Thesis review prepered by Krystyna Nadachowska-Brzyska
Thesis title: *The role of hybridization in the evolutionary response to environmental change in the genus Canis*
Thesis written by Roya Adavoudi Jolfaei

General assessment

This thesis explores the roles of hybridization and introgression within the genus Canis, focusing on free-ranging dogs, Eurasian gray wolves, and golden jackals. The PhD candidate uses genome-wide SNP data to assess evolutionary relationships, hybridization rates, and introgression patterns, as well as their ecological and evolutionary consequences. The results are presented in four chapters, one of which has already been published, and three of which have been prepared as manuscripts. The thesis progresses logically through five clearly defined objectives, ranging from the potential consequences of hybridization to the evaluation of methods, the adaptive significance, and the influence of environmental variables. The first chapter, a systematic review, successfully sets the stage for the subsequent three chapters.

The thesis's main original contribution is the inclusion of golden jackal samples in the wolf and dog hybridization and introgression analysis. This approach offers new insights into interspecies interactions and expands our understanding of hybridization and introgression among Canis species. The thesis's strengths include its impressive sample sizes and broad methodological scope, ranging from classical genomic approaches and selection scans to environmental association analyses. The candidate clearly grasped a wide range of skills from the "scientific toolbox," including literature searches, systematic reviews, bioinformatics, genomic analyses, and scientific writing. The candidate has demonstrated broad knowledge of topics directly related to her research questions. Additionally, I would like to congratulate the candidate on including code to ensure the reproducibility of the results.

That said, the thesis is not without some weaknesses. After a strong general introduction, chapters are more difficult to follow. As a reviewer, I sometimes got lost in the details provided because they lacked much needed clarity in terms of structure, well-described tables, and readable figures. Due to this I may have missed some points (or message). The text also feels somewhat inflated. The three empirical chapters are based on the same dataset. Chapters 3 and 4, in particular, could easily be combined into a single, coherent manuscript. Chapter 4 builds directly on the results of Chapter 3, but the additional analyses are not entirely convincing. In my opinion, this makes the division into four chapters unnecessary. Additionally, many of the novel results are based on low-quality samples, so the interpretation of such data must be approached with extra caution.

Despite this critique, **I am convinced that the thesis written by Roya Adavoudi Jolfaei meets the criteria for PhD candidate** pursuant to art.187 of Act of 20 July 2018 The Law on Higher

Education and Science (Journal of Laws of 2018, item1668, as amended) and my evaluation is positive.

Detailed comments

General introduction

The general introduction is clear. It starts with much-needed definitions of hybridization and introgression and continues with a well-thought-out overview of the literature, ranging from historical background to state-of-the-art approaches developed and updated during the molecular methods revolution. Although brief, the introduction adequately sets the stage for the work and presents clear aims.

Chapter 1

This chapter presents a systematic review of the consequences of hybridization and is based on an already published work. It nicely sets the stage for the subsequent three chapters. However, given the problematic status of MDPI journals, I think the choice of journal could have been better. **I have one question arising out of my curiosity: How do mammalian hybridization rates compare with those of other vertebrate groups?**

Chapter 2

This chapter is lengthy and contains many technical details. This made it difficult at times to follow the main line of reasoning. Reducing or moving the technical details to supplementary material would improve readability and strengthen the narrative, especially if the chapter is prepared for publication later on.

Introduction

The chapter begins by classifying methods into six categories, such as sequence similarity and clinal changes. However, these categories are not revisited when discussing global versus local ancestry inference or in the subsequent evaluation using Canis data. Integrating this classification throughout the text — or alternatively, omitting it and keeping only the local and global ancestry perspectives — would provide better coherence.

Global inference methods were presented as methods that "provide an overview of admixture proportions across the entire genome, but lack the resolution to detect admixture at finer scales, such as along individual chromosomes or at specific loci." **In my opinion, this is not entirely true, since both admixture and principal component analysis (PCA) methods can be used with a subset of data (e.g., chromosomes or parts of chromosomes).** Such an analysis can

easily identify local differences in structure, e.g., local PCA. However, I admit that this method is probably more widely used to identify structural variation within the genome.

Methods

SNP genotyping The rationale for relaxing strict thresholds in SNP genotyping is understandable; however**, it remains unclear whether the applied thresholds are in line with common practice**. For example, how many samples would have been excluded under stricter filtering? Furthermore, the manuscript alternates between reporting thresholds as "82" and "0.82," which introduces confusion and should be standardized.

Regional datasets **The reasoning behind extracting regional datasets was not entirely clear for me**. The DAPC results show India forming a separate cluster—was this the main justification? The text also states that the choice of two regions aimed to test the influence of genotyping quality and sample size. However, since the two regions differ in sample size, quality, and substructure, it is difficult to disentangle the effects of each of the three factors. A more controlled approach — e.g., subsampling the same region to create datasets with different sample sizes and/or including vs. excluding lower-quality samples—would allow these influences to be tested separately**. Would it be possible here?**

LAMP-LD I was surprised by the fact that using a high density recombination map made it impossible to obtain reliable admixture results in LAMP-LD. **It would be helpful to clarify what "reliable" means in this context and to discuss possible explanations for this outcome.**

Results

The candidate reports 25 putative F1 hybrids between wolves and dogs, but they are not clearly visible in Figure 2.3. Plotting the hybrids together or marking them explicitly would make the results easier to interpret. The same suggestion applies to other ADMIXTURE plots.

Identifying hybrids in PCA plots appears to be based solely on visual inspection, which can be subjective, especially when points are dispersed (e.g., Figure 2.4, blue dots). **Would a more formal clustering method applied to the PCA results be effective in this context? What are the standards in the field?**

Figures 2.5, 2.7, and 2.9 are good, although they are a bit small. It would be great if the plots also gave information on the estimated ancestry proportions and showed the F1 hybrids (or indicated them). Ideally, both local ancestry methods would include the same set of individuals to enable direct comparisons of the methods' performance and possible differences.

Discussion

The author correctly emphasizes the need to interpret PCA results with caution and to view PCA primarily as an initial exploratory tool. However, the text presents this recommendation as if it

were a novel conclusion. The limitations of interpreting PCA results have been well documented and criticized for quite some time. **It would be more accurate to view this as a reinforcement of existing knowledge rather than a new insight. But maybe we did lack proper testing in the context of hybridization and introgression?**

The author emphasizes that individuals with low-quality SNP genotypes may confound ancestry analysis and offers valuable recommendations. This is definitely not trivial, given the potential problems with arrays mentioned by the author. Nevertheless, an important practical question arises: **To what extent can data quality be reduced while still allowing for the reliable identification of hybrids? Could the author suggest a quality threshold below which results are likely to become misleading? Could this approach be an alternative to excluding lower-quality individuals after analysis? Removing low-quality data from the beginning may underestimate the number of hybrids, but could including low-quality data influence the results obtained for high-quality samples or clustering? What does the candidate think?**

Finally, while the importance of population substructure is acknowledged, the effects are difficult to disentangle given the study design. As mentioned above, the selected regional datasets differ in both sample size and data quality. **The specific contributions of substructure versus sampling or genotyping effects remain unclear. Clearly separating these factors would help clarify the interpretation of the results.**

Chapter 3

Introduction

After a methodological investigations the candidates moves towards investigation of evolutionary consequences of hybridization in grey wolves, golden jackals and domestic dogs The introduction offers a clear overview of hybridization and adaptive introgression, setting the stage for the investigation. While the main aims of the work are well presented, they currently appear twice (in a shorter and longer paragraph) - likely due to an editing mistake.

Methods

The size of the datasets used in this study is impressive. I am particularly interested in seeing the results derived from the whole-genome data. Overall, the methodological framework appears adequate.

However, I have some **concerns about the ELAI analysis**. The researcher chose to include only ancestry blocks containing at least ten consecutive SNPs to minimize false positives. **While this is an understandable decision, could we miss some blocks under recent selection that swept the variation this way? For example, what if these 10 SNPs are far apart and in a region with low variation? What are the PhD candidate's thoughts on this?**

Figures

Figures 3.3–3.6 contain many individual plots. Comparing results would be much easier if they were presented as whole-genome plots rather than broken down as they are now. Additionally, the axes are not consistently scaled, which creates the impression that the averages were computed per chromosome. Standardizing the scales would improve interpretability.

Discussion

I appreciate that the author pays close attention to data quality and to potential biases associated with the use of arrays, and that these considerations are carefully integrated into the interpretation of the results. **I was, however, somewhat surprised by the limited overlap in chromosomal ancestry blocks compared to the findings of Pilot et al. (2021). The way it was written on page 116 made me also wonder if the overlap was between chromosomes or the same regions on the same chromosomes?**

General question

Selection scans combined with Gene Ontology (GO) enrichment analyses often yield extensive lists of candidate genes. **To what extent can these results be meaningfully interpreted without further validation? This is a common critique in the field, and I am curious how the candidate would engage in such a debate.** Additionally, were the genes identified as candidates for selection (adaptive introgression) found to include non-synonymous changes?

Chapter 4

Introduction

The first paragraph is difficult to follow because it contains many specific examples and only a few citations. For instance, the sentence "studies on plants showed that hybridisation can affect and be affected by organism–environment interactions" reads as if such evidence is exclusive to plants. The single supporting citation reinforces that impression. Broadening the scope and balancing the examples would make the introduction clearer.

Results

In many places (e.g. Figure 4.2), climatic variables such as *bio_13* are not explained. In some parts of the text these abbreviations are clarified, but in other places they are not, which makes it difficult to follow. In general, using bio abbreviations and long loci numbers in many places made it difficult to grasp the main message and made me go back and forth between several tables and plots.

The statement *"Four of the seven candidate loci were mapped within 100 kb of protein-coding genes"* raises some questions. **A distance of 100 kb is really large — how was this threshold**

chosen? What is known about gene density and recombination landscape in dogs that could justify such choice?

Of the four candidate loci discussed, two seem quite convincing: one contains the outlier SNP within a gene, and another lies only ~700 bp away. By contrast, the other two loci are much farther from outlier SNPs **- I am really not sure if I would call it a candidate**.

In addition, the part of the story describing the candidate genes feels somewhat unfinished. **For instance, does the DOCK1 gene itself show any signs of selection? Could this be tested more directly, for example using a McDonald–Kreitman test or by examining nonsynonymous variation among the species in question?**

Discussion

The author describes the functions of the genes in question from Table 4.5. The sentence reads: "The functions of these genes suggest potential roles in immune response, metabolic adaptation, and behavioral or neurological processes, which may explain their association with environmental variables and adaptive advantage in admixed wolves." **However, we cannot judge whether the genes truly suggest the proposed roles. Nor are we given an explanation as to why they may explain their association with adaptive potential.** Additionally, the results are mixed, and the discussion of olfactory receptor genes comes as a surprise in the third paragraph of this section (*Environmental associations of adaptive introgressed regions)*. Are they linked to any of the discussed genes?

The discussion suggests that the results indicate adaptive introgression would be particularly beneficial in human-modified landscapes, **but this seems more like storytelling than evidence-based interpretation to me**. In particular, there are only a few genes. There are also problems with sample size and genetic divergence (mentioned in the discussion).

In my opinion, this chapter cannot really stand alone. For future publication, I strongly suggest combining it with Chapter 3. Overall, however, this is a valuable piece of work that meaningfully contributes to our understanding of hybridization and introgression among Canis species. With revisions, particularly streamlining the presentation and moderating the interpretation, the manuscript(s) derived from this thesis have the potential to be good publication(s).

Krystyna Nadachowska-Brzyska