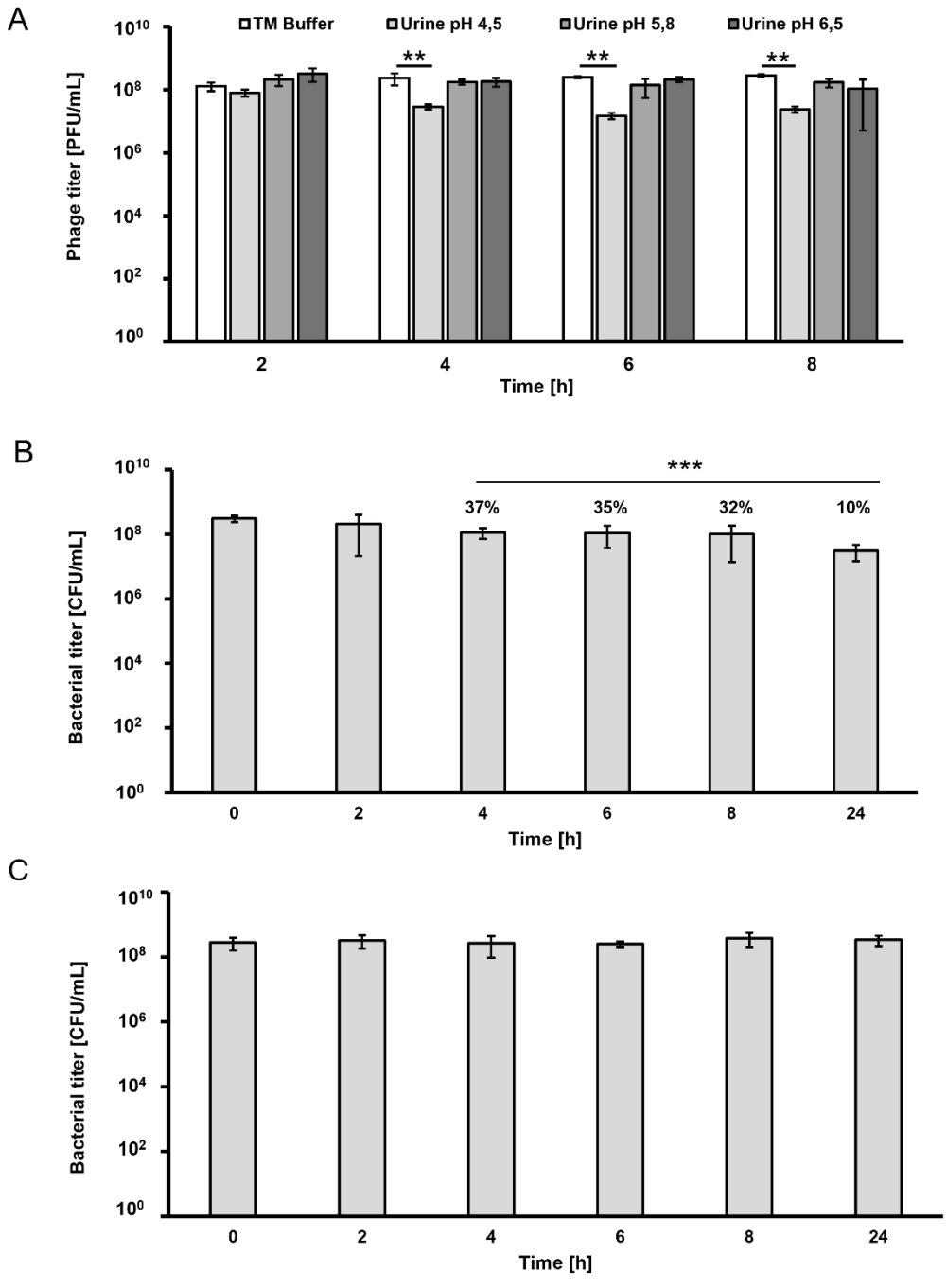
bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Figure 7



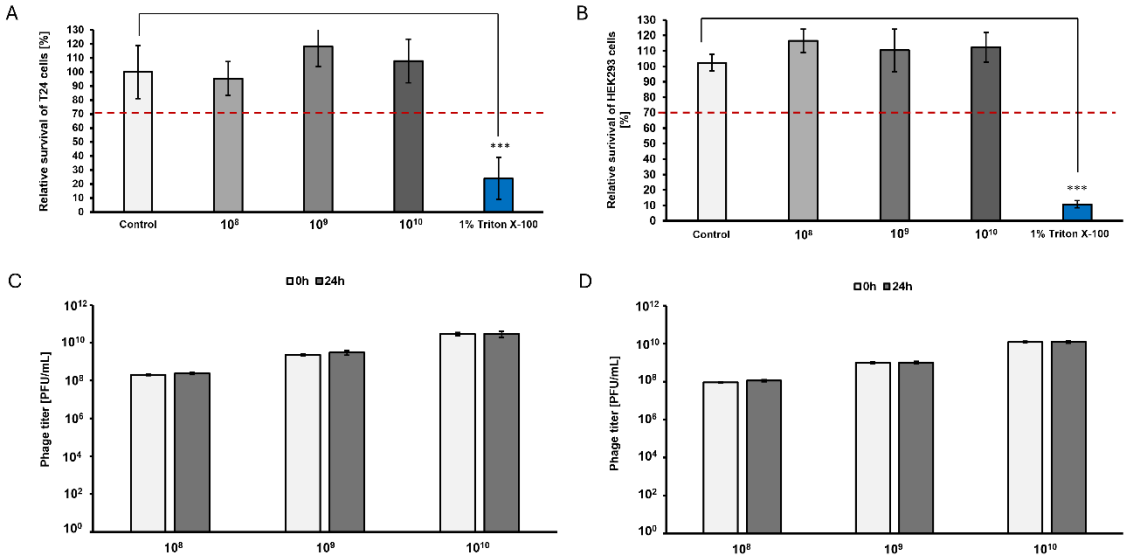
bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Figure 8



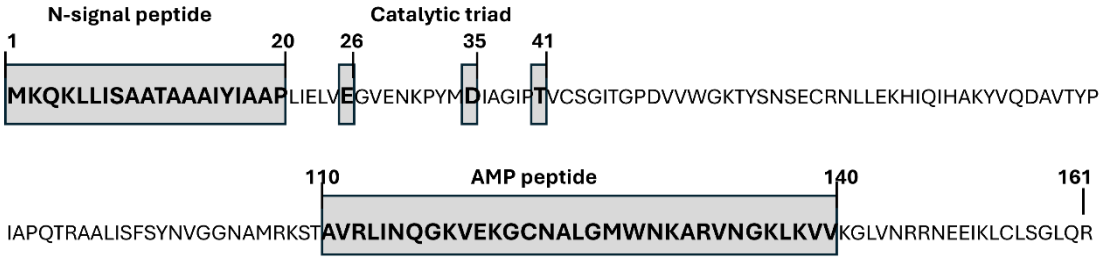
bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Figure 9



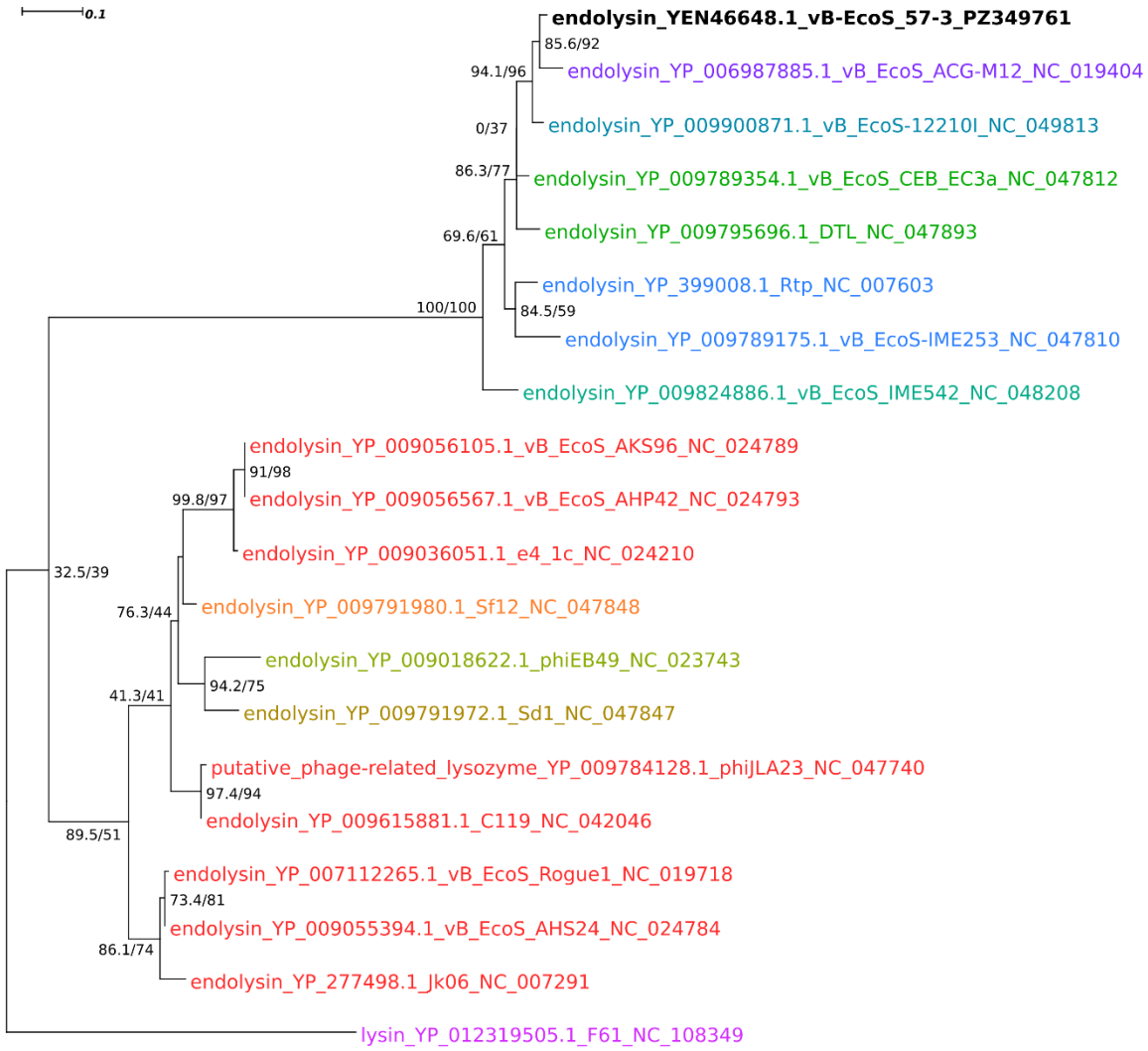
bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Figure 10



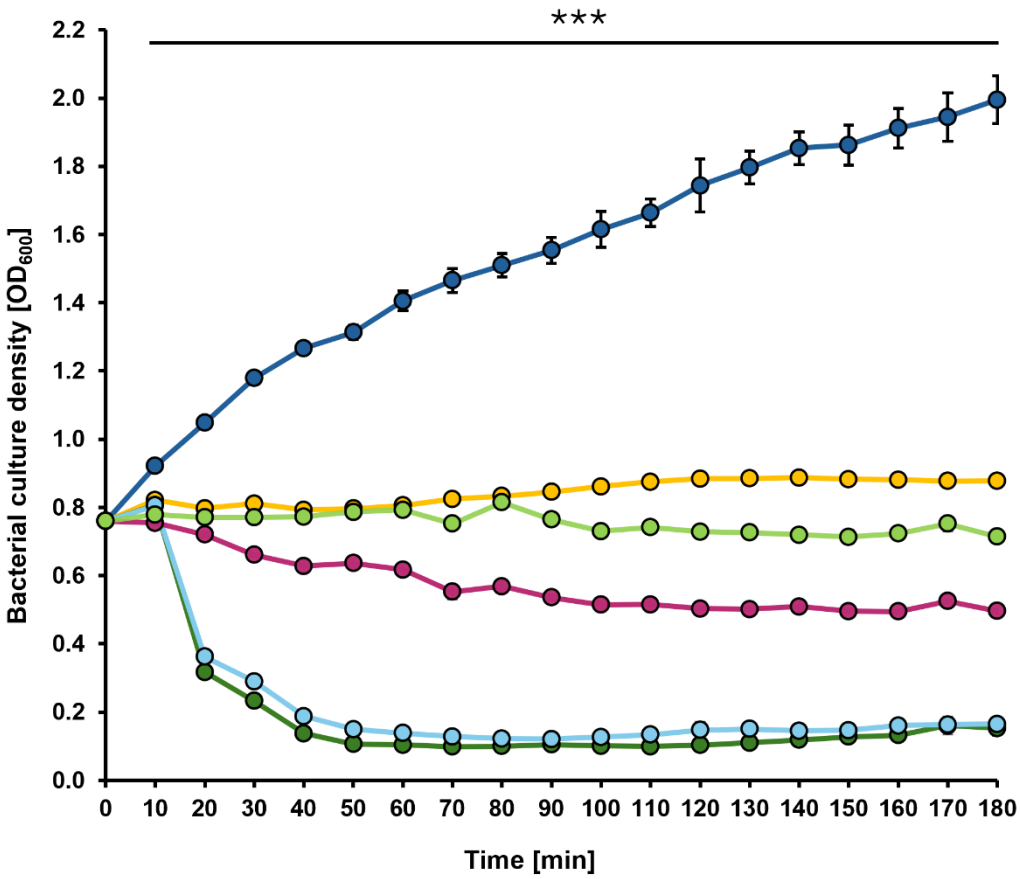
bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Figure 11



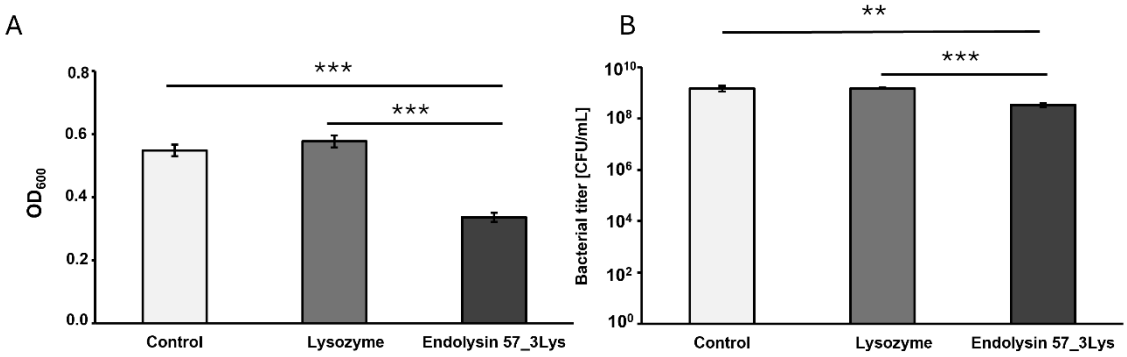
bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Figure 12



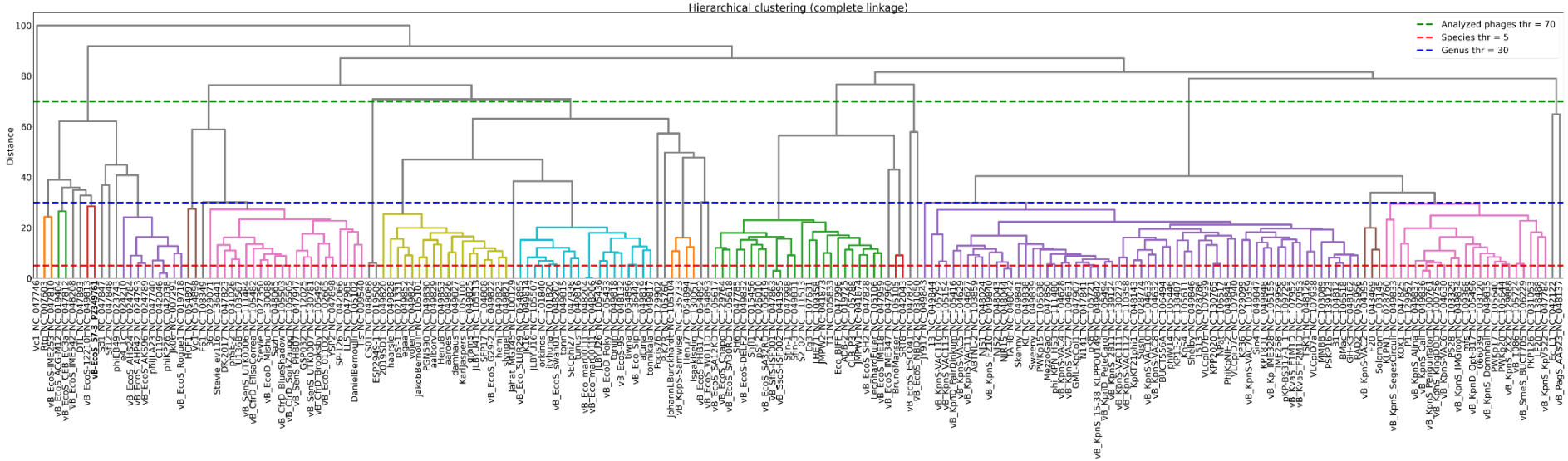
bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Figure 13



Suplement do pracy nr 3

Supplementary Figure S1



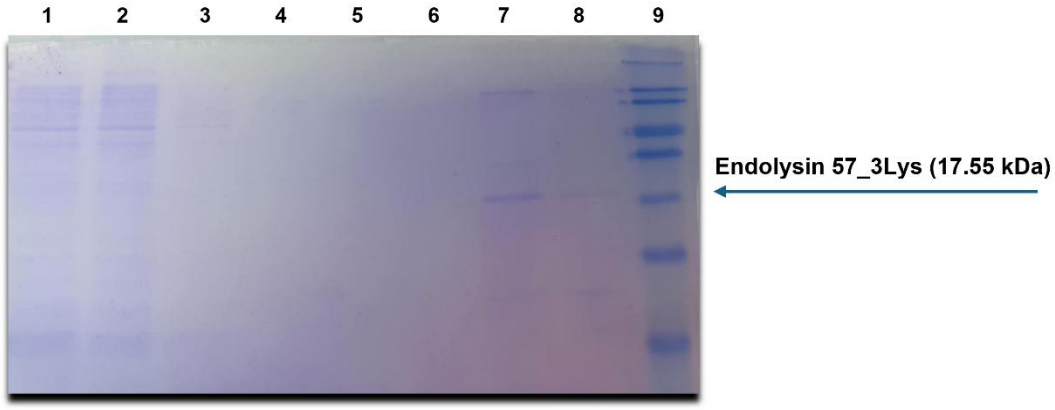
bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Supplementary Figure S2

>DeepTMHMM
SSSSSSSSSSSSSSSSSSSSSOOO

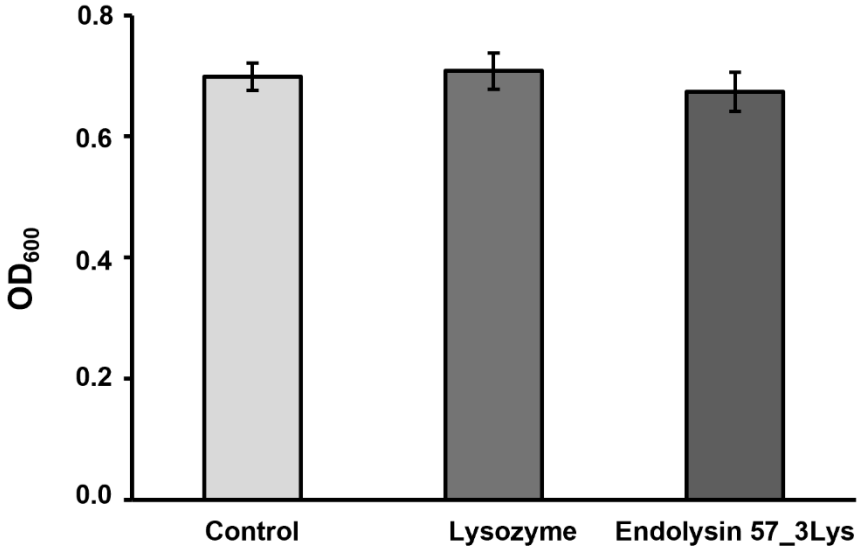
bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Supplementary Figure S3



bioRxiv preprint doi: https://doi.org/10.64898/2026.09.09.750462; this version posted September 10, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Supplementary Figure S4



Supplementary Table S2. Orthologous protein groups used for phylogenetic analysis of vB-EcoS_57-3.

Alignment length	Parsimony	Singleton	Constant	Included in phylogeny	Evolution model	Sf12_NC_047848	phiEB49_NC_023743	vB_EcoS_IME542_NC_048208
302	201	44	57	+	VT+G4	DNA_primase YP_009791997.1	DNA_primase YP_009018637.1	DNA_primase YP_009824872.1
313	128	73	112	+	Q.PFAM+G4	major_head_protein YP_009792027.1	major_head_protein YP_009018666.1	major_head_protein YP_009824918.1
159	93	19	47	+	WAG+G4	endolysin YP_009791980.1	endolysin YP_009018622.1	endolysin YP_009824886.1
131	79	29	23	+	VT	virion_structural_protein YP_009792025.1	virion_structural_protein YP_009018663.1	virion_structural_protein YP_009824920.1
137	68	30	39	+	LG+G4	putative_single-stranded_DNA_binding_protein YP_009791999.1	gp26 YP_009018640.1	hypothetical_protein YP_009824942.1
248	64	78	106	+	LG+G4	minor_tail_protein YP_009792012.1	minor_tail_protein YP_009018650.1	minor_tail_protein YP_009824932.1
116	60	26	30	+	VT+G4	hypothetical_protein YP_009791976.1	gp4 YP_009018618.1	hypothetical_protein YP_009824890.1
187	58	48	81	+	LG+G4	tail_protein YP_009792011.1	tail_protein YP_009018649.1	tail_protein YP_009824933.1
128	53	27	48	+	WAG+I	tail_terminator YP_009792022.1	tail_terminator YP_009018660.1	tail_terminator YP_009824923.1
115	44	37	34	+	LG+R3	minor_tail_protein YP_009792017.1	minor_tail_protein YP_009018653.1	minor_tail_protein YP_009824930.1
104	44	20	40	+	LG	tail_length_tape_measure_protein YP_009792020.1	tail_length_tape_measure_protein YP_009018657.1	tail_length_tape_measure_protein YP_009824926.1

The first six columns provide information on the alignments and their use in the phylogenetic analysis. Alignment length – length of the protein sequence alignment (number of amino acid positions); Parsimony – number of parsimony-informative sites in the alignment; Singleton – number of singleton sites in the alignment; Constant – number of constant sites in the alignment; Included in phylogeny – indicates whether the corresponding orthologous protein group was included in the phylogenetic analysis, together with the justification for its inclusion (+) or exclusion (-), where applicable; Evolution model – best-fit amino acid substitution model selected for the corresponding alignment partition by ModelFinder Plus implemented in IQ-TREE. The remaining columns correspond to individual phages, with each column header providing the phage name and its GenBank accession number. Cells in these columns contain the annotated protein assigned to the corresponding orthologous group and its protein accession number.

vB_EcoS_CEB_EC3a_NC_047812	vB_EcoS_AKS96_NC_024789	DTL_NC_047893	e4_1c_NC_024210	phiJLA23_NC_047740	C119_NC_042046
DNA_primase YP_009789341.1	DNA_primase YP_009056088.1	DNA_primase YP_009795743.1	DNA_primase YP_009036036.1	DNA_primase YP_009784144.1	DNA_primase YP_009615865.1
major_head_protein YP_009789317.1	major_head_protein YP_009056064.1	major_head_protein YP_009795719.1	major_head_protein YP_009036011.1	major_head_protein YP_009784102.1	major_head_protein YP_009615839.1
endolysin YP_009789354.1	endolysin YP_009056105.1	endolysin YP_009795696.1	endolysin YP_009036051.1	putative_phage-related_lysozyme YP_009784128.1	endolysin YP_009615881.1
virion_structural_protein YP_009789319.1	virion_structural_protein YP_009056066.1	virion_structural_protein YP_009795721.1	virion_structural_protein YP_009036013.1	virion_structural_protein YP_009784100.1	virion_structural_protein YP_009615842.1
single-stranded_DNA_binding_protein YP_009789339.1	single-stranded_DNA_binding_protein_Ssb YP_009056086.1	hypothetical_protein YP_009795741.1	hypothetical_protein YP_009036034.1	hypothetical_protein YP_009784146.1	single-stranded_DNA_binding_protein YP_009615863.1
minor_tail_protein YP_009789330.1	minor_tail_protein YP_009056076.1	minor_tail_protein YP_009795732.1	minor_tail_protein YP_009036023.1	minor_tail_protein YP_009784090.1	minor_tail_protein YP_009615853.1
hypothetical_protein YP_009789358.1	hypothetical_protein YP_009056112.1	hypothetical_protein YP_009795700.1	hypothetical_protein YP_009036057.1	hypothetical_protein YP_009784122.1	hypothetical_protein YP_009615887.1
tail_protein YP_009789331.1	tail_protein YP_009056077.1	tail_protein YP_009795733.1	tail_protein YP_009036024.1	tail_protein YP_009784089.1	tail_protein YP_009615854.1
tail_terminator YP_009789322.1	tail_terminator YP_009056069.1	tail_terminator YP_009795724.1	tail_terminator YP_009036016.1	tail_terminator YP_009784097.1	tail_terminator YP_009615845.1
minor_tail_protein YP_009789328.1	minor_tail_protein YP_009056074.1	minor_tail_protein YP_009795730.1	minor_tail_protein YP_009036021.1	minor_tail_protein YP_009784092.1	minor_tail_protein YP_009615850.1
tail_length_tape_measure_protein YP_009789325.1	tail_length_tape_measure_protein YP_009056071.1	tail_length_tape_measure_protein YP_009795727.1	tail_length_tape_measure_protein YP_009036018.1	tail_length_tape_measure_protein YP_009784095.1	tail_length_tape_measure_protein YP_009615847.1

vB_EcoS-12210I_NC_049813	Rtp_NC_007603	Sd1_NC_047847	vB_EcoS_AHP42_NC_024793	vB_EcoS_Rogue1_NC_019718	Jk06_NC_007291
DNA_primase YP_009900883.1	DNA_primase YP_398995.1	DNA_primase YP_009791914.1	DNA_primase YP_009056551.1	DNA_primase YP_007112249.1	DNA_primase YP_277514.1
major_head_protein YP_009900843.1	major_head_protein YP_398971.1	major_head_protein YP_009791942.1	major_head_protein YP_009056526.1	major_head_protein YP_007112224.1	major_head_protein YP_277465.1
endolysin YP_009900871.1	endolysin YP_399008.1	endolysin YP_009791972.1	endolysin YP_009056567.1	endolysin YP_007112265.1	endolysin YP_277498.1
virion_structural_protein YP_009900841.1	virion_structural_protein YP_398973.1	virion_structural_protein YP_009791939.1	virion_structural_protein YP_009056528.1	virion_structural_protein YP_007112227.1	virion_structural_protein YP_277463.1
hypothetical_protein YP_009900885.1	putative_single-stranded_DNA_binding_protein YP_398993.1	nucleic_acid_binding_protein YP_009791916.1	single-stranded_DNA_binding_protein_Ssb YP_009056549.1	putative_single-stranded_DNA_binding_protein_Ssb YP_007112247.1	hypothetical_protein YP_277517.1
minor_tail_protein YP_009900830.1	minor_tail_protein YP_398985.1	minor_tail_protein YP_009791929.1	minor_tail_protein YP_009056539.1	minor_tail_protein YP_007112238.1	minor_tail_protein YP_277445.1
hypothetical_protein YP_009900866.1	hypothetical_protein YP_399013.1	hypothetical_protein YP_009791967.1	hypothetical_protein YP_009056574.1	hypothetical_protein YP_007112271.1	hypothetical_protein YP_277490.1
tail_protein YP_009900829.1	tail_protein YP_398986.1	tail_protein YP_009791928.1	tail_protein YP_009056540.1	tail_protein YP_007112239.1	tail_protein YP_277443.1
tail_terminator YP_009900838.1	tail_terminator YP_398976.1	tail_terminator YP_009791936.1	tail_terminator YP_009056531.1	tail_terminator YP_007112230.1	tail_terminator YP_277458.1
minor_tail_protein YP_009900832.1	minor_tail_protein YP_398982.1	minor_tail_protein YP_009791931.1	minor_tail_protein YP_009056536.1	minor_tail_protein YP_007112235.1	minor_tail_protein YP_277448.1
tail_length_tape_measure_protein YP_009900835.1	tail_length_tape_measure_protein YP_398979.1	tail_length_tape_measure_protein YP_009791934.1	tail_length_tape_measure_protein YP_009056533.1	tail_length_tape_measure_protein YP_007112232.1	tail_length_tape_measure_protein YP_277456.1

vB-EcoS_57-3_PZ349761	vB_EcoS_AHS24_NC_024784	vB_EcoS-IME253_NC_047810	vB_EcoS_ACG-M12_NC_019404	F61_NC_108349
DNA_primase YEN46632.1	DNA_primase YP_009055377.1	DNA_primase YP_009789236.1	DNA_primase YP_006987867.1	DNA_primase/helicase YP_012319516.1
major_head_protein YEN46604.1	major_head_protein YP_009055350.1	major_head_protein YP_009789206.1	major_head_protein YP_006987838.1	hypothetical_protein YP_012319467.1
endolysin YEN46648.1	endolysin YP_009055394.1	endolysin YP_009789175.1	endolysin YP_006987885.1	lysin YP_012319505.1
virion_structural_protein YEN46606.1	virion_structural_protein YP_009055352.1	virion_structural_protein YP_009789208.1	virion_structural_protein YP_006987840.1	hypothetical_protein YP_012319465.1
single_strand_DNA_binding_protein YEN46630.1	single-stranded_DNA_binding_protein_Ssb YP_009055375.1	hypothetical_protein YP_009789234.1	putative_single-stranded_DNA_binding_protein YP_006987864.1	single-stranded_DNA-binding_protein YP_012319448.1
minor_tail_protein YEN46618.1	minor_tail_protein YP_009055363.1	minor_tail_protein YP_009789221.1	minor_tail_protein YP_006987853.1	tail_tip_assembly_protein_K YP_012319455.1
hypothetical_protein YEN46654.1	hypothetical_protein YP_009055401.1	hypothetical_protein YP_009789180.1	hypothetical_protein YP_006987889.1	hypothetical_protein YP_012319499.1
tail_protein YEN46619.1	tail_protein YP_009055364.1	tail_protein YP_009789222.1	tail_protein YP_006987854.1	tail_tip_assembly_protein_I YP_012319454.1
tail_terminator YEN46610.1	tail_terminator YP_009055355.1	tail_terminator YP_009789211.1	tail_terminator YP_006987843.1	hypothetical_protein YP_012319462.1
minor_tail_protein YEN46616.1	minor_tail_protein YP_009055360.1	minor_tail_protein YP_009789218.1	minor_tail_protein YP_006987850.1	minor_tail_protein YP_012319457.1
tail_length_tape_measure_protein YEN46613.1	tail_length_tape_measure_protein YP_009055357.1	tail_length_tape_measure_protein YP_009789214.1	tail_length_tape_measure_protein YP_006987846.1	hypothetical_protein YP_012319460.1

Supplementary Table S3

Analysis of orthologically common endolysins.

Phage name with accession	Protein name	Protein accession number	Protein length (aa)	Molecular weight (kDa)	Instability index	Isoelectric point (pI)	Gravy	Aliphatic index	Name of the predicted domain	Position of the domain in protein sequence
vB-EcoS_57-3_PZ349761	endolysin	YEN46648.1	161	17,55	32,544	9,7	-0,057	101,18	Endolysin. (MF_04110) endolysin_R21-like (cd16900) Lysozyme-like (SSF53955) SAR-endolysin. (MF_04136) Phage lysozyme (PF00959)	4-158 26-156 6-159 3-158 25-151
vB_EcoS_ACG-M12_NC_019404	endolysin	YP_006987885.1	161	17,63	33,838	9,83	-0,108	99,938	endolysin_R21-like (cd16900) Phage lysozyme (PF00959) Lysozyme-like (SSF53955) Endolysin. (MF_04110) SAR-endolysin. (MF_04136)	20-156 25-151 6-159 4-158 3-158
vB_EcoS_Rogue1_NC_019718	endolysin	YP_007112265.1	160	17,26	17,401	9,46	0,068	97,5	Endolysin. (MF_04110) Lysozyme-like (SSF53955) Phage lysozyme (PF00959) endolysin_R21-like (cd16900) SAR-endolysin. (MF_04136)	5-159 6-159 25-152 13-156 4-159
phiEB49_NC_023743	endolysin	YP_009018622.1	160	17,53	28,014	9,61	-0,032	95,562	Endolysin. (MF_04110) endolysin_R21-like (cd16900) SAR-endolysin. (MF_04136) Phage lysozyme (PF00959) Lysozyme-like (SSF53955)	5-159 13-156 4-159 25-152 8-159
e4_1c_NC_024210	endolysin	YP_009036051.1	160	17,41	22,44	9,88	0,06	99,875	Phage lysozyme (PF00959) endolysin_R21-like (cd16900) SAR-endolysin. (MF_04136) Endolysin. (MF_04110) Lysozyme-like (SSF53955)	25-152 13-156 4-159 5-159 6-159
vB_EcoS_AHS24_NC_024784	endolysin	YP_009055394.1	160	17,32	14,654	9,51	0,044	97,5	Phage lysozyme (PF00959) SAR-endolysin. (MF_04136) Endolysin. (MF_04110) endolysin_R21-like (cd16900) Lysozyme-like (SSF53955)	25-152 4-159 5-159 13-156 6-159
vB_EcoS_AKS96_NC_024789	endolysin	YP_009056105.1	160	17,33	23,295	9,83	0,086	101,062	endolysin_R21-like (cd16900) Phage lysozyme (PF00959) SAR-endolysin. (MF_04136) Endolysin. (MF_04110) Lysozyme-like (SSF53955)	13-156 25-152 8-159 5-159 7-159
vB_EcoS_AHP42_NC_024793	endolysin	YP_009056567.1	160	17,33	23,295	9,83	0,086	101,062	endolysin_R21-like (cd16900) Phage lysozyme (PF00959) SAR-endolysin. (MF_04136) Endolysin. (MF_04110) Lysozyme-like (SSF53955)	13-156 25-152 8-159 5-159 7-159

C119_NC_042046	endolysin	YP_009615881.1	161	17,58	21,594	10,07	-0,002	94,472	SAR-endolysin. (MF_04136) Endolysin. (MF_04110) Phage lysozyme (PF00959) Lysozyme-like (SSF53955) endolysin_R21-like (cd16900)	5-160 6-160 26-153 7-160 14-157
phiJLA23_NC_047740	putative_phage-related_lysozyme	YP_009784128.1	131	14,3	17,415	9,76	0,007	89,389	Lysozyme-like (SSF53955) Phage lysozyme (PF00959) Endolysin. (MF_04110) endolysin_R21-like (cd16900)	7-129 26-129 6-131 14-130
vB_EcoS-IME253_NC_047810	endolysin	YP_009789175.1	161	17,49	35,15	9,76	-0,078	99,876	endolysin_R21-like (cd16900) SAR-endolysin. (MF_04136) Phage lysozyme (PF00959) Lysozyme-like (SSF53955) Endolysin. (MF_04110)	20-156 3-158 25-151 5-159 4-158
vB_EcoS_CEB_EC3a_NC_047812	endolysin	YP_009789354.1	161	17,51	36,793	9,61	0,01	104,161	Phage lysozyme (PF00959) endolysin_R21-like (cd16900) Endolysin. (MF_04110) SAR-endolysin. (MF_04136) Lysozyme-like (SSF53955)	25-151 12-156 4-158 3-158 6-159
Sd1_NC_047847	endolysin	YP_009791972.1	161	17,44	30,394	9,66	0,091	98,696	SAR-endolysin. (MF_04136) Phage lysozyme (PF00959) Endolysin. (MF_04110) endolysin_R21-like (cd16900) Lysozyme-like (SSF53955)	5-160 26-153 6-160 14-157 6-160
Sf12_NC_047848	endolysin	YP_009791980.1	160	17,44	21,849	9,87	-0,006	93,812	Phage lysozyme (PF00959) endolysin_R21-like (cd16900) Lysozyme-like (SSF53955) SAR-endolysin. (MF_04136) Endolysin. (MF_04110)	25-152 13-156 8-159 4-159 5-159
DTL_NC_047893	endolysin	YP_009795696.1	161	17,52	41,81	9,61	-0,052	102,36	Lysozyme-like (SSF53955) endolysin_R21-like (cd16900) Endolysin. (MF_04110) Phage lysozyme (PF00959) SAR-endolysin. (MF_04136)	6-159 26-156 4-158 25-151 3-158
vB_EcoS_IME542_NC_048208	endolysin	YP_009824886.1	161	17,6	34,935	9,62	-0,081	95,714	endolysin_R21-like (cd16900) SAR-endolysin. (MF_04136) Lysozyme-like (SSF53955) Endolysin. (MF_04110) Phage lysozyme (PF00959)	26-156 3-158 6-159 4-158 25-151
vB_EcoS-12210I_NC_049813	endolysin	YP_009900871.1	161	17,62	35,273	9,78	-0,075	101,18	endolysin_R21-like (cd16900) Phage lysozyme (PF00959) Lysozyme-like (SSF53955) SAR-endolysin. (MF_04136) Endolysin. (MF_04110)	20-156 25-151 6-159 3-158 4-158

F61_NC_108349	lysin	YP_012319505.1	164	18,17	16,04	9,62	-0,248	82,744	Endolysin. (MF_04110) Lysozyme-like (SSF53955) Phage lysozyme (PF00959) endolysin_R21-like (cd16900) SAR-endolysin. (MF_04136)	5-161 12-162 25-154 13-158 4-161
Jk06_NC_007291	endolysin	YP_277498.1	160	17,39	17,571	9,63	-0,051	93,812	Lysozyme-like (SSF53955) endolysin_R21-like (cd16900) SAR-endolysin. (MF_04136) Endolysin. (MF_04110) Phage lysozyme (PF00959)	6-159 13-156 4-159 5-159 25-152
Rtp_NC_007603	endolysin	YP_399008.1	161	17,49	34,328	9,7	-0,061	100,497	endolysin_R21-like (cd16900) Endolysin. (MF_04110) Lysozyme-like (SSF53955) SAR-endolysin. (MF_04136) Phage lysozyme (PF00959)	20-156 4-158 5-159 3-158 25-151

Oświadczenia współautorów

Wojciech Wesołowski
.....
IMIĘ I NAZWISKO

28.09.2026
.....
DATA

Katedra Biologii Molekularnej,
Wydział Biologii,
Uniwersytet Gdański
.....
AFILIACJA

OŚWIADCZENIE

Niniejszym oświadczam, iż mój wkład w manuskrypt:

Wesołowski W, Bloch S, Czerwonka G, Jagieła E, Grabowski Ł, Jeschke M, Jacewicz J, Piwnicka E, Morcinek-Orłowska J, Loika H, Czaplewska P, Węgrzyn G, Adamus-Białek W, Łukasiak A, Nejman-Faleńczyk B. Antibacterial activity of a newly identified phage-derived endolysin and its parental bacteriophage against clinical uropathogenic *Escherichia coli*,

polegał na:

- przygotowaniu, zagęszczeniu i oczyszczeniu wirionów faga vB-EcoS_57-3 z wykorzystaniem ultrawirowania w gradiencie chlorku cezu celem wykonania zdjęć oraz przeprowadzeniu analizy na obecność zanieczyszczeń w postaci endotoksyn (Rycina nr 2),
- przygotowaniu, rozcieńczeniu i zmiareczkowaniu lizatu faga vB-EcoS_57-3 celem wykonania zdjęć łysinek (Rycina nr 2),
- złożeniu i wykonaniu funkcjonalnej adnotacji genomu faga vB-EcoS_57-3, przygotowaniu na jej podstawie mapy kołowej genomu (Rycina nr 3) oraz zdeponowaniu fagowej sekwencji genomowej w bazie GenBank,
- przeprowadzeniu analizy porównawczej organizacji genomu vB-EcoS_57-3 oraz fagów blisko z nim spokrewnionych (Rycina nr 4),
- wykonaniu analizy filogenetycznej i wygenerowaniu drzew (Rycina S1, Rycina 5 i Rycina 11),
- współudziale w zbadaniu kinetyki lizy faga i jego rozwoju wewnątrzkomórkowego z określeniem plonu fagowego w doświadczeniu „one step growth" (Rycina nr 6),
- współudziale w zbadaniu wydajności adsorpcji faga do komórek bakteryjnych (Rycina nr 7),
- określeniu wydajności infekcji fagiem vB-EcoS_57-3 klinicznych izolatów UPEC (Tabela 2),
- współudziale w analizie stabilności wirionów faga vB-EcoS_57-3 w obecności różnych czynników chemicznych i fizycznych (Tabela nr 3),
- zidentyfikowaniu genu endolizyny 57_3Lys w genomie faga i wykonaniu analizy *in silico* sekwencji tego białka, która umożliwiła m.in. określenie pozycji peptydu sygnałowego, peptydu AMP oraz triady katalitycznej (Tabela nr 4, Rycina nr 10, Tabela S2),
- współudziale w przygotowaniu konstruktu do nadprodukcji endolizyny 57_3Lys oraz jej oczyszczeniu na złożu niklowym (Rycina nr S3) i zbadaniu jej aktywności (Rycina nr 13 oraz Rycina nr S4),
- wykazaniu zależności endolizyny 57_3Lys od systemu translokacji Sec (Rycina nr 12),
- współudziale w przygotowaniu odpowiedzi na uwagi recenzentów,
- a także na obróbce i interpretacji wyników oraz przygotowaniu rycin i tabel.

Wesołowski
.....
PODPIS

Sylwia Bloch
.....
IMIĘ I NAZWISKO

08.09.2026 r.
.....
DATA

Katedra Biologii Molekularnej,
Wydział Biologii,
Uniwersytet Gdański

.....
AFILIACJA

OŚWIADCZENIE

Niniejszym oświadczam, iż mój wkład w manuskrypt:

Wesołowski W, Bloch S, Czerwonka G, Jagieła E, Grabowski Ł, Jeschke M, Jacewicz J, Piwnicka E, Morcinek-Orłowska J, Loika H, Czaplewska P, Węgrzyn G, Adamus-Białek W, Łukasiak A, Nejman-Faleńczyk B. Antibacterial activity of a newly identified phage-derived endolysin and its parental bacteriophage against clinical uropathogenic *Escherichia coli*,

polegał na:

(i) współtworzeniu i opisaniu metodologii;

(ii) współudziale w nadzorowaniu przebiegu realizacji prac eksperymentalnych;

(ii) współudziale w redagowaniu ostatecznej wersji manuskryptu.

Bloch
.....
PODPIS

Grzegorz Czerwonka
.....
IMIĘ I NAZWISKO

14.08.2026
DATA

Zakład Mikrobiologii,
Instytut Biologii,
Uniwersytet Jana Kochanowskiego w Kielcach
.....
AFILIACJA

OŚWIADCZENIE

Niniejszym oświadczam, iż mój wkład w manuskrypt:

Wesołowski W, Bloch S, Czerwonka G, Jagieła E, Grabowski Ł, Jeschke M, Jacewicz J, Piwnicka E, Morcinek-Orłowska J, Loika H, Czaplewska P, Węgrzyn G, Adamus-Białek W, Łukasiak A, Nejman-Faleńczyk B. Antibacterial activity of a newly identified phage-derived endolysin and its parental bacteriophage against clinical uropathogenic *Escherichia coli*,

polegał na:

gromadzeniu, opracowywaniu i weryfikacji danych badawczych (Data Curation), przeprowadzaniu analiz formalnych oraz interpretacji uzyskanych wyników (Formal Analysis), realizacji prac badawczych i eksperymentalnych (Investigation), a także współtworzeniu metodologii badań (Methodology). Ponadto uczestniczyłem w przygotowaniu pierwszej wersji manuskryptu (Writing – Original Draft) oraz w jego merytorycznej recenzji, redakcji i korekcie (Writing – Review and Editing).

Grzegorz Czerwonka

PODPIS

Ernest Jagieła
.....
IMIĘ I NAZWISKO

Kielce, 24.08.2026 r.
.....
DATA

Zakład Mikrobiologii,
Instytut Biologii,
Uniwersytet Jana Kochanowskiego w Kielcach
.....
AFILIACJA

OŚWIADCZENIE

Niniejszym oświadczam, iż mój wkład w manuskrypt:

Wesołowski W, Bloch S, Czerwonka G, Jagieła E, Grabowski Ł, Jeschke M, Jacewicz J, Piwnicka E, Morcinek-Orłowska J, Loika H, Czaplewska P, Węgrzyn G, Adamus-Białek W, Łukasiak A, Nejman-Faleńczyk B. Antibacterial activity of a newly identified phage-derived endolysin and its parental bacteriophage against clinical uropathogenic *Escherichia coli*,

polegał na:

Udziale w analizach dotyczących charakterystyki genetycznej szczepu *E. Coli.*

.....
PODPIS

Łukasz Grabowski
.....
IMIĘ I NAZWISKO

08.09.26
.....
DATA

Katedra Biologii Molekularnej,
Wydział Biologii,
Uniwersytet Gdański

.....
AFILIACJA

OŚWIADCZENIE

Niniejszym oświadczam, iż mój wkład w manuskrypt:

Wesołowski W, Bloch S, Czerwonka G, Jagieła E, Grabowski Ł, Jeschke M, Jacewicz J, Piwnicka E, Morcinek-Orłowska J, Loika H, Czaplewska P, Węgrzyn G, Adamus-Białek W, Łukasiak A, Nejman-Faleńczyk B. Antibacterial activity of a newly identified phage-derived endolysin and its parental bacteriophage against clinical uropathogenic *Escherichia coli*,

polegał na współudziale w przeprowadzeniu doświadczenia mającego na celu zbadania poziomu toksyczności lizatu fagowego wobec komórek eukariotycznych.

Ł. Grabowski
.....
PODPIS

Marta Jeschke
.....
IMIĘ I NAZWISKO

10.08.2026
.....
DATA

Katedra Biologii Molekularnej,
Wydział Biologii,
Uniwersytet Gdański

.....
AFILIACJA

OŚWIADCZENIE

Niniejszym oświadczam, iż mój wkład w manuskrypt:

Wesołowski W, Bloch S, Czerwonka G, Jagieła E, Grabowski Ł, Jeschke M, Jacewicz J, Piwnicka E, Morcinek-Orłowska J, Loika H, Czaplewska P, Węgrzyn G, Adamus-Białek W, Łukasiak A, Nejman-Faleńczyk B. Antibacterial activity of a newly identified phage-derived endolysin and its parental bacteriophage against clinical uropathogenic *Escherichia coli*,

polegał na współudziale w wykonaniu analizy stabilności wirionów faga vB-EcoS_57-3 w sztucznym moczu.

Jeschke Marta
.....
PODPIS

Jakub Jacewicz
.....
IMIĘ I NAZWISKO

30.08.2026
.....
DATA

Katedra Biologii Molekularnej,
Wydział Biologii,
Uniwersytet Gdański

.....
AFILIACJA

OŚWIADCZENIE

Niniejszym oświadczam, iż mój wkład w manuskrypt:

Wesołowski W, Bloch S, Czerwonka G, Jagieła E, Grabowski Ł, Jeschke M, Jacewicz J, Piwnicka E, Morcinek-Orłowska J, Loika H, Czaplewska P, Węgrzyn G, Adamus-Białek W, Łukasiak A, Nejman-Faleńczyk B. Antibacterial activity of a newly identified phage-derived endolysin and its parental bacteriophage against clinical uropathogenic *Escherichia coli*,

polegał na współudziale w przygotowaniu konstruktu do nadprodukcji endolizyny 57_3Lys. Prace te wykonywałem pod bezpośrednią opieką mgr. Wojciecha Wesołowskiego.



.....
PODPIS

KIR Szafir elektronicznypodpis.pl

Raport z weryfikacji podpisu

INFORMACJE O DOKUMENCIE:	
Nazwa pliku	JJ_20praca_20nr_204.pdf
Data weryfikacji podpisów	2026-08-30 20:39:49

SZCZEGÓŁY WERYFIKACJI:

Jakub Jacewicz
.....
IMIĘ I NAZWISKO

30.08.2026
.....
DATA

Katedra Biologii Molekularnej,
Wydział Biologii,
Uniwersytet Gdański

.....
AFILIACJA

OŚWIADCZENIE

Niniejszym oświadczam, iż mój wkład w manuskrypt:

Wesołowski W, Bloch S, Czerwonka G, Jagiełła E, Grabowski Ł, Jeschke M, Jacewicz J, Piwnicka E, Moreinek-Orłowska J, Loika H, Czaplewska P, Węgrzyn G, Adamus-Białek W, Łukasiak A, Nejman-Faleńczyk B. Antibacterial activity of a newly identified phage-derived endolysin and its parental bacteriophage against clinical uropathogenic *Escherichia coli*.

polegał na współudziale w przygotowaniu konstruktu do nadprodukcji endolizyny 57_3Lys. Prace te wykonywałem pod bezpośrednią opieką mgr. Wojciecha Wesołowskiego.

Podpisano przez JACEWICZ JAKUB w dniu 30.08.2026 20:04

PODPIS

Podpis: Pozytywny	
Rodzaj certyfikatu	Kwalifikowany
Format podpisu	PAdES-LTV
Data złożenia podpisu	2026-08-30 20:04:55
Podpis zawiera znacznik czasu	Data znacznika czasu: 2026-08-30 20:04:57 Rodzaj certyfikatu: Kwalifikowany Data ważności certyfikatu: 2031-05-29 13:41:20 Wystawca znacznika czasu: organizationIdentifier=VATPL-5250001090,CN=CUZ Sigillum - QTSU1,OU=Centrum Usług Zaufania Sigillum,O=Polska Wytwórnia Papierów Wartościowych S.A.,C=PL
Znacznik archiwalny dokumentu	Data znacznika czasu: 2026-08-30 20:04:58 Rodzaj certyfikatu: Kwalifikowany Data ważności certyfikatu: 2031-05-29 13:43:53 Wystawca znacznika czasu: organizationIdentifier=VATPL-5250001090,CN=CUZ Sigillum - QTSU2,OU=Centrum Usług Zaufania Sigillum,O=Polska Wytwórnia Papierów Wartościowych S.A.,C=PL
Certyfikat podpisującego / składającego pieczęć	Nazwa powszechna: JACEWICZ JAKUB
Numer seryjny certyfikatu	279276056478171234160913650596935310743353078453
Wystawca certyfikatu	Nazwa powszechna: CUZ Sigillum - QCA2 Nazwa organizacji: Polska Wytwórnia Papierów Wartościowych S.A. Kraj: PL Identyfikator organizacji: VATPL-5250001090
Lista CRL wykorzystana do weryfikacji certyfikatu	-
Odpowiedź OCSP wykorzystana do weryfikacji certyfikatu	Status: GOOD ArchiveCutOff: 14 Mar 2026 09:57:20 GMT 30 Aug 2026 18:04:57 GMT
Zweryfikowano na dzień	2026-08-30 20:04:57 (Znacznik czasu)
Status weryfikacji	**Pozytywny**
	Kwalifikowany podpis poprawnie zweryfikowany certyfikatem kwalifikowanym
Uwagi	Weryfikowany podpis został uznany za kwalifikowany. Zgodnie z Art. 25. Punkt 2 rozporządzenia Parlamentu Europejskiego i Rady (UE) nr 910/2014 z 23 lipca 2014 r. kwalifikowany podpis elektroniczny ma skutek prawny równoważny podpisowi własnoręcznemu.

Emilia Piwnicka (Węglińska)
.....
IMIĘ I NAZWISKO

01.09.2026
.....
DATA

Katedra Biologii Molekularnej,
Wydział Biologii,
Uniwersytet Gdański

.....
AFILIACJA

OŚWIADCZENIE

Niniejszym oświadczam, iż mój wkład w manuskrypt:

Wesołowski W, Bloch S, Czerwonka G, Jagieła E, Grabowski Ł, Jeschke M, Jacewicz J, Piwnicka E, Morcinek-Orłowska J, Loika H, Czaplewska P, Węgrzyn G, Adamus-Białek W, Łukasiak A, Nejman-Faleńczyk B. Antibacterial activity of a newly identified phage-derived endolysin and its parental bacteriophage against clinical uropathogenic *Escherichia coli*,

polegał na zebraniu danych literaturowych i przygotowaniu na ich podstawie rozdziału *Introduction.*

Emilia Piwnicka
.....
PODPIS

Joanna Morcinek-Orłowska
.....
IMIĘ I NAZWISKO

Zakład Biologii Strukturalnej,
Międzyuczelniany Wydział Biotechnologii UG i GUMed
.....
AFILIACJA

31/08/2026
.....
DATA

OŚWIADCZENIE

Niniejszym oświadczam, iż mój wkład w manuskrypt:

Wesołowski W, Bloch S, Czerwonka G, Jagieła E, Grabowski Ł, Jeschke M, Jacewicz J, Piwnicka E, Morcinek-Orłowska J, Loika H, Czaplewska P, Węgrzyn G, Adamus-Białek W, Łukasiak A, Nejman-Faleńczyk B. Antibacterial activity of a newly identified phage-derived endolysin and its parental bacteriophage against clinical uropathogenic *Escherichia coli*,

polegał na współudziale w przygotowaniu procedury oczyszczania endolizyny 57_3Lys.

Joanna Morcinek-Orłowska
.....
PODPIS

Hanna Loika
.....
IMIĘ I NAZWISKO

24.08.2026
DATA

Katedra i Zakład Mikrobiologii,
Wydział Lekarski,
Gdański Uniwersytet Medyczny

.....
AFILIACJA

OŚWIADCZENIE

Niniejszym oświadczam, iż mój wkład w manuskrypt:

Wesołowski W, Bloch S, Czerwonka G, Jagieła E, Grabowski Ł, Jeschke M, Jacewicz J, Piwnicka E, Morcinek-Orłowska J, Loika H, Czaplewska P, Węgrzyn G, Adamus-Białek W, Łukasiak A, Nejman-Faleńczyk B. Antibacterial activity of a newly identified phage-derived endolysin and its parental bacteriophage against clinical uropathogenic *Escherichia coli*,

polegał na współudziale w przeprowadzeniu analizy wydajności adsorpcji cząstek bakteriofaga vB-EcoS_57-3 do powierzchni komórek bakteryjnych.

Hanna Loika
.....
PODPIS

Dr hab. Paulina Czaplewska, prof. UG
Laboratorium Spektrometrii Mas
Międzyuczelniany Wydział Biotechnologii UG i GUMed
Uniwersytet Gdański

07.09.2026

OŚWIADCZENIE

Niniejszym oświadczam, iż mój wkład w manuskrypt:

Wesołowski W, Bloch S, Czerwonka G, Jagieła E, Grabowski Ł, Jeschke M, Jacewicz J, Piwnicka E, Morcinek-Orłowska J, Loika H, Czaplewska P, Węgrzyn G, Adamus-Białek W, Łukasiak A, Nejman-Faleńczyk B. Antibacterial activity of a newly identified phage-derived endolysin and its parental bacteriophage against clinical uropathogenic *Escherichia coli*,

polegał na przygotowaniu i wykonaniu doświadczenia identyfikacji endolizyny 57_3Lys z wykorzystaniem LC-MS/MS.

Paulina Czaplewska

PODPIS

Grzegorz Węgrzyn
.....
IMIĘ I NAZWISKO

27 sierpnia 2026 r.
.....
DATA

Katedra Biologii Molekularnej,
Wydział Biologii,
Uniwersytet Gdański

.....
AFILIACJA

OŚWIADCZENIE

Niniejszym oświadczam, iż mój wkład w manuskrypt:

Wesołowski W, Bloch S, Czerwonka G, Jagieła E, Grabowski Ł, Jeschke M, Jacewicz J, Piwnicka E, Morcinek-Orłowska J, Loika H, Czaplewska P, Węgrzyn G, Adamus-Białek W, Łukasiak A, Nejman-Faleńczyk B. Antibacterial activity of a newly identified phage-derived endolysin and its parental bacteriophage against clinical uropathogenic *Escherichia coli* (manuskrypt wysłany do redakcji *Virology*)

polegał na konsultacjach merytorycznych przy interpretacji wyników eksperymentów oraz pomocy w redagowaniu tekstu manuskryptu.

KIEROWNIK
KATEDRY BIOLOGII MOLEKULARNEJ
prof. dr hab. Grzegorz Węgrzyn

.....
PODPIS

Wioletta Adamus-Białek

.....
IMIĘ I NAZWISKO

23.08.2026

.....
DATA

Zakład Genetyki Medycznej i Diagnostyki Laboratoryjnej,
Wydział Lekarski,
Uniwersytet Jana Kochanowskiego w Kielcach

.....
AFILIACJA

OŚWIADCZENIE

Niniejszym oświadczam, iż mój wkład w manuskrypt:

Wesołowski W, Bloch S, Czerwonka G, Jagieła E, Grabowski Ł, Jeschke M, Jacewicz J, Piwnicka E, Morcinek-Orłowska J, Loika H, Czaplewska P, Węgrzyn G, Adamus-Białek W, Łukasiak A, Nejman-Faleńczyk B. Antibacterial activity of a newly identified phage-derived endolysin and its parental bacteriophage against clinical uropathogenic *Escherichia coli*.

polegał na: selekcji i dostarczeniu materiału do badań (izolaty bakteryjne), wraz z ich charakterystyką, nadzorowaniu pracy związanej z sekwencjonowaniem genomu bakteryjnego, współudziałem w przygotowaniu części manuskryptu, a następnie jego redakcji i korekcie.



Signed by /
Podpisano przez:

Wioletta Adamus-Białek

Date / Data: 2026-08-23 20:50

.....
PODPIS

KIR Szafir | elektronicznypodpis.pl

Raport z weryfikacji podpisu

INFORMACJE O DOKUMENCIE:	
Nazwa pliku	WAB praca nr 4-sig.pdf
Data weryfikacji podpisów	2026-08-24 15:57:04

SZCZEGÓŁY WERYFIKACJI:

Wioletta Adamus-Białek

IMIĘ I NAZWISKO

23.08.2026

DATA

Zakład Genetyki Medycznej i Diagnostyki Laboratoryjnej,
Wydział Lekarski,
Uniwersytet Jana Kochanowskiego w Kielcach

AFILIACJA

OŚWIADCZENIE

Niniejszym oświadczam, iż mój wkład w manuskrypt:

Wesołowski W, Bloch S, Czerwonka G, Jagiela E, Grabowski L, Jeschke M, Jacewicz J, Piwnicka E, Morcinek-Orlowska J, Loika H, Czaplewska P, Węgrzyn G, Adamus-Białek W, Lukasiak A, Nejman-Faleńczyk B. Antibacterial activity of a newly identified phage-derived endolysin and its parental bacteriophage against clinical uropathogenic *Escherichia coli*.

polegał na: selekcji i dostarczeniu materiału do badań (izolaty bakteryjne), wraz z ich charakterystyką, nadzorowaniu pracy związanej z sekwencjonowaniem genomu bakteryjnego, współudziałem w przygotowaniu części manuskryptu, a następnie jego redakcji i korekcie.

Signed by / Podpisano przez: Wioletta Adamus-Białek
Date / Data: 2026-08-23 20:50

PODPIS

Podpis: Pozytywny	
Rodzaj certyfikatu	Kwalifikowany
Format podpisu	PAdES-BES
Data złożenia podpisu	2026-08-23 20:50:47
Podpis zawiera znacznik czasu	Brak
Certyfikat podpisującego / składającego pieczęć	Nazwa powszechna: Wioletta Adamus-Białek Kraj: PL
Numer seryjny certyfikatu	183280257084507262196234348068648645422160390452
Wystawca certyfikatu	Nazwa powszechna: COPE SZAFIR - Kwalifikowany Nazwa organizacji: Krajowa Izba Rozliczeniowa S.A. Kraj: PL Identyfikator organizacji: VATPL-5260300517
Lista CRL wykorzystana do weryfikacji certyfikatu	-
Odpowiedź OCSP wykorzystana do weryfikacji certyfikatu	Status: GOOD ArchiveCutOff: 13 Mar 2017 14:07:47 GMT CN=COPE SZAFIR OCSP Responder OZK62 O=Krajowa Izba Rozliczeniowa S.A. L=Warszawa ST=mazowieckie C=PL - 24 Aug 2026 13:56:57 GMT
Zweryfikowano na dzień	2026-08-24 15:56:57 (Bieżący czas systemowy)
Status weryfikacji	**Pozytywny**
	Kwalifikowany podpis poprawnie zweryfikowany certyfikatem kwalifikowanym
Uwagi	Weryfikowany podpis został uznany za kwalifikowany. Zgodnie z Art. 25. Punkt 2 rozporządzenia Parlamentu Europejskiego i Rady (UE) nr 910/2014 z 23 lipca 2014 r. kwalifikowany podpis elektroniczny ma skutek prawny równoważny podpisowi własnoręcznemu.

Aleksandra Łukasiak
.....
IMIĘ I NAZWISKO

... 27.08.2026 ...
DATA

Katedra Biologii Molekularnej,
Wydział Biologii,
Uniwersytet Gdański

.....
AFILIACJA

OŚWIADCZENIE

Niniejszym oświadczam, iż mój wkład w manuskrypt:

Wesołowski W, Bloch S, Czerwonka G, Jagieła E, Grabowski Ł, Jeschke M, Jacewicz J, Piwnicka E, Morcinek-Orłowska J, Loika H, Czaplewska P, Węgrzyn G, Adamus-Białek W, Łukasiak A, Nejman-Faleńczyk B. Antibacterial activity of a newly identified phage-derived endolysin and its parental bacteriophage against clinical uropathogenic *Escherichia coli*,

polegał na

(i) współudziale w określeniu efektu cytotoksycznego wywoływanego przez lizat fagowy na komórki eukariotyczne;

(ii) pozyskaniu funduszy.

.....
PODPIS

Bożena Nejman-Faleńczyk
.....
IMIĘ I NAZWISKO

09.09.2026.
.....
DATA

Katedra Biologii Molekularnej,
Wydział Biologii,
Uniwersytet Gdański

.....
AFILIACJA

OŚWIADCZENIE

Niniejszym oświadczam, iż mój wkład w manuskrypt:

Wesołowski W, Bloch S, Czerwonka G, Jagieła E, Grabowski Ł, Jeschke M, Jacewicz J, Piwnicka E, Morcinek-Orłowska J, Loika H, Czaplewska P, Węgrzyn G, Adamus-Białek W, Łukasiak A, Nejman-Faleńczyk B. Antibacterial activity of a newly identified phage-derived endolysin and its parental bacteriophage against clinical uropathogenic *Escherichia coli*,

polegał na:

- opracowaniu i konsultowaniu koncepcji badań oraz założeń naukowych niniejszej publikacji,
- współudziale w sprawowaniu merytorycznego nadzoru nad realizacją prac,
- udziel w doborze metod badawczych, interpretacji uzyskanych wyników oraz formułowaniu wniosków,
- koordynowaniu prac związanych z przygotowaniem manuskryptu i dbaniu o jego spójność merytoryczną i redakcyjną,
- koordynowaniu współpracy pomiędzy współautorami i określeniu zakresu ich pracy,
- przygotowaniu oraz edycji pierwotnej i ostatecznej wersji manuskryptu.

B. Nejman-Faleńczyk
.....
PODPIS

Praca nr 4

Phage endolysins as promising and effective candidates for use against uropathogenic *Escherichia coli*.

Wesołowski W, Łukasiak A, Bloch S, Kuligowska K, Neumann J, Lewandowska N, Węglińska E, Węgrzyn G, Nejman-Faleńczyk B.

Viruses. 2025 Apr 13;17(4):560. doi: 10.3390/v17040560.

Impact Factor: 3.8

Punktacja Ministerstwa Nauki i Szkolnictwa Wyższego: 100

viruses

MDPI

Review

Phage Endolysins as Promising and Effective Candidates for Use Against Uropathogenic *Escherichia coli*

Wojciech Wesołowski [1], Aleksandra Łukasiak [1], Sylwia Bloch [1,2], Kaja Kuligowska [1], Julia Neumann [3], Natalia Lewandowska [1], Emilia Węglińska [1], Grzegorz Węgrzyn [1] and Bożena Nejman-Faleńczyk [1,2,*]

[1] Laboratory of Biology and Biotechnology of Bacteriophages, Department of Molecular Biology, Faculty of Biology, University of Gdańsk, Wita Stwosza 59, 80-308 Gdansk, Poland; wojciech.wesolowski@phdstud.ug.edu.pl (W.W.); aleksandra.lukasiak@phdstud.ug.edu.pl (A.Ł.); sylwia.bloch@ug.edu.pl (S.B.); k.kuligowska.370@studms.ug.edu.pl (K.K.); natalia.lewandowska@phdstud.ug.edu.pl (N.L.); e.weglinska.166@studms.ug.edu.pl (E.W.); grzegorz.wegrzyn@ug.edu.pl (G.W.)
[2] BNF—New Bio Force Ltd., Kartuska 420a, 80-125 Gdańsk, Poland
[3] Laboratory of Environmental Chemoinformatics, Faculty of Chemistry, University of Gdansk (UG), Wita Stwosza 63, 80-309 Gdansk, Poland; j.neumann.618@studms.ug.edu.pl
* Correspondence: bozena.nejman-falenczyk@ug.edu.pl; Tel.: +48-(58)-523-60-38

Abstract: The presented in silico and phylogenetic analysis of putative endolysins potentially produced by phages infecting uropathogenic *Escherichia coli* (UPEC) demonstrates their remarkable diversity. These proteins exhibit significant variations in sequence length, molecular weight, isoelectric point, and stability, as well as diverse functional domains determining their enzymatic activity, including lysin, lysozyme, hydrolase, amidase, and peptidase functions. Due to their predicted lytic properties, endolysins hold great promise in combating UPEC bacteria, including those within biofilms, which are often highly resistant to conventional treatments. Despite their potential, several challenges hinder the full utilization of endolysins. These include the relatively small number of identified proteins, challenges in the annotation process, and the scarcity of studies evaluating their efficacy in vitro and in vivo against Gram-negative bacteria. In this work, we emphasize these challenges while also underlining the potential of endolysins as an effective tool against UPEC infections. Their effectiveness could be significantly enhanced when combined with agents that disrupt the outer membrane of these bacteria, making them a promising alternative or complement to existing antimicrobial strategies. Further research is necessary to fully explore their therapeutic potential.

Keywords: bacteriophages; phage endolysins; uropathogenic *Escherichia coli* (UPEC); urinary tract infections (UTIs); phage therapy

check for updates

Academic Editors: Zuzanna Drulis-Kawa and Tomasz Olszak

Received: 27 February 2025
Revised: 26 March 2025
Accepted: 11 April 2025
Published: 13 April 2025

Citation: Wesołowski, W.; Łukasiak, A.; Bloch, S.; Kuligowska, K.; Neumann, J.; Lewandowska, N.; Węglińska, E.; Węgrzyn, G.; Nejman-Faleńczyk, B. Phage Endolysins as Promising and Effective Candidates for Use Against Uropathogenic *Escherichia coli*. *Viruses* **2025**, *17*, 560. https://doi.org/10.3390/v17040560

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Phage endolysins are the best-studied peptidoglycan-degrading enzymes [1,2]. These molecules are produced by bacteriophages during viral multiplication within the host. They belong to the lytic system that is responsible for the digestion of the bacterial cell wall from inside of the bacterium, thereby inducing fast osmotic lysis of host cells and the release of phage progeny into the environment [3–5]. When developed as a drug, endolysins must be stable, soluble, and able to degrade peptidoglycan from outside [5]. Taking into account the enzymatic activity and mechanism of the disruption of peptidoglycan, phage-derived endolysins can be subdivided into five different groups: (I) glucosaminidases (N-acetyl-β-D-glucosamidases), (II) lysozymes or muramidases (N-acetylmuramidases),

(III) lytic transglycosylases, (IV) endopeptidases (L-alanoyl-D-glutamate endopeptidases), and (V) amidases (N-acetylmuramoyl-L-alanine amidases) [6]. The action of endolysins is frequently supported by other accessory proteins such as holins and pinholins. These proteins form pores in the cytoplasmic membranes and enable endolysins to enter and degrade the peptidoglycan layer [7,8].

Proteins functioning as endolysins during the phage life cycle have a globular or modular structure [9,10]. Moreover, their architecture differs between enzymes specific for Gram-positive and Gram-negative bacteria which is related to the different structures of the cell walls of these microorganisms [11]. The bacteriophage-derived endolysins of Gram-positive hosts have a molecular mass between 25 and 40 kDa and are composed of an N-terminal enzymatic activity domain (EAD) and a C-terminal cell wall-binding domain (CBD). However, some of these enzymes may also have more than one EAD and CBD domains in different orders. The EAD and CBD domains are connected by a short, flexible linker region [9,10,12,13]. The N-terminal EAD contributes to the simple enzymatic cleavage of different bonds in peptidoglycan structure, while C-terminal CBD can recognize and bind to specific ligand molecules on the cell wall [11,14,15]. Importantly, the modular structure of these Gram-positive phage-derived products allows protein engineering to improve the production process, enhance the activity against bacteria, widen the lytic spectrum of action of enzymes, and modify physicochemical properties of endolysins, such as stability and solubility [16–18]. In contrast, Gram-negative endolysins are mostly small, single-domain globular proteins with molecular mass between 15 and 20 kDa. Most of them contain only a single EAD domain, without a CBD [8,14]. Similarly to endolysins active against Gram-positive bacteria, some Gram-negative endolysins were found to have a modular structure. Examples include endolysins encoded by *Pseudomonas*, *Salmonella*, and *Burkholderia* bacteriophages [8,14]. Unfortunately, the presence of the outer membrane at the cell wall of Gram-negative pathogens makes the exogenous application of endolysins less attractive and effective compared to their counterparts produced by phages infecting Gram-positive bacteria. However, due to the huge problem of antibiotic resistance among Gram-negative bacteria, recently, the increasing attention of researchers has been devoted to developing new strategies for overcoming the obstacle of the outer membrane by Gram-negative targeted endolysins [14,19–21].

Antimicrobial resistance is a huge problem among UPEC that are responsible for urinary tract infections (UTIs). It is estimated that UPEC bacteria are responsible for approximately 50–65% of complicated cases and 75–85% of uncomplicated infections [22]. Although antibiotics are effective in treating uncomplicated UTIs, such a therapy becomes problematic for complicated infections, especially those caused by resistant bacterial strains. UPEC bacteria are hazardous due to their adhesive abilities, which enable the colonization of the urinary tract, as well as the formation of biofilm.

Biofilm formation is a crucial stage in the pathogenesis of UTIs, allowing UPEC bacteria to expand within the urinary tract effectively. Biofilm development is facilitated by the production of specific adhesins by UPEC, such as type 1 and P fimbriae, which recognize receptors on the surface of the host's epithelial cells. This enables the bacteria to adhere to the epithelial cell layer, such as that of the bladder, and initiate the production of an extracellular matrix (EPS), which effectively protects the bacteria from the action of antibiotics and the host's immune cells. Moreover, the biofilm contains persister cells, whose presence enables the immediate restoration/regrowth of the bacterial population after antibiotic treatment [23,24]. For this reason, biofilm removal is extremely challenging, and infection relapses are frequent. Biofilm can also be formed on abiotic surfaces, such as urinary catheters, making the treatment of UTIs in catheterized patients especially challenging. In the hospital environment, biofilm serves as a source of chronic infections,

and the spread of resistant strains, due to the intensive transfer of genes between cells, which promotes the selection and dissemination of antibiotic resistance [25].

Bacteriophages and the endolysins they produce may help address the problem of biofilm removal due to their ability to go through the EPS structure and increase the exposure of bacterial cells to the action of the phages and their enzymes, as well as to applied antibiotics [11,26]. Phage therapy targeting biofilms, incorporating both phages and phage-derived enzymes, should address several challenges. The most significant are the following: insufficient number of characterized bacteriophages and phage proteins, the development of bacterial resistance to phages, and safety concerns. Importantly, there are no studies clearly indicating that phage lytic enzymes, including endolysins, induce the development of resistance similar to the one observed in the case of antibiotics or even phages. Therefore, great hopes are placed in them, among others, for combating UPEC. For this reason, in this review, we decided to summarize our knowledge about putative endolysins identified in the genomes of bacteriophages infecting UPEC bacteria to show their great application potential and indicate the main research problems. Importantly, we focus on both endolysins produced by phages specific only to UPEC, but also those derived from phages that, apart from UPEC, infect other types of *E. coli* bacteria. To our surprise, we could not find reports concerning in vitro and in vivo applications of these endolysins; thus, we focused mainly on the in silico characteristics of their sequences found in the GenBank database. We gathered 39 proteins annotated by authors as endolysins or proteins having endolysin domains and performed their detailed in silico analysis and phylogenetic searches. In addition, we summarized the literature reports on the use of endolysins derived from phages other than anti-UPEC to combat uropathogens, concentrating on the endolysin action against bacterial biofilms.

2. In Silico Analysis of Anti-UPEC Phage Proteins Annotated as Endolysins

This analysis considered 75 lytic anti-UPEC bacteriophages with DNA sequences found in the GenBank database on 15.12.2024. Table 1 summarizes the most important information and properties of the putative endolysins (39 in total) annotated by different authors in the genomes of tested bacteriophages [27–41]. In cases where the phage genomic sequence was available, but no annotations were made, we performed the annotation ourselves using the Pharokka version 1.6.1 [42]. All putative endolysins annotated by us were labeled as "self", with the appropriate number indicating the open reading frame number assigned during the annotation process. For all proteins, annotated both by us and other scientists, we gathered data, such as the name of the protein, position of the gene in the virus genome, gene length, protein accession number, protein length, molecular weight, isoelectric point (pI), melting point (Tm), instability index, aliphatic index, grand average of hydropathy score (Gravy), name of the predicted domain, and its position in the protein sequence (Table 1).

Table 1. Characteristics of endolysins annotated in genomes of anti-UPEC bacteriophages.

Phage name [References]	Protein Name	Position of the Gene in Virus Genome	Gene Length (nt)	Protein Accession Number	Protein Length (aa)	Molecular Weight (kDa)	Instability Index	Isoelectric Point (pI)	Gravy	Aliphatic Index	Melting Point (Tm)	Name of the Predicted Domain *	Position of the Domain in Protein Sequence
vB_EcoM_LNA1 [27]	endolysin	77244–77708	464	self_ZKRH133	154	17.21	33.482	9.36	−0.381	81.104	47.0 °C	lyz_endolysin_autolysin (cd00737) Phage lysozyme (PF00959) Lysozyme-like (SSF53955) Endolysin (MF_04110)	7–148 13–143 1–152 1–150
590B [28]	endolysin	39903–40391	488	QVR48622.1	162	17.11	20.515	9.86	−0.125	87.531	51.0 °C	lyz_endolysin_autolysin (cd00737) Phage lysozyme (PF00959) Lysozyme-like (SSF53955)	9–146 15–138 3–149
vB_EcoS_UTEC10 [29]	endolysin	1847–2332	485	WWY66481.1	161	17.03	20.046	9.81	−0.12	88.696	50.0 °C	lyz_endolysin_autolysin (cd00737) Phage lysozyme (PF00959) Lysozyme-like (SSF53955)	9–146 15–138 3–149
vB_EcoS-phiEc3 [30]	endolysin	41533–42021	488	YP_010749568.1	162	17.05	19.526	9.81	−0.101	88.148	51.0 °C	lyz_endolysin_autolysin (cd00737) Phage lysozyme (PF00959) Lysozyme-like (SSF53955)	9–146 15–138 3–149
NTEC3 [31]	endolysin	11540–12028	488	YP_010749155.1	162	17.14	21.851	9.86	−0.137	87.531	50.0 °C	lyz_endolysin_autolysin (cd00737) Phage lysozyme (PF00959) Lysozyme-like (SSF53955)	9–146 15–138 3–149
vB_EcoS-phiEc4 [30]	endolysin	41414–41899	485	QYU43915.1	161	17.11	22.631	9.92	−0.139	88.075	50.0 °C	lyz_endolysin_autolysin (cd00737) Phage lysozyme (PF00959) Lysozyme-like (SSF53955)	9–146 15–138 3–149
VB_EcoS-Golestan [32]	endolysin	41444–41902	458	YP_009620148.1	152	16.09	17.289	9.86	−0.069	88.816	49.0 °C	lyz_endolysin_autolysin (cd00737) Phage lysozyme (PF00959) Lysozyme-like (SSF53955)	3–137 6–129 2–141
vB_EcoM_SHAK7858 [33]	endolysin	40989–41474	485	WVW77834.1	161	17.11	19.781	9.62	−0.141	84.472	50.0 °C	lyz_endolysin_autolysin (cd00737) Phage lysozyme (PF00959) Lysozyme-like (SSF53955)	8–145 14–137 4–148

Table 1. *Cont.*

Phage name [References]	Protein Name	Position of the Gene in Virus Genome	Gene Length (nt)	Protein Accession Number	Protein Length (aa)	Molecular Weight (kDa)	Instability Index	Isoelectric Point (pI)	Gravy	Aliphatic Index	Melting Point (Tm)	Name of the Predicted Domain *	Position of the Domain in Protein Sequence
vB_EcoS_ACG-M12 [34]	endolysin	41386–41871	485	YP_006987885.1	161	17.63	33.838	9.83	−0.108	99.938	49.0 °C	endolysin_R21-like (cd16900) SAR-endolysin (MF_04136) Phage lysozyme (PF00959) Lysozyme-like (SSF53955) Endolysin (MF_04110)	20–156 3–158 25–151 6–159 4–158
Killian [35]	endolysin	77692–78186	494	WEU67899.1	164	18.59	37.155	9.59	−0.37	88.049	53.0 °C	T4-like_lys (cd00735) Phage lysozyme (PF00959) Lysozyme-like (SSF53955) Endolysin (MF_04110) Endolysin (G3DSA:1.10.530.40:FF:000002)	2–160 11–141 1–161 1–159 1–164
vB_EcoM_SHAK7693 [33]	endolysin	42651–42953	302	WVW77754.1	100	10.36	21.694	10.21	0.061	96.9	49.0 °C	Phage lysozyme (PF00959) Lysozyme-like (SSF53955)	2–76 2–86
vB_EcoM_SHAK7693 [33]	endolysin	42419–42658	239	WVW77753.1	79	8.73	25.656	7.91	−0.537	70.38	49.0 °C	Phage lysozyme (PF00959) Lysozyme-like (SSF53955)	20–67 6–68
vB_EcoP-101136BS1 [36]	endolysin	19728–20186	458	QZI79903.1	152	17.06	44.325	8.83	−0.391	78.158	51.0 °C	PGRP (cd06583) N-acetylmuramoyl-L-alanine amidase-like (SSF55846) N-acetylmuramoyl-L-alanine amidase (PF01510) Endolysin (MF_04111)	14–133 11–147 11–134 2–150
vB_EcoP-101120B2 [36]	endolysin	20895–21353	458	QZI79844.1	152	17.1	44.325	8.47	−0.46	76.25	51.0 °C	PGRP (cd06583) N-acetylmuramoyl-L-alanine amidase-like (SSF55846) N-acetylmuramoyl-L-alanine amidase (PF01510) Endolysin (MF_04111)	14–133 11–147 11–134 2–150
vB_EcoP-101120B1-2 [36]	endolysin	20840–21298	458	QZI79784.1	152	17.1	44.325	8.47	−0.46	76.25	51.0 °C	PGRP (cd06583) N-acetylmuramoyl-L-alanine amidase-like (SSF55846) N-acetylmuramoyl-L-alanine amidase (PF01510) Endolysin (MF_04111)	14–133 11–147 11–134 2–150

Table 1. *Cont.*

Phage name [References]	Protein Name	Position of the Gene in Virus Genome	Gene Length (nt)	Protein Accession Number	Protein Length (aa)	Molecular Weight (kDa)	Instability Index	Isoelectric Point (pI)	Gravy	Aliphatic Index	Melting Point (Tm)	Name of the Predicted Domain *	Position of the Domain in Protein Sequence
vB_EcoP-101118UKE1 [36]	endolysin	20127–20585	458	QZI79721.1	152	17.04	44.76	7.79	−0.38	78.158	51.0 °C	PGRP (cd06583) N-acetylmuramoyl-L-alanine amidase-like (SSF55846) N-acetylmuramoyl-L-alanine amidase (PF01510) Endolysin (MF_04111)	14–133 11–147 11–134 2–150
vB_EcoP-22664UKE3-2 [36]	endolysin	21058–21516	458	QZI78611.1	152	17.12	45.137	8.47	−0.458	76.908	51.0 °C	PGRP (cd06583) N-acetylmuramoyl-L-alanine amidase-like (SSF55846) N-acetylmuramoyl-L-alanine amidase (PF01510) Endolysin (MF_04111)	14–133 11–147 11–134 2–150
vB_EcoP-22664BS1 [36]	endolysin	19873–20226	353	QZI78465.1	117	13.16	50.611	7.83	−0.531	77.436	49.0 °C	PGRP (cd06583) N-acetylmuramoyl-L-alanine amidase-like (SSF55846) N-acetylmuramoyl-L-alanine amidase (PF01510)	5–98 2–112 7–99
vB_EcoP_ACG-C91 [34]	endolysin	39340–39684	344	YP_006987811.1	114	12.75	25.157	9.51	−0.333	72.632	52.0 °C	Peptidase M15 (PF08291) Hedgehog/DD-peptidase (SSF55166)	4–106 5–107
vB_EcoM_SHAK7163 [33]	endolysin	39749–40093	344	WVW77668.1	114	12.59	29.711	9.32	−0.348	71.754	50.0 °C	Peptidase M15 (PF08291) Hedgehog/DD-peptidase (SSF55166)	5–106 5–107
vB-EcoP-KMB14 [37]	endolysin	38777–39121	344	WPK27911.1	114	12.65	40.639	9.35	−0.392	67.544	49.0 °C	Peptidase M15 (PF08291) Hedgehog/DD-peptidase (SSF55166)	4–106 5–108
vB_EcoM_SHAK7704 [33]	endolysin	3327–3671	344	self_RMKP5	114	12.65	25.787	9.32	−0.401	70.088	51.0 °C	Peptidase M15 (PF08291) Hedgehog/DD-peptidase (SSF55166)	5–106 5–107
MLP1 [38]	endolysin	32790–33185	395	YP_010844467.1	131	14.96	47.821	8.79	−0.303	84.122	57.0 °C	L-Ala-D-Glu_peptidase_like (cd14845) Hedgehog/DD-peptidase (SSF55166) D-alanyl-D-alanine carboxypeptidase (PF13539)	9–130 4–129 58–110
vB-EcoM-KMB43 [37]	endolysin	55240–55629	389	WQN06429.1	129	14.72	12.666	9.3	−0.412	79.922	53.0 °C	L-Ala-D-Glu_peptidase_like (cd14845) Hedgehog/DD-peptidase (SSF55166) D-alanyl-D-alanine carboxypeptidase (PF13539)	11–120 2–120 55–120

Table 1. *Cont.*

Phage name [References]	Protein Name	Position of the Gene in Virus Genome	Gene Length (nt)	Protein Accession Number	Protein Length (aa)	Molecular Weight (kDa)	Instability Index	Isoelectric Point (pI)	Gravy	Aliphatic Index	Melting Point (Tm)	Name of the Predicted Domain *	Position of the Domain in Protein Sequence
vB_EcoP_EG1 [39]	internal virion protein with endolysin domain	30249–34223	3974	YP_009795842.1	1324	145.33	36.856	6.58	−0.397	78.52	41.0 °C	Transglycosylase SLT domain (PF01464) Peptidoglycan transglycosylase gp16. (MF_04121) Lysozyme-like (SSF53955) LT-like (cd00254)	15–127 2–1314 5–146 28–144
vB_EcoM_LNA4 [27]	endolysin	21617–22171	554	self_PFXJ59	184	20.08	23.118	9.3	−0.335	88.098	53.0 °C	Lysozyme-like (SSF53955) Chitinase class I (PF00182)	3–182 41–146
vB_EcoM_LNA4 [27]	endolysin	117557–118129	572	self_PFXJ204	190	21.65	30.802	9.67	−0.374	81.105	41.0 °C	Cell Wall Hydrolase (PF07486)	70–181
Killian [35]	endolysin	54687–56504	1817	WEU67851.1	605	67.92	35.65	6.61	−0.302	82.033	44.0 °C	anaerobic ribonucleoside-triphosphate reductase (TIGR02487) RNR_III (cd01675) PFL-like glycyl radical enzymes (SSF51998) Anaerobic ribonucleoside-triphosphate reductase (PF13597)	21–604 21–584 26–585 21–604
vB_Ec_ZCEC14 [40]	endolysin	13508–14425	917	UVD33329.1	305	35.56	28.885	8.61	−0.448	85.967	42.0 °C	T4 RNase H, C terminal (PF09293) PIN_53EXO-like (cd00008) PIN domain-like (SSF88723) H3TH_T4-like (cd09899) 5′-3′ exonuclease, N-terminal resolvase-like domain (PF02739) 5′ to 3′ exonuclease, C-terminal subdomain (SSF47807)	183–305 17–180 13–179 184–259 51–174 183–305
Killian [35]	endolysin	163120–164037	917	WEU68017.1	305	35.56	28.885	8.61	−0.448	85.967	42.0 °C	T4 RNase H, C terminal (PF09293) PIN_53EXO-like (cd00008) PIN domain-like (SSF88723) H3TH_T4-like (cd09899) 5′-3′ exonuclease, N-terminal resolvase-like domain (PF02739) 5′ to 3′ exonuclease, C-terminal subdomain (SSF47807)	183–305 17–180 13–179 184–259 51–174 183–305
vB_Ec_ZCEC14 [40]	endolysin	51539–52219	680	UVD33388.1	226	25.57	45.489	4.43	−0.378	91.947	48.0 °C	Bacteriophage T4, inh (IPR016594)	1–226

Table 1. *Cont.*

Phage name [References]	Protein Name	Position of the Gene in Virus Genome	Gene Length (nt)	Protein Accession Number	Protein Length (aa)	Molecular Weight (kDa)	Instability Index	Isoelectric Point (pI)	Gravy	Aliphatic Index	Melting Point (Tm)	Name of the Predicted Domain *	Position of the Domain in Protein Sequence
Killian [35]	endolysin	46919–47182	263	WEU67833.1	87	10.24	48.485	4.21	−0.197	74.943	40.0 °C	-	-
vB_EcoM_SHAK7163 [33]	internal virion protein with endolysin domain	31648–34968	3320	WVW77660.1	1106	121.66	35.115	5.61	−0.423	79.322	42.0 °C	-	-
vB_EcoM_SHAK7704 [33]	internal virion protein with endolysin domain	39690–42998	3308	self_RMKP62	1102	121.26	35.421	5.69	−0.429	78.63	42.0 °C	-	-
vB_EcoM_SHAK7854 [33]	internal virion protein with endolysin domain	32030–35917	3887	self_DNYN53	1295	140.9	37.494	5.67	−0.396	79.012	41.0 °C	-	-
vB_EcoM_SHAK9454 [33]	internal virion protein with endolysin domain	32160–36047	3887	self_ZENB53	1295	140.83	37.467	5.87	−0.39	78.942	41.0 °C	-	-
vB_EcoP_LNA3 [27]	internal virion protein with endolysin domain	5943–9830	3887	self_GXIJ6	1295	140.96	36.837	5.68	−0.399	79.459	41.0 °C	-	-
vB_EcoP_LNA5 [27]	internal virion protein with endolysin domain	23530–27417	3887	self_TCEH45	1295	140.81	37.737	5.93	−0.398	77.892	41.0 °C	-	-
vB_Ec_ZCEC14 [40]	endolysin	127670–127933	263	UVD33512.1	87	10.29	48.816	4.3	−0.277	74.943	40.0 °C	-	-

* Endolysins are arranged based on the type of domain they possess.

The data in Table 1 were obtained by us using several in silico tools. The Prot-Param module from the Biopython library was used to predict molecular weight, isoelectric point (pI), instability index, and Gravy score [43]. The DeepSTABp software was used to predict protein melting points [44]. The aliphatic index was calculated using a previously described formula [45]. A domain search was performed using InterProScan version 5.72-103 [46]. This tool integrated predictive information about proteins' function from several partner resources (such as CDD, Coil, FunFam, Gene3D, Hamap, MobiDBLite, NCBIfam, Panther, Pfam, Phobius, PIRSF, PRINTS, ProSitePatterns, ProSiteProfiles, SMART, and SUPERFAMILY), giving an overview of the domains that a protein contains and families that it belongs to. All data generated by this program are collected in Supplementary Table S1. Notably, this analysis showed that some of the analyzed proteins do not have domains typical for phage endolysins, and they are described in more detail below.

In silico analyses indicated that genomes of 30 anti-UPEC bacteriophages contain genes encoding potential endolysins or proteins with endolysin domain [27–41]. In total, 39 putative enzymes were listed in Table 1. Interestingly, seven bacteriophages—EG1, LNA3, LNA5, SHAK9454, SHAK7163, SHAK7854, and SHAK7704 [27,33,39]—were identified to produce the internal virion proteins with endolysin domains (Table 1). In this group of 30 bacterial viruses, we can distinguish eleven bacteriophages that produce enzymes with lysozyme domains: 590B, Killian, NTEC3, LNA1, SHAK7693, SHAK7858, ACG-M12, UTEC10, Golestan, phiEc3, phiEc4 [27–35]; six phages with enzymes of amidase activity: 101118UKE1, 101120B1-2, 101120B2, 101136BS1, 22664BS1, 22664UKE3-2 [36]; six phages having endolysins with peptidase domain: ACG-C91, SHAK7704, SHAK7163, KMB14, KMB43 MLP1 [33,34,37,38]; EG1 phage [39] with enzyme having transglycosylase domain; and LNA4 phage [27] producing enzymes with hydrolase activity, as indicated in Table 1. The exceptions are seven internal virion proteins (YP_009795842.1, self_TCEH45, self_ZENB53, self_GXIJ6, self_DNYN53; self_RMKP62, and WVW77660.1) identified in phages SHAK7163, SHAK7704, SHAK7854, SHAK8454, LNA3, LNA5 [27,33], and five endolysins (WEU68017.1, WEU67833.1, UVD33329.1, UVD33388.1, and UVD33512.1) originated from phages Killian [35] and ZCEC14 [40], in which the in silico tools were unable to identify the domains typically characteristic of endolysins (Table 1, Supplementary Table S1).

As we mentioned above, the first considered group of endolysins are enzymes with domains of lysozymes (Table 1), and they are annotated in genomes of the 11 anti-UPEC phages: 590B, Kilian, NTEC3, LNA1, SHAK7693, SHAK7858, ACG-M12, UTEC10, Golestan, phiEc3, and phiEc4 [27–35]. Phages probably produce these potential N-acetylmuramidases to kill Gram-negative bacteria through targeted hydrolysis. As a result of such action, the integrity of peptidoglycan is interrupted and an imbalance of turgor pressure occurs, causing the lysis of bacterial cells [47–49]. The endolysins (YP_010749568.1, QYU43915.1, YP_009620148.1, YP_010749155.1, QVR48622.1, WWY66481.1, WVW77834.1, and self_ZKRH133) of the phages phiEc3 [30], phiE4 [30], Golestan [32], NTEC3 [31], 590B [28], UTEC10 [29], SHAK7858 [33], and LNA1 [27] demonstrate the presence of endolysin and autolysin domains (cd00737, Table 1), and based on their domains, they are classified as lysozyme-like proteins or phage lysozymes (SSF53955, PF00959, respectively). These in silico-identified proteins have a predicted molecular mass of 8.73 to 18.59 kDa and an isoelectric point from 7.91 to 10.21 (Table 1). Interestingly, the endolysins (self_ ZKRH133, YP_006987885.1, YP_009620148.1, WVW77753.1, and WVW77754.1) encoded in genomes of the LNA1 [27], ACG-M12 [34], Golestan [32], and SHAK7693 [33] bacteriophages probably melt below 50 °C (Table 1). The remaining enzymes may denature at 50 °C and above (Table 1). The grand average of hydropathy (Gravy) of all proteins is below 0, indicating that they are soluble in water. The aliphatic index spans between 70.380 and 99.938, which indicates that these proteins are thermally stable and contain high amounts of hydrophobic

amino acids. Moreover, the instability index of all of them is below 40, confirming these proteins' stability (Table 1).

An interesting case among endolysins with a lysozyme domain is the endolysin (YP_006987885.1) synthesized by phage ACG-M12 [34]. It is an alkaline protein with a pI of 9.83, a melting temperature of 49 °C, and an aliphatic index of 99.938. This protein has a molecular weight of 17.63 kDa, is composed of 161 amino acids, seems to be stable as its instability index is equal to 33.838, and possesses the endolysin R21-like protein domain (cd16900, Table 1) that encompasses residues from 20 to 156. In addition to this, analyses from Phobius reveal that this protein contains a signal peptide spanning amino acids 1 to 27, and Hamap shows that it resembles SAR-endolysins (Supplementary Table S1) and may anchor to the membrane in an inactive form, which acts to prevent premature lysis of the infected host [50–52]. Another interesting protein with predicted lysozyme activity is the enzyme (WEU67899.1) encoded by the gene located in the DNA of the Killian phage [35]. This anti-UPEC virus possesses in its genome as many as four potential genes responsible for the production of endolysins. However, only one of them has the bacteriophage T4-like lysozyme domain (cd00735) that encompasses amino acid residues from 2 to 160 and is characteristic of N-acetylmuramidases (Table 1). The endolysin of the Killian phage has 164 amino acids, a molecular mass of 18.59 kDa, and the theoretical isoelectric point is estimated to be 9.59. Additionally, the critical melting point of the Killian-derived enzyme with the lysozyme activity equals 53 °C, the aliphatic index equals 88.049, and its instability index is 37.155, which, with the high value of the aliphatic index, may indicate its good stability (Table 1). In addition, it is worth emphasizing that all enzymes belonging to the group of proteins with lysozyme activity are involved in the peptidoglycan (GO:0009253) and cell wall macromolecule catabolic (GO:00016998) processes (Supplementary Table S1).

The second group of endolysins potentially produced by phages infecting UPEC bacteria are enzymes with an amidase domain (IPR002502; Supplementary Table S1). These proteins are probably produced by six viruses—22664UKE3-2, 101120B2, 101118UKE1, 101120B1-2, 101136BS1, and 22664BS1 [36]—and possess N-acetylmuramoyl-L-alanine amidase activity (PF01510; Table 1) and a peptidoglycan recognition protein (PGRP) domain (cd06583, Table 1). It is known that the peptidoglycan amidases act by breaking the amide bond that separates the glycan strand from the stem peptide located between the N-acetylmuramic acid and L-alanine residues [48,49]. Five of these proteins (QZI79721.1, QZI79903.1, QZI79784.1, QZI79844.1, QZI78611.1) are composed of 152 amino acids, their predicted molecular weight is approximately 17 kDa, and they are predicted to be unstable as the instability index remains over 40 (Table 1). The exception is amidase encoded by the 22664BS1 phage (QZI78465.1). This protein is slightly shorter (117 amino acids) and has a molecular mass of 13.16 kDa and an instability index equal to 50.611. Interestingly, all these enzymes possess similar isoelectric point values, ranging from 7.79 to 8.83, a theoretical melting point which ranges from 49 °C to 51 °C, and a Gravy score below 0, which indicates they are hydrophilic. Taking into account gene ontology, the analyzed endolysins with amidase properties are also involved in the peptidoglycan catabolic process (GO:0009253), as indicated in Supplementary Table S1.

In the analyzed group of putative endolysins encoded by phages infecting UPEC strains, there are also six proteins with peptidase activity (SSF55166, Table 1). These proteins (WPK27911.1, WQN06429.1, self_ RMKP5, WVW77668.1, YP_010844467.1, and YP_006987811.1) are encoded by genes located in the genomes of six phages: KMB14, KMB43, SHAK7704, SHAK7163, MLP1, and ACG-C91 [33,34,37,38]. The target of such proteins is a peptide that is made up of the L-lysine and D-alanine link [29,30]. Interestingly, the MLP1- [38] and KMB43- [37] derived peptidases (YP_010844467.1 and WQN06429.1) have the L-Ala-D-Glu peptidase-like domain (cd148945, Table 1), whereas ACG-C91- [34],

KMB14- [37], SHAK7704- [33], and SHAK7163- [33] derived endolysins (YP__006987811.1, WPK27911.1, self_RMKP5, and WVW77668.1) have been identified to have the peptidase M15 domain (PF08291, Table 1). These enzymes' length spans from 114 to 131 amino acid residues, and their molecular weights are between 12.59 and 14.96 kDa (Table 1). Moreover, the theoretical isoelectric point of potential peptidases is above 8.79, which indicates their alkaline nature. The predicted melting point of KMB14 putative peptidase (WQN06429.1) is 49 °C, which is the lowest predicted melting temperature, whereas the highest one is calculated for MLP1 putative peptidase (YP_010844467.1) and equals 57 °C. Importantly, most of these peptidases seem to be stable, besides KMB14- and MLP1- derived peptidases (WPK27911.1 and YP_010844467.1) with stability indexes of 40.639 and 47.821, respectively. Additionally, all of them show Gravy values below 0 and an aliphatic index above 65, indicating their thermal stability.

Interesting cases are also two putative endolysins encoded by genes located in the genome of the LNA4 bacteriophage [27]. Both of them seem to have hydrolase activity; however, in one of these two enzymes (annotated by us as self_PFXJ59), the chitinase class I domain (PF00182, Table 1) was found between amino acids 41 and 146. The presence of such a domain may suggest that this protein participates in the chitin catabolic process (GO: 0006032, Supplementary Table S1). This is a rather unusual feature for a phage, but a similar phenomenon has been observed previously in *Bacillus* infecting phage [53]. As indicated in Table 1, the endolysin has a molecular weight of 20.08 kDa, a pI of 9.3, and a Tm of 53 °C. Its instability index is low and equals 23.118, which demonstrates that the protein is stable. In addition, its aliphatic index is 88.098, a high value that also indicates the thermal stability of this protein. The Gravy index score equals -0.335, suggesting that this endolysin is more hydrophilic and soluble in water (Table 1). Interestingly, the second LNA4-derived endolysin (annotated by us as self_PFXJ204) probably has the cell wall hydrolase domain between amino acids 70 and 181 (PF07486, Table 1). This is a protein with a molecular mass of 21.65 kDa, a pI of 9.67, and a Tm of 41 °C. Its instability index equals 30.802, and the aliphatic index is 81.105, both suggesting the stability of this protein; and the Gravy score is −0.374 (Table 1).

It is worth emphasizing that anti-UPEC bacteriophages can potentially produce not only typical endolysins, but also the internal virion proteins with endolysin domains (Table 1). As we mentioned above, this phenomenon concerns seven bacterial viruses: SHAK9454, SHAK7163, SHAK7704, SHAK7854, EG1, LNA3, and LNA5 [27,33,39]. These enzymes are longer than the typical endolysins synthesized by bacteriophages. They are composed of over 1100 amino acids, and their molecular weight is estimated to be approximately between 121 and 145 kDa (Table 1). Moreover, their theoretical isoelectric point is within the range of about 5.61 to 6.58, and almost all of them seem to be stable. All of these proteins probably melt at temperatures between 41 °C and 42 °C and have an aliphatic index score between 77.892 and 79.459. Interestingly, among these bacteriophages [27,33], only the EG1-derived internal virion protein (YP_009795842.1) exhibits transglycosylase activity (PF01464, Table 1). This protein has the putative lytic transglycosylase (LT)-like domain between amino acids 28 and 144 (cd00254, Table 1). On the other hand, LNA3, LNA5-, SHAK7854-, and SHAK9454- derived internal virion proteins (self_GXIJ6, self_TCEH45, self_DNYN53, and self_ZENB53) are equipped with transmembrane regions (Supplementary Table S1), whereas the SHAK7163-, SHAK7704- derived proteins (WVW77660.1 and self_RMKP62) have coiled regions (Supplementary Table S1). Interestingly, the results of the Blast search indicated the highest similarity (100% query cover, 0.0 e-value, and 95.62% identity) of the EG1-derived protein to the DNA ejectosome component of *Yersinia* phage vB_YenP_WW2 and the outer membrane protein of *Yersinia* phage vB_YenP_WX1. On the other hand, Blast search against a non-redundant database of *E. coli* (taxid:562)

records revealed that this protein is also similar to transglycosylase SLT domain-containing protein, with 100% query cover, 0.0 e-value, and 65.63% identity. In turn, LNA3-, LNA5-, SHAK9454-, and SHAK7854-derived proteins (self_GXIJ6, self_TCEH45, self_ZENB53, and self_DNYN53) are mostly common to internal virion protein D, with 100% query cover, 0.0 e-value, and identity varying from 96.76 to 98.46%. Lastly, SHAK7163- and SHAK7704-derived proteins (WVW77660.1 and self_RMKP62) are mostly similar to *E. coli* hypothetical proteins, also with 100% query cover, e-value equal to 0.0, and identity varying from 93.22% to 94.58%.

As mentioned before, in the Killian phage genome [35] there are four putative endolysins, but only one (WEU67899.1) with a phage lysozyme domain is described (Table 1). Other proteins, WEU67851.1 and WEU68017.1, have motifs not typical of endolysins, such as anaerobic ribonucleoside-triphosphate reductase domain (PF13597, Table 1) or a T4 RNase H motif (PF09293, Table 1), respectively. In the case of ZCEC14 phage [40], one of its putative endolysins (UVD33329.1) also possesses a T4 RNase H motif (PF09293, Table 1), and the second one (UVD33388.1) is considered to have a prohead protease inhibitor (IPR016594, Supplementary Table S1). Importantly, the identified domains suggest that these proteins may have other activities and are probably not endolysins.

Among these, we analyzed sequences of putative endolysins for the presence of antimicrobial peptides (AMPs) and found 39 such peptides with a high probability > 0.8, as indicated in Supplementary Table S2. The presence of such peptides indicates the lytic properties of analyzed endolysins against bacterial cells. Interestingly, some sequences of the identified peptides occur in more than one endolysin, which may suggest their evolutionary preservation.

The obtained results indicate a large diversity of domains present in endolysins and potentially high stability of the identified proteins. Moreover, we demonstrate that it is extremely important to thoroughly verify the annotated proteins in terms of functions they may perform.

3. Phylogenetic Analysis of Putative Endolysins Annotated in Genomes of Anti-UPEC Phages

The analyzed putative endolysins differ significantly, as indicated in the previous paragraph. To establish their common evolutionary origin, the OMA [54] software version 2.6.0 was used to determine 11 orthologous groups for all 39 annotated endolysins and other proteins identified in 75 genomic sequences of anti-UPEC phages found in the NCBI database on 15.12.2024 [27–41]. Then, based on such orthologous groups containing endolysin sequences, maximum likelihood trees were prepared using IQ-TREE [55]. The alignments of nucleotide sequences were performed using Clustal Omega version 1.2.4 [56], and the MEGA [57] program was used to perform visualization with midpoint rooting. This analysis allowed us to generate 11 separate trees (Figures 1–11), which are discussed in detail below. In addition, we also identified proteins that were not annotated by the authors as endolysins, but turned out to be closely related to them and may have similar functions. Importantly, phylogenetic trees were not obtained for some of the described endolysins (WVW77754.1, WVW77753.1, self_ZKRH133, WQN06429.1, YP_010844467.1, self_PFXJ59, and self_PFXJ204) due to the lack of their sequence similarity to other annotated proteins.

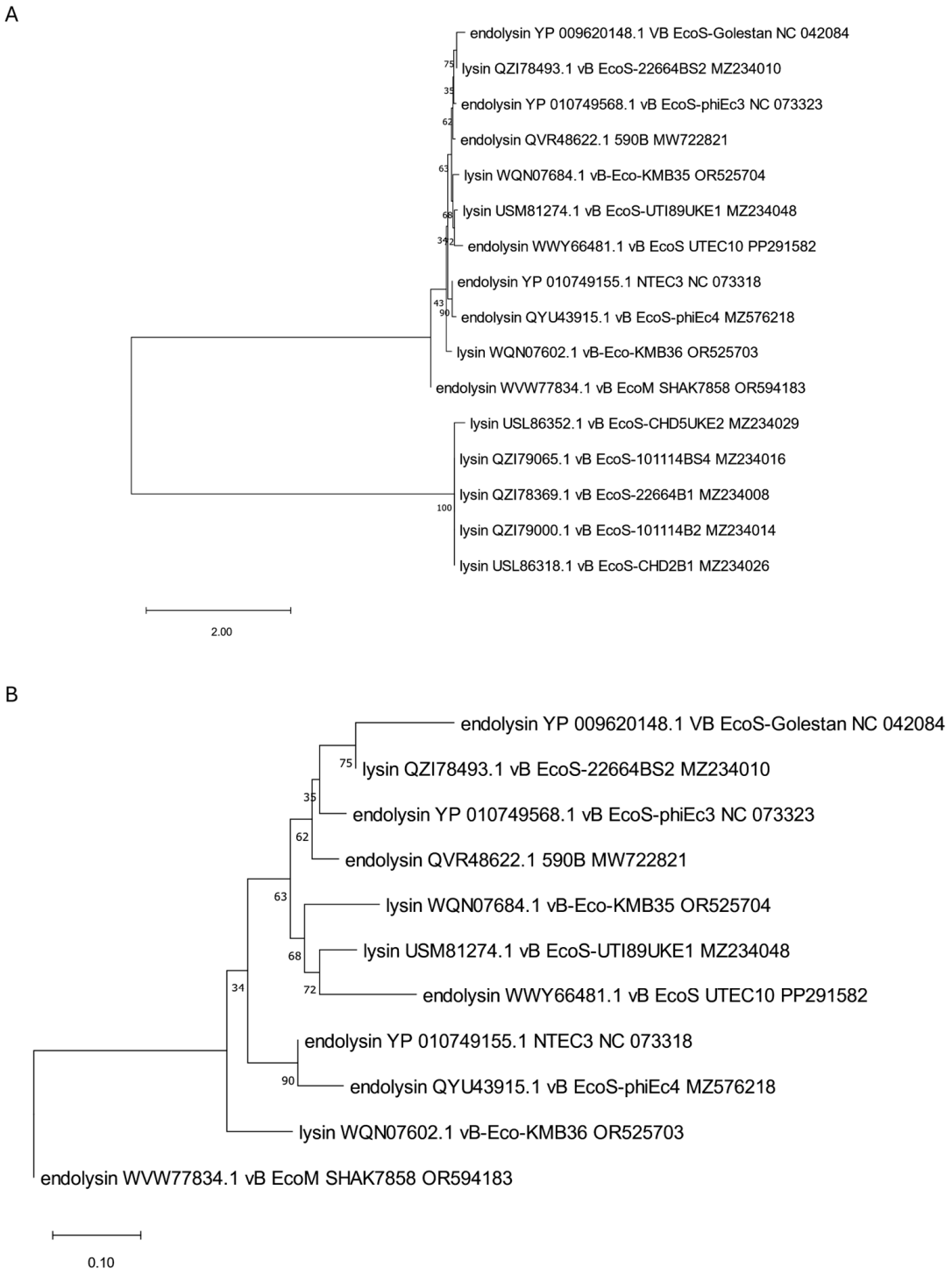


Figure 1. The phylogenetic tree is generated based on the first orthologous group determined by OMA and contains endolysins (YP_009620148.1, YP_010749568.1, QVR48622.1, WWY66481.1, YP_010749155.1, QYU43915.1, and WVW77834.1), with autolysin domains derived from phages Golestan, phiEc3, 590B, UTEC10, NTEC3, phiEc4, and SHAK7858, respectively. The tree was made as described in the text, using the maximum likelihood method with 1000 bootstrap replications. Panel (**A**) presents a tree with all grouped proteins. Panel (**B**) shows a modified tree, where proteins derived from phages CH5UKE2, 101114BS4, 22664B1, 101114B2, and CHD2B1 were removed to highlight the particular distances on the tree. The names indicated on the tree include the protein name and its accession number, followed by the phage name and its accession number.

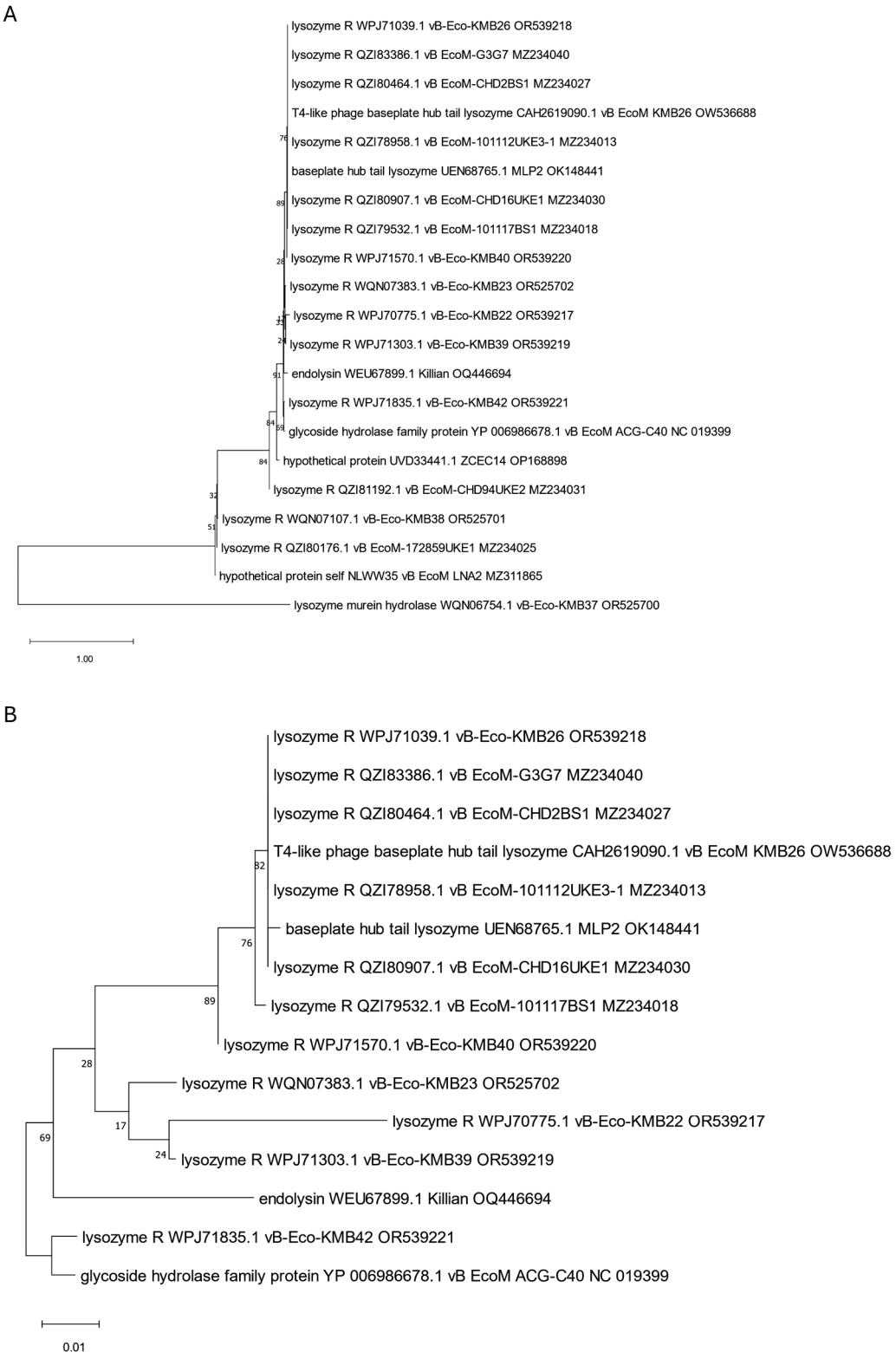


Figure 2. The phylogenetic tree is generated based on the second orthologous group estimated by OMA and contains a selected endolysin (WEU67899.1) of the Killian anti-UPEC phage. The tree was made using the maximum likelihood method with 1000 bootstrap replications. Panel (**A**) presents a tree with all grouped proteins. Panel (**B**) shows a modified tree, where proteins derived from phages KMB37, LNA2, 172859UKE1, KMB38, CHD94UKE2, and ZCEC14 were removed to highlight the particular distances on the tree. The names indicated on the tree include the protein name and its accession number, followed by the phage name and its accession number.

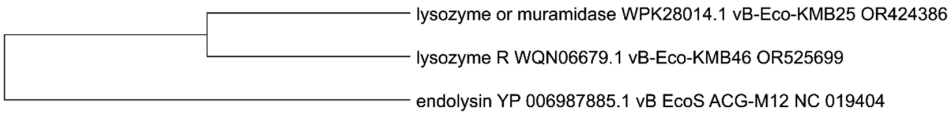


Figure 3. The phylogenetic tree is generated based on the third orthologous group determined by OMA and contains endolysin (YP_006987885.1) derived from phage ACG-M12. The tree was made using the maximum likelihood method. The names indicated on the tree include the protein name and its accession number, followed by the phage name and its accession number.

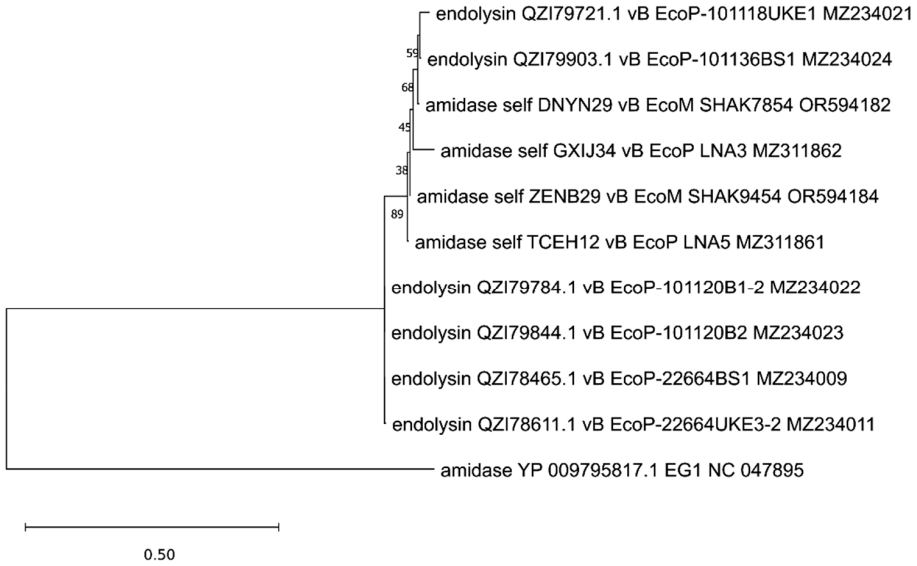


Figure 4. The phylogenetic tree is generated based on the fourth orthologous group estimated by OMA and contains endolysins with an amidase domain (QZI79721.1, QZI79903.1, QZI79784.1, QZI79844.1, QZI78465.1, and QZI78611.1). The tree was made using the maximum likelihood method with 1000 bootstrap replications. The names indicated on the tree include the protein name and its accession number, followed by the phage name and its accession number. The 'self'-labeled proteins were annotated by us using Pharokka version 1.6.1 [42].

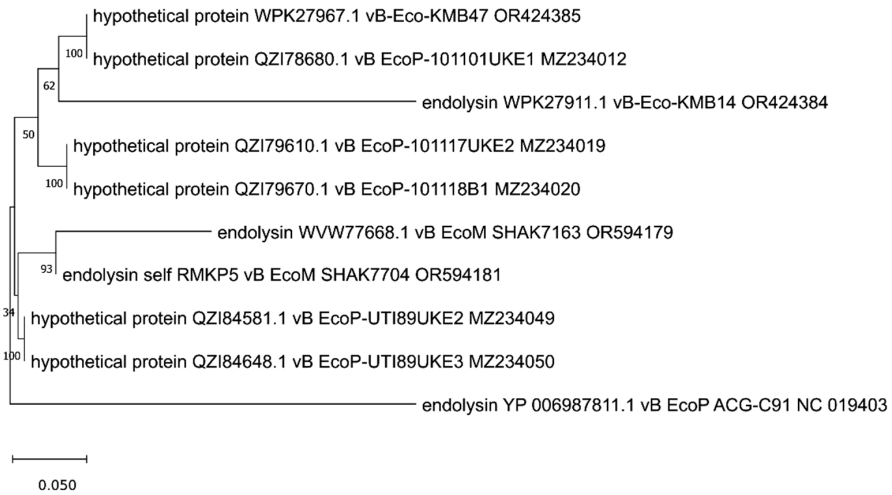


Figure 5. The phylogenetic tree is generated based on the fifth orthologous group determined by OMA and contains endolysins with peptidase domain (WPK27911.1, WVW77668.1, self_RMKP5, and YP006987811.1) derived from phages KMB14, SHAK7163, SHAK7704, and ACG-C91, respectively. The tree was made using the maximum likelihood method with 1000 bootstrap replications. The names indicated on the tree include the protein name and its accession number, followed by the phage name and its accession number. The 'self'-labeled endolysin was annotated by us using Pharokka version 1.6.1 [42].

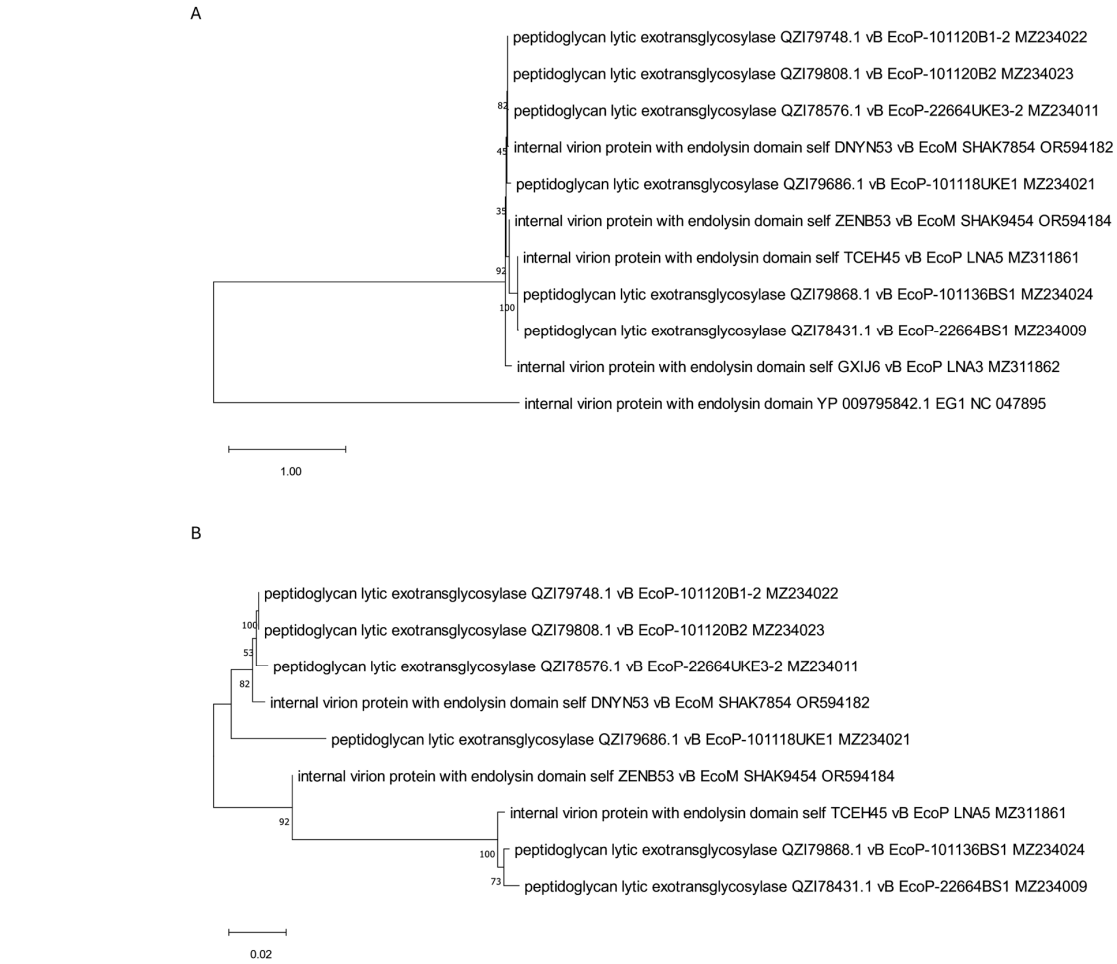


Figure 6. The phylogenetic tree is generated based on the sixth orthologous group determined by OMA and contains internal virion proteins with endolysin domain, such as EG-derived protein YP_009795842.1, and four other annotated and assigned by us: self_TCEH45, self_ZENB53, self_GXIJ6, and self_DNYN53. The tree was constructed using the maximum likelihood method with 1000 bootstrap replications. Panel (**A**) presents a tree with all grouped proteins. Panel (**B**) shows a modified tree, where proteins derived from phages EG1 and LNA3 were removed to highlight the particular distances on the tree. The names indicated on the tree include the protein name and its accession number, followed by the phage name and its accession number.

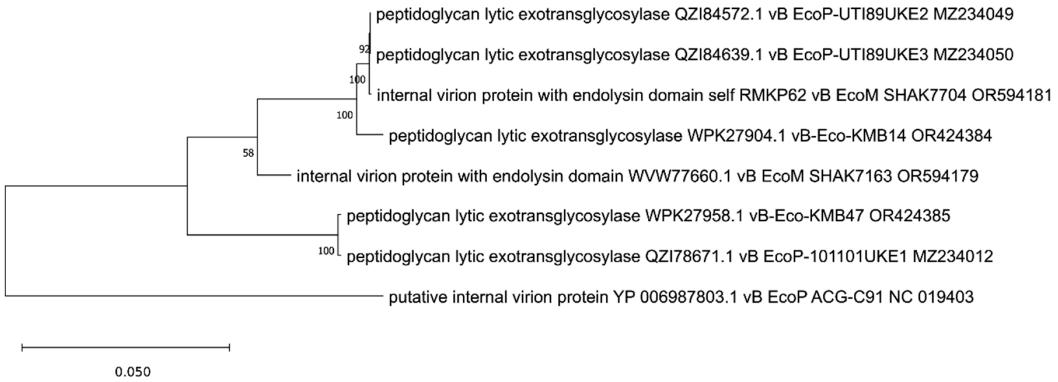


Figure 7. The phylogenetic tree is generated based on the seventh orthologous group established by OMA and contains internal virion proteins with a coiled-coil domain (WVW77660.1 and, annotated by us, self_RMKP62). It was made using the maximum likelihood method with 1000 bootstrap replications. The names indicated on the tree include the protein name and its accession number, followed by the phage name and its accession number.

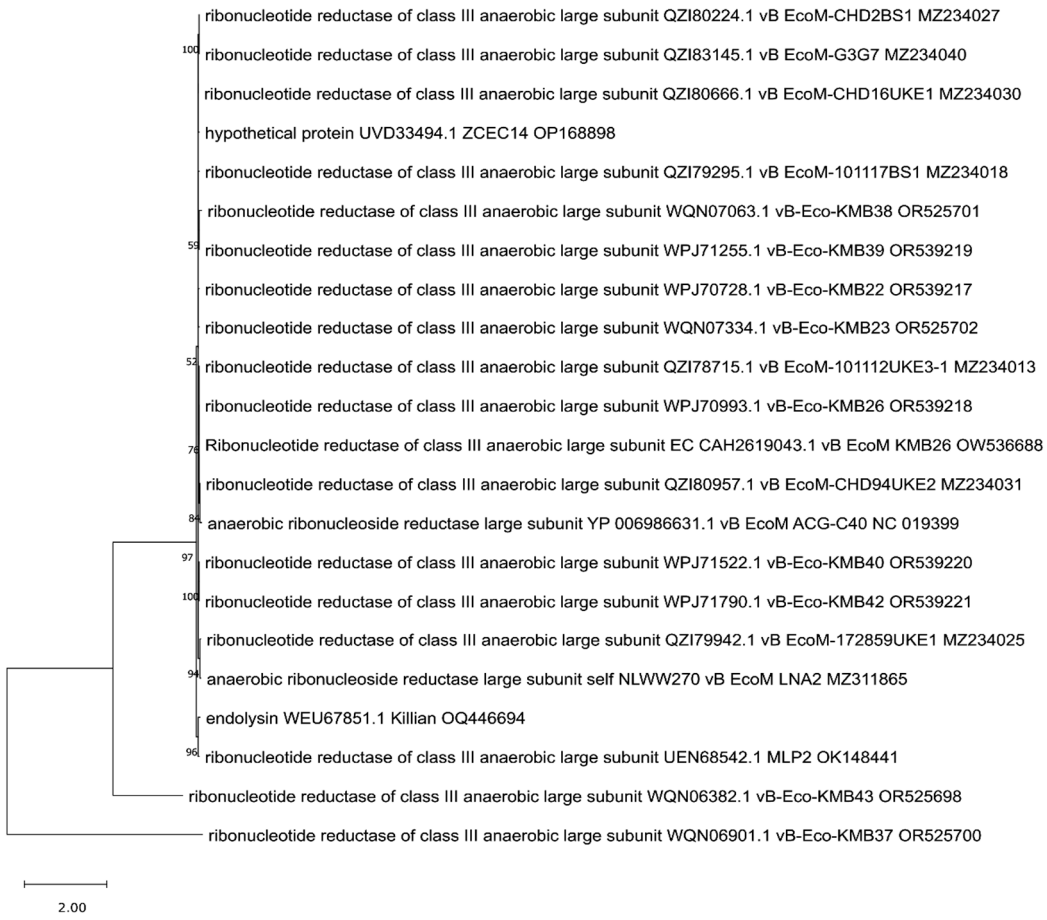


Figure 8. The phylogenetic tree is generated based on the eighth orthologous group established by OMA and contains putative endolysin of Killian phage (WEU67851.1). The tree was made using the maximum likelihood method with 1000 bootstrap replications. The names indicated on the tree include the protein name and its accession number, followed by the phage name and its accession number.

For convenience, we gathered all potential endolysins that appeared in our phylogenetic analysis in Supplementary Table S3. These were annotated differently by the authors (e.g., lysins, lysozymes, amidases, hydrolases, or hypothetical proteins). We performed analogous analyses for them, similar to those conducted for the originally annotated endolysins in Table 1. The results are presented in Supplementary Table S3. In addition, we performed a detailed domain search for all these proteins using InterProScan version 5.72-103 [47] and listed the obtained data in Supplementary Table S4. Among these, we analyzed the sequences of these proteins for the presence of AMPs and found 42 such peptides with high probability (>0.8). Interestingly, a few sequences of these peptides seem to be conserved among found proteins, as they repeat and occur in unchanging form in different putative enzymes (Supplementary Table S5).

The tree presented in Figure 1 includes almost all endolysins (YP_009620148.1, YP_010749568.1, QVR48622.1, WWY66481.1, YP_010749155.1, QYU43915.1, and WVW77834.1) with endolysin and autolysin domain (cd00737, Table 1) beyond endolysin from phage LNA1 [28] which, despite having the same domain type, has no common evolution origin. Interestingly, all endolysins included on the tree are closely related to proteins annotated as lysins in other anti-UPEC phages and probably act as their endolysins (Supplementary Tables S3 and S4).

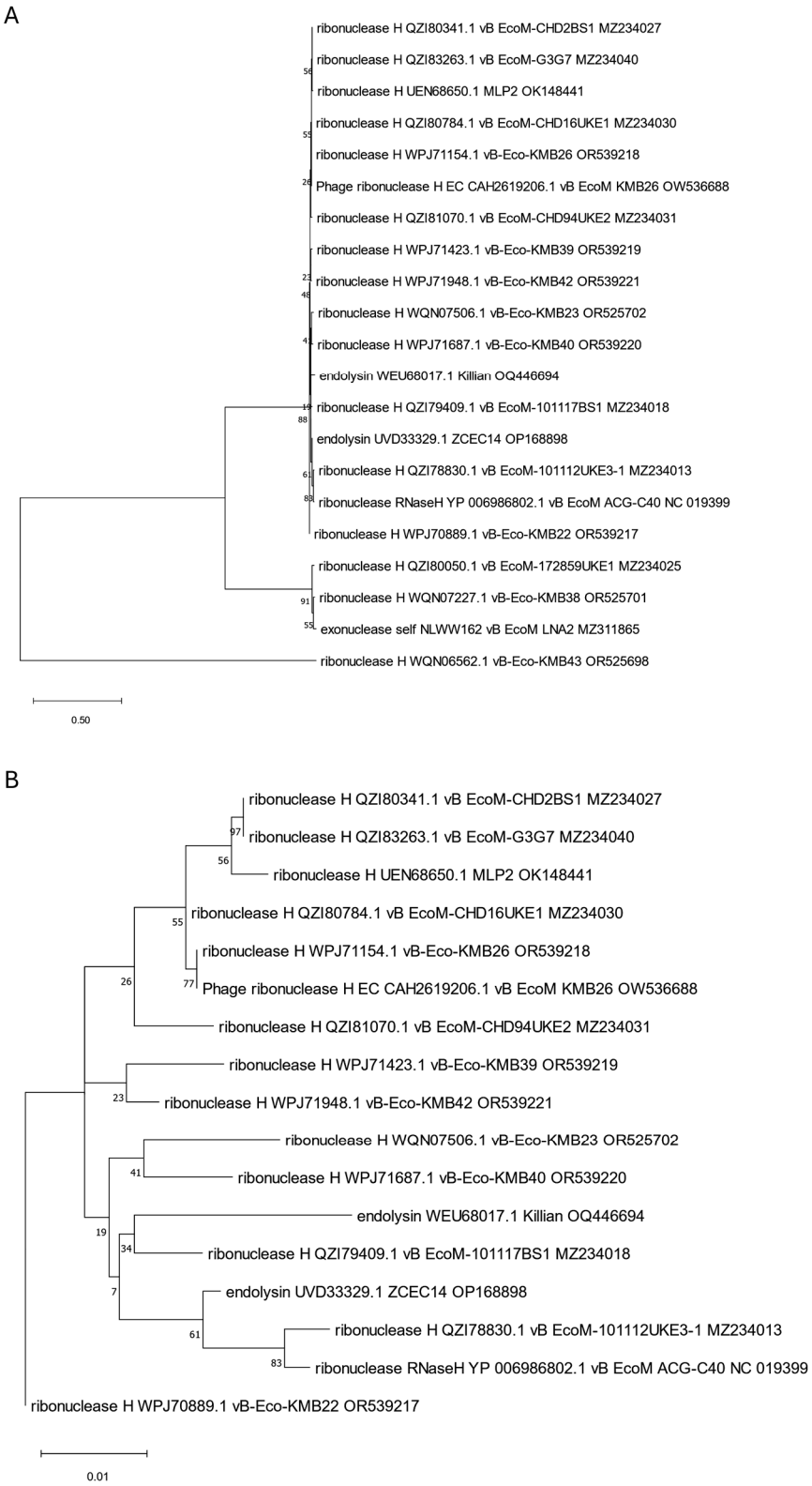


Figure 9. The phylogenetic tree is generated based on the ninth orthologous group determined by OMA and contains Killian's putative endolysin no. WEU68017.1 and ZCEC14 putative endolysin no. UVD33329.1. It was constructed using the maximum likelihood method with 1000 bootstrap replications. Panel (**A**) presents a tree with all grouped proteins. Panel (**B**) shows a modified tree, in which proteins of phages KMB43, LNA2, KMB38, and 172859UKE1 were removed to highlight the particular distances on the tree. The names indicated on the tree include the protein name and its accession number, followed by the phage name and its accession number.

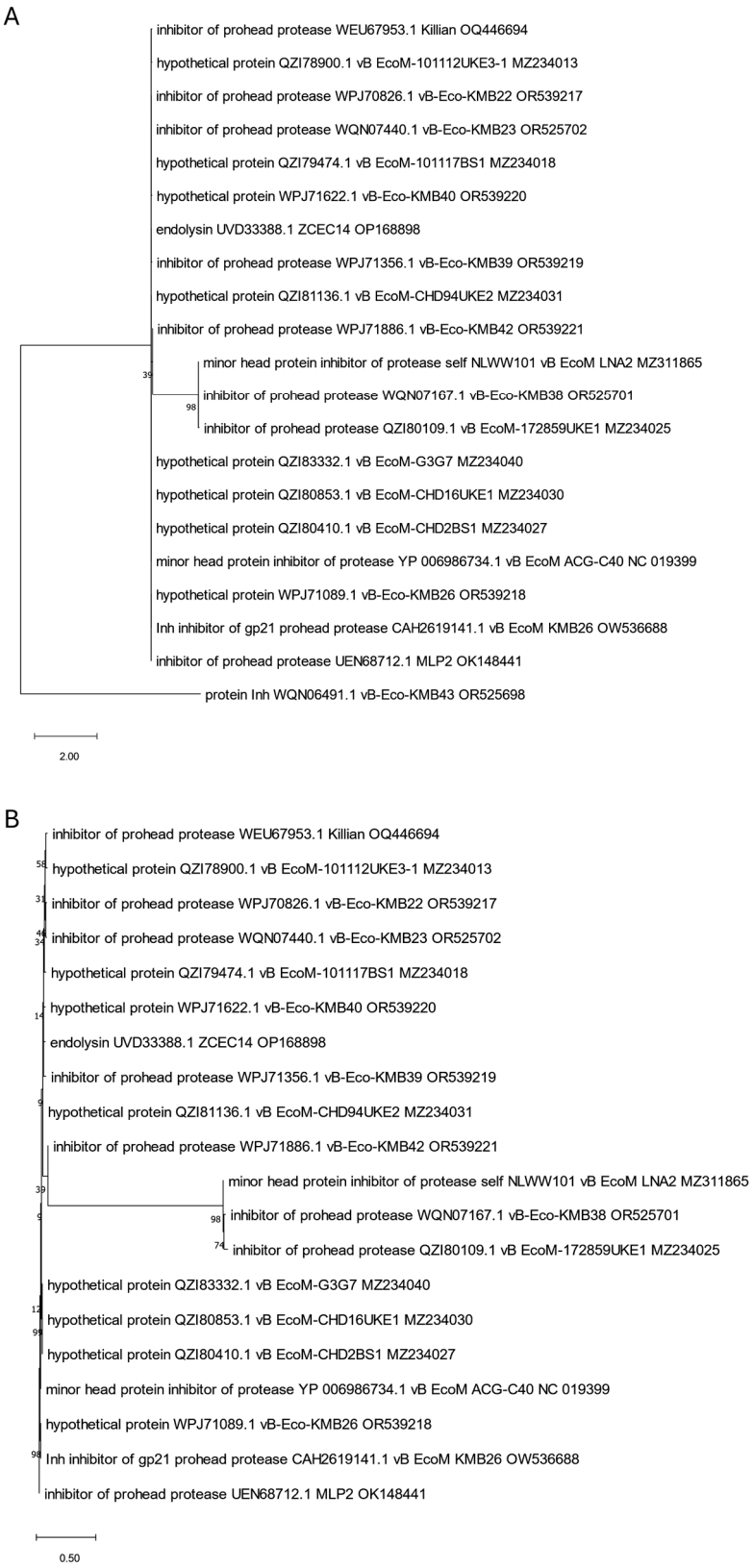


Figure 10. The phylogenetic tree is based on the tenth orthologous group identified by OMA and includes the ZCEC14-derived putative endolysin (UVD33388.1). It was constructed using the maximum likelihood method with 1000 bootstrap replications. Panel (**A**) displays the tree with all grouped proteins, while panel (**B**) presents a modified version in which the phage KMB43 protein was removed to emphasize specific distances within the tree. The names on the tree consist of the protein name and its accession number, followed by the phage name and its accession number.

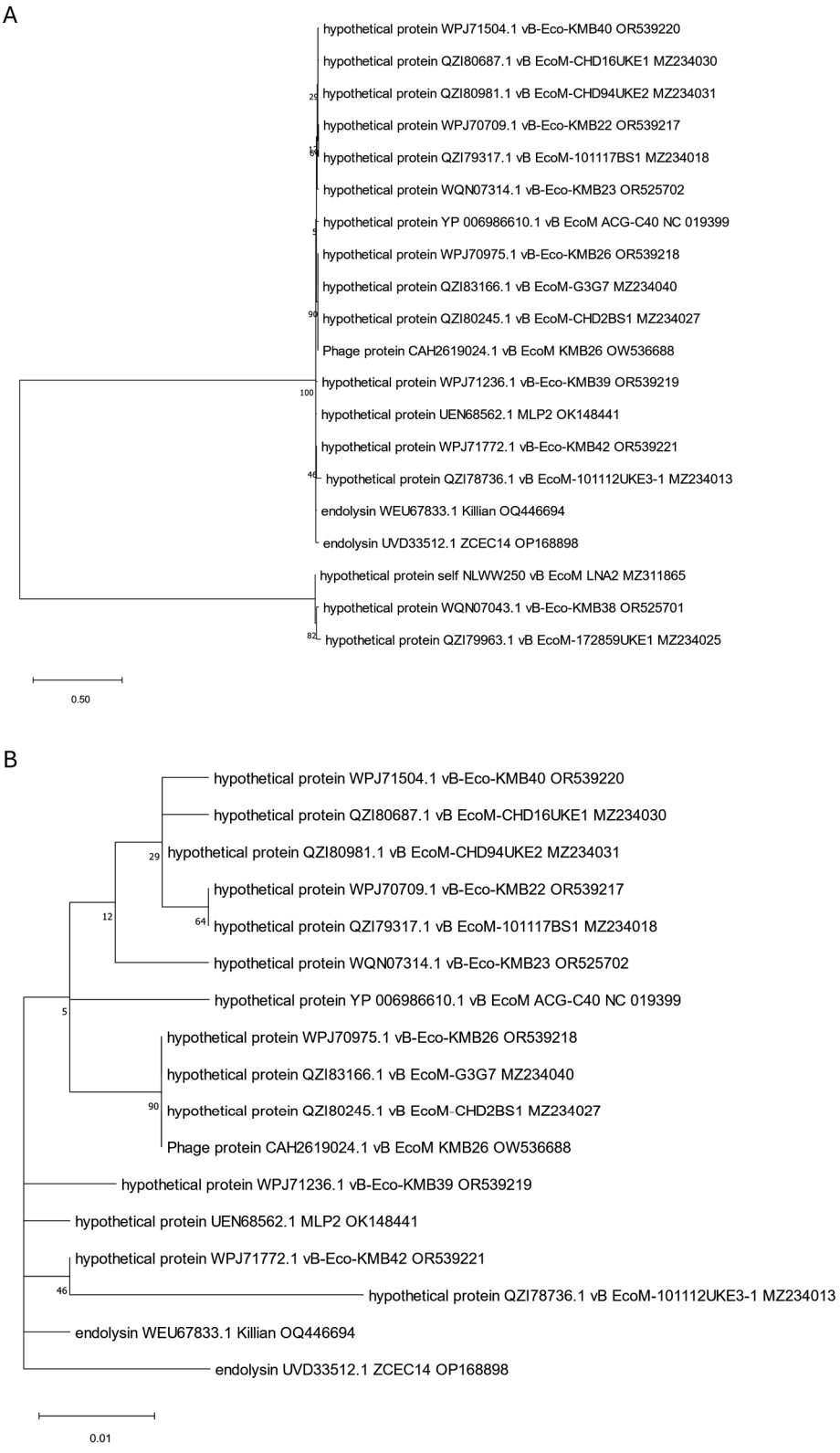


Figure 11. The phylogenetic tree is generated based on the eleventh orthologous group determined by OMA and contains two putative endolysins of Killian and ZCEC14 phages (WEU67833.1 and UVD33512.1). It was made using the maximum likelihood method with 1000 bootstrap replications. Panel (**A**) presents a tree with all grouped proteins. Panel (**B**) shows a modified tree, in which the proteins of phages LNA2, 172859UKE1, and KMB38 were removed to highlight the particular distances on the tree. The names indicated on the tree include the protein name and its accession number, followed by the phage name and its accession number.

Another two phylogenetic trees were prepared for the endolysins no. WEU67899.1 and YP_006987885.1 that have lysozyme domains and are derived from Killian [36] and ACG-M12 [35] phages, respectively (Figures 2 and 3). Within endolysin WEU67899.1 of the Killian phage, we identified, using different programs, three similar domains—a T4-like lysozyme domain (cd00735), a phage lysozyme (PF00959), and a lysozyme-like domain (SSF53955)—as indicated in Table 1. Phylogenetic analysis shows this protein is closely related to other phage proteins annotated as lysozyme R (Figure 2A,B). In turn, the ACG-M12-derived endolysin with a R21-like domain (cd16900), a phage lysozyme (PF00959), and a lysozyme-like domain (SSF53955) are grouped on the tree (presented in Figure 3) with proteins WPK28014.1 and WQN06679.1 annotated as lysozymes in genomes of KMB25 and KMB46 phages [37] that can serve as potential endolysins in these phages (Supplementary Tables S3 and S4).

Lastly, there are two SHAK7693- derived [33] unique endolysins (WVW77754.1 and WVW77753.1) with function confirmed by InterProScan but showing no similarity to other proteins.

Another tree, presented in Figure 4, encompasses endolysins containing amidase domain (PF01510, Table 1). All six originally annotated endolysins [36] with this domain (QZI79721.1, QZI79903.1, QZI79784.1, QZI79844.1, QZI78465.1, and QZI78611.1) are grouped with other proteins annotated as amidases, which can serve as potential endolysins for these phages and are indicated in Supplementary Tables S3 and S4. Importantly, four of the analyzed endolysins (QZI79784.1, QZI79844.1, QZI78465.1, and QZI78611.1) are almost identical in sequence, while the two others (QZI79721.1 and QZI79903.1) show a little disparity.

The next tree, presented in Figure 5, is based on endolysins with the domain classified as peptidase M15 (PF08291, Table 1). Interestingly, three of the four analyzed endolysins (WPK27911.1, WVW77668.1, and YP006987811.1) derived from phages KMB14 [37], SHAK7163 [33], and ACG-C91 [34], respectively, have longer branch lengths, indicating more genetic changes. This implies faster evolution of these proteins compared to hypothetical proteins. On the other hand, domain analysis conducted with InterPro shows that these hypothetical proteins have common peptidase M15 and hedgehog/DD-peptidase domains with the analyzed endolysins (Supplementary Tables S1 and S4). Unfortunately, two other endolysins (WQN06429.1 and YP_010844467.1) with peptidase domain no. PF13539, derived from phages KMB43 [37] and MLP1 [38], respectively, were not collected into the orthologous group by OMA, so they were not included in this analysis.

A separate tree was created, including proteins annotated as internal virion proteins with the endolysin domain (Figure 6). Importantly, all these proteins are significantly bigger than typical phage endolysins and present molecular weights from 120 to 145 kDa (Table 1). Among them, one interesting protein (YP_009795842.1) derived from EG1 phage [39] contains the SLT domain (PF01464, Table 1). This protein is grouped on the tree with four other internal virion proteins (annotated and assigned by us: self_GXIJ6, self_TCEH45, self_ZENB53, and self_DNYN53) originating from phages LNA3, LNA5, SHAK9454, and SHAK7854 [27,33] and showing transmembrane regions in their sequence (Supplementary Table S1). All of them seem to be closely related to phage proteins annotated as peptidoglycan lytic exotransglycosylases. However, the internal virion protein (YP_009795842.1) of EG1 phage [39] is the protein least related to others. Importantly, apart from internal virion proteins, the phages analyzed here have in their genomes genes encoding amidases, which were not annotated directly as endolysins but may serve such functions, as indicated in Supplementary Tables S3 and S4.

Two other internal virion proteins (WVW77660.1 and, annotated by us, self_RMKP62) derived from phage SHAK7163 [33] and phage SHAK7704 [33], respectively, have been sur-

prisingly classified by OMA into another orthologous group. The generated tree (Figure 7) shows that these two proteins with established coiled-coil structural motifs (Supplementary Table S1) are also closely related to proteins annotated as peptidoglycan lytic exotransglycosylases, but the least related protein, in this case, is a putative internal virion protein originating from phage ACG-C91 [34]. Importantly, except for internal virion proteins, the typical phage endolysins (WVW77668.1 and self_RMKP5) were also identified in genomes of phages SHAK7163 and SHAK7704 (Table 1). This may suggest that, although internal virion proteins may degrade peptidoglycan similarly to endolysins, they should not be considered classical endolysins participating in the lysis of bacterial cells that allow phage particles to release [58].

Since more than one endolysin was identified in Killian [35] and ZCEC14 [40] phages, we suspected that some of them might be annotated incorrectly and decided to take a closer look at the phylogenetic positions of all of them to clarify this. We created trees for them based on the orthologous group identified by OMA. We observed that one of the Killian putative endolysins (WEU67851.1), with predicted anaerobic ribonucleoside-triphosphate reductase feature (PF13597, Table 1), was grouped and closely related to many proteins annotated as large sub-units of ribonucleotide reductases (Figure 8). Another Killian putative endolysin (WEU68017.1), as well as a ZCEC14 putative endolysin (UVD33329.1) with a predicted 5′-3′ exonuclease, N-terminal resolvase-like domain (PF02739, Table 1), and T4 RNase H motif (PF09293, Table 1), turned out to be related to many ribonucleases H and one exonuclease (Figure 9). Another ZCEC14-derived putative endolysin (UVD33388.1), with prohead protease inhibitor (IPR016594, Supplementary Table S1) was, in turn, mostly related to inhibitors of head proteases and a few hypothetical proteins (Figure 10). Finally, the last two putative endolysins of Killian and ZCEC14 phages (WEU67833.1 and UVD33512.1) were grouped on the tree with many hypothetical proteins (Figure 11). Importantly, no specific endolysin domains were found within their sequences (Table 1).

Based on the presented results of phylogenetic analysis, we suggest that some proteins of anti-UPEC phages originally annotated as endolysins may play other roles. Phages in which more than one endolysin was annotated were particularly suspicious to us, and these doubts were clarified above. Our assumptions were confirmed because some of the originally annotated phage endolysins were found to be closely related to proteins with completely different functions (Figures 8–10). In addition, we found that proteins annotated as internal virion proteins with endolysin domains differ significantly from typical endolysins in size and domain content and are known to play a different role than typical endolysins. As reviewed by Rodriguez-Rubio [58], virion-associated peptidoglycan hydrolases (VAPGHs) are enzymes that locally degrade the peptidoglycan of the bacterial cell wall during the beginning of the phage infection. In contrast to endolysins that mediate the lysis of the host bacteria at the end of the lytic cycle, the action of VAPGHs generates a small hole through which the phage tail tube crosses the cell envelope to inject the phage genetic material at the beginning of the infection cycle.

On the other hand, phylogenetic analysis performed here allowed us to identify other putative endolysins that were initially annotated by the authors as lysins, lysozymes, amidases, hydrolases, or hypothetical proteins but which may serve as phage endolysins. All of them are gathered in Supplementary Tables S3 and S4 and are shown to have AMPs in their sequences (Supplementary Table S5).

Undoubtedly, the annotation of phage endolysins is not an easy process due to the diversity of domains and motifs present in their sequences. Nevertheless, such analysis should be performed more carefully and be verified by other methods, including phylogenetic position determination. Such an approach may allow us to obtain more reliable results. In addition, the nomenclature of endolysins should be standardized. Currently, the name

of an annotated protein is frequently assigned based on the identified domains and not the function it probably performs in the phage life cycle. Endolysins are very problematic in this respect as some authors annotate these proteins according to their purpose, and others according to the identified domain, which introduces confusion and difficulties in their identification.

Importantly, phage endolysins might also be grouped based on the similarity to endolysins produced by model phages, simply as (I) lysozymes of T4-like phages, (II) endopeptidases of T5-like phages, (III) amidases of T7-like phages, (IV) small peptidases of SP6-like phages, and (V) lysosymes of lambda-like phages. Considering this, we decided to perform detailed characteristics of model endolysins that confirmed the presence of appropriate domains and predicted their high stability (Supplementary Table S6). Then, we analyzed all the endolysins tested in this work in terms of their similarity to endolysins of model phages. Interestingly, we revealed that sequences of analyzed proteins were similar to model endolysins at varied levels, from 0.85% to 96.97% (Supplementary Table S7). However, the similarity at the level of about 1% was calculated for internal virion proteins, which, as we considered above, do not belong to typical endolysins. Among the classical endolysins, the protein originated from the LNA4 phage was the least similar to any reference endolysin, showing only 4.32% and 9.95% similarity to endolysins of phage T5 and T4, respectively. Importantly, we did not find any endolysin that was similar to λ phage protein in more than 5%. This observation coincides with the fact that only endolysins derived from lytic phages are considered in this work. Interestingly, 18 of the studied proteins showed very high sequence similarity (> 90%) to the reference endolysins. In total, a similarity above 70% was noted in 41 cases. For details, please see Supplementary Table S7.

4. In Vitro and In Vivo Testing of Endolysin Properties, Including Their Capability to Eradicate Biofilms

The formation of bacterial biofilms is a huge problem in UTI treatment. The harsh conditions within the urinary system, such as the flushing of cells during urination, force bacteria to develop strong adhesion mechanisms. In addition to the walls of the human bladder, bacterial biofilms can also form on medical surfaces, like catheters, leading to catheter-associated urinary tract infections (CAUTIs) [25]. Importantly, bacteria living in biofilm structures are frequently highly resistant to antibiotics; thus, addressing biofilms is crucial in the treatment of urinary tract infections.

Phage enzymes, including endolysins, show great promise in combating infections caused by antibiotic-resistant bacterial strains. Research on bactericidal proteins targeting Gram-negative bacteria is challenging due to the presence of the additional outer membrane that complicates access to peptidoglycan, the substrate for these enzymes [48]. The situation becomes even more complex when biofilms are involved, as the matrix serves as an additional layer of protection for bacterial cells, making it difficult for antibacterial agents to penetrate it [59]. Such challenges forced the necessity of searching for new effective treatment options. An interesting direction seems to study the effects of phage endolysins on biofilm-forming Gram-negative bacteria, also in combination with bacteriophages or antibiotics. One significant advantage of using phage enzymes is the lower likelihood of bacteria developing resistance to them, unlike traditional bacteriophage or antibiotic applications, where resistance cannot be completely ruled out [60]. The literature contains numerous reports on the activity of endolysins against bacteria from the ESKAPE group, such as endolysins LysAB1245 [61], LysJEP8 [62], and LysPA26 [63], targeting biofilms produced by *Pseudomonas aeruginosa*, or endolysins LysAm24, LysAp22, LysECD7, and LysSi3 [64] against biofilms of *Klebsiella pneumoniae*, *Acinetobacter baumannii*, and *P. aerug-*

inosa. However, there are relatively few reports on endolysins targeting uropathogens, particularly UPEC strains.

In the study conducted by Kajsikova et al. [65], the impact of an endolysin named EN572-5 on clinical uropathogenic isolates, primarily Gram-positive *Streptococcus agalactiae* bacteria, was investigated. The study also included four Gram-negative UPEC strains. The endolysin EN572-5, derived from *S. agalactiae* prophage KMB-572-E, exhibited lytic activity against 32 strains of *S. agalactiae*, averaging an impressive 83% effectiveness. The authors emphasized that no correlation could be observed between the enzyme's activity and the type of bacterial strain, confirming that phage enzymes are characterized by a broader spectrum of activity, in contrast to bacteriophages. On the other hand, endolysin EN572-5 showed activity against UPEC strains at low levels of 16-20%. However, its action was tested without the addition of compounds that destabilize the outer membrane—a primary barrier for this type of enzyme. This highlights how the outer membrane hinders the enzyme's access to peptidoglycan in Gram-negative bacteria. Importantly, this enzyme was also inactive against bacteria playing a protective role, like *Lactobacillus* spp., whose absence as a result of antibiotic treatments for UTIs can lead to complications such as bacterial vaginosis [65].

In the study conducted by Fursov et al. [66], the focus was on comparing the effects of the coliphage endolysin LysECD7 and amikacin against biofilm formation by a clinical multidrug-resistant strain of *K. pneumoniae* Ts 141-14, isolated from a patient's urine. It was demonstrated that both the endolysin and the antibiotic exhibited similar efficacy against forming biofilms in vitro. However, the enzyme showed greater bactericidal activity against mature biofilm, whose density was decreased by 60% and 68% compared to the control group. Notably, the concentrations used in this work (1000 and 3000 μg/mL) were highlighted by the authors as a potential limitation for the therapeutic application of the enzyme. The results of the in vivo experiments, however, indicated significant therapeutic potential for this endolysin. The researchers used rats implanted with specially designed chambers containing a nitrocellulose membrane filled with a suspension of *K. pneumoniae* Ts 141-14 cells. On the fourth day post-infection, treatment was initiated with LysECD7 administered daily for seven days, delivering 50 μg. This resulted in a reduction in biofilm biomass as early as two days after starting the therapy. On the eighth day of treatment, the colony count decreased from 1000 viable bacterial cells to an average of 58 cells. Amikacin produced a similar effect, although in this case, the number of bacteria increased compared to the fifth day, when the first measurement was taken [66].

Interestingly, in another study [67], the activity of three different endolysins derived from *Pseudomonas* phages was examined against various bacterial species, including clinical isolates obtained from patients with UTIs. The tested group included uropathogens such as *E. coli* strain 531, *K. pneumoniae* strains identified as ESBL (extended-spectrum beta-lactamases) producers, and *P. aeruginosa* strain 471. Additionally, other bacterial strains belonging to the ESKAPE group were also included. Notably, one of these endolysins, named PlyKp104 [66], exhibited the best bactericidal activity and a broad spectrum against Gram-negative bacteria. Its action reduced the number of viable cells (CFU/mL) of *P. aeruginosa*, *K. pneumoniae*, *E. coli*, and *A. baumannii* strains by five or six orders of magnitude in most tested variants. The authors did not use additional permeabilizing agents and attributed the activity of the PlyKp104 endolysin against Gram-negative bacteria to the presence of a positively charged C-terminal region, which interacts with the outer membrane. This natural ability to kill Gram-negative bacteria undoubtedly highlights its great potential in combating Gram-negative pathogens, including uropathogenic strains. The knowledge about PlyKp104 biofilm-eradicating capabilities is highly valuable and the researchers tested its lytic activity in vivo using a mouse skin wound infection model with *K. pneumo-*

niae. A single application of 300 µg of PlyKp104 was used, and after three hours, a reduction in the number of viable bacterial cells by two orders of magnitude was observed [67].

An interesting approach to overcoming the outer membrane of Gram-negative bacteria involved the use of a modified endolysin Lysep3 [68] with additional hydrophobic amino acid residues. The study examined the native endolysin and variants with added hydrophobic residues at the C-terminus, consisting of three, five, seven, and twelve residues. Enzymes with 12 residues were tested in two variants differing in amino acid composition. All proteins were applied externally. Endolysins with a higher number of hydrophobic residues exhibited greater bactericidal activity against *E. coli* bacteria compared to the unmodified enzyme. The activity increased with the number of residues, while no differences were noted regarding amino acid composition, as the degree of hydrophobicity was the same [68]. Meanwhile, Hasan et al. [69] used lactic acid in combination with two endolysins, LysEP114 and LysEP135, to combat biofilm formation by the *E. coli* KCCM 40405 strain. When applied together, the enzymes and lactic acid demonstrated significantly greater antibacterial and antibiofilm effects than when used separately. The numbers of biofilm cells were reduced by more than three and five logs for LysEP114 and LysEP135, respectively. The authors highlight an important observation that lactic acid's ability to permeabilize the outer membrane of Gram-negative bacteria gives it an advantage, particularly for applications such as food safety control [69]. In the field of food safety, significant attention is focused on the *E. coli* O157:H7 strain, a member of the STEC (Shiga toxin-producing *E. coli*) group, which causes severe gastrointestinal symptoms and is often linked to contaminated fruits and vegetables. It has been demonstrated that the endolysins LysECP26 [70], LysPECP14, and LysPECP20 [71] can effectively kill bacterial cells after pre-treatment with EDTA. However, the referenced studies examined only the bactericidal effect on liquid cultures, and the biofilm-disrupting potential of these enzymes remains unknown. Polymyxin B or cecropin A fused at the N-terminus, which is an antibacterial peptide, are capable of disrupting the structure of the outer membrane and have also been shown to effectively eradicate planktonic cells of *K. pneumoniae*, *P. aeruginosa*, *A. baumannii*, and *E. coli* in combination with the endolysin LysKP213 both in vitro and in the *Galleria mellonella* model [72]. A particularly interesting fact is that the mentioned endolysin, LysKP213, is highly thermostable. Its lytic activity remains at the level of 44.4% after 20 h of incubation at 95 °C, while after 30 min of autoclaving at 121 °C, it retains activity at up to 57.5%. This is valuable information regarding the potential applications of this protein in biotechnology, the medical industry, and food preservation [72].

In the context of the safety of therapies using endolysins against Gram-negative bacteria, including *E. coli*, the findings presented by Hwang et al. [73] are particularly intriguing. The researchers demonstrated that the endolysin LysPA90, encoded by the bacteriophage PBPA40, which infects *P. aeruginosa* strains, affects the motility of *E. coli* AIEC strain (adherent invasive *E. coli*). Sub-inhibitory concentrations of this protein caused stress in bacterial cells and increased the expression of genes responsible for flagella biosynthesis. This resulted in enhanced adhesion to intestinal epithelial cells of the Caco-2 line. The increased adhesion of bacterial cells was accompanied by elevated levels of proinflammatory cytokines in eukaryotic cells [73]. Therefore, in addition to studying the activity of enzymes against pathogens, very thorough and precise research is necessary to assess the safety of such therapies. The study mentioned earlier demonstrates that endolysins, even if they do not exhibit bactericidal activity, can influence the bacterial phenotype in ways that may be unfavorable for the patient.

The growing antibiotic resistance is a global problem that creates an urgent need to find new or improved known alternative methods for treating infections caused by MDR strains. Endolysins, despite their limitations in terms of selective activity against Gram-

negative bacteria, represent an interesting and potentially effective solution, especially when combined with agents that disrupt the structure of the outer membrane of these bacteria. Structural modifications of endolysins to enhance their bactericidal potential are also worth noting. Such an alternative strategy involves fusing endolysins with components that facilitate their access to the peptidoglycan layer. For instance, the *E. coli* phage lysin Lysep3 has been fused with the endolysin-binding domain D8 from *Bacillus amyloliquefaciens*. On its own, Lysep3 cannot break through *E. coli*'s outer membrane. However, when combined with the D8 domain, it gains the ability to destroy the bacteria externally. This is because D8 weakens the outer membrane's structure, allowing the enzyme to reach and degrade the peptidoglycan layer [74]. Other solutions of this problem are artilysin, which involves the fusion of LPS-disrupting molecules to a lysin; lysocins, which are combined endolysins and bacteriocins; innolysins, which exploit phage receptor-binding proteins (RBPs) to bind surface receptors to destabilize the outer membrane, synergy with permeabilizing chemicals; other phage enzymes, such as depolymerases, that degrade capsular polysaccharides, or even antibiotics; and finally, adding cationic compounds or antimicrobial peptides (AMPs) to the N- or C- terminus, as reviewed recently [48,75]. Importantly, the combined antimicrobial effect of phage-derived endolysin and depolymerase against biofilm-forming *Salmonella* Typhimurium was reported by Kim et al. [76]. However, studies accounting for fusions of these proteins or the combined effect of depolymerases and phage endolysins on Gram-negative bacteria are still missing.

To fully assess the potential of phage-derived enzymes in combating infections caused by antibiotic-resistant bacteria, including UPEC, it is essential to optimize their action against biofilms, conduct comprehensive safety studies, and test the efficacy of these proteins in clinical practice.

5. Conclusions

Interestingly, not only whole bacterial viruses, but also phage-derived proteins with lytic properties, named endolysins, have gained researchers' interest in recent years. Endolysins seem to possess greater application potential than intact phage particles, making them very promising and effective candidates for use as alternatives to antibiotics, also in UTI treatment. The major advantages of endolysins are the following: rapid bacterial cell lysis, a low probability of developing resistance, synergistic action with other antimicrobial agents, and high efficacy in planktonic and biofilm cells. Moreover, through genetic modification of endolysins, their lytic spectrum of action may be changed. Due to the above-mentioned unique features, parental and recombinant endolysins have been applied in much research.

We observe a growing interest in both phages and their enzymes; however, the number of identified and well-characterized phage particles or lytic proteins is still too small and not sufficiently studied for them to be introduced into standard treatment. The first main problem of endolysins concerns the difficulties with their identification in phage genomes. Our analyses have shown that annotation programs can make mistakes and mark proteins as endolysins that finally turn out not to be ones. Moreover, the nomenclature of phage proteins is not unified, which introduces confusion in the marking. The second problem is a broad spectrum of endolysins, which, when applied as antimicrobial agents, could affect the patient's microbiota and induce dysbiosis. Interesting solutions in this case can be the addition of a cell wall-binding domain to target only specific bacteria or phage RBPs. Other problems concern the difficulties with the purification of endolysins and their limited action from the outside of the cell, especially in the case of Gram-negative bacteria. The endolysin diffusion through the outer membrane of such bacteria is limited and, when administered exogenously, they are not able to reach their site of action—

peptidoglycan. In general, three main approaches have recently been explored to bypass the outer membrane of Gram-negative bacteria. These include utilizing lysins with intrinsic activity stemming from a positively charged C-terminus that destabilizes the membrane, employing physical or chemical methods to disrupt membrane integrity, and applying protein engineering to obtain the fusion of endolysins to components that allow them to gain access to the peptidoglycan layer. Undoubtedly, phage endolysins still face many challenges; however, they seem to be promising and effective candidates for use against bacteria, including UPEC.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/v17040560/s1: Supplementary Table S1. Data obtained by InterProScan version 5.72-103 software, giving an overview of the domains identified within the annotated endolysins. Supplementary Table S2. The list of AreMP peptides identified within sequences of putative endolysins. Supplementary Table S3. Characteristics of the proteins identified by phylogenetic analysis as closely related to the identified endolysins. Supplementary Table S4. Data obtained by InterProScan version 5.72-103 software, giving an overview of the domains identified within the proteins found by phylogenetic analysis as closely related to the tested endolysins. Supplementary Table S5. The list of AMP peptides found within sequences of the proteins identified by phylogenetic analysis as closely related to the tested endolysins. Supplementary Table S6. Data providing an overview of the size, genome position, stability, and domains identified within the endolysins derived from model phages λ, T4, T5, T7, and SP6. Supplementary Table S7. Results of the comparative analysis of all endolysin sequences tested in this work and those derived from model phages λ, T4, T5, T7, and SP6.

Author Contributions: Conceptualization, B.N-F. and W.W.; methodology, W.W.; investigation, W.W., A.Ł., S.B. and J.N.; data curation, W.W., A.Ł., S.B. and K.K.; writing—original draft preparation, W.W., A.Ł., S.B. and B.N.-F.; writing—review and editing, B.N.-F., G.W., E.W. and N.L.; visualization, W.W.; supervision, B.N.-F., S.B. and G.W.; project administration, A.Ł. and W.W.; funding acquisition, A.Ł. and W.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Science Centre, grant number 2023/49/N/NZ9/04058 and UG Young Reserecher grant number 539-D020-B084-23 to A.Ł., and by the UGrants-Start grant number 533-BG10-GS0I-24 to W.W.

Conflicts of Interest: Author Sylwia Bloch and Bożena Nejman-Faleńczyk were employed by the company BNF—New Bio Force Ltd., The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Gondil, V.S.; Chhibber, S. Bacteriophage and Endolysin Encapsulation Systems: A Promising Strategy to Improve Therapeutic Outcomes. *Front. Pharmacol.* **2021**, *12*, 675440. [CrossRef]
2. Abdelrahman, F.; Easwaran, M.; Daramola, O.I.; Ragab, S.; Lynch, S.; Oduselu, T.J.; Khan, F.M.; Ayobami, A.; Adnan, F.; Torrents, E.; et al. Phage-Encoded Endolysins. *Antibiotics* **2021**, *10*, 124. [CrossRef]
3. Principi, N.; Silvestri, E.; Esposito, S. Advantages and Limitations of Bacteriophages for the Treatment of Bacterial Infections. *Front. Pharmacol.* **2019**, *10*, 513. [CrossRef] [PubMed]
4. Fischetti, V.A. Development of Phage Lysins as Novel Therapeutics: A Historical Perspective. *Viruses* **2018**, *10*, 310. [CrossRef] [PubMed]
5. Theuretzbacher, U.; Piddock, L.J.V. Non-Traditional Antibacterial Therapeutic Options and Challenges. *Cell Host Microbe* **2019**, *26*, 61–72. [CrossRef] [PubMed]
6. Love, M.J.; Abeysekera, G.S.; Muscroft-Taylor, A.C.; Billington, C.; Dobson, R.C.J. On the Catalytic Mechanism of Bacteriophage Endolysins: Opportunities for Engineering. *Biochim. Biophys. Acta Proteins Proteom.* **2020**, *1868*, 140302. [CrossRef]
7. Li, X.; Zhang, C.; Wei, F.; Yu, F.; Zhao, Z. Bactericidal Activity of a Holin-Endolysin System Derived from Vibrio Alginolyticus Phage HH109. *Microb. Pathog.* **2021**, *159*, 105135. [CrossRef]
8. Nazir, A.; Xu, X.; Liu, Y.; Chen, Y. Phage Endolysins: Advances in the World of Food Safety. *Cells* **2023**, *12*, 2169. [CrossRef]

9. Khan, F.M.; Chen, J.-H.; Zhang, R.; Liu, B. A Comprehensive Review of the Applications of Bacteriophage-Derived Endolysins for Foodborne Bacterial Pathogens and Food Safety: Recent Advances, Challenges, and Future Perspective. *Front. Microbiol.* **2023**, *14*, 1259210. [CrossRef]
10. Oliveira, H.; São-José, C.; Azeredo, J. Phage-Derived Peptidoglycan Degrading Enzymes: Challenges and Future Prospects for In Vivo Therapy. *Viruses* **2018**, *10*, 292. [CrossRef]
11. Schmelcher, M.; Donovan, D.M.; Loessner, M.J. Bacteriophage Endolysins as Novel Antimicrobials. *Future Microbiol.* **2012**, *7*, 1147–1171. [CrossRef] [PubMed]
12. Microbial Products for Health, Environment and Agriculture. Available online: https://www.researchgate.net/publication/354735414_Microbial_Products_for_Health_Environment_and_Agriculture (accessed on 26 February 2025).
13. Liu, H.; Hu, Z.; Li, M.; Yang, Y.; Lu, S.; Rao, X. Therapeutic Potential of Bacteriophage Endolysins for Infections Caused by Gram-Positive Bacteria. *J. Biomed. Sci.* **2023**, *30*, 29. [CrossRef] [PubMed]
14. Lai, W.C.B.; Chen, X.; Ho, M.K.Y.; Xia, J.; Leung, S.S.Y. Bacteriophage-Derived Endolysins to Target Gram-Negative Bacteria. *Int. J. Pharm.* **2020**, *589*, 119833. [CrossRef]
15. Oechslin, F.; Zhu, X.; Dion, M.B.; Shi, R.; Moineau, S. Phage Endolysins Are Adapted to Specific Hosts and Are Evolutionarily Dynamic. *PLoS Biol.* **2022**, *20*, e3001740. [CrossRef]
16. Haddad Kashani, H.; Schmelcher, M.; Sabzalipoor, H.; Seyed Hosseini, E.; Moniri, R. Recombinant Endolysins as Potential Therapeutics against Antibiotic-Resistant Staphylococcus Aureus: Current Status of Research and Novel Delivery Strategies. *Clin. Microbiol. Rev.* **2018**, *31*, e00071-17. [CrossRef]
17. Rahman, M.U.; Wang, W.; Sun, Q.; Shah, J.A.; Li, C.; Sun, Y.; Li, Y.; Zhang, B.; Chen, W.; Wang, S. Endolysin, a Promising Solution against Antimicrobial Resistance. *Antibiotics* **2021**, *10*, 1277. [CrossRef]
18. São-José, C. Engineering of Phage-Derived Lytic Enzymes: Improving Their Potential as Antimicrobials. *Antibiotics* **2018**, *7*, 29. [CrossRef] [PubMed]
19. Ghose, C.; Euler, C.W. Gram-Negative Bacterial Lysins. *Antibiotics* **2020**, *9*, 74. [CrossRef]
20. Gontijo, M.T.P.; Jorge, G.P.; Brocchi, M. Current Status of Endolysin-Based Treatments against Gram-Negative Bacteria. *Antibiotics* **2021**, *10*, 1143. [CrossRef]
21. Gontijo, M.T.P.; Vidigal, P.M.P.; Lopez, M.E.S.; Brocchi, M. Bacteriophages That Infect Gram-Negative Bacteria as Source of Signal-Arrest-Release Motif Lysins. *Res. Microbiol.* **2021**, *172*, 103794. [CrossRef]
22. Zalewska-Piątek, B.; Piątek, R. Phage Therapy as a Novel Strategy in the Treatment of Urinary Tract Infections Caused by *E. coli*. *Antibiotics* **2020**, *9*, 304. [CrossRef]
23. Melican, K.; Sandoval, R.M.; Kader, A.; Josefsson, L.; Tanner, G.A.; Molitoris, B.A.; Richter-Dahlfors, A. Uropathogenic Escherichia Coli P and Type 1 Fimbriae Act in Synergy in a Living Host to Facilitate Renal Colonization Leading to Nephron Obstruction. *PLoS Pathog.* **2011**, *7*, e1001298. [CrossRef] [PubMed]
24. Chen, L.; Wen, Y. The Role of Bacterial Biofilm in Persistent Infections and Control Strategies. *Int. J. Oral Sci.* **2011**, *3*, 66–73. [CrossRef] [PubMed]
25. Lila, A.S.A.; Rajab, A.A.H.; Abdallah, M.H.; Rizvi, S.M.D.; Moin, A.; Khafagy, E.-S.; Tabrez, S.; Hegazy, W.A.H. Biofilm Lifestyle in Recurrent Urinary Tract Infections. *Life* **2023**, *13*, 148. [CrossRef]
26. Necel, A.; Bloch, S.; Topka-Bielecka, G.; Janiszewska, A.; Łukasiak, A.; Nejman-Faleńczyk, B.; Węgrzyn, G. Synergistic Effects of Bacteriophage vB_Eco4-M7 and Selected Antibiotics on the Biofilm Formed by Shiga Toxin-Producing *Escherichia coli*. *Antibiotics* **2022**, *11*, 712. [CrossRef]
27. Ngiam, L.; Schembri, M.A.; Weynberg, K.; Guo, J. Bacteriophage Isolated from Non-Target Bacteria Demonstrates Broad Host Range Infectivity against Multidrug-Resistant Bacteria. *Environ. Microbiol.* **2021**, *23*, 5569–5586. [CrossRef] [PubMed]
28. Chaudhary, N.; Maurya, R.K.; Singh, D.; Mohan, B.; Taneja, N. Genome Analysis and Antibiofilm Activity of Phage 590B against Multidrug-Resistant and Extensively Drug-Resistant Uropathogenic *Escherichia coli* Isolates, India. *Pathogens* **2022**, *11*, 1448. [CrossRef]
29. Rajab, A.A.H.; Fahmy, E.K.; Esmaeel, S.E.; Yousef, N.; Askoura, M. In Vitro and in Vivo Assessment of the Competence of a Novel Lytic Phage vB_EcoS_UTEC10 Targeting Multidrug Resistant *Escherichia coli* with a Robust Biofilm Eradication Activity. *Microb. Pathog.* **2024**, *197*, 107058. [CrossRef]
30. González-Villalobos, E.; Ribas-Aparicio, R.M.; Montealegre, G.E.R.; Belmont-Monroy, L.; Ortega-García, Y.; Aparicio-Ozores, G.; Balcázar, J.L.; Eslava-Campos, C.A.; Hernández-Chiñas, U.; Molina-López, J. Isolation and Characterization of Novel Bacteriophages as a Potential Therapeutic Option for *Escherichia coli* Urinary Tract Infections. *Appl. Microbiol. Biotechnol.* **2021**, *105*, 5617–5629. [CrossRef]
31. Chaudhary, N.; Singh, D.; Maurya, R.K.; Mohan, B.; Mavuduru, R.S.; Taneja, N. Whole Genome Sequencing and in Vitro Activity Data of *Escherichia* Phage NTEC3 against Multidrug-Resistant Uropathogenic and Extensively Drug-Resistant Uropathogenic *E. coli* Isolates. *Data Brief* **2022**, *43*, 108479. [CrossRef]

32. Yazdi, M.; Bouzari, M.; Ghaemi, E.A.; Shahin, K. Isolation, Characterization and Genomic Analysis of a Novel Bacteriophage VB_EcoS-Golestan Infecting Multidrug-Resistant *Escherichia coli* Isolated from Urinary Tract Infection. *Sci. Rep.* **2020**, *10*, 7690. [CrossRef] [PubMed]
33. Asgharzadeh Kangachar, S.; Logel, D.Y.; Trofimova, E.; Zhu, H.X.; Zaugg, J.; Schembri, M.A.; Weynberg, K.D.; Jaschke, P.R. Discovery and Characterisation of New Phage Targeting Uropathogenic *Escherichia coli*. *Virology* **2024**, *597*, 110148. [CrossRef]
34. Chibeu, A.; Lingohr, E.J.; Masson, L.; Manges, A.; Harel, J.; Ackermann, H.-W.; Kropinski, A.M.; Boerlin, P. Bacteriophages with the Ability to Degrade Uropathogenic *Escherichia coli* Biofilms. *Viruses* **2012**, *4*, 471–487. [CrossRef]
35. Khunti, P.; Chantakorn, K.; Tantibhadrasapa, A.; Htoo, H.H.; Thiennimitr, P.; Nonejuie, P.; Chaikeeratisak, V.A. Novel Coli Myophage and Antibiotics Synergistically Inhibit the Growth of the Uropathogenic *E. coli* Strain CFT073 in Stoichiometric Niches. *Microbiol. Spectr.* **2023**, *11*, e00889-23. [CrossRef]
36. Loose, M.; Sáez Moreno, D.; Mutti, M.; Hitzenhammer, E.; Visram, Z.; Dippel, D.; Schertler, S.; Tišáková, L.P.; Wittmann, J.; Corsini, L.; et al. Natural Bred E2-Phages Have an Improved Host Range and Virulence against Uropathogenic *Escherichia coli* over Their Ancestor Phages. *Antibiotics* **2021**, *10*, 1337. [CrossRef]
37. Markusková, B.; Elnwrani, S.; Andrezál, M.; Sedláčková, T.; Szemes, T.; Slobodníková, L.; Kajsik, M.; Drahovská, H. Characterization of Bacteriophages Infecting Multidrug-Resistant Uropathogenic *Escherichia coli* Strains. *Arch. Virol.* **2024**, *169*, 142. [CrossRef] [PubMed]
38. Vera-Mansilla, J.; Sánchez, P.; Silva-Valenzuela, C.A.; Molina-Quiroz, R.C. Isolation and Characterization of Novel Lytic Phages Infecting Multidrug-Resistant *Escherichia coli*. *Microbiol. Spectr.* **2022**, *10*, e01678-21. [CrossRef]
39. Gu, Y.; Xu, Y.; Xu, J.; Yu, X.; Huang, X.; Liu, G.; Liu, X. Identification of Novel Bacteriophage vB_EcoP-EG1 with Lytic Activity against Planktonic and Biofilm Forms of Uropathogenic *Escherichia coli*. *Appl. Microbiol. Biotechnol.* **2019**, *103*, 315–326. [CrossRef] [PubMed]
40. Ismael, N.M.; Azzam, M.; Abdelmoteleb, M.; El-Shibiny, A. Phage vB_Ec_ZCEC14 to Treat Antibiotic-Resistant Escherichia Coli Isolated from Urinary Tract Infections. *Virol. J.* **2024**, *21*, 44. [CrossRef]
41. Slobodníková, L.; Markusková, B.; Kajsík, M.; Andrezál, M.; Straka, M.; Liptáková, A.; Drahovská, H. Characterization of Anti-Bacterial Effect of the Two New Phages against Uropathogenic *Escherichia coli*. *Viruses* **2021**, *13*, 1348. [CrossRef]
42. Bouras, G.; Nepal, R.; Houtak, G.; Psaltis, A.J.; Wormald, P.-J.; Vreugde, S. Pharokka: A Fast Scalable Bacteriophage Annotation Tool. *Bioinformatics* **2023**, *39*, btac776. [CrossRef] [PubMed]
43. Cock, P.J.A.; Antao, T.; Chang, J.T.; Chapman, B.A.; Cox, C.J.; Dalke, A.; Friedberg, I.; Hamelryck, T.; Kauff, F.; Wilczynski, B.; et al. Biopython: Freely Available Python Tools for Computational Molecular Biology and Bioinformatics. *Bioinformatics* **2009**, *25*, 1422–1423. [CrossRef]
44. Jung, F.; Frey, K.; Zimmer, D.; Mühlhaus, T. DeepSTABp: A Deep Learning Approach for the Prediction of Thermal Protein Stability. *Int. J. Mol. Sci.* **2023**, *24*, 7444. [CrossRef] [PubMed]
45. Ikai, A. Thermostability and Aliphatic Index of Globular Proteins. *J. Biochem.* **1980**, *88*, 1895–1898. [PubMed]
46. Jones, P.; Binns, D.; Chang, H.-Y.; Fraser, M.; Li, W.; McAnulla, C.; McWilliam, H.; Maslen, J.; Mitchell, A.; Nuka, G.; et al. InterProScan 5: Genome-Scale Protein Function Classification. *Bioinformatics* **2014**, *30*, 1236–1240. [CrossRef]
47. Blum, M.; Chang, H.-Y.; Chuguransky, S.; Grego, T.; Kandasaamy, S.; Mitchell, A.; Nuka, G.; Paysan-Lafosse, T.; Qureshi, M.; Raj, S.; et al. The InterPro Protein Families and Domains Database: 20 Years on. *Nucleic Acids Res.* **2021**, *49*, D344–D354. [CrossRef]
48. Murray, E.; Draper, L.A.; Ross, R.P.; Hill, C. The Advantages and Challenges of Using Endolysins in a Clinical Setting. *Viruses* **2021**, *13*, 680. [CrossRef]
49. Fenton, M.; Ross, P.; McAuliffe, O.; O'Mahony, J.; Coffey, A. Recombinant Bacteriophage Lysins as Antibacterials. *Bioeng. Bugs* **2010**, *1*, 9–16. [CrossRef]
50. Marchler-Bauer, A.; Bo, Y.; Han, L.; He, J.; Lanczycki, C.J.; Lu, S.; Chitsaz, F.; Derbyshire, M.K.; Geer, R.C.; Gonzales, N.R.; et al. CDD/SPARCLE: Functional Classification of Proteins via Subfamily Domain Architectures. *Nucleic Acids Res.* **2017**, *45*, D200–D203. [CrossRef]
51. Lu, S.; Wang, J.; Chitsaz, F.; Derbyshire, M.K.; Geer, R.C.; Gonzales, N.R.; Gwadz, M.; Hurwitz, D.I.; Marchler, G.H.; Song, J.S.; et al. CDD/SPARCLE: The Conserved Domain Database in 2020. *Nucleic Acids Res.* **2020**, *48*, D265–D268. [CrossRef]
52. Wang, J.; Chitsaz, F.; Derbyshire, M.K.; Gonzales, N.R.; Gwadz, M.; Lu, S.; Marchler, G.H.; Song, J.S.; Thanki, N.; Yamashita, R.A.; et al. The Conserved Domain Database in 2023. *Nucleic Acids Res.* **2023**, *51*, D384–D388. [CrossRef] [PubMed]
53. Turnau, K.; Fiałkowska, E.; Ważny, R.; Rozpądek, P.; Tylko, G.; Bloch, S.; Nejman-Faleńczyk, B.; Grabski, M.; Węgrzyn, A.; Węgrzyn, G. Extraordinary Multi-Organismal Interactions Involving Bacteriophages, Bacteria, Fungi, and Rotifers: Quadruple Microbial Trophic Network in Water Droplets. *Int. J. Mol. Sci.* **2021**, *22*, 2178. [CrossRef]
54. Altenhoff, A.M.; Levy, J.; Zarowiecki, M.; Tomiczek, B.; Warwick Vesztrocy, A.; Dalquen, D.A.; Müller, S.; Telford, M.J.; Glover, N.M.; Dylus, D.; et al. OMA Standalone: Orthology Inference among Public and Custom Genomes and Transcriptomes. *Genome Res.* **2019**, *29*, 1152–1163. [CrossRef] [PubMed]

55. Minh, B.Q.; Schmidt, H.A.; Chernomor, O.; Schrempf, D.; Woodhams, M.D.; von Haeseler, A.; Lanfear, R. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **2020**, *37*, 1530–1534. [CrossRef]
56. Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T.J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J.; et al. Fast, Scalable Generation of High-Quality Protein Multiple Sequence Alignments Using Clustal Omega. *Mol. Syst. Biol.* **2011**, *7*, 539. [CrossRef]
57. Tamura, K.; Stecher, G.; Kumar, S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol. Biol. Evol.* **2021**, *38*, 3022–3027. [CrossRef]
58. Rodríguez-Rubio, L.; Martínez, B.; Donovan, D.M.; Rodríguez, A.; García, P. Bacteriophage virion-associated peptidoglycan hydrolases: Potential new enzybiotics. *Crit. Rev. Microbiol.* **2013**, *39*, 427–434. [CrossRef]
59. Zhao, A.; Sun, J.; Liu, Y. Understanding Bacterial Biofilms: From Definition to Treatment Strategies. *Front. Cell. Infect. Microbiol.* **2023**, *13*, 1137947. [CrossRef] [PubMed]
60. Gondil, V.S.; Harjai, K.; Chhibber, S. Endolysins as Emerging Alternative Therapeutic Agents to Counter Drug-Resistant Infections. *Int. J. Antimicrob. Agents* **2020**, *55*, 105844. [CrossRef]
61. Soontarach, R.; Srimanote, P.; Voravuthikunchai, S.P.; Chusri, S. Antibacterial and Anti-Biofilm Efficacy of Endolysin LysAB1245 against a Panel of Important Pathogens. *Pharmaceuticals* **2024**, *17*, 155. [CrossRef]
62. Carratalá, J.V.; Ferrer-Miralles, N.; Garcia-Fruitós, E.; Arís, A. LysJEP8: A Promising Novel Endolysin for Combating Multidrug-Resistant Gram-Negative Bacteria. *Microb. Biotechnol.* **2024**, *17*, e14483. [CrossRef] [PubMed]
63. Guo, M.; Feng, C.; Ren, J.; Zhuang, X.; Zhang, Y.; Zhu, Y.; Dong, K.; He, P.; Guo, X.; Qin, J. A Novel Antimicrobial Endolysin, LysPA26, against Pseudomonas Aeruginosa. *Front. Microbiol.* **2017**, *8*, 293. [CrossRef]
64. Vasina, D.V.; Antonova, N.P.; Grigoriev, I.V.; Yakimakha, V.S.; Lendel, A.M.; Nikiforova, M.A.; Pochtovyi, A.A.; Remizov, T.A.; Usachev, E.V.; Shevlyagina, N.V.; et al. Discovering the Potentials of Four Phage Endolysins to Combat Gram-Negative Infections. *Front. Microbiol.* **2021**, *12*, 748718. [CrossRef] [PubMed]
65. Kajsikova, M.; Kajsik, M.; Bocanova, L.; Papayova, K.; Drahovska, H.; Bukovska, G. Endolysin EN572-5 as an Alternative to Treat Urinary Tract Infection Caused by Streptococcus Agalactiae. *Appl. Microbiol. Biotechnol.* **2024**, *108*, 79. [CrossRef]
66. Fursov, M.V.; Abdrakhmanova, R.O.; Antonova, N.P.; Vasina, D.V.; Kolchanova, A.D.; Bashkina, O.A.; Rubalsky, O.V.; Samotrueva, M.A.; Potapov, V.D.; Makarov, V.V.; et al. Antibiofilm Activity of a Broad-Range Recombinant Endolysin LysECD7: In Vitro and In Vivo Study. *Viruses* **2020**, *12*, 545. [CrossRef]
67. Euler, C.W.; Raz, A.; Hernandez, A.; Serrano, A.; Xu, S.; Andersson, M.; Zou, G.; Zhang, Y.; Fischetti, V.A.; Li, J. PlyKp104, a Novel Phage Lysin for the Treatment of Klebsiella Pneumoniae, Pseudomonas Aeruginosa, and Other Gram-Negative ESKAPE Pathogens. *Antimicrob. Agents Chemother.* **2023**, *67*, e0151922. [CrossRef] [PubMed]
68. Yan, G.; Yang, R.; Fan, K.; Dong, H.; Gao, C.; Wang, S.; Yu, L.; Cheng, Z.; Lei, L. External Lysis of *Escherichia coli* by a Bacteriophage Endolysin Modified with Hydrophobic Amino Acids. *AMB Express* **2019**, *9*, 106. [CrossRef]
69. Hasan, M.; Kim, J.; Liao, X.; Ding, T.; Ahn, J. Antibacterial Activity of Bacteriophage-Encoded Endolysins against Planktonic and Biofilm Cells of Pathogenic *Escherichia coli*. *Microb. Pathog.* **2024**, *193*, 106780. [CrossRef]
70. Park, D.-W.; Park, J.-H. Characterization of Endolysin LysECP26 Derived from rV5-like Phage vB_EcoM-ECP26 for Inactivation of *Escherichia coli* O157:H7. *J. Microbiol. Biotechnol.* **2020**, *30*, 1552–1558. [CrossRef]
71. Oh, M.; Cevallos-Urena, A.; Kim, B.S. Bacteriophages PECP14, PECP20, and Their Endolysins as Effective Biocontrol Agents for *Escherichia coli* O157:H7 and Other Foodborne Pathogens. *Int. J. Food Microbiol.* **2024**, *409*, 110460. [CrossRef]
72. Chu, D.; Lan, J.; Liang, L.; Xia, K.; Li, L.; Yang, L.; Liu, H.; Zhang, T. The Antibacterial Activity of a Novel Highly Thermostable Endolysin, LysKP213, against Gram-Negative Pathogens Is Enhanced When Combined with Outer Membrane Permeabilizing Agents. *Front. Microbiol.* **2024**, *15*, 1454618. [CrossRef] [PubMed]
73. Hwang, Y.J.; Jo, J.; Kim, E.; Yoon, H.; Hong, H.; Kim, M.S.; Myung, H. Motility Increase of Adherent Invasive *Escherichia coli* (AIEC) Induced by a Sub-Inhibitory Concentration of Recombinant Endolysin LysPA90. *Front. Microbiol.* **2022**, *13*, 1093670. [CrossRef] [PubMed]
74. Wang, S.; Gu, J.; Lv, M.; Guo, Z.; Yan, G.; Yu, L.; Du, C.; Feng, X.; Han, W.; Sun, C. The antibacterial activity of *E. coli* bacteriophage lysin lysep3 is enhanced by fusing the *Bacillus amyloliquefaciens* bacteriophage endolysin binding domain D8 to the C-terminal region. *J. Microbiol.* **2017**, *55*, 403–408. [CrossRef] [PubMed]
75. Sisson, H.M.; Jackson, S.A.; Fagerlund, R.D.; Warring, S.L.; Fineran, P.C. Gram-negative endolysins: Overcoming the outer membrane obstacle. *Curr. Opin. Microbiol.* **2024**, *78*, 102433. [CrossRef]
76. Kim, J.; Wang, J.; Ahn, J. Combined antimicrobial effect of phage-derived endolysin and depolymerase against biofilm-forming *Salmonella* Typhimurium. *Biofouling* **2023**, *39*, 763–774. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Suplement do pracy nr 4

Supplementary Table S1. Data obtained by InterProScan version 5.72-103 software, giving an overview of the domains identified within the annotated endolysins.

Supplementary Table S2. The list of AMP peptides identified within sequences of putative endolysins.

Supplementary Table S3. Characteristics of the proteins identified by phylogenetic analysis as closely related to the identified endolysins.

Supplementary Table S4. Data obtained by InterProScan version 5.72-103 software, giving an overview of the domains identified within the proteins found by phylogenetic analysis as closely related to the tested endolysins.

Supplementary Table S5. The list of AMP peptides found within sequences of the proteins identified by phylogenetic analysis as closely related to the tested endolysins.

Supplementary Table S6. Data providing an overview of the size, genome position, stability, and domains identified within the endolysins derived from model phages λ, T4, T5, T7, and SP6.

Supplementary Table S7. Results of the comparative analysis of all endolysin sequences tested in this work and those derived from model phages λ, T4, T5, T7, and SP6.

Supplementary Table S1. Data obtained by InterProScan version 5.72-103 software, giving an overview of the domains identified within the annotated endolysins. The domains marked in yellow are the ones selected and listed in Table 1.

Phage name	Protein name	Protein accession number	Position of the gene in virus genome	Sequence length	Analysis	Signature accession	Signature description	Start location	Stop location	Score	InterPro annotations - accession	InterPro annotations - description	GO annotations
590B_MW722821	endolysin	QVR48622.1	39903 - 40391	162	CDD	cd00737	lyz_endolysin_autolysin	9	146	1.67708E-30	IPR033907	Endolysin/autolysin	GO:0003796(InterPro)
590B_MW722821	endolysin	QVR48622.1	39903 - 40391	162	Gene3D	G3DSA:1.10.530.40	-	2	147	3.5E-32	IPR023347	Lysozyme domain superfamily	GO:0003796(InterPro)
590B_MW722821	endolysin	QVR48622.1	39903 - 40391	162	PANTHER	PTHR38107	-	3	144	1.6E-22	IPR051018	Bacteriophage Glycosyl Hydrolase 24	GO:0003796(PANTHER)
590B_MW722821	endolysin	QVR48622.1	39903 - 40391	162	Pfam	PF00959	Phage lysozyme	15	138	2.4E-23	IPR002196	Glycoside hydrolase, family 24	GO:0003796(InterPro)|GO:0009253(InterPro)|GO:0016998(InterPro)
590B_MW722821	endolysin	QVR48622.1	39903 - 40391	162	SUPERFAMILY	SSF53955	Lysozyme-like	3	149	5.12E-32	IPR023346	Lysozyme-like domain superfamily	-
vB_EcoS-phiEc4_MZ576218	endolysin	QYU43915.1	41414 - 41899	161	CDD	cd00737	lyz_endolysin_autolysin	9	146	8.68255E-29	IPR033907	Endolysin/autolysin	GO:0003796(InterPro)
vB_EcoS-phiEc4_MZ576218	endolysin	QYU43915.1	41414 - 41899	161	Gene3D	G3DSA:1.10.530.40	-	2	147	1.7E-30	IPR023347	Lysozyme domain superfamily	GO:0003796(InterPro)
vB_EcoS-phiEc4_MZ576218	endolysin	QYU43915.1	41414 - 41899	161	PANTHER	PTHR38107	-	3	144	1.2E-21	IPR051018	Bacteriophage Glycosyl Hydrolase 24	GO:0003796(PANTHER)
vB_EcoS-phiEc4_MZ576218	endolysin	QYU43915.1	41414 - 41899	161	Pfam	PF00959	Phage lysozyme	15	138	1.9E-22	IPR002196	Glycoside hydrolase, family 24	GO:0003796(InterPro)|GO:0009253(InterPro)|GO:0016998(InterPro)
vB_EcoS-phiEc4_MZ576218	endolysin	QYU43915.1	41414 - 41899	161	SUPERFAMILY	SSF53955	Lysozyme-like	3	149	2.3E-30	IPR023346	Lysozyme-like domain superfamily	-
vB_EcoP-22664BS1_MZ234009	endolysin	QZI78465.1	19873 - 20226	117	CDD	cd06583	PGRP	5	98	6.48269E-24	IPR002502	N-acetylmuramoyl-L-alanine amidase domain	GO:0008745(InterPro)|GO:0009253(InterPro)
vB_EcoP-22664BS1_MZ234009	endolysin	QZI78465.1	19873 - 20226	117	Coils	Coil	Coil	67	87	-	-	-	-
vB_EcoP-22664BS1_MZ234009	endolysin	QZI78465.1	19873 - 20226	117	Gene3D	G3DSA:3.40.80.10	-	1	117	2.3E-41	IPR036505	N-acetylmuramoyl-L-alanine amidase/PGRP domain superfamily	GO:0008745(InterPro)|GO:0009253(InterPro)
vB_EcoP-22664BS1_MZ234009	endolysin	QZI78465.1	19873 - 20226	117	PANTHER	PTHR11022	PEPTIDOGLYCAN RECOGNITION PROTEIN	4	103	4.3E-17	IPR015510	Peptidoglycan recognition protein	-
vB_EcoP-22664BS1_MZ234009	endolysin	QZI78465.1	19873 - 20226	117	Pfam	PF01510	N-acetylmuramoyl-L-alanine amidase	7	99	2.5E-11	IPR002502	N-acetylmuramoyl-L-alanine amidase domain	GO:0008745(InterPro)|GO:0009253(InterPro)
vB_EcoP-22664BS1_MZ234009	endolysin	QZI78465.1	19873 - 20226	117	SMART	SM00644	ami_2	1	99	0.008	IPR002502	N-acetylmuramoyl-L-alanine amidase domain	GO:0008745(InterPro)|GO:0009253(InterPro)
vB_EcoP-22664BS1_MZ234009	endolysin	QZI78465.1	19873 - 20226	117	SMART	SM00701	pgrp	2	93	9.1E-6	IPR006619	Peptidoglycan recognition protein family domain, metazoa/bacteria	GO:0008270(InterPro)|GO:0008745(InterPro)|GO:0009253(InterPro)
vB_EcoP-22664BS1_MZ234009	endolysin	QZI78465.1	19873 - 20226	117	SUPERFAMILY	SSF55846	N-acetylmuramoyl-L-alanine amidase-like	2	112	4.45E-37	IPR036505	N-acetylmuramoyl-L-alanine amidase/PGRP domain superfamily	GO:0008745(InterPro)|GO:0009253(InterPro)

vB_EcoP-22664UKE3-2_MZ234011	endolysin	QZI78611.1	21058 - 21516	152	CDD	cd06583	PGRP	14	133	4.70396E-27	IPR002502	N-acetylmuramoyl-L-alanine amidase domain	GO:0008745(InterPro)|GO:0009253(InterPro)
vB_EcoP-22664UKE3-2_MZ234011	endolysin	QZI78611.1	21058 - 21516	152	Coils	Coil	Coil	102	122	-	-	-	-
vB_EcoP-22664UKE3-2_MZ234011	endolysin	QZI78611.1	21058 - 21516	152	Gene3D	G3DSA:3.40.80.10	-	8	152	1.3E-49	IPR036505	N-acetylmuramoyl-L-alanine amidase/PGRP domain superfamily	GO:0008745(InterPro)|GO:0009253(InterPro)
vB_EcoP-22664UKE3-2_MZ234011	endolysin	QZI78611.1	21058 - 21516	152	Hamap	MF_04111	Endolysin.	2	150	47.00668	IPR034689	Endolysin T7 type	GO:0008745(InterPro)|GO:0009253(InterPro)|GO:0032897(InterPro)|GO:0044659(InterPro)
vB_EcoP-22664UKE3-2_MZ234011	endolysin	QZI78611.1	21058 - 21516	152	PANTHER	PTHR11022	PEPTIDOGLYCAN RECOGNITION PROTEIN	15	137	4.0E-18	IPR015510	Peptidoglycan recognition protein	-
vB_EcoP-22664UKE3-2_MZ234011	endolysin	QZI78611.1	21058 - 21516	152	Pfam	PF01510	N-acetylmuramoyl-L-alanine amidase	11	134	1.4E-13	IPR002502	N-acetylmuramoyl-L-alanine amidase domain	GO:0008745(InterPro)|GO:0009253(InterPro)
vB_EcoP-22664UKE3-2_MZ234011	endolysin	QZI78611.1	21058 - 21516	152	SMART	SM00644	ami_2	2	134	1.2E-19	IPR002502	N-acetylmuramoyl-L-alanine amidase domain	GO:0008745(InterPro)|GO:0009253(InterPro)
vB_EcoP-22664UKE3-2_MZ234011	endolysin	QZI78611.1	21058 - 21516	152	SMART	SM00701	pgrp	2	128	7.1E-18	IPR006619	Peptidoglycan recognition protein family domain, metazoa/bacteria	GO:0008270(InterPro)|GO:0008745(InterPro)|GO:0009253(InterPro)
vB_EcoP-22664UKE3-2_MZ234011	endolysin	QZI78611.1	21058 - 21516	152	SUPERFAMILY	SSF55846	N-acetylmuramoyl-L-alanine amidase-like	11	147	4.58E-42	IPR036505	N-acetylmuramoyl-L-alanine amidase/PGRP domain superfamily	GO:0008745(InterPro)|GO:0009253(InterPro)
vB_EcoP-101118UKE1_MZ234021	endolysin	QZI79721.1	20127 - 20585	152	CDD	cd06583	PGRP	14	133	6.62816E-26	IPR002502	N-acetylmuramoyl-L-alanine amidase domain	GO:0008745(InterPro)|GO:0009253(InterPro)
vB_EcoP-101118UKE1_MZ234021	endolysin	QZI79721.1	20127 - 20585	152	Coils	Coil	Coil	102	122	-	-	-	-
vB_EcoP-101118UKE1_MZ234021	endolysin	QZI79721.1	20127 - 20585	152	Gene3D	G3DSA:3.40.80.10	-	8	152	1.8E-48	IPR036505	N-acetylmuramoyl-L-alanine amidase/PGRP domain superfamily	GO:0008745(InterPro)|GO:0009253(InterPro)
vB_EcoP-101118UKE1_MZ234021	endolysin	QZI79721.1	20127 - 20585	152	Hamap	MF_04111	Endolysin.	2	150	47.156376	IPR034689	Endolysin T7 type	GO:0008745(InterPro)|GO:0009253(InterPro)|GO:0032897(InterPro)|GO:0044659(InterPro)
vB_EcoP-101118UKE1_MZ234021	endolysin	QZI79721.1	20127 - 20585	152	PANTHER	PTHR11022	PEPTIDOGLYCAN RECOGNITION PROTEIN	16	138	2.1E-17	IPR015510	Peptidoglycan recognition protein	-
vB_EcoP-101118UKE1_MZ234021	endolysin	QZI79721.1	20127 - 20585	152	Pfam	PF01510	N-acetylmuramoyl-L-alanine amidase	11	134	7.1E-13	IPR002502	N-acetylmuramoyl-L-alanine amidase domain	GO:0008745(InterPro)|GO:0009253(InterPro)
vB_EcoP-101118UKE1_MZ234021	endolysin	QZI79721.1	20127 - 20585	152	SMART	SM00644	ami_2	2	134	3.9E-18	IPR002502	N-acetylmuramoyl-L-alanine amidase domain	GO:0008745(InterPro)|GO:0009253(InterPro)
vB_EcoP-101118UKE1_MZ234021	endolysin	QZI79721.1	20127 - 20585	152	SMART	SM00701	pgrp	2	128	2.0E-16	IPR006619	Peptidoglycan recognition protein family domain, metazoa/bacteria	GO:0008270(InterPro)|GO:0008745(InterPro)|GO:0009253(InterPro)

vB_EcoP-101118UKE1_MZ234021	endolysin	QZI79721.1	20127 - 20585	152	SUPERFAMILY	SSF55846	N-acetylmuramoyl-L-alanine amidase-like	11	147	1.31E-41	IPR036505	N-acetylmuramoyl-L-alanine amidase/PGRP domain superfamily	GO:0008745(InterPro)|GO:0009253(InterPro)
vB_EcoP-101120B1-2_MZ234022	endolysin	QZI79784.1	20840 - 21298	152	CDD	cd06583	PGRP	14	133	9.31561E-27	IPR002502	N-acetylmuramoyl-L-alanine amidase domain	GO:0008745(InterPro)|GO:0009253(InterPro)
vB_EcoP-101120B1-2_MZ234022	endolysin	QZI79784.1	20840 - 21298	152	Coils	Coil	Coil	102	122	-	-	-	-
vB_EcoP-101120B1-2_MZ234022	endolysin	QZI79784.1	20840 - 21298	152	Gene3D	G3DSA:3.40.80.10	-	8	152	1.6E-49	IPR036505	N-acetylmuramoyl-L-alanine amidase/PGRP domain superfamily	GO:0008745(InterPro)|GO:0009253(InterPro)
vB_EcoP-101120B1-2_MZ234022	endolysin	QZI79784.1	20840 - 21298	152	Hamap	MF_04111	Endolysin.	2	150	47.136852	IPR034689	Endolysin T7 type	GO:0008745(InterPro)|GO:0009253(InterPro)|GO:0032897(InterPro)|GO:0044659(InterPro)
vB_EcoP-101120B1-2_MZ234022	endolysin	QZI79784.1	20840 - 21298	152	PANTHER	PTHR11022	PEPTIDOGLYCAN RECOGNITION PROTEIN	16	137	5.4E-18	IPR015510	Peptidoglycan recognition protein	-
vB_EcoP-101120B1-2_MZ234022	endolysin	QZI79784.1	20840 - 21298	152	Pfam	PF01510	N-acetylmuramoyl-L-alanine amidase	11	134	2.3E-13	IPR002502	N-acetylmuramoyl-L-alanine amidase domain	GO:0008745(InterPro)|GO:0009253(InterPro)
vB_EcoP-101120B1-2_MZ234022	endolysin	QZI79784.1	20840 - 21298	152	SMART	SM00644	ami_2	2	134	3.4E-19	IPR002502	N-acetylmuramoyl-L-alanine amidase domain	GO:0008745(InterPro)|GO:0009253(InterPro)
vB_EcoP-101120B1-2_MZ234022	endolysin	QZI79784.1	20840 - 21298	152	SMART	SM00701	pgrp	2	128	6.8E-18	IPR006619	Peptidoglycan recognition protein family domain, metazoa/bacteria	GO:0008270(InterPro)|GO:0008745(InterPro)|GO:0009253(InterPro)
vB_EcoP-101120B1-2_MZ234022	endolysin	QZI79784.1	20840 - 21298	152	SUPERFAMILY	SSF55846	N-acetylmuramoyl-L-alanine amidase-like	11	147	9.03E-42	IPR036505	N-acetylmuramoyl-L-alanine amidase/PGRP domain superfamily	GO:0008745(InterPro)|GO:0009253(InterPro)
vB_EcoP-101120B2_MZ234023	endolysin	QZI79844.1	20895 - 21353	152	CDD	cd06583	PGRP	14	133	9.31561E-27	IPR002502	N-acetylmuramoyl-L-alanine amidase domain	GO:0008745(InterPro)|GO:0009253(InterPro)
vB_EcoP-101120B2_MZ234023	endolysin	QZI79844.1	20895 - 21353	152	Coils	Coil	Coil	102	122	-	-	-	-
vB_EcoP-101120B2_MZ234023	endolysin	QZI79844.1	20895 - 21353	152	Gene3D	G3DSA:3.40.80.10	-	8	152	1.6E-49	IPR036505	N-acetylmuramoyl-L-alanine amidase/PGRP domain superfamily	GO:0008745(InterPro)|GO:0009253(InterPro)
vB_EcoP-101120B2_MZ234023	endolysin	QZI79844.1	20895 - 21353	152	Hamap	MF_04111	Endolysin.	2	150	47.136852	IPR034689	Endolysin T7 type	GO:0008745(InterPro)|GO:0009253(InterPro)|GO:0032897(InterPro)|GO:0044659(InterPro)
vB_EcoP-101120B2_MZ234023	endolysin	QZI79844.1	20895 - 21353	152	PANTHER	PTHR11022	PEPTIDOGLYCAN RECOGNITION PROTEIN	16	137	5.4E-18	IPR015510	Peptidoglycan recognition protein	-
vB_EcoP-101120B2_MZ234023	endolysin	QZI79844.1	20895 - 21353	152	Pfam	PF01510	N-acetylmuramoyl-L-alanine amidase	11	134	2.3E-13	IPR002502	N-acetylmuramoyl-L-alanine amidase domain	GO:0008745(InterPro)|GO:0009253(InterPro)
vB_EcoP-101120B2_MZ234023	endolysin	QZI79844.1	20895 - 21353	152	SMART	SM00644	ami_2	2	134	3.4E-19	IPR002502	N-acetylmuramoyl-L-alanine amidase domain	GO:0008745(InterPro)|GO:0009253(InterPro)

vB_EcoP-101120B2_MZ234023	endolysin	QZI79844.1	20895 - 21353	152	SMART	SM00701	pgrp	2	128	6.8E-18	IPR006619	Peptidoglycan recognition protein family domain, metazoa/bacteria	GO:0008270(InterPro)|GO:0008745(InterPro)|GO:0009253(InterPro)
vB_EcoP-101120B2_MZ234023	endolysin	QZI79844.1	20895 - 21353	152	SUPERFAMILY	SSF55846	N-acetylmuramoyl-L-alanine amidase-like	11	147	9.03E-42	IPR036505	N-acetylmuramoyl-L-alanine amidase/PGRP domain superfamily	GO:0008745(InterPro)|GO:0009253(InterPro)
vB_EcoP-101136BS1_MZ234024	endolysin	QZI79903.1	19728 - 20186	152	CDD	cd06583	PGRP	14	133	2.01169E-26	IPR002502	N-acetylmuramoyl-L-alanine amidase domain	GO:0008745(InterPro)|GO:0009253(InterPro)
vB_EcoP-101136BS1_MZ234024	endolysin	QZI79903.1	19728 - 20186	152	Coils	Coil	Coil	102	122	-	-	-	-
vB_EcoP-101136BS1_MZ234024	endolysin	QZI79903.1	19728 - 20186	152	Gene3D	G3DSA:3.40.80.10	-	8	152	7.6E-49	IPR036505	N-acetylmuramoyl-L-alanine amidase/PGRP domain superfamily	GO:0008745(InterPro)|GO:0009253(InterPro)
vB_EcoP-101136BS1_MZ234024	endolysin	QZI79903.1	19728 - 20186	152	Hamap	MF_04111	Endolysin.	2	150	47.195431	IPR034689	Endolysin T7 type	GO:0008745(InterPro)|GO:0009253(InterPro)|GO:0032897(InterPro)|GO:0044659(InterPro)
vB_EcoP-101136BS1_MZ234024	endolysin	QZI79903.1	19728 - 20186	152	PANTHER	PTHR11022	PEPTIDOGLYCAN RECOGNITION PROTEIN	16	138	1.7E-17	IPR015510	Peptidoglycan recognition protein	-
vB_EcoP-101136BS1_MZ234024	endolysin	QZI79903.1	19728 - 20186	152	Pfam	PF01510	N-acetylmuramoyl-L-alanine amidase	11	134	7.2E-13	IPR002502	N-acetylmuramoyl-L-alanine amidase domain	GO:0008745(InterPro)|GO:0009253(InterPro)
vB_EcoP-101136BS1_MZ234024	endolysin	QZI79903.1	19728 - 20186	152	SMART	SM00644	ami_2	2	134	1.8E-18	IPR002502	N-acetylmuramoyl-L-alanine amidase domain	GO:0008745(InterPro)|GO:0009253(InterPro)
vB_EcoP-101136BS1_MZ234024	endolysin	QZI79903.1	19728 - 20186	152	SMART	SM00701	pgrp	2	128	1.3E-17	IPR006619	Peptidoglycan recognition protein family domain, metazoa/bacteria	GO:0008270(InterPro)|GO:0008745(InterPro)|GO:0009253(InterPro)
vB_EcoP-101136BS1_MZ234024	endolysin	QZI79903.1	19728 - 20186	152	SUPERFAMILY	SSF55846	N-acetylmuramoyl-L-alanine amidase-like	11	147	1.18E-41	IPR036505	N-acetylmuramoyl-L-alanine amidase/PGRP domain superfamily	GO:0008745(InterPro)|GO:0009253(InterPro)
vB_EcoM_SHAK7854_OR594182	internal_virion_protein_with_endolysin_domain	self_DNYN53	32030 - 35917	1295	Phobius	TRANSMEMBRANE	Region of a membrane-bound protein predicted to be embedded in the membrane.	1114	1131	-	-	-	-
vB_EcoM_SHAK7854_OR594182	internal_virion_protein_with_endolysin_domain	self_DNYN53	32030 - 35917	1295	Phobius	CYTOPLASMIC_DOMAIN	Region of a membrane-bound protein predicted to be outside the membrane, in the cytoplasm.	1132	1295	-	-	-	-
vB_EcoM_SHAK7854_OR594182	internal_virion_protein_with_endolysin_domain	self_DNYN53	32030 - 35917	1295	Phobius	NON_CYTOPLASMIC_DOMAIN	Region of a membrane-bound protein predicted to be outside the membrane, in the extracellular region.	1	1113	-	-	-	-
vB_EcoP_LNA3_MZ311862	internal_virion_protein_with_endolysin_domain	self_GXIJ6	5943 - 9830	1295	Phobius	TRANSMEMBRANE	Region of a membrane-bound protein predicted to be embedded in the membrane.	1114	1131	-	-	-	-

vB_EcoP_LNA3_MZ311862	internal_virion_protein_with_endolysin_domain	self_GXIJ6	5943 - 9830	1295	Phobius	CYTOPLASMIC_DOMAIN	Region of a membrane-bound protein predicted to be outside the membrane, in the cytoplasm.	1132	1295	-	-	-	-
vB_EcoP_LNA3_MZ311862	internal_virion_protein_with_endolysin_domain	self_GXIJ6	5943 - 9830	1295	Phobius	NON_CYTOPLASMIC_DOMAIN	Region of a membrane-bound protein predicted to be outside the membrane, in the extracellular region.	1	1113	-	-	-	-
vB_EcoM_LNA4_MZ311864	endolysin	self_PFXJ204	117557 - 118129	190	Gene3D	G3DSA:1.10.10.2520	Cell wall hydrolase SleB, domain 1	55	131	3.1E-18	IPR042047	Cell wall hydrolase SleB, domain 1	-
vB_EcoM_LNA4_MZ311864	endolysin	self_PFXJ204	117557 - 118129	190	Pfam	PF07486	Cell Wall Hydrolase	70	181	2.7E-17	IPR011105	Cell wall hydrolase, SleB	GO:0016787(InterPro)
vB_EcoM_LNA4_MZ311864	endolysin	self_PFXJ204	117557 - 118129	190	Phobius	SIGNAL_PEPTIDE_C_REGION	C-terminal region of a signal peptide.	32	37	-	-	-	-
vB_EcoM_LNA4_MZ311864	endolysin	self_PFXJ204	117557 - 118129	190	Phobius	SIGNAL_PEPTIDE_H_REGION	Hydrophobic region of a signal peptide.	20	31	-	-	-	-
vB_EcoM_LNA4_MZ311864	endolysin	self_PFXJ204	117557 - 118129	190	Phobius	SIGNAL_PEPTIDE_N_REGION	N-terminal region of a signal peptide.	1	19	-	-	-	-
vB_EcoM_LNA4_MZ311864	endolysin	self_PFXJ204	117557 - 118129	190	Phobius	NON_CYTOPLASMIC_DOMAIN	Region of a membrane-bound protein predicted to be outside the membrane, in the extracellular region.	38	190	-	-	-	-
vB_EcoM_LNA4_MZ311864	endolysin	self_PFXJ204	117557 - 118129	190	Phobius	SIGNAL_PEPTIDE	Signal peptide region	1	37	-	-	-	-
vB_EcoM_LNA4_MZ311864	endolysin	self_PFXJ59	21617 - 22171	184	Gene3D	G3DSA:1.10.530.10	-	3	183	7.1E-45	-	-	-
vB_EcoM_LNA4_MZ311864	endolysin	self_PFXJ59	21617 - 22171	184	PANTHER	PTHR34408	FAMILY PROTEIN, PUTATIVE-RELATED	31	178	1.5E-19	IPR052354	Bacterial Cell Wall Dynamics Protein	-
vB_EcoM_LNA4_MZ311864	endolysin	self_PFXJ59	21617 - 22171	184	Pfam	PF00182	Chitinase class I	41	146	7.3E-7	IPR000726	Glycoside hydrolase, family 19, catalytic	GO:0004568(InterPro)|GO:0006032(InterPro)|GO:0016998(InterPro)
vB_EcoM_LNA4_MZ311864	endolysin	self_PFXJ59	21617 - 22171	184	SUPERFAMILY	SSF53955	Lysozyme-like	3	182	4.22E-45	IPR023346	Lysozyme-like domain superfamily	-
vB_EcoM_SHAK7704_OR594181	endolysin	self_RMKP5	3327 - 3671	114	Gene3D	G3DSA:3.30.1380.10	-	5	114	1.1E-25	IPR009045	Hedgehog signalling/DD-peptidase zinc-binding domain superfamily	-
vB_EcoM_SHAK7704_OR594181	endolysin	self_RMKP5	3327 - 3671	114	Pfam	PF08291	Peptidase M15	5	106	3.4E-29	IPR013230	Peptidase M15A, C-terminal	-
vB_EcoM_SHAK7704_OR594181	endolysin	self_RMKP5	3327 - 3671	114	SUPERFAMILY	SSF55166	Hedgehog/DD-peptidase	5	107	1.29E-24	IPR009045	Hedgehog signalling/DD-peptidase zinc-binding domain superfamily	-
vB_EcoM_SHAK7704_OR594181	internal_virion_protein_with_endolysin_domain	self_RMKP62	39690 - 42998	1102	Coils	Coil	Coil	604	624	-	-	-	-
vB_EcoM_SHAK7704_OR594181	internal_virion_protein_with_endolysin_domain	self_RMKP62	39690 - 42998	1102	Coils	Coil	Coil	638	665	-	-	-	-
vB_EcoP_LNA5_MZ311861	internal_virion_protein_with_endolysin_domain	self_TCEH45	23530 - 27417	1295	MobiDBLite	mobidb-lite	consensus disorder prediction	1	21	-	-	-	-
vB_EcoP_LNA5_MZ311861	internal_virion_protein_with_endolysin_domain	self_TCEH45	23530 - 27417	1295	MobiDBLite	mobidb-lite	consensus disorder prediction	1	12	-	-	-	-

vB_EcoP_LNA5_MZ311861	internal_virion_protein_with_endolysin_domain	self_TCEH45	23530 - 27417	1295	Phobius	TRANSMEMBRANE	Region of a membrane-bound protein predicted to be embedded in the membrane.	1114	1131	-	-	-	-
vB_EcoP_LNA5_MZ311861	internal_virion_protein_with_endolysin_domain	self_TCEH45	23530 - 27417	1295	Phobius	CYTOPLASMIC_DOMAIN	Region of a membrane-bound protein predicted to be outside the membrane, in the cytoplasm.	1132	1295	-	-	-	-
vB_EcoP_LNA5_MZ311861	internal_virion_protein_with_endolysin_domain	self_TCEH45	23530 - 27417	1295	Phobius	NON_CYTOPLASMIC_DOMAIN	Region of a membrane-bound protein predicted to be outside the membrane, in the extracellular region.	1	1113	-	-	-	-
vB_EcoM_SHAK9454_OR594184	internal_virion_protein_with_endolysin_domain	self_ZENB53	32160 - 36047	1295	Phobius	TRANSMEMBRANE	Region of a membrane-bound protein predicted to be embedded in the membrane.	1114	1131	-	-	-	-
vB_EcoM_SHAK9454_OR594184	internal_virion_protein_with_endolysin_domain	self_ZENB53	32160 - 36047	1295	Phobius	CYTOPLASMIC_DOMAIN	Region of a membrane-bound protein predicted to be outside the membrane, in the cytoplasm.	1132	1295	-	-	-	-
vB_EcoM_SHAK9454_OR594184	internal_virion_protein_with_endolysin_domain	self_ZENB53	32160 - 36047	1295	Phobius	NON_CYTOPLASMIC_DOMAIN	Region of a membrane-bound protein predicted to be outside the membrane, in the extracellular region.	1	1113	-	-	-	-
vB_EcoM_LNA1_MZ311863	endolysin	self_ZKRH133	77244 - 77708	154	CDD	cd00737	lyz_endolysin_autolysin	7	148	2.64914E-63	IPR033907	Endolysin/autolysin	GO:0003796(InterPro)
vB_EcoM_LNA1_MZ311863	endolysin	self_ZKRH133	77244 - 77708	154	Gene3D	G3DSA:1.10.530.40	-	1	149	7.2E-47	IPR023347	Lysozyme domain superfamily	GO:0003796(InterPro)
vB_EcoM_LNA1_MZ311863	endolysin	self_ZKRH133	77244 - 77708	154	Hamap	MF_04110	Endolysin.	1	150	22.414141	IPR034690	Endolysin T4 type	GO:0003796(InterPro)|GO:0009253(InterPro)
vB_EcoM_LNA1_MZ311863	endolysin	self_ZKRH133	77244 - 77708	154	PANTHER	PTHR38107	-	2	151	3.4E-38	IPR051018	Bacteriophage Glycosyl Hydrolase 24	GO:0003796(PANTHER)
vB_EcoM_LNA1_MZ311863	endolysin	self_ZKRH133	77244 - 77708	154	Pfam	PF00959	Phage lysozyme	13	143	3.1E-37	IPR002196	Glycoside hydrolase, family 24	GO:0003796(InterPro)|GO:0009253(InterPro)|GO:0016998(InterPro)
vB_EcoM_LNA1_MZ311863	endolysin	self_ZKRH133	77244 - 77708	154	SUPERFAMILY	SSF53955	Lysozyme-like	1	152	7.16E-45	IPR023346	Lysozyme-like domain superfamily	-
ZCEC14_OP168898	endolysin	UVD33329.1	13508 - 14425	305	CDD	cd09899	H3TH_T4-like	184	259	3.49981E-12	-	-	-
ZCEC14_OP168898	endolysin	UVD33329.1	13508 - 14425	305	CDD	cd00008	PIN_53EXO-like	17	180	5.97673E-76	-	-	-
ZCEC14_OP168898	endolysin	UVD33329.1	13508 - 14425	305	Gene3D	G3DSA:1.10.150.20	-	183	259	6.2E-121	-	-	-
ZCEC14_OP168898	endolysin	UVD33329.1	13508 - 14425	305	Gene3D	G3DSA:3.40.50.1010	-	17	292	6.2E-121	-	-	-
ZCEC14_OP168898	endolysin	UVD33329.1	13508 - 14425	305	PANTHER	PTHR42646	FLAP ENDONUCLEASE XNI	60	202	5.4E-12	IPR038969	Flap endonuclease	GO:0017108(InterPro)|GO:0033567(InterPro)
ZCEC14_OP168898	endolysin	UVD33329.1	13508 - 14425	305	Pfam	PF02739	5'-3' exonuclease, N-terminal resolvase-like domain	51	174	9.3E-11	IPR020046	5'-3' exonuclease, alpha-helical arch, N-terminal	GO:0003677(InterPro)
ZCEC14_OP168898	endolysin	UVD33329.1	13508 - 14425	305	Pfam	PF09293	T4 RNase H, C terminal	183	305	3.0E-66	IPR036276	T4 RNase H, C-terminal	-
ZCEC14_OP168898	endolysin	UVD33329.1	13508 - 14425	305	SMART	SM00475	53exo3	38	258	0.0062	IPR002421	5'-3' exonuclease	GO:0003677(InterPro)|GO:0008409(InterPro)
ZCEC14_OP168898	endolysin	UVD33329.1	13508 - 14425	305	SMART	SM00279	HhH_4	185	220	0.0029	IPR008918	Helix-hairpin-helix motif, class 2	GO:0003677(InterPro)|GO:0003824(InterPro)
ZCEC14_OP168898	endolysin	UVD33329.1	13508 - 14425	305	SUPERFAMILY	SSF47807	5' to 3' exonuclease, C-terminal subdomain	183	305	8.89E-51	IPR036279	5'-3' exonuclease, C-terminal domain superfamily	-

ZCEC14_OP168898	endolysin	UVD33329.1	13508 - 14425	305	SUPERFAMILY	SSF88723	PIN domain-like	13	179	3.36E-39	IPR029060	PIN-like domain superfamily	-
ZCEC14_OP168898	endolysin	UVD33388.1	51539 - 52219	226	PIRSF	PIRSF012159	Inh_gp21	1	226	7.9E-103	IPR016594	Bacteriophage T4, inh	-
Killian_OQ446694	endolysin	WEU67851.1	54687 - 56504	605	CDD	cd01675	RNR_III	21	584	0.0	IPR012833	Ribonucleoside-triphosphate reductase, anaerobic	GO:0006260(InterPro)|GO:0008998(InterPro)|GO:0016491(InterPro)
Killian_OQ446694	endolysin	WEU67851.1	54687 - 56504	605	Gene3D	G3DSA:3.20.70.20	-	26	586	7.2E-207	-	-	-
Killian_OQ446694	endolysin	WEU67851.1	54687 - 56504	605	NCBIfam	TIGR02487	anaerobic ribonucleoside-triphosphate reductase	21	604	3.9E-165	IPR012833	Ribonucleoside-triphosphate reductase, anaerobic	GO:0006260(InterPro)|GO:0008998(InterPro)|GO:0016491(InterPro)
Killian_OQ446694	endolysin	WEU67851.1	54687 - 56504	605	PANTHER	PTHR21075	ANAEROBIC RIBONUCLEOSIDE-TRIPHOSPHATE REDUCTASE	5	605	5.6E-133	-	-	GO:0004748(PANTHER)|GO:0008998(PANTHER)|GO:0009265(PANTHER)|GO:0031250(PANTHER)
Killian_OQ446694	endolysin	WEU67851.1	54687 - 56504	605	Pfam	PF13597	Anaerobic ribonucleoside-triphosphate reductase	21	604	9.3E-167	IPR012833	Ribonucleoside-triphosphate reductase, anaerobic	GO:0006260(InterPro)|GO:0008998(InterPro)|GO:0016491(InterPro)
Killian_OQ446694	endolysin	WEU67851.1	54687 - 56504	605	ProSitePatterns	PS00850	Glycine radical domain signature.	575	583	-	IPR019777	Formate C-acetyltransferase glycine radical, conserved site	-
Killian_OQ446694	endolysin	WEU67851.1	54687 - 56504	605	ProSiteProfiles	PS51149	Glycine radical domain profile.	482	605	13.676711	IPR001150	Glycine radical domain	GO:0003824(InterPro)
Killian_OQ446694	endolysin	WEU67851.1	54687 - 56504	605	SUPERFAMILY	SSF51998	PFL-like glycyl radical enzymes	26	585	3.97E-194	-	-	-
Killian_OQ446694	endolysin	WEU67899.1	77692 - 78186	164	CDD	cd00735	T4-like_lys	2	160	9.7315E-78	-	-	-
Killian_OQ446694	endolysin	WEU67899.1	77692 - 78186	164	FunFam	G3DSA:1.10.530.40:FF:000002	Endolysin	1	164	2.6E-114	-	-	-
Killian_OQ446694	endolysin	WEU67899.1	77692 - 78186	164	Gene3D	G3DSA:1.10.530.40	-	1	164	1.6E-74	IPR023347	Lysozyme domain superfamily	GO:0003796(InterPro)
Killian_OQ446694	endolysin	WEU67899.1	77692 - 78186	164	Hamap	MF_04110	Endolysin.	1	159	16.697023	IPR034690	Endolysin T4 type	GO:0003796(InterPro)|GO:0009253(InterPro)
Killian_OQ446694	endolysin	WEU67899.1	77692 - 78186	164	PANTHER	PTHR37406	T4-TYPE LYSOZYME 1-RELATED	4	161	7.3E-28	IPR052619	Bacteriophage lysozyme-like	-
Killian_OQ446694	endolysin	WEU67899.1	77692 - 78186	164	Pfam	PF00959	Phage lysozyme	11	141	4.8E-10	IPR002196	Glycoside hydrolase, family 24	GO:0003796(InterPro)|GO:0009253(InterPro)|GO:0016998(InterPro)
Killian_OQ446694	endolysin	WEU67899.1	77692 - 78186	164	PRINTS	PR00684	Phage T4 lysozyme signature	95	114	3.9E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
Killian_OQ446694	endolysin	WEU67899.1	77692 - 78186	164	PRINTS	PR00684	Phage T4 lysozyme signature	24	42	3.9E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
Killian_OQ446694	endolysin	WEU67899.1	77692 - 78186	164	PRINTS	PR00684	Phage T4 lysozyme signature	72	92	3.9E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
Killian_OQ446694	endolysin	WEU67899.1	77692 - 78186	164	PRINTS	PR00684	Phage T4 lysozyme signature	4	23	3.9E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
Killian_OQ446694	endolysin	WEU67899.1	77692 - 78186	164	PRINTS	PR00684	Phage T4 lysozyme signature	118	137	3.9E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
Killian_OQ446694	endolysin	WEU67899.1	77692 - 78186	164	PRINTS	PR00684	Phage T4 lysozyme signature	51	70	3.9E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
Killian_OQ446694	endolysin	WEU67899.1	77692 - 78186	164	PRINTS	PR00684	Phage T4 lysozyme signature	138	159	3.9E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
Killian_OQ446694	endolysin	WEU67899.1	77692 - 78186	164	SUPERFAMILY	SSF53955	Lysozyme-like	1	161	8.19E-54	IPR023346	Lysozyme-like domain superfamily	-
Killian_OQ446694	endolysin	WEU68017.1	163120 - 164037	305	CDD	cd09899	H3TH_T4-like	184	259	3.49981E-12	-	-	-

Killian_OQ446694	endolysin	WEU68017.1	163120 - 164037	305	CDD	cd00008	PIN_53EXO-like	17	180	5.97673E-76	-	-	-
Killian_OQ446694	endolysin	WEU68017.1	163120 - 164037	305	Gene3D	G3DSA:1.10.150.20	-	183	259	6.2E-121	-	-	-
Killian_OQ446694	endolysin	WEU68017.1	163120 - 164037	305	Gene3D	G3DSA:3.40.50.1010	-	17	292	6.2E-121	-	-	-
Killian_OQ446694	endolysin	WEU68017.1	163120 - 164037	305	PANTHER	PTHR42646	FLAP ENDONUCLEASE XNI	60	202	5.4E-12	IPR038969	Flap endonuclease	GO:0017108(InterPro)|GO:0033567(InterPro)
Killian_OQ446694	endolysin	WEU68017.1	163120 - 164037	305	Pfam	PF02739	5'-3' exonuclease, N-terminal resolvase-like domain	51	174	9.3E-11	IPR020046	5'-3' exonuclease, alpha-helical arch, N-terminal	GO:0003677(InterPro)
Killian_OQ446694	endolysin	WEU68017.1	163120 - 164037	305	Pfam	PF09293	T4 RNase H, C terminal	183	305	3.0E-66	IPR036276	T4 RNase H, C-terminal	-
Killian_OQ446694	endolysin	WEU68017.1	163120 - 164037	305	SMART	SM00475	53exo3	38	258	0.0062	IPR002421	5'-3' exonuclease	GO:0003677(InterPro)|GO:0008409(InterPro)
Killian_OQ446694	endolysin	WEU68017.1	163120 - 164037	305	SMART	SM00279	HhH_4	185	220	0.0029	IPR008918	Helix-hairpin-helix motif, class 2	GO:0003677(InterPro)|GO:0003824(InterPro)
Killian_OQ446694	endolysin	WEU68017.1	163120 - 164037	305	SUPERFAMILY	SSF47807	5' to 3' exonuclease, C-terminal subdomain	183	305	8.89E-51	IPR036279	5'-3' exonuclease, C-terminal domain superfamily	-
Killian_OQ446694	endolysin	WEU68017.1	163120 - 164037	305	SUPERFAMILY	SSF88723	PIN domain-like	13	179	3.36E-39	IPR029060	PIN-like domain superfamily	-
vB-Eco-KMB14_OR424384	endolysin	WPK27911.1	38777 - 39121	114	Gene3D	G3DSA:3.30.1380.10	-	5	114	2.0E-26	IPR009045	Hedgehog signalling/DD-peptidase zinc-binding domain superfamily	-
vB-Eco-KMB14_OR424384	endolysin	WPK27911.1	38777 - 39121	114	Pfam	PF08291	Peptidase M15	4	106	1.3E-29	IPR013230	Peptidase M15A, C-terminal	-
vB-Eco-KMB14_OR424384	endolysin	WPK27911.1	38777 - 39121	114	SUPERFAMILY	SSF55166	Hedgehog/DD-peptidase	5	108	2.04E-27	IPR009045	Hedgehog signalling/DD-peptidase zinc-binding domain superfamily	-
vB-Eco-KMB43_OR525698	endolysin	WQN06429.1	55240 - 55629	129	CDD	cd14845	L-Ala-D-Glu_peptidase_like	11	120	1.31964E-27	-	-	-
vB-Eco-KMB43_OR525698	endolysin	WQN06429.1	55240 - 55629	129	Gene3D	G3DSA:3.30.1380.10	-	2	120	1.7E-20	IPR009045	Hedgehog signalling/DD-peptidase zinc-binding domain superfamily	-
vB-Eco-KMB43_OR525698	endolysin	WQN06429.1	55240 - 55629	129	Pfam	PF13539	D-alanyl-D-alanine carboxypeptidase	55	120	1.8E-4	IPR039561	Peptidase M15C	GO:0008233(InterPro)
vB-Eco-KMB43_OR525698	endolysin	WQN06429.1	55240 - 55629	129	SUPERFAMILY	SSF55166	Hedgehog/DD-peptidase	2	120	3.53E-23	IPR009045	Hedgehog signalling/DD-peptidase zinc-binding domain superfamily	-
vB_EcoM_SHAK7163_OR594179	internal_virion_protein_with_endolysin_domain	WVW77660.1	31648 - 34968	1106	Coils	Coil	Coil	608	628	-	-	-	-
vB_EcoM_SHAK7163_OR594179	internal_virion_protein_with_endolysin_domain	WVW77660.1	31648 - 34968	1106	Coils	Coil	Coil	642	669	-	-	-	-
vB_EcoM_SHAK7163_OR594179	endolysin	WVW77668.1	39749 - 40093	114	Gene3D	G3DSA:3.30.1380.10	-	5	114	4.2E-25	IPR009045	Hedgehog signalling/DD-peptidase zinc-binding domain superfamily	-

vB_EcoM_SHAK7163_OR594179	endolysin	WVW77668.1	39749 - 40093	114	Pfam	PF08291	Peptidase M15	5	106	5.4E-29	IPR013230	Peptidase M15A, C-terminal	-
vB_EcoM_SHAK7163_OR594179	endolysin	WVW77668.1	39749 - 40093	114	SUPERFAMILY	SSF55166	Hedgehog/DD-peptidase	5	107	9.02E-24	IPR009045	Hedgehog signalling/DD-peptidase zinc-binding domain superfamily	-
vB_EcoM_SHAK7693_OR594180	endolysin	WVW77753.1	42419 - 42658	79	Gene3D	G3DSA:1.10.530.40	-	6	72	7.9E-10	IPR023347	Lysozyme domain superfamily	GO:0003796(InterPro)
vB_EcoM_SHAK7693_OR594180	endolysin	WVW77753.1	42419 - 42658	79	MobiDBLite	mobidb-lite	consensus disorder prediction	56	79	-	-	-	-
vB_EcoM_SHAK7693_OR594180	endolysin	WVW77753.1	42419 - 42658	79	Pfam	PF00959	Phage lysozyme	20	67	2.3E-4	IPR002196	Glycoside hydrolase, family 24	GO:0003796(InterPro)|GO:0009253(InterPro)|GO:0016998(InterPro)
vB_EcoM_SHAK7693_OR594180	endolysin	WVW77753.1	42419 - 42658	79	SUPERFAMILY	SSF53955	Lysozyme-like	6	68	5.63E-12	IPR023346	Lysozyme-like domain superfamily	-
vB_EcoM_SHAK7693_OR594180	endolysin	WVW77754.1	42651 - 42953	100	Gene3D	G3DSA:1.10.530.40	-	1	85	4.7E-14	IPR023347	Lysozyme domain superfamily	GO:0003796(InterPro)
vB_EcoM_SHAK7693_OR594180	endolysin	WVW77754.1	42651 - 42953	100	Pfam	PF00959	Phage lysozyme	2	76	2.3E-8	IPR002196	Glycoside hydrolase, family 24	GO:0003796(InterPro)|GO:0009253(InterPro)|GO:0016998(InterPro)
vB_EcoM_SHAK7693_OR594180	endolysin	WVW77754.1	42651 - 42953	100	SUPERFAMILY	SSF53955	Lysozyme-like	2	86	1.33E-13	IPR023346	Lysozyme-like domain superfamily	-
vB_EcoM_SHAK7858_OR594183	endolysin	WVW77834.1	40989 - 41474	161	CDD	cd00737	lyz_endolysin_autolysin	8	145	7.46015E-29	IPR033907	Endolysin/autolysin	GO:0003796(InterPro)
vB_EcoM_SHAK7858_OR594183	endolysin	WVW77834.1	40989 - 41474	161	Gene3D	G3DSA:1.10.530.40	-	1	146	6.3E-30	IPR023347	Lysozyme domain superfamily	GO:0003796(InterPro)
vB_EcoM_SHAK7858_OR594183	endolysin	WVW77834.1	40989 - 41474	161	PANTHER	PTHR38107	-	2	143	5.5E-20	IPR051018	Bacteriophage Glycosyl Hydrolase 24	GO:0003796(PANTHER)
vB_EcoM_SHAK7858_OR594183	endolysin	WVW77834.1	40989 - 41474	161	Pfam	PF00959	Phage lysozyme	14	137	1.4E-20	IPR002196	Glycoside hydrolase, family 24	GO:0003796(InterPro)|GO:0009253(InterPro)|GO:0016998(InterPro)
vB_EcoM_SHAK7858_OR594183	endolysin	WVW77834.1	40989 - 41474	161	SUPERFAMILY	SSF53955	Lysozyme-like	4	148	2.61E-28	IPR023346	Lysozyme-like domain superfamily	-
vB_EcoS_UTEC10_PP291582	endolysin	WWY66481.1	1847 - 2332	161	CDD	cd00737	lyz_endolysin_autolysin	9	146	6.37925E-31	IPR033907	Endolysin/autolysin	GO:0003796(InterPro)
vB_EcoS_UTEC10_PP291582	endolysin	WWY66481.1	1847 - 2332	161	Gene3D	G3DSA:1.10.530.40	-	2	147	1.6E-32	IPR023347	Lysozyme domain superfamily	GO:0003796(InterPro)
vB_EcoS_UTEC10_PP291582	endolysin	WWY66481.1	1847 - 2332	161	PANTHER	PTHR38107	-	3	144	2.4E-22	IPR051018	Bacteriophage Glycosyl Hydrolase 24	-
vB_EcoS_UTEC10_PP291582	endolysin	WWY66481.1	1847 - 2332	161	Pfam	PF00959	Phage lysozyme	15	138	1.2E-23	IPR002196	Glycoside hydrolase, family 24	GO:0003796(InterPro)|GO:0009253(InterPro)|GO:0016998(InterPro)
vB_EcoS_UTEC10_PP291582	endolysin	WWY66481.1	1847 - 2332	161	SUPERFAMILY	SSF53955	Lysozyme-like	3	149	5.01E-32	IPR023346	Lysozyme-like domain superfamily	-
vB_EcoP_ACG-C91_NC_019403	endolysin	YP_006987811.1	39340 - 39684	114	Gene3D	G3DSA:3.30.1380.10	-	5	114	7.2E-26	IPR009045	Hedgehog signalling/DD-peptidase zinc-binding domain superfamily	-
vB_EcoP_ACG-C91_NC_019403	endolysin	YP_006987811.1	39340 - 39684	114	Pfam	PF08291	Peptidase M15	4	106	3.0E-29	IPR013230	Peptidase M15A, C-terminal	-

vB_EcoP_ACG-C91_NC_019403	endolysin	YP_006987811.1	39340 - 39684	114	SUPERFAMILY	SSF55166	Hedgehog/DD-peptidase	5	107	2.39E-26	IPR009045	Hedgehog signalling/DD-peptidase zinc-binding domain superfamily	-
vB_EcoS_ACG-M12_NC_019404	endolysin	YP_006987885.1	41386 - 41871	161	CDD	cd16900	endolysin_R21-like	20	156	1.10155E-60	-	-	-
vB_EcoS_ACG-M12_NC_019404	endolysin	YP_006987885.1	41386 - 41871	161	Gene3D	G3DSA:1.10.530.40	-	1	160	4.9E-44	IPR023347	Lysozyme domain superfamily	GO:0003796(InterPro)
vB_EcoS_ACG-M12_NC_019404	endolysin	YP_006987885.1	41386 - 41871	161	Hamap	MF_04110	Endolysin.	4	158	27.074312	IPR034690	Endolysin T4 type	GO:0003796(InterPro)|GO:0009253(InterPro)
vB_EcoS_ACG-M12_NC_019404	endolysin	YP_006987885.1	41386 - 41871	161	Hamap	MF_04136	SAR-endolysin.	3	158	30.775623	IPR043688	SAR-endolysin-like	GO:0003796(InterPro)|GO:0016998(InterPro)
vB_EcoS_ACG-M12_NC_019404	endolysin	YP_006987885.1	41386 - 41871	161	PANTHER	PTHR38107	-	8	159	1.9E-32	IPR051018	Bacteriophage Glycosyl Hydrolase 24	GO:0003796(PANTHER)
vB_EcoS_ACG-M12_NC_019404	endolysin	YP_006987885.1	41386 - 41871	161	Pfam	PF00959	Phage lysozyme	25	151	3.9E-32	IPR002196	Glycoside hydrolase, family 24	GO:0003796(InterPro)|GO:0009253(InterPro)|GO:0016998(InterPro)
vB_EcoS_ACG-M12_NC_019404	endolysin	YP_006987885.1	41386 - 41871	161	Phobius	SIGNAL_PEPTIDE_C_REGION	C-terminal region of a signal peptide.	16	27	-	-	-	-
vB_EcoS_ACG-M12_NC_019404	endolysin	YP_006987885.1	41386 - 41871	161	Phobius	SIGNAL_PEPTIDE_H_REGION	Hydrophobic region of a signal peptide.	5	15	-	-	-	-
vB_EcoS_ACG-M12_NC_019404	endolysin	YP_006987885.1	41386 - 41871	161	Phobius	SIGNAL_PEPTIDE_N_REGION	N-terminal region of a signal peptide.	1	4	-	-	-	-
vB_EcoS_ACG-M12_NC_019404	endolysin	YP_006987885.1	41386 - 41871	161	Phobius	NON_CYTOPLASMIC_DOMAIN	Region of a membrane-bound protein predicted to be outside the membrane, in the extracellular region.	28	161	-	-	-	-
vB_EcoS_ACG-M12_NC_019404	endolysin	YP_006987885.1	41386 - 41871	161	Phobius	SIGNAL_PEPTIDE	Signal peptide region	1	27	-	-	-	-
vB_EcoS_ACG-M12_NC_019404	endolysin	YP_006987885.1	41386 - 41871	161	SUPERFAMILY	SSF53955	Lysozyme-like	6	159	1.02E-43	IPR023346	Lysozyme-like domain superfamily	-
VB_EcoS-Golestan_NC_042084	endolysin	YP_009620148.1	41444 - 41902	152	CDD	cd00737	lyz_endolysin_autolysin	3	137	6.04237E-30	IPR033907	Endolysin/autolysin	GO:0003796(InterPro)
VB_EcoS-Golestan_NC_042084	endolysin	YP_009620148.1	41444 - 41902	152	Gene3D	G3DSA:1.10.530.40	-	1	138	3.6E-31	IPR023347	Lysozyme domain superfamily	GO:0003796(InterPro)
VB_EcoS-Golestan_NC_042084	endolysin	YP_009620148.1	41444 - 41902	152	PANTHER	PTHR38107	-	2	135	5.1E-21	IPR051018	Bacteriophage Glycosyl Hydrolase 24	-
VB_EcoS-Golestan_NC_042084	endolysin	YP_009620148.1	41444 - 41902	152	Pfam	PF00959	Phage lysozyme	6	129	1.6E-23	IPR002196	Glycoside hydrolase, family 24	GO:0003796(InterPro)|GO:0009253(InterPro)|GO:0016998(InterPro)
VB_EcoS-Golestan_NC_042084	endolysin	YP_009620148.1	41444 - 41902	152	SUPERFAMILY	SSF53955	Lysozyme-like	2	141	3.48E-29	IPR023346	Lysozyme-like domain superfamily	-
EG1_NC_047895	internal_virion_protein_with_endolysin_domain	YP_009795842.1	30249 - 34223	1324	CDD	cd00254	LT-like	28	144	5.06794E-25	-	-	-
EG1_NC_047895	internal_virion_protein_with_endolysin_domain	YP_009795842.1	30249 - 34223	1324	Gene3D	G3DSA:1.10.530.10	-	5	151	9.6E-31	-	-	-
EG1_NC_047895	internal_virion_protein_with_endolysin_domain	YP_009795842.1	30249 - 34223	1324	Hamap	MF_04121	Peptidoglycan transglycosylase gp16.	2	1314	10.10026	IPR038994	Internal virion protein Gp16	GO:0039678(InterPro)
EG1_NC_047895	internal_virion_protein_with_endolysin_domain	YP_009795842.1	30249 - 34223	1324	PANTHER	PTHR37423	SOLUBLE LYTIC MUREIN TRANSGLYCOSYLASE-RELATED	7	143	2.9E-18	-	-	-

EG1_NC_047895	internal_virion_protein_with_endolysin_domain	YP_009795842.1	30249 - 34223	1324	Pfam	PF01464	Transglycosylase SLT domain	15	127	1.6E-25	IPR008258	Transglycosylase SLT domain 1	-
EG1_NC_047895	internal_virion_protein_with_endolysin_domain	YP_009795842.1	30249 - 34223	1324	ProSitePatterns	PS00922	Prokaryotic transglycosylases signature.	33	61	-	IPR000189	Prokaryotic transglycosylase, active site	GO:0000270(InterPro)|GO:0008933(InterPro)|GO:0016020(InterPro)
EG1_NC_047895	internal_virion_protein_with_endolysin_domain	YP_009795842.1	30249 - 34223	1324	SUPERFAMILY	SSF53955	Lysozyme-like	5	146	3.46E-28	IPR023346	Lysozyme-like domain superfamily	-
NTEC3_NC_073318	endolysin	YP_010749155.1	11540 - 12028	162	CDD	cd00737	lyz_endolysin_autolysin	9	146	8.49828E-29	IPR033907	Endolysin/autolysin	GO:0003796(InterPro)
NTEC3_NC_073318	endolysin	YP_010749155.1	11540 - 12028	162	Gene3D	G3DSA:1.10.530.40	-	2	147	1.7E-30	IPR023347	Lysozyme domain superfamily	GO:0003796(InterPro)
NTEC3_NC_073318	endolysin	YP_010749155.1	11540 - 12028	162	PANTHER	PTHR38107	-	3	144	1.3E-21	IPR051018	Bacteriophage Glycosyl Hydrolase 24	GO:0003796(PANTHER)
NTEC3_NC_073318	endolysin	YP_010749155.1	11540 - 12028	162	Pfam	PF00959	Phage lysozyme	15	138	2.0E-22	IPR002196	Glycoside hydrolase, family 24	GO:0003796(InterPro)|GO:0009253(InterPro)|GO:0016998(InterPro)
NTEC3_NC_073318	endolysin	YP_010749155.1	11540 - 12028	162	SUPERFAMILY	SSF53955	Lysozyme-like	3	149	2.35E-30	IPR023346	Lysozyme-like domain superfamily	-
vB_EcoS-phiEc3_NC_073323	endolysin	YP_010749568.1	41533 - 42021	162	CDD	cd00737	lyz_endolysin_autolysin	9	146	1.75147E-30	IPR033907	Endolysin/autolysin	GO:0003796(InterPro)
vB_EcoS-phiEc3_NC_073323	endolysin	YP_010749568.1	41533 - 42021	162	Gene3D	G3DSA:1.10.530.40	-	2	147	2.9E-32	IPR023347	Lysozyme domain superfamily	GO:0003796(InterPro)
vB_EcoS-phiEc3_NC_073323	endolysin	YP_010749568.1	41533 - 42021	162	PANTHER	PTHR38107	-	3	144	1.3E-22	IPR051018	Bacteriophage Glycosyl Hydrolase 24	GO:0003796(PANTHER)
vB_EcoS-phiEc3_NC_073323	endolysin	YP_010749568.1	41533 - 42021	162	Pfam	PF00959	Phage lysozyme	15	138	2.4E-23	IPR002196	Glycoside hydrolase, family 24	GO:0003796(InterPro)|GO:0009253(InterPro)|GO:0016998(InterPro)
vB_EcoS-phiEc3_NC_073323	endolysin	YP_010749568.1	41533 - 42021	162	SUPERFAMILY	SSF53955	Lysozyme-like	3	149	1.02E-31	IPR023346	Lysozyme-like domain superfamily	-
MLP1_NC_079176	endolysin	YP_010844467.1	32790 - 33185	131	CDD	cd14845	L-Ala-D-Glu_peptidase_like	9	130	1.8598E-23	-	-	-
MLP1_NC_079176	endolysin	YP_010844467.1	32790 - 33185	131	Gene3D	G3DSA:3.30.1380.10	-	3	129	9.1E-21	IPR009045	Hedgehog signalling/DD-peptidase zinc-binding domain superfamily	-
MLP1_NC_079176	endolysin	YP_010844467.1	32790 - 33185	131	Pfam	PF13539	D-alanyl-D-alanine carboxypeptidase	58	110	0.004	IPR039561	Peptidase M15C	GO:0008233(InterPro)
MLP1_NC_079176	endolysin	YP_010844467.1	32790 - 33185	131	SUPERFAMILY	SSF55166	Hedgehog/DD-peptidase	4	129	3.92E-19	IPR009045	Hedgehog signalling/DD-peptidase zinc-binding domain superfamily	-

Supplementary Table S2. The list of AMP peptides identified within sequences of putative endolysins.

Phage name	Protein name	Protein accession number	Position	Sequence	Class	AMP Probability
590B-MW722821	endolysin	QVR48622.1	123-142	FINQNGRPLLGLRRRTAGRL	AMP	0.950
VB-ECOS-PHIEC4-MZ576218	endolysin	QYU43915.1	123-142	FINQNGRPLLGLRRRTAGRL	AMP	0.950
VB-ECOP-22664BS1-MZ234009	endolysin	QZI78465.1	33-52	GSHVKDWNSRSVGVCLVGGI	AMP	0.840
VB-ECOP-22664UKE3-2-MZ234011	endolysin	QZI78611.1	68-87	GSHVKDWNSRSVGVCLVGGI	AMP	0.840
VB-ECOP-101118UKE1-MZ234021	endolysin	QZI79721.1	68-87	GSHVKDWNSRSVGVCLVGGI	AMP	0.840
VB-ECOP-101120B1-2-MZ234022	endolysin	QZI79784.1	68-87	GSHVKDWNSRSVGVCLVGGI	AMP	0.840
VB-ECOP-101120B2-MZ234023	endolysin	QZI79844.1	68-87	GSHVKDWNSRSVGVCLVGGI	AMP	0.840
VB-ECOP-101136BS1-MZ234024	endolysin	QZI79903.1	68-87	GSHVKDWNSRSVGVCLVGGI	AMP	0.840
VB-ECOM-SHAK7854-OR594182	internal-virion-protein-with-endolysin-domain	SELF-DNYN53	323-342	IAGQVGKGGKLVNKMFTVAA	AMP	0.987
VB-ECOP-LNA3-MZ311862	internal-virion-protein-with-endolysin-domain	SELF-GXIJ6	323-342	IAGQIGKGGKLVNKMFTVAA	AMP	0.990
VB-ECOM-LNA4-MZ311864	endolysin	SELF-PFXJ204	15-34	LKVKNLILATACAFSLATTP	AMP	0.937
VB-ECOM-LNA4-MZ311864	endolysin	SELF-PFXJ59	98-117	GLIQVTGLANYKACGKALGL	AMP	0.969
VB-ECOM-SHAK7704-OR594181	endolysin	SELF-RMKP5	50-69	AKHNANVGGAKNSMHLTGKA	AMP	0.892
VB-ECOM-SHAK7704-OR594181	internal-virion-protein-with-endolysin-domain	SELF-RMKP62	38-57	LQEQVLEAKADADILGAVGA	AMP	0.989
VB-ECOP-LNA5-MZ311861	internal-virion-protein-with-endolysin-domain	SELF-TCEH45	323-342	IAGQVGKGGKLVNKMFTVAA	AMP	0.987
VB-ECOM-SHAK9454-OR594184	internal-virion-protein-with-endolysin-domain	SELF-ZENB53	323-342	IAGQVGKGGKLVNKMFTVAA	AMP	0.987
VB-ECOM-LNA1-MZ311863	endolysin	SELF-ZKRH133	109-128	GNKVGCAEAMQWWNKITVKG	AMP	0.941
ZCEC14-OP168898	endolysin	UVD33329.1	135-154	IAVLVKKFSLEGHKILIISS	AMP	0.977
ZCEC14-OP168898	endolysin	UVD33388.1	94-113	LIDSPVKSDIKIEVVETDKI	AMP	0.963
ZCEC14-OP168898	endolysin	UVD33512.1	42-61	RIKKVTVCTDWAVSFNFNIL	AMP	0.532
KILLIAN-OQ446694	endolysin	WEU67833.1	42-61	CIKKVTVCTDWAVSFNFNIL	AMP	0.715
KILLIAN-OQ446694	endolysin	WEU67851.1	38-57	AGIVSKHIACRILPSHIVNA	AMP	0.940
KILLIAN-OQ446694	endolysin	WEU67899.1	41-60	VAKSELDKAIGRNCNGVITK	AMP	0.858
KILLIAN-OQ446694	endolysin	WEU68017.1	135-154	IAVLVKKFSLEGHKILIISS	AMP	0.977
VB-ECO-KMB14-OR424384	endolysin	WPK27911.1	50-69	AKHNANVGGAKNSMHLTGKA	AMP	0.892
VB-ECO-KMB43-OR525698	endolysin	WQN06429.1	53-72	GKSKTMKSRHLHGFAVDVVA	AMP	0.582
VB-ECOM-SHAK7163-OR594179	internal-virion-protein-with-endolysin-domain	WVW77660.1	38-57	LQEQVLEAKADADILGAVGA	AMP	0.989
VB-ECOM-SHAK7163-OR594179	endolysin	WVW77668.1	50-69	AKHNANVGGVNNSMHLTGKA	AMP	0.846
VB-ECOM-SHAK7693-OR594180	endolysin	WVW77753.1	28-47	AYRATPNEKYLTIGYGHYGP	NAMP	0.455
VB-ECOM-SHAK7693-OR594180	endolysin	WVW77754.1	56-75	AKLALFINQNGKPLLGLRRR	AMP	0.980
VB-ECOM-SHAK7858-OR594183	endolysin	WVW77834.1	117-136	AKLALFINQNGKPLLGLRRR	AMP	0.980
VB-ECOS-UTEC10-PP291582	endolysin	WWY66481.1	118-137	AKLALFINQNGKPLLGLRRR	AMP	0.980
VB-ECOP-ACG-C91-NC-019403	endolysin	YP-006987811.1	50-69	AKHNANVGGAKNSMHLTGKA	AMP	0.892
VB-ECOS-ACG-M12-NC-019404	endolysin	YP-006987885.1	125-144	ALGMWNKARVNGKLKVVKGL	AMP	0.975
VB-ECOS-GOLESTAN-NC-042084	endolysin	YP-009620148.1	109-128	AKLALFINQNGKPILGLRRR	AMP	0.983
EG1-NC-047895	internal-virion-protein-with-endolysin-domain	YP-009795842.1	354-373	GFKLINKALVVGAQSAAINA	AMP	0.993
NTEC3-NC-073318	endolysin	YP-010749155.1	123-142	FINQNGRPLLGLRRRTAGRL	AMP	0.950
VB-ECOS-PHIEC3-NC-073323	endolysin	YP-010749568.1	118-137	AKLALFINQNGKPILGLRRR	AMP	0.983
MLP1-NC-079176	endolysin	YP-010844467.1	102-121	IRWGGTWERSFTKPVKWAKF	AMP	0.828

Supplementary Table S3 giving an overview of the selected features of the proteins identified by phylogenetic analysis as closely related to the identified endolysins.

Phage name [Reference]	Protein name	Position of the gene in virus genome	Gene length (nt)	Protein accession number	Protein length (aa)	Molecular weight (kDa)	Instability index	Isoelectric point (pI)	Gravy	Aliphatic index	Melting point (Tm)	Name of the predicted domain	Position of the domain in protein sequence
vB_EcoS-22664B1 [36]	lysin	2556 - 3047	491	QZI78369.1	163	18,18	31,013	9,22	-0,456	77,914	46.0°C	lyz_endolysin_autolysin (cd00737) Phage lysozyme (PF00959) Lysozyme-like (SSF53955)	7-148 14-139 1-153
vB_EcoS-22664BS2 [36]	lysin	817 - 1305	488	QZI78493.1	162	17,08	19,002	9,81	-0,117	87,531	50.0°C	lyz_endolysin_autolysin (cd00737) Phage lysozyme (PF00959) Lysozyme-like (SSF53955)	9-146 15-138 3-149
vB_EcoS-UTI89UKE1[36]	lysin	361 - 849	488	USM81274.1	162	17,08	18,337	9,81	-0,121	87,531	50.0°C	lyz_endolysin_autolysin (cd00737) Phage lysozyme (PF00959) Lysozyme-like (SSF53955)	9-146 15-138 3-149
vB_EcoS-101114B2 [36]	lysin	2526 - 3017	491	QZI79000.1	163	18,18	31,013	9,22	-0,456	77,914	46.0°C	lyz_endolysin_autolysin (cd00737) Phage lysozyme (PF00959) Lysozyme-like (SSF53955)	7-148 14-139 1-153
vB_EcoS-101114BS4 [36]	lysin	1030 - 1521	491	QZI79065.1	163	18,18	31,013	9,22	-0,456	77,914	46.0°C	lyz_endolysin_autolysin (cd00737) Phage lysozyme (PF00959) Lysozyme-like (SSF53955)	7-148 14-139 1-153
vB-Eco-KMB35 [37]	lysin	40675 - 41163	488	WQN07684.1	162	17,05	18,337	9,81	-0,075	89,938	51.0°C	lyz_endolysin_autolysin (cd00737) Phage lysozyme (PF00959) Lysozyme-like (SSF53955)	9-146 15-138 3-149
vB-Eco-KMB36 [37]	lysin	41598 - 42101	503	WQN07602.1	167	17,77	16,158	9,59	-0,151	85,509	51.0°C	lyz_endolysin_autolysin (cd00737) Phage lysozyme (PF00959) Lysozyme-like (SSF53955)	14-151 20-143 7-154
vB_EcoS-CHD5UKE2 [36]	lysin	14512 - 15003	491	USL86352.1	163	18,3	31,731	9,43	-0,478	77,914	47.0°C	lyz_endolysin_autolysin (cd00737) Phage lysozyme (PF00959) Lysozyme-like (SSF53955)	7-148 14-139 1-153
vB_EcoS-CHD2B1 [36]	lysin	42121 - 42612	491	USL86318.1	163	18,18	31,013	9,22	-0,456	77,914	46.0°C	lyz_endolysin_autolysin (cd00737) Phage lysozyme (PF00959) Lysozyme-like (SSF53955)	7-148 14-139 1-153
vB-Eco-KMB46 [37]	lysozyme R	44328 - 44819	491	WQN06679.1	163	18,33	33,344	9,55	-0,351	83,19	49.0°C	endolysin_R21-like (cd16900) SAR-endolysin. (MF_04136) Phage lysozyme (PF00959) Lysozyme-like (SSF53955) Endolysin. (MF_04110)	19-158 4-161 25-154 15-161 5-161
vB-Eco-KMB25 [37]	lysozyme or muramidase	33375 - 33869	494	WPK28014.1	164	18,3	24,023	9,65	-0,318	82,744	50.0°C	endolysin_R21-like (cd16900) SAR-endolysin. (MF_04136) Phage lysozyme (PF00959) Lysozyme-like (SSF53955) Endolysin. (MF_04110)	13-158 4-161 25-154 12-162 5-161

vB-Eco-KMB39 [37]	lysozyme R	61730 - 62224	494	WPJ71303.1	164	18,61	37,155	9,59	-0,368	88,659	52.0°C	T4-like_lys (cd00735) Phage lysozyme (PF00959) Lysozyme-like (SSF53955) Endolysin. (MF_04110) Endolysin (G3DSA:1.10.530.40:FF:000002)	2-160 11-141 1-161 1-159 1-164
vB_EcoM_KMB26 [41]	T4-like phage baseplate hub + tail lysozyme	64196 - 64690	494	CAH2619090.1	164	18,61	37,155	9,59	-0,368	88,659	52.0°C	T4-like_lys (cd00735) Phage lysozyme (PF00959) Lysozyme-like (SSF53955) Endolysin. (MF_04110) Endolysin (G3DSA:1.10.530.40:FF:000002)	2-160 11-141 1-161 1-159 1-164
vB-Eco-KMB42 [37]	lysozyme R	62848 - 63342	494	WPJ71835.1	164	18,61	37,155	9,59	-0,368	88,659	52.0°C	T4-like_lys (cd00735) Phage lysozyme (PF00959) Lysozyme-like (SSF53955) Endolysin. (MF_04110) Endolysin (G3DSA:1.10.530.40:FF:000002)	2-160 11-141 1-161 1-159 1-164
vB-Eco-KMB22 [37]	lysozyme R	64127 - 64621	494	WPJ70775.1	164	18,63	36,638	9,59	-0,365	89,878	52.0°C	T4-like_lys (cd00735) Phage lysozyme (PF00959) Lysozyme-like (SSF53955) Endolysin. (MF_04110) Endolysin (G3DSA:1.10.530.40:FF:000002)	2-160 11-141 1-161 1-159 1-164
vB-Eco-KMB23 [37]	lysozyme R	63387 - 63881	494	WQN07383.1	164	18,61	37,155	9,59	-0,368	88,659	52.0°C	T4-like_lys (cd00735) Phage lysozyme (PF00959) Lysozyme-like (SSF53955) Endolysin. (MF_04110) Endolysin (G3DSA:1.10.530.40:FF:000002)	2-160 11-141 1-161 1-159 1-164
vB_EcoM_ACG-C40 [34]	glycoside hydrolase family protein	63339 - 63833	494	YP_006986678.1	164	18,61	37,155	9,59	-0,368	88,659	52.0°C	T4-like_lys (cd00735) Phage lysozyme (PF00959) Lysozyme-like (SSF53955) Endolysin. (MF_04110) Endolysin (G3DSA:1.10.530.40:FF:000002)	2-160 11-141 1-161 1-159 1-164
vB-Eco-KMB26 [37]	lysozyme R	64196 - 64690	494	WPJ71039.1	164	18,61	37,155	9,59	-0,368	88,659	52.0°C	T4-like_lys (cd00735) Phage lysozyme (PF00959) Lysozyme-like (SSF53955) Endolysin. (MF_04110) Endolysin (G3DSA:1.10.530.40:FF:000002)	2-160 11-141 1-161 1-159 1-164
vB-Eco-KMB40 [37]	lysozyme R	62190 - 62684	494	WPJ71570.1	164	18,61	37,155	9,59	-0,368	88,659	52.0°C	T4-like_lys (cd00735) Phage lysozyme (PF00959) Lysozyme-like (SSF53955) Endolysin. (MF_04110) Endolysin (G3DSA:1.10.530.40:FF:000002)	2-160 11-141 1-161 1-159 1-164

vB_EcoM-101112UKE3-1 [36]	lysozyme R	158103 - 158597	494	QZI78958.1	164	18,61	37,155	9,59	-0,368	88,659	52.0°C	T4-like_lys (cd00735) Phage lysozyme (PF00959) Lysozyme-like (SSF53955) Endolysin. (MF_04110) Endolysin (G3DSA:1.10.530.40:FF:000002)	2-160 11-141 1-161 1-159 1-164
vB_EcoM-CHD2BS1 [36]	lysozyme R	157128 - 157622	494	QZI80464.1	164	18,61	37,155	9,59	-0,368	88,659	52.0°C	T4-like_lys (cd00735) Phage lysozyme (PF00959) Lysozyme-like (SSF53955) Endolysin. (MF_04110) Endolysin (G3DSA:1.10.530.40:FF:000002)	2-160 11-141 1-161 1-159 1-164
MLP2 [38]	baseplate hub + tail lysozyme	156422 - 156916	494	UEN68765.1	164	18,61	37,673	9,39	-0,366	88,659	51.0°C	T4-like_lys (cd00735) Phage lysozyme (PF00959) Lysozyme-like (SSF53955) Endolysin. (MF_04110) Endolysin (G3DSA:1.10.530.40:FF:000002)	2-160 11-141 1-161 1-159 1-164
vB_EcoM-G3G7 [36]	lysozyme R	157200 - 157694	494	QZI83386.1	164	18,61	37,155	9,59	-0,368	88,659	52.0°C	T4-like_lys (cd00735) Phage lysozyme (PF00959) Lysozyme-like (SSF53955) Endolysin. (MF_04110) Endolysin (G3DSA:1.10.530.40:FF:000002)	2-160 11-141 1-161 1-159 1-164
vB_EcoM-101117BS1 [36]	lysozyme R	154626 - 155120	494	QZI79532.1	164	18,61	37,155	9,59	-0,368	88,659	52.0°C	T4-like_lys (cd00735) Phage lysozyme (PF00959) Lysozyme-like (SSF53955) Endolysin. (MF_04110) Endolysin (G3DSA:1.10.530.40:FF:000002)	2-160 11-141 1-161 1-159 1-164
vB_EcoM-CHD16UKE1 [36]	lysozyme R	157094 - 157588	494	QZI80907.1	164	18,61	37,155	9,59	-0,368	88,659	52.0°C	T4-like_lys (cd00735) Phage lysozyme (PF00959) Lysozyme-like (SSF53955) Endolysin. (MF_04110) Endolysin (G3DSA:1.10.530.40:FF:000002)	2-160 11-141 1-161 1-159 1-164
ZCEC14 [40]	hypothetical protein	96330 - 96818	488	UVD33441.1	162	18,39	36,563	9,66	-0,354	89,136	52.0°C	T4-like_lys (cd00735) Phage lysozyme (PF00959) Lysozyme-like (SSF53955) Endolysin. (MF_04110)	2-160 11-141 1-161 1-159
vB_EcoM-CHD94UKE2 [36]	lysozyme R	154846 - 155334	488	QZI81192.1	162	18,38	35,161	9,66	-0,338	89,136	52.0°C	T4-like_lys (cd00735) Phage lysozyme (PF00959) Lysozyme-like (SSF53955) Endolysin. (MF_04110)	2-160 11-140 1-161 1-159
vB-Eco-KMB38 [37]	lysozyme R	66043 - 66531	488	WQN07107.1	162	18,31	29,657	9,66	-0,269	88,519	51.0°C	T4-like_lys (cd00735) Phage lysozyme (PF00959) Lysozyme-like (SSF53955)	2-160 10-140 1-161

vB-Eco-KMB37 [37]	lysozyme murein hydrolase	22896 - 23396	500	WQN06754.1	166	18,86	21,257	9,67	-0,281	95,12	44.0°C	T4-like_lys (cd00735) Phage lysozyme (PF00959) Lysozyme-like (SSF53955)	7-164 15-142 7-166
vB_EcoM-172859UKE1 [36]	lysozyme R	157547 - 158035	488	QZI80176.1	162	18,32	29,133	9,66	-0,263	88,519	52.0°C	T4-like_lys (cd00735) Phage lysozyme (PF00959) Lysozyme-like (SSF53955)	2-160 10-140 1-161
vB_EcoM_LNA2 [27]	hypothetical protein	12924 - 13412	488	self_NLWW35	162	18,29	26,867	9,73	-0,28	87,346	52.0°C	T4-like_lys (cd00735) Phage lysozyme (PF00959) Lysozyme-like (SSF53955)	2-160 10-140 1-161
EG1 [39]	amidase	11177 - 11632	455	YP_009795817.1	151	17,01	36,916	8,83	-0,501	75,497	49.0°C	PGRP (cd06583) N-acetylmuramoyl-L-alanine amidase-like (SSF55846) N-acetylmuramoyl-L-alanine amidase (PF01510) Endolysin. (MF_04111) Endolysin (G3DSA:3.40.80.10:FF:000009)	13-132 10-146 11-132 1-149 6-151
vB_EcoM_SHAK7854 [33]	amidase	13662 - 14120	458	self_DNYN29	152	17,06	44,325	8,83	-0,391	78,158	51.0°C	PGRP (cd06583) N-acetylmuramoyl-L-alanine amidase-like (SSF55846) N-acetylmuramoyl-L-alanine amidase (PF01510) Endolysin. (MF_04111)	14-133 11-147 11-134 2-150
vB_EcoP_LNA3 [27]	amidase	26726 - 27184	458	self_GXIJ34	152	17,02	44,826	8,47	-0,418	75,592	51.0°C	PGRP (cd06583) N-acetylmuramoyl-L-alanine amidase-like (SSF55846) N-acetylmuramoyl-L-alanine amidase (PF01510) Endolysin. (MF_04111)	14-133 11-147 11-134 2-150
vB_EcoP_LNA5 [27]	amidase	6617 - 7075	458	self_TCEH12	152	17,08	44,884	8,47	-0,442	76,25	51.0°C	PGRP (cd06583) N-acetylmuramoyl-L-alanine amidase-like (SSF55846) N-acetylmuramoyl-L-alanine amidase (PF01510) Endolysin. (MF_04111)	14-133 11-147 11-134 2-150
vB_EcoM_SHAK9454 [33]	amidase	13615 - 14073	458	self_ZENB29	152	17,08	45,883	8,83	-0,404	75,592	51.0°C	PGRP (cd06583) N-acetylmuramoyl-L-alanine amidase-like (SSF55846) N-acetylmuramoyl-L-alanine amidase (PF01510) Endolysin. (MF_04111)	14-133 11-147 11-134 2-150
vB_EcoP-101101UKE1 [36]	hypothetical protein	36863 - 37207	344	QZI78680.1	114	12,8	39,555	9,45	-0,439	70,088	50.0°C	Peptidase M15 (PF08291) Hedgehog/DD-peptidase (SSF55166)	4-106 5-110
vB_EcoP-101117UKE2 [36]	hypothetical protein	39216 - 39587	371	QZI79610.1	123	13,64	26,292	9,02	-0,349	72,846	52.0°C	Peptidase M15 (PF08291) Hedgehog/DD-peptidase (SSF55166)	14-115 14-116
vB_EcoP-101118B1 [36]	hypothetical protein	39216 - 39587	371	QZI79670.1	123	13,64	26,292	9,02	-0,349	72,846	52.0°C	Peptidase M15 (PF08291) Hedgehog/DD-peptidase (SSF55166)	14-115 14-116
vB_EcoP-UTI89UKE2 [36]	hypothetical protein	36944 - 37288	344	QZI84581.1	114	12,65	29,187	9,42	-0,404	70,088	50.0°C	Peptidase M15 (PF08291) Hedgehog/DD-peptidase (SSF55166)	4-106 5-107
vB_EcoP-UTI89UKE3 [36]	hypothetical protein	36945 - 37289	344	QZI84648.1	114	12,65	29,187	9,42	-0,404	70,088	50.0°C	Peptidase M15 (PF08291) Hedgehog/DD-peptidase (SSF55166)	4-106 5-107
vB-Eco-KMB47 [37]	hypothetical protein	38771 - 39115	344	WPK27967.1	114	12,8	39,555	9,45	-0,439	70,088	50.0°C	Peptidase M15 (PF08291) Hedgehog/DD-peptidase (SSF55166)	4-106 5-110

Supplementary Table S4. Data obtained by InterProScan version 5.72-103 software, giving an overview of the domains identified within the proteins found by phylogenetic analysis as closely related to the tested endolysins. The domains marked in yellow were selected and listed in Supplementary Table S3.

Phage name	Protein name	Protein accession number	Position of the gene in virus genome	Sequence length	Analysis	Signature accession	Signature description	Start location	Stop location	Score	InterPro annotations - accession	InterPro annotations - description	GO annotations
vB EcoM KMB26 OW536688	T4-like phage baseplate hub + tail lysozyme	CAH2619090.1	64196 - 64690	164	Gene3D	G3DSA:1.10.530.40	-	1	164	6.6E-75	IPR023347	Lysozyme domain superfamily	GO:0003796(InterPro)
vB EcoM KMB26 OW536688	T4-like phage baseplate hub + tail lysozyme	CAH2619090.1	64196 - 64690	164	FunFam	G3DSA:1.10.530.40:FF:000002	Endolysin	1	164	1.1E-114	-	-	-
vB EcoM KMB26 OW536688	T4-like phage baseplate hub + tail lysozyme	CAH2619090.1	64196 - 64690	164	Hamap	MF 04110	Endolysin.	1	159	16.677807	IPR034690	Endolysin T4 type	GO:0003796(InterPro)|GO:0009253(InterPro)
vB EcoM KMB26 OW536688	T4-like phage baseplate hub + tail lysozyme	CAH2619090.1	64196 - 64690	164	SUPERFAMILY	SSF53955	Lysozyme-like	1	161	4.81E-54	IPR023346	Lysozyme-like domain superfamily	-
vB EcoM KMB26 OW536688	T4-like phage baseplate hub + tail lysozyme	CAH2619090.1	64196 - 64690	164	Pfam	PF00959	Phage lysozyme	11	141	4.4E-10	IPR002196	Glycoside hydrolase, family 24	GO:0003796(InterPro)|GO:0009253(InterPro)|GO:0016998(InterPro)
vB EcoM KMB26 OW536688	T4-like phage baseplate hub + tail lysozyme	CAH2619090.1	64196 - 64690	164	PRINTS	PR00684	Phage T4 lysozyme signature	95	114	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM KMB26 OW536688	T4-like phage baseplate hub + tail lysozyme	CAH2619090.1	64196 - 64690	164	PRINTS	PR00684	Phage T4 lysozyme signature	72	92	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM KMB26 OW536688	T4-like phage baseplate hub + tail lysozyme	CAH2619090.1	64196 - 64690	164	PRINTS	PR00684	Phage T4 lysozyme signature	4	23	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM KMB26 OW536688	T4-like phage baseplate hub + tail lysozyme	CAH2619090.1	64196 - 64690	164	PRINTS	PR00684	Phage T4 lysozyme signature	24	42	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM KMB26 OW536688	T4-like phage baseplate hub + tail lysozyme	CAH2619090.1	64196 - 64690	164	PRINTS	PR00684	Phage T4 lysozyme signature	118	137	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM KMB26 OW536688	T4-like phage baseplate hub + tail lysozyme	CAH2619090.1	64196 - 64690	164	PRINTS	PR00684	Phage T4 lysozyme signature	51	70	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM KMB26 OW536688	T4-like phage baseplate hub + tail lysozyme	CAH2619090.1	64196 - 64690	164	PRINTS	PR00684	Phage T4 lysozyme signature	138	159	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM KMB26 OW536688	T4-like phage baseplate hub + tail lysozyme	CAH2619090.1	64196 - 64690	164	CDD	cd00735	T4-like lys	2	160	6.07455E-78	-	-	-

vB EcoM KMB26 OW536688	T4-like phage baseplate hub + tail lysozyme	CAH2619090.1	64196 - 64690	164	PANTHER	PTHR37406	T4-TYPE LYSOZYME 1-RELATED	4	161	1.5E-27	IPR052619	Bacteriophage lysozyme-like	-
vB EcoS-22664B1 MZ234008	lysin	QZI78369.1	2556 - 3047	163	PANTHER	PTHR38107	-	2	147	3.6E-18	IPR051018	Bacteriophage Glycosyl Hydrolase 24	GO:0003796(PANTHER)
vB EcoS-22664B1 MZ234008	lysin	QZI78369.1	2556 - 3047	163	Gene3D	G3DSA:1.10.530.40	-	2	154	9.0E-31	IPR023347	Lysozyme domain superfamily	GO:0003796(InterPro)
vB EcoS-22664B1 MZ234008	lysin	QZI78369.1	2556 - 3047	163	SUPERFAMILY	SSF53955	Lysozyme-like	1	153	3.79E-32	IPR023346	Lysozyme-like domain superfamily	-
vB EcoS-22664B1 MZ234008	lysin	QZI78369.1	2556 - 3047	163	CDD	cd00737	lyz endolysin autolysin	7	148	1.42938E-36	IPR033907	Endolysin/autolysin	GO:0003796(InterPro)
vB EcoS-22664B1 MZ234008	lysin	QZI78369.1	2556 - 3047	163	Pfam	PF00959	Phage lysozyme	14	139	4.6E-18	IPR002196	Glycoside hydrolase, family 24	GO:0003796(InterPro)|GO:0009253(InterPro)|GO:0016998(InterPro)
vB EcoS-22664BS2 MZ234010	lysin	QZI78493.1	817 - 1305	162	Gene3D	G3DSA:1.10.530.40	-	2	147	1.1E-31	IPR023347	Lysozyme domain superfamily	GO:0003796(InterPro)
vB EcoS-22664BS2 MZ234010	lysin	QZI78493.1	817 - 1305	162	PANTHER	PTHR38107	-	3	144	3.2E-22	IPR051018	Bacteriophage Glycosyl Hydrolase 24	GO:0003796(PANTHER)
vB EcoS-22664BS2 MZ234010	lysin	QZI78493.1	817 - 1305	162	SUPERFAMILY	SSF53955	Lysozyme-like	3	149	1.53E-31	IPR023346	Lysozyme-like domain superfamily	-
vB EcoS-22664BS2 MZ234010	lysin	QZI78493.1	817 - 1305	162	CDD	cd00737	lyz endolysin autolysin	9	146	2.67399E-30	IPR033907	Endolysin/autolysin	GO:0003796(InterPro)
vB EcoS-22664BS2 MZ234010	lysin	QZI78493.1	817 - 1305	162	Pfam	PF00959	Phage lysozyme	15	138	6.2E-23	IPR002196	Glycoside hydrolase, family 24	GO:0003796(InterPro)|GO:0009253(InterPro)|GO:0016998(InterPro)
vB EcoP-101101UKE1 MZ234012	hypothetical protein	QZI78680.1	36863 - 37207	114	Gene3D	G3DSA:3.30.1380.10	-	5	113	3.5E-26	IPR009045	Hedgehog signalling/DD-peptidase zinc-binding domain superfamily	-
vB EcoP-101101UKE1 MZ234012	hypothetical protein	QZI78680.1	36863 - 37207	114	SUPERFAMILY	SSF55166	Hedgehog/DD-peptidase	5	110	6.28E-25	IPR009045	Hedgehog signalling/DD-peptidase zinc-binding domain superfamily	-
vB EcoP-101101UKE1 MZ234012	hypothetical protein	QZI78680.1	36863 - 37207	114	Pfam	PF08291	Peptidase M15	4	106	1.4E-29	IPR013230	Peptidase M15A, C-terminal	-

vB EcoM-101112UKE3-1 MZ234013	lysozyme R	QZI78958.1	158103 - 158597	164	Gene3D	G3DSA:1.10.530.40	-	1	164	6.6E-75	IPR023347	Lysozyme domain superfamily	GO:0003796(InterPro)
vB EcoM-101112UKE3-1 MZ234013	lysozyme R	QZI78958.1	158103 - 158597	164	FunFam	G3DSA:1.10.530.40:FF:000002	Endolysin	1	164	1.1E-114	-	-	-
vB EcoM-101112UKE3-1 MZ234013	lysozyme R	QZI78958.1	158103 - 158597	164	Hamap	MF 04110	Endolysin.	1	159	16.677807	IPR034690	Endolysin T4 type	GO:0003796(InterPro)|GO:0009253(InterPro)
vB EcoM-101112UKE3-1 MZ234013	lysozyme R	QZI78958.1	158103 - 158597	164	SUPERFAMILY	SSF53955	Lysozyme-like	1	161	4.81E-54	IPR023346	Lysozyme-like domain superfamily	-
vB EcoM-101112UKE3-1 MZ234013	lysozyme R	QZI78958.1	158103 - 158597	164	Pfam	PF00959	Phage lysozyme	11	141	4.4E-10	IPR002196	Glycoside hydrolase, family 24	GO:0003796(InterPro)|GO:0009253(InterPro)|GO:0016998(InterPro)
vB EcoM-101112UKE3-1 MZ234013	lysozyme R	QZI78958.1	158103 - 158597	164	PRINTS	PR00684	Phage T4 lysozyme signature	95	114	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM-101112UKE3-1 MZ234013	lysozyme R	QZI78958.1	158103 - 158597	164	PRINTS	PR00684	Phage T4 lysozyme signature	72	92	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM-101112UKE3-1 MZ234013	lysozyme R	QZI78958.1	158103 - 158597	164	PRINTS	PR00684	Phage T4 lysozyme signature	4	23	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM-101112UKE3-1 MZ234013	lysozyme R	QZI78958.1	158103 - 158597	164	PRINTS	PR00684	Phage T4 lysozyme signature	24	42	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM-101112UKE3-1 MZ234013	lysozyme R	QZI78958.1	158103 - 158597	164	PRINTS	PR00684	Phage T4 lysozyme signature	118	137	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM-101112UKE3-1 MZ234013	lysozyme R	QZI78958.1	158103 - 158597	164	PRINTS	PR00684	Phage T4 lysozyme signature	51	70	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM-101112UKE3-1 MZ234013	lysozyme R	QZI78958.1	158103 - 158597	164	PRINTS	PR00684	Phage T4 lysozyme signature	138	159	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM-101112UKE3-1 MZ234013	lysozyme R	QZI78958.1	158103 - 158597	164	CDD	cd00735	T4-like lys	2	160	6.07455E-78	-	-	-
vB EcoM-101112UKE3-1 MZ234013	lysozyme R	QZI78958.1	158103 - 158597	164	PANTHER	PTHR37406	T4-TYPE LYSOZYME 1-RELATED	4	161	1.5E-27	IPR052619	Bacteriophage lysozyme-like	-
vB EcoS-101114B2 MZ234014	lysin	QZI79000.1	2526 - 3017	163	PANTHER	PTHR38107	-	2	147	3.6E-18	IPR051018	Bacteriophage Glycosyl Hydrolase 24	GO:0003796(PANTHER)

vB EcoS-101114B2 MZ234014	lysin	QZI79000.1	2526 - 3017	163	Gene3D	G3DSA:1.10.530.40	-	2	154	9.0E-31	IPR023347	Lysozyme domain superfamily	GO:0003796(InterPro)
vB EcoS-101114B2 MZ234014	lysin	QZI79000.1	2526 - 3017	163	SUPERFAMILY	SSF53955	Lysozyme-like	1	153	3.79E-32	IPR023346	Lysozyme-like domain superfamily	-
vB EcoS-101114B2 MZ234014	lysin	QZI79000.1	2526 - 3017	163	CDD	cd00737	lyz endolysin autolysin	7	148	1.42938E-36	IPR033907	Endolysin/autolysin	GO:0003796(InterPro)
vB EcoS-101114B2 MZ234014	lysin	QZI79000.1	2526 - 3017	163	Pfam	PF00959	Phage lysozyme	14	139	4.6E-18	IPR002196	Glycoside hydrolase, family 24	GO:0003796(InterPro)|GO:0009253(InterPro)|GO:0016998(InterPro)
vB EcoS-101114BS4 MZ234016	lysin	QZI79065.1	1030 - 1521	163	PANTHER	PTHR38107	-	2	147	3.6E-18	IPR051018	Bacteriophage Glycosyl Hydrolase 24	GO:0003796(PANTHER)
vB EcoS-101114BS4 MZ234016	lysin	QZI79065.1	1030 - 1521	163	Gene3D	G3DSA:1.10.530.40	-	2	154	9.0E-31	IPR023347	Lysozyme domain superfamily	GO:0003796(InterPro)
vB EcoS-101114BS4 MZ234016	lysin	QZI79065.1	1030 - 1521	163	SUPERFAMILY	SSF53955	Lysozyme-like	1	153	3.79E-32	IPR023346	Lysozyme-like domain superfamily	-
vB EcoS-101114BS4 MZ234016	lysin	QZI79065.1	1030 - 1521	163	CDD	cd00737	lyz endolysin autolysin	7	148	1.42938E-36	IPR033907	Endolysin/autolysin	GO:0003796(InterPro)
vB EcoS-101114BS4 MZ234016	lysin	QZI79065.1	1030 - 1521	163	Pfam	PF00959	Phage lysozyme	14	139	4.6E-18	IPR002196	Glycoside hydrolase, family 24	GO:0003796(InterPro)|GO:0009253(InterPro)|GO:0016998(InterPro)
vB EcoM-101117BS1 MZ234018	lysozyme R	QZI79532.1	154626 - 155120	164	Gene3D	G3DSA:1.10.530.40	-	1	164	6.6E-75	IPR023347	Lysozyme domain superfamily	GO:0003796(InterPro)
vB EcoM-101117BS1 MZ234018	lysozyme R	QZI79532.1	154626 - 155120	164	FunFam	G3DSA:1.10.530.40:FF:000002	Endolysin	1	164	1.1E-114	-	-	-
vB EcoM-101117BS1 MZ234018	lysozyme R	QZI79532.1	154626 - 155120	164	Hamap	MF 04110	Endolysin.	1	159	16.677807	IPR034690	Endolysin T4 type	GO:0003796(InterPro)|GO:0009253(InterPro)
vB EcoM-101117BS1 MZ234018	lysozyme R	QZI79532.1	154626 - 155120	164	SUPERFAMILY	SSF53955	Lysozyme-like	1	161	4.81E-54	IPR023346	Lysozyme-like domain superfamily	-
vB EcoM-101117BS1 MZ234018	lysozyme R	QZI79532.1	154626 - 155120	164	Pfam	PF00959	Phage lysozyme	11	141	4.4E-10	IPR002196	Glycoside hydrolase, family 24	GO:0003796(InterPro)|GO:0009253(InterPro)|GO:0016998(InterPro)
vB EcoM-101117BS1 MZ234018	lysozyme R	QZI79532.1	154626 - 155120	164	PRINTS	PR00684	Phage T4 lysozyme signature	95	114	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM-101117BS1 MZ234018	lysozyme R	QZI79532.1	154626 - 155120	164	PRINTS	PR00684	Phage T4 lysozyme signature	72	92	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM-101117BS1 MZ234018	lysozyme R	QZI79532.1	154626 - 155120	164	PRINTS	PR00684	Phage T4 lysozyme signature	4	23	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM-101117BS1 MZ234018	lysozyme R	QZI79532.1	154626 - 155120	164	PRINTS	PR00684	Phage T4 lysozyme signature	24	42	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)

vB EcoM-101117BS1 MZ234018	lysozyme R	QZI79532.1	154626 - 155120	164	PRINTS	PR00684	Phage T4 lysozyme signature	118	137	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM-101117BS1 MZ234018	lysozyme R	QZI79532.1	154626 - 155120	164	PRINTS	PR00684	Phage T4 lysozyme signature	51	70	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM-101117BS1 MZ234018	lysozyme R	QZI79532.1	154626 - 155120	164	PRINTS	PR00684	Phage T4 lysozyme signature	138	159	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM-101117BS1 MZ234018	lysozyme R	QZI79532.1	154626 - 155120	164	CDD	cd00735	T4-like lys	2	160	6.07455E-78	-	-	-
vB EcoM-101117BS1 MZ234018	lysozyme R	QZI79532.1	154626 - 155120	164	PANTHER	PTHR37406	T4-TYPE LYSOZYME 1-RELATED	4	161	1.5E-27	IPR052619	Bacteriophage lysozyme-like	-
vB EcoP-101117UKE2 MZ234019	hypothetical protein	QZI79610.1	39216 - 39587	123	Gene3D	G3DSA:3.30.1380.10	-	14	123	1.0E-25	IPR009045	Hedgehog signalling/DD-peptidase zinc-binding domain superfamily	-
vB EcoP-101117UKE2 MZ234019	hypothetical protein	QZI79610.1	39216 - 39587	123	SUPERFAMILY	SSF55166	Hedgehog/DD-peptidase	14	116	2.86E-24	IPR009045	Hedgehog signalling/DD-peptidase zinc-binding domain superfamily	-
vB EcoP-101117UKE2 MZ234019	hypothetical protein	QZI79610.1	39216 - 39587	123	Pfam	PF08291	Peptidase M15	14	115	3.2E-29	IPR013230	Peptidase M15A, C-terminal	-
vB EcoP-101118B1 MZ234020	hypothetical protein	QZI79670.1	39216 - 39587	123	Gene3D	G3DSA:3.30.1380.10	-	14	123	1.0E-25	IPR009045	Hedgehog signalling/DD-peptidase zinc-binding domain superfamily	-
vB EcoP-101118B1 MZ234020	hypothetical protein	QZI79670.1	39216 - 39587	123	SUPERFAMILY	SSF55166	Hedgehog/DD-peptidase	14	116	2.86E-24	IPR009045	Hedgehog signalling/DD-peptidase zinc-binding domain superfamily	-
vB EcoP-101118B1 MZ234020	hypothetical protein	QZI79670.1	39216 - 39587	123	Pfam	PF08291	Peptidase M15	14	115	3.2E-29	IPR013230	Peptidase M15A, C-terminal	-
vB EcoM-172859UKE1 MZ234025	lysozyme R	QZI80176.1	157547 - 158035	162	Gene3D	G3DSA:1.10.530.40	-	2	162	1.0E-89	IPR023347	Lysozyme domain superfamily	GO:0003796(InterPro)

vB EcoM-172859UKE1 MZ234025	lysozyme R	QZI80176.1	157547 - 158035	162	SUPERFAMILY	SSF53955	Lysozyme-like	1	161	6.14E-53	IPR023346	Lysozyme-like domain superfamily	-
vB EcoM-172859UKE1 MZ234025	lysozyme R	QZI80176.1	157547 - 158035	162	Pfam	PF00959	Phage lysozyme	10	140	2.6E-10	IPR002196	Glycoside hydrolase, family 24	GO:0003796(InterPro)|GO:0009253(InterPro)|GO:0016998(InterPro)
vB EcoM-172859UKE1 MZ234025	lysozyme R	QZI80176.1	157547 - 158035	162	PRINTS	PR00684	Phage T4 lysozyme signature	4	23	3.3E-76	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM-172859UKE1 MZ234025	lysozyme R	QZI80176.1	157547 - 158035	162	PRINTS	PR00684	Phage T4 lysozyme signature	118	137	3.3E-76	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM-172859UKE1 MZ234025	lysozyme R	QZI80176.1	157547 - 158035	162	PRINTS	PR00684	Phage T4 lysozyme signature	138	159	3.3E-76	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM-172859UKE1 MZ234025	lysozyme R	QZI80176.1	157547 - 158035	162	PRINTS	PR00684	Phage T4 lysozyme signature	24	42	3.3E-76	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM-172859UKE1 MZ234025	lysozyme R	QZI80176.1	157547 - 158035	162	PRINTS	PR00684	Phage T4 lysozyme signature	72	92	3.3E-76	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM-172859UKE1 MZ234025	lysozyme R	QZI80176.1	157547 - 158035	162	PRINTS	PR00684	Phage T4 lysozyme signature	95	114	3.3E-76	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM-172859UKE1 MZ234025	lysozyme R	QZI80176.1	157547 - 158035	162	PRINTS	PR00684	Phage T4 lysozyme signature	51	70	3.3E-76	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM-172859UKE1 MZ234025	lysozyme R	QZI80176.1	157547 - 158035	162	CDD	cd00735	T4-like lys	2	160	1.16396E-73	-	-	-
vB EcoM-172859UKE1 MZ234025	lysozyme R	QZI80176.1	157547 - 158035	162	PANTHER	PTHR37406	T4-TYPE LYSOZYME 1-RELATED	4	160	5.7E-26	IPR052619	Bacteriophage lysozyme-like	-
vB EcoM-CHD2BS1 MZ234027	lysozyme R	QZI80464.1	157128 - 157622	164	Gene3D	G3DSA:1.10.530.40	-	1	164	6.6E-75	IPR023347	Lysozyme domain superfamily	GO:0003796(InterPro)
vB EcoM-CHD2BS1 MZ234027	lysozyme R	QZI80464.1	157128 - 157622	164	FunFam	G3DSA:1.10.530.40:FF:000002	Endolysin	1	164	1.1E-114	-	-	-
vB EcoM-CHD2BS1 MZ234027	lysozyme R	QZI80464.1	157128 - 157622	164	Hamap	MF 04110	Endolysin.	1	159	16.677807	IPR034690	Endolysin T4 type	GO:0003796(InterPro)|GO:0009253(InterPro)
vB EcoM-CHD2BS1 MZ234027	lysozyme R	QZI80464.1	157128 - 157622	164	SUPERFAMILY	SSF53955	Lysozyme-like	1	161	4.81E-54	IPR023346	Lysozyme-like domain superfamily	-
vB EcoM-CHD2BS1 MZ234027	lysozyme R	QZI80464.1	157128 - 157622	164	Pfam	PF00959	Phage lysozyme	11	141	4.4E-10	IPR002196	Glycoside hydrolase, family 24	GO:0003796(InterPro)|GO:0009253(InterPro)|GO:0016998(InterPro)

vB EcoM-CHD2BS1 MZ234027	lysozyme R	QZI80464.1	157128 - 157622	164	PRINTS	PR00684	Phage T4 lysozyme signature	95	114	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM-CHD2BS1 MZ234027	lysozyme R	QZI80464.1	157128 - 157622	164	PRINTS	PR00684	Phage T4 lysozyme signature	72	92	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM-CHD2BS1 MZ234027	lysozyme R	QZI80464.1	157128 - 157622	164	PRINTS	PR00684	Phage T4 lysozyme signature	4	23	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM-CHD2BS1 MZ234027	lysozyme R	QZI80464.1	157128 - 157622	164	PRINTS	PR00684	Phage T4 lysozyme signature	24	42	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM-CHD2BS1 MZ234027	lysozyme R	QZI80464.1	157128 - 157622	164	PRINTS	PR00684	Phage T4 lysozyme signature	118	137	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM-CHD2BS1 MZ234027	lysozyme R	QZI80464.1	157128 - 157622	164	PRINTS	PR00684	Phage T4 lysozyme signature	51	70	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM-CHD2BS1 MZ234027	lysozyme R	QZI80464.1	157128 - 157622	164	PRINTS	PR00684	Phage T4 lysozyme signature	138	159	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM-CHD2BS1 MZ234027	lysozyme R	QZI80464.1	157128 - 157622	164	CDD	cd00735	T4-like lys	2	160	6.07455E-78	-	-	-
vB EcoM-CHD2BS1 MZ234027	lysozyme R	QZI80464.1	157128 - 157622	164	PANTHER	PTHR37406	T4-TYPE LYSOZYME 1-RELATED	4	161	1.5E-27	IPR052619	Bacteriophage lysozyme-like	-
vB EcoM-CHD16UKE1 MZ234030	lysozyme R	QZI80907.1	157094 - 157588	164	Gene3D	G3DSA:1.10.530.40	-	1	164	6.6E-75	IPR023347	Lysozyme domain superfamily	GO:0003796(InterPro)
vB EcoM-CHD16UKE1 MZ234030	lysozyme R	QZI80907.1	157094 - 157588	164	FunFam	G3DSA:1.10.530.40:FF:000002	Endolysin	1	164	1.1E-114	-	-	-
vB EcoM-CHD16UKE1 MZ234030	lysozyme R	QZI80907.1	157094 - 157588	164	Hamap	MF 04110	Endolysin.	1	159	16.677807	IPR034690	Endolysin T4 type	GO:0003796(InterPro)|GO:0009253(InterPro)
vB EcoM-CHD16UKE1 MZ234030	lysozyme R	QZI80907.1	157094 - 157588	164	SUPERFAMILY	SSF53955	Lysozyme-like	1	161	4.81E-54	IPR023346	Lysozyme-like domain superfamily	-
vB EcoM-CHD16UKE1 MZ234030	lysozyme R	QZI80907.1	157094 - 157588	164	Pfam	PF00959	Phage lysozyme	11	141	4.4E-10	IPR002196	Glycoside hydrolase, family 24	GO:0003796(InterPro)|GO:0009253(InterPro)|GO:0016998(InterPro)
vB EcoM-CHD16UKE1 MZ234030	lysozyme R	QZI80907.1	157094 - 157588	164	PRINTS	PR00684	Phage T4 lysozyme signature	95	114	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM-CHD16UKE1 MZ234030	lysozyme R	QZI80907.1	157094 - 157588	164	PRINTS	PR00684	Phage T4 lysozyme signature	72	92	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM-CHD16UKE1 MZ234030	lysozyme R	QZI80907.1	157094 - 157588	164	PRINTS	PR00684	Phage T4 lysozyme signature	4	23	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM-CHD16UKE1 MZ234030	lysozyme R	QZI80907.1	157094 - 157588	164	PRINTS	PR00684	Phage T4 lysozyme signature	24	42	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)

vB EcoM-CHD16UKE1 MZ234030	lysozyme R	QZI80907.1	157094 - 157588	164	PRINTS	PR00684	Phage T4 lysozyme signature	118	137	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM-CHD16UKE1 MZ234030	lysozyme R	QZI80907.1	157094 - 157588	164	PRINTS	PR00684	Phage T4 lysozyme signature	51	70	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM-CHD16UKE1 MZ234030	lysozyme R	QZI80907.1	157094 - 157588	164	PRINTS	PR00684	Phage T4 lysozyme signature	138	159	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM-CHD16UKE1 MZ234030	lysozyme R	QZI80907.1	157094 - 157588	164	CDD	cd00735	T4-like lys	2	160	6.07455E-78	-	-	-
vB EcoM-CHD16UKE1 MZ234030	lysozyme R	QZI80907.1	157094 - 157588	164	PANTHER	PTHR37406	T4-TYPE LYSOZYME 1-RELATED	4	161	1.5E-27	IPR052619	Bacteriophage lysozyme-like	-
vB EcoM-CHD94UKE2 MZ234031	lysozyme R	QZI81192.1	154846 - 155334	162	Gene3D	G3DSA:1.10.530.40	-	2	162	8.2E-104	IPR023347	Lysozyme domain superfamily	GO:0003796(InterPro)
vB EcoM-CHD94UKE2 MZ234031	lysozyme R	QZI81192.1	154846 - 155334	162	Hamap	MF 04110	Endolysin.	1	159	16.158941	IPR034690	Endolysin T4 type	GO:0003796(InterPro)|GO:0009253(InterPro)
vB EcoM-CHD94UKE2 MZ234031	lysozyme R	QZI81192.1	154846 - 155334	162	SUPERFAMILY	SSF53955	Lysozyme-like	1	161	2.86E-53	IPR023346	Lysozyme-like domain superfamily	-
vB EcoM-CHD94UKE2 MZ234031	lysozyme R	QZI81192.1	154846 - 155334	162	Pfam	PF00959	Phage lysozyme	11	140	6.4E-10	IPR002196	Glycoside hydrolase, family 24	GO:0003796(InterPro)|GO:0009253(InterPro)|GO:0016998(InterPro)
vB EcoM-CHD94UKE2 MZ234031	lysozyme R	QZI81192.1	154846 - 155334	162	PRINTS	PR00684	Phage T4 lysozyme signature	24	42	1.9E-88	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM-CHD94UKE2 MZ234031	lysozyme R	QZI81192.1	154846 - 155334	162	PRINTS	PR00684	Phage T4 lysozyme signature	95	114	1.9E-88	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM-CHD94UKE2 MZ234031	lysozyme R	QZI81192.1	154846 - 155334	162	PRINTS	PR00684	Phage T4 lysozyme signature	72	92	1.9E-88	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM-CHD94UKE2 MZ234031	lysozyme R	QZI81192.1	154846 - 155334	162	PRINTS	PR00684	Phage T4 lysozyme signature	118	137	1.9E-88	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM-CHD94UKE2 MZ234031	lysozyme R	QZI81192.1	154846 - 155334	162	PRINTS	PR00684	Phage T4 lysozyme signature	138	159	1.9E-88	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM-CHD94UKE2 MZ234031	lysozyme R	QZI81192.1	154846 - 155334	162	PRINTS	PR00684	Phage T4 lysozyme signature	51	70	1.9E-88	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)

vB EcoM-CHD94UKE2 MZ234031	lysozyme R	QZI81192.1	154846 - 155334	162	PRINTS	PR00684	Phage T4 lysozyme signature	4	23	1.9E-88	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM-CHD94UKE2 MZ234031	lysozyme R	QZI81192.1	154846 - 155334	162	CDD	cd00735	T4-like lys	2	160	4.4557E-77	-	-	-
vB EcoM-CHD94UKE2 MZ234031	lysozyme R	QZI81192.1	154846 - 155334	162	PANTHER	PTHR37406	T4-TYPE LYSOZYME 1-RELATED	4	160	1.8E-27	IPR052619	Bacteriophage lysozyme-like	-
vB EcoM-G3G7 MZ234040	lysozyme R	QZI83386.1	157200 - 157694	164	Gene3D	G3DSA:1.10.530.40	-	1	164	6.6E-75	IPR023347	Lysozyme domain superfamily	GO:0003796(InterPro)
vB EcoM-G3G7 MZ234040	lysozyme R	QZI83386.1	157200 - 157694	164	FunFam	G3DSA:1.10.530.40:FF:000002	Endolysin	1	164	1.1E-114	-	-	-
vB EcoM-G3G7 MZ234040	lysozyme R	QZI83386.1	157200 - 157694	164	Hamap	MF 04110	Endolysin.	1	159	16.677807	IPR034690	Endolysin T4 type	GO:0003796(InterPro)|GO:0009253(InterPro)
vB EcoM-G3G7 MZ234040	lysozyme R	QZI83386.1	157200 - 157694	164	SUPERFAMILY	SSF53955	Lysozyme-like	1	161	4.81E-54	IPR023346	Lysozyme-like domain superfamily	-
vB EcoM-G3G7 MZ234040	lysozyme R	QZI83386.1	157200 - 157694	164	Pfam	PF00959	Phage lysozyme	11	141	4.4E-10	IPR002196	Glycoside hydrolase, family 24	GO:0003796(InterPro)|GO:0009253(InterPro)|GO:0016998(InterPro)
vB EcoM-G3G7 MZ234040	lysozyme R	QZI83386.1	157200 - 157694	164	PRINTS	PR00684	Phage T4 lysozyme signature	95	114	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM-G3G7 MZ234040	lysozyme R	QZI83386.1	157200 - 157694	164	PRINTS	PR00684	Phage T4 lysozyme signature	72	92	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM-G3G7 MZ234040	lysozyme R	QZI83386.1	157200 - 157694	164	PRINTS	PR00684	Phage T4 lysozyme signature	4	23	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM-G3G7 MZ234040	lysozyme R	QZI83386.1	157200 - 157694	164	PRINTS	PR00684	Phage T4 lysozyme signature	24	42	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM-G3G7 MZ234040	lysozyme R	QZI83386.1	157200 - 157694	164	PRINTS	PR00684	Phage T4 lysozyme signature	118	137	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM-G3G7 MZ234040	lysozyme R	QZI83386.1	157200 - 157694	164	PRINTS	PR00684	Phage T4 lysozyme signature	51	70	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM-G3G7 MZ234040	lysozyme R	QZI83386.1	157200 - 157694	164	PRINTS	PR00684	Phage T4 lysozyme signature	138	159	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM-G3G7 MZ234040	lysozyme R	QZI83386.1	157200 - 157694	164	CDD	cd00735	T4-like lys	2	160	6.07455E-78	-	-	-
vB EcoM-G3G7 MZ234040	lysozyme R	QZI83386.1	157200 - 157694	164	PANTHER	PTHR37406	T4-TYPE LYSOZYME 1-RELATED	4	161	1.5E-27	IPR052619	Bacteriophage lysozyme-like	-

vB EcoP-UTI89UKE2 MZ234049	hypothetical protein	QZI84581.1	36944 - 37288	114	Gene3D	G3DSA:3.30.1380.10	-	5	114	1.1E-25	IPR009045	Hedgehog signalling/DD-peptidase zinc-binding domain superfamily	-
vB EcoP-UTI89UKE2 MZ234049	hypothetical protein	QZI84581.1	36944 - 37288	114	SUPERFAMILY	SSF55166	Hedgehog/DD-peptidase	5	107	1.29E-24	IPR009045	Hedgehog signalling/DD-peptidase zinc-binding domain superfamily	-
vB EcoP-UTI89UKE2 MZ234049	hypothetical protein	QZI84581.1	36944 - 37288	114	Pfam	PF08291	Peptidase M15	4	106	2.9E-29	IPR013230	Peptidase M15A, C-terminal	-
vB EcoP-UTI89UKE3 MZ234050	hypothetical protein	QZI84648.1	36945 - 37289	114	Gene3D	G3DSA:3.30.1380.10	-	5	114	1.1E-25	IPR009045	Hedgehog signalling/DD-peptidase zinc-binding domain superfamily	-
vB EcoP-UTI89UKE3 MZ234050	hypothetical protein	QZI84648.1	36945 - 37289	114	SUPERFAMILY	SSF55166	Hedgehog/DD-peptidase	5	107	1.29E-24	IPR009045	Hedgehog signalling/DD-peptidase zinc-binding domain superfamily	-
vB EcoP-UTI89UKE3 MZ234050	hypothetical protein	QZI84648.1	36945 - 37289	114	Pfam	PF08291	Peptidase M15	4	106	2.9E-29	IPR013230	Peptidase M15A, C-terminal	-
vB EcoM SHAK7854 OR594182	amidase	self DNYN29	13662 - 14120	152	Gene3D	G3DSA:3.40.80.10	-	8	152	7.6E-49	IPR036505	N-acetylmuramoyl-L-alanine amidase/PGRP domain superfamily	GO:0008745(InterPro)|GO:0009253(InterPro)
vB EcoM SHAK7854 OR594182	amidase	self DNYN29	13662 - 14120	152	SMART	SM00644	ami 2	2	134	1.8E-18	IPR002502	N-acetylmuramoyl-L-alanine amidase domain	GO:0008745(InterPro)|GO:0009253(InterPro)
vB EcoM SHAK7854 OR594182	amidase	self DNYN29	13662 - 14120	152	Coils	Coil	Coil	102	122	-	-	-	-

vB EcoM SHAK7854 OR594182	amidase	self DNYN29	13662 - 14120	152	Hamap	MF 04111	Endolysin.	2	150	47.195431	IPR034689	Endolysin T7 type	GO:0008745(InterPro)|GO:0009253(InterPro)|GO:0032897(InterPro)|GO:0044659(InterPro)
vB EcoM SHAK7854 OR594182	amidase	self DNYN29	13662 - 14120	152	Pfam	PF01510	N-acetylmuramoyl-L-alanine amidase	11	134	7.2E-13	IPR002502	N-acetylmuramoyl-L-alanine amidase domain	GO:0008745(InterPro)|GO:0009253(InterPro)
vB EcoM SHAK7854 OR594182	amidase	self DNYN29	13662 - 14120	152	SUPERFAMILY	SSF55846	N-acetylmuramoyl-L-alanine amidase-like	11	147	1.18E-41	IPR036505	N-acetylmuramoyl-L-alanine amidase/PGRP domain superfamily	GO:0008745(InterPro)|GO:0009253(InterPro)
vB EcoM SHAK7854 OR594182	amidase	self DNYN29	13662 - 14120	152	PANTHER	PTHR11022	PEPTIDOGLYCAN RECOGNITION PROTEIN	16	138	1.7E-17	IPR015510	Peptidoglycan recognition protein	-
vB EcoM SHAK7854 OR594182	amidase	self DNYN29	13662 - 14120	152	CDD	cd06583	PGRP	14	133	2.01169E-26	IPR002502	N-acetylmuramoyl-L-alanine amidase domain	GO:0008745(InterPro)|GO:0009253(InterPro)
vB EcoM SHAK7854 OR594182	amidase	self DNYN29	13662 - 14120	152	SMART	SM00701	pgrp	2	128	1.3E-17	IPR006619	Peptidoglycan recognition protein family domain, metazoa/bacteria	GO:0008270(InterPro)|GO:0008745(InterPro)|GO:0009253(InterPro)
vB EcoP LNA3 MZ311862	amidase	self GXIJ34	26726 - 27184	152	Gene3D	G3DSA:3.40.80.10	-	8	152	4.2E-49	IPR036505	N-acetylmuramoyl-L-alanine amidase/PGRP domain superfamily	GO:0008745(InterPro)|GO:0009253(InterPro)
vB EcoP LNA3 MZ311862	amidase	self GXIJ34	26726 - 27184	152	SMART	SM00644	ami 2	2	134	2.0E-19	IPR002502	N-acetylmuramoyl-L-alanine amidase domain	GO:0008745(InterPro)|GO:0009253(InterPro)
vB EcoP LNA3 MZ311862	amidase	self GXIJ34	26726 - 27184	152	Coils	Coil	Coil	102	122	-	-	-	-

vB EcoP LNA3 MZ311862	amidase	self GXIJ34	26726 - 27184	152	Hamap	MF 04111	Endolysin.	2	150	47.110817	IPR034689	Endolysin T7 type	GO:0008745(InterPro)|GO:0009253(InterPro)|GO:0032897(InterPro)|GO:0044659(InterPro)
vB EcoP LNA3 MZ311862	amidase	self GXIJ34	26726 - 27184	152	Pfam	PF01510	N-acetylmuramoyl-L-alanine amidase	11	134	2.0E-13	IPR002502	N-acetylmuramoyl-L-alanine amidase domain	GO:0008745(InterPro)|GO:0009253(InterPro)
vB EcoP LNA3 MZ311862	amidase	self GXIJ34	26726 - 27184	152	SUPERFAMILY	SSF55846	N-acetylmuramoyl-L-alanine amidase-like	11	147	7.2E-42	IPR036505	N-acetylmuramoyl-L-alanine amidase/PGRP domain superfamily	GO:0008745(InterPro)|GO:0009253(InterPro)
vB EcoP LNA3 MZ311862	amidase	self GXIJ34	26726 - 27184	152	PANTHER	PTHR11022	PEPTIDOGLYCAN RECOGNITION PROTEIN	16	138	1.0E-17	IPR015510	Peptidoglycan recognition protein	-
vB EcoP LNA3 MZ311862	amidase	self GXIJ34	26726 - 27184	152	CDD	cd06583	PGRP	14	133	1.00503E-26	IPR002502	N-acetylmuramoyl-L-alanine amidase domain	GO:0008745(InterPro)|GO:0009253(InterPro)
vB EcoP LNA3 MZ311862	amidase	self GXIJ34	26726 - 27184	152	SMART	SM00701	pgrp	2	128	2.7E-16	IPR006619	Peptidoglycan recognition protein family domain, metazoa/bacteria	GO:0008270(InterPro)|GO:0008745(InterPro)|GO:0009253(InterPro)
vB EcoM LNA2 MZ311865	hypothetical protein	self NLWW35	12924 - 13412	162	Gene3D	G3DSA:1.10.530.40	-	2	162	2.7E-88	IPR023347	Lysozyme domain superfamily	GO:0003796(InterPro)
vB EcoM LNA2 MZ311865	hypothetical protein	self NLWW35	12924 - 13412	162	SUPERFAMILY	SSF53955	Lysozyme-like	1	161	2.66E-53	IPR023346	Lysozyme-like domain superfamily	-
vB EcoM LNA2 MZ311865	hypothetical protein	self NLWW35	12924 - 13412	162	Pfam	PF00959	Phage lysozyme	10	140	1.1E-10	IPR002196	Glycoside hydrolase, family 24	GO:0003796(InterPro)|GO:0009253(InterPro)|GO:0016998(InterPro)
vB EcoM LNA2 MZ311865	hypothetical protein	self NLWW35	12924 - 13412	162	PRINTS	PR00684	Phage T4 lysozyme signature	4	23	4.1E-75	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM LNA2 MZ311865	hypothetical protein	self NLWW35	12924 - 13412	162	PRINTS	PR00684	Phage T4 lysozyme signature	138	159	4.1E-75	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)

vB EcoM LNA2 MZ311865	hypothetical protein	self NLWW35	12924 - 13412	162	PRINTS	PR00684	Phage T4 lysozyme signature	24	42	4.1E-75	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM LNA2 MZ311865	hypothetical protein	self NLWW35	12924 - 13412	162	PRINTS	PR00684	Phage T4 lysozyme signature	72	92	4.1E-75	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM LNA2 MZ311865	hypothetical protein	self NLWW35	12924 - 13412	162	PRINTS	PR00684	Phage T4 lysozyme signature	118	137	4.1E-75	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM LNA2 MZ311865	hypothetical protein	self NLWW35	12924 - 13412	162	PRINTS	PR00684	Phage T4 lysozyme signature	95	114	4.1E-75	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM LNA2 MZ311865	hypothetical protein	self NLWW35	12924 - 13412	162	PRINTS	PR00684	Phage T4 lysozyme signature	51	70	4.1E-75	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM LNA2 MZ311865	hypothetical protein	self NLWW35	12924 - 13412	162	CDD	cd00735	T4-like lys	2	160	4.11032E-74	-	-	-
vB EcoM LNA2 MZ311865	hypothetical protein	self NLWW35	12924 - 13412	162	PANTHER	PTHR37406	T4-TYPE LYSOZYME 1-RELATED	4	160	2.4E-26	IPR052619	Bacteriophage lysozyme-like	-
vB EcoP LNA5 MZ311861	amidase	self TCEH12	6617 - 7075	152	Gene3D	G3DSA:3.40.80.10	-	8	152	2.2E-49	IPR036505	N-acetylmuramoyl-L-alanine amidase/PGRP domain superfamily	GO:0008745(InterPro)|GO:0009253(InterPro)
vB EcoP LNA5 MZ311861	amidase	self TCEH12	6617 - 7075	152	SMART	SM00644	ami 2	2	134	3.4E-19	IPR002502	N-acetylmuramoyl-L-alanine amidase domain	GO:0008745(InterPro)|GO:0009253(InterPro)
vB EcoP LNA5 MZ311861	amidase	self TCEH12	6617 - 7075	152	Coils	Coil	Coil	102	122	-	-	-	-
vB EcoP LNA5 MZ311861	amidase	self TCEH12	6617 - 7075	152	Hamap	MF 04111	Endolysin.	2	150	47.084782	IPR034689	Endolysin T7 type	GO:0008745(InterPro)|GO:0009253(InterPro)|GO:0032897(InterPro)|GO:0044659(InterPro)
vB EcoP LNA5 MZ311861	amidase	self TCEH12	6617 - 7075	152	Pfam	PF01510	N-acetylmuramoyl-L-alanine amidase	11	134	2.3E-13	IPR002502	N-acetylmuramoyl-L-alanine amidase domain	GO:0008745(InterPro)|GO:0009253(InterPro)
vB EcoP LNA5 MZ311861	amidase	self TCEH12	6617 - 7075	152	SUPERFAMILY	SSF55846	N-acetylmuramoyl-L-alanine amidase-like	11	147	1.01E-41	IPR036505	N-acetylmuramoyl-L-alanine amidase/PGRP domain superfamily	GO:0008745(InterPro)|GO:0009253(InterPro)

vB EcoP LNA5 MZ311861	amidase	self TCEH12	6617 - 7075	152	PANTHER	PTHR11022	PEPTIDOGLYCAN RECOGNITION PROTEIN	16	138	5.2E-18	IPR015510	Peptidoglycan recognition protein	-
vB EcoP LNA5 MZ311861	amidase	self TCEH12	6617 - 7075	152	CDD	cd06583	PGRP	14	133	6.1692E-27	IPR002502	N-acetylmuramoyl-L-alanine amidase domain	GO:0008745(InterPro)|GO:0009253(InterPro)
vB EcoP LNA5 MZ311861	amidase	self TCEH12	6617 - 7075	152	SMART	SM00701	pgrp	2	128	6.8E-18	IPR006619	Peptidoglycan recognition protein family domain, metazoa/bacteria	GO:0008270(InterPro)|GO:0008745(InterPro)|GO:0009253(InterPro)
vB EcoM SHAK9454 OR594184	amidase	self ZENB29	13615 - 14073	152	Gene3D	G3DSA:3.40.80.10	-	8	152	8.7E-49	IPR036505	N-acetylmuramoyl-L-alanine amidase/PGRP domain superfamily	GO:0008745(InterPro)|GO:0009253(InterPro)
vB EcoM SHAK9454 OR594184	amidase	self ZENB29	13615 - 14073	152	SMART	SM00644	ami 2	2	134	1.2E-19	IPR002502	N-acetylmuramoyl-L-alanine amidase domain	GO:0008745(InterPro)|GO:0009253(InterPro)
vB EcoM SHAK9454 OR594184	amidase	self ZENB29	13615 - 14073	152	Coils	Coil	Coil	102	122	-	-	-	-
vB EcoM SHAK9454 OR594184	amidase	self ZENB29	13615 - 14073	152	Hamap	MF 04111	Endolysin.	2	150	47.260513	IPR034689	Endolysin T7 type	GO:0008745(InterPro)|GO:0009253(InterPro)|GO:0032897(InterPro)|GO:0044659(InterPro)
vB EcoM SHAK9454 OR594184	amidase	self ZENB29	13615 - 14073	152	Pfam	PF01510	N-acetylmuramoyl-L-alanine amidase	11	134	8.3E-13	IPR002502	N-acetylmuramoyl-L-alanine amidase domain	GO:0008745(InterPro)|GO:0009253(InterPro)
vB EcoM SHAK9454 OR594184	amidase	self ZENB29	13615 - 14073	152	SUPERFAMILY	SSF55846	N-acetylmuramoyl-L-alanine amidase-like	11	147	1.96E-41	IPR036505	N-acetylmuramoyl-L-alanine amidase/PGRP domain superfamily	GO:0008745(InterPro)|GO:0009253(InterPro)

vB EcoM SHAK9454 OR594184	amidase	self ZENB29	13615 - 14073	152	PANTHER	PTHR11022	PEPTIDOGLYCAN RECOGNITION PROTEIN	16	138	2.3E-17	IPR015510	Peptidoglycan recognition protein	-
vB EcoM SHAK9454 OR594184	amidase	self ZENB29	13615 - 14073	152	SMART	SM00701	pgrp	2	128	3.8E-17	IPR006619	Peptidoglycan recognition protein family domain, metazoa/bacteria	GO:0008270(InterPro)|GO:0008745(InterPro)|GO:0009253(InterPro)
vB EcoM SHAK9454 OR594184	amidase	self ZENB29	13615 - 14073	152	CDD	cd06583	PGRP	14	133	2.36694E-26	IPR002502	N-acetylmuramoyl-L-alanine amidase domain	GO:0008745(InterPro)|GO:0009253(InterPro)
MLP2 OK148441	baseplate hub + tail lysozyme	UEN68765.1	156422 - 156916	164	Gene3D	G3DSA:1.10.530.40	-	2	162	2.2E-106	IPR023347	Lysozyme domain superfamily	GO:0003796(InterPro)
MLP2 OK148441	baseplate hub + tail lysozyme	UEN68765.1	156422 - 156916	164	FunFam	G3DSA:1.10.530.40:FF:000002	Endolysin	1	164	6.0E-114	-	-	-
MLP2 OK148441	baseplate hub + tail lysozyme	UEN68765.1	156422 - 156916	164	Hamap	MF 04110	Endolysin.	1	159	16.677807	IPR034690	Endolysin T4 type	GO:0003796(InterPro)|GO:0009253(InterPro)
MLP2 OK148441	baseplate hub + tail lysozyme	UEN68765.1	156422 - 156916	164	SUPERFAMILY	SSF53955	Lysozyme-like	1	161	5.12E-54	IPR023346	Lysozyme-like domain superfamily	-
MLP2 OK148441	baseplate hub + tail lysozyme	UEN68765.1	156422 - 156916	164	Pfam	PF00959	Phage lysozyme	11	141	4.9E-10	IPR002196	Glycoside hydrolase, family 24	GO:0003796(InterPro)|GO:0009253(InterPro)|GO:0016998(InterPro)
MLP2 OK148441	baseplate hub + tail lysozyme	UEN68765.1	156422 - 156916	164	PRINTS	PR00684	Phage T4 lysozyme signature	95	114	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
MLP2 OK148441	baseplate hub + tail lysozyme	UEN68765.1	156422 - 156916	164	PRINTS	PR00684	Phage T4 lysozyme signature	72	92	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
MLP2 OK148441	baseplate hub + tail lysozyme	UEN68765.1	156422 - 156916	164	PRINTS	PR00684	Phage T4 lysozyme signature	4	23	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
MLP2 OK148441	baseplate hub + tail lysozyme	UEN68765.1	156422 - 156916	164	PRINTS	PR00684	Phage T4 lysozyme signature	24	42	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
MLP2 OK148441	baseplate hub + tail lysozyme	UEN68765.1	156422 - 156916	164	PRINTS	PR00684	Phage T4 lysozyme signature	118	137	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
MLP2 OK148441	baseplate hub + tail lysozyme	UEN68765.1	156422 - 156916	164	PRINTS	PR00684	Phage T4 lysozyme signature	51	70	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
MLP2 OK148441	baseplate hub + tail lysozyme	UEN68765.1	156422 - 156916	164	PRINTS	PR00684	Phage T4 lysozyme signature	138	159	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
MLP2 OK148441	baseplate hub + tail lysozyme	UEN68765.1	156422 - 156916	164	CDD	cd00735	T4-like lys	2	160	5.15369E-78	-	-	-

MLP2 OK148441	baseplate hub + tail lysozyme	UEN68765.1	156422 - 156916	164	PANTHER	PTHR37406	T4-TYPE LYSOZYME 1-RELATED	4	161	2.6E-27	IPR052619	Bacteriophage lysozyme-like	-
vB EcoS-CHD2B1 MZ234026	lysin	USL86318.1	42121 - 42612	163	PANTHER	PTHR38107	-	2	147	3.6E-18	IPR051018	Bacteriophage Glycosyl Hydrolase 24	GO:0003796(PANTHER)
vB EcoS-CHD2B1 MZ234026	lysin	USL86318.1	42121 - 42612	163	Gene3D	G3DSA:1.10.530.40	-	2	154	9.0E-31	IPR023347	Lysozyme domain superfamily	GO:0003796(InterPro)
vB EcoS-CHD2B1 MZ234026	lysin	USL86318.1	42121 - 42612	163	SUPERFAMILY	SSF53955	Lysozyme-like	1	153	3.79E-32	IPR023346	Lysozyme-like domain superfamily	-
vB EcoS-CHD2B1 MZ234026	lysin	USL86318.1	42121 - 42612	163	CDD	cd00737	lyz endolysin autolysin	7	148	1.42938E-36	IPR033907	Endolysin/autolysin	GO:0003796(InterPro)
vB EcoS-CHD2B1 MZ234026	lysin	USL86318.1	42121 - 42612	163	Pfam	PF00959	Phage lysozyme	14	139	4.6E-18	IPR002196	Glycoside hydrolase, family 24	GO:0003796(InterPro)|GO:0009253(InterPro)|GO:0016998(InterPro)
vB EcoS-CHD5UKE2 MZ234029	lysin	USL86352.1	14512 - 15003	163	Gene3D	G3DSA:1.10.530.40	-	2	153	2.0E-30	IPR023347	Lysozyme domain superfamily	GO:0003796(InterPro)
vB EcoS-CHD5UKE2 MZ234029	lysin	USL86352.1	14512 - 15003	163	PANTHER	PTHR38107	-	2	147	1.7E-17	IPR051018	Bacteriophage Glycosyl Hydrolase 24	GO:0003796(PANTHER)
vB EcoS-CHD5UKE2 MZ234029	lysin	USL86352.1	14512 - 15003	163	SUPERFAMILY	SSF53955	Lysozyme-like	1	153	5.12E-32	IPR023346	Lysozyme-like domain superfamily	-
vB EcoS-CHD5UKE2 MZ234029	lysin	USL86352.1	14512 - 15003	163	CDD	cd00737	lyz endolysin autolysin	7	148	1.1715E-34	IPR033907	Endolysin/autolysin	GO:0003796(InterPro)
vB EcoS-CHD5UKE2 MZ234029	lysin	USL86352.1	14512 - 15003	163	Pfam	PF00959	Phage lysozyme	14	139	1.1E-17	IPR002196	Glycoside hydrolase, family 24	GO:0003796(InterPro)|GO:0009253(InterPro)|GO:0016998(InterPro)
vB EcoS-UTI89UKE1 MZ234048	lysin	USM81274.1	361 - 849	162	Gene3D	G3DSA:1.10.530.40	-	2	147	3.3E-32	IPR023347	Lysozyme domain superfamily	GO:0003796(InterPro)
vB EcoS-UTI89UKE1 MZ234048	lysin	USM81274.1	361 - 849	162	PANTHER	PTHR38107	-	3	144	1.4E-22	IPR051018	Bacteriophage Glycosyl Hydrolase 24	GO:0003796(PANTHER)
vB EcoS-UTI89UKE1 MZ234048	lysin	USM81274.1	361 - 849	162	SUPERFAMILY	SSF53955	Lysozyme-like	3	149	5.12E-32	IPR023346	Lysozyme-like domain superfamily	-
vB EcoS-UTI89UKE1 MZ234048	lysin	USM81274.1	361 - 849	162	CDD	cd00737	lyz endolysin autolysin	9	146	6.59567E-31	IPR033907	Endolysin/autolysin	GO:0003796(InterPro)
vB EcoS-UTI89UKE1 MZ234048	lysin	USM81274.1	361 - 849	162	Pfam	PF00959	Phage lysozyme	15	138	1.7E-23	IPR002196	Glycoside hydrolase, family 24	GO:0003796(InterPro)|GO:0009253(InterPro)|GO:0016998(InterPro)

ZCEC14 OP168898	hypothetical protein	UVD33441.1	96330 - 96818	162	Gene3D	G3DSA:1.10.530.40	-	2	162	1.3E-106	IPR023347	Lysozyme domain superfamily	GO:0003796(InterPro)
ZCEC14 OP168898	hypothetical protein	UVD33441.1	96330 - 96818	162	Hamap	MF 04110	Endolysin.	1	159	16.495243	IPR034690	Endolysin T4 type	GO:0003796(InterPro)|GO:0009253(InterPro)
ZCEC14 OP168898	hypothetical protein	UVD33441.1	96330 - 96818	162	SUPERFAMILY	SSF53955	Lysozyme-like	1	161	1.18E-53	IPR023346	Lysozyme-like domain superfamily	-
ZCEC14 OP168898	hypothetical protein	UVD33441.1	96330 - 96818	162	Pfam	PF00959	Phage lysozyme	11	141	3.8E-10	IPR002196	Glycoside hydrolase, family 24	GO:0003796(InterPro)|GO:0009253(InterPro)|GO:0016998(InterPro)
ZCEC14 OP168898	hypothetical protein	UVD33441.1	96330 - 96818	162	PRINTS	PR00684	Phage T4 lysozyme signature	95	114	6.6E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
ZCEC14 OP168898	hypothetical protein	UVD33441.1	96330 - 96818	162	PRINTS	PR00684	Phage T4 lysozyme signature	24	42	6.6E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
ZCEC14 OP168898	hypothetical protein	UVD33441.1	96330 - 96818	162	PRINTS	PR00684	Phage T4 lysozyme signature	72	92	6.6E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
ZCEC14 OP168898	hypothetical protein	UVD33441.1	96330 - 96818	162	PRINTS	PR00684	Phage T4 lysozyme signature	51	70	6.6E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
ZCEC14 OP168898	hypothetical protein	UVD33441.1	96330 - 96818	162	PRINTS	PR00684	Phage T4 lysozyme signature	4	23	6.6E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
ZCEC14 OP168898	hypothetical protein	UVD33441.1	96330 - 96818	162	PRINTS	PR00684	Phage T4 lysozyme signature	118	137	6.6E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
ZCEC14 OP168898	hypothetical protein	UVD33441.1	96330 - 96818	162	PRINTS	PR00684	Phage T4 lysozyme signature	138	159	6.6E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
ZCEC14 OP168898	hypothetical protein	UVD33441.1	96330 - 96818	162	CDD	cd00735	T4-like lys	2	160	1.15612E-76	-	-	-
ZCEC14 OP168898	hypothetical protein	UVD33441.1	96330 - 96818	162	PANTHER	PTHR37406	T4-TYPE LYSOZYME 1-RELATED	4	160	1.8E-27	IPR052619	Bacteriophage lysozyme-like	-
vB-Eco-KMB22 OR539217	lysozyme R	WPJ70775.1	64127 - 64621	164	Gene3D	G3DSA:1.10.530.40	-	2	162	2.1E-106	IPR023347	Lysozyme domain superfamily	GO:0003796(InterPro)
vB-Eco-KMB22 OR539217	lysozyme R	WPJ70775.1	64127 - 64621	164	FunFam	G3DSA:1.10.530.40:FF:000002	Endolysin	1	164	5.8E-114	-	-	-
vB-Eco-KMB22 OR539217	lysozyme R	WPJ70775.1	64127 - 64621	164	Hamap	MF 04110	Endolysin.	1	159	16.600937	IPR034690	Endolysin T4 type	GO:0003796(InterPro)|GO:0009253(InterPro)
vB-Eco-KMB22 OR539217	lysozyme R	WPJ70775.1	64127 - 64621	164	SUPERFAMILY	SSF53955	Lysozyme-like	1	161	1.18E-53	IPR023346	Lysozyme-like domain superfamily	-
vB-Eco-KMB22 OR539217	lysozyme R	WPJ70775.1	64127 - 64621	164	Pfam	PF00959	Phage lysozyme	11	141	6.8E-10	IPR002196	Glycoside hydrolase, family 24	GO:0003796(InterPro)|GO:0009253(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB22 OR539217	lysozyme R	WPJ70775.1	64127 - 64621	164	PRINTS	PR00684	Phage T4 lysozyme signature	95	114	3.8E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)

vB-Eco-KMB22 OR539217	lysozyme R	WPJ70775.1	64127 - 64621	164	PRINTS	PR00684	Phage T4 lysozyme signature	51	70	3.8E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB22 OR539217	lysozyme R	WPJ70775.1	64127 - 64621	164	PRINTS	PR00684	Phage T4 lysozyme signature	72	92	3.8E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB22 OR539217	lysozyme R	WPJ70775.1	64127 - 64621	164	PRINTS	PR00684	Phage T4 lysozyme signature	4	23	3.8E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB22 OR539217	lysozyme R	WPJ70775.1	64127 - 64621	164	PRINTS	PR00684	Phage T4 lysozyme signature	24	42	3.8E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB22 OR539217	lysozyme R	WPJ70775.1	64127 - 64621	164	PRINTS	PR00684	Phage T4 lysozyme signature	118	137	3.8E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB22 OR539217	lysozyme R	WPJ70775.1	64127 - 64621	164	PRINTS	PR00684	Phage T4 lysozyme signature	138	159	3.8E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB22 OR539217	lysozyme R	WPJ70775.1	64127 - 64621	164	CDD	cd00735	T4-like lys	2	160	1.21164E-77	-	-	-
vB-Eco-KMB22 OR539217	lysozyme R	WPJ70775.1	64127 - 64621	164	PANTHER	PTHR37406	T4-TYPE LYSOZYME 1-RELATED	4	161	1.7E-27	IPR052619	Bacteriophage lysozyme-like	-
vB-Eco-KMB26 OR539218	lysozyme R	WPJ71039.1	64196 - 64690	164	Gene3D	G3DSA:1.10.530.40	-	1	164	6.6E-75	IPR023347	Lysozyme domain superfamily	GO:0003796(InterPro)
vB-Eco-KMB26 OR539218	lysozyme R	WPJ71039.1	64196 - 64690	164	FunFam	G3DSA:1.10.530.40:FF:000002	Endolysin	1	164	1.1E-114	-	-	-
vB-Eco-KMB26 OR539218	lysozyme R	WPJ71039.1	64196 - 64690	164	Hamap	MF 04110	Endolysin.	1	159	16.677807	IPR034690	Endolysin T4 type	GO:0003796(InterPro)|GO:0009253(InterPro)
vB-Eco-KMB26 OR539218	lysozyme R	WPJ71039.1	64196 - 64690	164	SUPERFAMILY	SSF53955	Lysozyme-like	1	161	4.81E-54	IPR023346	Lysozyme-like domain superfamily	-
vB-Eco-KMB26 OR539218	lysozyme R	WPJ71039.1	64196 - 64690	164	Pfam	PF00959	Phage lysozyme	11	141	4.4E-10	IPR002196	Glycoside hydrolase, family 24	GO:0003796(InterPro)|GO:0009253(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB26 OR539218	lysozyme R	WPJ71039.1	64196 - 64690	164	PRINTS	PR00684	Phage T4 lysozyme signature	95	114	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB26 OR539218	lysozyme R	WPJ71039.1	64196 - 64690	164	PRINTS	PR00684	Phage T4 lysozyme signature	72	92	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB26 OR539218	lysozyme R	WPJ71039.1	64196 - 64690	164	PRINTS	PR00684	Phage T4 lysozyme signature	4	23	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB26 OR539218	lysozyme R	WPJ71039.1	64196 - 64690	164	PRINTS	PR00684	Phage T4 lysozyme signature	24	42	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB26 OR539218	lysozyme R	WPJ71039.1	64196 - 64690	164	PRINTS	PR00684	Phage T4 lysozyme signature	118	137	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB26 OR539218	lysozyme R	WPJ71039.1	64196 - 64690	164	PRINTS	PR00684	Phage T4 lysozyme signature	51	70	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB26 OR539218	lysozyme R	WPJ71039.1	64196 - 64690	164	PRINTS	PR00684	Phage T4 lysozyme signature	138	159	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB26 OR539218	lysozyme R	WPJ71039.1	64196 - 64690	164	CDD	cd00735	T4-like lys	2	160	6.07455E-78	-	-	-

vB-Eco-KMB26 OR539218	lysozyme R	WPJ71039.1	64196 - 64690	164	PANTHER	PTHR37406	T4-TYPE LYSOZYME 1-RELATED	4	161	1.5E-27	IPR052619	Bacteriophage lysozyme-like	-
vB-Eco-KMB39 OR539219	lysozyme R	WPJ71303.1	61730 - 62224	164	Gene3D	G3DSA:1.10.530.40	-	1	164	6.6E-75	IPR023347	Lysozyme domain superfamily	GO:0003796(InterPro)
vB-Eco-KMB39 OR539219	lysozyme R	WPJ71303.1	61730 - 62224	164	FunFam	G3DSA:1.10.530.40:FF:000002	Endolysin	1	164	1.1E-114	-	-	-
vB-Eco-KMB39 OR539219	lysozyme R	WPJ71303.1	61730 - 62224	164	Hamap	MF 04110	Endolysin.	1	159	16.677807	IPR034690	Endolysin T4 type	GO:0003796(InterPro)|GO:0009253(InterPro)
vB-Eco-KMB39 OR539219	lysozyme R	WPJ71303.1	61730 - 62224	164	SUPERFAMILY	SSF53955	Lysozyme-like	1	161	4.81E-54	IPR023346	Lysozyme-like domain superfamily	-
vB-Eco-KMB39 OR539219	lysozyme R	WPJ71303.1	61730 - 62224	164	Pfam	PF00959	Phage lysozyme	11	141	4.4E-10	IPR002196	Glycoside hydrolase, family 24	GO:0003796(InterPro)|GO:0009253(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB39 OR539219	lysozyme R	WPJ71303.1	61730 - 62224	164	PRINTS	PR00684	Phage T4 lysozyme signature	95	114	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB39 OR539219	lysozyme R	WPJ71303.1	61730 - 62224	164	PRINTS	PR00684	Phage T4 lysozyme signature	72	92	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB39 OR539219	lysozyme R	WPJ71303.1	61730 - 62224	164	PRINTS	PR00684	Phage T4 lysozyme signature	4	23	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB39 OR539219	lysozyme R	WPJ71303.1	61730 - 62224	164	PRINTS	PR00684	Phage T4 lysozyme signature	24	42	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB39 OR539219	lysozyme R	WPJ71303.1	61730 - 62224	164	PRINTS	PR00684	Phage T4 lysozyme signature	118	137	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB39 OR539219	lysozyme R	WPJ71303.1	61730 - 62224	164	PRINTS	PR00684	Phage T4 lysozyme signature	51	70	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB39 OR539219	lysozyme R	WPJ71303.1	61730 - 62224	164	PRINTS	PR00684	Phage T4 lysozyme signature	138	159	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB39 OR539219	lysozyme R	WPJ71303.1	61730 - 62224	164	CDD	cd00735	T4-like lys	2	160	6.07455E-78	-	-	-
vB-Eco-KMB39 OR539219	lysozyme R	WPJ71303.1	61730 - 62224	164	PANTHER	PTHR37406	T4-TYPE LYSOZYME 1-RELATED	4	161	1.5E-27	IPR052619	Bacteriophage lysozyme-like	-
vB-Eco-KMB40 OR539220	lysozyme R	WPJ71570.1	62190 - 62684	164	Gene3D	G3DSA:1.10.530.40	-	1	164	6.6E-75	IPR023347	Lysozyme domain superfamily	GO:0003796(InterPro)
vB-Eco-KMB40 OR539220	lysozyme R	WPJ71570.1	62190 - 62684	164	FunFam	G3DSA:1.10.530.40:FF:000002	Endolysin	1	164	1.1E-114	-	-	-
vB-Eco-KMB40 OR539220	lysozyme R	WPJ71570.1	62190 - 62684	164	Hamap	MF 04110	Endolysin.	1	159	16.677807	IPR034690	Endolysin T4 type	GO:0003796(InterPro)|GO:0009253(InterPro)
vB-Eco-KMB40 OR539220	lysozyme R	WPJ71570.1	62190 - 62684	164	SUPERFAMILY	SSF53955	Lysozyme-like	1	161	4.81E-54	IPR023346	Lysozyme-like domain superfamily	-

vB-Eco-KMB40 OR539220	lysozyme R	WPJ71570.1	62190 - 62684	164	Pfam	PF00959	Phage lysozyme	11	141	4.4E-10	IPR002196	Glycoside hydrolase, family 24	GO:0003796(InterPro)|GO:0009253(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB40 OR539220	lysozyme R	WPJ71570.1	62190 - 62684	164	PRINTS	PR00684	Phage T4 lysozyme signature	95	114	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB40 OR539220	lysozyme R	WPJ71570.1	62190 - 62684	164	PRINTS	PR00684	Phage T4 lysozyme signature	72	92	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB40 OR539220	lysozyme R	WPJ71570.1	62190 - 62684	164	PRINTS	PR00684	Phage T4 lysozyme signature	4	23	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB40 OR539220	lysozyme R	WPJ71570.1	62190 - 62684	164	PRINTS	PR00684	Phage T4 lysozyme signature	24	42	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB40 OR539220	lysozyme R	WPJ71570.1	62190 - 62684	164	PRINTS	PR00684	Phage T4 lysozyme signature	118	137	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB40 OR539220	lysozyme R	WPJ71570.1	62190 - 62684	164	PRINTS	PR00684	Phage T4 lysozyme signature	51	70	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB40 OR539220	lysozyme R	WPJ71570.1	62190 - 62684	164	PRINTS	PR00684	Phage T4 lysozyme signature	138	159	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB40 OR539220	lysozyme R	WPJ71570.1	62190 - 62684	164	CDD	cd00735	T4-like lys	2	160	6.07455E-78	-	-	-
vB-Eco-KMB40 OR539220	lysozyme R	WPJ71570.1	62190 - 62684	164	PANTHER	PTHR37406	T4-TYPE LYSOZYME 1-RELATED	4	161	1.5E-27	IPR052619	Bacteriophage lysozyme-like	-
vB-Eco-KMB42 OR539221	lysozyme R	WPJ71835.1	62848 - 63342	164	Gene3D	G3DSA:1.10.530.40	-	1	164	6.6E-75	IPR023347	Lysozyme domain superfamily	GO:0003796(InterPro)
vB-Eco-KMB42 OR539221	lysozyme R	WPJ71835.1	62848 - 63342	164	FunFam	G3DSA:1.10.530.40:FF:000002	Endolysin	1	164	1.1E-114	-	-	-
vB-Eco-KMB42 OR539221	lysozyme R	WPJ71835.1	62848 - 63342	164	Hamap	MF 04110	Endolysin.	1	159	16.677807	IPR034690	Endolysin T4 type	GO:0003796(InterPro)|GO:0009253(InterPro)
vB-Eco-KMB42 OR539221	lysozyme R	WPJ71835.1	62848 - 63342	164	SUPERFAMILY	SSF53955	Lysozyme-like	1	161	4.81E-54	IPR023346	Lysozyme-like domain superfamily	-
vB-Eco-KMB42 OR539221	lysozyme R	WPJ71835.1	62848 - 63342	164	Pfam	PF00959	Phage lysozyme	11	141	4.4E-10	IPR002196	Glycoside hydrolase, family 24	GO:0003796(InterPro)|GO:0009253(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB42 OR539221	lysozyme R	WPJ71835.1	62848 - 63342	164	PRINTS	PR00684	Phage T4 lysozyme signature	95	114	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB42 OR539221	lysozyme R	WPJ71835.1	62848 - 63342	164	PRINTS	PR00684	Phage T4 lysozyme signature	72	92	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB42 OR539221	lysozyme R	WPJ71835.1	62848 - 63342	164	PRINTS	PR00684	Phage T4 lysozyme signature	4	23	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB42 OR539221	lysozyme R	WPJ71835.1	62848 - 63342	164	PRINTS	PR00684	Phage T4 lysozyme signature	24	42	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB42 OR539221	lysozyme R	WPJ71835.1	62848 - 63342	164	PRINTS	PR00684	Phage T4 lysozyme signature	118	137	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)

vB-Eco-KMB42 OR539221	lysozyme R	WPJ71835.1	62848 - 63342	164	PRINTS	PR00684	Phage T4 lysozyme signature	51	70	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB42 OR539221	lysozyme R	WPJ71835.1	62848 - 63342	164	PRINTS	PR00684	Phage T4 lysozyme signature	138	159	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB42 OR539221	lysozyme R	WPJ71835.1	62848 - 63342	164	CDD	cd00735	T4-like lys	2	160	6.07455E-78	-	-	-
vB-Eco-KMB42 OR539221	lysozyme R	WPJ71835.1	62848 - 63342	164	PANTHER	PTHR37406	T4-TYPE LYSOZYME 1-RELATED	4	161	1.5E-27	IPR052619	Bacteriophage lysozyme-like	-
vB-Eco-KMB47 OR424385	hypothetical protein	WPK27967.1	38771 - 39115	114	Gene3D	G3DSA:3.30.1380.10	-	5	113	3.5E-26	IPR009045	Hedgehog signalling/DD-peptidase zinc-binding domain superfamily	-
vB-Eco-KMB47 OR424385	hypothetical protein	WPK27967.1	38771 - 39115	114	SUPERFAMILY	SSF55166	Hedgehog/DD-peptidase	5	110	6.28E-25	IPR009045	Hedgehog signalling/DD-peptidase zinc-binding domain superfamily	-
vB-Eco-KMB47 OR424385	hypothetical protein	WPK27967.1	38771 - 39115	114	Pfam	PF08291	Peptidase M15	4	106	1.4E-29	IPR013230	Peptidase M15A, C-terminal	-
vB-Eco-KMB25 OR424386	lysozyme or muramidase	WPK28014.1	33375 - 33869	164	Gene3D	G3DSA:1.10.530.40	-	1	161	4.2E-45	IPR023347	Lysozyme domain superfamily	GO:0003796(InterPro)
vB-Eco-KMB25 OR424386	lysozyme or muramidase	WPK28014.1	33375 - 33869	164	PANTHER	PTHR38107	-	18	162	5.1E-37	IPR051018	Bacteriophage Glycosyl Hydrolase 24	GO:0003796(PANTHER)
vB-Eco-KMB25 OR424386	lysozyme or muramidase	WPK28014.1	33375 - 33869	164	Phobius	SIGNAL PEPTIDE C REGION	C-terminal region of a signal peptide.	17	21	-	-	-	-
vB-Eco-KMB25 OR424386	lysozyme or muramidase	WPK28014.1	33375 - 33869	164	CDD	cd16900	endolysin R21-like	13	158	2.12923E-65	-	-	-
vB-Eco-KMB25 OR424386	lysozyme or muramidase	WPK28014.1	33375 - 33869	164	Hamap	MF 04110	Endolysin.	5	161	26.81488	IPR034690	Endolysin T4 type	GO:0003796(InterPro)|GO:0009253(InterPro)
vB-Eco-KMB25 OR424386	lysozyme or muramidase	WPK28014.1	33375 - 33869	164	Phobius	SIGNAL PEPTIDE H REGION	Hydrophobic region of a signal peptide.	6	16	-	-	-	-
vB-Eco-KMB25 OR424386	lysozyme or muramidase	WPK28014.1	33375 - 33869	164	SUPERFAMILY	SSF53955	Lysozyme-like	12	162	1.33E-46	IPR023346	Lysozyme-like domain superfamily	-
vB-Eco-KMB25 OR424386	lysozyme or muramidase	WPK28014.1	33375 - 33869	164	Phobius	SIGNAL PEPTIDE N REGION	N-terminal region of a signal peptide.	1	5	-	-	-	-
vB-Eco-KMB25 OR424386	lysozyme or muramidase	WPK28014.1	33375 - 33869	164	Pfam	PF00959	Phage lysozyme	25	154	6.2E-32	IPR002196	Glycoside hydrolase, family 24	GO:0003796(InterPro)|GO:0009253(InterPro)|GO:0016998(InterPro)

vB-Eco-KMB25 OR424386	lysozyme or muramidase	WPK28014.1	33375 - 33869	164	Phobius	NON CYTOPLASMIC DOMAIN	Region of a membrane-bound protein predicted to be outside the membrane, in the extracellular region.	22	164	-	-	-	-
vB-Eco-KMB25 OR424386	lysozyme or muramidase	WPK28014.1	33375 - 33869	164	Hamap	MF 04136	SAR-endolysin.	4	161	30.038778	IPR043688	SAR-endolysin-like	GO:0003796(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB25 OR424386	lysozyme or muramidase	WPK28014.1	33375 - 33869	164	Phobius	SIGNAL PEPTIDE	Signal peptide region	1	21	-	-	-	-
vB-Eco-KMB46 OR525699	lysozyme R	WQN06679.1	44328 - 44819	163	PANTHER	PTHR38107	-	19	160	1.2E-39	IPR051018	Bacteriophage Glycosyl Hydrolase 24	GO:0003796(PANTHER)
vB-Eco-KMB46 OR525699	lysozyme R	WQN06679.1	44328 - 44819	163	Gene3D	G3DSA:1.10.530.40	-	1	160	2.1E-47	IPR023347	Lysozyme domain superfamily	GO:0003796(InterPro)
vB-Eco-KMB46 OR525699	lysozyme R	WQN06679.1	44328 - 44819	163	Phobius	SIGNAL PEPTIDE C REGION	C-terminal region of a signal peptide.	19	23	-	-	-	-
vB-Eco-KMB46 OR525699	lysozyme R	WQN06679.1	44328 - 44819	163	CDD	cd16900	endolysin R21-like	19	158	1.14887E-70	-	-	-
vB-Eco-KMB46 OR525699	lysozyme R	WQN06679.1	44328 - 44819	163	Hamap	MF 04110	Endolysin.	5	161	28.505993	IPR034690	Endolysin T4 type	GO:0003796(InterPro)|GO:0009253(InterPro)
vB-Eco-KMB46 OR525699	lysozyme R	WQN06679.1	44328 - 44819	163	Phobius	SIGNAL PEPTIDE H REGION	Hydrophobic region of a signal peptide.	7	18	-	-	-	-
vB-Eco-KMB46 OR525699	lysozyme R	WQN06679.1	44328 - 44819	163	SUPERFAMILY	SSF53955	Lysozyme-like	15	161	7.16E-48	IPR023346	Lysozyme-like domain superfamily	-
vB-Eco-KMB46 OR525699	lysozyme R	WQN06679.1	44328 - 44819	163	Phobius	SIGNAL PEPTIDE N REGION	N-terminal region of a signal peptide.	1	6	-	-	-	-
vB-Eco-KMB46 OR525699	lysozyme R	WQN06679.1	44328 - 44819	163	Pfam	PF00959	Phage lysozyme	25	154	1.1E-35	IPR002196	Glycoside hydrolase, family 24	GO:0003796(InterPro)|GO:0009253(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB46 OR525699	lysozyme R	WQN06679.1	44328 - 44819	163	Phobius	NON CYTOPLASMIC DOMAIN	Region of a membrane-bound protein predicted to be outside the membrane, in the extracellular region.	24	163	-	-	-	-
vB-Eco-KMB46 OR525699	lysozyme R	WQN06679.1	44328 - 44819	163	Hamap	MF 04136	SAR-endolysin.	4	161	30.766636	IPR043688	SAR-endolysin-like	GO:0003796(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB46 OR525699	lysozyme R	WQN06679.1	44328 - 44819	163	Phobius	SIGNAL PEPTIDE	Signal peptide region	1	23	-	-	-	-

vB-Eco-KMB37 OR525700	lysozyme murein hydrolase	WQN06754.1	22896 - 23396	166	Gene3D	G3DSA:1.10.530.40	-	5	166	1.5E-57	IPR023347	Lysozyme domain superfamily	GO:0003796(InterPro)
vB-Eco-KMB37 OR525700	lysozyme murein hydrolase	WQN06754.1	22896 - 23396	166	SUPERFAMILY	SSF53955	Lysozyme-like	7	166	2.81E-47	IPR023346	Lysozyme-like domain superfamily	-
vB-Eco-KMB37 OR525700	lysozyme murein hydrolase	WQN06754.1	22896 - 23396	166	Pfam	PF00959	Phage lysozyme	15	142	1.9E-8	IPR002196	Glycoside hydrolase, family 24	GO:0003796(InterPro)|GO:0009253(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB37 OR525700	lysozyme murein hydrolase	WQN06754.1	22896 - 23396	166	PRINTS	PR00684	Phage T4 lysozyme signature	122	141	1.5E-45	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB37 OR525700	lysozyme murein hydrolase	WQN06754.1	22896 - 23396	166	PRINTS	PR00684	Phage T4 lysozyme signature	76	96	1.5E-45	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB37 OR525700	lysozyme murein hydrolase	WQN06754.1	22896 - 23396	166	PRINTS	PR00684	Phage T4 lysozyme signature	55	74	1.5E-45	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB37 OR525700	lysozyme murein hydrolase	WQN06754.1	22896 - 23396	166	PRINTS	PR00684	Phage T4 lysozyme signature	8	27	1.5E-45	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB37 OR525700	lysozyme murein hydrolase	WQN06754.1	22896 - 23396	166	PRINTS	PR00684	Phage T4 lysozyme signature	99	118	1.5E-45	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB37 OR525700	lysozyme murein hydrolase	WQN06754.1	22896 - 23396	166	PRINTS	PR00684	Phage T4 lysozyme signature	28	46	1.5E-45	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB37 OR525700	lysozyme murein hydrolase	WQN06754.1	22896 - 23396	166	PRINTS	PR00684	Phage T4 lysozyme signature	142	163	1.5E-45	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB37 OR525700	lysozyme murein hydrolase	WQN06754.1	22896 - 23396	166	CDD	cd00735	T4-like lys	7	164	8.82227E-56	-	-	-
vB-Eco-KMB37 OR525700	lysozyme murein hydrolase	WQN06754.1	22896 - 23396	166	PANTHER	PTHR37406	T4-TYPE LYSOZYME 1-RELATED	4	165	2.3E-28	IPR052619	Bacteriophage lysozyme-like	-
vB-Eco-KMB38 OR525701	lysozyme R	WQN07107.1	66043 - 66531	162	Gene3D	G3DSA:1.10.530.40	-	2	162	2.4E-89	IPR023347	Lysozyme domain superfamily	GO:0003796(InterPro)
vB-Eco-KMB38 OR525701	lysozyme R	WQN07107.1	66043 - 66531	162	SUPERFAMILY	SSF53955	Lysozyme-like	1	161	1.28E-52	IPR023346	Lysozyme-like domain superfamily	-
vB-Eco-KMB38 OR525701	lysozyme R	WQN07107.1	66043 - 66531	162	Pfam	PF00959	Phage lysozyme	10	140	3.6E-10	IPR002196	Glycoside hydrolase, family 24	GO:0003796(InterPro)|GO:0009253(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB38 OR525701	lysozyme R	WQN07107.1	66043 - 66531	162	PRINTS	PR00684	Phage T4 lysozyme signature	4	23	6.4E-76	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB38 OR525701	lysozyme R	WQN07107.1	66043 - 66531	162	PRINTS	PR00684	Phage T4 lysozyme signature	118	137	6.4E-76	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB38 OR525701	lysozyme R	WQN07107.1	66043 - 66531	162	PRINTS	PR00684	Phage T4 lysozyme signature	138	159	6.4E-76	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB38 OR525701	lysozyme R	WQN07107.1	66043 - 66531	162	PRINTS	PR00684	Phage T4 lysozyme signature	24	42	6.4E-76	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)

vB-Eco-KMB38 OR525701	lysozyme R	WQN07107.1	66043 - 66531	162	PRINTS	PR00684	Phage T4 lysozyme signature	72	92	6.4E-76	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB38 OR525701	lysozyme R	WQN07107.1	66043 - 66531	162	PRINTS	PR00684	Phage T4 lysozyme signature	95	114	6.4E-76	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB38 OR525701	lysozyme R	WQN07107.1	66043 - 66531	162	PRINTS	PR00684	Phage T4 lysozyme signature	51	70	6.4E-76	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB38 OR525701	lysozyme R	WQN07107.1	66043 - 66531	162	CDD	cd00735	T4-like lys	2	160	1.18768E-72	-	-	-
vB-Eco-KMB38 OR525701	lysozyme R	WQN07107.1	66043 - 66531	162	PANTHER	PTHR37406	T4-TYPE LYSOZYME 1-RELATED	4	160	3.3E-26	IPR052619	Bacteriophage lysozyme-like	-
vB-Eco-KMB23 OR525702	lysozyme R	WQN07383.1	63387 - 63881	164	Gene3D	G3DSA:1.10.530.40	-	1	164	6.6E-75	IPR023347	Lysozyme domain superfamily	GO:0003796(InterPro)
vB-Eco-KMB23 OR525702	lysozyme R	WQN07383.1	63387 - 63881	164	FunFam	G3DSA:1.10.530.40:FF:000002	Endolysin	1	164	1.1E-114	-	-	-
vB-Eco-KMB23 OR525702	lysozyme R	WQN07383.1	63387 - 63881	164	Hamap	MF 04110	Endolysin.	1	159	16.677807	IPR034690	Endolysin T4 type	GO:0003796(InterPro)|GO:0009253(InterPro)
vB-Eco-KMB23 OR525702	lysozyme R	WQN07383.1	63387 - 63881	164	SUPERFAMILY	SSF53955	Lysozyme-like	1	161	4.81E-54	IPR023346	Lysozyme-like domain superfamily	-
vB-Eco-KMB23 OR525702	lysozyme R	WQN07383.1	63387 - 63881	164	Pfam	PF00959	Phage lysozyme	11	141	4.4E-10	IPR002196	Glycoside hydrolase, family 24	GO:0003796(InterPro)|GO:0009253(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB23 OR525702	lysozyme R	WQN07383.1	63387 - 63881	164	PRINTS	PR00684	Phage T4 lysozyme signature	95	114	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB23 OR525702	lysozyme R	WQN07383.1	63387 - 63881	164	PRINTS	PR00684	Phage T4 lysozyme signature	72	92	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB23 OR525702	lysozyme R	WQN07383.1	63387 - 63881	164	PRINTS	PR00684	Phage T4 lysozyme signature	4	23	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB23 OR525702	lysozyme R	WQN07383.1	63387 - 63881	164	PRINTS	PR00684	Phage T4 lysozyme signature	24	42	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB23 OR525702	lysozyme R	WQN07383.1	63387 - 63881	164	PRINTS	PR00684	Phage T4 lysozyme signature	118	137	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB23 OR525702	lysozyme R	WQN07383.1	63387 - 63881	164	PRINTS	PR00684	Phage T4 lysozyme signature	51	70	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB23 OR525702	lysozyme R	WQN07383.1	63387 - 63881	164	PRINTS	PR00684	Phage T4 lysozyme signature	138	159	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB23 OR525702	lysozyme R	WQN07383.1	63387 - 63881	164	CDD	cd00735	T4-like lys	2	160	6.07455E-78	-	-	-
vB-Eco-KMB23 OR525702	lysozyme R	WQN07383.1	63387 - 63881	164	PANTHER	PTHR37406	T4-TYPE LYSOZYME 1-RELATED	4	161	1.5E-27	IPR052619	Bacteriophage lysozyme-like	-
vB-Eco-KMB36 OR525703	lysin	WQN07602.1	41598 - 42101	167	Gene3D	G3DSA:1.10.530.40	-	4	152	1.2E-30	IPR023347	Lysozyme domain superfamily	GO:0003796(InterPro)

vB-Eco-KMB36 OR525703	lysin	WQN07602.1	41598 - 42101	167	PANTHER	PTHR38107	-	7	149	1.9E-20	IPR051018	Bacteriophage Glycosyl Hydrolase 24	-
vB-Eco-KMB36 OR525703	lysin	WQN07602.1	41598 - 42101	167	SUPERFAMILY	SSF53955	Lysozyme-like	7	154	1.64E-30	IPR023346	Lysozyme-like domain superfamily	-
vB-Eco-KMB36 OR525703	lysin	WQN07602.1	41598 - 42101	167	CDD	cd00737	lyz endolysin autolysin	14	151	1.4534E-29	IPR033907	Endolysin/autolysin	GO:0003796(InterPro)
vB-Eco-KMB36 OR525703	lysin	WQN07602.1	41598 - 42101	167	Pfam	PF00959	Phage lysozyme	20	143	1.0E-22	IPR002196	Glycoside hydrolase, family 24	GO:0003796(InterPro)|GO:0009253(InterPro)|GO:0016998(InterPro)
vB-Eco-KMB35 OR525704	lysin	WQN07684.1	40675 - 41163	162	Gene3D	G3DSA:1.10.530.40	-	2	147	9.0E-32	IPR023347	Lysozyme domain superfamily	GO:0003796(InterPro)
vB-Eco-KMB35 OR525704	lysin	WQN07684.1	40675 - 41163	162	PANTHER	PTHR38107	-	3	144	3.7E-22	IPR051018	Bacteriophage Glycosyl Hydrolase 24	GO:0003796(PANTHER)
vB-Eco-KMB35 OR525704	lysin	WQN07684.1	40675 - 41163	162	SUPERFAMILY	SSF53955	Lysozyme-like	3	149	3.22E-31	IPR023346	Lysozyme-like domain superfamily	-
vB-Eco-KMB35 OR525704	lysin	WQN07684.1	40675 - 41163	162	CDD	cd00737	lyz endolysin autolysin	9	146	7.59506E-31	IPR033907	Endolysin/autolysin	GO:0003796(InterPro)
vB-Eco-KMB35 OR525704	lysin	WQN07684.1	40675 - 41163	162	Pfam	PF00959	Phage lysozyme	15	138	4.1E-23	IPR002196	Glycoside hydrolase, family 24	GO:0003796(InterPro)|GO:0009253(InterPro)|GO:0016998(InterPro)
vB EcoM ACG-C40 NC 019399	glycoside hydrolase family protein	YP 006986678.1	63339 - 63833	164	Gene3D	G3DSA:1.10.530.40	-	1	164	6.6E-75	IPR023347	Lysozyme domain superfamily	GO:0003796(InterPro)
vB EcoM ACG-C40 NC 019399	glycoside hydrolase family protein	YP 006986678.1	63339 - 63833	164	FunFam	G3DSA:1.10.530.40:FF:000002	Endolysin	1	164	1.1E-114	-	-	-
vB EcoM ACG-C40 NC 019399	glycoside hydrolase family protein	YP 006986678.1	63339 - 63833	164	Hamap	MF 04110	Endolysin.	1	159	16.677807	IPR034690	Endolysin T4 type	GO:0003796(InterPro)|GO:0009253(InterPro)
vB EcoM ACG-C40 NC 019399	glycoside hydrolase family protein	YP 006986678.1	63339 - 63833	164	SUPERFAMILY	SSF53955	Lysozyme-like	1	161	4.81E-54	IPR023346	Lysozyme-like domain superfamily	-
vB EcoM ACG-C40 NC 019399	glycoside hydrolase family protein	YP 006986678.1	63339 - 63833	164	Pfam	PF00959	Phage lysozyme	11	141	4.4E-10	IPR002196	Glycoside hydrolase, family 24	GO:0003796(InterPro)|GO:0009253(InterPro)|GO:0016998(InterPro)
vB EcoM ACG-C40 NC 019399	glycoside hydrolase family protein	YP 006986678.1	63339 - 63833	164	PRINTS	PR00684	Phage T4 lysozyme signature	95	114	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM ACG-C40 NC 019399	glycoside hydrolase family protein	YP 006986678.1	63339 - 63833	164	PRINTS	PR00684	Phage T4 lysozyme signature	72	92	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM ACG-C40 NC 019399	glycoside hydrolase family protein	YP 006986678.1	63339 - 63833	164	PRINTS	PR00684	Phage T4 lysozyme signature	4	23	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM ACG-C40 NC 019399	glycoside hydrolase family protein	YP 006986678.1	63339 - 63833	164	PRINTS	PR00684	Phage T4 lysozyme signature	24	42	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)

vB EcoM ACG-C40 NC 019399	glycoside hydrolase family protein	YP 006986678.1	63339 - 63833	164	PRINTS	PR00684	Phage T4 lysozyme signature	118	137	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM ACG-C40 NC 019399	glycoside hydrolase family protein	YP 006986678.1	63339 - 63833	164	PRINTS	PR00684	Phage T4 lysozyme signature	51	70	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM ACG-C40 NC 019399	glycoside hydrolase family protein	YP 006986678.1	63339 - 63833	164	PRINTS	PR00684	Phage T4 lysozyme signature	138	159	2.7E-91	IPR001165	T4-type lysozyme	GO:0003796(InterPro)|GO:0016998(InterPro)
vB EcoM ACG-C40 NC 019399	glycoside hydrolase family protein	YP 006986678.1	63339 - 63833	164	CDD	cd00735	T4-like lys	2	160	6.07455E-78	-	-	-
vB EcoM ACG-C40 NC 019399	glycoside hydrolase family protein	YP 006986678.1	63339 - 63833	164	PANTHER	PTHR37406	T4-TYPE LYSOZYME 1-RELATED	4	161	1.5E-27	IPR052619	Bacteriophage lysozyme-like	-
EG1 NC 047895	amidase	YP 009795817.1	11177 - 11632	151	Gene3D	G3DSA:3.40.80.10	-	6	151	1.2E-52	IPR036505	N-acetylmuramoyl-L-alanine amidase/PGRP domain superfamily	GO:0008745(InterPro)|GO:0009253(InterPro)
EG1 NC 047895	amidase	YP 009795817.1	11177 - 11632	151	SMART	SM00644	ami 2	1	151	2.9E-27	IPR002502	N-acetylmuramoyl-L-alanine amidase domain	GO:0008745(InterPro)|GO:0009253(InterPro)
EG1 NC 047895	amidase	YP 009795817.1	11177 - 11632	151	FunFam	G3DSA:3.40.80.10:FF:000009	Endolysin	6	151	5.0E-99	-	-	-
EG1 NC 047895	amidase	YP 009795817.1	11177 - 11632	151	Hamap	MF 04111	Endolysin.	1	149	50.983433	IPR034689	Endolysin T7 type	GO:0008745(InterPro)|GO:0009253(InterPro)|GO:0032897(InterPro)|GO:0044659(InterPro)
EG1 NC 047895	amidase	YP 009795817.1	11177 - 11632	151	Pfam	PF01510	N-acetylmuramoyl-L-alanine amidase	11	132	8.1E-14	IPR002502	N-acetylmuramoyl-L-alanine amidase domain	GO:0008745(InterPro)|GO:0009253(InterPro)
EG1 NC 047895	amidase	YP 009795817.1	11177 - 11632	151	SUPERFAMILY	SSF55846	N-acetylmuramoyl-L-alanine amidase-like	10	146	1.26E-44	IPR036505	N-acetylmuramoyl-L-alanine amidase/PGRP domain superfamily	GO:0008745(InterPro)|GO:0009253(InterPro)
EG1 NC 047895	amidase	YP 009795817.1	11177 - 11632	151	PANTHER	PTHR11022	PEPTIDOGLYCAN RECOGNITION PROTEIN	13	133	2.4E-16	IPR015510	Peptidoglycan recognition protein	-

EG1 NC 047895	amidase	YP 009795817.1	11177 - 11632	151	CDD	cd06583	PGRP	13	132	9.23323E-28	IPR002502	N-acetylmuramoyl-L-alanine amidase domain	GO:0008745(InterPro)|GO:0009253(InterPro)
EG1 NC 047895	amidase	YP 009795817.1	11177 - 11632	151	SMART	SM00701	pgrp	1	124	3.5E-32	IPR006619	Peptidoglycan recognition protein family domain, metazoa/bacteria	GO:0008270(InterPro)|GO:0008745(InterPro)|GO:0009253(InterPro)

Supplementary Table S5. The list of AMP peptides found within sequences of the proteins identified by phylogenetic analysis as closely related to the tested endolysins.

Phage name	Protein name	Protein accession number	Position	Sequence	Class	AMP Probability
VB-ECOM-KMB26-OW536688	t4-like-phage-baseplate-hub-+-tail-lysozyme	CAH2619090.1	41-60	VAKSELDKAIGRNCNGVITK	AMP	0.858
VB-ECOS-22664B1-MZ234008	lysin	QZI78369.1	90-109	INAGQGPIVPDNIANDFDDA	AMP	0.978
VB-ECOS-22664BS2-MZ234010	lysin	QZI78493.1	118-137	AKLALFINQNGKPILGLRRR	AMP	0.983
VB-ECOP-101101UKE1-MZ234012	hypothetical-protein	QZI78680.1	50-69	AKHNANVGGAKNSMHLTGKA	AMP	0.892
VB-ECOM-101112UKE3-1-MZ234013	lysozyme-r	QZI78958.1	41-60	VAKSELDKAIGRNCNGVITK	AMP	0.858
VB-ECOS-101114B2-MZ234014	lysin	QZI79000.1	88-107	LVINAGQGPIVPDNIANDFD	AMP	0.978
VB-ECOS-101114BS4-MZ234016	lysin	QZI79065.1	90-109	INAGQGPIVPDNIANDFDDA	AMP	0.978
VB-ECOM-101117BS1-MZ234018	lysozyme-r	QZI79532.1	41-60	VAKSELDKAIGRNCNGVITK	AMP	0.858
VB-ECOP-101117UKE2-MZ234019	hypothetical-protein	QZI79610.1	59-78	AKHNANVGGAKNSMHLTGKA	AMP	0.892
VB-ECOP-101118B1-MZ234020	hypothetical-protein	QZI79670.1	59-78	AKHNANVGGAKNSMHLTGKA	AMP	0.892
VB-ECOM-172859UKE1-MZ234025	lysozyme-r	QZI80176.1	25-44	WTIGIGHLVTKNPSLAVAKA	AMP	0.942
VB-ECOM-CHD2BS1-MZ234027	lysozyme-r	QZI80464.1	41-60	VAKSELDKAIGRNCNGVITK	AMP	0.858
VB-ECOM-CHD16UKE1-MZ234030	lysozyme-r	QZI80907.1	41-60	VAKSELDKAIGRNCNGVITK	AMP	0.858
VB-ECOM-CHD94UKE2-MZ234031	lysozyme-r	QZI81192.1	41-60	VAKSELDKAIGRNCNGVITK	AMP	0.858
VB-ECOM-G3G7-MZ234040	lysozyme-r	QZI83386.1	41-60	VAKSELDKAIGRNCNGVITK	AMP	0.858
VB-ECOP-UTI89UKE2-MZ234049	hypothetical-protein	QZI84581.1	50-69	AKHNANVGGAKNSMHLTGKA	AMP	0.892
VB-ECOP-UTI89UKE3-MZ234050	hypothetical-protein	QZI84648.1	50-69	AKHNANVGGAKNSMHLTGKA	AMP	0.892
VB-ECOM-SHAK7854-OR594182	amidase	SELF-DNYN29	68-87	GSHVKDWNSRSVGVCLVGGI	AMP	0.840
VB-ECOP-LNA3-MZ311862	amidase	SELF-GXIJ34	68-87	GSHVKDWNSRSVGVCLVGGV	AMP	0.805
VB-ECOM-LNA2-MZ311865	hypothetical-protein	SELF-NLWW35	25-44	WTIGIGHLVTKNPSLAVAKA	AMP	0.942
VB-ECOP-LNA5-MZ311861	amidase	SELF-TCEH12	68-87	GSHVKDWNSRSVGVCLVGGI	AMP	0.840
VB-ECOM-SHAK9454-OR594184	amidase	SELF-ZENB29	68-87	GSHVKDWNSRSVGVCLVGGI	AMP	0.840
MLP2-OK148441	baseplate-hub-+-tail-lysozyme	UEN68765.1	66-85	LFNQDVDAAVRGILRNAKLK	AMP	0.843
VB-ECOS-CHD2B1-MZ234026	lysin	USL86318.1	90-109	INAGQGPIVPDNIANDFDDA	AMP	0.978
VB-ECOS-CHD5UKE2-MZ234029	lysin	USL86352.1	85-104	LVDLVINAGQGPIVPDHIAN	AMP	0.871
VB-ECOS-UTI89UKE1-MZ234048	lysin	USM81274.1	118-137	AKLALFINQNGKPLLGLRRR	AMP	0.980
ZCEC14-OP168898	hypothetical-protein	UVD33441.1	39-58	LNVAKSELDKAIGRNCNGVI	AMP	0.903
VB-ECO-KMB22-OR539217	lysozyme-r	WPJ70775.1	41-60	VAKSELDKAIGRNCNGIITK	AMP	0.892
VB-ECO-KMB26-OR539218	lysozyme-r	WPJ71039.1	41-60	VAKSELDKAIGRNCNGVITK	AMP	0.858
VB-ECO-KMB39-OR539219	lysozyme-r	WPJ71303.1	41-60	VAKSELDKAIGRNCNGVITK	AMP	0.858
VB-ECO-KMB40-OR539220	lysozyme-r	WPJ71570.1	41-60	VAKSELDKAIGRNCNGVITK	AMP	0.858
VB-ECO-KMB42-OR539221	lysozyme-r	WPJ71835.1	41-60	VAKSELDKAIGRNCNGVITK	AMP	0.858
VB-ECO-KMB47-OR424385	hypothetical-protein	WPK27967.1	50-69	AKHNANVGGAKNSMHLTGKA	AMP	0.892
VB-ECO-KMB25-OR424386	lysozyme-or-muramidase	WPK28014.1	134-153	RNPKTGKKEKSRGLKNRRAV	AMP	0.946
VB-ECO-KMB46-OR525699	lysozyme-r	WQN06679.1	135-154	NPKTGKREKSKGLLNRRKVE	AMP	0.970
VB-ECO-KMB37-OR525700	lysozyme-murein-hydrolase	WQN06754.1	31-50	VGIGHLLTKLKDKAEAIRIL	AMP	0.934
VB-ECO-KMB38-OR525701	lysozyme-r	WQN07107.1	25-44	WTIGIGHLVTKNPSLAVAKA	AMP	0.942
VB-ECO-KMB23-OR525702	lysozyme-r	WQN07383.1	41-60	VAKSELDKAIGRNCNGVITK	AMP	0.858
VB-ECO-KMB36-OR525703	lysin	WQN07602.1	128-147	FINQNGKPLLGLRRRTAGRL	AMP	0.947
VB-ECO-KMB35-OR525704	lysin	WQN07684.1	118-137	AKLALFINQNGKPLLGLRRR	AMP	0.980
VB-ECOM-ACG-C40-NC-019399	glycoside-hydrolase-family-protein	YP-006986678.1	41-60	VAKSELDKAIGRNCNGVITK	AMP	0.858
EG1-NC-047895	amidase	YP-009795817.1	67-86	GSHVKGHNHNSIGVCLVGGI	AMP	0.961

Supplementary Table S6. Data giving an overview of the size, genome position, stability, and domains identified within the endolysins derived from model phages λ, T4, T5, T7, and SP6.

Phage name	Phage accession number	Protein name	Position of the gene in virus genome	Gene length (nt)	Protein accession number	Protein length (aa)	Molecular weight (kDa)	Instability index	Isoelectric point (pI)	Gravy	Aliphatic index	Melting point (Tm)	Name of the predicted domain	Position of the domain in protein sequence
Lambda	NC_001416.1	endolysin	45493 - 45969	477	NP_040645.1	158	17,82	22,188	9,21	-0,631	77,848	43.0°C	lambda_lys-like (cd00736) Lysozyme-like (SSF53955) Lysozyme (G3DSA:1.10.530.10:FF:000009) Endolysin. (MF_04109)	7-151 2-152 1-158 1-154
T4	NC_000866.4	glycoside_hydrolase_family_protein	66503 - 66997	495	NP_049736.1	164	18,69	32,488	9,59	-0,385	89,268	52.0°C	T4-like_lys (cd00735) Phage lysozyme (PF00959) Lysozyme-like (SSF53955) Endolysin. (MF_04110) Endolysin (G3DSA:1.10.530.40:FF:000002)	2-160 11-135 1-161 1-159 1-164
SP6	NC_004831.2	endolysin	38976 - 39332	357	NP_853606.1	118	13,13	34,402	9,42	-0,424	69,407	52.0°C	Peptidase M15 (PF08291) Hedgehog/DD-peptidase (SSF55166)	6-108 7-115
T7	NC_001604.1	amidase	10706 - 11161	456	NP_041973.1	151	16,98	40,307	8,81	-0,482	74,901	50.0°C	PGRP (cd06583) N-acetylmuramoyl-L-alanine amidase-like (SSF55846) N-acetylmuramoyl-L-alanine amidase (PF01510) Endolysin. (MF_04111) Endolysin (G3DSA:3.40.80.10:FF:000009)	13-132 10-146 11-132 1-149 6-151
T5	NC_005859.1	endolysin	19931 - 20344	414	YP_006868.1	137	15,26	35,778	7,91	-0,495	76,277	46.0°C	L-Ala-D-Glu_peptidase_like (cd14845) Hedgehog/DD-peptidase (SSF55166) Endolysin (G3DSA:3.30.1380.10:FF:000004)	9-136 3-120 3-136

Supplementary Table S7. Results of the comparative analysis of all, tested in this work, endolysin sequences and those derived from model phages λ, T4, T5, T7, and SP6. Sequence comparison was performed by using BLAST. The % of similarity was calculated by dividing the number of similar amino acids (selected as positives by BLAST) by the length of the tested protein and multiplying by 100%.

Phage name_acc. number	Protein accession number	Protein length	Reference phage	Percentage of similarity	Score	E-value	Identities	Positives	Gaps
590B_MW722821	QVR48622.1	163	T4	33.13	27.3	2e-04	33/133	54/133	25/133
EG1_NC_047895	YP_009795842.1	1325	T7	0.98	18.1	2.2	9/22	13/22	0/22
EG1_NC_047895	YP_009795817.1	152	T7	96.71	301	2e-111	142/151	147/151	0/151
Killian_OQ446694	WEU67851.1	606	T7	2.97	18.5	0.79	12/47	18/47	0/47
Killian_OQ446694	WEU67899.1	165	T4	96.97	328	1e-121	158/164	160/164	0/164
MLP1_NC_079176	YP_010844467.1	132	T5	59.09	89.0	7e-28	50/137	78/137	9/137
MLP2_OK148441	UEN68765.1	165	T4	96.97	327	3e-121	158/164	160/164	0/164
NTEC3_NC_073318	YP_010749155.1	163	T4	34.97	23.9	0.003	36/152	57/152	23/152
vB_EcoM_ACG-C40_NC_019399	YP_006986678.1	165	T4	96.97	329	7e-122	159/164	160/164	0/164
vB_EcoM_KMB26_OW536688	CAH2619090.1	165	T4	96.97	329	7e-122	159/164	160/164	0/164
vB_EcoM_LNA1_MZ311863	self_ZKRH133	155	T4	29.03	30.4	1e-05	25/108	45/108	18/108
vB_EcoM_LNA2_MZ311865	self_NLWW35	163	T4	88.34	277	2e-101	131/161	144/161	0/161
vB_EcoM_LNA4_MZ311864	self_PFXJ59	185	T5	4.32	15.8	1.9	5/12	8/12	0/12
vB_EcoM_LNA4_MZ311864	self_PFXJ204	191	T4	9.95	23.9	0.004	7/28	19/28	0/28
vB_EcoM_SHAK7163_OR594179	WVW77660.1	1107	T5	1.17	20.8	0.26	12/20	13/20	2/20
vB_EcoM_SHAK7163_OR594179	WVW77668.1	115	SP6	82.61	185	1e-66	88/114	95/114	0/114
vB_EcoM_SHAK7693_OR594180	WVW77753.1	80	T4	20.0	18.5	0.055	13/36	16/36	5/36
vB_EcoM_SHAK7693_OR594180	WVW77754.1	101	T7	6.93	13.5	5.2	5/12	7/12	0/12
vB_EcoM_SHAK7704_OR594181	self_RMKP5	115	SP6	83.48	184	3e-66	88/114	96/114	0/114
vB_EcoM_SHAK7704_OR594181	self_RMKP62	1103	T5	1.18	20.8	0.24	12/20	13/20	2/20
vB_EcoM_SHAK7854_OR594182	self_DNYN53	1296	SP6	0.85	20.8	0.32	9/17	11/17	0/17
vB_EcoM_SHAK7854_OR594182	self_DNYN29	153	T7	77.78	225	3e-81	105/151	119/151	0/151
vB_EcoM_SHAK7858_OR594183	WVW77834.1	162	T4	34.57	21.6	0.020	36/152	56/152	23/152
vB_EcoM_SHAK9454_OR594184	self_ZENB53	1296	SP6	1.08	20.8	0.31	11/23	14/23	1/23
vB_EcoM_SHAK9454_OR594184	self_ZENB29	153	T7	77.78	226	2e-81	105/151	119/151	0/151
vB_EcoM-101112UKE3-1_MZ234013	QZI78958.1	165	T4	96.97	329	7e-122	159/164	160/164	0/164

vB_EcoM-101117BS1_MZ234018	QZI79532.1	165	T4	96.97	329	7e-122	159/164	160/164	0/164
vB_EcoM-172859UKE1_MZ234025	QZI80176.1	163	T4	88.96	280	1e-102	133/161	145/161	0/161
vB_EcoM-CHD16UKE1_MZ234030	QZI80907.1	165	T4	96.97	329	7e-122	159/164	160/164	0/164
vB_EcoM-CHD2BS1_MZ234027	QZI80464.1	165	T4	96.97	329	7e-122	159/164	160/164	0/164
vB_EcoM-CHD94UKE2_MZ234031	QZI81192.1	163	T4	95.71	317	2e-117	152/161	156/161	0/161
vB_EcoM-G3G7_MZ234040	QZI83386.1	165	T4	96.97	329	7e-122	159/164	160/164	0/164
vB_EcoP_ACG-C91_NC_019403	YP_006987811.1	115	SP6	85.22	185	2e-66	86/114	98/114	0/114
vB_EcoP_LNA3_MZ311862	self_GXIJ6	1296	SP6	1.08	20.8	0.30	11/23	14/23	1/23
vB_EcoP_LNA3_MZ311862	self_GXIJ34	153	T7	77.78	223	3e-80	103/151	119/151	0/151
vB_EcoP_LNA5_MZ311861	self_TCEH45	1296	SP6	1.08	20.8	0.32	11/23	14/23	1/23
vB_EcoP_LNA5_MZ311861	self_TCEH12	153	T7	77.78	224	6e-81	105/151	119/151	0/151
vB_EcoP-101101UKE1_MZ234012	QZI78680.1	115	SP6	83.48	184	6e-66	89/114	96/114	0/114
vB_EcoP-101117UKE2_MZ234019	QZI79610.1	124	SP6	78.23	186	1e-66	89/115	97/115	0/115
vB_EcoP-101118B1_MZ234020	QZI79670.1	124	SP6	78.23	186	1e-66	89/115	97/115	0/115
vB_EcoP-101118UKE1_MZ234021	QZI79721.1	153	T7	77.12	221	9e-80	103/151	118/151	0/151
vB_EcoP-101120B1-2_MZ234022	QZI79784.1	153	T7	77.78	224	6e-81	105/151	119/151	0/151
vB_EcoP-101120B2_MZ234023	QZI79844.1	153	T7	77.78	224	6e-81	105/151	119/151	0/151
vB_EcoP-101136BS1_MZ234024	QZI79903.1	153	T7	77.78	225	3e-81	105/151	119/151	0/151
vB_EcoP-22664BS1_MZ234009	QZI78465.1	118	T7	78.81	177	4e-63	84/116	93/116	0/116
vB_EcoP-22664UKE3-2_MZ234011	QZI78611.1	153	T7	77.78	224	9e-81	104/151	119/151	0/151
vB_EcoP-UTI89UKE2_MZ234049	QZI84581.1	115	SP6	83.48	185	1e-66	89/114	96/114	0/114

vB_EcoP-UTI89UKE3_MZ234050	QZI84648.1	115	SP6	83.48	185	1e-66	89/114	96/114	0/114
vB_EcoS_ACG-M12_NC_019404	YP_006987885.1	162	Lambda	4.32	14.2	5.4	4/12	7/12	0/12
vB_EcoS_UTEC10_PP291582	WWY66481.1	162	T4	33.95	30.0	3e-05	34/133	55/133	25/133
vB_EcoS-101114B2_MZ234014	QZI79000.1	164	T4	21.34	25.4	0.001	23/89	35/89	23/89
vB_EcoS-101114BS4_MZ234016	QZI79065.1	164	T4	21.34	25.4	0.001	23/89	35/89	23/89
vB_EcoS-22664B1_MZ234008	QZI78369.1	164	T4	21.34	25.4	0.001	23/89	35/89	23/89
vB_EcoS-22664BS2_MZ234010	QZI78493.1	163	T4	34.36	27.3	2e-04	37/152	56/152	23/152
vB_EcoS-CHD2B1_MZ234026	USL86318.1	164	T4	21.34	25.4	0.001	23/89	35/89	23/89
vB_EcoS-CHD5UKE2_MZ234029	USL86352.1	164	T4	21.34	23.9	0.003	22/86	35/86	23/86
VB_EcoS-Golestan_NC_042084	YP_009620148.1	153	T4	37.25	28.1	1e-04	38/151	57/151	23/151
vB_EcoS-phiEc3_NC_073323	YP_010749568.1	163	T4	34.36	27.7	2e-04	38/152	56/152	23/152
vB_EcoS-phiEc4_MZ576218	QYU43915.1	162	T4	35.19	23.9	0.003	36/152	57/152	23/152
vB_EcoS-UTI89UKE1_MZ234048	USM81274.1	163	T4	34.36	27.7	1e-04	38/152	56/152	23/152
vB-Eco-KMB14_OR424384	WPK27911.1	115	SP6	83.48	183	8e-66	88/114	96/114	0/114
vB-Eco-KMB22_OR539217	WPJ70775.1	165	T4	96.97	328	2e-121	157/164	160/164	0/164
vB-Eco-KMB23_OR525702	WQN07383.1	165	T4	96.97	329	7e-122	159/164	160/164	0/164
vB-Eco-KMB25_OR424386	WPK28014.1	165	Lambda	4.24	13.5	9.6	4/8	7/8	0/8
vB-Eco-KMB26_OR539218	WPJ71039.1	165	T4	96.97	329	7e-122	159/164	160/164	0/164
vB-Eco-KMB35_OR525704	WQN07684.1	163	T4	34.36	27.3	2e-04	38/152	56/152	23/152
vB-Eco-KMB36_OR525703	WQN07602.1	168	T4	24.4	23.5	0.004	26/103	41/103	5/103
vB-Eco-KMB37_OR525700	WQN06754.1	167	T4	64.67	171	1e-59	81/161	108/161	0/161
vB-Eco-KMB38_OR525701	WQN07107.1	163	T4	88.96	280	2e-102	133/161	145/161	0/161
vB-Eco-KMB39_OR539219	WPJ71303.1	165	T4	96.97	329	7e-122	159/164	160/164	0/164
vB-Eco-KMB40_OR539220	WPJ71570.1	165	T4	96.97	329	7e-122	159/164	160/164	0/164

vB-Eco-KMB42_OR539221	WPJ71835.1	165	T4	96.97	329	7e-122	159/164	160/164	0/164
vB-Eco-KMB43_OR525698	WQN06429.1	130	T5	55.38	87.8	2e-27	48/128	72/128	5/128
vB-Eco-KMB46_OR525699	WQN06679.1	164	SP6	4.27	15.4	2.1	5/9	7/9	0/9
vB-Eco-KMB47_OR424385	WPK27967.1	115	SP6	83.48	184	6e-66	89/114	96/114	0/114
ZCEC14_OP168898	UVD33441.1	163	T4	96.32	322	2e-119	155/161	157/161	0/161

Oświadczania współautorów

Wojciech Wesołowski
.....
IMIĘ I NAZWISKO

28.08.2026
.....
DATA

Katedra Biologii Molekularnej,
Wydział Biologii,
Uniwersytet Gdański

.....
AFILIACJA

OŚWIADCZENIE

Niniejszym oświadczam, iż mój wkład w publikację:

Wesołowski W, Łukasiak A, Bloch S, Kuligowska K, Neumann J, Lewandowska N, Węglińska E, Węgrzyn G, Nejman-Faleńczyk B. Phage Endolysins as Promising and Effective Candidates for Use Against Uropathogenic *Escherichia coli*. Viruses. 2025 Apr 13;17(4):560. doi: 10.3390/v17040560,

polegał na pozyskaniu funduszy, zebraniu danych, napisaniu rozdziałów 2 i 3 oraz na wykonaniu opisanych w tych rozdziałach analiz bioinformatycznych, w tym na:

- wykonaniu analizy *in silico* zebranych w Tabeli nr 1 białek,
- identyfikacji domen białkowych z użyciem oprogramowania InterProScan (Tabela S1 i Tabela S4),
- identyfikacji peptydów antydrobnoustrojowych (AMP) w obrębie sekwencji analizowanych endolizyn (Tabela S2 i Tabela S5),
- wykonaniu analizy *in silico* endolizyn fagów modelowych λ, T4, T5, T7 i SP6,
- wykonaniu analizy filogenetycznej i konstrukcji drzew dla 11 grup ortologicznych, w które zostały pogrupowane badane białka z użyciem oprogramowania OMA (Ryciny 1-11) oraz scharakteryzowaniu białek wyselekcjonowanych w analizie filogenetycznej jako białka pokrewne do testowanych potencjalnych endolizyn (Tabela S3),
- przeprowadzeniu analizy porównawczej badanych endolizyn z endolizynami fagów modelowych λ, T4, T5, T7 i SP6 (Tabela S7),
- współudziale w przygotowaniu odpowiedzi na uwagi recenzentów,
- a także na obróbce i interpretacji wyników oraz przygotowaniu rycin i tabel.

Wesołowski
.....
PODPIS

Aleksandra Łukasiak
.....
IMIĘ I NAZWISKO

Katedra Biologii Molekularnej,
Wydział Biologii,
Uniwersytet Gdański

.....
AFILIACJA

..... 27.08.2026 ....
DATA

OŚWIADCZENIE

Niniejszym oświadczam, iż mój wkład w publikację:

Wesołowski W, Łukasiak A, Bloch S, Kuligowska K, Neumann J, Lewandowska N, Węglińska E, Węgrzyn G, Nejman-Faleńczyk B. Phage Endolysins as Promising and Effective Candidates for Use Against Uropathogenic *Escherichia coli*. Viruses. 2025 Apr 13;17(4):560. doi: 10.3390/v17040560,

polegał na

(i) przygotowaniu rozdziału numer 4;

(ii) pozyskaniu funduszy.

.....
PODPIS

Sylwia Bloch
.....
IMIĘ I NAZWISKO

Katedra Biologii Molekularnej,
Wydział Biologii,
Uniwersytet Gdański

.....
AFILIACJA

08.09.2026 r.
.....
DATA

OŚWIADCZENIE

Niniejszym oświadczam, iż mój wkład w publikację:

Wesołowski W, Łukasiak A, Bloch S, Kuligowska K, Neumann J, Lewandowska N, Węglińska E, Węgrzyn G, Nejman-Faleńczyk B. Phage Endolysins as Promising and Effective Candidates for Use Against Uropathogenic *Escherichia coli*. Viruses. 2025 Apr 13;17(4):560. doi: 10.3390/v17040560,

polegał na współudziale w:

(i) nadzorowaniu prac polegających na zbieraniu danych do przygotowania rozdziału numer 1 oraz 2;

(ii) nadzorowaniu prac związanych z powstawaniem rozdziału nr 4;

(iii) przygotowaniu pierwotnej i ostatecznej wersji manuskryptu.

Bloch
.....
PODPIS

Kaja Stromska (Kuligowska)
.....
IMIĘ I NAZWISKO

24.08.2026 r. Gdańsk
DATA

Katedra Biologii Molekularnej,
Wydział Biologii,
Uniwersytet Gdański
.....
AFILIACJA

OŚWIADCZENIE

Niniejszym oświadczam, iż mój wkład w publikację:

Wesołowski W, Łukasiak A, Bloch S, Kuligowska K, Neumann J, Lewandowska N, Węglińska E, Węgrzyn G, Nejman-Faleńczyk B. Phage Endolysins as Promising and Effective Candidates for Use Against Uropathogenic *Escherichia coli*. Viruses. 2025 Apr 13;17(4):560. doi: 10.3390/v17040560,

polegał na zbieraniu danych literaturowych na temat endolizyn fagowych.

Kaja Stromska
PODPIS

Julia Neumann
.....
IMIĘ I NAZWISKO

25/08/26r.
.....
DATA

Pracownia Chemoinformatyki Środowiska,
Wydział Chemii,
Uniwersytet Gdański

.....
AFILIACJA

OŚWIADCZENIE

Niniejszym oświadczam, iż mój wkład w publikację:

Wesołowski W, Łukasiak A, Bloch S, Kuligowska K, Neumann J, Lewandowska N, Węglińska E, Węgrzyn G, Nejman-Faleńczyk B. Phage Endolysins as Promising and Effective Candidates for Use Against Uropathogenic *Escherichia coli*. Viruses. 2025 Apr 13;17(4):560. doi: 10.3390/v17040560,

polegał na współudziale w zbieraniu i walidacji danych do analiz filogenetycznych.

Julia Neuma
.....
PODPIS

Natalia Lewandowska
.....
IMIĘ I NAZWISKO

24. 08. 2026
.....
DATA

Katedra Biologii Molekularnej,
Wydział Biologii,
Uniwersytet Gdański

.....
AFILIACJA

OŚWIADCZENIE

Niniejszym oświadczam, iż mój wkład w publikację:

Wesołowski W, Łukasiak A, Bloch S, Kuligowska K, Neumann J, Lewandowska N, Węglińska E, Węgrzyn G, Nejman-Faleńczyk B. Phage Endolysins as Promising and Effective Candidates for Use Against Uropathogenic *Escherichia coli*. Viruses. 2025 Apr 13;17(4):560. doi: 10.3390/v17040560,

polegał na współudziale w przygotowaniu oraz edycji pierwotnej i ostatecznej wersji manuskryptu.

Natalia Lewandowska
.....
PODPIS

Emilia Piwnicka (Węglińska)
.....
IMIĘ I NAZWISKO

01.09.2026
.....
DATA

Katedra Biologii Molekularnej,
Wydział Biologii,
Uniwersytet Gdański

.....
AFILIACJA

OŚWIADCZENIE

Niniejszym oświadczam, iż mój wkład w publikację:

Wesołowski W, Łukasiak A, Bloch S, Kuligowska K, Neumann J, Lewandowska N, Węglińska E, Węgrzyn G, Nejman-Faleńczyk B. Phage Endolysins as Promising and Effective Candidates for Use Against Uropathogenic *Escherichia coli*. Viruses. 2025 Apr 13;17(4):560. doi: 10.3390/v17040560,

polegał na współudziale w przygotowaniu oraz edycji pierwotnej i ostatecznej wersji manuskryptu.

Emilia Piwnicka
.....
PODPIS

Grzegorz Węgrzyn
.....
IMIĘ I NAZWISKO

Katedra Biologii Molekularnej,
Wydział Biologii,
Uniwersytet Gdański

.....
AFILIACJA

27 sierpnia 2026 r.
.....
DATA

OŚWIADCZENIE

Niniejszym oświadczam, iż mój wkład w publikację:

Wesołowski W, Łukasiak A, Bloch S, Kuligowska K, Neumann J, Lewandowska N, Węglińska E, Węgrzyn G, Nejman-Faleńczyk B. Phage Endolysins as Promising and Effective Candidates for Use Against Uropathogenic *Escherichia coli*. Viruses. 2025 Apr 13;17(4):560. doi: 10.3390/v17040560,

polegał na dyskusjach merytorycznych dotyczących analiz wcześniej opublikowanych artykułów w celu opracowanie spójnej koncepcji i wniosków, a także pomocy w redagowaniu tekstu manuskryptu.

KIEROWNIK
KATEDRY BIOLOGII MOLEKULARNEJ
prof. dr hab. Grzegorz Węgrzyn

.....
PODPIS

Bożena Nejman-Faleńczyk
.....
IMIĘ I NAZWISKO

09.08.2026
.....
DATA

Katedra Biologii Molekularnej,
Wydział Biologii,
Uniwersytet Gdański

.....
AFILIACJA

OŚWIADCZENIE

Niniejszym oświadczam, iż mój wkład w publikację:

Wesołowski W, Łukasiak A, Bloch S, Kuligowska K, Neumann J, Lewandowska N, Węglińska E, Węgrzyn G, Nejman-Faleńczyk B. Phage Endolysins as Promising and Effective Candidates for Use Against Uropathogenic *Escherichia coli*. Viruses. 2025 Apr 13;17(4):560. doi: 10.3390/v17040560,

polegał na:

- opracowaniu i konsultowaniu koncepcji badań oraz założeń naukowych niniejszej publikacji,
- współudziale w sprawowaniu merytorycznego nadzoru nad realizacją prac,
- interpretacji uzyskanych wyników oraz formułowaniu wniosków,
- koordynowaniu prac związanych z przygotowaniem manuskryptu i dbaniu o jego spójność merytoryczną i redakcyjną,
- koordynowaniu współpracy pomiędzy współautorami i określeniu zakresu ich pracy,
- przygotowaniu oraz edycji pierwotnej i ostatecznej wersji manuskryptu,
- komunikacji z redakcją czasopisma, w tym prowadzeniu korespondencji na etapie zgłoszenia, recenzowania i publikacji manuskryptu,
- koordynowaniu prac związanych z przygotowaniem odpowiedzi na uwagi recenzentów oraz wprowadzeniem wymaganych zmian do manuskryptu.

B. Nejman-Faleńczyk
.....
PODPIS

Działalność naukowa

Publikacje naukowe

(pogrubione pozycje wchodzą w skład rozprawy)

1. **Wesołowski W, Czerwonka G, Zegadło K, Jagieła E, Łukasiak A, Grabowski Ł, Mantej J, Dziuba A, Bloch S, Topka-Bielecka G, Necel A, Kostovski D, Węglińska E, Narajczyk M, Węgrzyn G, Nejman-Faleńczyk B, Adamus-Białek W. Significant variability in virulence factors and antibiotic resistance of uropathogenic *Escherichia coli* strains may be counteracted by their susceptibility to lytic bacteriophages. Sci Rep. 2026 Jul 21;16(1):26677. doi: 10.1038/s41598-026-62793-9.**

2. Necel A, Stasiłojć M, **Wesołowski W**, Narajczyk M, Kosznik-Kwaśnicka K, Kaźmierczak N, Naumiuk Ł, Loika H, Piechowicz L, Żywicka A. Newly isolated Pakpunavirus: efficacy and safety assessment in light of alternative therapies against P. aeruginosa skin infections. Front Microbiol. 2026 Apr;17. doi: 10.3389/fmicb.2026.1807725.

3. Bloch S, Lewandowska N, Zwolenkiewicz-Nowak J, Mazur M, Chylewska A, Dąbrowska AM, Musiał N, Czaplewska P, **Wesołowski W**, Łukasiak A, Ferenc M, Pietrzycki M, Węglińska E, Mach P, Węgrzyn G, Olejniczak M, Nejman-Faleńczyk B. UpRoi1 – the phage-derived microRNA-size non-coding RNA molecule regulating both phage and its bacterial host development. Biochim Biophys Acta Gene Regul Mech. 2026 Mar;1869(1):195135. doi: 10.1016/j.bbagrm.2026.195135.

4. **Łukasiak A, Wesołowski W, Neumann J, Lewandowska N, Węglińska E, Bloch S, Węgrzyn G, Nejman-Faleńczyk B. Groundwork for phage therapy: multi-faceted comparative analysis of lytic bacteriophages infecting uropathogenic *Escherichia coli*. Virology. 2026 Feb;615:110738. doi: 10.1016/j.virol.2025.110738.**

5. Lewandowska N, Bloch S, Chylewska A, Dąbrowska AM, Mazur M, Zwolenkiewicz J, **Wesołowski W**, Łukasiak A, Węglińska E, Olejniczak M, Węgrzyn G, Nejman-Faleńczyk B. Application of Ultraviolet (UV) and Circular Dichroism (CD) Spectroscopies to Study RNA-RNA Interactions of Regulatory RNAs of Bacteriophage Origin. In: RNA Spectroscopy. Springer US; 2026. p. 123–136. doi: 10.1007/978-1-0716-5084-4_8.

6. Knefel M, Zielenkiewicz M, Jeleniewska W, Rąbalski Ł, Kapusta M, Górniak M, **Wesołowski W**, Malinowska K, Szota M, Lepionka T, Węgrzyn G, Węgrzyn A,

Grabowski Ł. Evaluation of the effectiveness and safety of bacteriophage vB_Sen-Miriam1 against Salmonella enterica in in vitro and in vivo models. Virology. 2025 Oct;611:110660. doi: 10.1016/j.virol.2025.110660.

7. Lewandowska N, Bloch S, Łukasiak A, **Wesołowski W**, Węgrzyn G, Nejman-Faleńczyk B. The Role of Bacteriophage-Derived Small RNA Molecules in Bacterial and Phage Interactions. Viruses. 2025 Jun;17(6):834. doi: 10.3390/v17060834.

8. Wesołowski W, Łukasiak A, Bloch S, Kuligowska K, Neumann J, Lewandowska N, Węglińska E, Węgrzyn G, Nejman-Faleńczyk B. Phage Endolysins as Promising and Effective Candidates for Use Against Uropathogenic *Escherichia coli*. Viruses. 2025 Apr 13;17(4):560. doi: 10.3390/v17040560.

9. Necel A, Dydecka A, Topka-Bielecka G, **Wesołowski W**, Lewandowska N, Bloch S, Nejman-Faleńczyk B. What, how, and why? – anti-EHEC phages and their application potential in medicine and food industry. J Appl Genet. 2024 Nov;66(1):219–240. doi: 10.1007/s13353-024-00918-4.

10. Bloch S, Lewandowska N, **Wesołowski W**, Łukasiak A, Mach P, Nejman-Faleńczyk B, Węgrzyn G. Analysis of Phage Regulatory RNAs: Sequencing Library Construction from the Fraction of Small Prokaryotic RNAs Less Than 50 Nucleotides in Length. In: Bacterial Regulatory RNA. Springer US; 2024. p. 25–34. doi: 10.1007/978-1-0716-3565-0_3.

11. Witkowska M, **Wesołowski W**, Markiewicz M, Pakizer J, Neumann J, Ożarowska A, Meissner W. The intensity of supplementary feeding in an urban environment impacts overwintering Mallards (Anas platyrhynchos) as wintering conditions get harsher. Avian Res. 2024;15:100205. doi: 10.1016/j.avrs.2024.100205.

Wystąpienia na konferencjach naukowych

Wystąpienia ustne

1. *The role of sRNAs molecules in an antirepressor-mediated initiation of the phage lytic cycle*. V International Sopot Youth Conference 2021 11 czerwca 2021, Sopot, Polska
2. *Influence of the newly discovered sRNA molecules on the development of phage carrying toxins dangerous to humans*. 8th Central European Congress of Life Sciences EUROBIOTECH 21 czerwca 2022, Kraków, Polska
3. *Potencjał nowo odkrytych bakteriofagów w zwalczaniu zakażeń spowodowanych lekoopornymi uropatogenami*. III Sympozjum Bakteriofagowe 5-7 września 2024, Gdańsk, Polska

Wystąpienia posterowe

1. *The role of sRNAs molecules in an antirepressor-mediated initiation of the phage lytic cycle*. IXth Intercollegiate Biotechnology Symposium "Symbioza" 23 maja 2021, Poznań, Polska
2. *Herpeswirusy i bakteriofagi, co mają wspólnego?*. XXIX Zjazd Polskiego Towarzystwa Mikrobiologów 15-17 września 2022, Warszawa, Polska
3. *Porównanie nowo wyizolowanych bakteriofagów mających potencjalne zastosowanie w terapii fagowej skierowanej przeciwko uropatogennym bakteriom Escherichia coli*. II Sympozjum Bakteriofagowe 7-9 czerwca 2023, Gdańsk, Polska
4. *Bacteriophages – a new biological force in the fight against uropathogenic bacteria resistant to antibiotics*. 9th Central European Congress of Life Sciences EUROBIOTECH 27-28 czerwca 2024, Kraków, Polska
5. *Wieloaspektowa charakterystyka bakteriofaga vB-EcoS_57-3 infekującego uropatogenne bakterie Escherichia coli*. IV Sympozjum Bakteriofagowe 4-6 września 2025, Gdańsk, Polska

Kursy, staże i szkolenia

1. Uniwersytet Gdański, Wydział Matematyki, Fizyki i Informatyki, studia licencjackie z bioinformatyki w latach 2020 – 2025. Praca dyplomowa wykonana pod kierunkiem dr. hab. Stanisława Ołdzieja pt. *Identyfikacja wariantów w sekwencjach aminokwasowych białek w obrębie centrów aktywnych i fragmentów konserwowanych*.
2. Uniwersytet Jana Kochanowskiego, Collegium Medicum, Kielce, Polska, 17-22 września 2023 roku, opiekunowie; Wioletta Adamus-Białek i Grzegorz Czerwonka. Staż naukowy mający na celu wykonanie analiz bioinformatycznych szczepów UPEC i uzyskanych z nich bakteriofagów w celu opracowania wyników potrzebnych do publikacji w czasopiśmie naukowym.
3. Polska Akademia Nauk, Instytut Immunologii i Terapii Doświadczalnej im. Ludwika Hirszfelda, Wrocław, 8-9 kwietnia 2024, opiekun Ryan Cook. Szkolenie z zakresu analizy bioinformatycznej danych metagenomowych.

Projekty naukowe

Tytuł: *Assembly and annotation of phage genomes capable of combating uropathogenic Escherichia coli.* Grant na badania otrzymany w ramach IV edycji programu małych grantów – UGrants, realizowanego w ramach Inicjatywy Doskonałości Uczelnia Badawcza (IDUB) w Uniwersytecie Gdańskim.