



Institut des Sciences de l'Evolution-Montpellier

SEKRETARIAT
INSTYTUTU OCEANOGRAPHII

Wplynelo dnia: 16.11.2017

Zarejestrowano pod numerem:

305/17

Dr. Nicolas Bierne, DR2 CNRS (HDR)
Institut des Sciences de l'Evolution UMR5554
Université de Montpellier
Bâtiment 24, 1er étage
Place Eugène Bataillon
34095 Montpellier Cedex 05
France

Tel : +33 (0)4 67 14 36 88
e-mail : n-bierne@univ-montp2.fr

Subject: Favorable recommendation to the defense of the PhD thesis of LUDMIŁA KATARZYNA SROMEK "PHYLOGEOGRAPHY AND POPULATION GENOMICS OF THE LAGOON COCKLE CERASTODERMA GLAUCUM"

Report

During her PhD work, Ludmila Sromek studied the population genomics of the lagoon cockle *Cerastoderma glaucum*. The objectives of the study were (i) to evaluate the level of genetic differentiation along the European coasts in a marine species with a patchily distributed habitat and a moderate larval stage duration, and elucidate the influence of lagoon habitats on connectivity with a comparison with a related species that inhabits the shoreline, (ii) to localize major genetic breaks and discuss their origin and maintenance and (iii) to describe divergence patterns across the genome. A major achievement of Ludmila Sromek's work is the development of a high-throughput genotyping technique (RAD sequencing) in a non-model bivalve species. Important results are (i) the identification of genetic barriers to gene flow in an understudied part of the Mediterranean Sea, between Aegean, Ionian and Black seas, and (ii) the observation of an isolation-by-distance (IBD) pattern of genetic subdivision from the Mediterranean Sea to the Baltic Sea, with an overall level of genetic differentiation well above what is usually observed in other marine bivalve species. These results provide new information that shades light on the population genetics of marine bivalves in general and will probably be welcomed by the scientific community. The lagoon cockle indeed provides a new example in the increasing list of old divergence breaks associated with two young postglacial marine transgressions in the Black and Baltic Seas. It also provides an example of a much more pronounced IBD pattern than usually observed in other marine invertebrates.

The thesis begins with a detailed introduction that reviews many aspects of the marine population genetics literature of the studied area and system, some information about the geological histories, a review of genetic differentiation at major marine transition zones, of genetic structure in post-glacially recolonized areas, some information about the population genomics approach and about the

Institut des Sciences de l'Evolution de Montpellier (ISEM) - Université Montpellier 2 - Place Eugène Bataillon
Case courrier 065 - 34095 MONTPELLIER cedex 5, France

Code UMR ISEM, UMR 5554 Université de Montpellier - CNRS - IRD - CIRAD - EPHE

Tél. (3 3) 04 67 46 33 75, Fax (33) 04 67 46 33 99, email : nicolas.bierne@univ-montp2.fr



ecology of lagoon and of *Cerastoderma glaucum*, and a review of previous works on the population genetics of the lagoon cockle. Although I am impressed by the literature reviewed, my overall feeling is that it has remained mainly descriptive while I would have liked a more focused exposé of the evolutionary processes that underlie each interpreting hypothesis. Genetic breaks can be explained by hydrographic (connectivity), genetic (reproductive isolation), environmental (local adaptation or resistance distance) or historical (secondary contact) factors, or more likely a combination of some of them (if not all). Often only one of them is invoked in a given publication (mainly hydrography or environment) while other factors are not addressed. One message of this PhD thesis is that some species are not differentiated at some transition zones (eg. the Almeria-Oran front) suggesting that hydrography alone is not sufficient to explain the genetic differentiation observed at other species. Hydrography alone rather generates a large scale IBD pattern, even when sampling across emblematic barriers to dispersal. To my point of view it would have been more straightforward to use this review to emphasize that marine genetic divides are multifactorial (although I acknowledge that this is well explained at some places). The interest of studying a lagoon species is well explained and put into context.

After exposing the objectives, the results are synthesized in a section that refers to three articles provided in annexes. Here I will address some concerns about a few points although overall the PhD report is very good.

- RADseq data filtering. One can adjust the stringency with which NGS data are filtered in order to adjust the trade-off between quantity and quality. And actually this is what was partly done in the present work. The study of the spatial structure and genomewide level of differentiation does not require intending to keep as many loci as possible. After filtering, strong departures from Hardy-Weinberg proportions still remain ($F_{is} \sim 0.3$). Such departure cannot easily be explained by biological processes and are most likely the consequence of technical artefacts. In addition, the starshaped genetic relationship among individuals of the same sample suggests that, once intra-individual genetic correlations are eluded, there is no variance in relatedness between pairs of individuals, which would be expected for any biological reasons of positive F_{is} . The literature on heterozygote deficiency in marine bivalves is misleading. There is no reasonable reason why we should observe heterozygote deficiencies outside technical problem. To my point of view Hardy-Weinberg proportion should be used itself as a filtering criterion. In addition, despite a high average coverage per loci (>140X) the rate of missing data is much too strong (50%). I would find it safer to keep only loci with a missing rate below 10%. With small sample sizes and individual-based analysis, one

Institut des Sciences de l'Evolution de Montpellier (ISEM) - Université Montpellier 2 - Place Eugène Bataillon
Case courrier 065 - 34095 MONTPELLIER cedex 5, France

Code UMR ISEM, UMR 5554 Université de Montpellier - CNRS - IRD - CIRAD - EPHE

Tél. (3 3) 04 67 46 33 75, Fax (33) 04 67 46 33 99, email : nicolas.bierne@univ-montp2.fr

would not want to use a dataset full of holes. Actually the LEA outputs are likely to be strongly influenced by non-panmictic loci and they are very difficult to interpret. An LEA analysis on a much cleaner dataset matrix would be highly valuable.

- Mito-nuclear discordance. The surprising pattern observed at the mitochondrial locus is a focus of the first article that is not so much discussed with the RADseq data. This is not because the average nuclear structure does not correlate with the mtDNA locus that some nuclear loci / genome regions could not correlate. A scan of the mito-nuclear disequilibrium would have been welcomed. In addition, I would have appreciated in the PhD report (not necessarily in the article) a more thorough discussion of the alternative explanations by explaining why a mito-nuclear incompatibility could have moved and been trapped at unexpected place, how gene surfing during population colonization could create this pattern and why it could be more pronounced at the mitochondrial locus etc...

- Demographic history. The demographic history of populations is important to decipher the various interpreting scenarios that can explain the patterns observed. Much care should be taken to the interpretation of a phylogenetic tree approach in a within species context. The basic phylogenetic approach considers correlated genealogies, either because of using a single non-recombining locus or because we make the assumption that speciation is short enough that recombination between genome regions can be neglected. Here recombination occurs, within populations and also between populations (at least in the IBD area). Every individuals of a sample are equally distant to other individuals of the same sample as expected in a panmictic population. The genetic proximity between two samples is not necessarily because of low divergence from a recent common ancestral population but because gene flow is stronger. To reconstruct a demographic history one needs to analyze the phylogeny of nonrecombining loci (gene trees) or to use coalescent-based or coalescent-derived methods. Even the coalescent-based species tree says little about the demographic history, because it assumes no gene flow. With RADseq data, the best are methods that use the joint site frequency spectrum (JSFS) as for example in the software dadi.

- Co-existence in sympatry. I have been surprised that the observation of a coexistence of individuals from other genetic clusters in the same sample not been discussed more than that. I agree they are just one individual each time, but sample sizes are small. This is important information, if they are not the consequence of manipulation errors, because they support the hypothesis of reproductively isolated lineages (Note however that they cannot explain Fis values). To my point of view, if there are three cryptic partially isolated lineages in the Eastern Mediterranean sea, these three individuals should be thoroughly discussed.

Institut des Sciences de l'Evolution de Montpellier (ISEM) - Université Montpellier 2 - Place Eugène Bataillon
Case courrier 065 - 34095 MONTPELLIER cedex 5, France

Code UMR ISEM, UMR 5554 Université de Montpellier - CNRS - IRD - CIRAD - EPHE

Tél. (3 3) 04 67 46 33 75, Fax (33) 04 67 46 33 99, email : nicolas.bierne@univ-montp2.fr



Institut des Sciences de l'Evolution-Montpellier

- Genome scan. I would have preferred F_{st} scans to be performed between identified genetic clusters, rather than using this unusual decomposition according to overall F_{st} bins. The difference between figure 5b and 5c is nothing than a visual representation that you have chosen loci for which the ratio of the intrapop versus interpop genetic variance is stronger, hence internal branches are longer in 5c than in 5b. There are nothing to deduce from this procedure. What is interesting are loci that show an extreme differentiation level and what are their pattern of genetic structure. If a high rate of outliers is found between the three isolated cryptic lineages (figure 5C), this is an argument for their mere existence, and this should be investigated more thoroughly. As said above, scanning for mito-nuclear disequilibrium could also be interesting.

Despite these few concerns, that I understand as a consequence of a lack of time to analyze such big datasets produced by NGS, I repeat that I found this PhD report very good. I also emphasize that I had no concern about the comparison between the genetic structure found in *C. glaucum* and *C. edule*. I think this result of a much pronounced genetic differentiation in the lagoon species when compare to a shoreline species, and that is not explained by cryptic subdivision, is an important result for people interested in marine genetic connectivity.

To conclude, I consider the PhD work of Ludmila Sromek as excellent, and I am supportive and authorize the PhD thesis defense.

Nicolas Bierne



Institut des Sciences de l'Evolution de Montpellier (ISEM) - Université Montpellier 2 - Place Eugène Bataillon
Case courrier 065 - 34095 MONTPELLIER cedex 5, France

Code UMR ISEM, UMR 5554 Université de Montpellier - CNRS - IRD - CIRAD - EPHE

Tél. (3 3) 04 67 46 33 75, Fax (33) 04 67 46 33 99, email : nicolas.bierne@univ-montp2.fr

TLUMACZENIE POŚWIADCZONE Z JĘZYKA ANGIELSKIEGO

Logo: ISEM

Institut des Sciences de l'Evolution – Montpellier

[Uniwersytet Nauk o Ewolucji]

Dr. Nicolas Bierne, DR2 CNRS (HDR)

Tel : +33 (0)4 67 14 36 88

Institut des Sciences de l'Evolution UMR5554

e-mail : n-bierne@univ-montp2.fr

Université de Montpellier

Bâtiment 24, 1er étage

Place Eugene Bataillon

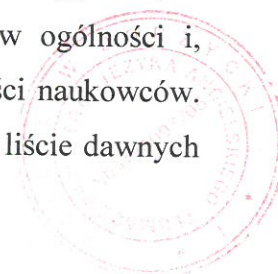
34095 Montpellier Cedex 05

Francja

Temat: Pozytywna rekomendacja w sprawie przyszłej obrony Rozprawy Doktorskiej Ludmiły Katarzyny Sromek pt. „FILOGEOGRAFIA I GENOMIKA POPULACJI SERCÓWKI - CERASTODERMA GLAUCUM”

Recenzja

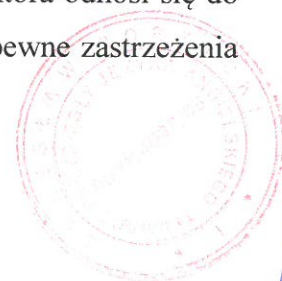
W swojej pracy doktorskiej Ludmiła Sromek przebadła genomikę populacji sercówki – *Cerastoderma glaucum*. Celem badań była (i) ocena poziomu zróżnicowania genetycznego gatunku morskiego występującego wzdłuż wybrzeży europejskich w nieregularnie rozmieszczonych siedliskach o umiarkowanej długości stadium larwalnego, a także wyjaśnienie wpływu siedlisk lagunowych na łączność pomiędzy populacjami w porównaniu z pokrewnym gatunkiem występującym wzdłuż wybrzeży, (ii) lokalizacja głównych przełomów w strukturze genetycznej oraz omówienie ich powstawania i utrzymywania się, oraz (iii) opis wzorców zróżnicowania w genomie. Głównym osiągnięciem pracy Ludmiły Sromek jest opracowanie wysokowydajnej techniki genotypowania (sekwencjonowanie RAD) u niemodelowego gatunku małża. Do ważnych osiągnięć zaliczam: (i) identyfikację barier genetycznych dla przepływu genów w niewystarczająco przebadanej części Morza Śródziemnego, pomiędzy Morzem Egejskim, Jońskim i Czarnym, oraz (ii) obserwacja wzorca izolacji przez dystans (IBD) kolejnych grup genetycznych od Morza Śródziemnego do Bałtyku przy ogólnym poziomie zróżnicowania genetycznego znacznie powyżej tego, co jest zwykle obserwowane u innych gatunków morskich małży. Powyższe wyniki dostarczają nowych informacji, które rzucają światło na genetykę populacji małży morskich w ogólności i, prawdopodobnie, spotkają się z pozytywnym przyjęciem ze strony społeczności naukowców. Istotnie, sercówka lagunowa dostarcza nowego przykładu na wydłużającej się liście dawnych



[Handwritten signature]

dywergencji związanych z dwoma młodymi polodowcowymi transgresjami morskimi w Morzu Czarnym i Bałtyckim. Dostarcza ona również przykład dużo bardziej wyrazistego wzorca IBD niż zwykle obserwowanego u innych bezkręgowców morskich. Rozprawę otwiera szczegółowy wstęp w którym autorka dokonuje przeglądu wielu aspektów piśmiennictwa dotyczącego genetyki populacji morskich w zakresie badanego rejonu i systemu, informacji dotyczących historii geologicznej, przeglądu (i) zróżnicowania genetycznego w głównych morskich strefach przejściowych, (ii) struktury genetycznej w rejonach rekolonizowanych w okresie polodowcowym, informacji z perspektywy genomiki populacji oraz informacji dotyczących ekologii laguny i gatunku *Cerastoderma glaucum*, a także przegląd prac wcześniejszych, zajmujących się genetyką populacji sercówki lagunowej. Mimo, iż jestem pod dużym wrażeniem dokonanego przeglądu piśmiennictwa, moje ogólne odczucie jest takie, że przegląd ten jest głównie opisowy, podczas gdy życzylibym sobie przedstawienia bardziej sprecyzowanej prezentacji procesów ewolucyjnych, które leżą u podstaw każdej hipotezy interpretującej. Przełomy w strukturze genetycznej mogą być wyjaśnione czynnikami hydrograficznymi (łączność), genetycznymi (izolacja rozrodcza), środowiskowymi (adaptacja lokalna lub dystans odporności) lub też czynnikami historycznymi (kontakt wtórny) lub, co najbardziej prawdopodobne, kombinacją kilku z nich (jeśli nie wszystkich). Często tylko jeden z nich jest rozpatrywany w danej publikacji (głównie czynnik hydrograficzny lub środowiskowy) natomiast pozostałe są pominięte. Głównym przesłaniem Rozprawy Doktorskiej jest stwierdzenie, że niektóre gatunki nie są zróżnicowane w niektórych strefach przejściowych (np. front Almeria – Oran), co sugeruje, że sama hydrografia nie wystarcza aby wyjaśnić zróżnicowanie genetyczne obserwowane u innych gatunków. Czynniki hydrograficzne, jako takie generują raczej wzorzec IBD o dużej skali, nawet jeśli próbkowanie przeprowadzane jest z różnych stron typowych barier przeciwdziałających rozprzestrzenianiu. Z mojego punktu widzenia o wiele prościej byłoby wykorzystać ten przegląd, aby podkreślić fakt, że podziały w genetyce morskiej są wieloczynnikowe (choć, z drugiej strony, muszę przyznać, że w niektórych miejscach Rozprawy problem ten jest dobrze wyjaśniony). Cel badań nad gatunkami lagunowymi został dobrze wyjaśniony i osadzony we właściwym kontekście.

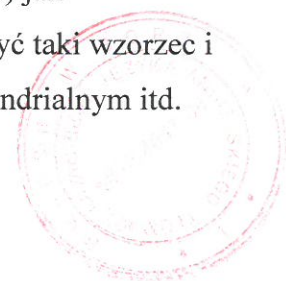
Po przedstawieniu celów Rozprawy, następuje synteza wyników w sekcji, która odnosi się do trzech artykułów przedstawionych w Aneksach. Pozwolę sobie przedstawić pewne zastrzeżenia dotyczące kilku punktów, chociaż całość Rozprawy jest bardzo dobra.



Filtrowanie danych RADseq Można dostosować stopień rygorystyczności podejścia do filtrowania danych NGS, aby uzyskać kompromis pomiędzy ilością i jakością, co zresztą zostało częściowo zrobione w tej Rozprawie. Badania struktury przestrzennej i zróżnicowania na poziomie całego genomu nie wymaga założenia o utrzymaniu jak największej ilości loci. Po przefiltrowaniu, w dalszym ciągu pozostają znaczne odstępstwa od proporcji Hardy'ego-Weinberga ($F_{is} \sim 0.3$). Takiego odstępstwa nie można łatwo wyjaśnić procesami biologicznymi a, co najbardziej prawdopodobne, są one konsekwencjami artefaktów technicznych. Ponadto, gwiazdziście ukształtowane powiązania pomiędzy osobnikami tej samej próbki sugeruje, że kiedy już wyeliminujemy międzyosobnicze powiązania genetyczne, nie będzie już odchyień w pokrewieństwie par osobników, których można by oczekiwać, z jakichkolwiek biologicznych przyczyn dodatniego F_{is} . Piśmiennictwo zajmujące się deficytem heterozygoty u małży morskich jest mylące. Nie istnieje żaden uzasadniony powód dla którego powinniśmy obserwować deficyty heterozygot, poza problemem technicznym. Według mnie to proporcja Hardy'ego-Weinberga, powinna być użyta sama w sobie, jako kryterium filtrowania. Ponadto, pomimo wysokiego średniego próbkowania przypadającego na dane miejsce ($>140x$) ilość brakujących danych jest zbyt znaczna (50%). Uznałbym, jako bardziej bezpieczne zachowanie wyłącznie tych miejsc gdzie ilość brakujących danych jest mniejsza niż 10%. Przy małych wielkościach próbek oraz analizie osobniczej, nikt nie chciałby korzystać z zestawu danych pełnego luk. W rzeczywistości wyniki LEA są, prawdopodobnie pod silnym wpływem miejsc nie-panmiktycznych i są one bardzo trudne do zinterpretowania. Analiza LEA na dużo bardziej „czystej” matrycy zestawu danych byłaby bardziej pożądana.

- Niezgodność mito-nuklearna.

Zaskakujący wzorzec, zaobserwowany w locus mitochondrialnym jest przedmiotem pierwszego artykułu, który nie został omówiony w wystarczający sposób w kontekście danych RADseq. To nie jest dlatego, że przeciętna struktura nuklearna nie koreluje z loci mtDNA, tylko dlatego, że pewne loci nuklearne / obszary genomu mogą nie korelować. Bardzo pożądanym byłby skan, ukazujący brak równowagi mito-nuklearnej. Ponadto bardziej ceniałbym sobie dogłębnierze omówienie w pracy doktorskiej (niekoniecznie w samym artykule) (i) alternatywnych wyjaśnień uzasadniających dlaczego ta niekompatybilność mito-nuklearna mogła się przemieszczać i „wpaść w pułapkę” w nieoczekiwanym miejscu, (ii) jak „surfowanie” genu podczas kolonizacji przez daną populację mogło wytworzyć taki wzorzec i (iii) dlaczego mógłby on być bardziej wyraźnie uwidoczniiony w loci mitochondrialnym itd.



Historia demograficzna.

Historia demograficzna populacji jest istotna dla odtworzenia rozmaitych scenariuszy interpretacyjnych, które mogłyby wyjaśnić obserwowane wzorce. Należałoby ostrożnie podchodzić do interpretacji drzewa filogenetycznego w kontekście wewnątrz-gatunkowym. Podstawowe podejście filogenetyczne bierze pod uwagę genealogie skorelowane, albo z uwagi na wykorzystanie pojedynczego, nierekombinującego loci albo z uwagi na to, iż zakładamy, że ewolucyjne powstawanie nowego gatunku jest na tyle krótkie, że można pominąć rekombinację pomiędzy obszarami genomu. Tutaj rekombinacja ta ma miejsce, wewnątrz populacji, a także pomiędzy tymi populacjami (a co najmniej - w obszarze IBD). Wszystkie osobniki w danej próbce są równo oddalone od innych osobników z tej samej próbki, jak tego oczekujemy w przypadku populacji panmiktycznej. Bliskość genetyczna pomiędzy dwiema próbkami nie jest koniecznie powodowana niskim zróżnicowaniem względem najbliższej wspólnej populacji ancestralnej, lecz bardziej intensywnym przepływem genów. W celu zrekonstruowania historii demograficznej należy przeanalizować filogenezę nie-rekombinujących loci (drzewa genów) lub zastosować metody oparte na koalescencji lub pochodne od koalescencji. Jednakże drzewo gatunków oparte na metodzie koalescencji niewiele nam mówi o historii demograficznej, ponieważ zakłada ono brak przepływu genów. Dla danych RADseq najlepsze są metody, wykorzystujące spektrum częstotliwości wspólnego miejsca (JSFS), jak na przykład w programie „dadi”.

- Koegzystencja w ramach sympatryczności

Zaskoczyło mnie to, że obserwacja koegzystencji osobników z różnych klastrow genetycznych w tej samej próbce nie została szerzej omówiona. Zgadzam się z tym, że jest to za każdym razem jeden osobnik, ale wielkości próbek są nieznaczące. Informacja, że nie są one wynikiem błędów manipulacyjnych, jest istotna, ponieważ wspierają one hipotezę linii genetycznych izolowanych reprodukcyjnie (należy jednak zauważyć, że nie są one w stanie wyjaśnić wartości F_{is}). Z mojego punktu widzenia, jeżeli występują trzy kryptyczne, częściowo izolowane linie genetyczne we wschodniej części Morza Śródziemnego, to te trzy osobniki powinny zostać dokładnie omówione.

- Skanowanie genomu

Preferowałbym tu raczej wykonanie skanów F_{st} pomiędzy zidentyfikowanymi klastrami genetycznymi, a nie wykorzystanie tej niezwyklej dekompozycji według całościowego F_{st} bloku. Różnica pomiędzy Rysunkiem 5b a 5c jest niczym innym, jak przedstawieniem wizualnym tego, że autorka Pracy wybrała loci, dla których stosunek wewnątrzpopulacyjnych odchyleń genetycznych do międzypopulacyjnych jest silniejszy, stąd gałęzie wewnętrzne są

dłuższe na Rysunku 5c, niż na Rysunku 5b. Z tej procedury nie można niczego wydedukować. To, co jest tu interesujące, to loci ukazujące skrajny poziom zróżnicowania i to, jaki jest ich wzorzec struktury genetycznej. Jeżeli wśród trzech izolowanych lini wykryto wysoki stopień wartości odstających (Rysunek 5c), to jest to argument, świadczący za ledwie o ich istnieniu a powinno to zostać przebadane bardziej dokładnie. Jak to zostało wyżej powiedziane, interesujące byłoby także przeskanowanie pod kątem nierównowagi mito-nuklearnej.

Pomimo tych kilku zastrzeżeń, które uważam za wynikłe z braku czasu na przeanalizowanie tak rozległych zestawów danych, wygenerowanych przez NGS, powtarzam, iż ocena pracy doktorskiej przedstawiona w tej recenzji jest bardzo dobra. Chciałbym także podkreślić, iż nie mam zastrzeżeń co do porównania struktur genetycznych stwierdzonych u sercówek *C. glaucum* i *C. edule*. Uważam, że to bardzo wyraziste zróżnicowanie genetyczne u gatunku lagunowego w porównaniu z gatunkiem występującym wzdłuż wybrzeży i to, że nie wyjaśnia tego kryptyczny podział jest ważną informacją dla osób, zainteresowanych łącznością genetyczną pomiędzy populacjami organizmów morskich.

W konkluzji uważam rozprawę doktorską pani Ludmiły Sromek za doskonałą. Uznaję tę Rozprawę i wyrażam zgodę na przeprowadzenie obrony tejże pracy.

Nicolas Bierne

/-/ podpis nieczytelny

Institut des Sciences de l'Evolution de Montpellier (ISEM) - Université Montpellier 2 - Place Eugène Bataillon
Case courrier 065 - 34095 MONTPELLIER cedex 5, Francja

Code UMR ISEM, UMR 5554 Université de Montpellier - CNRS - IRD - CIRAD - EPHE

Tél. (3 3) 04 67 46 33 75, Fax (33) 04 67 46 33 99, email : nicolas.bierne@univ-montp2.fr

.....
Ja, tłumacz przysięgły języka angielskiego, wpisany na listę tłumaczy przysięgłych pod numerem TP/4097/05, prowadzoną przez Ministra Sprawiedliwości, niniejszym podpisem i pieczęcią stwierdzam zgodność i poprawność tłumaczenia przedstawionego mi dokumentu (-ów) sporządzonego (-ych) w języku angielskim.

.....
Dr Bolesław Borzycki

Gdynia, 2018-01-23

Repertorium Nr

67/2018

Lingua Care International, Cross Cultural & Language Services, Ul. Świętojańska 49/9, 81-391 Gdynia, Poland,
Tel: 620 62 87 Fax: 661 68 29 E-mail: borzycki@lingua-care.home.pl, www.lingua-care.home.pl