

The effect of environmental factors on adaptive genetic variation in grey wolves (*Canis lupus*) and free-ranging dogs (*Canis lupus familiaris*)

Doctoral Thesis

Francelly Martínez Sosa

THE EFFECT OF ENVIRONMENTAL FACTORS ON ADAPTIVE GENETIC VARIATION IN IN GREY WOLVES AND FREE-RANGING DOGS

Doctoral thesis

Francelly Martínez Sosa

The effect of environmental factors on adaptive genetic variation in grey wolves (*Canis lupus*) and free-ranging dogs (*Canis lupus familiaris*)

Francelly Martínez Sosa

Museum and Institute of Zoology

Polish Academy of Sciences

Warsaw, Poland



PhD Dissertation

Warsaw 2025

Table of Contents

Authors' Addresses 3
Abstract 4
Streszczenie 6
Acknowledgements 9
General introduction and summary of results 10
Chapter 1 | Molecular mechanisms underlying vertebrate adaptive evolution: a systematic review 15
Chapter 2 | Comparative analysis of population structure and diversifying selection patterns between Eurasian grey wolves (*Canis lupus*) and free-ranging dogs (*Canis lupus familiaris*) 50
Chapter 3 | Comparative genomics of local adaptive evolution in Eurasian grey wolves and free-ranging-dogs 81
Chapter 4 | South-north decline in adaptive potential in of free-ranging dogs points to dog domestication in southern Eurasia 126
Summary of results 156
General discussion and general conclusions 160
References 165

Author's and supervisor's affiliations

Author:

Francelly Martínez Sosa

Museum and Institute of Zoology

Polish Academy of Sciences

ul. Twarda 51/55, 00-818 Warsaw, Poland

Supervisor:

Małgorzata Pilot

Faculty of Biology

University of Gdańsk

ul. Wita Stwosza 59, 80-308 Gdańsk, Poland

Funding

The DNA extractions and genotyping resulting in the genome-wide SNP dataset used in this project were funded by the Polish National Science Centre grant SONATA BIS no. 2019/34/E/NZ8/00246 awarded to the thesis supervisor Dr. Małgorzata Pilot.

Abstract

Adaptive evolution is the process by which organisms change from generation to generation through genomic modifications to improve survival and reproductive rates under environmental selective pressures. Investigating this process allows for us to learn how organisms evolve to adapt to their enviroments. This thesis encompasses an investigation into the adaptive evolution of two representatives of the genus *Canis*, grey wolves and domestic dogs. Domestic dogs are the descendants of grey wolves and owe their ecological success in great part to their relationship with humans. Most of the dogs in the world (i.e., ~ 1 billion), are free-ranging dogs (FRD) that live unconfined with minimal human assistance. Grey wolves and FRDs are closely related taxa with overlapping geographic ranges that are adapted to different habitats throughout their wide ranges. These canids also exploit different ecological niches, with grey wolves being apex predators and FRDs human commensals. Thus, these canids are good models for investigating molecular adaptations under selective pressures from varied habitats as well as adaptive potential under environmental shifts.

This thesis aimed to investigate whether grey wolves and FRDs show similar molecular adaptations under selective pressures within similar habitats, and whether they retain adaptive potential under climate change. To answer these questions I investigated genetic differentiation, genes under diversifying selection, genes associated with environmental variables and adaptive potential with SNP genotypes of 370 FRDs and 232 grey wolves from 10 different biomes across Eurasia.

Four different approaches employed to investigate population genetic structure consistently showed that FRDs have weaker genetic differentiation than grey wolves, implying a stronger connectivity among FRDs subpopulations. The stronger connectivity in FRDs has been likely facilitated by their relationship with humans. I also investigated genes under diversifying selection i.e., genes with variants associated with different phenotypes being favoured by natural selection in different subpopulations. The approaches applied for this purpose were based on detection of outliers using the measures of genetic differentiation among populations; hence the approaches are influenced by population structure. There was a strong overlap among the genes with signatures of diversifying selection in FRDs and wolves. Thus, FRDs and wolves have evolved under selective pressures that favour many of the same phenotypic traits (e.g., olfaction, morphological traits).

To investigate the genes with signatures of local adaptation I used a dual approach to detect SNPs associated with environmental variables. The approaches detect high correlations between SNPs and environmental variables, while accounting for population structure. There was little overlap among the genes with signatures of local adaptation in grey wolves and FRDs. However, the phenotypic traits associated with these genes were similar in both taxa. Furthermore, the genes under diversifying selection and local adaptive evolution showed overlap in FRDs, but not in grey wolves. The difference among the genes with signatures of diversifying selection and with signatures of local adaptive variation, could be due the presence of genes that are under other types of selection (e.g., balancing), population genetic factors (e.g., genetic drift,

historic bottlenecks) and environmental gradients. These results may indicate that both taxa have unique genetic mechanisms enabling adaptation to their local enviroments.

Dogs are human commensals and therefore it may be predicted that they are not well adapted to natural environments. However, dogs may retain an adaptive potential for the natural landscapes in which their ancestral wild population lived. To that end, I investigated the adaptive potential of domestic dogs and attempted to identify environments where they are best suited to survive with minimal human assistance. The combined analysis of dogs' adaptive potential and suitable ecological niche modelling contributed evidence into the location of geographical centre of dog domestication.

Based on genomic offset and risk of maladaptation analyses, FRDs are predicted to retain lower adaptive potential than grey wolves under climate change. Populations of FRDs in Southern Eurasia are predicted to retain the highest adaptive potential under climate change. If dogs have retained the adaptive potential from the ancestral population that underwent domestication, our results suggest that dogs may have been domesticated in Southern Eurasia. The modelling of the most suitable habitats based on modern dogs' distribution showed Southern Eurasia as the geographical region with the most suitable habitats for dogs. The combined findings of adaptive potential and the most suitable habitat for dogs suggest that dogs were domesticated in Southern Eurasia, although it is not possible to point to a precise location of the domestication centre for the species.

This work shows the complex dynamics of selective pressures in two widespread canid species with different ecological niches. Our findings may be useful in informing conservation efforts for grey wolves and population control of FRDs. Our results also suggest that local adaptation of closely related taxa may involve similar phenotypic traits but not necessarily the same genes. Furthermore, this work contributes knowledge to reduce the uncertainty surrounding the dog domestication place.

Streszczenie

Ewolucja adaptacyjna to proces, w którym organizmy zmieniają się z pokolenia na pokolenie poprzez modyfikacje ich genomów, aby poprawić przeżywalność i wskaźniki reprodukcji pod presją selekcyjną środowiska. Badanie tego procesu pozwala nam dowiedzieć się, w jaki sposób organizmy ewoluują, aby przystosować się do swojego środowiska. Praca ta obejmuje badania nad ewolucją adaptacyjną dwóch przedstawicieli rodzaju *Canis*, wilków szarych i psów domowych. Psy domowe są potomkami wilków szarych i swój sukces ekologiczny zawdzięczają w dużej mierze relacjom z ludźmi. Większość psów na świecie (tj. ~ 1 miliard) to psy żyjące na wolności (FRD), które żyją bez ograniczeń przy minimalnej pomocy człowieka. Wilki szare i FRD są blisko spokrewnionymi taksonami o nakładających się zasięgach geograficznych i przystosowanymi do różnych siedlisk w obrębie swoich szerokich zasięgów. Te psowate wykorzystują również różne nisze ekologiczne, przy czym szare wilki są drapieżnikami szczytowymi, a FRD ludzkimi komensalami. W związku z tym te psowate są dobrymi modelami do badania adaptacji molekularnych pod presją selekcyjną z różnych siedlisk, a także potencjału adaptacyjnego w warunkach zmian środowiskowych.

Celem tej pracy było zbadanie, czy wilki szare i FRD wykazują podobne adaptacje molekularne pod presją selekcyjną w obrębie podobnych siedlisk i czy zachowują potencjał adaptacyjny w obliczu zmian klimatu. Aby odpowiedzieć na te pytania, zbadałam zróżnicowanie genetyczne, geny podlegające różnicującemu doborowi naturalnemu, geny związane ze zmiennymi środowiskowymi, oraz potencjał adaptacyjny za pomocą genotypów SNP dla 370 FRD i 232 wilków szarych z 10 różnych biomów w obrębie Eurazji.

Cztery różne podejścia wykorzystane do zbadania genetycznej struktury populacji wykazały zgodnie, że FRD mają słabsze zróżnicowanie genetyczne niż wilki, co sugeruje silniejszą łączność między subpopulacjami FRD. Silniejsza łączność w FRD jest prawdopodobnie ułatwiana przez ich relacje z ludźmi. Badałem również geny pod wpływem doboru różnicującego, tj. geny z wariantami związanymi z różnymi fenotypami, które są faworyzowane przez dobór naturalny w różnych subpopulacjach. Podejścia wykorzystywane w tym celu opierały się na wykrywaniu skrajnych wartości za pomocą miar zróżnicowania genetycznego między populacjami; w związku z tym na podejścia te ma wpływ struktura populacji. Stwierdzono dużą proporcję wspólnych genów z oznakami doboru różnicującego u FRD i wilków. W związku z tym FRD i wilki ewoluowały pod presją selekcyjną, która faworyzuje wiele z tych samych cech fenotypowych (np. węch, cechy morfologiczne).

Aby zbadać geny z sygnaturami lokalnej adaptacji, zastosowałem podwójne podejście do wykrywania markerów SNP związanych ze zmiennymi środowiskowymi. Podejścia te wykrywają wysokie korelacje między SNP a zmiennymi środowiskowymi, uwzględniając jednocześnie strukturę populacji. Proporcja genów z sygnaturami lokalnej adaptacji wspólnych dla wilków szarych i FRD była niewielka. Jednak cechy fenotypowe związane z tymi genami były podobne u obu taksonów. Co więcej, geny podlegające selekcji różnicującej i lokalnej ewolucji adaptacyjnej pokrywały się w

znacznym stopniu u FRD, lecz nie u wilków. Różnica między genami z sygnaturami doboru różnicującego i z sygnaturami lokalnej zmienności adaptacyjnej może wynikać z obecności genów, które podlegają innym typom selekcji (np. równoważącej), czynników populacyjno-genetycznych (np. dryf genetyczny, historyczne wąskie gardła) i gradientów środowiskowych. Wyniki te mogą wskazywać, że oba taksony mają unikalne mechanizmy genetyczne umozliwiające przystosowanie się do lokalnego środowiska.

Psy są komensalami człowieka i dlatego można przewidzieć, że nie są dobrze przystosowane do naturalnych środowisk. Jednak psy mogły zachować potencjał adaptacyjny do naturalnych krajobrazów, w których żyła dzika populacja ich przodków. W tym celu zbadałam potencjał adaptacyjny psów domowych i podjęłam próbę zidentyfikowania środowisk, w których są one najlepiej przystosowane do przetrwania przy minimalnej pomocy człowieka. Połączona Analiza potencjału adaptacyjnego psów i połączona z modelowaniem nisz ekologicznych odpowiednich dla psów dostarczyły dowodów na położenie geograficznego centrum udomowienia psa.

Na podstawie analiz kompensacji genomowej i ryzyka niedostosowania przewiduje się, że w warunkach zmian klimatu FRD zachowają niższy potencjał adaptacyjny niż wilki. Przewiduje się, że populacje FRD w południowej Eurazji zachowają najwyższy potencjał adaptacyjny w obliczu zmian klimatu. Jeśli psy zachowały potencjał adaptacyjny populacji przodków, która została udomowiona, nasze wyniki sugerują, że psy mogły zostać udomowione w południowej Eurazji. Modelowanie najbardziej odpowiednich siedlisk w oparciu o rozmieszczenie współczesnych psów wykazało, że południowa Eurazja jest regionem geograficznym o najbardziej odpowiednich siedliskach dla psów. Połączone wyniki dotyczące potencjału adaptacyjnego i najbardziej odpowiednich siedlisk dla psów sugerują, że psy zostały udomowione w południowej Eurazji, chociaż nie jest możliwe wskazanie dokładnej lokalizacji centrum udomowienia tego gatunku.

Praca ta pokazuje złożoną dynamikę presji selekcyjnych u dwóch szeroko rozpowszechnionych gatunkach psowatych o różnych niszach ekologicznych. Nasze wyniki mogą dostarczyć przydatnych informacji dla działań ochronnych w odniesieniu do wilków i dla kontroli populacji wolno-żyjących psów. Nasze wyniki sugerują również, że lokalna adaptacja u blisko spokrewnionych taksonów może obejmować podobne cechy fenotypowe, lecz niekoniecznie te same geny. Co więcej, praca ta przyczynia się do zmniejszenia niepewności co do miejsca udomowienia psa.

Acknowledgements

I have wanted to work with canids since I was 4 years old and while pursuing my bachelor's degree, I became fascinated with genomics. Thus, for me working on this project was truly a dream come true. The best lesson this experience taught me was to move through feeling uncomfortable. Working in bioinformatics, a field I was unfamiliar with, pushed me out of my "laboratory work" comfort zone. I now enjoy working pipelines, writing scripts and visualizing data. Moving forward I will take that lesson with me and try new things even though it can be intimidating to do so. Because on the other side of it I may find a new passion and learn more about myself.

Achieving this work would not have been possible without the support of my family. My daughter Areyza has been my youngest teammate, making me laugh and teaching me to enjoy the little moments. Learning to be her mom while working towards a PhD would not have been possible without her amazing personality and her incredible dad, my husband Francisco. My husband was the first to support my desire to pursue my degree and has been patient and supportive when I had to be away for months or been fused to my computer to complete my work. My father Frank, along with my sisters, Lycelli and Mia Maricelly, were always sure I would achieve this, even when my confidence faltered. Even though she is no longer with us, I hope my mother Eyra, is proud and happy to see me realize our shared dream. Finally, my Abuela Mimi who has been the best teacher on what unconditional love is, even when she didn't understand my choices she always rooted for me.

This project would not have been possible without the Evolutionary Genomics research team. My supervisor Dr. Pilot gave me the opportunity to work in this team, and I truly appreciate both the opportunity and her guidance throughout this process. I am also thankful to my now friends Karolina, Milomir, Alejandro and Morgane who were willing to teach me the basics and helped feel at home in Poland even with the pandemic restrictions. I am especially thankful to Roya and Carolina, without whom I would not have been able to complete this project. Roya has not only been my friend but shared all the laboratory work with me and helped me make sense of things when I first started to work in bioinformatics. Carolina has been an amazing friend and teacher in the most supportive way, offering me thoughtful and useful insights whenever I needed them.

I would also like to acknowledge our collaborators whose contributions were essential for the success of this project; Wieslaw Bogdanowicz, Srinivas Yellapu, Abdulaziz N. Alagaili, Elena Bykova, Duško Ćirović, Silviu Chirac, Mihajla Djan, Sergey Gashev, Vahram Hayrapetyan, Kanstantsin Homel, Yadvendradev Vikramsinh Jhala, Natia Kopaliani, Nikolay V. Mamaev, Vladislav Molchan, Carmela Musto, Henryk Okarma, Innokentiy M. Okhlopkov, Stéphane Ostrowski, Massimo Scandura, Dragana Šnjegota, Elena Tsingarska, Anatoliy M. Volok, Maria Yakovleva and Four Paws Ukraine.

Finally, I am thankful to the BioPlanet Doctoral School at the Museum and Institute of Zoology of the Polish Academy of Sciences. Especially, to the doctoral school director, Dr. Chylarecki, the professors and my fellow students. I would also like to acknowledge the BioPlanet Doctoral School for funding my stipend. Additionally, I am thankful to the administrative personnel at the Stacja Ornitologiczna, particularly Mr. Rohde who

helped me in many of the logistics as a foreign student in Poland. I became a mom in the middle of my PhD journey, and they were all very supportive, understanding and accommodating whenever an issue arose. Being both a PhD student and a new mom would not have been possible without their kindness and empathy. I have learned so much on this journey not only about canids, but also about how amazing people in this community can be.

General introduction

This thesis presents a comparative genomics study on two closely related canids grey wolves and domestic dogs, which occur in sympatry in the northern hemisphere. The overachieving aim of this study was to understand similarities and differences in adaptive strategies between a wild and a domestic canid. My findings also provide insights into adaptive potential for both wolves and free ranging dogs for adapting to climate change. Finally, this thesis provides insights into dogs' domestication origin, and the adaptations free-ranging dogs populations have evolved to survive with minimal human assistance. This involved testing three predictions: (I) Both species show the strongest genetic differentiation in a north - south axis consistent with habitat differentiation which occurs strongest in this axis; (II) Genes with signatures of selection in both species have functions related to morphology, immunity, coat-coloration and sensory perception based on previous findings and the role of these types of genes in interacting with their enviroments. And (III) free-ranging dogs will show the differences in adaptive potential between habitats, the habitats for which they show the highest adaptive potential, corresponds to the one they were first domesticated. This work contributes to our understanding of the mechanisms involved in local adaptive evolutionary processes, domestication of dogs and wolf conservation management.

Comparative genomics among a domesticated and wild species

Comparative genomics is an effective tool for understanding how wild and domestic populations are genetically differentiated. This knowledge has applications in understanding evolutionary trajectories of wild and free ranging domesticated populations. The systematic review of the molecular mechanisms underlying adaptive evolution I carried out in Chapter 1 (Martinez-Sosa & Pilot 2023), showed that adaptive evolution is not only driven by natural selection but also gene flow, genetic drift, epigenetic changes and pseudogenization. The underpinnings of adaptive evolution are complex not only because most phenotypic traits are polygenic, but also because gene expression is highly regulated by molecular mechanisms. Molecular mechanisms which regulate gene expression include but are not limited to changes in expression of transcription factors, chromatin remodelling modifying enzyme access to genomic elements which impact transcription (e.g., enhancers, silencers) or inhibition of transcription by expressing RNA molecules that interfere with the process (e.g., microRNA).

Adaptations can be context-dependant since the environmental pressures each population experiences vary among populations or habitats, i.e., a trait can be beneficial in one habitat and detrimental in another (Lenormand, 2002). This is particularly true for widespread species, as their ranges subject them to a myriad of environmental conditions. Local adaptive evolution is the evolutionary process in which selection operates within a population in the context of this population's immediate habitat. Regulation of gene expression and cellular process have emerged as essential mechanisms for the evolution of favourable traits under local environmental pressures (Chapter 1). Widespread ecologically successful lineages such as canids present an opportunity to investigate adaptive evolutionary process, particularly local adaptations.

Canids represent one of the most successful mammalian lineages. Their adaptiveness is evident due to their 56 extant species with ranges that cover all continents except Antarctica, and an evolutionary history of successful colonization of new environments (Porto et al., 2023). The ecological success of the lineage is owed in part to its genetic variability and phenotypic plasticity. The Canidae lineage exhibits a wide range of morphological variation among species. Canidae species have adaptations that increase their survival and reproductive success in their respective habitats (Machado, 2020; Tangredi & Lawler, 2024) Our study is part of a broader effort to understand genetic differentiation and the underlying factors driving adaptation in canid species. Canids are unique among carnivores as they are widespread, have diverse trophic roles and most species live in social groups (Flemming et al., 2017). Despite their ecological success, ranges of various canid species have been reduced due conflicts with humans (Chavez, et al., 2022; Torres et al., 2020). Thus, insights in canid evolutionary dynamics broadens our understanding into how carnivores undergo adaptive evolution, furthermore it has applications for managing both native and invasive populations.

Study species: Eurasian grey wolves (*Canis lupus*) and domestic dogs (*Canis familiaris*)

This study is focused on two of the most ecologically successful canid taxa, grey wolves and domestic dogs. Grey wolves have a wide distribution across the northern hemisphere, which highlights their adaptability (Gonzalez Bernal et al., 2022; Wolf & Ripple, 2017). In grey wolves, habitat differentiation is one of the main drivers of their genetic differentiation (Muñoz-Fuentes et al., 2009; Pilot et al., 2006, Leonard et al. 2013). Thus, selection under local environmental pressures has been proposed as the main factor driving their adaptive processes. Ecological specialization is evidenced by the presence of different ecotypes among grey wolf populations (Musiani & Randi, 2024).

Habitat differentiation (i.e., climate variation and prey availability) has been proposed as the determinant of phenotypic variation among ecotypes (Hendricks, et al., 2019; Randi, 2011). Body size in grey wolves follow Bergmann's Rule, i.e. it increases from the southern to the northern parts of the species range (Musiani & Randi, 2024). Prey availability has operated as a selective pressure in skull morphology among ecotypes (Musiani & Randi, 2024). Grey wolves are hypercarnivores, and their diet is determined by diversity and density of ungulates in their habitats; they show a preference for wild ungulates but consume domestic species when wild prey density is low (Janeiro-Otero et al., 2020; Newsome et al., 2016). Grey wolf adaptability is underlined by morphological and immune adaptations which have played a pivotal role in their evolutionary history (Pilot et al., 2014; Schweizer et al. 2016).

Dogs are a domesticated evolutionary linage descendent from grey wolves. Domesticated dogs are human commensals, and their diet is mainly composed of human-derived food. (reviewed in Tancredi & Cardinali, 2023) Most of the contemporary dog populations i.e., ~75% of over 1 billion individuals, are free-ranging dogs (FRDs) that survive with minimal human assistance but typically depend on anthropogenic food sources (Gompper, 2014). Isolation by distance and anthropogenic factors have been described as the main drivers of genetic differentiation among dog

populations (Perri et al., 2021; Shannon et al., 2015; Vychodilova et al., 2018). However, FRDs' ecology is not well known, with most studies focusing on their behavioral ecology (Lazarroni et al., 2020; Range & Marshall-Pescini, 2022; Wheat & Wynne, 2025). Domestic dog populations can invade natural landscapes and threaten wildlife as competitors for resources, transmitters of disease (Huges & Macdonald, 2013; Marshall-Pescini, 2022; Mendoza Roldan & Otranto, 2023), and predators. Thus, it is important to gain a better understanding of how FRD populations have adapted to natural landscapes, to adequately assess their impact on these landscapes.

The domestic dog is the first animal subjected to artificial selection, i.e., humans acting as the selective pressure (Pierotti, 2024). Dog domestication emerged from a commensal relationship with humans (Losey et al., 2022; Pierotti, 2024), from a Eurasian grey wolf population (which is either extinct or changed its genetic composition due to admixture), ~40,000-16,000 years ago (Bergström et al. 2022, Tancredi & Cardinali, 2023). The region where dogs were first domesticated in has been suggested to be in Europe (Thalmann et al., 2013), Central Asia (Shannon et al., 2015), and Southeastern Asia (Wang et al., 2016). Alternatively, two independent events have been suggested in both Eastern and Western Eurasia (Franz et al., 2016). Due to the lack of knowledge of when precisely dogs were domesticated (Bergström et al. 2022, Tancredi & Cardinali, 2023), and admixture with other canids impacting their genetic composition (Bergström et al. 2022), a consensus has not been reached on the geographical centre of dog domestication.

Investigative approach

The overarching approach in this thesis is a comparative genomic study between two related taxa, grey wolves and FRDs, based on single nucleotide polymorphism (SNP) genotype and environmental data. The goals of this thesis are to (1) review the current knowledge on the molecular mechanisms underlying adaptive evolution in vertebrates in response to environmental selective pressures and identify the knowledge gaps, (2) explore patterns of genetic differentiation in Eurasian grey wolves and FRDs and add to our knowledge of the genes driving their evolutionary trajectories, (3) address a gap in our knowledge of adaptations in contemporary grey wolves and FRDs under similar environmental pressures and (4) assess the potential of both species to adapt to climate change in addition to offering new insights into the geographic region where dogs were first domesticated.

To achieve these goals, I analysed 73360 quality-controlled SNP markers from 232 grey wolves and 370 FRDs from across 10 different biomes. First, I estimated genetic differentiation and inferred populations clusters. Then I investigated whether the same genes showed signatures of selection in both taxa. To achieve this, I employed approaches to detect signatures of selection at two different scales. Broader patterns of selection were investigated by detection of genes with signatures of diversifying selection, a type of selection which favours different phenotypes in different parts of the species range and/or different environmental conditions. The approaches employed to detect loci with signatures of diversifying selection, identify the outlier loci based on genetic differentiation statistics, thus are influenced by patterns of genetic differentiation and not solely environmental selective pressures. To investigate

selection in the context of local environmental conditions, I investigated genes associated with environmental variables with approaches that account for genetic differentiation. I then used the SNPs with signatures of local adaption in each species to estimate adaptive potential and risk of maladaptiveness. In this work I found similarities in some patterns (e.g., genetic differentiation and genes with signatures of diversifying selection) alluding to the shared evolutionary history of grey wolves and domestic dogs. Conversely, I detected differences in local adaptations and adaptive potentials among these closely related taxa, implying differences in their adaptive strategies that may have contributed to their divergence. The findings from this study may have applications for management of protected grey wolf populations and free breeding populations of domestic dogs.

Chapter 1 | Molecular mechanisms underlying vertebrate adaptive evolution: A Systematic Review

genes

MDPI

Review

Molecular Mechanisms Underlying Vertebrate Adaptive Evolution: A Systematic Review

Francelly Martínez Sosa [1] and Małgorzata Pilot [1,2,*]

[1] Museum and Institute of Zoology, Polish Academy of Sciences, 80-680 Gdańsk, Poland
[2] Faculty of Biology, University of Gdańsk, 80-308 Gdańsk, Poland
* Correspondence: mpilot@miiz.waw.pl

Abstract: Adaptive evolution is a process in which variation that confers an evolutionary advantage in a specific environmental context arises and is propagated through a population. When investigating this process, researchers have mainly focused on describing advantageous phenotypes or putative advantageous genotypes. A recent increase in molecular data accessibility and technological advances has allowed researchers to go beyond description and to make inferences about the mechanisms underlying adaptive evolution. In this systematic review, we discuss articles from 2016 to 2022 that investigated or reviewed the molecular mechanisms underlying adaptive evolution in vertebrates in response to environmental variation. Regulatory elements within the genome and regulatory proteins involved in either gene expression or cellular pathways have been shown to play key roles in adaptive evolution in response to most of the discussed environmental factors. Gene losses were suggested to be associated with an adaptive response in some contexts. Future adaptive evolution research could benefit from more investigations focused on noncoding regions of the genome, gene regulation mechanisms, and gene losses potentially yielding advantageous phenotypes. Investigating how novel advantageous genotypes are conserved could also contribute to our knowledge of adaptive evolution.

Keywords: adaptive evolution; vertebrates; regulatory mechanisms; environmental adaptations; gene loss

Citation: Martínez Sosa, F.; Pilot, M. Molecular Mechanisms Underlying Vertebrate Adaptive Evolution: A Systematic Review. *Genes* **2023**, *14*, 416. https://doi.org/10.3390/genes14020416

Academic Editors: Jeffrey M. Kidd and Alexander S. Graphodatsky

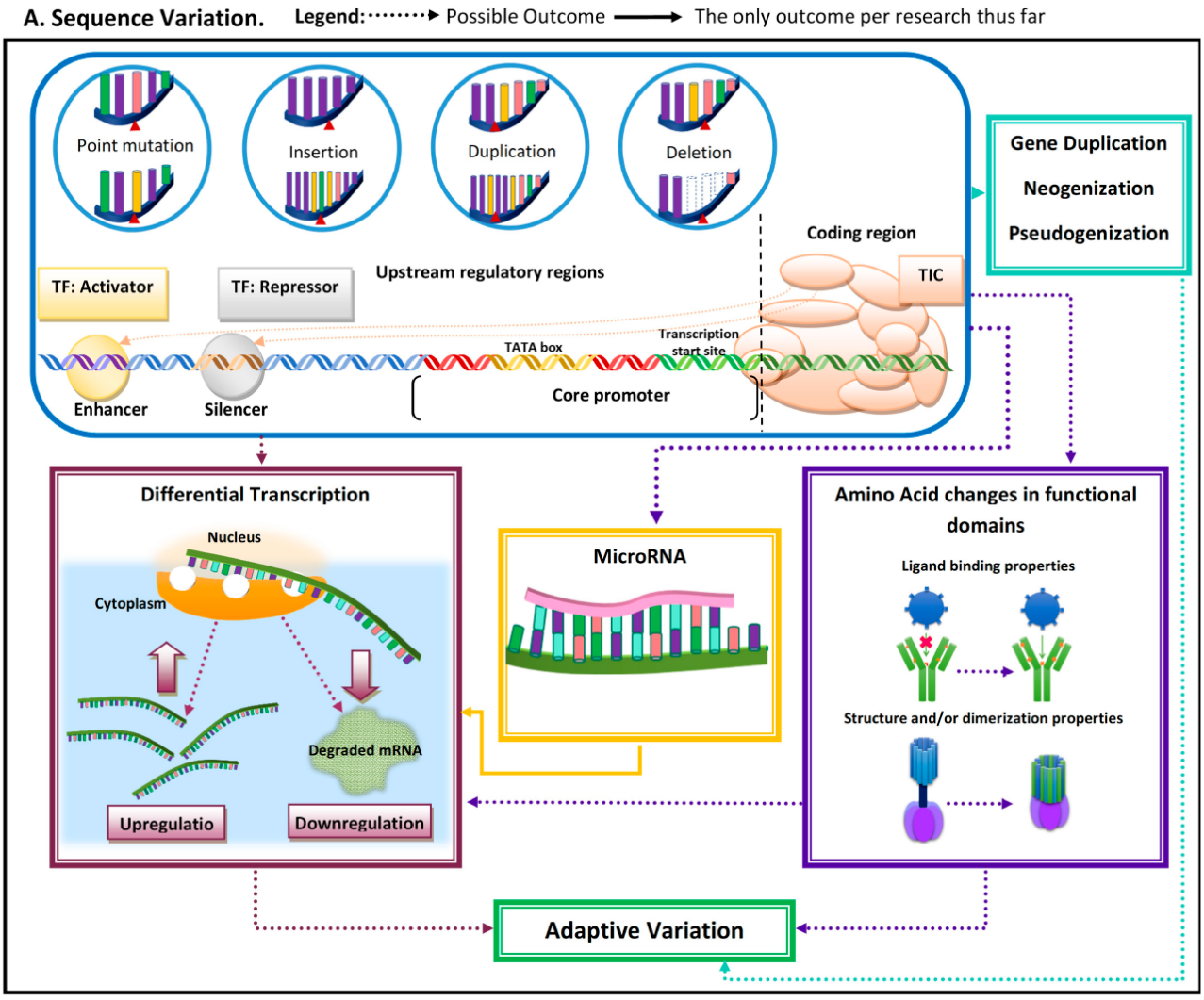
Received: 3 December 2022
Revised: 24 January 2023
Accepted: 1 February 2023
Published: 5 February 2023

Copyright: © 2023 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

Vertebrates colonized a broad range of habitats with varying ecological conditions (e.g., lighting conditions, oxygen content, and water salinity). This process was associated with adaptations to these varied environments. Adaptive evolution is a process by which advantageous phenotypes, i.e., traits which increase fitness, are propagated through positive selection or environmentally induced variation [1]. Advantageous traits are determined by evolutionary pressures and habitat dynamics; consequently, a phenotype that is advantageous in certain ecological contexts might be neutral or deleterious in others.

Adaptive evolutionary processes occur in two general steps: (1) variation arises and (2) advantageous phenotype is propagated and maintained in the population. Genetic variation arises from mutations that can occur due to single base changes, insertions, deletions, and duplications [2]. Mutations can have multiple implications, from pseudogenization to de novo emergence of genes and/or regulatory elements (Figure 1 and Table 1). Mutations can give rise to novel phenotypes, and the advantageous ones become propagated in a population [2]. Thus, functionally important genomic regions are assumed to be highly conserved and adaptive alleles undergo rapid fixation [3,4]. The conserved nature of these functionally important elements can be used to identify them as being adaptive.


A. Sequence Variation.
Legend: Possible Outcome
The only outcome per research thus far
Point mutation
Insertion
Duplication
Deletion
Gene Duplication
Neogenization
Pseudogenization
Coding region
TIC
TF: Activator
TF: Repressor
Upstream regulatory regions
TATA box
Transcription start site
Enhancer
Silencer
Core promoter
Differential Transcription
Nucleus
Cytoplasm
Degraded mRNA
Upregulatio
Downregulation
MicroRNA
Amino Acid changes in functional domains
Ligand binding properties
Structure and/or dimerization properties
Adaptive Variation

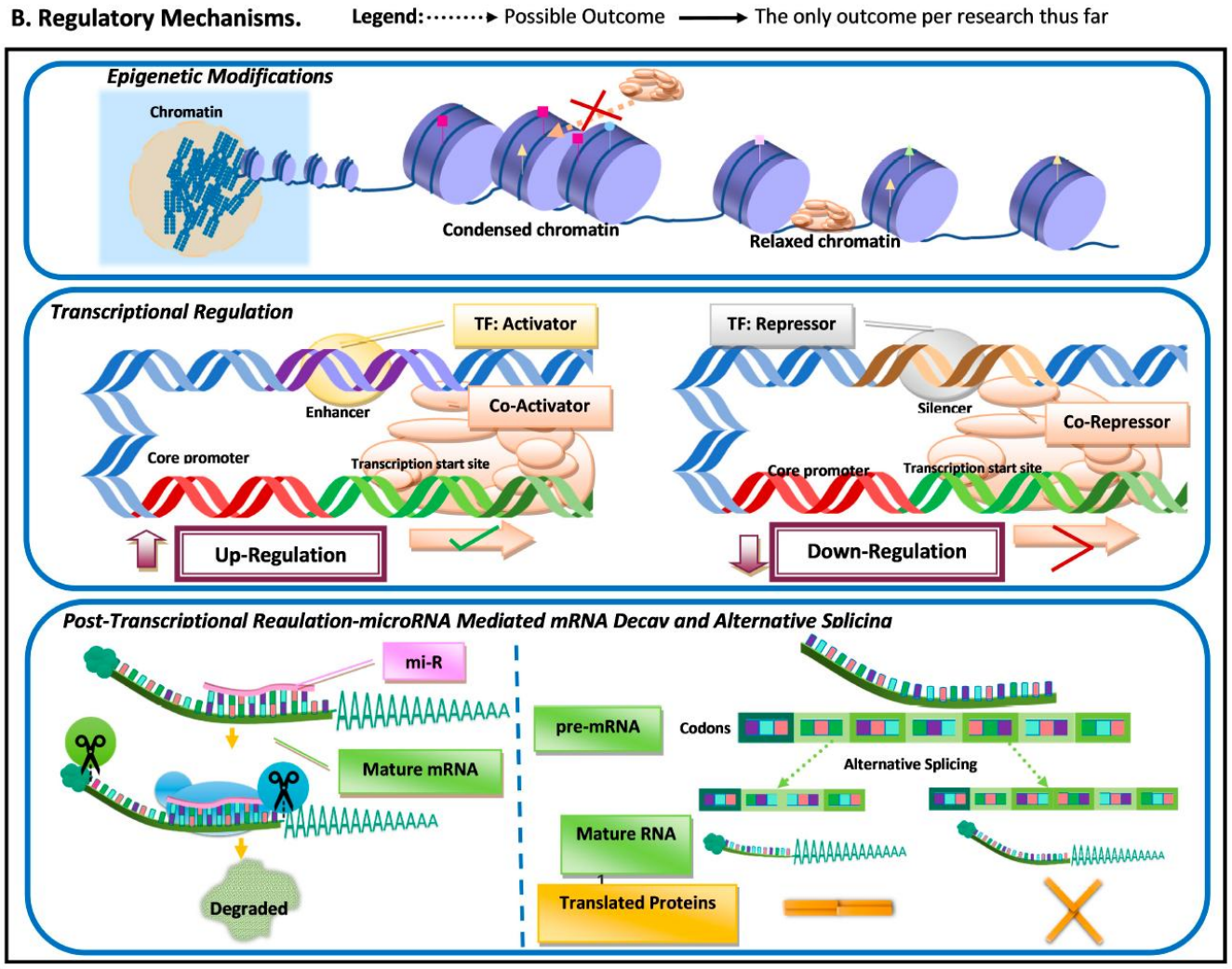


Figure 1. Schematic representation of the most common molecular mechanisms underlying adaptive variation in vertebrates. (**A**) Sequence variation. Sequence variation refers to any change in DNA sequence that commonly occurs due to DNA polymerase errors (e.g., slippage and erroneous excisions) and/or insertion or deletion of transposable elements. DNA sequence variation in conserved elements (i.e., regulatory elements and coding regions) is the most common mechanism underlying adaptive variation in vertebrates and can lead to gene duplication, neogenization, and pseudogenization. Sequence variation in coding regions can lead to amino acid changes at or near functional domains, which can impact protein function and/or structure. Additionally, sequence variation at the coding regions can also lead to the emergence of microRNA (miR) or modification in existing miRs. MiRs can hybridize with target mRNA and can lead to its decay, thus downregulating gene expression. Sequence variation in regulatory elements can lead to increases (upregulation) or decreases (downregulation) in transcription of a gene. (**B**). Regulatory mechanisms. The three most common regulatory mechanisms underlying adaptive variation in vertebrates described thus far are epigenetic modifications, transcriptional regulation, and post-transcriptional regulation. Epigenetic modifications are modifications to histones, the protein component of nucleosomes. These modifications can impact an RNA polymerase's ability to find core promoters as well the activity of transcription factors. Transcriptional regulation involves cis-regulatory elements and trans-regulatory elements, which impact the function of the transcription initiation complex (TIC). Post-transcriptional regulation involves elements that impact the messenger RNA transcript itself. One of these mechanisms is microRNA (mi-R)-mediated mRNA decay, a process that leads to the degradation and subsequent downregulation of said transcript. Another mechanism that can impact the translated protein structure and/or function is alternative splicing.

Table 1. Brief descriptions of the molecular mechanisms illustrated in Figure 1.

Concept	Brief Description
Point mutation	Change in DNA sequence by a single nucleotide; it can be an insertion, duplication, or deletion.
Insertion	DNA sequence variation by insertion of one or more nucleotides.
Duplication	Modification in DNA sequence by duplication of a stretch of nucleotides (i.e., one or more).
Deletion	Change in DNA sequence by excision of one or more nucleotides.
Upstream regulatory regions	Regions upstream (before) from the core promoter to which transcription factors (TF) or coactivators can cis-regulate. Transcription factors can activate or increase transcription when bound to enhancers, while repressors can decrease transcription when bound to silencers.
Core promoter	Regulatory region where the transcription initiation complex (TIC) binds, as it has the transcription start site and can include a thymidine adenine-rich region (i.e., TATA box) where the TIC is recruited.
Coding region	DNA sequence that codes for the pre-mRNA, which is later modified to the mRNA and translated into proteins.
Upregulation	Regulatory change that leads to an increase in transcription of specific mRNA.
Gene duplication	Refers to the duplication of a complete gene.
Neogenization	DNA sequence gains a novel gene function and can occur in previously non-coding regions, more commonly in duplicated genes.
Pseudogenization	Process by which a gene either loses a function or becomes completely non-functional.
Downregulation	Regulatory change which leads to a decrease in transcription of specific messenger RNA (mRNA).
microRNAs (miRs)	Small RNA fragments that can hybridize with target mRNA and can lead to its decay, thus downregulating it.
Epigenetic modifications	Molecular modifications (e.g., acetylation, methylation, and phosphorylation) to histones, proteins involved in the packing of DNA. These modifications can impact chromatin condensation and/or RNA polymerase activity.
Chromatin	A protein–DNA molecular complex that is the natural state of Eukaryotic genetic material.
Transcriptional regulation	Processes that impact the rate of transcription, either upregulating or downregulating; these can be cis-regulatory elements (e.g., enhancers or silencers) or components of the transcription initiation complex (TIC) (e.g., co-activators and co-repressors).
Enhancer	Upstream regulatory element that can cis-regulate transcription (upregulation) by interacting with activator transcription factors, which in turn can interact with co-activators in the TIC.
Silencer	Upstream regulatory element that can cis-regulate transcription by interacting with repressor transcription factors, which in turn can interact with a co-repressor in the TIC, for disassociation from DNA.
Post-transcriptional regulation	Regulatory mechanisms that operate upon the mRNA transcript itself or in its maturation process.
microRNA-mediated mRNA decay	A mechanism in which mi-R hybridize with target DNA and recruit the Argo complex, which leads to removal of the protective elements of the mRNA: its cap and polyadenylation tail, which leads to degradation and subsequent downregulation of said transcript.
Splicing	A process that occurs during mRNA maturation, where segments of the raw transcript (pre-mRNA) are removed.
Alternative splicing	Alternative mRNA maturation processes that result in variations in the final mature RNA transcript and subsequent variation in the translated amino acid sequence, which in turn can result in functional variation in protein properties (e.g., dimerization, folding, and ligand affinity).

As novel approaches continue to emerge, we are now able not only to identify positive selection on genomic regions but also to investigate the molecular mechanisms underlying adaptive phenotypes. The aim of this systematic review is to discuss specific examples of molecular mechanisms driving adaptive evolution in vertebrates in response to

ecological factors, such as differences in lighting conditions, aquatic and terrestrial environments, extreme conditions, diet diversification, pathogens, reproductive adaptations, and others. We focus here on recent studies on adaptive evolution (from 2016 to 2022), which were not covered by earlier reviews on this topic [1,2]. A more recent review [5] was focused specifically on adaptive convergent evolution, while the current review has a broader scope, providing descriptions of the molecular mechanisms involved in the emergence of adaptive variation.

2. Materials and Methods: Evidence Acquisition

The database Web of Science was searched using the following Boolean string: "("adaptive evolution" or "adaptive evolution mechanism$" or "adaptive evolutionary mechanism$" or "molecular adaptive evolution" or "transgenerational adaptation" or "adaptat$ epigenetic" or "adaptat$ epigenetics" or "local adapatation$") AND ("mammal$" or "vertebrat$" or "fish$" or "bird$" or "reptile$" or "amphibian$" or "avian reptile$" or "non avian reptile$") NOT (plant$ or invertebrate$ or fung$ or drosophila or "information technology")". To address a concern of one of the study reviewers that studies on reproductive adaptations are underrepresented in this review, we carried out an additional search specifically focused on this topic, using the following query: "("reproduction or reproductive") AND ("adaptive evolution" or "molecular adaptation") AND ("mechanism" or "epigenetic" or "regulatory") AND ("mammal$" or "vertebrat$" or "fish$" or "bird$" or "reptile$" or "amphibian$" or "avian reptile$" or "non avian reptile$")".

2.1. Study Selection

The search was refined to only include records from relevant research areas and yielded 1117 records. This was further narrowed down to only cover the years 2016–2022, resulting in 487 records. The search was limited to the last seven years because three previous reviews [1,2,5] discussed similar topics based on the earlier literature. Furthermore, the largest number of relevant papers was published within the covered period (Figure 2). From 571 studies published in 2016–2022, 148 were selected using the inclusion and exclusion criteria (Figure 3).

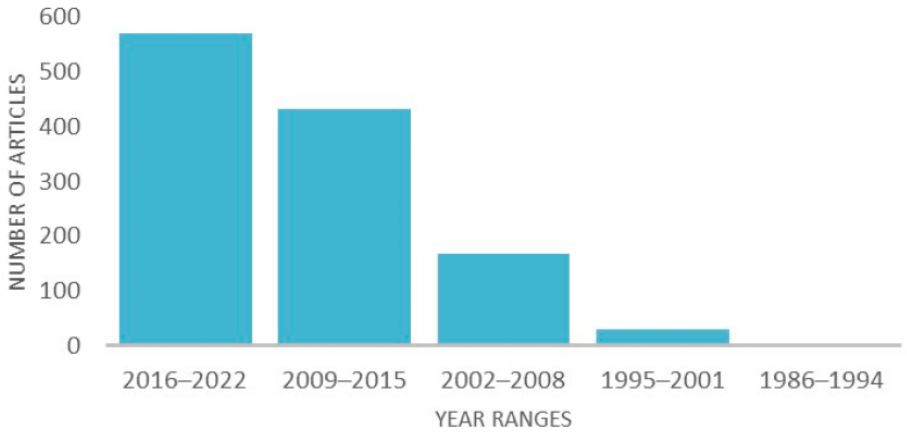


Figure 2. Number of articles found in the systematic search in Web of Science grouped by period of time.

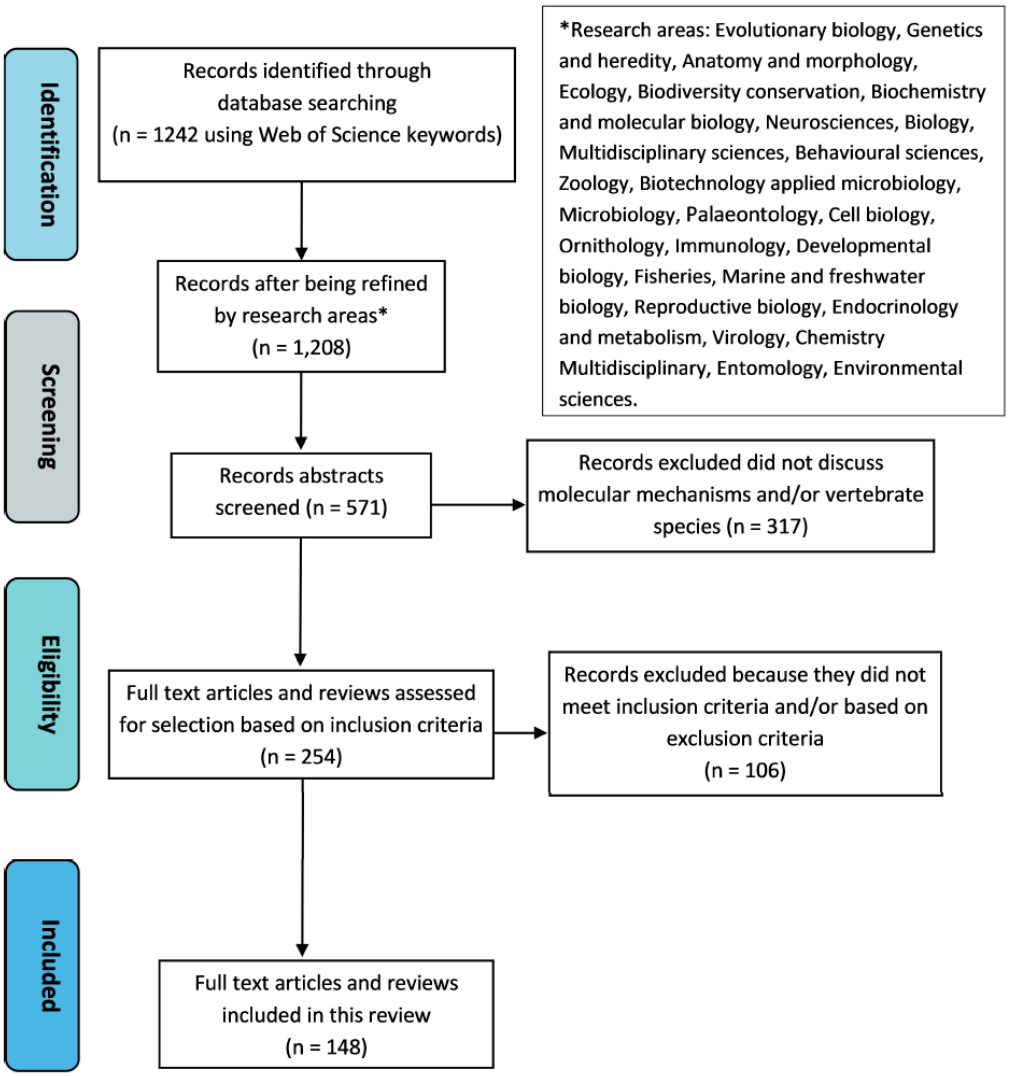


Figure 3. PRISMA flow diagram for systematic review of molecular mechanisms underlying vertebrate adaptive evolution. Records were identified for the years 2016–2022 using the Web of Science database. They were further refined to only include records from relevant research areas. Inclusion criteria: research articles or reviews that investigated or discussed molecular mechanisms involved in adaptive evolution and/or evolution of a trait that was shown to increase fitness. Exclusion criteria: research articles or reviews that described observations suggesting adaptive evolution of a trait (i.e., signatures of selection) but neither offered detailed information about the underlying mechanisms for adaptive evolution nor were supported by other articles that did investigate the underlying mechanism; research articles that deemed their results inconclusive or in need of validation, and/or found no evidence to support the adaptive molecular mechanisms they were investigating; and research articles or reviews that described the evolution of a trait that had not been classified as adaptive. In this search, we selected 148 papers, and 2 additional papers were later included, following the second literature search (see the main text).

The additional search focused on reproductive adaptations identified 515 studies, but only 2 studies newly identified in this search met the inclusion criteria. Moreover, 51 studies appeared in both searches, accounting for 9% of papers found in the original search and 10% in the new one. Of these 51 studies, 41 either did not meet the scope of the review, i.e., did not discuss molecular mechanisms in response to environmental factors, and/or

did not meet one of the criteria discussed in the methodology. Five studies identified in this additional search are discussed in the section on reproductive adaptations (of which three appeared in both searches and two in the new search only) and the remaining seven studies in other sections.

The addition of two new studies on reproductive adaptations increased the total number of studies included in the systematic review to 150. All these studies are listed in the reference list. The reference list has 152 records, as the first 2 studies cited [1,2] are previous reviews on the topic, which were not identified in the systematic search. The 150 selected studies were grouped into categories corresponding to the six discussed environmental factors and an "others" section for studies that could not be classified into any of these categories (Figure 4). Fourteen studies were assigned to more than one category.

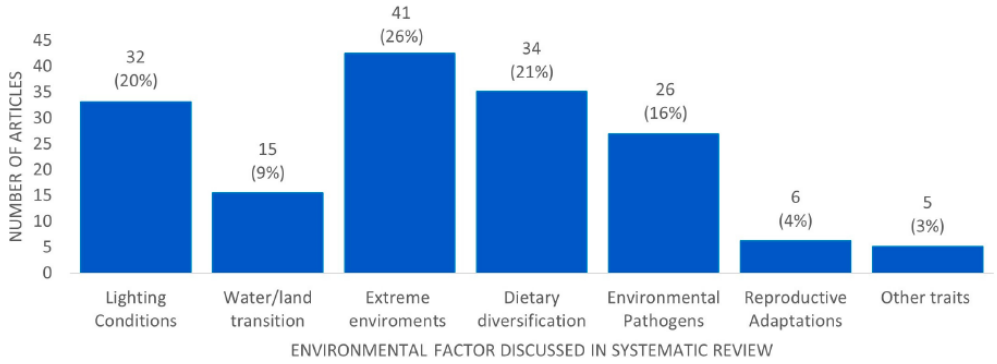


Figure 4. Division of articles by topic, as discussed in this systematic review. The 150 selected articles were grouped into six categories corresponding to the six discussed environmental factors and "other traits" pulled together. A total of 14 articles investigated or reviewed more than one of these environmental factors; consequently, 13 of these articles were discussed in two chapters, and the remaining article was discussed in three.

2.2. *Eligibility Criteria*

I. We included research articles or reviews that investigated or discussed molecular mechanisms involved in adaptive evolution and/or the evolution of a trait that was shown to increase fitness.

II. The following articles were excluded:

   i. Research articles or reviews that described observations suggesting the adaptive evolution of a trait (i.e., signatures of selection) but neither offered detailed information about the underlying mechanisms for adaptive evolution nor were supported by other articles that did investigate the underlying mechanism.

   ii. Research articles that deemed their results inconclusive or in need of validation, and/or found no evidence to support the adaptive molecular mechanisms they were investigating.

   iii. Research articles or reviews that described the evolution of a trait that had not been classified as adaptive.

2.3. *Limitations*

Although we aimed to identify as many articles as possible that met the established criteria, it is possible that some relevant articles may be missing from this review if they were not identified based on the set of keywords used in the search and/or were unavailable in the Web of Science database.

3. General Aspects of the Molecular Mechanisms Underlying Adaptive Evolution in Vertebrates

Advances in sequencing, proteomics, and bioinformatic technologies enabled researchers to investigate the molecular mechanisms underlying adaptive variation. Sequence variation (e.g., indels (i.e., insertion–deletions) and point mutations) is one of these mechanisms (Figure 1). In some environmental contexts, sequence variation in key conserved sites can result in adaptions, but molecular implications of sequence variation/mutations vary depending on the type of conserved element impacted [6]. Mutations may arise via the insertion of transposable elements, which can result in variations in transcriptional regulation, DNA methylation patterns, and chromosome stability [7].

When mutations occur within the coding region, it has the potential to impact amino acid sequences near or within functional protein domains, and/or to affect post-transcriptional regulatory elements such as microRNAs (miRs). MiRs can be tissue-specific, can be developmental-time-specific, and may even be a response to stimuli [8–10]. MiR-mediated adaptive evolution can enable a lineage to gain or lose a regulatory sequence rather quickly, modifying regulatory patterns under novel pressure [9]. The modification of regulatory patterns leads to differential transcription. This can also occur due to (1) indirect factors (i.e., mutations leading to functional variation in proteins regulating transcription (e.g., transcription factors)) and (2) mutations in noncoding regulatory regions (e.g., enhancers and silencers) (Figure 1).

Another regulatory mechanism that can result in differential transcription is epigenetic modifications (Figure 1). Epigenetic changes are molecular modifications that affect genome structure, altering transcription rates by impacting the DNA binding ability of transcription factors and RNA polymerases [11]. Epigenetic modifications have been suggested to be essential for a rapid adaptive response to environmental factors [12]. Epigenetic changes do not always require novel mutations and are reversible (i.e., transient); thus, it can be an effective mechanism for differential gene regulation when organisms are subjected to sudden environmental pressures. Epigenetic variation can provide an immediate response to a sudden environmental change and can be transient.

Gene losses and losses of function can also give rise to evolutionary novelties, which can be advantageous in some ecological contexts [13]. The idea that gene loss can result in an increase in fitness is a controversial topic. Non-functionality of a gene leads to loss of the protein or proteins coded by this gene, affecting associated phenotypes. Some regulatory elements (e.g., silencers and corepressors) that contribute to gene loss of function have been deemed evolutionary conserved units, implying that gene loss may be adaptive. Exploring the possibility of how gene loss can be advantageous by yielding alternative phenotypes could provide further insight into the complexities of adaptive evolution.

4. Adaptive Evolution in Response to Variation in Lighting Conditions

Visual systems are important sensory systems used for navigation and receiving environmental stimuli. Lighting conditions vary considerably across environments, thus successful colonization of new environments requires adaptation of visual systems. Ecological factors shape visual sensitivities and have an impact on the molecular mechanisms that drive the adaptive evolution of visual structures and/or regulatory mechanisms in vertebrates [6,14–17]. Adaptation to a novel lighting environment is a complex process that requires changes in lens transmittance and retinal cell structure (i.e., chromatophore ratios and opsin gene expression) [14,15,18,19]. Opsins are photoreceptor proteins that bind to photons, thus interacting directly with an environmental stimulus.

Adaptive variation affecting retinae structure in response to variation in lighting conditions has been observed across multiple vertebrate species. One of the most common mechanisms is modifications near or within the functional domains of photoreceptors

[17,20,21]. A well-described adaptive phenomenon in this regard is spectral tuning. Spectral tuning occurs when sequence variations impact photoreceptor binding sites, modifying the wavelength of the absorbed light that can be detected by photoreceptors [6,22].

Mutations in rhodopsin genes (*RH1* and *RH2*) have been observed to cause spectral shifts in fishes, reptiles, and marine mammals as an adaptive response to variation in aquatic lighting conditions or lighting variation associated with living on the forest floor [15,16,22–26]. Spectral tuning associated with mutations in opsin genes (*LWS* and *OPN1LW*) has been proposed as an adaptive mechanism in vision acuity in giraffes [27]. Spectral tuning in the opsins LWS and SWS2 has also been observed as an adaptive mechanism contributing to nocturnal or benthic vertebrates' ability to navigate low-lighted environments [20,25,28,29]. Both in reptile and fish lineages, relaxed constraints, at times leading to pseudogenization or gene loss, have been observed across species living in dimly lit environments [21,22]. SW1 loss has been proposed to be a beneficial phenotype, contributing to improving opsin sensibility towards longer wavelengths [20]. The shifts in spectral sensitivity are postulated to have occurred due to an increase in crepuscular and/or nocturnal activities of these species. Conversely, spectral tuning in SWS1 was observed in nocturnal reptiles and was proposed as an adaptive mechanism for diurnality [22].

Other molecular mechanisms underlying adaptive variation in visual systems are regulatory changes. In various fish species, the differential expression of enzymes and photoreceptors (e.g., CYP27C1, SWS1, RH2, RHO, and SWS2), and mutations in regulatory regions are suggested to be important molecular mechanisms underlying adaptive visual variation [6,21,25,30–34]. In cichlids, mutations in non-coding regions, specifically miR and transcription binding sites, have been found to be associated with differential transcription of opsin genes [33]. This highlights how a better understanding of where regulatory elements bind in the genome can provide essential information on how differential transcription occurs.

Regulatory modifications resulting in variations in the photoreceptor expression patterns are also suggested as a molecular mechanism underlying visual adaptations to nocturnal/dimly lit environments. In cichlids from dimly lit environments, mutations impacting dimerization properties, subsequently affecting downstream pathways regulated by these receptors, have been observed and are proposed to be adaptive [21]. Regulatory changes have impacted the retinal structure of nocturnal vertebrates to exhibit a higher proportion of rod cells, typically expressing RH1 [16,33,35]. RH1 may be essential for visual adaptations to dimly lit environments [24]. An electric knifefish species from dimly lit environments has been shown to exhibit compensatory substitutions in RH1 thought to be involved in re-establishing the dimerization properties of RH1, compensating for a RH1 mutation associated with the visual defects they exhibit in humans [36].

Mutations in functional domains have also been observed in non-photoreceptor visual systems genes. In the forest-dwelling okapi *LUM*, a gene involved in the phototransduction process exhibits a mutation suggested to impact the protein's interaction with collagen, subsequently affecting UV transmission in dimly lit environments [27]. Additionally, the functional enhancement of retinae cell surface proteins (GRK1 and SLC24A1) has been proposed as a part of the underlying mechanisms of adaptations of vision in nocturnal birds [37].

Adaptive variations in visual systems are complex and can also result in adaptions impacting other aspects of phototransduction processes. In teleost fishes (Teleostei), molecular variations resulting in metabolic changes have been suggested to be involved in these visual system adaptations. A mutation in ATPase VHA involved in the increase in acidification of fish blood cells, leading to an increase in oxygen production and subsequent secretion into the retinae, has been proposed as a key adaptation [38]. This could have contributed to the morphological adaptations within teleost retinas and the adaptive evolution of their visual systems.

As previously mentioned, nocturnal vertebrates exhibit a higher proportion of rod cells within their retinae. Conversely, relaxed constraints of the visual system loci have been suggested to be involved in the variation and degeneration of nocturnal vertebrates' visual systems [15,39–43]. Specifically, photoreceptor genes such as *RH1, RH2, SWS1*, or *SWS2* have been lost under relaxed constraints in some of these species [8,41,43]. However, a unique mechanism of transmutation has been suggested as a means for vertebrates to adapt back to diurnal environments [8,36]. In members of the reptilian *Colubridae* family, which lost retinae cone cells, re-gaining a function has been demonstrated to occur through transmutation, resulting in evolutionary modifications of rods to become cone-like in function [8].

Overall, the adaptive variation of vertebrate visual systems is underlined by molecular modifications in retinae. Adapting to varying lighting conditions is a complex process, where many regulatory and structural genes are involved. Research has focused mainly on functional variation within photoreceptors that directly interact with environmental stimuli. However, adaptive variation has also been observed in proteins involved in phototransduction processes and in the expression patterns of photoreceptors. Thus, future studies should also investigate the putative regulatory mechanism, which may impact the expression patterns of photoreceptors as well as the phototransduction process itself.

5. Adaptations for Colonization of Aquatic and Terrestrial Environments

Ancestral vertebrates inhabited oceans and, later, transitioned into terrestrial environments. Conversely, some vertebrates regressed toward aquatic environments from terrestrial environments. Multiple molecular mechanisms involved in the vertebrate transition from water to land have been identified [44,45]. Li et al. (2018) investigated the underlaying molecular mechanisms that allowed vertebrates to invade terrestrial habitats by comparing the walking catfish and non-air breathing catfish. Their findings suggested that genes involved in DNA repair, enzyme activation, and small GTPase regulator activity were part of the molecular mechanisms to overcome the increase in DNA damage in terrestrial environments and the variation in metabolic processes.

The colonization of terrestrial environments also required vertebrates to adapt to hypoxic conditions, terrestrial xenobiotics, and novel environmental stimuli. Gene expansions and tissue-specific regulatory modifications of myoglobin genes (*MB*), sulfotransferase genes (e.g., *SULT6B1*), and olfactory receptor genes (e.g., *ORA1*) have been suggested to be involved in these processes [45]. The tissue-specific differential transcription of genes associated with ion homeostasis, acid–base balance, hemoglobin genes, angiogenesis, elastic fiber formation genes, and mutations in the functional domains of MB are thought to have contributed to overcoming hypoxic conditions in terrestrial environments [45,46]. Adaptations to hypoxic environments are further discussed in a subsequent section of this review.

The transition of vertebrates back to aquatic environments presented a unique evolutionary challenge, as many traits advantageous in these environments had been lost. Interestingly, for some complex traits, the relaxed constraints of specific genes yielded phenotypes that are advantageous in the return to aquatic environments. Specifically, the miR-based downregulation of genes has been suggested to be involved in adaptive variation for diving and overcoming hypoxic conditions. Tissue-specific variation in the microRNome of the deep diving Weddell seals suggests an important role of post-transcriptional regulatory mechanisms in marine mammal adaption to the aquatic environment. The tissue-specific differential expression of miRs targeting genes associated with hypoxia tolerance, anti-apoptotic pathways, and nitric oxide signal transduction was observed in Weddell seals [10]. MiR-mediated post-transcriptional regulation has been shown to be involved in downregulation, a mechanism that results in a substantial decrease in the

amount of protein translated, which mimics the gene being "turned down" or "off" (Figure 1A) [9]. This suggests that the functional loss of some genes could potentially yield advantageous traits in marine mammal diving adaptions [10].

Eukaryotic regulatory pathways are complex, and losing a protein in the pathway which modifies it could potentially yield an advantageous phenotype by modifying instead of losing the trait [47]. This is the case for matrix metalloproteinase (*MMP12*), epidermal and hair development genes (*DSC1, DSG4, TGM5, GSDMA, LYG1*, and *LYG2*), and keratin genes (*KRT9* and *KRT20*). MMP12 is involved in extracellular matrix breakdown, and its loss impacts pulmonary elasticity in a way that allows marine mammals to renew ~90% of their air in a single breath, an advantageous trait for diving [47]. The loss of the aforementioned epidermal and hair developmental genes resulted in modifications of the respective developmental pathways, yielding alternative traits of a thicker epidermis and hair loss traits in cetaceans and sirenians, suggested to be advantageous for the aquatic environments [47–49].

Loss of other functional proteins has also been suggested to have yielded advantageous traits in marine vertebrates. Relaxed constraints and subsequent gene loss of some olfactory and taste receptors genes (e.g., *GNAT3* and *CALHM1*) observed in marine birds and mammals have been suggested to be advantageous traits by masking the smell and taste imbued onto their prey by the seawater [50,51]. Another example is the relaxed constraint of erythrocyte specific enzyme (AMPD3), which causes a reduced affinity to oxygen and a threefold increase in ATP in erythrocytes, an advantageous trait for diving [47].

Mutations in proteins with regulatory functions in cellular pathways (*ACAN*, a growth inhibitor), osteogenesis, and gene regulation (*PIT-1, HOXD11, HOXD12, HOXD13* and *MLL*) have been proposed to be molecular mechanisms involved in the adaptive variation of the cetacean body architecture. The lineage-specific mutations are predicted to impact the structure and functionality of these proteins [52]. As in other vertebrates, they result in a body size reduction as well as shortened limbs and trunk [53]. Mutations in the aforementioned transcriptional factors impact their regulatory functions (e.g., binding affinity to target genes, chromatin remodeling, and histone modification), consequently modifying the transcriptional patterns of their target genes [11,54]. Target genes include genes involved in determining vertebrate body size (e.g., *ELK1, LHX3*, and *PITX1*), suggesting that these mutations are involved in the adaptive variation of cetacean body size and skeletal morphology [11,53,55].

The functional variation and expansion of genes involved in multiple cellular functions have also been proposed as underlying molecular mechanisms involved in adaptions to aquatic environments. Mutations in cetacean ATPase (*ATP8*) have been suggested to result in metabolic adjustments beneficial in their marine environments [56]. Mutations in a reproductive gene (*FSHR*) are suggested to be involved in adaptations of marine mammals' reproductive systems [48]. Finally, the expansions of genes involved in multiple cellular processes (i.e., cellular response, oxidative stress, oxidation reduction, and hydrogen peroxide response) have been identified in cetaceans, other aquatic mammals, and semi aquatic mammals. They are likely involved in adaptations to hypoxic conditions, aquatic pathogens, novel energetic metabolic demands and nervous system adaptations [48,57–59].

Transitions to terrestrial and aquatic environments are complex evolutionary processes that require variation associated with both environmental stimuli (e.g., xenobiotics) and navigating these novel environments (e.g., diving). Due to the complex regulatory dynamics of eukaryotic gene expression, modification or even loss of regulatory proteins have yielded phenotypes that are advantageous in specific environmental contexts. Thus, future research in this topic would benefit from approaches that investigate the role of specific biochemical pathways involved in relevant cellular functions and gene regulation.

6. Adapting to Extreme Environmental Conditions

Extreme environments can occur due to natural ecological factors as well as anthropogenic factors (e.g., pollution). Pollution can cause an array of environmental stressors for organisms. Amazonian teleost fishes living in polluted environments have been observed to exhibit an increase in transposable elements, which are thought to contribute to their ability to endure these stressful conditions [7]. In snow finches from the Tibetan Plateau, mutations in a DNA repair gene (*DTL*) have been suggested to be an adaption for their high-altitude environments [59]. Furthermore, mutations in metabolic enzymes (*GST*) are thought to be involved in modulating specificity or potency of the detoxification and metabolic activation of xenobiotic pathways in mammals [60].

6.1. Hypoxic Environments

Hypoxic conditions (i.e., low oxygen) exist both in terrestrial and aquatic environments and can negatively impact biological processes. Functional mutations and regulatory modifications are suggested to be molecular mechanisms underlying vertebrate adaptations to hypoxic conditions [61,62]. Mutations improving binding affinity in the oxygen-transporting molecule (HB) have been identified in multiple vertebrate lineages [5,46,63]. The convergent evolution of a hypoxia inducible factor (*EPAS1*) has been observed in multiple unrelated vertebrate species living in hypoxic conditions in Tibet [64]. This suggests that functional mutations in proteins that either directly interact with oxygen molecules or are involved in the modification of cellular oxygenation, have been essential for adapting to hypoxic environments. Physiological adaptations within respiratory systems can also be essential in hypoxic environments. Variations in regulatory pathways have also been proposed as a mechanism underlying lung adaptations in yaks living in hypoxic environments [62].

In Tibetan fish species living in hypoxic environments, rapid evolution was inferred for genes enriched in hypoxia and energy metabolism [61,65]. The co-evolution of genes that enhance the function of the hypoxia-inducible factor was also suggested to be involved in adaption to these conditions in fishes [61]. Mutations inhibiting target protein-binding affinity in a regulatory protein (VHL), whose function is targeting the hypoxia-inducible factor to degradation, has been observed in tree unrelated fish lineages living in hypoxic conditions [61]. Mutations resulting in the functional variation of proteins involved in metabolic processes and/or pathways have also been suggested to be involved in mammalian adaptations to hypoxic environments. Mutations in genes involved in metabolic processes (e.g., rate limiting enzymes within the gluconeogenesis pathway (LDHA, LDHD, PC, PCK1, FBP1, and GPI), anaerobic respiration (ALDOA and ENO), oxidative phosphorylation proteins (NDUFA9, NDUFA10, NDUFAB1, NDUFC2, and NDUFV3), the hypoxia inducible pathway, and the mitochondrial NADH dehydrogenase complex have been suggested to be involved in mammalian adaptations to hypoxic environments [63,66,67].

Modifications in transcriptional patterns have also been proposed as molecular mechanisms involved in adaptive variation to hypoxic environments. For instance, both tissue-specific differential expression and mutations of cytokine erythropoietin (*EPO*) enabled schizothoracine fish to adapt to their hypoxic environment [68,69]. Some of the mutations identified in EPO have been shown to confer antioxidative and antiapoptotic properties to the protein, both of which are beneficial in hypoxic environments [69]. Finally, developmental and tissue-specific differential expression of miRs enriched for genes involved in multiple cardiac processes (e.g., ischemic postconditioning, perfusion, and angiogenesis) have been associated with managing challenges associated with hypoxic conditions in Weddell seals [10].

6.2. Extreme Aquatic Environments

An environmental factor for which aquatic vertebrates had to adapt is the salinity content of the water (i.e., freshwater vs. saltwater). For most saltwater species, freshwater is an extreme environment and vice versa. Transitioning from one to the other requires specific adaptations. The differential expression of genes associated with metabolic processes (e.g., homoeostasis, transmembrane transport, ion-exchange, osmoregulation, blood traits, and tooth morphology) has been observed between freshwater and saltwater stickleback ecotypes [70–73]. Additionally, in sticklebacks, it has been proposed that some of the marine-environment-specific advantageous variation can be inherited in clusters as a unit [74]. The differential expression of genes including transcription factors associated with stress response and immune responses has been observed in the tongue sole fish in response to a variation in the salinity content of the water [75]. Both cis- and trans-regulatory variations are suggested to modify the expression patterns of the genes (e.g., *GLA* and *IGFLR1*) associated with these processes [70,71,73]. The environment-specific adaptive variation in both sticklebacks and cod fishes has been suggested to be mediated by chromosomal rearrangements [76]. In reptiles and amphibians, gene expression variation mediated by transposon activity has been proposed as one of the mechanisms for adaptations to marine environments [77].

Regulatory variation is another molecular mechanism involved in adaptive morphological variation to freshwater environments. Mutations in cis-acting regulatory elements (i.e., transposon insertion and/or deletions) of genes associated with morphological and skeletal development (i.e., *PITX1, EDA, KITLG, BMP6,* and *GDF6*) and nutrient intake (i.e., *CSAD* and *CTH*) are thought to be involved in the emergence of adaptations for freshwater environments in sticklebacks [72,78–81]. Regulatory variation in one of these genes (*PITX1*), which underwent deletion of an enhancer region, is suggested to have occurred due to increased mutation rates resulting from the DNA breakage propensity of said region [82]. Similarly, the deletion of an enhancer of *EDA* is thought to be pleiotropically involved in variation in multiple traits associated with morphological and feeding adaptions to freshwater environments (i.e., lateral plate count, number, and patterning of posterior lateral line neuromasts) [81,83].

Mutations resulting in the functional variation of *SLC12A3* and the truncation of *MSX2* have also been suggested to be involved in adaptations to freshwater environments. Mutations observed in *SLC12A3* in Atlantic and Baltic herring have been proposed as necessary adaptive variations for osmotic balance in response to environmental differences in salinity [15]. Mutations in *MSX2* resulting in alternative splicing, which in turn results in higher expression of a truncated isomorph, are thought to be involved in the adaptive reduction in spine length in freshwater sticklebacks [84]. Finally, mutations in choriolytic enzyme loci impacting hatching in herring have been suggested to be involved in reproductive adaptive variation to salinity content in aquatic environments [15].

Another environmental challenge in aquatic environments is highly acidic or highly alkaline waters. Mutations in *VIPR1* and mitochondrial proteins (e.g., COX3 and COX1) of Amur ide fish and *Poeciliidae* fish have been associated with adaptations to environments with high sulfuric acid concentrations and alkaline environments, respectively [85–87]. The duplication of *NPR1, ZP,* and *VMO1* genes have also been proposed as mechanisms involved in adaptations to alkaline environments. The expansion of these genes yielded an adaptive variation of osmoregulatory mechanisms and a thickening of the egg chorion, respectively [88].

The differential expression of *NPR1, CA, GST,* and *SOD* genes is suggested to underlie homeostatic adaptive variation to alkaline environments in teleost fish [87]. The differential expression of metabolic genes (e.g., *PKLR; FBP2, G6PCB,* and *PCXB*) and genes associated with DNA replication and repair, and proteasome assembly have been suggested to be involved in compensatory growth [89]. Compensatory growth, i.e., rapid growth occurring after adverse environmental conditions have dissipated, is an advantageous trait in transient extreme environmental conditions.

6.3. Temperature Extremes

Extreme high or low temperatures are another environmental factor that can create adverse conditions. These extremes can lead to water loss, metabolic changes, and/or cell viability. Adaptations in functional domains of proteins (*ABCA12*) and differential expression (*ATP1A1*) have both been proposed as molecular mechanisms underlying adaptations to high temperatures [90,91]. In birds and mammals, mutations in a temperature-sensitive receptor TRPM8 have been suggested to be involved in cold-tolerance adaptations [41]. The combined effect of mutations in *PGC-1*, which regulate the expression of mutated variants of thermogenin (*UCP1*), affected the heat-generating capacity of brown adipose tissue, enabling mammals to adapt to low-temperature environments [63].

Expansions of the 118 gene families involved in metabolic processes have been suggested to be involved in adaptation to low-temperature environments [15]. Epigenetic modifications (e.g., DNA methylation) are thought to be involved in the adaptation to transient low temperatures in winter skates [92]. De novo gene birth has been identified as the mechanism by which similar anti-freeze proteins (AFGP) arose in two unrelated teleost fish species [93]. The rare event of de novo gene formation occurred in both species: duplication of noncoding DNA elements, which over time gained the functional elements to become a gene (i.e., TATA box, open reading frame, and a promoter).

7. Adaptations to Dietary Changes

7.1. Metabolic Adaptations

Access to food/nutrients is an important ecological factor that often influences evolutionary adaptations. The mechanisms underlying food selection can be complex, requiring adaptations in digestive systems and/or feeding apparatus. Teleost fishes' adaptive variation within insulin-based metabolic processes arose via a combination of gene duplications, pseudogenization, and retrotransposition [94]. In the Gekkonidae lineage, functional variation in IGF-1 resulting in a switch between metabolic pathways is suggested to be involved in their dietary diversification [95]. The adaptive evolution of mammalian metabolic systems has been found to have occurred mainly due to mutations leading to structural and functional variations in proteins [4].

Adaptations for diet diversification in birds resulted from the combined effect of gene duplications, losses, and functional protein variation in digestive enzymes [96]. Mutations in a carbohydrase (*AMY*) and a lipase (*CYP7A1*) were commonly observed amongst specific bird groups (i.e., carbohydrase in seed eaters and lipase in meat eaters) suggesting an association with diet diversification [97]. In mammals, duplication and functional variation in an immune system gene, *RNASE1*, and a G-Protein-Coupled-Receptor have been proposed as mechanisms underlying adaptations toward herbivory [97,98].

Cetaceans experienced a dietary shift from herbivory to carnivory. Mutations in functional domains of multiple proteins (PGA, LIPF, ACAD9, EHHADH, and PLRP2) have been proposed to be associated with this shift [54,99]. The functional implications of these mutations have not been described for all these proteins. However, the mutation in PGA increases binding to the ligand, while the one in PLRP2 is thought to be involved in enabling calves to digest fattier milk. Another mammal that experienced a dietary shift was the domestic dog, who adapted to feed on human waste. This transition could have been facilitated by an emergence of novel retrogenes involved in lutathione biosynthetic/metabolic processes and toxic substance response [100].

7.2. Non-Metabolic Adaptations

In giant pandas, the pseudogenization of bitter taste receptors (TAS2R) was suggested to have facilitated their shift to an almost exclusive bamboo diet. Bamboo has bitter-tasting substances as a defense mechanism, and the loss of bitter taste receptors is thought to have enabled giant pandas to overcome this challenge [101]. Additionally, epigenetic modifications increasing the expression of thiosulfate sulfurtransferase have also

been proposed as an adaptation in red and giant pandas to overcome the toxicity of bamboo [102]. Polar bears underwent a dietary shift to a hypercarnivorous diet, which is thought to have resulted in losses in the copy number of genes associated with a digestive system, (e.g., *AMY1B* and *NOX4*) and olfactory receptor genes [103]. These copy number losses are proposed to have yielded advantageous traits, enabling polar bears to increase fat storage (*NOX4*) and to develop a more acute olfactory system [103].

Dietary shifts can inadvertently cause imbalances in nutrient availability or toxicity issues. Vampire bats exhibit a diet exclusively based on blood and have a unique evolutionary strategy to overcome associated nutrient deficiencies. Vampire bats have a mutational bias favoring cytosine and not favoring thymine, which subsequently increases threonine in their proteomes, enabling them to overcome inherent amino acid limitations in their diet [104]. Convergent epigenetic changes in both red and giant pandas have been proposed as the molecular mechanism behind the differential expression patterns of genes involved in lipid digestion (*DGAT2, SLC2A3,* and *PARL*), enabling them to overcome nutritional challenges of their low-lipid diet [102].

In 31 bird species, mutations in alanine-glyoxylate aminotransferase (AGT) modifying protein localization have been suggested to be involved in enabling them to overcome carnivory-associated toxicity issues [105]. Similarly, both in colubrid snakes and mammals, mutations in voltage-gated sodium channels pores (*SCN4A, SCN8A*, and *SCN9A*) have immunized some species to tetrodotoxin (i.e., toxic compound secreted by some amphibians (e.g., *Taricha* newts) as a defense mechanism against predators [8,106,107]. Amphibians that secrete tetrodotoxin also have mutations in voltage-gated sodium channels, conferring them resistance to the toxin [108]. These mutations are suggested to have been facilitated by duplication events.

7.3. Anatomical and Behavioural Adaptations

Dietary shifts can also require the emergence of novel traits to catch and/or consume novel food items. Snakes have developed venom-delivery systems to subdue their prey. The molecular mechanisms underlying this adaptation are complex and vary amongst lineages. The current consensus is that a combination of neofunctionalized genes and tissue-specific differential transcription is the main mechanism underlying the emergence of this trait [8]. Some birds have evolved tactile foraging behaviors to catch their prey. An increase in expression mechanoreceptors associated with tactile sensitivity has been observed in three species of birds exhibiting this behavior [109].

Dietary shifts often require changes in craniofacial structures, and regulatory modifications have been proposed to play a part in such changes. In sticklebacks and cichlids, an adaptive variation in craniofacial development, including teeth number and morphology, is suggested to occur due to differential expression patterns of developmental genes (*ODAM, UNK, SCPP5*, and *RPFA*), including cis-regulatory (*TFAP2A* and *BMP6*) and/or trans-regulatory changes [72,110–112]. Adaptive variation in mammalian teeth shape has also been associated with mutations resulting in functional variation in developmental (*BARX1, EVE1, LHX7, LHX8*, and seven *HOX* genes) and regulatory (*BMP2, BMP4, DLX2, EDA, EDAR, PAX9, PITX2, RUNX2*, and *SHH*) genes [113,114]. In sea horses, African cichlids, and some mammalian species, the loss of conserved elements and various morphological genes (e.g., *TBX4, SALL1A, SHOX*, and *IRX5A*), and a negative regulation leading to effective loss of genes affecting morphological features (BMP) have been associated with adaptive phenotypes [13,114,115]. Loss of specific genes and regulatory elements contributed to morphological changes in the head (e.g., teeth and jawbone), resulting in advantageous evolutionary feeding traits/novelties [13,114,115].

Behavioral adaptations are involved in dietary shifts and often require complex variation involving multiple genes. Differential expression of the developmental gene *EDAR* is thought to impact their predatory behavior in teleost fish. Lower levels of *EDAR* expression have been associated with an increase in predatory behavior in both zebrafishes and Mandarin fishes [116]. Conversely, in cave fish species, mutations in MAO and OCA2

are suggested to be involved in adaptive behavioral changes (e.g., reducing predatory, schooling, or sleeping behaviors) [43,117]. In cetaceans pseudogenization (*MCPH1*) and mutations in neurodevelopment genes (*WDR62, CDK5RAP2, CEP152*, and *ASPM*) are thought be involved in an incrementation in brain size [118]. These developmental and anatomical modifications can modify or add complexity to feeding behaviors.

8. Environmental Pathogens

Vertebrate immune systems are characterized by adaptiveness under the co-evolutionary arms race with pathogens. Sticklebacks display habitat specific (i.e., lake vs. river) differential expression of immune system genes; thus, gene regulation underlies adaptations in the presence of habitat-specific pathogens [119]. Amino acid changes near or within the functional domains of immune system proteins are one of the main molecular mechanisms underlying adaptations to environmental pathogens [4,120]. Mutations giving rise to habitat-specific functional adaptive variation in immune system genes (*IGHM, IGHE IGH, CD14, CD40, CD80, IFNAR2, LY96, TAB 1, TICAM1, TLR4, MDA5*, chicken type lysozyme, and *pIGR*) have been observed in different vertebrates and can happen through insertions (e.g., MUC7 in primates) or gene duplications [121–128]. For the pIGR specifically, research suggests that the observed mutations occurred through an insertion of transposable elements [127].

Functional variation has also been observed in non-immune system proteins. In several species of Galliformes birds, a single amino acid change in MX Dynamin Like GTPase, confers resistance to avian influenza [129]. In gaur, an Asian bovid, mutations causing a structural conversion in the mRNA-editing protein APOBEC3Z3 enabled the protein to acquire resistance to degradation by the infectivity factor of the Jembrana disease virus [130].

Gene duplications and the neofunctionalization of genes have occurred in multiple vertebrate immune system gene families (e.g., *MHC, NLPRP3, CD22*, immunoglobins, *TLRs*, and *SSC4D*) [126,131,132]. Among the most studied gene families are immune pattern recognition receptors, and *TLR* gene family is of particular interest due to their involvement in pathogen detection. Environmentally specific duplication and/or neofunctionalization of TLR genes have been observed across fish, reptile, and bird species [133–137]. Environmentally specific functional variation in pattern receptors including TLR family genes, impacting binding affinities, and/or imparting resistance to environmentally specific pathogens has been observed in mammals, birds, and reptiles [123,134,137–141]. The emergence of novel pathogen resistances enabled vertebrates to colonize novel environments [141].

Within the *Rhinolophoidae* and *Pteropodidae* bat lineages, indels have been suggested to lead to functional variation, impacting the binding affinity of TLR8, thus improving pathogen recognition [142]. Gene losses are also thought to be essential in the adaptive evolution of vertebrate's immune systems. Negative selection leading to the loss of genes' *NF-kB*, targets of TLRs, has led to conservancy of functional domains and copy numbers from fish to mammalian vertebrates [143]. Gene expansions, gene losses, and mutations leading to neogenization and/or functional variation, particularly impacting binding affinities and protein–protein interactions, are the main mechanisms known to produce immune systems adaptive variation.

9. Reproductive Adaptations

Reproductive adaptations have also been necessary for the colonization of some environments. In bony fishes, the duplication of zona pellucida genes (*ZP*) has been proposed to play a role in reproduction via external fertilization and by successfully hatching eggs, providing an additional layer of protection under the constraints of the environmental factors [144,145]. Multiple gene-duplication mechanisms have been identified in this

process (i.e., genome duplication, gene block duplication, and tandem duplication). Conversely, vertebrates with internal fertilization exhibit signs of relaxed purifying selection, pseudogenization, and subsequent reduction in number of *ZP* genes [144].

In platyfish, the duplication of *C6AST* genes has been associated with the adaptive evolution of internal fertilization and ovoviviparity [13]. In seahorses, the duplication of *C6AST* arose independently, and it is involved in the evolution of male pregnancy in these fish [13]. The species also exhibits duplication and neofunctionalization of the *PASTN* genes, which are thought to be involved in the adaptive evolution of internal fertilization [13]. These molecular mechanisms (i.e., gene duplications and neofunctionalization) are also involved in mammalian reproductive adaptations. *CRISP* genes are involved in sperm adaptions to increase male reproductive success [146].

The evolution of novel reproductive strategies in response to environmental factors can require adaptations in non-reproductive structures. Some species of anglerfish have evolved a unique reproductive strategy to overcome challenges in their deep-sea environments: parasitic males. Males in these species have evolved to be parasitic and are permanently attached to the females. Without the necessary immune system adaptations, the parasitism of the male would illicit an immune response in the female, which could potentially kill the males. These adaptations involved mutations impacting the function of immune system proteins (e.g., substrate binding affinity), differential regulation (e.g., reduced expression), and gene losses (e.g., pseudogenization) (*MHC* and *CD* gene families) [147].

Another reproductive strategy that required adaptation of non-reproductive structures is acoustic communication used to find and/or compete for mates. Concave-eared frogs live in environments with background noise that can obscure their acoustic communication. Supersonic hearing has been proposed as an adaption phenotype to background noises in their environments. Differential transcription in genes involved in neurogenesis and sensory perception has been proposed as one of the mechanisms underlying the evolution of this trait [148]. Multiple genes observed to be differentially transcribed in male and female anglerfish are involved in gene regulation or have roles within cellular pathways [148]. Thus, gene duplication, neofunctionalization, differential transcription, and gene losses have been important molecular mechanisms in reproductive adaptations.

10. Other Environmental Factors

10.1. Changes in Circadian Rhythm

Another environmental factor that vertebrates contend with is day/night cycles. In Artic reindeer, mutations in a clock gene (*PER2*) resulting in loss in binding affinity to another circadian rhythm protein (CRY1) are suggested to impact their sleep/wake cycles, enabling them to adapt to unusually long nights in their environments [3].

10.2. Morphological Adaptations

The colonization of novel habitats often requires morphological adaptations. Indels in EDA have been associated with the advantageous loss of scales in underground snow trout inhabiting the Tibetan plateau [149]. Mutations in the transcription activation domain of mammalian MLL, a methyltransferase involved in epigenetic regulation by chromatin remodeling of *HOX* genes, imply modifications in *HOX* expression [11]. *HOX* genes play a key role in morphological development, and differential transcription of these genes due to MLL mutations is one of the mechanisms underlying adaptive morphological variation in mammals [11]. The unique morphological features in seahorses have been suggested to have emerged due to the loss of noncoding regulatory elements in genes associated with vertebrate morphology (e.g., regulation of transcription, pectoral fin morphogenesis, steroid hormone receptor activity, regulation of the fibroblast growth factor receptor signaling pathway, and O-acetyltransferase activity) [13]. Epigenetic regulation, methylation, impacting gene expression, and subsequent regulatory pathways have also

been proposed as a mechanism underlying adaptive physiological changes in muscle properties of pelagic migratory fishes [150].

11. Limitations of the Study

A systematic review approach has many advantages, including clearly defined selection criteria and reproducible results. This approach also eliminates conscious or unconscious bias in the article selection, which can occur in the case of a traditional review, where the list of articles cited depends on the choice made by the authors. However, systematic reviews also have some inherent limitations, as they can be subject to errors associated with the search algorithm as well as human error. In this instance, the literature review was limited to the results of a search with a specific query and a specific search engine. Consequently, if publications within the scope of this review were unavailable in Web of Science and/or did not have the keywords in the query within their titles, keywords section, and/or abstracts, they were not included. This could have led to the seeming underrepresentation of some topics (e.g., reproductive adaptations). Additionally, there is an apparent overrepresentation of findings based on the use of model organisms. This is unlikely to result from any bias in the choice of keywords but instead reflects the fact that studies on model organisms tend to be at the forefront of the progress in any biological discipline and are overrepresented in the literature in general. Finally, the review cites multiple papers by the same authors or research groups. Given that the studies were identified using objective criteria, this outcome does not result from any preferences of the review authors but reflects productivity of particular researchers or groups working in the field discussed in the review.

12. Future Research

There are multiple molecular mechanisms underlying adaptive evolution in vertebrates. For the 157 genes presented in this review, 4 major molecular mechanisms underlying adaptive evolution were described: sequence variation/mutation (56%), gene expansions (17%), differential transcription (12%), and gene loss or pseudogenization (12%) (Table 2). The most frequent molecular mechanism of adaptive variation is mutation, either in conserved coding or non-coding regions [2]. Mutation mainly impacts eukaryotic transcriptional regulatory mechanisms (e.g., enhancers and miRs) or functional domains of proteins (e.g., ligand binding and dimerization) [4,32,33,120]. Interestingly, positively selected sites tend to occur close to each other within the proteins, resulting in functionally important regions [4]. Thus, it is important to analyze the location of such regions within the genome, which could provide further insight into the molecular mechanisms associated with adaptive variation. For instance, DNA fragility has been proposed as the reason for increased mutation rates within specific regions of the genome [82].

Mutations within non-coding regions can result in the emergence of novel regulatory elements (e.g., miRs). Duplication is one of the main molecular mechanisms by which novel miRs arise and undergo neofunctionalization in vertebrates [9]. Mutations within regulatory elements can affect gene expression, and these modifications in gene expression can yield adaptive variation [32]. Mutations that modify translational and post-translational processes can also yield adaptive variation. Novel ubiquitination sites can impact both protein localization and protein structure. As vertebrate lineages diverged, novel ubiquitination sites emerged in proteins associated with multiple cell processes (e.g., differentiation and motility), suggesting that this was a novel adaptive variation [151]. To gain a better understanding of the mechanisms underlying vertebrate adaptive evolution, a deeper look into regulatory modifications in both genetic and cellular pathways is needed. The gene loss of a function should also be considered, since the loss of regulatory elements in cellular pathways can yield advantageous phenotypes [81,82].

Adaptive variation can occur in specific regions and even more so in clusters, meaning that larger functionally important regions instead of specific sites need to be conserved [74]. There is a growing body of research proposing that chromosomal rearrangements

and inversions are involved in maintaining adaptive genotypes [76]. This could also be possible in some immune system proteins and xenobiotic-metabolic enzymes, as these proteins tend to have evolutionary conserved functional domains [4]. In teolosts, RNA interference specifically of piwi-interacting RNA (piRNA) has been suggested to be involved in regulating the insertion of transposable elements [152]. PiRNA are a class of noncoding RNA molecules involved in maintaining germline integrity through transposon silencing, and consequent transcriptional and post-transcriptional regulation [152]. This mechanism has been shown to be a key element in ensuring genome integrity, which includes genomic regions with adaptive variation [152]. Thus, RNA interference could be an important molecular mechanism affecting the conservation of adaptive genomic regions.

Epigenetic regulatory mechanisms present a promising area of study for understanding rapid adaptive evolution [5]. There are multiple examples of adaptive variation due to differential transcription [32,68,112,145]. However, there is a gap in our knowledge of how epigenetic variation arises. Transposon activity has also been proposed as a mechanism involved in adaptive variation within regulatory mechanisms [77,127]. Altogether, these developments suggest that there could be mechanisms in place that leads to an increased occurrence of adaptive variation within specific regions of the genome. Investigating the genome structure could provide further insights into how adaptive differential transcription occurs, particularly transient adaptive variation, which is likely regulated by epigenetic regulatory mechanisms.

Investigations into the mechanisms of adaptive evolution have "evolved" with novel technologies and the continued increase in accessibility to genomic, transcriptomic, and proteomic data. Now, it is possible to go beyond identifying signatures of selection and to investigate the precise molecular mechanisms conferring adaptive variation under specific environmental contexts for particular traits. The key role of regulatory mechanisms in adaptive evolution has become evident. Thus, going forward, it may be worthwhile to continue employing approaches that investigate not only mutations and differential transcription but also how these changes have been conserved, and to consider the potential beneficial impact of losses (e.g., downregulation or loss of function). Further research into regulatory elements (e.g., silencers, enhancers, miRs, piwiRNAs, and epigenetic modifications) could provide significant insight into the mechanisms by which vertebrates permanently or transiently adapt to their environments.

Table 2. Complete list of genes/proteins names discussed in this review, their ontology, and their mechanisms of evolution. D: differential transcription; S: sequence variation/mutation; L: gene loss or pseudogenization; E: gene expansion; -: adaptive evolution of this gene not mentioned in the studies discussed.

Adaptive Evolution in Visual Systems "in Light of" Varying Lighting Conditions			
Abbreviation	**Full Name**	**Gene Ontology**	**Mechanisms**
CYP27C1	Cytochrome P450 Family 27	Cytochrome enzyme that catalyzes desaturation from Vitamin A1, converting it to Vitamin A2.	D
GRK1	G-Protein-Coupled Receptor Kinase	G-Protein receptor kinase, phosphorylates rhodopsin and initiates deactivation.	S
LUM	Lumican	Proteoglycan that, in retinae, regulates organization of collagen fibers.	S
LWS	Long-Wavelength-Sensitive Opsin	G-protein-coupled receptor involved in the phototransduction process, sensitive to "reds" and "greens" between 501 and 573 nm.	S
RH1	Rhodopsin	G-protein-coupled receptor involved in the phototransduction process expressed in rods, absorbs in 447–525 nm.	S, L
RH2	Rhodopsin	G-protein-coupled receptor involved in the phototransduction process expressed in rods, absorbs in 452–537 nm.	D, S, L

RHO	Rhodopsin	G-protein-coupled receptor involved in the phototransduction activity.	D
SWS1	Short-Wavelength-Sensitive Opsin	G-protein-coupled receptor involved in the phototransduction process, sensitive to "violet", including ultraviolet between 347 and 383 nm.	D, L
SWS2	Short-Wavelength-Sensitive Opsin 2	G-protein-coupled receptor involved in the phototransduction process, sensitive to "blue-violet" between 397 and 482 nm.	D, S, L
VHA	Vacuolar-Type H+ -ATPase	Mitochondrial protein that participates in the excretion of H+ from endothelial cells into the lumen.	S
Adaptations for colonization of aquatic and terrestrial environments			
ACAN	Aggrecan	Integral part of extracellular membrane within cartilaginous tissue.	S
AMPD3	Adenosine Monophosphate Deaminase 3	Highly regulated enzyme that hydrolytically deaminates adenosine monophosphate into inosine monophosphate within the adenylate catabolic pathway.	L
ATP8	Mitochondrially Encoded ATP Synthetase Membrane Subunit 8	ATP synthetase that produces ATP from ADP when a protein gradient is present.	S
CALHM1	Calcium Homeostasis Modulator 1	Calcium–ion channel involved in sweet, bitter, and umami taste transduction.	L
DSC1	Desmocollin 1	Calcium-dependent glycoprotein found primarily in epidermal cells constituting adhesive proteins within desmosome cell–cell junctions.	L
DSG4	Desmocollin 4	Transmembrane component of desmosomes.	L
ELK1	ETS Transcription Factor ELK1	Transcription factor that regulates expression by binding to the promoter of the serum response factor gene.	-
FSHR	Follicle-Stimulating Hormone Receptor	Receptor for the follicle-stimulating hormone and functions in gonad development.	S
GNAT3	G Protein Subunit α Transducin 3	G protein subunit involved in bitter, sweet, and umami taste transduction.	L
GSDMA	Gasdermin A	Precursor protein of a cell-pore-forming protein.	L
HOX	Homeobox Domain Genes	Family of transcription factor genes involved in body plan along animal bilateral axis.	-
HOXD11	Homeobox D11	Transcription factor part of a family involved in limb and genital development.	S
HOXD12	Homeobox D12	Transcription factor part of a family involved in limb and genital development.	S
HOXD13	Homeobox D13	Transcription factor part of a family involved in limb and genital development.	S
KRT20	Keratin 20	Intermediate filament conferring structural integrity to epidermal cells.	L
KRT9	Keratin 9	Intermediate filament chain expressed in terminally differentiated epidermal cells.	L
LHX3	LIM Homeobox 3	Transcription factor required for pituitary development and motor neuron specification.	-
LYG1	Lysozyme G1	Lysozyme activity	L
LYG2	Lysozyme G2	Lysozyme activity	
MB	Myoglobin	Iron- and oxygen-binding protein typically present in skeletal muscle tissue.	E

MLL	Lysine Methyltransferase 2A	Currently called KMT2A, co-activator involved in transcriptional regulation of genes during early development and hematopoiesis.	S
MMP12	Matrix Metalloproteinase 12	Involved in extracellular matrix breakdown	L
ORA1	Olfactory Receptor Class-A-Like-Protein-1	Olfactory receptor.	E
PIT-1	POU Class 1 Homeobox 1	Transcription factor involved in regulating expression of multiple pituitary development genes and hormones.	S
SULT6B1	Sulfotransferase Family 6B Member 1	Sulfotransferase.	E
TGM5	Transglutaminase 5	Transglutaminase, enzyme that catalyzes crosslinking between glutamine and lysine residues.	L
Adapting to extreme environmental conditions (i.e., hypoxia, salinity, and low temperatures)			
ABCA12	ATP-Binding-Cassette Subfamily A Member 12	ATP-Binding-Cassette transporter protein.	S
AFGP	Antifreeze Glycoprotein	Proteins that can inhibit growth of ice.	S
ALDOA	Aldolase	Glycolytic enzyme.	S
BMP6	Bone Morphogenetic Protein 6	Ligand of the transforming growth factor β, involved in multiple regulatory processes (e.g., iron homeostasis, fat and bone development, and ovulation).	S
CA	Carbonic Anhydrase	Family of genes of zinc metalloenzymes that catalyze reversible hydration of carbon dioxide.	E
COX1	Cytochrome C Oxidase I	Mitochondrial protein, component of the cytochrome C oxidase.	S
COX3	Cytochrome C Oxidase III	Mitochondrial protein, components of the cytochrome C oxidase.	S
CSAD	Cysteine Sulfinic Acid Decarboxylase	Member of the group 2 decarboxylase.	D
CTH	Cystathionine γ-Lyase	Cytoplasmic enzyme involved in the transculturation pathway.	D
EDA	Ectodysplasin A	Membrane protein thought to be involved in cell–cell signaling during development.	S
ENO	Enolase	Glycolytic enzyme.	S
EPAS1	Endothelial Pas Domain Protein 1	Transcription factor involved in the regulation of genes that are controlled by oxygen.	S
EPO	Erythropoietin	Secreted glycosylated cytokine involved in promoting red blood cell production.	D, S
FBP1	Fructose-Bisphosphatase 1	Glucogenesis regulatory enzyme.	S
FBP2	Fructose-Bisphosphatase 2	Glucogenesis regulatory enzyme.	D
G6PCB	Glucose-6-Phosphatase	Involved in the gluconeogenesis pathway.	D
GDF6	Growth Differentiation Factor 6	Ligand of the transforming growth factor β, involved in regulation of genes associated with formation of some bones, joints, limbs, skull, and axial skeleton.	S

GLA	Galactosidase α	Glycoprotein involved in termina hydrolyses from glycolipids and glycoproteins.	S
GPI	Glucose-6-Phosphate Isomerase	Multiple functions; glycolytic enzyme and neurotrophic factor that promote survival of skeletal motor neurons and sensory neurons intracellularly and extracellularly, respectively.	S
GST	Gluthathione-S-transferase	Conjugation of reduced glutathionone, important in detoxification.	E
HB	Hemoglobin	Transport of oxygen.	S
IGFLR1	IGF-Like Family Receptor 1	Possibly a cell membrane receptor for IGF-like proteins.	S
KITLG	KIT Ligand	Ligand of the tyrosine-kinase receptor, a pleiotropic factor involved in embryonic development.	S
LDHA	Lactate Dehydrogenase A	Involved in the pyruvate fermentation to lactate pathway.	S
LDHD	Lactate Dehydrogenase D	Lactate dehydrogenase.	S
NDUFA10	NADH Dehydrogenase 1 α Subcomplex Subunit 10	Subunit of the first enzyme complex in the electron transport chain in the inner mitochondrial membrane.	S
NDUFA9	NADH Dehydrogenase 1 α Subcomplex Subunit 9	Subunit of the first enzyme complex in the electron transport chain in the inner mitochondrial membrane.	S
NDUFAB1	NADH Dehydrogenase 1 α/β Subcomplex 1	Non-catalytic subunit of the NADH complex involved in the mitochondrial inner membrane electron transport.	S
NDUFC2	NADH Ubiquinone Oxidoreductase Subunit C2	Suspected to be non-catalytic, a subunit of the NADH complex involved in the mitochondrial inner membrane electron transport.	S
NDUFV3	NADH Ubiquinone Oxidoreductase Subunit V3	Suspected to be non-catalytic, a subunit of the NADH complex involved in the mitochondrial inner membrane electron transport.	S
NPR1	Natriuretic Peptide Receptor 1	Guanylyl cyclase involved in catalyzing the production of cGMP from GTP.	E
PC	Pyruvate Carboxylase	Involved in glucogenesis, lipogenesis, insulin secretion, and glutamate.	S
PCK1	Phosphoenolpyruvate Carboxykinase 1	Main control point in the glucogenesis.	S
PCXB	Pyruvate Carboxylase	Catalyzes the carboxylation of pyruvate to oxaloacetate a process that requires biotin and ATP.	D
PGC-1	PPARG Coactivator 1 α	Transcriptional coactivator involved in regulation of genes involved in energy metabolism.	S
PITX1	Paired-Like Homeodomain 1	Transcription factor part of a family involved in organ development and bilateral symmetry.	S
PKLR	Pyruvate Kinase L/R	Pyruvate kinase that catalyzes the rate-limiting step in glycolysis.	D
SLC12A3	Solute Carrier Family 12 Member 3	Important in electrolyte homeostasis, cotransporter that mediates sodium and chloride reabsorption.	S
SOD	Sodium Peroxide Dismutase 1	Binds copper and zinc ions. Functions to destroy free radicals in the body.	E

TRPM8	Transient Receptor Potential Cation Channer Subfamily Member 8	Receptor-activated non-selection cation channel, suggested to be involved in low temperature sensation.	S
UCP1	Uncoupling Protein 1	Mitochondrial anion carrier protein.	D
VIPR1	Vasoactive Intestinal Polypeptide Receptor 1	Receptor for the vasoactive intestinal neuropeptide.	S
VHL	Von Hippel–Lindau Tumor Suppressor	Involved in ubiquitination of the hypoxia-inducible factor.	S
VOM1	Vitelline Membrane Outer Layer 1	Vitelline membrane outer layer proteins.	E
ZP	Zona Pellucida Glycoproteins	Proteins involved in the composition and fertilization functions, and preimplantation development in the zona pellucida.	E
		Adaptations to dietary changes	
ACAD9	Acyl-CoA Dehydrogenase Family Member 9	Localized in mitochondria and involved in catalyzing the rate limiting step of the β-oxidation of fatty acyl-CoA.	S
AGT	Alanine-Glyoxylate Aminotransferase	Also known as AGXT, involved in glyoxylate detoxification. In carnivores, the AGT is needed in the mitochondria, while in herbivores peroxisomes, and omnivores, have AGT in both organelles.	E
AMY	α-Amylase	Carbohydrase, specifically amylase.	S
AMY1B	Amylase α 1B	Secreted protein that catalyzes the first steps in digestion of starch and glycogen.	L
ASPM	Assembly Factor For Spindle Microtubules	Essential in mitotic spindle function in embryonic neuroblasts.	S
BARX1	BARX Homeobox 1	Homeobox transcription factor, suggested to be involved in teeth and craniofacial mesenchyme of neural crest origin.	S
BMP2	Bone Morphogenetic Protein 2	Ligand of the transforming growth factor β, involved in bone and cartilage development.	S
BMP4	Bone Morphogenetic Protein 4	Ligand of the transforming growth factor β, involved in heart development and adipogenesis.	S
CDK5RAP2	Cyclin-Dependent-Kinase 5 Regulatory Subunit-Associated Protein 2	Plays a role in centriole engagement and microtubule nucleation.	S
CEP152	Centrosomal Protein 152	Thought to play a role in centrosome function.	S
CYP7A1	Cytochrome P450 Family 7 Subfamily A Member 1	Monooxygenase with lipase catalytic function.	S
DGAT2	Diacylglycerol O-Acyltransferase 2	Enzyme involved in catalyzing the final reaction in triglyceride synthesis.	D
DLX2	Distal-Less Homeobox 2	Homeobox transcription factor, postulated to play a role in forebrain and craniofacial development.	S
EDAR	Ectodysplasin A Receptor	Receptor for Ectodyplasin A and can activate multiple cell death pathways.	S

EHHADH	Enoyl-CoA Hydratase	Bifunctional enzyme part of the peroxisomal β-oxidation pathway.	E
EVE1	EVE1	Transcription factor.	S
IFG-1	Insulin Growth Factor 1	Similar to insulin in function and structure, involved in mediating growth and development.	S
IRX5A	Iroquis Homeobox 5A	Transcription factor involved in cell development, embryonic skeletal joint development, chondrocyte differentiation, and neuron differentiation.	L
LHX7	LIM Homeobox 7	Transcription factor with zinc-finger motifs, involved in patterning and differentiation of multiple tissue types.	S
LHX8	LIM Homeobox 8	Transcription factor with zinc-finger motifs, involved in patterning and differentiation of multiple tissue types.	S
LIPF	Lipase F Gastric Type	Gastric lipase involved in digestion of triglycerides.	S
MAO	Monoamine Oxidase A	Enzyme involved in the catalysis of the oxidative deamination of amines of three neurotransmitters (i.e., dopamine, norepinephrine, and serotonin).	S
MCPH1	Microcephalin 1	DNA damage response protein.	L
NOX4	NADPH Oxidase 4	Catalytic subunit of the NADPH oxidase complex.	L
ODAM	Odontogenic Ameloblast-Associated Protein Precursor	Tooth-associated epithelial protein thought to have a role in odontogenesis.	D
PAX9	Paired Box 9	Paired box domain transcription factor with a critical role in fetal development and cancer growth.	S
PGA	Pepsin A	Enzyme that participates in digestion by activating pepsinogen.	S
PITX2	Paired-Like Homeodomain 2	Homeodomain transcription factor, regulates procollagen lysyl hydroxylase gene expression, as well as basal and hormone regulated activity of prolactin. Involved in development of the eye, tooth, and abdominal organs.	S
PLRP2	Pancreatic Lipase-Related Protein 2	Lipase for hydrolyzation of galactolipids.	E
RNASE1	Ribonuclease A Family Member 1 Pancreatic	Pancreatic-type of secretory ribonuclease.	E
RPFA	Resuscitation-Promoting Factor A	Protein with peptidoglycan hydrolytic activity functions as a factor that stimulates resuscitation of dormant cells.	D
RUNX2	RUNX Family Transcription Factor 2	Transcription factor with Runt DNA-binding domain, essential for osteoblastic differentiation and skeletal morphogenesis.	S
SALL1A	Spalt-Like Transcription Factor 1	Transcription factor with zinc-finger motifs.	L
SCN8A	Sodium Voltage-Gated Channel α Subunit 8	Forms the ion-pore region of the voltage-gated sodium channel. Involved in depolarization during the formation of an action potential in excitable neurons.	S
SCN9A	Sodium Voltage-Gated Channel α Subunit 9	Ion channel activity and sodium ion binding.	S
SCNA4A	Sodium Voltage-Gated Channel α Subunit 4	Transmembrane glycoprotein, involved in generation and propagation of action potentials in neurons and muscle.	S

SCPP5	Secretory Calcium-Binding Phosphoprotein 5	Secretory Calcium-Binding Phosphoprotein.	D
SHH	Sonic Hedgehog	Protein in the Sonic Hedgehog signaling pathway, essential in patterning in early embryonic development.	S
SHOX	Short Stature Homeobox	Homeodomain transcription factor	L
SLC2A3	Solute Carrier Family 2 Member 3	Enables dehydroascorbic acid transmembrane transporter activity	D
TAS2R	Taste 2 Receptor	Transmembrane G-protein-coupled receptor involved in ability to taste glucosinolates and bitter compounds in plants.	L
TBX4	T-Box Transcription Factor 4	Transcription factor with a T-Box binding domain. It has been suggested to play a role in limb development.	L
TFAP2A	Transcription Factor AP 2 α	Transcription factor that works both as a suppressor and activator of multiple genes.	S
TST	Thiosulfate Sulfurtransferase	Catalyzes the conversion of thiosulfate and cyanide to thiocyanate and sulfite	D
UNK	Unk Zinc Finger	RNA-binding protein involved in establishment and maintenance of the early morphology of cortical neurons in embryonic development.	D
WDR62	WD Repent Domain 62	Proposed to play a role in cerebral cortical development.	S
		Environmental Pathogens	
APOBEC	Apolipoprotein B MRNA Editing Enzyme Catalytic Subunit 3G	Catalyzes site-specific deamination of RNA and single-stranded DNA.	S
CD14	Myeloid Cell-Specific Leucine-Rich	Surface antigen preferentially expressed on monocytes/macrophages.	S
CD22	Sialic Acid-Binding Ig-Like Lectin 2	Carbohydrate binding protein. Mediates B-cell/B-cell interactions. Thought to localize B-cells in lymphoid tissues.	E
CD40	Tumor Necrosis Factor Receptor Superfamily Member 5	Receptor on antigen-presenting cells of the immune system.	S
CD80	T-Lymphocyte Activation Antigen CD80	Membrane receptor activated by CD28 involved in costimulatory signal essential for T-lymphocyte activation.	S
IFNAR2	Interferon α and β Receptor Subunit 2	Type I membrane protein, forms one of the chains of a interferon α and β receptor.	S
IGH	Immunoglobulin Heavy Locus	Antigen-binding protein involved in the Lectin-induced complement pathway and NFAT immune response.	S
IGHE	Immunoglobulin Heavy Constant Epsilon	Antigen-binding protein involved in Interleukin 4-mediated and cytokine signaling.	S
IGHM	Immunoglobulin Heavy Constant Delta	Antigen binding.	S
LY96	Lymphocyte Antigen 96	Confers responsiveness to lipopolysaccharides when it associates with a Toll-like receptor.	S
MHC	Major Histocompatibility Complex	Family of genes associated with antigen processing.	E

	class I		
MDA5	Interferon Induced with Helicase C Domain 1	Intracellular sensor of viral RNA that triggers innate immune response.	S
MUC7	Mucin 7-Secreted	Salivary mucin thought to play a role in facilitating clearance of bacteria in the mouth.	S
MX	MX Dynamin-Like GTPase	Guanosine triphosphate metabolizing protein that participates in antiviral response.	S
NLPRP3	NLR Family Pyrin Domain-Containing 3	Peptidoglycan binding protein.	E
SSC4D	Scavenger Receptor Cysteine-Rich Family Member with 4 Domains	Scavenger receptor activity.	E
TAB 1	TGF β-Activated Kinase 1	Regulator of the MAP kinase kinase kinase. Thus, mediates various intracellular signaling pathways.	S
TICAM1	Toll-Like Receptor Adaptor Molecule 1	Adaptor protein with Toll/Interleukin 1-receptor homology domain. Protein kinase binding and obsolete signal transducer activity.	S
TLR	Toll-Like Receptors	Family of genes that play a fundamental role in pathogen recognition and activation of innate immunity.	-
TLR1	Toll-Like Receptor 1	Non-viral pathogen recognition, protein heterodimerization activity, and transmembrane signaling.	E
TLR2	Toll-Like Receptor 2	Non-viral pathogen recognition, protein heterodimerization activity, and transmembrane signaling.	E, S
TLR4	Toll-Like Receptor 4	Non-viral pathogen recognition, lipopolysaccharide binding.	S
TLR5	Toll-Like Receptor 5	Non-viral pathogen recognition, protein heterodimerization activity, and transmembrane signaling.	S
TLR7	Toll-Like Receptor 7	Transmembrane signaling receptor activity and double-stranded RNA binding.	E, S
TLR8	Toll-Like Receptor 8	RNA binding and drug binding.	E, S
TLR9	Toll-Like Receptor 9	Transmembrane signaling receptor activity and siRNA binding.	E
TLR22	Toll-Like Receptor 22	Transmembrane signaling receptor activity.	E
TLR23	Toll-Like Receptor 23	Receptor activity.	E
TLR25	Toll-Like Receptor 25	Transmembrane signaling receptor activity.	E
		Other Factors	
MSX2	Homeobox Protein MSX-2	Transcription factor with a homeobox-binding domain.	E
C6AST	Calpastatin	Endogenous calcium-dependent cysteine protease inhibitor.	E
PASTN	Pasttristacin	Metalloendopeptidase activity.	E
CRISP	Cysteine-Rich Secretory Proteins	Associated with reptilian venom production and mammalian reproduction.	E
PER2	Period Circadian Regulator 2	Transcription factor and "activator" activity. Primary circadian pacemaker in the mammalian brain.	S

13. Conclusions

Based on the research thus far, the main molecular mechanisms underlying adaptive evolution in vertebrates are sequence variation/mutations. These mutations commonly occur in regulatory elements or functional domains of proteins, leading to regulatory modifications such as changes in gene expression, or cellular and/or signaling pathways.

Therefore, modifications in gene expression patterns, and cellular and/or signaling pathways are important molecular mechanisms underlying adaptive variation identified in this review. Thus far, most studies on this topic have aimed to identify selected regions and mutations associated with adaptive phenotypes, although there has been an increase in studies investigating differential transcription. Studies on the molecular impact of mutations (e.g., the effect on the translated protein or what regulatory changes led to the differential transcription) have been rare. Only one of the reviewed studies employed both a genomic analysis and functional assays to investigate the impact of the mutation they identified [117]. The more frequent use of such an approach is needed to advance our knowledge of molecular mechanisms of adaptive evolution. More research into RNA mediated regulation (e.g., miRs and piwiRNAs) and epigenetics as pertaining to adaptive evolution is also needed. Currently, such studies may be limited by the accessibility to samples, the high costs of the necessary assays, and/or access to fully annotated genome sequences. As the field continues to grow, these limitations should continue to be reduced.

Cellular regulatory pathways are complex, and changes in function (e.g., binding affinity) or absence of a protein can considerably impact phenotypes. The possibility that gene losses could be adaptive is a debated topic and is subsequently also understudied [13]. There is evidence that miRs' sequences can be conserved elements and that they have only been reported to downregulate genes, which may have a similar effect as gene loss. There are also other mechanisms of gene downregulation [9,10]. Some of them can be permanent modifications (e.g., mutation in a regulatory element) or transient (e.g., epigenetic). In some contexts, the downregulation or loss of a gene may yield an advantageous phenotype, e.g., the downregulation or loss of a gene in a regulatory pathway may result in a downstream effect that changes the final molecular response and yields a novel phenotype. Finally, the genome architecture should also be further explored, as there have been some interesting biochemical findings explaining why mutations are more likely to occur in some areas over others.

Author Contributions: Conceptualization, F.M.S. and M.P.; methodology, F.M.S.; writing—original draft preparation, F.M.S.; writing—review and editing, F.M.S. and M.P.; supervision, M.P. Both authors have read and agreed to the published version of the manuscript.

Funding: M. Pilot was funded by the Polish National Agency for Academic Exchange (NAWA, Polish Returns Fellowship PPN/PPO/2018/1/00037).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created in this study. All the papers used in the systematic review are cited in the reference list [3–152].

Acknowledgments: We thank Francisco J. Cintrón-Jiménez for his assistance in figure development. We thank A. Moura for his comments on an early version of this manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Beldade, P.; Mateus, A.R.; Keller, R.A. Evolution and molecular mechanisms of adaptive developmental plasticity. *Mol. Ecol.* **2011**, *20*, 1347–1363. https://doi.org/10.1111/j.1365-294x.2011.05016.x.
2. Olson-Manning, C.F.; Wagner, M.R.; Mitchell-Olds, T. Adaptive evolution: Evaluating empirical support for theoretical predictions. *Nat. Rev. Genet.* **2012**, *13*, 867–877. https://doi.org/10.1038/nrg3322.
3. Wang, B.; Chen, L.; Wang, W. Genomic insights into ruminant evolution: From past to future prospects. *Zool. Res.* **2019**, *40*, 476–48. https://doi.org/10.24272/j.issn.2095-8137.2019.061.
4. Slodkowicz, G.; Goldman, N. Integrated structural and evolutionary analysis reveals common mechanisms underlying adaptive evolution in mammals. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 5977–5986. https://doi.org/10.1073/pnas.1916786117.
5. Hao, Y.; Qu, Y.; Song, G.; Lei, F. Genomic Insights into the Adaptive Convergent Evolution. *Curr. Genom.* **2019**, *20*, 81–89. https://doi.org/10.2174/1389202920666190313162702.

6. Carleton, K.L.; Escobar-Camacho, D.; Stieb, S.M.; Cortesi, F.; Marshall, N.J. Seeing the rainbow: Mechanisms underlying spectral sensitivity in teleost fishes. *J. Exp. Biol.* 2020, 223, jeb193334. https://doi.org/10.1242/jeb.193334.
7. Da Silva, F.A.; Corrêa Guimarães, E.M.; Carvalho, N.; Ferreira, A.; Schneider, C.H.; Carvalho-Zilse, G.A.; Feldberg, E.; Gross, M. Transposable DNA Elements in Amazonian Fish: From Genome Enlargement to Genetic Adaptation to Stressful Environments. *Cytogenet. Genome Res.* **2020**, *160*, 148–155. https://doi.org/10.1159/000507104.
8. Perry, B.W.; Card, D.C.; McGlothlin, J.W.; Pasquesi, G.I.M.; Adams, R.H.; Schield, D.R.; Hales, N.R.; Corbin, A.B.; Demuth, J.P.; Hoffmann, F.G. et al. Molecular Adaptations for Sensing and Securing Prey and Insight into Amniote Genome Diversity from the Garter Snake Genome. *Genome Biol. Evol.* **2018**, *10*, 2110–2129. https://doi.org/10.1093/gbe/evy157.
9. Luo, J.; Wang, Y.; Yuan, J.; Zhao, Z.; Lu, J. MicroRNA duplication accelerates the recruitment of new targets during vertebrate evolution. *RNA* **2018**, *24*, 787–802. https://doi.org/10.1261/rna.062752.117.
10. Penso-Dolfin, L.; Haerty, W.; Hindle, A.; Di Palma, F. microRNA profiling in the Weddell seal suggests novel regulatory mechanisms contributing to diving adaptation. *BMC Genom.* **2020**, *21*, 303. https://doi.org/10.1186/s12864-020-6675-0.
11. Li, K.; Sun, X.; Chen, M.; Sun, Y.; Tian, R.; Wang, Z. Evolutionary changes of Hox genes and relevant regulatory factors provide novel insights into mammalian morphological modifications. *Integr. Zool.* **2018**, *13*, 21–35. https://doi.org/10.1111/1749-4877.12271.
12. Sun, Y.B.; Zhang, Y.; Wang, K. Perspectives on studying molecular adaptations of amphibians in the genomic era. *Zool. Res.* **2020**, *41*, 351–364. https://doi.org/10.24272/j.issn.2095-8137.2020.046.
13. Lin, Q.; Fan, S.; Zhang, Y.; Xu, M.; Zhang, H.; Yang, Y.; Xia, W.; Liu, C.; Zhu, W.; Wang, H.; et al. The seahorse genome and the evolution of its specialized morphology. *Nature* **2016**, *540*, 395–399. https://doi.org/10.1038/nature20595.
14. Bloch, N.I. The evolution of opsins and color vision: Connecting genotype to a complex phenotype. *Acta Biol. Colomb.* **2016**, *21*, 481–494. https://doi.org/10.15446/abc.v21n3.53907.
15. Bian, C.; Huang, Y.; Li, J.; You, X.; Yi, Y.; Ge, W.; Shi, Q. Divergence, evolution and adaptation in ray-finned fish genomes. *Sci. China Life Sci.* **2019**, *62*, 1003–1018. https://doi.org/10.1007/s11427-018-9499-5.
16. Katti, C.; Stacey-Solis, M.; Coronel-Rojas, N.A.; Davies, W.I.L. The Diversity and Adaptive Evolution of Visual Photopigments in Reptiles. *Front. Ecol. Evol.* **2019**, *7*, 352. https://doi.org/10.3389/fevo.2019.00352.
17. Musilova, Z.; Salzburger, W.; Cortesi, F. The Visual Opsin Gene Repertoires of Teleost Fishes: Evolution, Ecology, and Function. *Annu. Rev. Cell Dev. Biol.* **2021**, *37*, 441–468. https://doi.org/10.1146/annurev-cellbio-120219-024915.
18. Torres-Dowdall, J.; Pierotti, M.E.R.; Härer, A.; Karagic, N.; Woltering, J.M.; Henning, F.; Elmer, K.R.; Meter, A. Rapid and Parallel Adaptive Evolution of the Visual System of Neotropical Midas Cichlid Fishes. *Mol. Biol. Evol.* **2017**, *34*, 2469–2485. https://doi.org/10.1093/molbev/msx143.
19. Lin, J.J.; Wang, F.Y.; Li, W.H.; Wang, T.Y. The rises and falls of opsin genes in 59 ray-finned fish genomes and their implications for environmental adaptation. *Sci. Rep.* **2017**, *7*, 15568. https://doi.org/10.1038/s41598-017-15868-7.
20. Carleton, K.L.; Dalton, B.E.; Escobar-Camacho, D.; Nandamuri, S.P. Proximate and ultimate causes of variable visual sensitivities: Insights from cichlid fish radiations. *Genesis* **2016**, *54*, 299–325. https://doi.org/10.1002/dvg.22940.
21. Hauser, F.E.; Ilves, K.L.; Schott, R.K.; Alvi, E.; López-Fernández, H.; Chang, B.S.W. Evolution, inactivation and loss of short wavelength-sensitive opsin genes during the diversification of Neotropical cichlids. *Mol. Ecol.* **2021**, *30*, 1688–1703. https://doi.org/10.1111/mec.15838.
22. Cai, Y.; Fan, Y.; Yue, Y.; Li, P.; Yan, J.; Zhou, K. Molecular Evolution of Visual Opsin Genes during the Behavioral Shifts between Different Photic Environments in Geckos. *Asian Herpetol. Res.* **2021**, *12*, 280–288. https://doi.org/10.16373/j.cnki.ahr.200124.
23. Dungan, S.Z.; Kosyakov, A..; Chang, B.S. Spectral Tuning of Killer Whale (*Orcinus orca*) Rhodopsin: Evidence for Positive Selection and Functional Adaptation in a Cetacean Visual Pigment. *Mol. Biol. Evol.* **2016**, *33*, 323–336. https://doi.org/10.1093/molbev/msv217.
24. Zheng, S.; Shao, F.; Tao, W.; Liu, Z.; Long, J.; Wang, X.; Zhang, S.; Zhao, Q.; Carleton, K.; L.; Kocher, T.D. Chromosome-level assembly of southern catfish (*Silurus meridionalis*) provides insights into visual adaptation to nocturnal and benthic lifestyles. *Mol. Ecol. Resour.* **2021**, *21*, 1575–1592. https://doi.org/10.1111/1755-0998.13338.
25. Zhang, Z.; Liu, Y.; Zhang, W.; Du, X.; Liu, J. Benthic visual adaptation by fine-tuning light sensitivity in Japanese flounder (*Paralichthys olivaceus*). *Front. Mar. Sci.* **2022**, *9*, 1019660. https://doi.org/10.3389/fmars.2022.1019660.
26. Ito, R.K.; Harada, S.; Tabata, R.; Watanabe, K. Molecular evolution and convergence of the rhodopsin gene in *Gymnogobius,* a goby group having diverged into coastal to freshwater habitats. *J. Evol. Biol.* **2021,** 35, 333–346. https://doi.org/10.1111/jeb.13955.
27. Ishengoma, E.; Agaba, M.; Cavener, D.R. Evolutionary analysis of vision genes identifies potential drivers of visual differences between giraffe and okapi. *PeerJ* **2017**, *5*, e3145. https://doi.org/10.7717/peerj.3145.
28. Wu, Y.; Hadly EA.; Teng, W.; Hao, Y.; Liang, W.; Liu, Y.; Wang, H. Retinal transcriptome sequencing sheds light on the adaptation to nocturnal and diurnal lifestyles in raptors. *Sci. Rep.* **2016**, *6*, 33578. https://doi.org/10.1038/srep33578.
29. Wu, Y.; Wang, H.; Hadly, E.A. Invasion of Ancestral Mammals into Dim-light Environments Inferred from Adaptive Evolution of the Phototransduction Genes. *Sci. Rep.* **2017**, *7*, 46542. https://doi.org/10.1038/srep46542.
30. Luehrmann, M.; Stieb, S.M.; Carleton, K.L.; Pietzker, A.; Cheney, K.L.; Marshall, N.J. Short-term colour vision plasticity on the reef: Changes in opsin expression under varying light conditions differ between ecologically distinct fish species. *J. Exp. Biol.* 2018, 221, jeb175281. https://doi.org/10.1242/jeb.175281.
31. Rennison, D.J.; Owens, G.L.; Heckman, N.; Schluter, D.; Veen, T. Rapid adaptive evolution of colour vision in the threespine stickleback radiation. *Proc. Biol. Sci.* **2016**, *283*, 20160242. https://doi.org/10.1098/rspb.2016.0242.

32. Mehta, T.K.; Koch, C.; Nash, W.; Knaack, S.A.; Sudhakar, P.; Olbei, M.; Bastkowski, S.; Penso-Dolfin, L.; Korcsmaros, T.; Haerty, W.; et al. Evolution of regulatory networks associated with traits under selection in cichlids. *Genome Biol.* **2021**, *22*, 25. https://doi.org/10.1186/s13059-020-02208-8.
33. Mehta, T.K.; Penso-Dolfin, L.; Nash, W.; Roy, S.; Di-Palma, F.; Haerty, W. Evolution of miRNA-Binding Sites and Regulatory Networks in Cichlids. *Mol. Biol. Evol.* **2022**, *39*, 146. https://doi.org/10.1093/molbev/msac146.
34. Wilwert, E.; Etienne, R.S.; van de Zande, L.; Maan, M.E. Contribution of opsins and chromophores to cone pigment variation across populations of Lake Victoria cichlids. *J. Fish Biol.* **2022**, *101*, 365–377. https://doi.org/10.1111/jfb.14969.
35. Mohun, S.M.; Davies, W.I.L. The Evolution of Amphibian Photoreception. *Front. Ecol. Evol.* **2019**, *7*, 321. https://doi.org/10.3389/fevo.2019.00321.
36. van Nynatten, A.; Janzen, F.H.; Brochu, K.; Maldonado-Ocampo, J.A.; Crampton, W.G.R.; Chang, B.S.W. To see or not to see: Molecular evolution of the rhodopsin visual pigment in neotropical electric fishes. *Proc. Biol. Sci.* **2019**, *286*, 20191182. https://doi.org/10.1098/rspb.2019.1182.
37. Wu, Y. Widespread nocturnality of living birds stemming from their common ancestor. *BMC Evol. Biol.* **2019**, *19*, 189. https://doi.org/10.1186/s12862-019-1508-y.
38. Damsgaard, C.; Lauridsen, H.; Harter, T.S.; Kwan, G.T.; Thomsen, J.S.; Funder, A.M.; Supuran, C.T.; Tresguerres, M.; Matthews, P.G.; Brauner, C.J. A novel acidification mechanism for greatly enhanced oxygen supply to the fish retina. *eLife* **2020**, *9*, e58995. https://doi.org/10.7554/elife.58995.
39. Partha, R.; Chauhan, B.K.; Ferreira, Z.; Robinson, J.D.; Lathrop, K.; Nischal, K.K.; Chikina, M.; Ckarmk, N.L. Subterranean mammals show convergent regression in ocular genes and enhancers, along with adaptation to tunneling. *eLife* **2017**, *6*, e25884. https://doi.org/10.7554/elife.25884.
40. Borges, R.; Fonseca, J.; Gomes, C.; Johnson, W.E.; O'Brien, S.J.; Zhang, G.; Gilbers, M.T.P.; Jarvis, E.D.; Atunes, A. Avian Binocularity and Adaptation to Nocturnal Environments: Genomic Insights from a Highly Derived Visual Phenotype. *Genome Biol. Evol.* **2019**, *11*, 2244–2255. https://doi.org/10.1093/gbe/evz111.
41. Baldwin, M.W.; Ko, M.C. Functional evolution of vertebrate sensory receptors. *Horm. Behav.* **2020**, *124*, 104771. https://doi.org/10.1016/j.yhbeh.2020.104771.
42. Jiang, M.; Shi, L.; Li, X.; Dong, Q.; Sun, H.; Du, Y.; Zhang, Y.; Shao, T.; Cheng, H.; Chen, W.; et al. Genome-wide adaptive evolution to underground stresses in subterranean mammals: Hypoxia adaption, immunity promotion, and sensory specialization. *Ecol. Evol.* **2020**, *10*, 7377–7388. https://doi.org/10.1002/ece3.6462.
43. Zhao, Y.; Chen, H.; Li, C.; Chen, S.; Xiao, H. Comparative Transcriptomics Reveals the Molecular Genetic Basis of Cave Adaptability in Sinocyclocheilus Fish Species. *Front. Ecol. Evol.* **2020**, *8*, 636503. https://doi.org/10.1002/ece3.7024.
44. Yuan, Z.; Liu, S.; Zhou, T.; Tian, C.; Bao, L.; Dunham, R.; Liu, Z. Comparative genome analysis of 52 fish species suggests differential associations of repetitive elements with their living aquatic environments. *BMC Genom.* **2018**, *19*, 141. https://doi.org/10.1186/s12864-018-4516-1.
45. Li, N.; Bao, L.; Zhou, T.; Yuan, Z.; Liu, S.; Dunham, R.; Li, Y.; Wang, K.; Xu, X.; Jin, Y.; et al. Genome sequence of walking catfish (*Clarias batrachus*) provides insights into terrestrial adaptation. *BMC Genom.* **2018**, *19*, 952. https://doi.org/10.1186/s12864-018-5355-9.
46. Tian, R.; Yin, D.; Liu, Y.; Seim, I.; Xu, S.; Yang, G. Adaptive Evolution of Energy Metabolism-Related Genes in Hypoxia-Tolerant Mammals. *Front. Genet.* **2017**, *8*, 205. https://doi.org/10.3389/fgene.2017.00205.
47. Sharma, V.; Hecker, N.; Roscito, J.G.; Foerster, L.; Langer, B.E.; Hiller, M. A genomics approach reveals insights into the importance of gene losses for mammalian adaptations. *Nat. Commun.* **2018**, *9*, 1215. https://doi.org/10.1038/s41467-018-03667-1.
48. Zhou, X.; Sun, D.; Guang, X.; Ma, S.; Fang, X.; Mariotti, M.; Nielsen, R.; Gladyshev, V.N.; Yang, G. Molecular Footprints of Aquatic Adaptation Including Bone Mass Changes in Cetaceans. *Genome Biol. Evol.* **2018**, *10*, 967–975. https://doi.org/10.1093/gbe/evy062.
49. Zhang, X.; Chi, H.; Li, G.; Irwin, D.M.; Zhang, S.; Rossiter, S.J.; Liu, Y. Parallel Independent Losses of G-Type Lysozyme Genes in Hairless Aquatic Mammals. *Genome Biol. Evol.* 2021, 13, evab201. https://doi.org/10.1093/gbe/evab201.
50. Chikina, M.; Robinson, J.D.; Clark, N.L. Hundreds of Genes Experienced Convergent Shifts in Selective Pressure in Marine Mammals. *Mol. Biol. Evol.* **2016**, *33*, 2182–2192. https://doi.org/10.1093/molbev/msw112.
51. Silva, M.C.; Chibucos, M.; Munro, J.B.; Daugherty, S.; Coelho, M.M.; Silva, J.C. Signature of adaptive evolution in olfactory receptor genes in Cory's Shearwater supports molecular basis for smell in procellariiform seabirds. *Sci. Rep.* **2020**, *10*, 543. https://doi.org/10.1038/s41598-019-56950-6.
52. Sun, D.; Zhou, X.; Yu, Z.; Xu, S.; Seim, I.; Yang, G. Accelerated evolution and diversifying selection drove the adaptation of cetacean bone microstructure. *BMC Evol. Biol.* **2019**, *19*, 194. https://doi.org/10.1186/s12862-019-1509-x.
53. Sun, Y.; Liu, Y.; Sun, X.; Lin, Y.; Yin, D.; Xu, S.; Yang, G. Insights into body size variation in cetaceans from the evolution of body-size-related genes. *BMC Evol. Biol.* **2019**, *19*, 157. https://doi.org/10.1186/s12862-019-1461-9.
54. Li, J.; Shang, S.; Fang, N.; Zhu, Y.; Zhang, J.; Irwin, D.M.; Zhang, S.; Wang, Z. Accelerated Evolution of Limb-Related Gene Hoxd11 in the Common Ancestor of Cetaceans and Ruminants (Cetruminantia). *G3 (Bethesda Md.)* **2020**, *10*, 515–524. https://doi.org/10.1534/g3.119.400512.
55. Li, G.; Wei, H.; Bi, J.; Ding, X.; Li, L.; Xu, S.; Yang, G.; Ren, W. Insights into Dietary Switch in Cetaceans: Evidence from Molecular Evolution of Proteinases and Lipases. *J. Mol. Evol.* **2020**, *88*, 521–535. https://doi.org/10.1007/s00239-020-09952-2.

56. Mori, S.; Matsunami, M. Signature of positive selection in mitochondrial DNA in Cetartiodactyla. *Genes Genet. Syst.* **2018**, *93*, 65–73. https://doi.org/10.1266/ggs.17-00015.
57. Tian, R.; Geng, Y.; Yang, Y.; Seim, I.; Yang, G. Oxidative stress drives divergent evolution of the glutathione peroxidase (GPX) gene family in mammals. *Integr. Zool.* **2021**, *16*, 696–711. https://doi.org/10.1111/1749-4877.12521.
58. Wang, L.; Zhou, S.; Lyu, T.; Shi, L.; Dong, Y.; He, S.; Zhang, H. Comparative Genome Analysis Reveals the Genomic Basis of Semi-Aquatic Adaptation in American Mink (*Neovison vison*). *Animals* **2022**, *12*, 2385. https://doi.org/10.3390/ani12182385.
59. Qu, Y.; Chen, C.; Chen, X.; Hao, Y.; She, H.; Wang, M.; Ericson, P.; Lin, H.; Cai, T.; Song, G.; et al. The evolution of ancestral and species-specific adaptations in snowfinches at the Qinghai-Tibet Plateau. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2012398118. https://doi.org/10.1073/pnas.2012398118.
60. Tian, R.; Seim, I.; Ren, W.; Xu, S.; Yang, G. Contraction of the ROS Scavenging Enzyme Glutathione S-Transferase Gene Family in Cetaceans. *G3 (Bethesda Md.)* **2019**, *9*, 2303–2315. https://doi.org/10.1534/g3.119.400224.
61. Yang, L.; Wang, Y.; Sun, N.; Chen, J.; He, S. Genomic and functional evidence reveals convergent evolution in fishes on the Tibetan Plateau. *Mol. Ecol*. **2021**, *30*, 5752–5764. https://doi.org/10.1111/mec.16171.
62. Gao, X.; Wang, S.; Wang, Y.F.; Li, S.; Wu, S.W.; Yan, R.G.; Zhang, Y.W.; Wan, R.D.; He, Z.; Song, R.D. et al. Long read genome assemblies complemented by single cell RNA-sequencing reveal genetic and cellular mechanisms underlying the adaptive evolution of yak. *Nat. Commun*. **2022**, *13*, 4887. https://doi.org/10.1038/s41467-022-32164-9.
63. Zhu, K.; Ge, D.; Wen, Z.; Xia, L.; Yang, Q. Evolutionary Genetics of Hypoxia and Cold Tolerance in Mammals. *J. Mol. Evol.* **2018**, *86*, 618–634. https://doi.org/10.1007/s00239-018-9870-8.
64. Wu, D.D.; Yang, C.P.; Wang, M.S.; Dong, K.Z.; Yan, D.W.; Hao, Z.Q.; Fan, S.; Chu, S.; Shen, Q.; Jiang, L. et al. Convergent genomic signatures of high-altitude adaptation among domestic mammals. *Natl. Sci. Rev*. **2020**, *7*, 952–963. https://doi.org/10.1093/nsr/nwz213.
65. Ma, X.; Dai, W.; Kang, J.; Yang, L.; He, S. Comprehensive Transcriptome Analysis of Six Catfish Species from an Altitude Gradient Reveals Adaptive Evolution in Tibetan Fishes. *G3 (Bethesda Md.)* **2016**, *6*, 141–148. https://doi.org/10.1534/g3.115.024448.
66. Tian, R.; Losilla, M.; Lu, Y.; Yang, G.; Zakon, H. Molecular evolution of globin genes in Gymnotiform electric fishes: Relation to hypoxia tolerance. *BMC Evol. Biol*. **2017**, *17*, 51. https://doi.org/10.1186/s12862-017-0893-3.
67. Bartáková, V.; Bryjová, A.; Nicolas, V.; Lavrenchenko, L.A.; Bryja, J. Mitogenomics of the endemic Ethiopian rats: Looking for footprints of adaptive evolution in sky islands. *Mitochondrion* **2021**, *57*, 182–191. https://doi.org/10.1016/j.mito.2020.12.015.
68. Xu, Q.; Zhang, C.; Zhang, D.; Jiang, H.; Peng, S.; Liu, Y.; Zhao, K.; Wang, C.; Chen, L. Analysis of the erythropoietin of a Tibetan Plateau schizothoracine fish (*Gymnocypris dobula*) reveals enhanced cytoprotection function in hypoxic environments. *BMC Evol. Biol*. **2016**, *16*, 11. https://doi.org/10.1186/s12862-015-0581-0.
69. Wang, C.; Zhang, Q.; Liu, Y.; Xu, Q. Characterization of EPO H131S as a key mutation site in the hypoxia-adaptive evolution of *Gymnocypris dobula*. *Fish Physiol. Biochem*. **2022**, *48*, 723–733. https://doi.org/10.1007/s10695-022-01080-6.
70. Gibbons, T.C.; Metzger, D.; Healy, T.M.; Schulte, P.M. Gene expression plasticity in response to salinity acclimation in threespine stickleback ecotypes from different salinity habitats. *Mol. Ecol.* **2017**, *26*, 2711–2725. https://doi.org/10.1111/mec.14065.
71. Verta, J.P.; Jones, F.C. Predominance of cis-regulatory changes in parallel expression divergence of sticklebacks. *eLife* **2019**, *8*, e43785. https://doi.org/10.7554/elife.43785.
72. Kitano, J.; Ishikawa, A.; Kusakabe, M. Parallel transcriptome evolution in stream threespine sticklebacks. *Dev. Growth Differ.* **2019**, *61*, 104–113. https://doi.org/10.1111/dgd.12576.
73. Hart, J.C.; Ellis, N.A.; Eisen, M.B.; Miller, C.T. Convergent evolution of gene expression in two high-toothed stickleback populations. *PLoS Genet*. **2018**, *14*, e1007443. https://doi.org/10.1371/journal.pgen.1007443.
74. Roberts Kingman, G.A.; Vyas, D.N.; Jones, F.C.; Brady, S.D.; Chen, H.I.; Reid, K.; Milhaven, M.; Bertino, T.S.; Aguirre, W.E.; Heins, D.C.; et al. Predicting future from past: The genomic basis of recurrent and rapid stickleback evolution. *Sci. Adv*. **2021**, *7*, eabg5285. https://doi.org/10.1126/sciadv.abg5285.
75. Deng, Z.; Hui, L.; He, C.; Shou, C.; Han, Z. Heat shock protein 70 (Hsp70) and heat shock transcription factor (Hsf) gene families in *Cynoglossus semilaevis*: Genome-wide identification and correlation analysis in response to low salinity stress. *Mar. Freshw. Resh*. **2021**, *72*, 1132–1141. https://doi.org/10.1071/MF20326.
76. Barth, J.; Berg, P.R.; Jonsson, P.R.; Bonanomi, S.; Corell, H.; Hemmer-Hansen, J.; Jakobsen, K.S.; Johannesson, K.; Jorde, P.E.; Knutsen, H.; et al. Genome architecture enables local adaptation of Atlantic cod despite high connectivity. *Mol. Ecol*. **2017**, *26*, 4452–4466. https://doi.org/10.1111/mec.14207.
77. Ahmad, S.F.; Singchat, W.; Panthum, T.; Srikulnath, K. Impact of Repetitive DNA Elements on Snake Genome Biology Evolution. *Cells* **2021**, *10*, 1707. https://doi.org/10.3390/cells10071707.
78. Indjeian, V.B.; Kingman, G.A.; Jones, F.C.; Guenther, C.A.; Grimwood, J.; Schmutz, J.; Myers, R.M.; Kingsley, D.M. Evolving New Skeletal Traits by cis-Regulatory Changes in Bone Morphogenetic Proteins. *Cell* **2016**, *164*, 45–56. https://doi.org/10.1016/j.cell.2015.12.007.
79. Peichel, C.L.; Marques, D.A. The genetic and molecular architecture of phenotypic diversity in sticklebacks. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2017**, *372*, 20150486. https://doi.org/10.1098/rstb.2015.0486.
80. Xie, P.; Yi, S.K.; Yao, H.; Chi, W.; Guo, Y.; Ma, X.F.; Wang, F.P. Comparative transcriptome analysis reveals potential evolutionary differences in adaptation of temperature and body shape among four Percidae species. *PLoS ONE* **2019**, *14*, e0215933. https://doi.org/10.1371/journal.pone.0215933.

81. Reid, K.; Bell, M.A.; Veeramah, K.R. Threespine Stickleback: A Model System For Evolutionary Genomics. *Annu. Rev. Genom. Hum. Genet*. **2021**, *22*, 357–383. https://doi.org/10.1146/annurev-genom-111720-081402.
82. Xie, K.T.; Wang, G.; Thompson, A.C.; Wucherpfennig, J.I.; Reimchen, T.E.; MacColl, A.D.C. DNA fragility in the parallel evolution of pelvic reduction in stickleback fish. *Science* **2018**, *363*, 81–84. https://doi.org/10.1126/science.aan1425.
83. Archambeault, S.L.; Bärtschi, L.R.; Merminod, A.D.; Peichel, C.L. Adaptation via pleiotropy and linkage: Association mapping reveals a complex genetic architecture within the stickleback Eda locus. *Evol Lett*. **2020**, *4*, 282–301. https://doi.org/10.1002/evl3.175.
84. Howes, T.R.; Summers, B.R.; Kingsley, D.M. Dorsal spine evolution in threespine sticklebacks via a splicing change in MSX2A. *BMC Biol*. **2017**, *15*, 115. https://doi.org/10.1186/s12915-017-0456-5.
85. Greenway, R.; Barts, N.; Henpita, C.; Brown, A.P.; Arias Rodriguez, L.; Rodríguez Peña, C.M.; Arndt, S.; Lau, G.Y.; Murphy, M.P.; Wu, L. et al. Convergent evolution of conserved mitochondrial pathways underlies repeated adaptation to extreme environments. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 16424–16430. https://doi.org/10.1073/pnas.2004223117.
86. Dong, C.; Duan, X.; Younis, L.M.; Zhang, M.; Ma, X.; Chen, B.; Li, X.; Xu, P. Mitogenomic Perspectives on the Adaptation to Extreme Alkaline Environment of Amur ide (*Leuciscus waleckii*). *Mar. Biotechnol.* **2020**, *22*, 220–232. https://doi.org/10.1007/s10126-020-09946-7.
87. Wang, S.; Kuang, Y.; Liang, L.; Sun, B.; Zhao, X.; Zhang, L.; Chang, Y. Resequencing and SNP discovery of Amur ide (*Leuciscus waleckii*) provides insights into local adaptations to extreme environments. *Sci. Rep*. **2021**, *11*, 5064. https://doi.org/10.1038/s41598-021-84652-5.
88. Xu, J.; Li, J.T.; Jiang, Y.; Peng, W.; Yao, Z.; Chen, B.; Jiang, L.; Feng, J.; Ji, P.; Liu, G.; et al. Genomic Basis of Adaptive Evolution: The Survival of Amur Ide (*Leuciscus waleckii*) in an Extremely Alkaline Environment. *Mol. Biol. Evol*. **2017**, *34*, 145–159. https://doi.org/10.1093/molbev/msw230.
89. Jia, J.; Qin, J.; Yuan, X.; Liao, Z.; Huang, J.; Wang, B.; Sun, C.; Li, W. Microarray and metabolome analysis of hepatic response to fasting and subsequent refeeding in zebrafish (*Danio rerio*). *BMC Genom.* **2019**, *20*, 919. https://doi.org/10.1186/s12864-019-6309-6.
90. Chebii, V.J.; Oyola, S.O.; Kotze, A.; Domelevo Entfellner, J.B.; Musembi Mutuku, J.; Agaba, M. Genome-Wide Analysis of Nubian Ibex Reveals Candidate Positively Selected Genes That Contribute to Its Adaptation to the Desert Environment. *Animals* **2020**, *10*, 2181. https://doi.org/10.3390/ani10112181.
91. Elayadeth-Meethal, M.; Thazhathu Veettil, A.; Asaf, M.; Pramod, S.; Maloney, S.K.; Martin, G.B.; Rivero, M.J.; Sejian, V.; Naseef, P.P.; Kuruniyan, M.S.; et al. Comparative Expression Profiling and Sequence Characterization of ATP1A1 Gene Associated with Heat Tolerance in Tropically Adapted Cattle. *Animals* **2021**, *11*, 2368. https://doi.org/10.3390/ani11082368.
92. Lighten, J.; Incarnato, D.; Ward, B.J.; van Oosterhout, C.; Bradbury, I.; Hanson, M.; Bentzen, P. Adaptive phenotypic response to climate enabled by epigenetics in a K-strategy species, the fish *Leucoraja ocellata* (Rajidae). *R. Soc. Open Sci*. **2016**, *3*, 160299. https://doi.org/10.1098/rsos.160299.
93. Zhuang, X.; Yang, C.; Murphy, K.R.; Cheng, C.C. Molecular mechanism and history of non-sense to sense evolution of antifreeze glycoprotein gene in northern gadids. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 4400–4405. https://doi.org/10.1073/pnas.1817138116.
94. Irwin, D.M. Duplication and diversification of insulin genes in ray-finned fish. *Zool. Res*. **2019**, *40*, 185–197. https://doi.org/10.24272/j.issn.2095-8137.2018.052.
95. Yamagishi, G.; Yoshida, A.; Kobayashi, A.; Park, M.K. Molecular characterization of insulin from squamate reptiles reveals sequence diversity and possible adaptive evolution. *Gen. Comp. Endocrinol.* **2016**, *225*, 197–211. https://doi.org/10.1016/j.ygcen.2015.08.021.
96. Chen, Y.H.; Zhao, H. Evolution of digestive enzymes and dietary diversification in birds. *PeerJ* **2019**, *7*, e6840. https://doi.org/10.7717/peerj.6840.
97. Janiak, M.C.; Burrell, A.S.; Orkin, J.D.; Disotell, T.R. Duplication and parallel evolution of the pancreatic ribonuclease gene (*RNASE1*) in folivorous non-colobine primates, the howler monkeys (*Alouatta* spp.). *Sci. Rep*. **2019**, *9*, 20366. https://doi.org/10.1038/s41598-019-56941-7.
98. Song, J.; Zheng, H.; Xue, J.; Liu, J.; Sun, Q.; Yang, W. GPR15-C10ORF99 functional pairing initiates colonic Treg homing in amniotes. *EMBO Rep*. **2022**, *23*, e53246. https://doi.org/10.15252/embr.202153246.
99. Endo, Y.; Kamei, K.I.; Inoue-Murayama, M. Genetic signatures of lipid metabolism evolution in Cetacea since the divergence from terrestrial ancestor. *J. Evol. Biol*. **2018**, *31*, 1655–1665. https://doi.org/10.1111/jeb. 13361.
100. Gao, X.; Li, Y.; Adetula, A.A.; Wu, Y.; Chen, H. Analysis of new retrogenes provides insight into dog adaptive evolution. *Ecol. Evol*. **2019**, *9*, 11185–11197. https://doi.org/10.1002/ece3.5620.
101. Shan, L.; Wu, Q.; Wang, L.; Zhang, L.; Wei, F. Lineage-specific evolution of bitter taste receptor genes in the giant and red pandas implies dietary adaptation. *Integr. Zool*. **2018**, *13*, 152–159. https://doi.org/10.1111/1749-4877.12291.
102. Li, L.; Shen, F.; Jie, X.; Zhang, L.; Yan, G.; Wu, H.; Huang, Y.; Hou, R.; Yue, B.; Zhang, X. Comparative Transcriptomics and Methylomics Reveal Adaptive Responses of Digestive and Metabolic Genes to Dietary Shift in Giant and Red Pandas. *Genes* **2022**, *13*, 1446. https://doi.org/10.3390/genes13081446.
103. Rinker, D.C.; Specian, N.K.; Zhao, S.; Gibbons, J.G. Polar bear evolution is marked by rapid changes in gene copy number in response to dietary shift. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 13446–13451. https://doi.org/10.1073/pnas.1901093116.

104. Botero-Castro, F.; Tilak, M.K.; Justy, F.; Catzeflis, F.; Delsuc, F.; Douzery, E. In Cold Blood: Compositional Bias and Positive Selection Drive the High Evolutionary Rate of Vampire Bats Mitochondrial Genomes. *Genome Biol. Evol.* **2018**, *10*, 2218–2239. https://doi.org/10.1093/gbe/evy120.
105. Wang, B.J.; Xia, J.M.; Wang, Q.; Yu, J.L.; Song, Z.; Zhao, H. Diet and Adaptive Evolution of Alanine-Glyoxylate Aminotransferase Mitochondrial Targeting in Birds. *Mol. Biol. Evol.* **2020**, *37*, 786–798. https://doi.org/10.1093/molbev/msz266.
106. McGlothlin, J.W.; Kobiela, M.E.; Feldman, C.R.; Castoe, T.A.; Geffeney, S.L.; Hanifin, C.T.; Toledo, G.; Vonk, F.J.; Richardson, M.K.; Brodie, E.D.; et al. Historical Contingency in a Multigene Family Facilitates Adaptive Evolution of Toxin Resistance. *Curr. Biol.* **2016**, *26*, 1616–1621. https://doi.org/10.1016/j.cub.2016.04.056.
107. Feldman, C.R.; Durso, A.M.; Hanifin, C.T.; Pfrender, M.E.; Ducey, P.K.; Stokes, A.N.; Barnett, K.E.; Brodie, E.D. III; Brodie, E.D., Jr. Is there more than one way to skin a newt? Convergent toxin resistance in snakes is not due to a common genetic mechanism. *Heredity* **2017**, *119*, 468. https://doi.org/10.1038/hdy.2015.73.
108. Gendreau, K.L.; Hornsby, A.D.; Hague, M.T.J.; McGlothlin, J.W. Gene Conversion Facilitates the Adaptive Evolution of Self-Resistance in Highly Toxic Newts. *Mol. Biol. Evol.* **2021**, *38*, 4077–4094. https://doi.org/10.1093/molbev/msab182.
109. Wang, L.; Sun, L.; Wan, Q.H.; Fang, S.G. Comparative Genomics Provides Insights into Adaptive Evolution in Tactile-Foraging Birds. *Genes* **2022**, *13*, 678. https://doi.org/10.3390/genes13040678.
110. Cleves, P.A.; Hart, J.C.; Agoglia, R.M.; Jimenez, M.T.; Erickson, P.A.; Gai, L.; Miller, C.T. An intronic enhancer of Bmp6 underlies evolved tooth gain in sticklebacks. *PLoS Genet.* **2018**, *14*, e1007449. https://doi.org/10.1371/journal.pgen.1007449.
111. Erickson, P.A.; Baek, J.; Hart, J.C.; Cleves, P.A.; Miller, C.T. Genetic Dissection of a Supergene Implicates Tfap2a in Craniofacial Evolution of Threespine Sticklebacks. *Genetics* **2018**, *209*, 591–605. https://doi.org/10.1534/genetics.118.300760.
112. Karagic, N.; Schneider, R.F.; Meyer, A.; Hulsey, C.D. A Genomic Cluster Containing Novel and Conserved Genes is Associated with Cichlid Fish Dental Developmental Convergence. *Mol. Biol. Evol.* **2020**, *37*, 3165–3174. https://doi.org/10.1093/molbev/msaa153.
113. Machado, J.P.; Philip, S.; Maldonado, E.; O'Brien, S.J.; Johnson, W.E.; Antunes, A. Positive Selection Linked with Generation of Novel Mammalian Dentition Patterns. *Genome Biol. Evol.* **2016**, *8*, 2748–2759. https://doi.org/10.1093/gbe/evw200.
114. Mu, Y.; Tian, R.; Xiao, L.; Sun, D.; Zhang, Z.; Xu, S.; Yang, G. Molecular Evolution of Tooth-Related Genes Provides New Insights into Dietary Adaptations of Mammals. *J. Mol. Evol.* **2021**, *89*, 458–471. https://doi.org/10.1007/s00239-021-10017-1.
115. Conith, A.J.; Albertson, R.C. The cichlid oral and pharyngeal jaws are evolutionarily and genetically coupled. *Nat. Commun.* **2021**, *12*, 5477. https://doi.org/10.1038/s41467-021-25755-5.
116. He, S.; Li, L.; Lv, L.Y.; Cai, W.J.; Dou, Y.Q.; Li, J.; Tang, S.L..; Chen, X.; Zhang, Z.; Xu, J.; et al. Mandarin fish (Sinipercidae) genomes provide insights into innate predatory feeding. *Commun. Biol.* **2020**, *3*, 361. https://doi.org/10.1038/s42003-020-1094-y.
117. O'Gorman, M.; Thakur, S.; Imrie, G.; Moran, R.L.; Choy, S.; Sifuentes-Romero, I.; Bilandžija H.; Renner, K.J.; Duboué, E.; Rohner, N. et al. Pleiotropic function of the oca2 gene underlies the evolution of sleep loss and albinism in cavefish. *Curr. Biol.* **2021**, *31*, 3694–3701. https://doi.org/10.1016/j.cub.2021.06.077.
118. Xu, S.; Sun, X.; Niu, X.; Zhang, Z.; Tian, R.; Ren, W.; Zhou, K.; Yang, G. Genetic basis of brain size evolution in cetaceans: Insights from adaptive evolution of seven primary microcephaly (MCPH) genes. *BMC Evol. Biol.* **2017**, *17*, 206. https://doi.org/10.1186/s12862-017-1051-7.
119. Huang, Y.; Chain, F.J.; Panchal, M.; Eizaguirre, C.; Kalbe, M.; Lenz, T.L.; Samonte, I.E.; Stoll, M.; Bornberg-Bauer, E.; Reusch, T.B. et al. Transcriptome profiling of immne tissues reveals habitat-specific gene expression between lake and river sticklebacks. *Mol. Ecol.* **2016**, *25*, 943–958. https://doi.org/10.1111/mec.13520.
120. Ahmad, H.I.; Zhou, J.; Ahmad, M.J.; Afzal, G.; Jiang, H.; Zhang, X.; Elokil, A.A.; Khan, M.A.; Li, L.; Li, H.; et al. Adaptive selection in the evolution of programmed cell death-1 and its ligands in vertebrates. *Aging* **2020**, *12*, 3516–3557. https://doi.org/10.18632/aging.102827.
121. Xu, D.; Pavlidis, P.; Thamadilok, S.; Redwood, E.; Fox, S.; Blekhman, R. Recent evolution of the salivary mucin MUC7. *Sci. Rep.* **2016**, *6*, 31791.
122. Li, L.; Rong, X.; Li, G.; Wang, Y.; Chen, B.; Ren, W.; Yang, G.; Xu, S. Genomic organization and adaptive evolution of IGHC genes in marine mammals. *Mol. Immunol.* **2018**, *99*, 75–81. https://doi.org/10.1016/j.molimm.2018.04.011.
123. Chen, J.; Shang, S.; Wu, X.; Zhong, H.; Zhao, C.; Wei, Q.; Zhang, H.; Xia, T.; Chen, Y.; Zhang, H.; et al. Genomic analysis and adaptive evolution of the RIG-I-like and NOD-like receptors in reptiles. *Int. J. Biol. Macromol.* **2019**, *134*, 1045–1051. https://doi.org/10.1016/j.ijbiomac.2019.05.172.
124. He, C.; Wei, Y.; Zhu, Y.; Xia, Y.; Irwin, D.M.; Liu, Y. Adaptive Evolution of C-Type Lysozyme in Vampire Bats. *J. Mol. Evol.* **2019**, *87*, 309–316. https://doi.org/10.1007/s00239-019-09910-7.
125. Tian, R.; Seim, I.; Zhang, Z.; Yang, Y.; Ren, W.; Xu, S.; Yang, G. Distinct evolution of toll-like receptor signaling pathway genes in cetaceans. *Genes Genom.* **2019**, *41*, 1417–1430. https://doi.org/10.1007/s13258-019-00861-3.
126. Palomar, G.; Dudek, K.; Wielstra, B.; Jockusch, E.L.; Vinkler, M.; Arntzen, J.W.; Ficetola, G.F.; Matsunami, M.; Waldman, B.; Těšický, M.; et al. Molecular Evolution of Antigen-Processing Genes in Salamanders: Do They Coevolve with MHC Class I Genes? *Genome Biol. Evol.* **2021**, *13*, 259. https://doi.org/10.1093/gbe/evaa259.
127. Ametrano, A.; Picchietti, S.; Guerra, L.; Giacomelli, S.; Oreste, U.; Coscia, M.R. Comparative Analysis of the pIgR Gene from the Antarctic Teleost *Trematomus bernacchii* Reveals Distinctive Features of Cold-Adapted Notothenioidei. *Int. J. Mol. Sci.* **2022**, *23*, 7783. https://doi.org/10.3390/ijms23147783.

128. Yang, Y.; Wang, T.; Chen, J.; Wu, L.; Wu, X.; Zhang, W.; Luo, J.; Xia, J.; Meng, Z.; Liu, X. Whole-genome sequencing of brown-marbled grouper (Epinephelus fuscoguttatus) provides insights into adaptive evolution and growth differences. *Mol. Ecol. Resour*. **2022**, *22*, 711–723. https://doi.org/10.1111/1755-0998.13494.
129. Wang, S.; Joka, F.R.; Wang, X.; Bai, B. Analysis of the Phylogeny and Evolutionary Selection Pressure of the Mx Gene in 10 Wild Birds. *Pak. J. Zool*. **2019**, *51*, 1299–1307. https://doi.org/10.17582/journal.pjz/2019.51.4.1299.1307.
130. Yamada, E.; Yoshikawa, R.; Nakano, Y.; Misawa, N.; Kobayashi, T.; Ren, F.; Izumi, T.; Miyazawa, T.; Koyanagi, Y.; Sato, K. A naturally occurring bovine APOBEC3 confers resistance to bovine lentiviruses: Implication for the co-evolution of bovids and their lentiviruses. *Sci. Rep*. **2016**, *6*, 33988. https://doi.org/10.1038/srep33988.
131. Yang, Y.; Wu, L.N.; Chen, J.F.; Wu, X.; Xia, J.H.; Meng, Z.N.; Liu, X.C.; Lin, H.R. Whole-genome sequencing of leopard coral grouper (*Plectropomus leopardus*) and exploration of regulation mechanism of skin color and adaptive evolution. *Zool. Res*. **2020**, *41*, 328–340. https://doi.org/10.24272/j.issn.2095-8137.2020.038.
132. Jaiswal, S.K.; Gupta, A.; Saxena, R.; Prasoodanan, V.P.K.; Sharma, A.K.; Mittal, P.; Roy, A.; Shafer, A.; Vijay, N.; Sharma, V.K. Genome Sequence of Peacock Reveals the Peculiar Case of a Glittering Bird. *Front. Genet.* **2018**, *9*, 392. https://doi.org/10.3389%2Ffgene.2018.00392.
133. Solbakken, M.H.; Voje, K.L.; Jakobsen, K.S.; Jentoft, S. Linking species habitat and past palaeoclimatic events to evolution of the teleost innate immune system. *Proc. Biol. Sci.* **2017**, *284*, 20162810. https://doi.org/10.1098/rspb.2016.2810.
134. Bagheri, M.; Zahmatkesh, A. Evolution and species-specific conservation of toll-like receptors in terrestrial vertebrates. *Int. Rev. Immunol*. **2018**, *37*, 217–228. https://doi.org/10.1080/08830185.2018.1506780.
135. Shang, S.; Zhong, H.; Wu, X.; Wei, Q.; Zhang, H.; Chen, J.; Chen, Y.; Tang, X.; Zhang, H. Genomic evidence of gene duplication and adaptive evolution of Toll like receptors (TLR2 and TLR4) in reptiles. *Int. J. Biol. Macromol*. **2018**, *109*, 698–703. https://doi.org/10.1016/j.ijbiomac.2017.12.123.
136. Tian, R.; Chen, M.; Chai, S.; Rong, X.; Chen, B.; Ren, W.; Xu, S.; Yang, G. Divergent Selection of Pattern Recognition Receptors in Mammals with Different Ecological Characteristics. *J. Mol. Evol.* **2018**, *86*, 138–149. https://doi.org/10.1007/s00239-018-9832-1.
137. Neves, F.; Águeda-Pinto, A.; Pinheiro, A.; Abrantes, J.; Esteves, P.J. Strong selection of the TLR2 coding region among the Lagomorpha suggests an evolutionary history that differs from other mammals. *Immunogenetics* **2019**, *71*, 437–443. https://doi.org/10.1007/s00251-019-01110-3.
138. Voogdt, C.G.; Bouwman, L.I.; Kik, M.J.; Wagenaar, J.A.; van Putten, J.P. Reptile Toll-like receptor 5 unveils adaptive evolution of bacterial flagellin recognition. *Sci. Rep*. **2016**, *6*, 19046. https://doi.org/10.1038/srep19046.
139. Liu, G.; Zhang, H.; Sun, G.; Zhao, C.; Shang, S.; Gao, X.; Xia, T.; Yang, X. Characterization of the peripheral blood transcriptome and adaptive evolution of the MHC I and TLR gene families in the wolf (*Canis lupus*). *BMC Genom.* **2017**, *18*, 584. https://doi.org/10.1186/s12864-017-3983-0.
140. Velová, H.; Gutowska-Ding, M.W.; Burt, D.W.; Vinkler, M. Toll-Like Receptor Evolution in Birds: Gene Duplication, Pseudogenization, and Diversifying Selection. *Mol. Biol. Evol.* **2018**, *35*, 2170–2184. https://doi.org/10.1093/molbev/msy119.
141. Xu, S.; Tian, R.; Lin, Y.; Yu, Z.; Zhang, Z.; Niu, X.; Wang, X.; Yan, G. Widespread positive selection on cetacean TLR extracellular domain. *Mol. Immunol*. **2019**, *106*, 135–142.
142. Schad, J.; Voigt, C.C. Adaptive evolution of virus-sensing toll-like receptor 8 in bats. *Immunogenetics* **2016**, *68*, 783–795. https://doi.org/10.1016/j.molimm.2018.12.022.
143. Wang, X.; Liu, S.; Yang, Y.; Fu, Q.; Abebe, A.; Liu, Z. Identification of NF-κB related genes in channel catfish and their expression profiles in mucosal tissues after columnaris bacterial infection. *Dev. Comp. Immunol*. **2017**, *70*, 27–38. https://doi.org/10.1016/j.dci.2017.01.003.
144. Wu, T.; Cheng, Y.; Liu, Z.; Tao, W.; Zheng, S.; Wang, D. Bioinformatic analyses of zona pellucida genes in vertebrates and their expression in Nile tilapia. *Fish Physiol. Biochem*. **2018**, *44*, 435–449. https://doi.org/10.1007/s10695-017-0434-4.
145. Killingbeck, E.E.; Wilburn, D.B.; Merrihew, G.E.; MacCoss, M.J.; Swanson, W.J. Proteomics support the threespine stickleback egg coat as a protective oocyte envelope. *Mol. Reprod. Dev*. **2021**, *88*, 500–515. https://doi.org/10.1002/mrd.23517.
146. Vicens, A.; Treviño, C.L. Positive Selection in the Evolution of Mammalian CRISPs. *J. Mol. Evol*. **2018**, *86*, 635–645. https://doi.org/10.1007/s00239-018-9872-6.
147. Isakov, N. Histocompatibility and Reproduction: Lessons from the Anglerfish. *Life* **2022**, *12*, 113. https://doi.org/10.3390/life12010113.
148. Chen, Z.; Liu, Y.; Liang, R.; Cui, C.; Zhu, Y.; Zhang, F.; Zhang, J.; Chen, X. Comparative transcriptome analysis provides insights into the molecular mechanisms of high-frequency hearing differences between the sexes of *Odorrana tormota*. *BMC Genom*. **2022**, *23*, 296. https://doi.org/10.1186/s12864-022-08536-2.
149. Zhang, C.; Tong, C.; Ludwig, A.; Tang, Y.; Liu, S.; Zhang, R.; Feng, C.; Li, G.; Peng, Z.; Zhao, K. Adaptive Evolution of the Eda Gene and Scales Loss in Schizothoracine Fishes in Response to Uplift of the Tibetan Plateau. *Int. J. Mol. Sci.* **2018**, *19*, 2953. https://doi.org/10.3390/ijms19102953.
150. Li, B.; Wang, H.; Li, A.; An, C.; Zhu, L.; Liu, S.; Zhuang, Z. The Landscape of DNA Methylation Generates Insight Into Epigenetic Regulation of Differences Between Slow-Twitch and Fast-Twitch Muscles in *Pseudocaranx dentex*. *Front. Mar. Sci*. **2022**, *9*, 916373. https://doi.org/10.3389/fmars.2022.916373.
151. Lu, L.; Li, Y.; Liu, Z.; Liang, F.; Guo, F.; Yang, S.; Wang, D.; He, Y.; Xiong, J.; Li, D. Functional constraints on adaptive evolution of protein ubiquitination sites. *Sci. Rep*. **2017**, *7*, 39949. https://doi.org/10.1038/srep39949.

152. Song, H.; Xing, C.; Lu, W.; Liu, Z.; Wang, X.; Cheng, J.; Zhang, Q. Rapid evolution of piRNA pathway and its transposon targets in Japanese flounder (*Paralichthys olivaceus*). *Comp. Biochem. Physiol. Part D Genom. Proteom.* **2019**, *31*, 100609. https://doi.org/10.1016/j.cbd.2019.100609.

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.

Chapter 2 | Comparative analysis of population structure and diversifying selection patterns between Eurasian grey wolves (*Canis lupus*) and free-ranging dogs (*Canis lupus familiaris*)

Contents

Abstract **52**

1. Introduction **53**

2. Materials and Methods **54**

2.1. DNA extractions and genotyping 54

2.2. Data filtering and quality control 54

2.3 Inference of population structure 56

2.4. Spatial Autocorrelation and Spatial Principal Component Analysis (sPCA) 56

2.5. Identification of SNPs under diversifying selection 57

2.6. Identification of genes under selection and gene ontology enrichment analysis 57

3. Results **58**

3.1. Population structure 58

3.2. Spatial autocorrelation and sPCA 59

3.3. Candidate SNPs under diversifying selection 59

3.4. Genes with signatures of diversifying selection 59

4. Discussion **64**

4.1. Regional variation among FRDs and grey wolves 64

4.2. Diversifying selection in FRDs and grey wolves 67

4.2.1. Genes with signatures of diversifying selection 68

4.2.2. Genes with missense mutations 68

4.2.3. Genes with missense mutations predicted to affect function 69

5. Conclusion **70**

References **71**

Abstract

Dogs are human commensals that diverged from their wolf ancestors due to domestication. Most contemporary domestic dogs are free-ranging (FRDs), thus their mate choice and habitat choice are not restricted by humans. FRDs and grey wolves have sympatric distributions across the wolf geographic range but occupy different niches. There is limited knowledge regarding adaptive strategies of these closely related wild and domestic canids in the areas where their distributions overlap. We genotyped genome-wide single-nucleotide-polymorphisms (SNPs) for 320 FRDs and 270 wolves from across Eurasia to investigate genetic structure and diversifying selection in both canids. We found that in both canids peninsular populations were genetically distinct, but wolves had a more pronounced population structure. Greater genetic connectivity in FRDs compared to wolves may reflect the human-facilitated gene flow and fragmentation of wolf habitats. Additionally, in both canids we identified signatures of selection in genes enriched for phenotypic traits impacting morphology and sensory perception, particularly the olfactory system. Moreover, we identified variants under selection predicted to affect protein function in the pathogen recognition receptor NOD1 in FRDs and in the neurotransmitter inhibitor SLC6A9 in both taxa. Despite their differences in connectivity, ecological niches and evolutionary histories, FRDs and grey wolves have experienced similar evolutionary trajectories, with diversifying selection favouring many of the same genes.

Keywords: free-ranging dogs, grey wolves, diversifying selection, genetic differentiation, fragmentation, evolutionary trajectories

1. Introduction

Dogs are descendants of grey wolves and were the first domesticated animals (Larsen & Fueller, 2014). Genomic and morphological research suggests an origin ~40000-16000 years ago from Eurasian grey wolves (reviewed in Tancredi & Cardinali, 2023). Evolutionary divergence from wolves and the dispersal of dogs is intertwined and influenced by domestication, human dispersal and evolutionary history (Feuerborn et al., 2021; Perri et al., 2020). Thus, dogs are almost ubiquitous world-wide, with an estimated population of 1 billion (Belsare & Vanak, 2020; Bryce, 2021). Globally, most dogs live without human supervision or confinement, i.e., they are free-ranging dogs (FRDs) (Gompper, 2014). FRDs feed mainly on anthropogenic foods and occupy a niche of human commensal within anthropogenic landscapes. Progressive encroachment of anthropogenic landscapes leads to increased presence of FRDs in natural landscapes (Bryce, 2021). FRDs threaten natural ecosystems by predating wildlife or spreading diseases, particularly to wild canids (Bryce, 2021; Plumer, et al., 2018; Hughes & Macdonald, 2013; Wierzbowska et al., 2016).

Grey wolves are apex predators and have been shown to impact trophic cascades in natural landscapes, thus their evolutionary mechanisms and history have long been a topic of interest (Fleming et al., 2017). Although they maintain a wide distribution throughout the Northern hemisphere, their current distribution is reduced from a wider ancestral range (Ericson et al., 2019; vonHoldt et al., 2023; Wolf & Ripple, 2017). Grey wolves were eradicated in most Western European countries in the 1900s and nearly extirpated across North America (Gerber et al., 2024; Smith & Peterson, 2021). Since then, they have recolonized and been re-introduced throughout the northern hemisphere (Gerber et al., 2024; Ericson et al., 2019; Šnjegota et al., 2021; vonHoldt et al., 2024). Eurasian grey wolves' range is discontinuous with divided populations (Kaczensky et al., 2013), with various potential factors impacting their connectivity, including but not limited to habitat differentiation, habitat loss, and human-wildlife conflicts (Gerber et al., 2014; Plumer et al., 2018). Consistently, wolves' genetic structure is also shaped by habitat differentiation, even in the absence of physical barriers to gene flow (Pilot et al. 2006, Schweizer et al. 2016; Leonard et al. 2014). FRDs genetic structure, on the other hand, has been proposed to be mainly influenced by isolation by distance and anthropogenic factors (Perri et al., 2021; Shannon et al., 2015; Vychodilova et al., 2018).

Grey wolves are characterised by high adaptability (Hendricks et al., 2019), and genetic adaptations related to body size and immune system have been identified in both European and North American populations (Pilot et al., 2014; Schweizer et al. 2016). While there is a large body of research into morphological and behavioural adaptations of pure-bred dogs (reviewed in Ostrander et al., 2017; Wynne, 2021), less is known about FRDs adaptations under natural selection. Grey wolves and FRDs occupy different niches and consequently have different ecological impacts on natural landscapes. We used these canids as a model to investigate population stratification and adaptations among wild and domestic animals under comparable environmental pressures.

The aim of our study was to elucidate how natural selection is operating upon two closely related canid taxa with overlapping distributions, but different ecological niches

and evolutionary histories. For this purpose, we investigated the genetic differentiation and signatures of diversifying selection in populations of FRDs and grey wolves from comparable habitats across their overlapping distribution in Eurasia. We employed four population structure estimation approaches to discern genetic structure using high-coverage SNP genotyping data from Eurasian grey wolves and FRDs. Two different approaches were employed to detect loci under diversifying selection based on the identified genetic populations in each canid. In addition, we used variant effect predictors and gene ontology enrichment analysis to explore the putative phenotypic impact of the variants with signatures of diversifying selection.

2. Materials and Methods

2.1. DNA extractions and genotyping

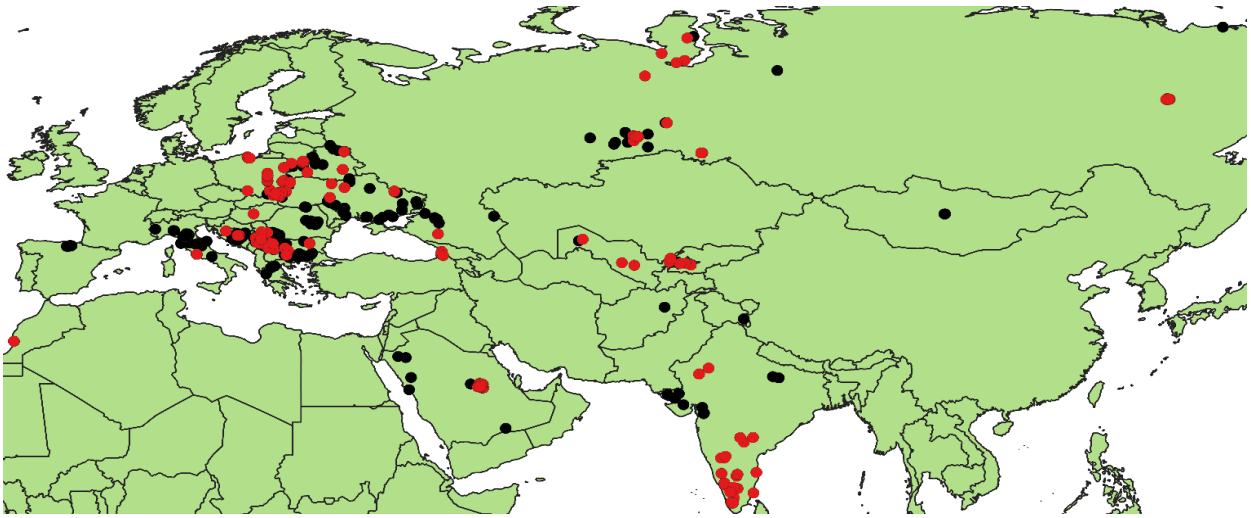
Tissue or saliva samples from 350 grey wolves and 714 FRDs were collected across Eurasia and Northern Africa (Fig. 1A) and stored at -20°C. We performed DNA extractions from muscle and skin samples using the NucleoSpin Genomic DNA Extraction kit, following manufacturer's instructions. We used the Performagene protocol to extract the DNA from saliva samples following the manufacturer's instructions. DNA Extracts with adequate DNA concentrations were genotyped with the Axiom Canine HD Array from Thermo Fischer Scientific (TFS). The Axiom Analysis Suite v5.1.1.1 (TFS) was used according to the Best Practices Genotyping Analysis Workflow for SNP genotype calling as described in Stefanovic et al., 2024. High-quality SNP genotyping data was obtained for 232 wolves and 370 FRDs (Table S1).

2.2. Data filtering and quality control

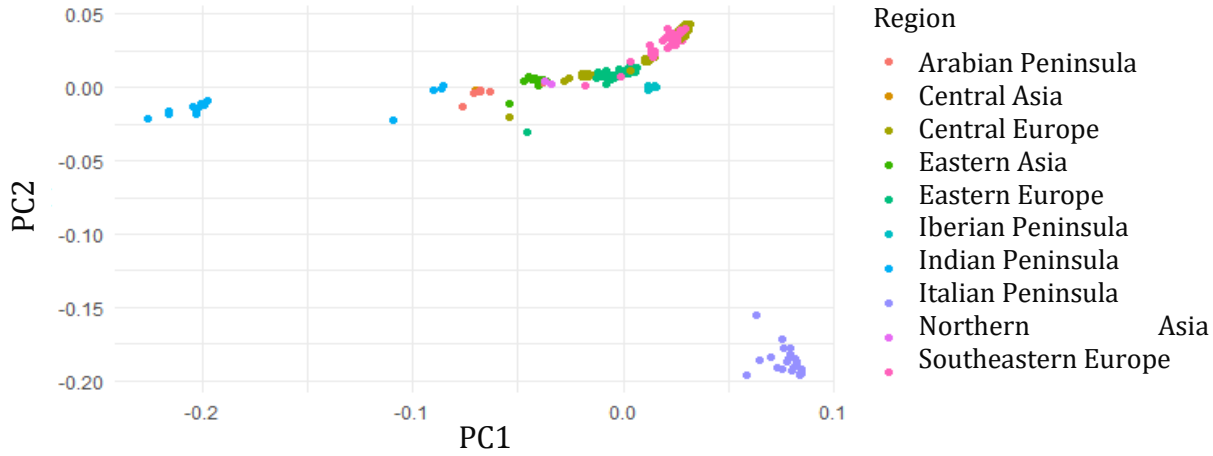
We performed a principal component analysis (PCA) with the pre-quality-controlled SNPs dataset (n=928 samples, Fig. S1) to address any potential taxonomic misidentification, using the Plink v2.0 software (Chang et al., 2015). Additionally, we performed quality controls on the dataset with Plink v1.90b6.21 software (Purcell et al., 2007): we pruned variants with minor allele frequency below 10%, 10% missingness per marker, markers with high degree of linkage disequilibrium (r2 >=0.1) and individuals with more than 20% missing data. To remove first-degree relatives, we calculated pairwise relatedness estimates with the KING v2.3.1 software (Manichaikul et al., 2010). We removed individuals based on 0.48 and 0.179 kinship thresholds, respectively. These thresholds were determined based on the identification of gaps along the kinship values across all individuals informed by the values recommended in the software manual (Manichaikul et al., 2010). To remove potential dog-wolf hybrids from the dataset we calculated admixture coefficients with ADMIXTURE v.1.3.0 (Alexaner et al., 2009) and removed F1 hybrids and F2 backcrosses based on a 20% admixture threshold. To avoid overrepresentation of individuals from specific regions, we also removed individuals from regions for which we had more than 50 individuals. The removed individuals were selected based on genotyping missingness, relatedness and randomly based on geographical coordinates. Our final quality-controlled dataset comprised 73360 SNP marker genotypes for 230 grey wolves and 370 FRDs (Fig. 1A, Table 1). We used this dataset for PCAs (i.e., one for grey wolves another for FRDs),

with Plink for outlier detection, using the top 5 principal components extracted from the variance-standardized relationship matrix to detect any additional outliers.

A. Distribution of sampled individuals



B. PCA of grey wolf populations



C. PCA of free-ranging dog populations

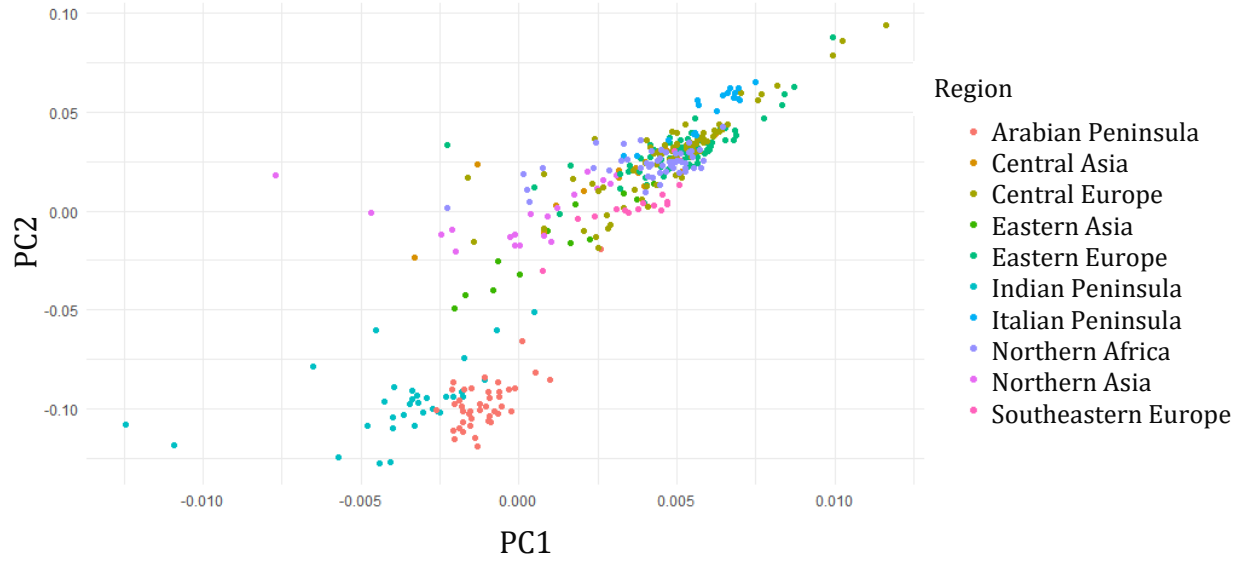


Figure 1. Map detailing distribution of locations from where samples were collected and principal component analysis (PCA) plots. A: Map detailing the geographical distribution from where grey wolves

and FRDs samples were collected. B: Principal component analysis (PCA) of quality-controlled Eurasian wolves, the eigen values for the first 5 principal components were: 10.9, 9.7, 5.7, 3.9 and 3.5. Four main clusters were detected Arabian Peninsula, Indian Peninsula, Eurasia and Italian Peninsula. C: PCA analysis of quality controlled-FRDs grouped by geographical regions, the eigen values for the first 5 principal components were: 16.1781, 7.81677, 4.37396, 3.37124 and 3.20545. Three main clusters were detected Arabian Peninsula, Indian Peninsula and Eurasia.

2.3 Inference of population structure

We employed four approaches to infer population structure in taxa: discriminant component analysis (DPCA) (Adegenet v2.1.10), maximum likelihood (ADMIXTURE), sparse nonnegative matrix factorization (sNMF) (LEA v3.14) and fast inference of spatial population structure (Tess3r v1.1.0). DAPC is a non-Bayesian method based on multivariate analysis where the genetic data is first transformed and then submitted to a PCA. Principal components of the PCA are then subjected to a linear discriminant (LD) analysis. We performed the DAPC analysis twice using the R package Adegenet (Jombart and Ahmed, 2011), first to establish the optimal number of principal components to retain from the initial 99 based on the Bayesian information criterion (BIC), and then with the optimal number of principal components retained (Fig. S3). We also measured genetic differentiation between clusters by calculating the proportion of genetic diversity partitioned among population clusters relative to the total genetic diversity (G_{ST}), which was estimated across the DAPC defined clusters, applying the Hedrick's G'ST method with the R package mmod v1.3.3 (White, 2012). ADMIXTURE was run with 10,000 permutations and 5 repetitions. The Tess3r R package (Caye et al, 2016) ancestry estimation function with the "projected.ls" method was used. Tess3r uses a non-negative matrix factorization algorithm, and it accounts for geographic data when estimating population structure. The LEA R package (Frichot and Francois, 2015) "snmf" function was used, which is based on sparse nonnegative matrix factorization (sNMF) approach to estimate ancestry coefficients. sNMF was run with 5 repetitions and up to 200 iterations with entropy=true. We ran these software from K1-K10 and selected optimal K estimations based on the lowest cross-validation or cross entropy error across the runs, respectively.

2.4. Spatial Autocorrelation and Spatial Principal Component Analysis (sPCA)

We performed spatial autocorrelation analysis and spatial principal component analysis (sPCA) to investigate the presence of spatial variance and the non-random distribution of genetic variation. The latter is a composite statistic which measures genetic diversity and spatial structure. We used the R packages sdep v1.3-3, dartR v2.9.7 to perform the spatial autocorrelation analysis (Monte-Carlo simulation of Moran I) and sPCA (Bivand, 2022; Mijanjos et al., 2022). The "moran.mc" function was used for the spatial autocorrelation with 999 permutations. The sPCA was done using manufacturers recommendations, 1000 random SNPs with 99 permutations and spatial autocorrelation values from the Moran I statistics. Eigen values in the sPCA measure both genetic variance and the spatial autocorrelation measured by Moran's I. We used decomposition of these values to select the interpreted principal component (PC).

2.5. Identification of SNPs under diversifying selection

We employed two approaches to identify the SNPs under diversifying selection using the R package PCAdapt v4.3.5 (Prive et al., 2020) and the Bayesacan v2.1 software (Foll & Gaggiotti, 2008), to mitigate the risk of misidentification. PCAdapt performs principal component analysis to calculate p-value and identify outliers based on the correlations between genetic variation and a defined number of principal components. We used a minor allele frequency threshold of 0.05 to select the candidate SNPs. Additionally, we used the Benjamini-Hochberg procedure to thin out false positives with an expected false discovery rate based on a 10% threshold. Bayescan's algorithm aims to identify candidate loci under selection based on differences in allele frequencies between populations. We defined the populations by individual assignments to snmf clusters described above. The Bayescan analysis was done using the default settings save for the prior odds neutral model, which was adjusted to 1000, to reduce false positives given the number of markers (i.e., >10000). Candidate SNPs under diversifying selection were identified based on 0.05 for q-value and alpha equal or higher than 0 thresholds. The list of SNPs under diversifying selection comprised outlier SNPs identified in both approaches.

2.6. Identification of genes under selection and gene ontology enrichment analysis

The ENSEMBL Variant Effect Predictor (VEP) r113 online tool (McLaren et al., 2016) was used to ascertain annotations for the genes for which there were SNP variants with signatures of diversifying selection and predict consequences of these variants. We lifted the SNPs under diversifying selection from the CanFam3 to CanFam6 genome using the Lift Genome Annotations software (Kent et al., 2002). Two SNPs were detected in the CanFam6 genome and were removed from subsequent analysis (Table S2). The positions of the SNPs under diversifying selection were updated and then extracted as the input file for the VEP analysis with Plink. We used the VEP tool with default settings save for the upstream/downstream distance from transcript which was increased to 100000 bp. Based on previous studies, this is recommended distance from the transcript to account for putative regulatory impacts (Axelsson et al., 2013; Pilot et al., 2016). In addition to providing ENSCAFG codes for the genes with SNP variants under selection it also included sorting intolerant from tolerant (SIFT) calculations. The SIFT calculations predicted the likelihood of SNP variants with signatures of diversifying selection to affect protein function.

We then used UnitProt r2024-03-06 mapping software (UnitProt Consortium, 2025) to convert the ENSCAFG codes to UnitProt identification codes. The software automatically excluded RNA genes. This produced the final list of genes under selection, which was the input for the gene ontology enrichment analysis. g:Profiler ve112_eg59_p19_25aa4782 (Kolberg et al., 2023) was used with default settings to infer gene ontology enrichment. Both the human and dog genomes options were used in our analysis and only genes with UnitProt gene identification codes were included; excluded genes are detailed in Table S3-S4. We used the setting which selected the ENSMBLE ID with most GO annotations and ignored ambiguous queries. Additionally, the software automatically excluded genes which either did not have GO annotations or were not in the corresponding annotated genome (Table S5). We also used the STRING v12 software

(Szklarczyk et al., 2023) to identify local clusters and protein associations between the products of the genes under diversifying selection. In this analysis we were able to use the ENSCAFG codes, therefore both protein and RNA genes were included in the analysis. It is important to mention that the genes without annotation or description of products were automatically ignored by the software. To further investigate the genes under diversifying selection and identify ontologies enriched within these sub-networks we performed the STRING network analysis for the 3 largest sub-networks independently.

3. Results

3.1. Population structure

Grey wolves and FRDs clustered separately in the first PCA, with FRDs showing less pronounced structure (Fig. S1). PCAs showed four and three clusters in wolves and FRDs respectively (Figs. 1B-C). More precise estimates for population structure were obtained with subsequent approaches. DAPC identified optimal K=4 for wolves and K=2 for FRDs, respectively, based on cross-entropy and cross-validation values (Figs. S2-S3). DAPC estimation of de novo groups with subtle population structure can differ from other approaches as it has been shown to be impacted by fixation index levels and sample sizes, the degree of which still needs to be further investigated (Miller et al., 2020). The highest genetic distance was between the wolves from the third (i.e., individuals from the Indian Peninsula) and fourth (i.e., individuals from the Italian peninsula) clusters (G_{ST} =0.07), while the lowest genetic distance was between the wolves from the first (i.e., individuals from Eurasia) and second (i.e., individuals from Southeastern Europe) clusters (G_{ST} = -0.001) (Fig. S4). The genetic distance between the first cluster and the third wolf cluster was (G_{ST} =0.02), between the fourth and the first (G_{ST} =0.04), between the second and the third (G_{ST} =0.05) and between the fourth and the second (G_{ST} =0.05) FRDs from the Indian and Arabian Peninsulas were assigned to one genetic cluster, and the remaining FRDs were assigned to the second cluster (Fig. S4). The genetic distance between these two clusters was G_{ST} =0.002.

The remaining approaches (i.e. sNMF, ADMIXTURE and Tess3r) consistently inferred K=8 as optimal for grey wolves and K=5 for FRDs, based on cross-validation, cross-entropy, and Bayesian information criterion (BIC), respectively (Fig. S2, Figs. 2-4). Most of the grey wolves from Southeastern Europe were assigned to the same two clusters, except for a few Romanian individuals which were assigned to one of the Eastern Europe/Asia clusters (Fig. 2, Fig. 4A-C). A few individuals from Belarus were assigned to their own cluster in contrast to the remaining Eurasian individuals, which were all assigned to varying degrees to the two Eurasian clusters. Both grey wolves and FRDs from the Italian and Indian Peninsulas clustered separately (Figs. 2-4). FRDs from the Arabian Peninsula also clustered separately in contrast to the grey wolves which only clustered separately with the Tess3r approach, while in the other two approaches they clustered with the other Eurasian individuals. The remaining FRDs were assigned to the same two clusters (Fig. 3, Fig. 4D-E). Clustering patterns remained consistent across Ks in each taxa. Specifically, K=5 in wolves showed similar patterns as seen for K=8, the main differences being the Iberian Peninsula individuals clustering with individuals

from Eastern Europe (Belarus) and ancestry proportions for the Southeastern Europe were assigned for 2-4 populations in contrast to the 3-8 at K=8 (Fig. S6). Conversely, FRDs showed the same clustering patterns at K=8 as at K=5, with the difference being the proportion of ancestry assignments (Fig. S7).

3.2. Spatial autocorrelation and sPCA

The spatial analysis and sPCA detected a spatial autocorrelation in both grey wolves and FRDs. In terms of variance and spatial autocorrelation λ1 was the largest eigenvalue, thus only the first global component was interpreted in both FRDs and grey wolves (Fig. S7). The first global component was structured spatially in both taxa and the geographic mapping of the sPCA eigen values showed a west-east cline in grey wolves and a north-south cline in FRDs (Figs. S8-S10). We found a significant positive autocorrelation between spatial distribution and genetic variance for both grey wolves (z-score=0.89542, p=0.001) and FRDs (z-score= 0.88547, p=0.001) (Fig. S9). The sPCA eigen values suggested that global patterns of spatial structure were more likely than local patterns in both taxa (Fig. 5). Monte Carlo tests confirmed global patterns in both FRDs and grey wolves (p=0.01) and no local structure (p=0.9) (Fig. S8).

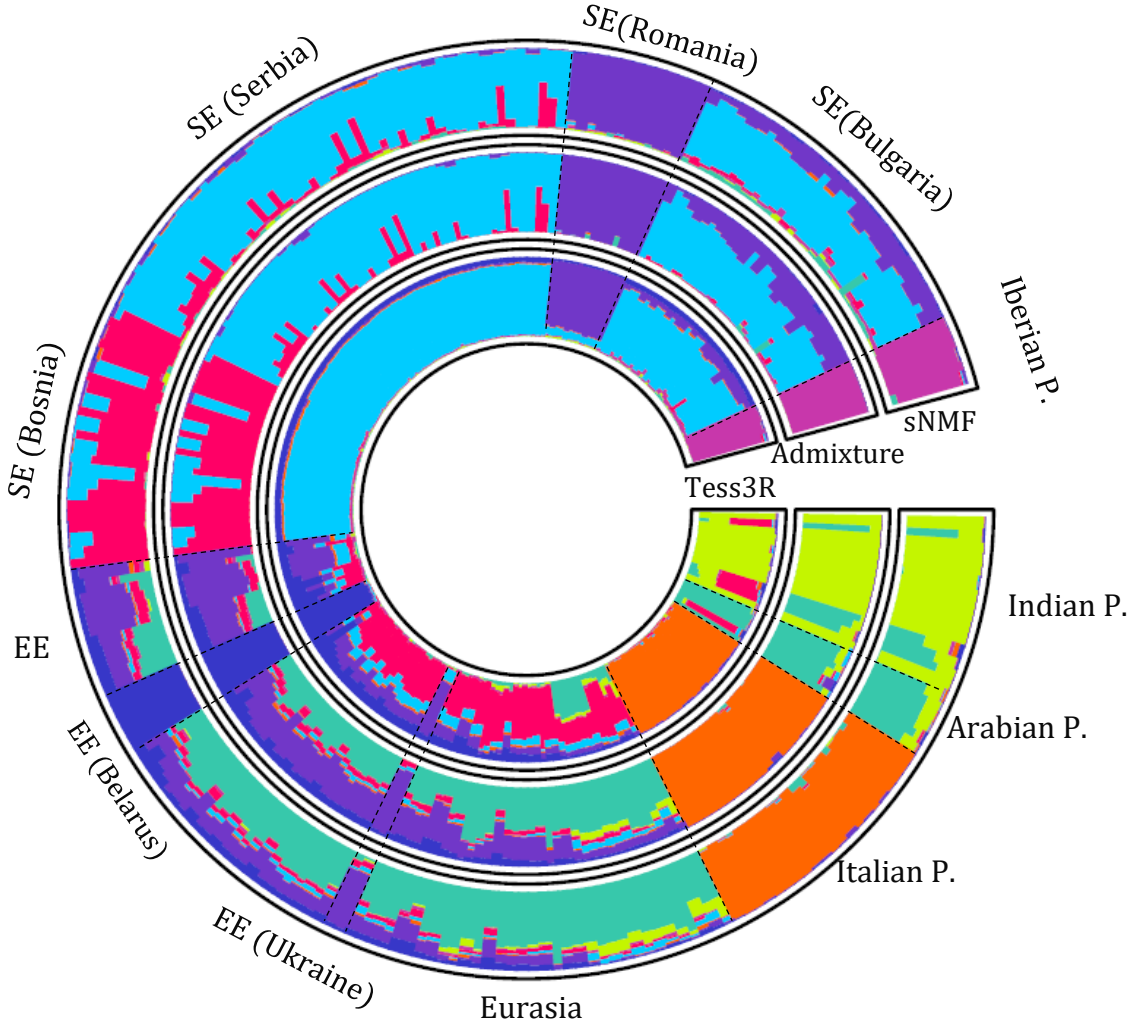
3.3. Candidate SNPs under diversifying selection

Two approaches were used to detect candidate SNPs under diversifying selection and candidate SNPs detected in both (i.e., overlapping SNPs) were utilized for the subsequent analysis. With PCAdapt, which utilized p-values computed with principal component analysis, 1060 candidate SNPs in grey wolves and 282 in FRDs were detected after the false discovery thinning (Fig. 6, Figs. S11-S12). Populations were defined based on individual assignments to the population structure clusters calculated with the sNMF. Thus, for this analysis there were 8 grey wolf populations and 5 FRD populations. Bayescan identified 188 candidate SNPs under diversifying selection in grey wolves and 185 in FRDs (Fig. S12). The final list of SNPs under diversifying selection comprised the candidates detected in both approaches, 116 SNPs for grey wolves and 106 for FRDs. These SNPs were used for subsequent analysis save for 1 SNP for FRDs and grey wolves, which failed the lifting (Fig. 6, Table S2). There were 62 SNPs common among both taxa with signatures of diversifying selection.

3.4. Genes with signatures of diversifying selection

471 genes were identified by the VEP tool to be under selection in grey wolves, including 121 RNA genes (i.e., LncRNAs, microRNA and snoRNAs), 8 pseudogenes and 1 riboenzyme gene (Table S3). The remaining 341 protein coding genes were involved in: gene regulation (i.e., transcription factors, RNA processing), cellular processes and structure (e.g., kinases, cellular structural proteins, transporters), stimulus response (i.e., immune system, olfactory receptors, visual system), metabolic processes, among others. The highest ranked enriched GO terms in the grey wolf genes under diversifying selection with both g:Profiler and STRING analysis were related to stimulus response (Fig. 7A, 7C, Figs. S13-S14, S16-S18), specifically olfactory receptor activity (Fig. 7A) and response to organic cyclic compounds (Fig. 7C). The STRING functional protein association network analysis predicted 55 protein interactions (Figs. S14, S16, S18).

A: Population structure histograms



B: DAPC: K-nearest neighbours network LD1

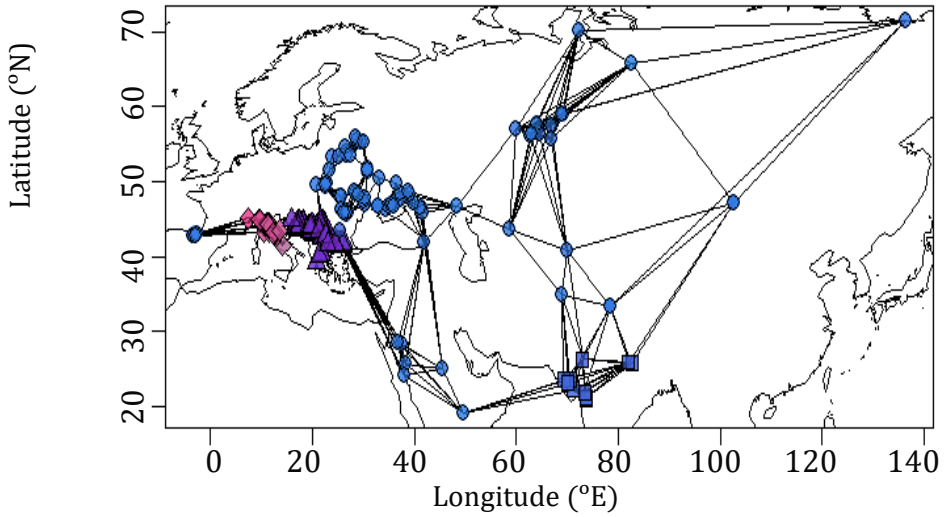


Figure 2. Ancestry assignment histograms for wolves based on Tess3r, LEA (sNMF) and ADMIXTURE assuming the number of clusters K=8. For Tess3r an optimal K could not be determined based on the cross-entropy values (Fig. S3), however with LEA and ADMIXTURE, K=8 was detected as optimal (Figs. S2-S30. Regions were abbreviated as follows: Eastern Europe (EE), Southeastern Europe (SE) and (P.) stands for Peninsula. B: DAPC K-nearest neighbours map considering 10 neighbours, with the first linear

discriminant. Wolves' individual cluster assignment DAPC (K=4), are denoted by shapes, wolves assigned to the first cluster have circle shape, second cluster triangle, third cluster square and fourth cluster diamond. Additionally, individuals' colours reflect coefficients for all the clusters (Fig. S5). Thus, individuals assigned coefficients of more than one cluster have intermediate colours to reflect that.

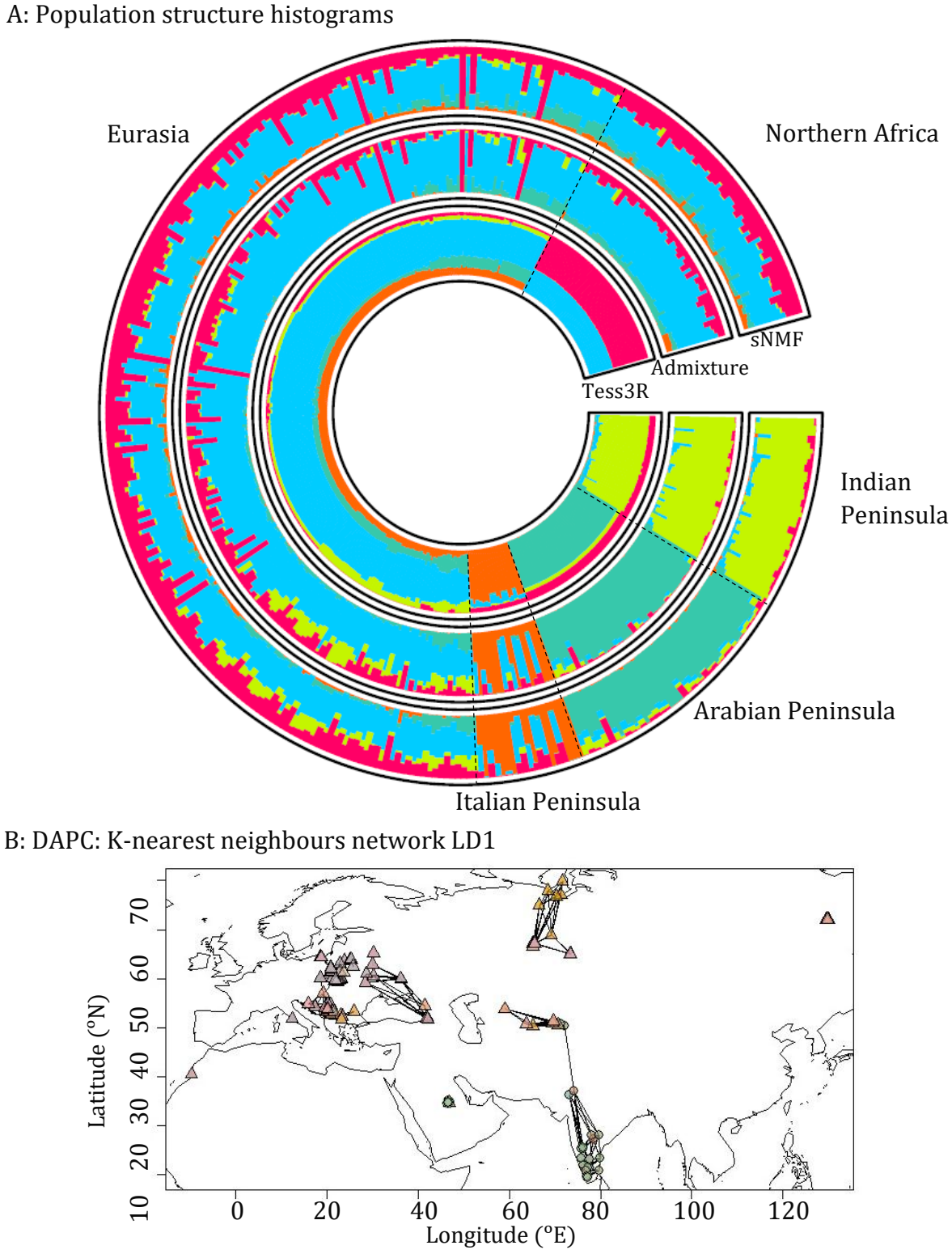


Figure 3. Ancestry assignments histograms for FRDs based on all three approaches and nearest neighbour network. A: ancestry histograms K=5 from Tess3r, LEA (sNMF) and ADMIXTURE. Comparison of histograms from all three different approaches. For Tess3r an optimal K could not be determined based on the cross-entropy values (Fig. S3), however with LEA and ADMIXTURE, K=5 was detected as optimal

(Figs. S2, S3). B: K-nearest neighbours map considering 10 neighbours, with the first linear discriminant. Each shape represents one of the two clusters in Fig. S5, in-between colours represent individuals with coefficients of more than one cluster.

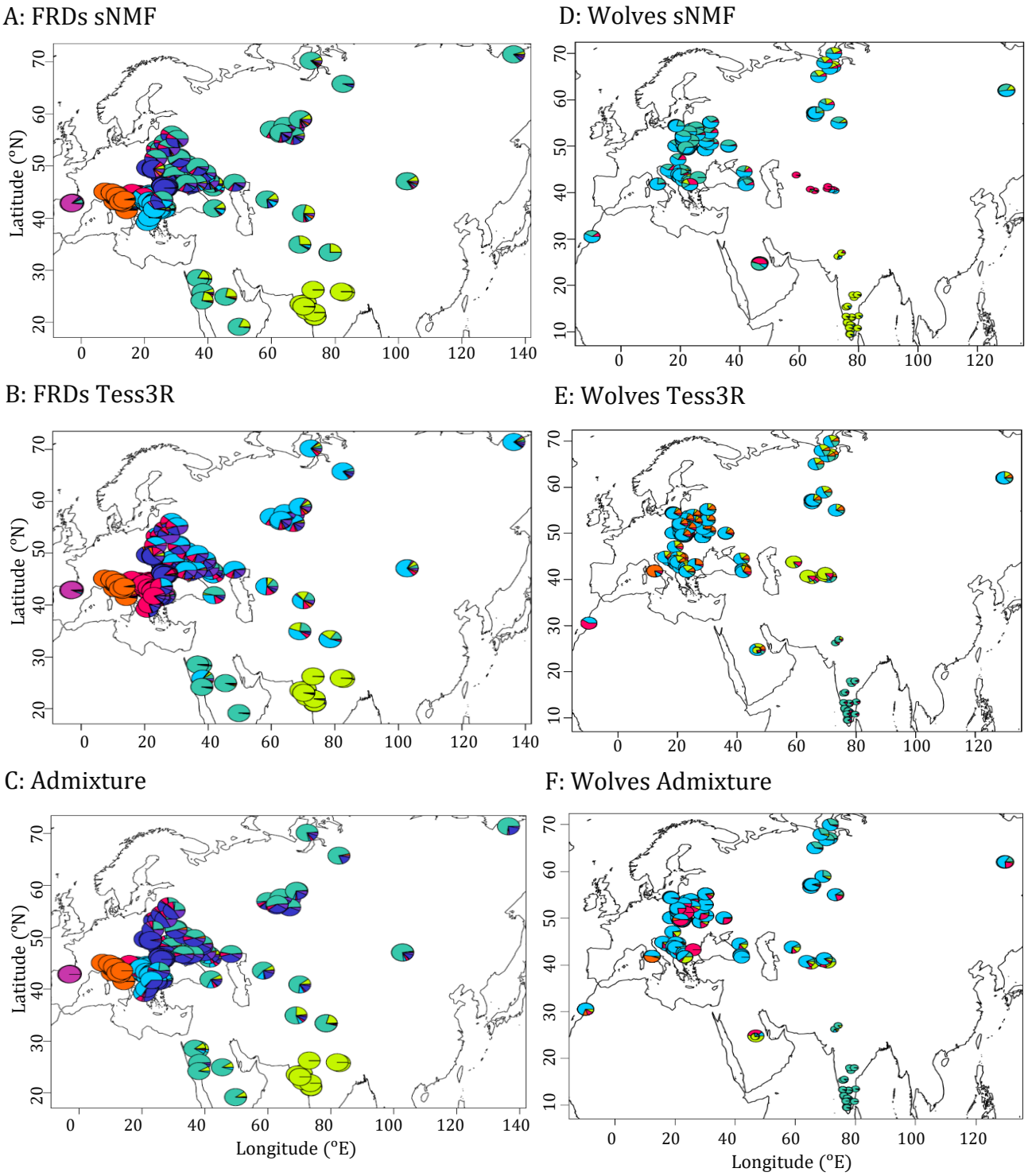


Figure 4. Geographic maps of genetic differentiation ancestry coefficients for wolves (A-C) and FRDs for sNMF (A, D), Tess3r (B, E) and Admixture (C, F). Using K-8 ancestral populations for wolves and K=5 ancestral populations for FRDs.

The VEP tool predicted 5 of the variants with signatures of selection to be missense mutations. The genes with these mutations coded for: a transcription repressor (INSM2), two olfactory receptors (LOC481603 and OR5M13), a pyruvate kinase (PKLR)

and a glycine transporter (SLC6A9 i.e., solute carrier family 6 member 9). The SLC6A9 variant allele A (Fig. S20) results in an amino acid change (G > S) which is predicted to occur in transporter domain of the protein. This mutation was predicted to be deleterious by SIFT score (0.01) (Fig. S13).

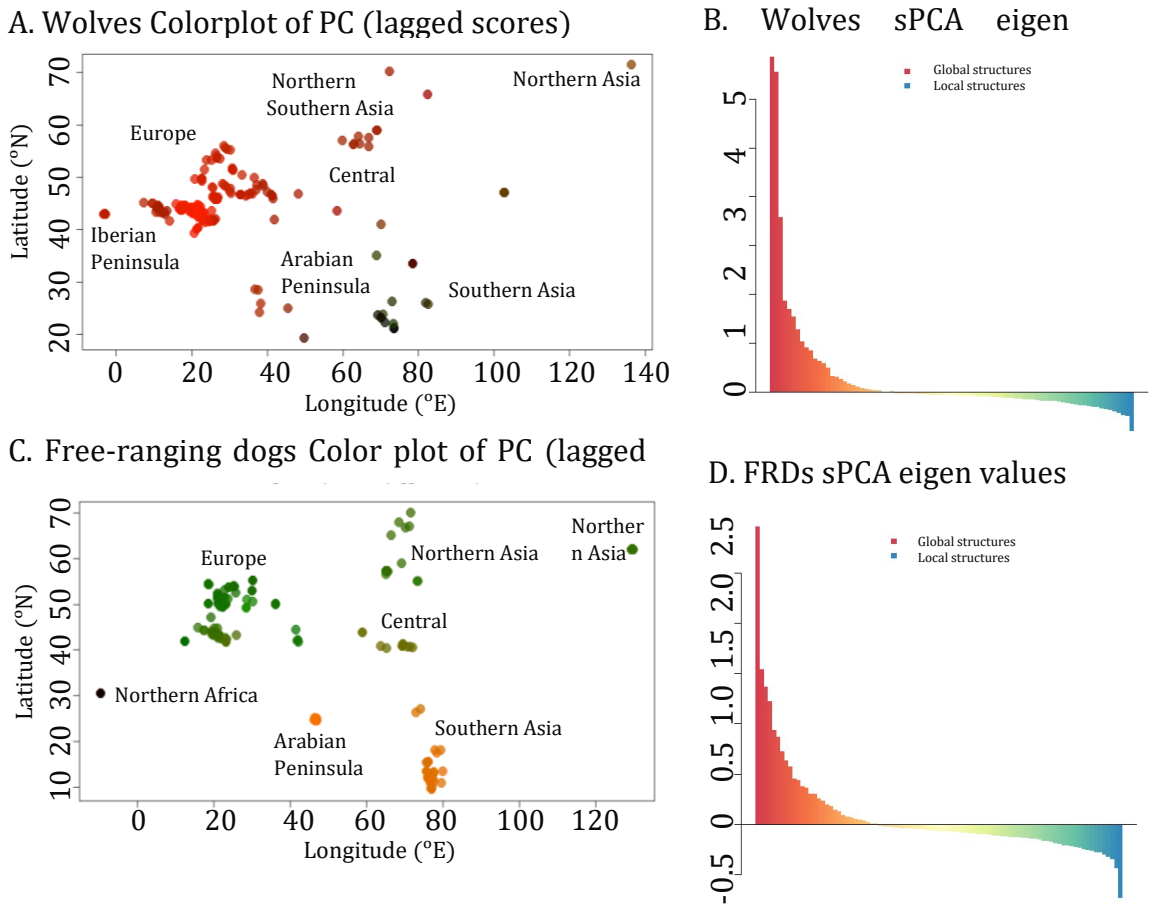


Figure 5. Colour plots for sPCA lagged scores onto geographical space for wolves (A) and FRDs (C). Each principal component score was translated into a channel of colour, each score into a channel of colour which are then used to compose a colour using the RGB system (red, green, and blue). Genetic variation in wolves and FRDs is suggested to be spatially structured as genetic variation clustering is seen in spatial groups. B: Histograms of sPCA eigen values for global and local structures. Based on the respective highest eigen values of 7 in wolves (B) and 3 in FRDs (D), a global structure is more likely than a local structure in both.

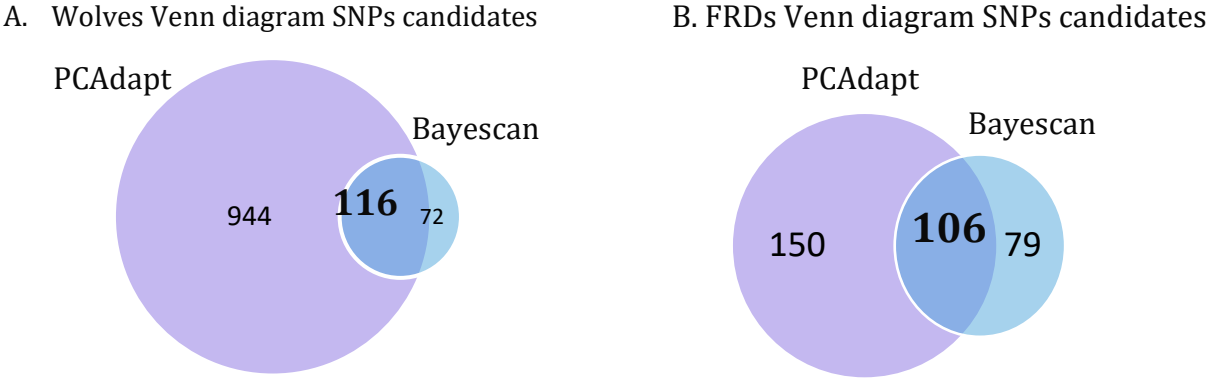


Figure 6. Candidate SNPs and proteins identified under diversifying selection. A-B: Venn diagrams for candidate SNPs identified to be under diversifying and balancing selection using two different approaches PCAdapt and Bayescan for grey wolves (A) and FRDs (B).

In FRDs 517 genes under diversifying selection were identified with VEP tool, of which 134 were RNA genes, 11 were pseudogenes and 372 protein coding genes (Table S4). The protein coding genes were enriched for GO terms related to the olfactory system and morphology (Fig. 7B, 7D, Fig. S13). STRING software detected 57 protein interaction clusters and 9 significant local clusters. These clusters were also enriched for GO terms related to morphology (Figs. S15, S17). The two largest protein interaction clusters were comprised of genes with enriched GO terms related to stimulus response, olfactory system, embryonic and neuronal development, and gene regulation (Fig. S19). Additionally, the VEP tool detected six variants with signatures of selection predicted to be missense mutations. Two of them were predicted to not be tolerated with SIFT scores of 0.01 (Fig. S13): one within the leucine rich repeats of the nucleotide binding oligomerization domain containing 1 (NOD1), and one in SLC6A9 which was also detected in wolves. Both missense mutations resulted in a G > S change, thus from the smallest ammino acid to a polar nonchanged amino acid.

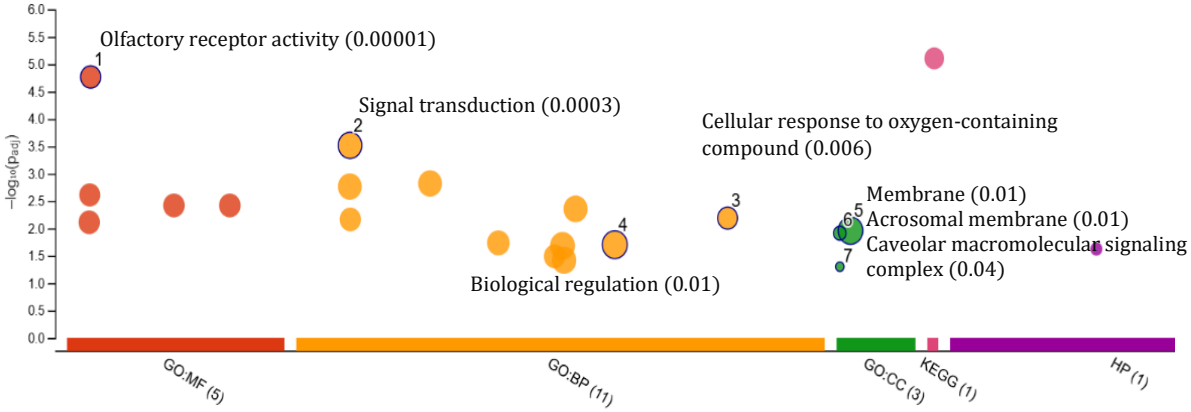
4. Discussion

Our results showed a contrast between genetic differentiation patterns among FRDs and grey wolves. FRDs, including the North African population, showed strong signature of Eurasian ancestry, while grey wolves displayed stronger genetic differentiation with clustering patterns in an East-West cline (Fig. S10). There was a substantial overlap between genes with signatures of diversifying selection in both FRDs and grey wolves (i.e., 322 genes), many of which were associated with phenotypic traits related to the sensory perception, particularly olfactory systems and detection of stimuli. A notable gene was SLC6A9, a transmembrane transporter domain of the solute carrier, for which a missense mutation predicted to impact protein function showed a signal of diversifying selection in both taxa.

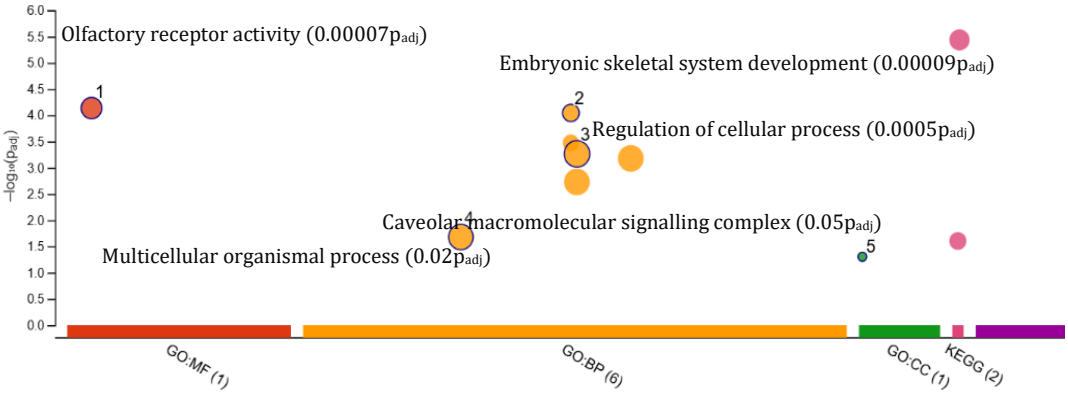
4.1. Regional variation among FRDs and grey wolves

Genetic diversity of both Eurasian grey wolves and FRDs showed clear spatial patterns (Fig. 5B, 5D) and individuals which were geographically close tended to share similar genotypes (Figs. 2B, 3B, 5A, 5C). Clustering patterns in non-peninsular wolf and FRD populations were consistent with previous findings (Bergström et al., 2020; Coutinho-Lima et al., 2024; Pilot et al. 2006), including separate clustering of wolves from the Carpathian Mountains (Ericson et al., 2020; Pilot et al., 2006; Stronen et al., 2015; Stronen et al., 2022) (Figs. 2A, 4A-C). Most non-Peninsular FRDs and North African individuals clustered together, which echoes previous studies on African dogs (Bergström et al., 2020; Boyko et al., 2009; Coutinho-Lima et al., 2024; Shannon et al., 2015). Eurasian ancestry patterns observed in North African individuals could be explained by recent and/or historical admixture with European breed dogs (Perri et al., 2021; Boyko et al., 2009). Dog dispersal is facilitated by anthropogenic factors (Smith et al., 2022; Perri et al., 2021), which allow dogs to highly expand their natural dispersal capabilities by providing methods of long-range transportation. Historically dog dispersal patterns have shadowed anthropogenic patterns (Perri et al., 2021).

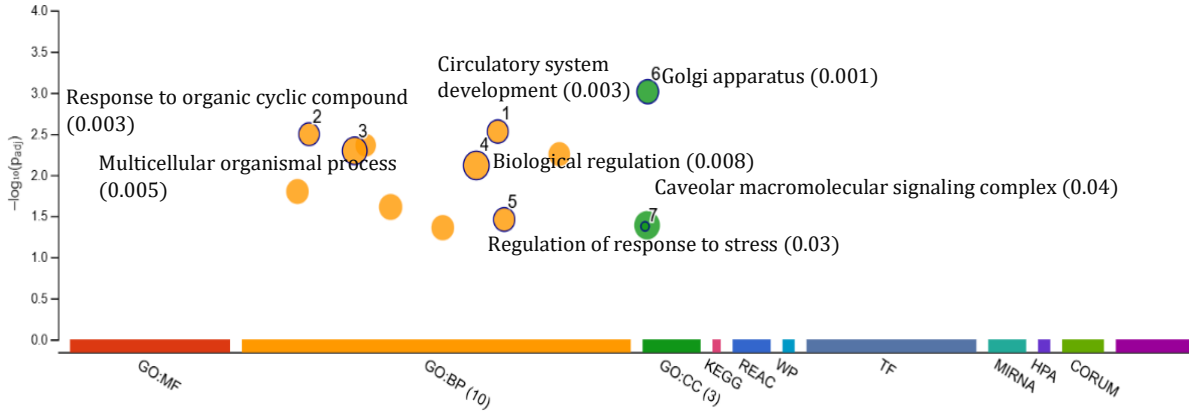
A. Wolves Bubble plot GO enrichment dog genome



B. FRDs Bubble plot GO enrichment dog genome



C. Wolves Bubble plot GO enrichment human genome



D. FRDs Bubble plot GO enrichment human genome

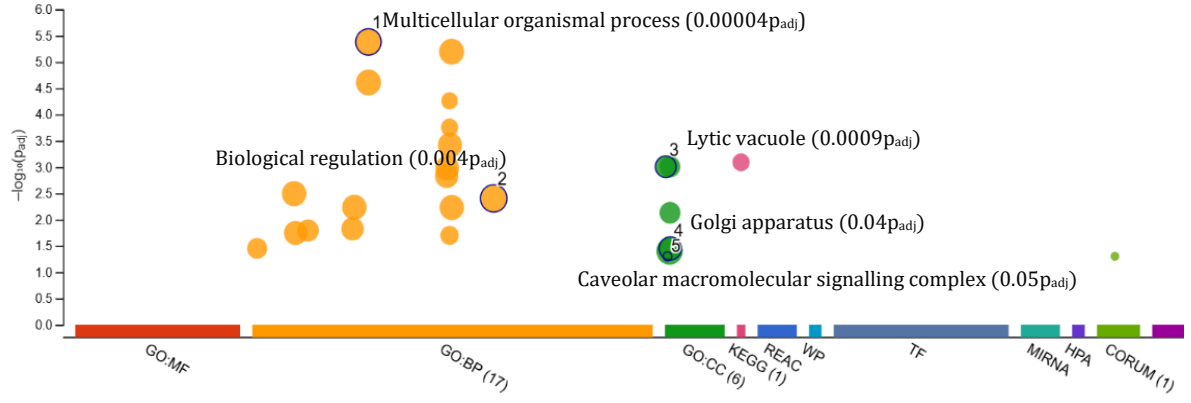


Figure 7. gProfiler gene ontology enrichment bubble plots with both the Canis lupus familiaris genome (A-B) and Homo sapiens sapiens genome (C-D) for grey wolves and FRDs respectively. The significantly enriched categories were highlighted and numbered based on significance and their respective adjusted p values are included in the parenthesis.

Table 1. Genes with signatures of selection previously detected under selection in dogs and wolves

Gene	Taxa	Consequence	Function	Previous study
DLA-12	D/W	Upstream/ Intron-Upstream	Part of the major histocompatibility complex class I (MHCI) gene family, highly polymorphic and part of the adaptive immunity (Wieczorek et al., 2017)	Liu et al., 2017
DLA88	D/W	Intron	Has been proposed as the ancestral gene of DLA-12 in canids (Feng et al., 2023; Miyamae et al., 2022). Detected to be very diverse and highly expressed among MHCI genes (Feng et al., 2023).	Liu et al., 2017
EYA1	D/W	Upstream/Downstream-Intron-Upstream	Involved in cell differentiation	Schweizer et al., 2016
GAA	D	Downstream	Involved in glycogen breakdown, GAA deficiency has been associated with the fatal metabolic disease Pompe in dogs (Seppälä et al., 2013)	Pilot et al., 2021
HNF4G	W	Downstream	Involved in anatomical structure development and cell differentiation via regulation of RNA polymerase	Plassais et al., 2019
IGF2BP2	D/W	Intron	Variants of these genes have been described to result in reduction of mean height and weight in pure-breed dogs (Doherty et al., 2020; Wade, 2025).	Plassais et al., 2019
LRP6	W	Intron	Has been identified as the functional entry receptor for canine distemper virus (Gradauskaite et al., 2023).	Schweizer et al., 2016
OR13J1	D/W	Downstream	Olfactory receptor	Zhang et al., 2020
PIBF1	D/W	Intron	Involved in cilium assembly	Zhang et al., 2020
SLC4A7	W	Upstream	Involved in auditory receptor cell development via transmembrane transporter activity	Zhang et al., 2020
SMAD2	D/W	Downstream/ Downstream-Intron	Has been associated with body weight variation in dogs (Bannasch et al., 2020; Doherty et al., 2020)	Plassais et al., 2019
TRDN	D/W	Downstream and Intron in Wolves	Expressed in dog cardiac tissue (Le Béguec et al., 2018), enables signaling receptor binding	Pilot et al., 2021
WNT10A	D	Upstream	Mammalian tooth development (Yin, 2015). WNT signalling pathway is involved in skeletal and mandible formation (Choe et al., 2022; Smolders et al., 2013).	Wei et al., 2015

Even temporary human dispersal patterns (i.e., tourism) have been shown to increase dispersal in FRDs between areas separated by physical barriers (Schüttler & Jiménez, 2022; Vychodilova et al., 2018). In addition to human-facilitated connectivity, the difference between evolutionary histories, i.e., a more recent evolutionary origin from 1-2 ancestral wolf population(s) (Freedman et al., 2014; Bergström et al., 2020, Bergström et al. 2022), might explain the weaker genetic differentiation observed in FRDs.

FRDs and wolves' populations from peninsular regions were consistently found to be genetically differentiated (Figs. 2-4). The observed genetic differentiation in peninsular wolves was expected, since similar patterns have been previously observed (Pilot et al, 2010; Pilot et al., 2014; Sastre et al., 2011; Vila et al., 1999; Stronen et al., 2022). Genetic differentiation in peninsular grey wolves is likely due to historical populations remaining in peninsular glacial refugia during Last Glacial Maximum, while northern Eurasia was subsequently recolonized by Siberian populations (Bergström et al. 2022). Conversely, the genetic differentiation found in the Italian Peninsula FRDs was unexpected. Previous studies on village dogs and FRDs from the Italian Peninsula showed them clustering with other European dogs (Boyko et al., 2009; Bergström et al., 2020; Natoli et al. 2021, Coutinho-Lima et al., 2024). The genetic differentiation was less pronounced in Italian FRDs than other peninsular populations. Considering that all Italian Peninsula FRDs originated from the same local population (Fig. 1A), a possible explanation for this discrepancy could be the fact that most individuals were distantly related (as closest relatives were removed) coupled with the relatively small sample size for this populations that could affect ancestry coefficient estimation algorithms (Caye et al., 2016; Frichot et al., 2014; Alexander & Lange, 2011).

Genetic differentiation in FRDs from the Arabian Peninsula posed an interesting contrast to the grey wolves from this peninsula, which clustered with Eurasian wolves (Figs. 2, 3D-F). The difference in patterns suggests that there are ecological factors impacting gene flow between contemporary Arabian FRDs and other populations (Fig. 3B), possibly due to habitat differentiation. Another possible explanation are anthropogenic factors e.g., cultural differentiation throughout this region, which may have reduced human-facilitated gene flow between FRDs from the Indian Peninsula and the rest of Eurasia. Overall, genetic differentiation of peninsular populations could be due to a varying degree of anthropogenic impact, orographic barriers and climate transitions limiting dispersal (Ficetola et al., 2021). Alternatively, since there were multiple routes for dispersal across Asia for both grey wolves and dogs (Bergström et al., 2020, 2022), perhaps in the southern Peninsulas older lineages were preserved.

4.2. Diversifying selection in FRDs and grey wolves

There was substantial overlap between the genes with signatures of diversifying selection among both taxa. Many of the genes with signatures of selection coded for RNA genes, including LncRNAs, i.e. RNA transcripts with over 200 nucleotides, which have been shown to have roles in a myriad of cellular processes from gene expression to metabolism (Bridges et al., 2021). Thus, detecting LncRNAs with signatures of

diversifying selection underscores the need for deeper understanding of the functions of RNA genes. Most of the protein coding genes under diversifying selection had products that function as receptor and regulatory proteins (Fig. 7). Some of these genes had functions in cellular processes and the olfactory system in both grey wolves and FRDs, and signatures of diversifying selection were found in genes related to embryonic development, nervous system and morphology in FRDs (Figs. S18-S19). We detected over 648 genes under diversifying selection in this study; hence we focused on these gene functions for which genes were significantly enriched, notable genes (i.e., genes detected to be under selection in previous studies) and genes with variants predicted to result in missense mutations.

4.2.1. Genes with signatures of diversifying selection

Our analysis detected multiple genes involved in canine morphology and immune system which have been previously detected to be under selection (Table 1), validating our findings. The highest-ranking GO term in both taxa was olfactory receptor activity (Fig. 7A-B). This is unsurprising as for canids sense of smell is important for navigating their environments, which is underlined by over 970 olfactory receptor genes identified in dogs (Olender et al., 2004). Mutations within olfactory receptors have been described as drivers of niche adaptation in vertebrates (Chapter 1, Huges et al., 2018). Our results emphasize the role of OR genes in canid adaptive evolution.

Signatures of selection in WNT10 (a developmental gene which has been implicated in tooth development (Table 1)) and GAA (essential for glycogen breakdown (Table 1)) in FRDs specifically allude to adaptations for dietary shifts, reiterating the need for a deeper understanding of FRDs diets. In addition to developmental and metabolic adaptations we also detected immune system adaptations. Adaptive variation in LRP6 (i.e., a functional entry receptor for canine distemper virus (CDV)) was detected in wolves and warrants further investigation, given the role of this protein in CDV infection. CDV is a disease akin to human measles (Rivera-Martínez et al., 2024) with multi-host capacity, although dogs are the ones most frequently affected (Wipf et al., 2025). CDV has global prevalence, and its infectivity is impacted by variation in a glycoprotein (i.e., H protein) which enables the virus to attach and enter host cells (Rivera-Martínez et al., 2024). This virus has been proposed to result in higher mortality in isolated wolf populations (Di Francesco et al., 2020). In this study we detected genetically differentiated wolf populations (e.g., peninsular and from the Carpathian Mountains) (Figs., 2, 4D-E), which may be more vulnerable to CDV. Additionally, incidence of the disease in wolves has been suggested to increase in par with more frequent interactions with dogs (Brandell et al., 2021). Further understanding of the variants under diversifying selection we identified could prove to be relevant in management of wolf populations.

4.2.2. Genes with missense mutations

Missense mutations were detected in PKLR, OR5M13, LOC481603, INSM2 and SLC6A9 in both FRDs and grey wolves (Fig. S13). An additional missense mutation was detected in NOD1 in FRDs. PKLR gene codes for the pyruvate kinase, a rate-limiting enzyme in the glycolytic pathway in blood and liver cells. In both humans and dogs, there are

mutations in this gene which have been associated with haemolytic diseases (Wang et al., 2024; Wynkoop et al., 2023). The identified missense variant does not lie within the predicted functional domains of the protein. Thus, it likely does not impact the proteins' current function.

We also detected missense variants in the olfactory receptors (OR) OR5M13 and LOC481603. The protein structures of these two receptors have not been extensively studied, hence the SIFT prediction was not possible in the former (Fig. S13). Loss of function (pseudogenization) in OR genes has been proposed as an evolutionary mechanism in mammals and has been associated with adaptations for shifts in ecological niches or shifts in sensory perception in response to environmental shifts (Chapter 1, Liu et al., 2019; Niimura & Nei, 2007; Zhou et al., 2023). In dogs, OR genes have been suggested to have experienced deletions and pseudogenization during the domestication (Mouton et al., 2025), thus their olfactory acumen was negatively impacted by the process. Additionally, patterns of pseudogenization vary within the domestic dogs, with differences in patters of pseudogenization of specific OR genes among breeds (Azzouzi et al., 2022; Jenkins et al., 2018; Robins et al., 2009; Singletary & Hagerty, 2023). It is possible that the detected variants arose as part of these complex evolutionary dynamics observed in vertebrate OR genes and are part of pseudogenization variability. Alternatively, receptor proteins' main function is recognition of ligands and changes in the receptors may yield beneficial phenotypes when encountering novel ligands (Chapter 1). Olfactory receptors are involved in recognition of environmental molecules, and adaptations in these proteins could have been beneficial for both FRDs and wolves as they encountered non-familiar substances in natural environments.

Another missense mutation was detected in the transcriptional repressor INSM2, whose function has not been fully investigated in canids. The variant detected in this gene did not have a SIFT score calculation, i.e., predicted impact on protein function, likely due to the knowledge gap in the function of this protein. Knock-out experiments in mice have suggested that INSM2 has a pivotal role in glucose tolerance, insulin secretion (Cao et al., 2022) and oocyte development (Li et al., 2022). Thus, it is possible that this protein has a role in adaptations for the consumption of anthropogenic foods and/or reproduction. Our findings underscore the need for further studies into the expression and function of this protein.

4.2.3. Genes with missense mutations predicted to affect function

There were two missense mutations predicted to impact protein function based on SIFT analysis (i.e., primary and secondary amino acid sequence analysis), in NOD1 in FRDs and SLC6A9 in both taxa. NOD1 has been described as a pattern receptor involved in pathogen recognition within the innate immune response signalling pathway (Bröer & Fairweather, 2018; Okanishi, et al., 2013). The leucine rich repeats in NOD1 are the bacterial ligand pattern recognition domain and this is where the identified missense variant is located (Fernández-García et al., 2021). The resulting amino acid change from glycine, the simplest of amino acids, to a serine which has an additional polar chain could be causing a change in the structure of the pattern recognition domain, impacting

this protein function. The mutation may be an adaptation to recognize a novel or recently evolved bacterial pathogens.

Our analysis also detected a variant predicted to affect protein function in SLC6A9 in both taxa. This protein belongs to a solute carrier family of proteins, which are involved in homeostatic processes across a variety of specialized cells (Nguyen et al., 2021). SLC6A9 is described as a central nervous system inhibitor linked with synaptic signalling (i.e., neurotransmitter), transporter of glycine. SLC6A9 may be involved in behavioural adaptations or stimuli processing (Mercier et al., 2012). Interestingly, another sodium-dependent glucose transporter SLC5A1 (Table 2) has been identified to be under selection during the development of European dog breeds (Zhang et al., 2020). SLC proteins could prove to be important in canid adaptive evolution and further investigations into their variation in the context of environmental pressures could shed some light on the matter.

5. Conclusion

Eurasian FRDs display weaker genetic differentiation than grey wolves, influenced by both anthropogenically facilitated dispersal and differences in evolutionary histories. Genetic differentiation patterns in peninsular grey wolves and FRDs indicated that either these southern populations have ecological barriers limiting their dispersal or older genetic diversity has been preserved in these populations due to variations in dispersal patterns during their range expansion. Similar mechanisms underlying adaptive variation under diversifying selection in FRDs and wolves is implied from the overlap between genes with signatures of diversifying selection. SLC6A9, a neurotransmitter transporter, may have an important role in canid evolution and warrants future investigation. Genes involved in stimulus response, particularly those related to olfaction play a critical role in canids' adaptative processes. Additionally, genes involved in morphological variation are important in the adaptive evolutionary landscape of FRDs. Overall, FRD and grey wolf populations exhibit parallel molecular adaptations under comparable selective pressures.

References

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19(10), 1655–1664. https://doi.org/10.1101/gr.094052.109

Axelsson, E., Ratnakumar, A., Arendt, M.-L., Maqbool, K., Webster, M. T., Perloski, M., Liberg, O., Arnemo, J. M., Hedhammar, Å., & Lindblad-Toh, K. (2013). The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature*, 495(7441), 360–364. https://doi.org/10.1038/nature11837

Azzouzi, N., Guillory, A. S., Chaudieu, G., Moisan, M. P., Bertaud, A., & Charpigny, G. (2022). Dog olfactory receptor gene expression profiling using samples derived from nasal epithelium brushing. *Canine Medicine and Genetics, 9*, 7. https://doi.org/10.1186/s40575-022-00116-7

Bannasch, D. L., Baes, C. F., & Leeb, T. (2020). Genetic variants affecting skeletal morphology in domestic dogs. *Trends in Genetics*, 36(8), 598–609. https://doi.org/10.1016/j.tig.2020.05.005

Brandell, E. E., Cross, P. C., Craft, M. E., Fountain-Jones, N. M., Munkhtsog, B., & Carver, S. (2021). Patterns and processes of pathogen exposure in gray wolves across North America. *Scientific Reports, 11*, 3722. https://doi.org/10.1038/s41598-021-81192-w

Belsare, A., & Vanak, A. T. (2020). Modelling the challenges of managing free-ranging dog populations. *Scientific Reports*, 10, 18874. https://doi.org/10.1038/s41598-020-75828-6

Bergström, A., Frantz, L., Schmidt, R., Ersmark, E., Lebrasseur, O., Girdland-Flink, L., Lin, A. T., Storå, J., Sjögren, K. G., Anthony, D., Antipina, E., Amiri, S., Bar-Oz, G., Bazaliiskii, V. I., Bulatović, J., Brown, D., Carmagnini, A., Davy, T., Fedorov, S., Fiore, I., … Skoglund, P. (2020). Origins and genetic legacy of prehistoric dogs. *Science (New York, N.Y.)*, *370*(6516), 557–564. https://doi.org/10.1126/science.aba9572

Bergström, A., Stanton, D. W. G., Taron, U. H., Frantz, L., Sinding, M. S., Ersmark, E., Pfrengle, S., Cassatt-Johnstone, M., Lebrasseur, O., Girdland-Flink, L., Fernandes, D. M., Ollivier, M., Speidel, L., Gopalakrishnan, S., Westbury, M. V., Ramos-Madrigal, J., Feuerborn, T. R., Reiter, E., Gretzinger, J., Münzel, S. C., … Skoglund, P. (2022). Grey wolf genomic history reveals a dual ancestry of dogs. *Nature*, *607*(7918), 313–320. https://doi.org/10.1038/s41586-022-04824-9

Bivand, R. (2022). R packages for analyzing spatial data: A comparative case study with areal data. Geographical Analysis, 54(3), 488–518. https://doi.org/10.1111/gean.12319

Boyko, A. R., Boyko, R. H., Boyko, C. M., Parker, H. G., Castelhano, M., Corey, L., Degenhardt, J. D., Auton, A., Hedimbi, M., Kityo, R., Ostrander, E. A., Schoenebeck, J., Todhunter, R. J., Jones, P., & Bustamante, C. D. (2009). Complex population structure in African village dogs and its implications for inferring dog domestication history.

Proceedings of the National Academy of Sciences, 106(33), 13903–13908. https://doi.org/10.1073/pnas.0902129106

Bridges, M. C., Daulagala, A. C., & Kourtidis, A. (2021). LNCcation: lncRNA localization and function. *Journal of Cell Biology*, 220(2), e202009045. https://doi.org/10.1083/jcb.202009045

Bröer, S., & Fairweather, S. J. (2018). Amino acid transport across the mammalian intestine. *Comprehensive Physiology*, 9(1), 343–373. https://doi.org/10.1002/cphy.c170041

Bryce, C. (2021). Dogs as pets and pests: Global patterns of canine abundance, activity, and health. *Integrative and Comparative Biology*, 61, https://doi.org/10.1093/icb/icab046

Cao, H., Zhuo, R., Zhang, Z., Wang, J., Tao, Y., Yang, R., Guo, X., Chen, Y., Jia, S., Yao, Y., Yang, P., Yu, J., Jiao, W., Li, X., Fang, F., Xie, Y., Li, G., Wu, D., Wang, H., Feng, C., … Wang, J. (2022). Super-enhancer-associated INSM2 regulates lipid metabolism by modulating mTOR signaling pathway in neuroblastoma. *Cell & bioscience*, *12*(1), 158. https://doi.org/10.1186/s13578-022-00895-3

Caye, K., Jumentier, B., Lepeule, J., & François, O. (2019). LFMM 2: Fast and accurate inference of gene-environment associations in genome-wide studies. *Molecular Biology and Evolution*, 36(4), 852–860. https://doi.org/10.1093/molbev/msz008

Caye, K., Deist, T., Martins, H., Michel, O., & Francois, O. (2016). TESS3: Fast inference of spatial population structure and genome scans for selection. *Molecular Ecology Resources*, 16(2), 540-548. https://doi.org/10.1111/1755-0998.12471

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, 4, 7. https://doi.org/10.1186/s13742-015-0047-8

Choe, Y. H., Hur, T. Y., Lee, S. L., Lee, S., Lim, D., Choi, B. H., Jeong, H., No, J. G., & Ock, S. A. (2022). Brachygnathia inferior in cloned dogs is possibly correlated with variants of Wnt signaling pathway initiators. *International Journal of Molecular Sciences*, 23(1), 475. https://doi.org/10.3390/ijms23010475

Coutinho-Lima, D., Dreger, D. L., Doadrio, I., Parker, H. G., Ghanavi, H. R., Frantz, L., Larson, G., Ostrander, E. A., & Godinho, R. (2024). Multiple ancestries and shared gene flow among modern livestock guarding dogs. *iScience*. https://doi.org/10.1016/j.isci.2024.110396

Di Francesco, C. E., Smoglica, C., & Angelucci, S. (2020). Infectious diseases and wildlife conservation medicine: The case of canine distemper in European wolf populations. *Animals*, 10(12), 2426. https://doi.org/10.3390/ani10122426

Ericson, H. S., Fedorca, A., Toderas, I., Hegyeli, Z., Plis, K., Dykyy, I., Jędrzejewska, B., Ionescu, G., Fedorca, M., Iacolina, L., & Stronen, A.V. (2020). Genome-wide profiles indicate wolf population connectivity within the eastern Carpathian Mountains. *Genetica*, 148(1), 33–39. https://doi.org/10.1007/s10709-019-00083-1

Feng, Y., Hess, P., Tompkins, S., Hildebrand, W., & Zhao, S. (2023). A Kmer-based paired-end read de novo assembler and genotyper for canine MHC class I genotyping. *iScience*, 26, 105996. https://doi.org/10.1016/j.isci.2023.105996

Fernández-García, V., González-Ramos, S., Martín-Sanz, P., García-Del Portillo, F., Laparra, J. M., & Boscá, L. (2021). NOD1 in the interplay between microbiota and gastrointestinal immune adaptations. *Pharmacological Research*, 171, 105775. https://doi.org/10.1016/j.phrs.2021.105775

Feuerborn, T. R., Carmagnini, A., Losey, R. J., Nomokonova, T., Askeyev, A., Askeyev, I., Askeyev, O., Antipina, E. E., Appelt, M., Bachura, O. P., Beglane, F., Bradley, D. G., Daly, K. G., Gopalakrishnan, S., Murphy, Gregersen, K., Guo, C., Gusev, A. V., Jones, C., Kosintsev, P. A., Kuzmin, Y. V., … Frantz, L. (2021). Modern Siberian dog ancestry was shaped by several thousand years of Eurasian-wide trade and human dispersal. *Proceedings of the National Academy of Sciences of the United States of America*, 118(39), e2100338118. https://doi.org/10.1073/pnas.2100338118

Fleming, P. J. S., Nolan, H., Jackson, S. M., Ballard, G.-A., Bengsen, A., Brown, W. Y., Meek, P. D., Mifsud, G., Pal, S. K., & Sparkes, J. (2017). Roles for the Canidae in food webs reviewed: Where do they fit? *Food Webs*, *12*, 14-34. https://doi.org/10.1016/j.fooweb.2017.03.001

Foll, M., & Gaggiotti, O. E. (2008). A genome scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics*, 180(2), 977–993. https://doi.org/10.1534/genetics.108.092221

Ficetola, G. F., Mazel, F., Falaschi, M., Marta, S., & Thuiller, W. (2021). Determinants of zoogeographical boundaries differ between vertebrate groups. *Global Ecology and Biogeography*, 30(10), 1796–1809. https://doi.org/10.1111/geb.13345

Freedman, A. H., Gronau, I., Schweizer, R. M., Ortega-Del Vecchyo, D., Han, E., Silva, P. M., Galaverni, M., Fan, Z., Marx, P., Lorente-Galdos, B., Beale, H., Ramirez, O., Hormozdiari, F., Alkan, C., Vilà, C., Squire, K., Geffen, E., Kusak, J., Boyko, A. R., Parker, H. G., … Novembre, J. (2014). Genome sequencing highlights the dynamic early history of dogs. *PLoS genetics*, 10(1), e1004016. https://doi.org/10.1371/journal.pgen.1004016

Frichot, E., & François, O. (2015). LEA: An R package for Landscape and Ecological Association studies. *Methods in Ecology and Evolution*. http://membres-timc.imag.fr/Olivier.Francois/lea.html

Gerber, N., Riesch, F., Bojarska, K., Zetsche, M., Rohwer, N.-K., Signer, J., Isselstein, J., Herzog, S., Okarma, H., Kuijper, D. P. J., & Balkenhol, N. (2024). Do recolonizing wolves trigger non-consumptive effects in European ecosystems? A review of evidence. *Wildlife Biology*, 2024, e01229. https://doi.org/10.1002/wlb3.01229

Gompper, M. E. (2014). Introduction: Outlining the ecological influences of a subsidized, domesticated predator. In Free-Ranging Dogs and Wildlife Conservation (pp. 1–8). Oxford University Press.

Hall, B. E., Mazhar, K., Macdonald, E., Cassidy, M., Doty, M., Judkins, C., Terse, A., Shiers, S., Tadros, S., Yun, S., Burton, M. D., Price, T. J., & Kulkarni, A. (2024). Transcriptome analysis of rheumatoid arthritis uncovers genes linked to inflammation-induced pain. *Research Square*. https://doi.org/10.21203/rs.3.rs-4218885/v1

Hendricks, S. A., Schweizer, R. M., & Wayne, R. K. (2019). Conservation genomics illuminates the adaptive uniqueness of North American gray wolves. Conservation *Genetics*, 20(1), 29–43. https://doi.org/10.1007/s10592-018-1118-z

Hijmans, R. J., Bivand, R., Pebesma, E., & Sumner, M. D. (2023). terra: Spatial Data Analysis. [R package]. https://cran.r-project.org/web/packages/terra/index.html

Hughes, J., & Macdonald, D. W. (2013). A review of the interactions between free-roaming domestic dogs and wildlife. *Biological Conservation*, 157, 341-351. https://doi.org/10.1016/j.biocon.2012.07.005

Hughes, G. M., Boston, E. S. M., Finarelli, J. A., Murphy, W. J., Higgins, D. G., & Teeling, E. C. (2018). The birth and death of olfactory receptor gene families in mammalian niche adaptation. *Molecular Biology and Evolution*, 35(6), 1390–1406. https://doi.org/10.1093/molbev/msy028

Jenkins, E. K., DeChant, M. T., & Perry, E. B. (2018). When the Nose Doesn't Know: Canine Olfactory Function Associated With Health, Management, and Potential Links to Microbiota. *Frontiers in veterinary science*, *5*, 56. https://doi.org/10.3389/fvets.2018.00056

Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, 27(21), 3038–3039. https://doi.org/10.1093/bioinformatics/btr521

Jumentier, B. (2021). lfmm: Latent Factor Mixed Models. R package version 1.1. https://CRAN.R-project.org/package=lfmm

Kolberg, L., Raudvere, U., Kuzmin, I., Adler, P., Vilo, J., & Peterson, H. (2023). g:Profiler—Interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). *Nucleic Acids Research*, 51(W1), W222–W227. https://doi.org/10.1093/nar/gkad347

Larson, G., & Fuller, D. Q. (2014). The evolution of animal domestication. *Annual Review of Ecology, Evolution, and Systematics*, *45*, 115-136. https://doi.org/10.1146/annurev-ecolsys-110512-135813

Letunic, I., & Bork, P. (2020). SMART: Recent updates, new developments and status in 2020. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkaa937

Leonard, J. (2014). Ecology drives evolution in grey wolves. *Evolutionary Ecology Research*, 16, 461–473.

Li, Z. M., Li, Y. Y., Fei, C. F., & Zhou, L. Q. (2022). Insm2 deficiency results in female infertility by disturbing steroid pathway and decreasing ovarian reserve in mice. Cell

Cycle (Georgetown, Tex.), 21(21), 2255–2267. https://doi.org/10.1080/15384101.2022.2092816

Liu, A., He, F., Shen, L., Zhang, J. & Zhou, J. (2019). Convergent degeneration of olfactory receptor gene repertoires in marine mammals. *BMC Genomics, 20*(1), 977. https://doi.org/10.1186/s12864-019-6290-0

Liu, G., Zhang, H., Sun, G., Zhao, C., Shang, S., Gao, X., Xia, T., & Yang, X. (2017). Characterization of the peripheral blood transcriptome and adaptive evolution of the MHC I and TLR gene families in the wolf (*Canis lupus*). *BMC Genomics*, 18, 584. https://doi.org/10.1186/s12864-017-3983-0

Lin, A. P., Abbas, S., Kim, S. W., Ortega, M., Bouamar, H., Escobedo, Y., Varadarajan, P., Qin, Y., Sudderth, J., Schulz, E., Deutsch, A., Mohan, S., Ulz, P., Neumeister, P., Rakheja, D., Gao, X., Hinck, A., Weintraub, S. T., DeBerardinis, R. J., Sill, H., … Aguiar, R. C. (2015). D2HGDH regulates alpha-ketoglutarate levels and dioxygenase function by modulating IDH2. *Nature communications*, *6*, 7768. https://doi.org/10.1038/ncomms8768

Kaczensky, P., Chapron, G., von Arx, M., Huber, D., Andrén, H., & Linnell, J. (2013). Status, management and distribution of large carnivores—bear, lynx, wolf & wolverine—in Europe. Verlag nicht ermittelbar.

Kummer, E., Schubert, K. N., Schoenhut, T., Scaiola, A., & Ban, N. (2021). Structural basis of translation termination, rescue, and recycling in mammalian mitochondria. *Molecular Cell*, 81(12), 2566–2582.e6. https://doi.org/10.1016/j.molcel.2021.03.042

Manichaikul, A., Mychaleckyj, J. C., Rich, S. S., Daly, K., Sale, M., & Chen, W. M. (2010). Robust relationship inference in genome-wide association studies. *Bioinformatics (Oxford, England)*, *26*(22), 2867–2873. https://doi.org/10.1093/bioinformatics/btq559.

McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R., Thormann, A., Flicek, P., & Cunningham, F. (2016). The Ensembl Variant Effect Predictor. *Genome biology*, 17(1), 122. https://doi.org/10.1186/s13059-016-0974-4

Mercier, E., Peters, I. R., Day, M. J., Clercx, C., & Peeters, D. (2012). Toll- and NOD-like receptor mRNA expression in canine sino-nasal aspergillosis and idiopathic lymphoplasmacytic rhinitis. *Veterinary Immunology and Immunopathology*, 145(3-4), 618-624. https://doi.org/10.1016/j.vetimm.2012.01.009

Miller, J. M., Cullingham, C. I., & Peery, R. M. (2020). The influence of a priori grouping on inference of genetic clusters: Simulation study and literature review of the DAPC method. *Heredity*, 125(5), 269–280. https://doi.org/10.1038/s41437-020-0348-2

Mijangos, J. L., Gruber, B., Berry, O. F., Pacioni, C., & Georges, A. (2022). dartR v2: An accessible genetic analysis platform for conservation, ecology, and agriculture. *Methods in Ecology and Evolution*. https://doi.org/10.1111/2041-210X.13918

Miyamae, J., Okano, M., Nishiya, K., Katakura, F., Kulski, J. K., Moritomo, T., & Shiina, T. (2022). Haplotype structures and polymorphisms of dog leukocyte antigen (DLA) class I loci shaped by intralocus and interlocus recombination events. *Immunogenetics,* 74(2), 245–259. https://doi.org/10.1007/s00251-021-01234-5

Mouton, A., Bird, D. J., Li, G., Craven, B. A., Levine, J. M., Morselli, M., Pellegrini, M., Van Valkenburgh, B., Wayne, R. K., & Murphy, W. J. (2025). Genetic and anatomical determinants of olfaction in dogs and wild canids. *Molecular Biology and Evolution, 42*(3), msaf035. https://doi.org/10.1093/molbev/msaf035

Natoli, E., Bonanni, R., Cafazzo, S., Mills, D. S., Pontier, D., & Pilot, M. (2021). Genetic inference of the mating system of free-ranging domestic dogs. *Behavioral Ecology*, 32(4), 646–656. https://doi.org/10.1093/beheco/arab011

Niimura, Y., & Nei, M. (2007). Extensive gains and losses of olfactory receptor genes in mammalian evolution. *PloS one*, *2*(8), e708. https://doi.org/10.1371/journal.pone.0000708

Nguyen, Y. T. K., Ha, H. T. T., Nguyen, T. H., & Nguyen, L. N. (2022). The role of SLC transporters for brain health and disease. *Cellular and Molecular Life Sciences*, 79, 20. https://doi.org/10.1007/s00018-021-04074-4

Oksanen, J., Simpson, G. L., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., . . . Weedon, J. (2022). vegan: Community Ecology Package (R package version 2.6-4). https://CRAN.R-project.org/package=vegan

Okanishi, H., Hayashi, K., Sakamoto, Y., Sano, T., Maruyama, H., Kagawa, Y., & Watari, T. (2013). NOD2 mRNA expression and NF-kappaB activation in dogs with lymphocytic plasmacytic colitis. *Journal of Veterinary Internal Medicine*, 27(3), 439–444. https://doi.org/10.1111/jvim.12082

Olender, T., Fuchs, T., Linhart, C., Shamir, R., Adams, M., Kalush, F., Khen, M., & Lancet, D. (2004). The canine olfactory subgenome. *Genomics*, 83(3), 361–372. https://doi.org/10.1016/j.ygeno.2003.08.009

Ostrander, E., Wayne, R., Freedman, A., & Davis, B.W. (2017). Demographic history, selection, and functional diversity of the canine genome. *Nature Reviews Genetics*, 18, 705–720. https://doi.org/10.1038/nrg.2017.67

Perri, A. R., Feuerborn, T. R., Frantz, L. A. F., Larson, G., Malhi, R. S., Meltzer, D. J., & Witt, K. E. (2021). Dog domestication and the dual dispersal of people and dogs into the Americas. *Proceedings of the National Academy of Sciences of the United States of America*, 118(6), e2010083118. https://doi.org/10.1073/pnas.2010083118

Pilot, M., Greco, C., vonHoldt, B. M., Jędrzejewska, B., Randi, E., Jędrzejewski, W., Sidorovich, V. E., Ostrander, E. A., & Wayne, R. K. (2014). Genome-wide signatures of population bottlenecks and diversifying selection in European wolves. *Heredity*, 112(4), 428–442. https://doi.org/10.1038/hdy.2013.122

Pilot, M., Jedrzejewski, W., Branicki, W., Sidorovich, V. E., Jedrzejewska, B., Stachura, K., & Funk, S. M. (2006). Ecological factors influence population genetic structure of

European grey wolves. *Molecular Ecology*, 15, 4533–4553. https://doi.org/10.1111/j.1365-294X.2006.03110.x

Pilot, M., Branicki, W., Jedrzejewski, W., Goszczyński, J., Jedrzejewska, B., Dykyy, I., Shkvyrya, M., & Tsingarska, E. (2010). Phylogeographic history of grey wolves in Europe. *BMC Evolutionary Biology*, 21, 10–104. https://doi.org/10.1186/1471-2148-10-104

Pilot, M., Moura, A. E., Okhlopkov, I. M., Mamaev, N. V., Manaseryan, N. H., Hayrapetyan, V., Kopaliani, N., Tsingarska, E., Alagaili, A. N., Mohammed, O. B., Ostrander, E. A., & Bogdanowicz, W. (2021). Human-modified canids in human-modified landscapes: The evolutionary consequences of hybridization for grey wolves and free-ranging domestic dogs. *Evolutionary applications*, *14*(10), 2433–2456. https://doi.org/10.1111/eva.13257

Plumer, L., Talvi, T., Männil, P., Urmas, S. (2018). Assessing the roles of wolves and dogs in livestock predation with suggestions for mitigating human–wildlife conflict and conservation of wolves. *Conservation Genetics*, 19, 665–672. https://doi.org/10.1007/s10592-017-1045-4

Privé, F., Luu, K., Vilhjálmsson, B. J., & Blum, M. G. (2020). Performing highly efficient genome scans for local adaptation with R package pcadapt version 4. *Molecular Biology and Evolution*. https://doi.org/10.1093/molbev/msaa053

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A toolset for whole-genome association and population-based linkage analysis. *American Journal of Human Genetics*, 81(3), 559–575. https://doi.org/10.1086/519795

QGIS.org. (n.d.). QGIS Geographic Information System. QGIS Association. http://www.qgis.org

Rivera-Martínez, A., Rodríguez-Alarcón, C. A., Adame-Gallegos, J. R., Laredo-Tiscareño, S. V., de Luna-Santillana, E. d. J., Hernández-Triana, L. M., & Garza-Hernández, J. A. (2024). Canine Distemper Virus: Origins, Mutations, Diagnosis, and Epidemiology in Mexico. *Life*, *14*(8), 1002. https://doi.org/10.3390/life14081002

Robin, S., Tacher, S., Rimbault, M., Vaysse, A., Dréano, S., André, C., Hitte, C., & Galibert, F. (2009). Genetic diversity of canine olfactory receptors. *BMC genomics*, *10*, 21. https://doi.org/10.1186/1471-2164-10-21

Sastre, N., Vila, C., Salinas, M., Bologov, V. V., Urios, V., Sanchez, A., Francino, O., & Ramírez, O. (2011). Signatures of demographic bottlenecks in European wolf populations. *Conservation Genetics*, 12, 701–712. https://doi.org/10.1007/s10592-010-0177-6

Schüttler, E., & Jiménez, J. E. (2022). Are tourists facilitators of the movement of free-ranging dogs? *Animals*, 12, 3564. https://doi.org/10.3390/ani12243564

Shannon, L. M., Boyko, R. H., Castelhano, M., Corey, E., Hayward, J. J., McLean, C., White, M. E., Abi Said, M., Anita, B. A., Bondjengo, N. I., Calero, J., Galov, A., Hedimbi,

M., Imam, B., Khalap, R., Lally, D., Masta, A., Oliveira, K. C., Pérez, L., Randall, J., … Boyko, A. R. (2015). Genetic structure in village dogs reveals a Central Asian domestication origin. *Proceedings of the National Academy of Sciences of the United States of America*, 112(44), 13639–13644. https://doi.org/10.1073/pnas.1516215112

Schweizer, R. M., vonHoldt, B. M., Harrigan, R., Knowles, J. C., Musiani, M., Coltman, D., Novembre, J., & Wayne, R. K. (2016). Genetic subdivision and candidate genes under selection in North American grey wolves. Molecular Ecology, 25(1), 380-402. https://doi.org/10.1111/mec.13364

Smith, L. M., Goold, C., Quinnell, R. J., Munteanu, A. M., Hartmann, S., Dalla Villa, P., & Collins, L. M. (2022). Population dynamics of free-roaming dogs in two European regions and implications for population control. *PLOS ONE*, 17(9), e0266636. https://doi.org/10.1371/journal.pone.0266636

Smith, D.W., & Peterson, R.O. (2021). Intended and unintended consequences of wolf restoration to Yellowstone and Isle Royale National Parks. Conservation *Science and Practice*, 3, e413. https://doi.org/10.1111/csp2.413

Smolders, L. A., Meij, B. P., Onis, D., Riemers, F. M., Bergknut, N., Wubbolts, R., Grinwis, G. C., Houweling, M., Groot Koerkamp, M. J., van Leenen, D., Holstege, F. C., Hazewinkel, H. A., Creemers, L. B., Penning, L. C., & Tryfonidou, M. A. (2013). Gene expression profiling of early intervertebral disc degeneration reveals a down-regulation of canonical Wnt signaling and caveolin-1 expression: Implications for development of regenerative strategies. *Arthritis Research & Therapy*, 15, R23. https://doi.org/10.1186/ar4157

Stefanovic, M., Bogdanowicz, W., Adavoudi, R., Martínez-Sosa, F., Doan, K., Flores-Manzanero, A., Yellapu, S., Gianluca, D., Djan, M., Giannatos, G., Hatlauf, J., Hayrapetyan, V., Heltai, M., Homel, K., Hulva, P., Ionică, A., & Jhala, Y., Juránková, J., Kaboli, M., & Pilot, M. (2024). Range-wide phylogeography of the golden jackals (*Canis aureus*) reveals multiple sources of recent spatial expansion and admixture with dogs at the expansion front. *Biological Conservation*, 290, 110448. https://doi.org/10.1016/j.biocon.2024.110448

Stronen, A. V., Jędrzejewska, B., Pertoldi, C., Demontis, D., Randi, E., Niedziałkowska, M., Borowik, T., Sidorovich, V. E., Kusak, J., Kojola, I., Karamanlidis, A. A., Ozolins, J., Dumenko, V., & Czarnomska, S. D. (2015). Genome-wide analyses suggest parallel selection for universal traits may eclipse local environmental selection in a highly mobile carnivore. *Ecology and Evolution*, 5, 4410–4425. https://doi.org/10.5061/dryad.9s1t9

Stronen, A. V., Mattucci, F., Fabbri, E., Galaverni, M., Cocchiararo, B., Nowak, C., Godinho, R., Ruiz-González, A., Kusak, J., Skrbinšek, T., Randi, E., Vlasseva, A., Mucci, N., & Caniglia, R. (2022). A reduced SNP panel to trace gene flow across southern European wolf populations and detect hybridization with other Canis taxa. *Scientific Reports*, 12(1), 4195. https://doi.org/10.1038/s41598-022-08132-0

Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., Gable, A. L., Fang, T., Doncheva, N. T., Pyysalo, S., Bork, P., Jensen, L. J., & von Mering, C. (2023). The STRING database in 2023: Protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Research*, 51(D1), D638–D646. https://doi.org/10.1093/nar/gkac1000

Šnjegota, D., Stronen, A. V., Boljte, B., Ćirović, D., Djan, M., Huber, D., Jelenčič, M., Konec, M., Kusak, J., & Skrbinšek, T. (2021). Population genetic structure of wolves in the northwestern Dinaric-Balkan region. *Ecology and Evolution*, 11(24), 18492–18504. https://doi.org/10.1002/ece3.8444

Tancredi, D., & Cardinali, I. (2023). Being a Dog: A Review of the Domestication Process. *Genes*, *14*(5), 992. https://doi.org/10.3390/genes14050992.

UniProt Consortium (2025). UniProt: the Universal Protein Knowledgebase in 2025. *Nucleic acids research*, *53*(D1), D609–D617. https://doi.org/10.1093/nar/gkae1010

Vila, C., Amorim, I. R., Leonard, J. A., Posada, D., Castroviejo, J., Petrucci-Fonseca, F., Crandall, K. A., Ellegren, H., & Wayne, R. K. (1999). Mitochondrial DNA phylogeography and population history of the grey wolf *Canis lupus*. *Molecular Ecology*, 8(12), 2089–2103. https://doi.org/10.1046/j.1365-294x.1999.00825.x

vonHoldt, B. M., Stahler, D. R., Brzeski, K. E., Musiani, M., Peterson, R., Phillips, M., Stephenson, J., Laudon, K., Meredith, E., Vucetich, J. A., Leonard, J. A., & Wayne, R. K. (2024). Demographic history shapes North American gray wolf genomic diversity and informs species' conservation. *Molecular Ecology*, 33(3), e17231. https://doi.org/10.1111/mec.17231

Vychodilova, L., Necesankova, M., Albrechtova, K., Hlavac, J., Modry, D., Janova, E., Vyskocil, M., Mihalca, A. D., Kennedy, L. J., & Horin, P. (2018). Genetic diversity and population structure of African village dogs based on microsatellite and immunity-related molecular markers. *PLOS ONE*, 13(6), e0199506. https://doi.org/10.1371/journal.pone.0199506

Wang, Y., Liu, J., Liu, T., An, X., Huang, L., Li, J., Zhang, Y., Xiang, Y., Xiao, L., Yi, W., Qin, J., Liu, L., Wang, C., & Yu, J. (2024). Pyruvate kinase deficiency and PKLR gene mutations: Insights from molecular dynamics simulation analysis. *Heliyon*, 10(5), e26368. https://doi.org/10.1016/j.heliyon.2024.e26368

Wierzbowska, I. A., Hędrzak, M., Popczyk, B., Okarma, H., & Crooks, K. R. (2016). Predation of wildlife by free-ranging domestic dogs in Polish hunting grounds and potential competition with the grey wolf. *Biological Conservation*, 201, 1–9. https://doi.org/10.1016/j.biocon.2016.06.016

Wei, C., Wang, H., Liu, G., Wu, M., Cao, J., Liu, Z., Liu, R., Zhao, F., Zhang, L., Lu, J., Liu, C., & Du, L. (2015). Genome-wide analysis reveals population structure and selection in Chinese indigenous sheep breeds. *BMC Genomics*, 16, 194. https://doi.org/10.1186/s12864-015-1384-9

Wieczorek, M., Abualrous, E. T., Sticht, J., Álvaro-Benito, M., Stolzenberg, S., Noé, F., & Freund, C. (2017). Major Histocompatibility Complex (MHC) Class I and MHC Class II proteins: Conformational plasticity in antigen presentation. *Frontiers in Immunology*, 8, 292. https://doi.org/10.3389/fimmu.2017.00292

Wipf, A., Perez-Cutillas, P., Ortega, N., Huertas-López, A., Martínez-Carrasco, C., & Candela, M. G. (2025). Geographical Distribution of Carnivore Hosts and Genotypes of Canine Distemper Virus (CDV) Worldwide: A Scoping Review and Spatial Meta-Analysis. *Transboundary and emerging diseases*, *2025*, 6632068. https://doi.org/10.1155/tbed/6632068

White, D. (2012). mmod : an R library for the calculation of population differentiation statistics. Molecular Ecology Resources, 12(6), 1158–1160. https://doi.org/10.1111/j.1755-0998.2012.03174.x

Wolf, C., & Ripple, W. J. (2017). Range contractions of the world's large carnivores. Royal Society open science, 4(7), 170052. https://doi.org/10.1098/rsos.170052

Wynkoop, M. R., Lalwani, S., Cipolli, W., & Jimenez, A. G. (2023). Scaling with body mass and age in glycolytic enzymes of domestic dogs. *Veterinary Research Communications*, 47(1), 39–50. https://doi.org/10.1007/s11259-022-09926-3

Wynne, C. D. L. (2021). The indispensable dog. *Frontiers in Psychology*, 12, 656529. https://doi.org/10.3389/fpsyg.2021.656529

Yin, W., & Bian, Z. (2015). The gene network underlying hypodontia. *Journal of Dental Research*, 94(7), 878–885. https://doi.org/10.1177/0022034515583999

Zhang, S. J., Wang, G. D., Ma, P., Zhang, L. L., Yin, T. T., Liu, Y. H., Otecko, N. O., Wang, M., Ma, Y. P., Wang, L., Mao, B., Savolainen, P., & Zhang, Y. P. (2020). Genomic regions under selection in the feralization of the dingoes. *Nature Communications*, 11(1), 671. https://doi.org/10.1038/s41467-020-14515-6

Zhou, C., Liu, Y., Zhao, G., Liu, Z., Chen, Q., Yue, B., Du, C., & Zhang, X. (2023). Comparative Analysis of Olfactory Receptor Repertoires Sheds Light on the Diet Adaptation of the Bamboo-Eating Giant Panda Based on the Chromosome-Level Genome. *Animals*, *13*(6), 979. https://doi.org/10.3390/ani13060979

Chapter 3 | Comparative Genomics of Local Adaptive Evolution in Eurasian Grey Wolves and Free-Ranging-Dogs

Contents

Abstract 83

1. Introduction **84**

Material and Methods **86**

2.1. Dataset 86

2.2. Identifying associations between environmental variables and genetic diversity 86

2.4. Gene ontology enrichment and protein association analysis 89

3. Results **89**

3.1. Genome wide SNP associations with WorldClim environmental variables 89

3.2. Genome wide SNP associations with mammalian species richness 91

3.3 Variant effect prediction and gene ontology analysis 93

3.3.1. Genes associated with climatic variables 93

3.3.2. Genes associated with mammalian species richness 102

3.3.3. Genes under both diversifying selection and local adaptation in FRDs 102

4. Discussion **102**

4.1. Associations between Genetic Diversity and Environmental Variables 102

4.2. Genes associated with environmental variables 105

4.2.1. Morphological, reproductive and cardiovascular adaptations 105

4.2.2. Energetic, metabolic and dietary adaptations 106

4.2.4. Immune system adaptations 107

4.2.5. Adaptations for prey/food detection 108

4.2.5. SLC genes in canid adaptations 109

4.3. Genes in FRDs with signatures of both diversifying selection and local adaptation 109

4.4. Common evolutionary patterns in grey wolves and FRDs 110

5. Conclusion **110**

References **112**

Abstract

Local adaptation is a process by which the populations' fitness increases in their specific environments. Recently diverged domesticated species and their wild relatives frequently experience comparable environmental conditions despite occupying different ecological niches. We may therefore expect similarities in their adaptive strategies under comparable environmental pressures, but our knowledge on this is limited. Dogs (Canis lupus familiaris) are among the most widespread mammals globally, with the majority being free-ranging (FRDs), not confined or owned by humans. Their wild relatives, grey wolves (Canis lupus), are distributed across the northern hemisphere, reflecting their adaptability to a wide range of biomes. To investigate how local selective pressures are impacting the evolutionary trajectory of canids we analysed single nucleotide polymorphism (SNP) genotypes from 232 Eurasian grey wolves and 370 FRDs across 10 biomes. Comparison among a wild and domestic canid allowed us to disentangle lineage specific patterns. Our findings revealed that precipitation and temperature gradients are driving local adaptation in both taxa, albeit in different ways. We detected 1,337 genes in wolves and 1,211 in FRDs that display signatures of local adaptation predominantly linked to metabolism, immune response, morphology, and sensory perception, albeit with minimal overlap among the two canids. Thus, local adaptive variation is unique to each taxa and not resulting from their recent shared origin. However, it impacts similar phenotypic traits. **Keywords**: Free-ranging dogs, grey wolves, adaptive evolution, species richness, molecular adaptations

1. Introduction

Local adaptation is an evolutionary process by which alleles favourable in their immediate (i.e., local) environments propagate and/or fixate in a population. Local adaptive evolution can be a precursor to speciation as it may lead to genetic divergence among populations (Wadgymar et al., 2022). Knowledge of the molecular mechanisms underlying these evolutionary processes is limited (reviewed in Chapter 1). Habitat differentiation is one of the drivers of local adaptation (Hendricks et al., 2018; Schweizer et al., 2016; Wadgymar et al., 2022), and phenotypic plasticity positively impacts this process (Espinosa-Soto et al., 2011, Gao et al., 2019). However, the specific mechanisms through which biotic and abiotic factors drive local adaptation remain elusive. Investigating this process is fundamental to deepening our understanding of the impact of environmental change, including climate change, upon genetic diversity. This is necessary for adequate management of both threatened and invasive species.

Large SNP genotype datasets, combined with novel statistical approaches, allow us to address a broad range of questions concerning the adaptive potential of populations under diverse environmental conditions. Gene-environment association (GEA) studies have become the standard for identification of the molecular mechanisms underlying adaptive traits in both plants and animals (Chapter 1, Dauphin et al., 2023; Lasky et al., 2023; Rocha et al., 2021). While these associations do not establish causality, they provide a foundational step for identifying potential causal relationships and elucidating phenotypes impacted by the variants (Gallagher and Chen-Plotkin, 2018; Lasky et al., 2023; Uffelman et al., 2021). Availability of environmental data from WorldClim (Fick & Hijmans, 2017); has further allowed the combined use of genomic and ecological data to delve into mechanisms underlying local adaptation (Chapter 1; Wadgymar et al., 2022). In this study we employed these methods to investigate local adaptive variation in two widespread sympatric carnivores.

Dogs (*Canis lupus familiaris*) and grey wolves (*Canis lupus*) are among the most ecologically successful canid lineages, widely distributed across the globe and subject to diverse environmental conditions. They are highly adaptable, closely related taxa with wide overlapping distributions (Gao et al., 2019; Hendricks et al., 2018). While dogs are the domesticated descendants of grey wolves, their evolutionary trajectories have diverged significantly due to human influence (Pierotti, 2024; Tancredi & Cardinali, 2023). The domestication of dogs from grey wolves is a well-documented process, yet it continues to reveal new insights into adaptations in canids (Bergström et al., 2022; Hecht et al., 2023; Tancredi & Cardinali, 2023). Originally stemming from a commensal relationship where early dogs scavenged human waste, this interaction has evolved into complex relationships ranging from pest status to home-living pets (Losey et al., 2022; Pierotti, 2024). In contrast, grey wolves have maintained a more natural evolutionary course in natural landscapes, less dependent on anthropogenic foods and consequently, less directly shaped by humans (Kuijper et al., 2016; Newsome et al., 2017). Wolves' ability to inhabit diverse biomes across the northern hemisphere—from arid deserts to tundra—underscores their

significant adaptive potential (Boitani et al., 2023; Schweizer et al., 2016). Dogs have an even broader, nearly global distribution facilitated by their cohabitation with humans (Chapter 2), making them arguably the most populous canids worldwide (Belsare & Vanak, 2020; Bryce, 2021).

An estimated 75-80% of the domestic dog population are free-ranging dogs (FRDs), (Belsare & Vanak 2020; Gompper, 2014), which are not confined by humans and not restricted with mate choice. FRDs rely on their own abilities to access resources, mainly anthropogenic foods, however they have also been documented to predate on livestock and wildlife (Capitain et al., 2024; Duarte et al., 2016; Hughes et al., 2013; Hussain et al., 2021; Lino et al., 2023; Wierzbowska et al., 2016). FRDs can interact and interbreed with pure-bred pet-dogs, which is facilitated by overlapping distributions in urban and rural habitats, thus enabling gene flow between these intraspecific groups (Belo et al., 2015; Pilot et al. 2015; Belsare & Vanak 2020). Such inter-breeding, although it's not very frequent, contributes to the genetic diversity of these dogs, as they navigate the spectrum from domesticity to a more independent existence (Belsare & Vanak 2020). Free-roaming populations, while reliant on anthropogenic food, are still subject to environmental pressures comparable to those faced by wild canids (Belo et al., 2015; Belsare & Vanak, 2020; Bryce, 2021; Hussain et al., 2021; Mahajan et al., 2024). Thus, grey wolves and FRDs are particularly well-suited for studying the genetic basis of local adaptation.

There is ample evidence for dogs' adaptations to human dominated landscapes. Behavioural, reproductive and foraging variation among dogs and wolves have been proposed as adaptations for their human commensal niche (Wynne, 2021). Enzymes involved in the digestion of anthropogenic foods (Freedman et al., 2016; Gao et al., 2019), specifically lactose (Liu et al., 2021) and starch have been detected to be under selection, the latter also showing an increase in copy number (Axelsson et al., 2013; Arendt et al. 2016; Freedman et al., 2016; Ollivier et al., 2016). These adaptations are also thought to be involved in the evolution of dogs as human commensals during the agricultural revolution. Signatures of selection have been detected in dogs in genes associated with morphology and coat colour in the context of artificial selection (Plassais et al., 2019) and local adaptation (Sinding et al., 2020; Li et al., 2023). Moreover, local adaptations in response to local pathogens (Liu et al., 2018) and high altitude (Wang et al., 2014) have also been detected in dogs.

Grey wolves' local adaptation has been investigated in more depth. Grey wolves' population genetic structure is shaped by differentiation of environmental conditions, even in the absence of physical barriers to gene flow (e.g. Pilot et al., 2006; Leonard et al. 2012; Schweizer et al., 2016). Additionally, signatures of selection in genes for coat coloration, metabolism, vision, and morphology (e.g., body size, skeletal) have been detected in European and North American populations (Pilot et al., 2014; Schweizer et al., 2015; Schweizer et al., 2016; Schweizer et al., 2018). North American wolves displayed signatures of selection in genes involved in the aforementioned phenotypic traits in association with temperature and precipitation variations (Schweizer et al., 2016). The shared evolutionary

history of both taxa, coupled with their remarkable adaptability, positions them as ideal subjects for investigating the selective pressures on populations of geographically widespread taxa across multiple biomes. This investigation aimed to determine whether FRDs and grey wolves in similar habitats shared molecular adaptations to reveal insights into their adaptive strategies. Our findings will reveal if wild and domestic species display similar adaptive solutions under comparable environmental conditions.

Our investigation focused on comparing the effects of environmental variables on the genetic diversity of FRD and grey wolf populations that occupy Eurasian biomes. Based on previous studies on dogs and grey wolves, we expected that adaptive genetic variations linked to environmental differentiation among biomes will be associated with phenotypic traits such as coat colour, body size, skull shape, immunity, and metabolism in both canids. Areas with a high human footprint index were hypothesized to show positive selection for heritable traits that enhance survival in anthropogenic landscapes in both taxa. We explored adaptations to local environmental conditions and provided insights into molecular adaptations of both taxa to environmental factors. Furthermore, we investigated if closely related wild and domestic taxa, showed signatures of selection within the same genes under similar environmental pressures. Our investigative approach is comprised of genome-environment association analysis, gene ontology (GO) analysis and variant effect prediction (VEP). VEP enabled us to infer putative phenotypic effects of genes near SNPs with signatures of adaptive variation. With this approach we were able to identify selective pressures driving intra and interspecific adaptative processes. Additionally, we identified both shared and lineage specific adaptations to local environmental conditions.

Material and Methods

2.1. Dataset

The quality-controlled 73360 SNPs genotypes datasets consisted of 232 grey wolves and 370 FRDs (Fig. 1, Table S1). Single nucleotide polymorphisms (SNPs) calling methodology and quality controls including hybrid detection are described in Chapter 2. The non-admixed individuals studied were sampled across Eurasia and grouped by biomes. Between both taxa there was representation from 10 different biomes from deserts and xeric shrublands to tundra. No FRDs samples were available from two of the biomes: flooded grasslands and savannas, and montane grasslands and shrublands (Fig. 1, Table S1).

2.2. Identifying associations between environmental variables and genetic diversity

We employed landscape genomic approaches using latent factor mixed models (LFMM) and redundancy analysis (RDA) to estimate associations between genetic variation and environmental variables. We applied this dual-modelling approach to mitigate the risk of misidentification of associations (Dauphin et al., 2023). We sourced twenty-five environmental variables from the WorldClim database at 30 arc-second resolution (Fick & Hijmans, 2017). Using the R package corrplot v0.92 (Wei & Simko, 2024), we assessed and

removed collinear variables based on a 70% threshold, retaining 14 predictors for wolves and 12 for dogs (Table 1). We used the R package terra v1.7-81 (Hijmans, 2024) to stack environmental variables and extract data points necessary for constructing the analysis matrices. To address missing values within our environmental matrix dataset, we manually imputed them by referencing the nearest neighbour based on GPS coordinates.

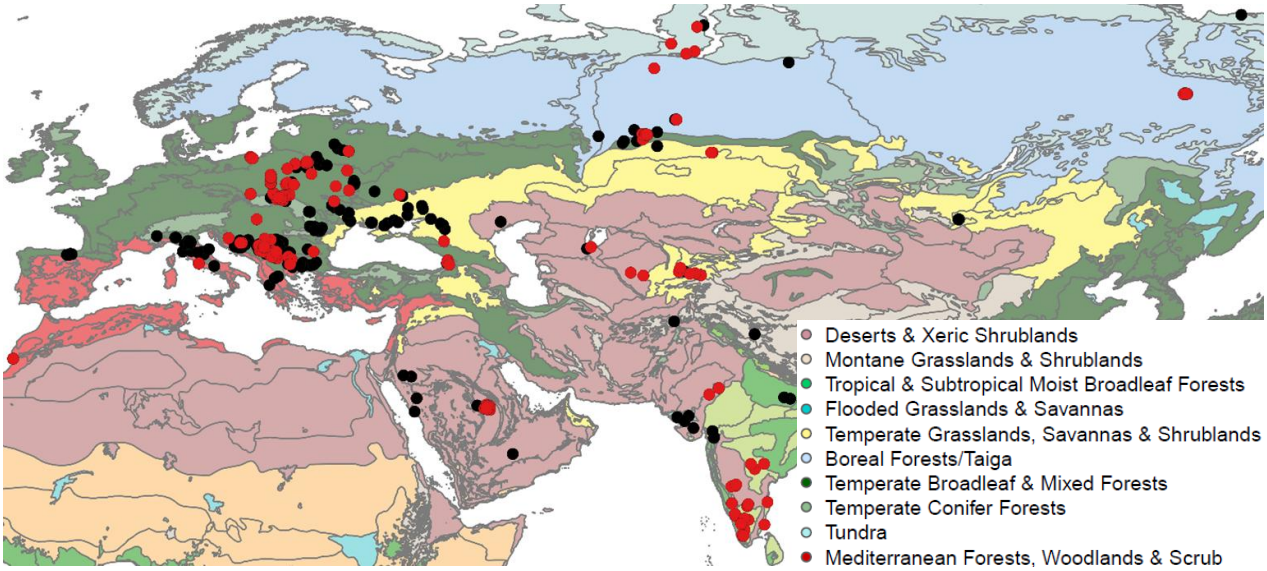


Figure 1. Biome map with the distribution of individuals included in the dataset. Each dot was plotted based on the GPS coordinates of the collected samples of free-ranging dogs (red) and wolves (black). The distribution of the dataset is mapped on the Resolve Ecoregions 2017 map, where each colour signifies a different biome (Dinnerstein et al., 2017).

Additionally, we generated mammalian species richness variables for gene-environment association studies, defining these layers based on taxonomic ranks—families for ungulates and orders for carnivores and other mammals. These variables were derived from spatial distribution data polygons of mammals, downloaded from the IUCN website, which rarely underrepresent species presence, but may overrepresent it at range edges (Chen et al., 2024). Four species richness layers were created: (1) large ungulates, (2) medium ungulates, (3) carnivores and (4) "other mammals" comprised of mammalian orders not included in the previous three layers (Table 1). This last category will be further referred to as "other mammals".

For the genomic analyses, we utilized the R package lfmm v1.1 (Proust-Lima et al., 2023), which applies latent factors mixed models to estimate genome-wide associations with environmental variables, correcting for population structure with predetermined K values (K=5 for dogs, K=8 for wolves). These values were derived from sNMF analyses (outlined in Chapter 2). Candidate SNPs were selected based on a q-value threshold of 0.05, derived from the calibrated p-values. Parallel to LFMM, we conducted RDA using the vegan package v2.6-6.1 (Oksane et al., 2024). Unlike LFMM, RDA does not adjust for population structure and was executed under default settings with a two-tailed p-value threshold of 0.0027 to

identify significant associations. Outliers were managed using the R package qvalue: Q-value estimation for false discovery rate control v2.40.0 (Storey et al., 2025), to reduce the risk of false positives. RDA analysis produces two types of results (1) associations between candidate SNPs and environmental variables (Fig. S9), and (2) ordinations in space of the associations between individual genotypes and the environmental variables, with the individuals being colour-coded by biomes (Fig. 5). For the triplot ordination plots the scaling =3 was used, which scales the SNP and individual scores by the square root of the eigenvalues. SNPs identified by both LFMM and RDA as displaying signatures of local adaptation were considered robust candidates for further investigation.

Table 1. List of WorldClim predictors and species richness layers used; The predictors retained after the collinearity analysis in the landscape genomic analysis.

Abrev	**Predictor**	**Retained FRDs**	**Retained W**
BIO2	Mean Diurnal Range (Mean of monthly (max temp - min temp))	Yes	Yes
BIO3	Isothermality (BIO2/BIO7) (×100)	-	Yes
BIO7	Temperature Annual Range (BIO5-BIO6)	Yes	Yes
BIO8	Mean Temperature of Wettest Quarter	Yes	Yes
BIO9	Mean Temperature of Driest Quarter	-	Yes
BIO13	Precipitation of Wettest Month	Yes	-
BIO14	Precipitation of Driest Month	Yes	-
BIO16	Precipitation of Wettest Quarter	-	Yes
BIO18	Precipitation of Warmest Quarter	-	Yes
BIO19	Precipitation of Coldest Quarter	Yes	Yes
Crops	Land used for crops	Yes	Yes
Elev	Elevation	Yes	Yes
Grass	Grassland	Yes	Yes
Shrub	Shrubland	Yes	Yes
Tree	Tree cover	Yes	Yes
Hfpi	Human footprint	Yes	Yes
LargUng	Large ungulates: species richness based on the distribution of the species in the families; Bovidae, Cervidae, Antilocapridae, Equidae and Tapiridae.	Yes	Yes
MediumUng	Medium ungulates: species richness based on the distribution of the species in the families; Tragulidae Moschidae, Suidae and Tayassuidae.	Yes	Yes
Carnivores	Carnivore: species richness based on the distribution of the species in the Carnivora order.	Yes	Yes
Other Mammals	Other mammals: species richness based on the distribution of the species from mammalian orders not included to the other categories; Dermoptera, Eulipotyphla Hyracoidea, Lagomorpha, Macroscelidea, Pholidota, Primates, Proboscidea, Rodentia, Scandentia and Tubulidentata.	Yes	Yes

2.4. Gene ontology enrichment and protein association analysis

We used gene ontology enrichment (GO) and protein association predictions to investigate genes with signals of local adaptation. For the GO enrichment analysis all the gene products without a UniProt code were automatically excluded: 33% genes (GEA -wolves), 31% (GEA-FRDs), 29% (wolves- mammalian species richness) and 39% (FRDs- mammalian species richness), which included RNA genes, pseudogenes and any protein-coding gene without the code (Table S3-S4). We performed the GO enrichment using g:Profiler (Kolberg, et al., 2023) with default settings, including a significance threshold of 0.05, and selected options to ignore ambiguous queries and chose IDs with the most GO annotations. The analysis was conducted using the *Canis lupus familiaris* (dog) and *Homo sapiens* (human) genome options. This approach enabled us to leverage the more comprehensive annotations available for the human genome, enriching our analysis with additional GO terms. Due to discrepancies in gene codes between species, UniProt identifiers were necessary to match dog genes with their human orthologs, as ENSCAFG codes do not align with human genomic data.

For protein association predictions, we employed STRING v12 (Szklarczyk et al., 2023), a software tool for building functional protein association networks. The analysis settings included a default false discovery rate below 0.05 and a minimum interaction strength and signal threshold of 0.01. We input ENSCAFG codes, permitting the inclusion of proteins lacking UniProt identifiers. Given the extensive number of genes analysed, STRING automatically activated settings to display interaction confidence on the network edges. We utilized the Markov Cluster Algorithm (MCL) clustering option to colour-code clusters and subsequently analysed the three largest clusters for each taxa individually to determine enriched GO terms for each sub-network.

3. Results

3.1. Genome wide SNP associations with WorldClim environmental variables

The use of dual analytical approaches allowed us to minimize false positives, though at the expense of excluding loci with weaker signals. Using latent factor mixed models (LFMM), we identified 2653 candidate SNPs for wolves and 1516 for dogs, applying a q-value threshold of 0.05 (Fig. 2A and 2B). These q-values were derived from calibrated p-values (Figs. S3-S6). Redundancy analysis (RDA) identified 1245 candidate SNPs for wolves and 1045 for FRDs, explaining 8% and 3% of the observed genomic variation in wolves and dogs, respectively (Figs. S7-S8). The spatial ordination of SNPs on the first three constrained axes revealed significant associations (two-tailed p-value threshold = 0.0027).

RDA analysis produces separate results of associations between environmental variables, for candidate SNPs and individual genotypes. The former allows us to identify environmental pressures operating the as strongest selective pressures i.e., drivers of adaptive variation. This is achieved by exploring the environmental variables vectors strongly correlated with the candidate SNPs on each RDA axis. Based on the first RDA axis

in both FRDs and grey wolves the main driver of adaptive variation in candidate SNPs was mean temperature of the wettest quarter (Bio8). With low values driving adaptive genetic differentiation between populations from colder vs warmer climate zones. Additionally, low values of precipitation of coldest quarter (Bio19) were a close second driver in wolves, based on this axis. These associations imply that the candidate SNPs are involved in adaptations to colder and rainier habitats (Fig. S9). Additionally, the main drivers of genetic variation among RDA candidate SNPs based on the second and third RDA axes for wolves were lower annual range of temperature (Bio7) and higher elevation, respectively (Fig. S9). These correlations imply that the candidate SNPs are involved in adaptations to more stable temperature regimes and higher altitudes. In FRDs, the second driver of genetic variation among RDA candidate SNPs was low precipitation of the wettest month (Bio13) based on the second and third RDA axes (Fig. S9). These associations imply adaptations in FRD populations for both wet cool environments (Bio8) and dry and hot habitats (Bio13).

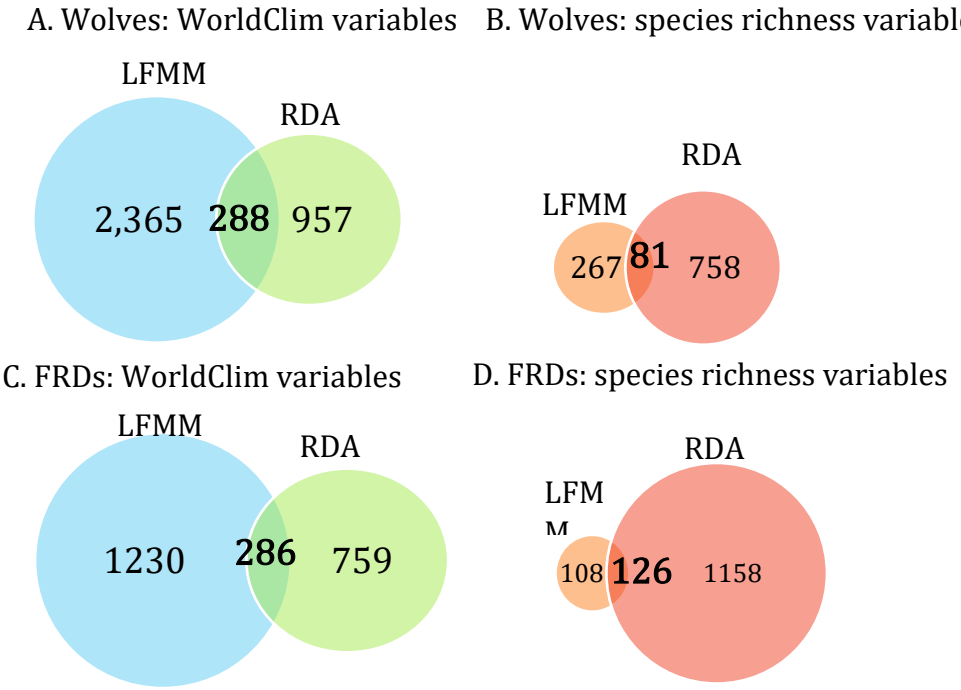


Figure 2. Venn diagrams of landscape genomic analysis. Venn diagrams for candidate SNPs identified with both LFMM and RDA using the WorldClim variables (A, C) and the mammalian species richness variables (B-D) for both wolves (A-B) and FRDs (C-D). The final list of SNPs displaying signatures of local adaptation used for subsequent analysis was comprised of the SNPs shared between the candidate lists of both LFMM and RDA.

The RDA results also showed associations between environmental variables and individual genotypes of candidate SNPs across various biomes in both taxa. Individual genotypes of wolves from most biomes were associated with high elevation, grass cover, and annual temperature range (Bio7). Individual genotypes of wolves from boreal forests/taiga and flooded grasslands and savanna were correlated with high precipitation of the coldest quarter (Bio19) (Fig. 3A, Fig. S17). For FRDs, individual genotypes from temperate

broadleaved and mixed forests showed associations with high tree cover and precipitation of the driest month (Bio14) (Fig. 3C, Fig. S17). Individual genotypes of FRDs from boreal forests/taiga and most FRDs from the desert and xeric shrublands biomes were correlated with high Bio7. Individual genotypes of FRDs from temperate grasslands, savannas, and shrublands were positively correlated with high human footprint, Bio19, and cover of trees and shrubs. Notably, genetic differentiation of wolves and FRDs from multiple biomes was associated with high annual temperature range (Bio7) (Fig. 3A, 3C).

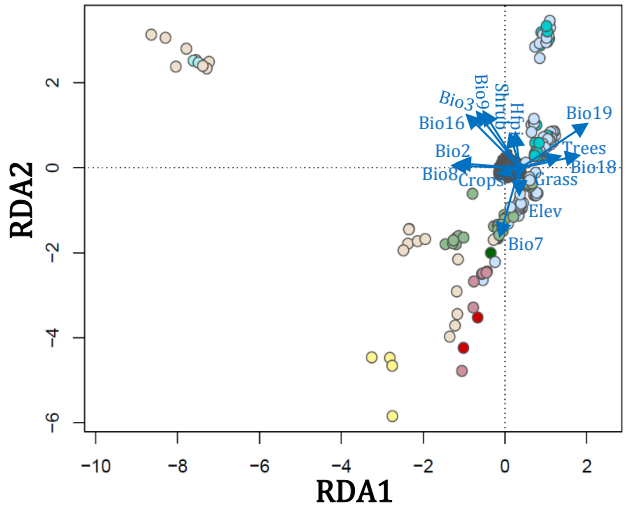
288 SNPs in wolves and 286 in FRDs appeared on both LFMM and RDA candidate lists; 63% of those in wolves and 55% in dogs were primarily associated with Bio16 (precipitation of the wettest quarter) and Bio13 (precipitation of wettest month), respectively (Fig. 4). Notably, only three SNPs (i.e., chromosome (chr) 8, 35744827 base pair (bp); chr 9, 1299833 bp and chr 30, 22931462 bp) displayed signatures of local adaptation common to both species (note that SNP coordinates are based on the CanFam 3 genome). Two of the shared SNPs are located within SIX4 and RPTOR genes, respectively, all three SNPs were located near SIX1, NPTX1 and POLR2M (Tables 2-3, Tables S5-S6).

3.2. Genome wide SNP associations with mammalian species richness

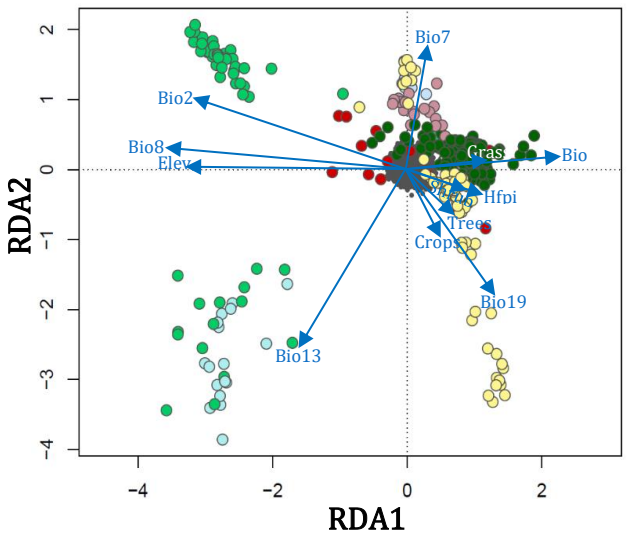
Species richness of four mammalian groups served as environmental variables, with SNP variants identified as being associated with these variables in both wolves and FRDs. Using the same methods as with the WorldClim variables, LFMM analysis identified 348 candidate SNPs for wolves and 234 for FRDs (Fig. 2C-D). RDA identified 839 candidates for wolves and 1,284 for FRDs based on spatial ordination of SNP locations on the first three constrained axes (Figs. S7-S8). The RDA analysis revealed an adjusted R^2 of 0.06 for wolves and 0.02 for FRDs, indicating moderate explanatory power of species richness on SNP variation.

In total, 81 outlier SNPs were identified in wolf populations and 126 in FRD populations using both approaches, with no overlap between the two canids. In wolves, ungulates' species richness accounted for 79% of the associations in LFMM, while “other mammals” and carnivores had the highest proportion of associations in RDA at 57%. Individual genotypes of wolves from most biomes were positively correlated with higher species richness of ungulates, except for individual genotypes of wolves from boreal forests/taiga, flooded grasslands and savanna (Fig. 3B, Fig. S17). For FRDs, species richness of “other mammals” and carnivores accounted for most observed SNP variant associations, with 81% and 76% for LFMM and RDA, respectively. Individual SNP genotypes of FRDs’ from most biomes were negatively correlated with high species richness of non-ungulate mammals (Fig. 3D). Conversely, individual genotypes of FRDs from temperate grasslands, savannas, and shrublands showed a positive correlation with high species richness of medium ungulates on the first two axes (Fig. 3D, Fig. S17). Individual genotypes of dogs from temperate broadleaved and mixed forests were associated with high diversity of non-ungulate mammals.

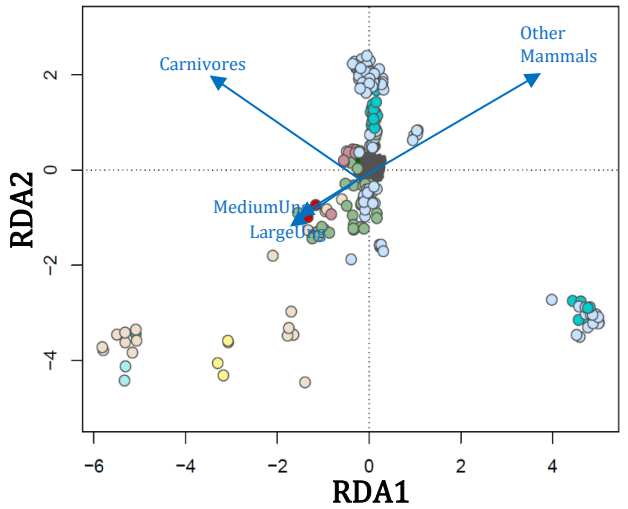
A. Wolves WorldClim variables RDA axes 1 and 2



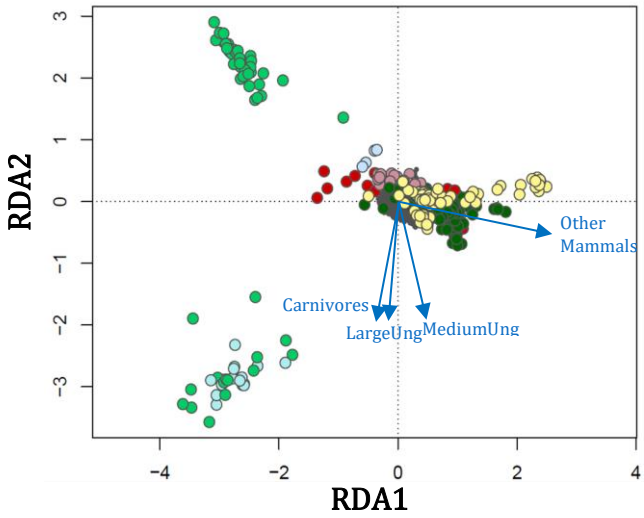
C. FRDs WorldClim variables RDA axes 1 and 2



B. Wolves species richness RDA axes 1 and 2



D. FRDs species richness RDA axes 1 and 2



- Deserts & Xeric Shrublands
- Montane Grasslands & Shrublands
- Tropical & Subtropical Moist Broadleaf Forests
- Flooded Grasslands & Savannas
- Temperate Grasslands, Savannas & Shrublands
- Boreal Forests/Taiga
- Temperate Broadleaf & Mixed Forests
- Temperate Conifer Forests

Figure 3. RDA ordination projection plots for candidate SNPs display signatures of local adaptive variation and environmental vectors. The output for RDA analysis shows the ordination in space of individual genotypes of wolves (A-B) and dogs (C-D) for the first two axes for both the WorldClim (A, C) and species

richness variables (B, D). Individuals are coloured by biome, SNPs are in grey, and the blue vectors are the environmental predictors.

3.3 Variant effect prediction and gene ontology analysis

VEP analysis identified 1413 genes with or near SNPs displaying signatures of local adaptation in wolves and 1537 in FRDs. Specifically, 343 genes in wolves and 502 in FRDs were associated with species richness and the remainder with WorldClim variables, with 78 and 283 genes, respectively, appearing on both lists (Table S5 for wolves and Table S6 for FRDs). The genes associated with WorldClim variables were enriched for 25 GO terms in wolves and 20 GO terms in FRDs, including "cytoplasm" and "organonitrogen compound metabolic process" for both taxa (Table S7, Fig. S10). The genes associated with species richness were enriched for 5 GO terms in wolves and 13 in FRDs, with no common terms (Table S7, Fig. S10). GO terms for genes with signatures of adaptive variation in both grey wolves and FRDs were related to muscular, hepatic and olfactory systems (Tables S5-S7).

3.3.1. Genes associated with climatic variables

STRING networks for the genes associated with WorldClim variables comprised 693 nodes (i.e., gene products) in wolves and 684 in FRDs, respectively. The wolves' STRING network showed the strongest functional enrichment in GO terms related to electron transfer, lipid and oxygen metabolism (Fig. S11). Additionally, significant enrichment was noted in KEGG pathways, including metabolism of xenobiotics, carcinogenic chemicals and drugs (Fig. S11). The top three wolves' STRING sub-networks displayed pronounced functional enrichment for various biological processes including sensory perception of chemical stimulus (Fig. S18). In contrast, the FRDs' STRING network for genes associated with WorldClim variables showed no functional enrichment for any GO term (Fig. S13). Nevertheless, the sub-networks revealed significant enrichment of various terms including these related to environmental stimulus detection, metabolic and immune processes (Fig. 13).

In grey wolves, highest-ranking SNPs were located near genes with varied functions including metabolic, immune and reproductive functions (Table 2, Table S5). Additionally, missense mutations predicted to impact protein function were also detected in an immune system gene, IL12B (Table 2, Fig. S14). In FRDs highest ranking SNPs detected with both approaches were located near genes with functions related to morphology, immune system and DNA regulation (Table 2, Table S6). In FRDs, five missense mutations predicted to be deleterious were identified in D2HGDH, HSPA14, MRRF, SLC6A9 genes, and a gene that codes for an uncharacterized protein (ENSCAFG00000049094) (Table 2, Fig. S14).

Table 2. Genes near highest ranking SNPs and with non-tolerated missense SNPs variants

Wolves' highest-ranking genes			
Gene name	**UnitProt Code**	**Gene and variant details**	**Approach information**
RNA Polymerase II Subunit M	POLR2M	Encodes a subunit of RNA polymerase, implicated in cardiomyopathy (Heliö et al., 2021). Also detected in FRDs, though the variant ranked lower	Ranked highest with LFMM associated variables Bio3, Bio8, Bio16, and Bio19
Acetylglucosaminyltransferase	MGAT4A	Plays a role in carbohydrate metabolism, particularly in glucose-stimulated insulin secretion (Osada et al., 2024).	Ranked second highest with LFMM associated with Bio2, Bio3, Bio8, Bio16, and Elevation. Ranked highest in RDA with mammalian species richness associated with "other mammals".
Solute Carrier Family 6 Member 6	SLC6A6	Plays a role in taurine transportation in the liver, an essential amino acid for various metabolic processes (Anderson et al., 2009b).	Ranked third highest with LFMM associated with Bio16.
Glyceraldehyde-3-Phosphate Dehydrogenase, Spermatogenic	GAPDHS	Encodes a catalytic enzyme crucial for spermatogenesis and sperm motility, essential for reproductive fitness (Margaryan et al., 2015). Besides the intron variant, two additional variants with signatures of local adaptation were detected near this gene for both WorldClim and species richness variables.	Ranked highest with RDA associated with Bio8. Ranked highest in species richness LFMM and associated with medium ungulates.
Beta-defensin 122	CBD122	Antimicrobial peptide, with differential transcription detected in dogs with skin conditions, suggesting a role in skin health and disease resistance (Lancto et al., 2013)	Ranked second highest with RDA and associated Bio16. Ranked second highest with LFMM in species richness associated with "other mammals"' species richness.
DNA Nucleotidylexotransferase	DNTT	Encodes terminal transferase, a template-independent DNA polymerase that catalyses primer synthesis. It is associated with GO terms related to DNA repair and modification.	Ranked third highest in RDA associated Bio3. Ranked third highest in species richness with LFMM and associated with medium ungulates species richness.
Ubiquitin-specific protease 36	USP36	Codes for a ubiquitin specific peptidase which has GO terms related to protein translation.	Ranked second highest in species richness RDA and associated with carnivores' species richness.
Methylenetetrahydrofolate dehydrogenase/cyclohydrolase	MTHFD2	Encodes a nuclear-encoded mitochondrial enzyme with methylenetetrahydrofolate dehydrogenase and methenyltetrahydrofolate cyclohydrolase functions.	Ranked third highest in species richness RDA and associated with other mammals' species richness.
Wolves' non-tolerated missense mutations			
Interleukin 12 subunit beta	IL12B	Notable for its role as a toll-receptor that mediates the host response to artic rabies, a finding that has also been observed in Arctic foxes (Baecklund et al.,	Missense mutation (G reference allele C), resulting in S18G amino acid change (SIFT 0.01). Associated with elevation in LFMM and Bio8 in RDA. The

		2021). Another variant with signatures of selection was detected upstream of this gene.	variant is located within the amino acids preceding the Ig-like C2-type domain, which is crucial for the protein's role in cell adhesion and signal transduction.
SSD domain-containing protein	ENSCAFG00000013789	Encodes a transmembrane protein with a sterol sensing domain with and has no UnitProt code. This family of proteins is involved in cholesterol metabolism, transport and signalling (Wu et al., 2022).	Missense mutation (T reference allele C), resulting in T408I amino acid change (SIFT 0). Associated with elevation in LFMM and Bio19 in RDA.
Calpain 13	CAPN13	Encodes for a calcium-activated neutral protease expressed in the stomach and small intestine (Litosh et al., 2017).	Missense mutation (A reference allele C), resulting in A518S amino acid change (SIFT 0.03). Associated with other mammals' species richness in both approaches. The mutation occurs 20 amino acids before the EF-hand, the calcium-binding domain of the protein.
Free-ranging dogs' highest-ranking genes			
Matrix Metallopeptidase 17	MMP17	Encodes for a peptidase that plays a significant role degradation of the extracellular matrix particularly in the vascular smooth muscle wall (Martine-Alonso et al., 2015; Xiao et al., 2022)	Ranked highest both LFMM and RDA associated with Bio13 and Bio8 respectively
Integrin Subunit Beta 6	ITGB6	Encodes for an integrin associated with fibrosis and cancer, important for maintaining tissue integrity and (Krafft et al., 2014; Zhang et al., 2024).	Ranked second highest with LFMM associated with Bio8
PX Domain Containing Serine/Threonine Kinase Like	PXK	Encodes for a phox domain-containing protein, involved in regulating growth factor receptor localization (Takeuchi et al., 2010).	Ranked third highest with LFMM associated with Bio13
Interphotoreceptor Matrix Proteoglycan 1	IMPG1	Encodes for a component which surrounds photoreceptor neurons with similar function to IMPG2, which has been linked to progressive retinal atrophies in dogs, highlighting a critical role in canine vision health (Mitchell et al., 2022; Hitti-Malin et al., 2020).	Ranked second highest with RDA associated with Bio13.
URB1 Ribosome Biogenesis Homolog	URB1	Protein involved in ribosome biogenesis, has been shown to be involved in adequate organ development in vertebrates (He et al., 2017).	Ranked third highest with RDA crops cover, missense variant.
Histone methyl-lysine binding protein 3	L3MBTL3*	Encodes for a chromatin interacting transcriptional repressor suggested to be involved with immune response in pure-breed dogs (Álvarez et al., 2022)	Ranked highest in species richness with LFMM associated with non-ungulate mammals and carnivores' species richness.
p21-activated kinase 4	PAK4*	Encodes for a serine/threonine p21-activating kinase	Ranked second highest in species richness with

		suggested to be involved in bone development (Choi et al., 2015)	LFMM associated with non-ungulate mammals and carnivores' species richness.
Ankyrin repeat domain-containing protein 26-like	LOC607646*	Uncharacterized protein.	Ranked third highest in species richness with LFMM associated with non-ungulate mammals and carnivores' species richness.
Solute Carrier Family 26 Member 11	SLC26A11*	Encodes for a solute carrier involved in anion exchange	Ranked highest in species richness with RDA associated with non-ungulate mammals' species richness.
Inhibitor Of DNA Binding 1	ID1*	Encodes for a helix-loop-helix transcription factor can inhibit DNA binding and transcriptional activation.	Ranked second highest in species richness with RDA associated with large ungulates' species richness.
Prune Homolog 2 With BCH Domain	PRUNE2*	Encodes for a protein part of the B-cell CLL/lymphoma 2 and adenovirus E1B 19 kDa interacting family and plays role in cellular processes.	Ranked third highest in species richness with RDA associated with large ungulates' species richness.
Free ranging dogs' non-tolerated missense mutations			
D-2-hydroxyglutarate dehydrogenase	D2HGDH	Encodes for a protein that modulates mitochondrial NADP-dependent isocitrate dehydrogenase enzyme (IDH2), which in turn is involved in mitochondrial energy production functions (Lin et al., 2015). The variant under selection lay within its flavin adenine dinucleotide (FAD)-binding p-cresol methylhydroxylase (PCMH)-type domain, a critical region for its catalytic activity involving the cofactor FAD. An upstream SNP variant of this gene was detected to be under selection in wolves.	Missense mutation (T reference allele C), resulting in T408I amino acid change (SIFT 0.03). Associated with elevation in LFMM and Bio19 in RDA.
heat shock protein family A (Hsp70) member 14	HSPA14	Encodes for a protein involved in response to temperature fluctuations, this family of proteins has been associated with heatstroke response in dogs (Bruchim et al., 2017).	Missense mutation (A reference allele G), resulting in T325I amino acid change (SIFT 0.01). Associated with Bio19 in LFMM and Bio7 in RDA.
Mitochondrial ribosome recycling factor	MRRF	Encodes for a component of the mitochondrial translational machinery which has been found to be upregulated in dogs on a raw meat diet (Anturaniemi et al., 2020). The variant is predicted to impact its functions across both isoforms	Missense mutation (T reference allele C), resulting in A158T amino acid change (SIFT 0.04). Associated with elevation in LFMM and Bio8 in RDA.
Solute carrier family 6 member 9	SLC6A9	Encodes for a solute carrier which plays a role in the hepatic system specifically in glycine uptake (Bankir et al., 2024; Bruchim et al., 2017; Lin et al., 2015)	Missense mutation (A reference allele G), resulting in G111S amino acid change (SIFT 0.01). Associated with Bio14 in LFMM and elevation in RDA.

Detected in both species			
RNA Polymerase II Subunit M	POL2RM	See above.	Downstream variant, associated with Bio3, Bio8, Bio16, Bio19 and Bio2 in wolves and Bio13 in FRDs
Six Homeobox 1	SIX1	Encodes for a transcription factor, involved in embryonic development of kidney and nervous systems in particular	Upstream variant, associated with elevation and Bio16 in wolves and Bio19 and Bio8 in FRDs
Six Homeobox 4	SIX4	Encodes for a transcription factor, required for muscle and gonad development (Chen et al., 2021)	Missense mutation associated with elevation and Bio16 in wolves and Bio19 and Bio8 in FRDs.
Neuronal Pentraxin	NPTX1	Encodes for a neuron specific pentraxin, involved in pro-apoptotic signalling cascade. It has been suggested to be involved in mitochondrial dysfunction in response to hypoxia or excitatory toxicity (Chew et al., 2020).	Upstream variant associated with elevation and Bio8 in wolves and Bio13 and shrubs cover in FRDs.
Regulatory associated protein of MTOR complex 1	RPTOR	Encodes for a component of a signalling pathway involved in cellular growth in response to nutrient and insulin levels (Sun et al., 2010).	Intron variant associated with elevation and Bio8 in wolves and Bio13 and shrubs cover in FRDs
Cluster of Differentiation	CD70	Encodes for a cytokine member of the tumour necrosis factor family and involved in macrophage infiltration (Ge et al., 2017)	Downstream variant associated with medium ungulate and carnivores' species richness in both species.
Beta-defensin 123	DEFB123	Encodes for a defensin involved in immunologic response to invading microorganisms.	Downstream variant associated with ungulate and carnivores' species richness in wolves and non-ungulate mammal and large ungulates' species.
Histocompatibility Minor 13	HM13	Encodes for an enzyme involved in proteolysis of some signal peptides which has been associated with cellular response to stimuli	Downstream variant associated with large ungulate, non-ungulate and carnivores' species richness in wolves and non-ungulate mammal and large ungulates' species.
Inhibitor of DNA binding 1	ID1	Encodes for a helix-loop-helix DNA inhibitor transcription factor involved in cellular response to stimuli	Downstream variant associated with large ungulate, non-ungulate and carnivores' species richness in wolves and non-ungulate mammal and large ungulates' species.
GET-binding protein REM1	REM1	Encodes for a GTPase expressed in endothelial cells involved in cellular morphological changes	Downstream variant associated with large ungulate, non-ungulate and carnivores' species richness in wolves and non-ungulate mammal and large ungulates' species.

Non-tolerated: proteins predicted by SIFT analysis to not be tolerated were the protein to keep its current function. *Gene associated with mammalian species richness, **Gene associated with both mammalian species richness and WorldClim.

Table 3. *Genes involved in local adaptation detected in this study which have been detected in previous studies*

Genes	Consequence	Details	LFMM Variable	RDA Variable	Previous association
Wolves					
ATRN	Upstream	Encodes for attractin of which there are membrane bound isoforms associated with body wight and coat color in mice (Kaelin et al., 2008). Additional GO terms include cerebellum development and signalling receptor activity.	LargeUng	Mammals	Local adaptive evolution in North American gray wolves (Schweizer et al., 2016).
CBD103	Upstream	Encodes for canine beta defensin 3, which has been implicated in both innate immunity and black coat coloration in wolves (Anderson et al., 2009; Candille et al., 2010; Hrckova Turnova et al., 2022; Schweizer et al., 2018).	MediumUng	LargeUng	Local adaptive evolution in North American gray wolves (Schweizer et al., 2016).
CRIM1	Intron	Encodes for cysteine rich transmembrane BMP regulator 1, has GO terms related to skeletal development and morphology.	Elev	Bio16	Local adaptive evolution in North American gray wolves (Schweizer et al., 2016).
FGF4	Downstream	Encodes for fibroblast growth factor 4 and has multiple GO terms related to morpho-skeletal and neuronal development.	Elev	Bio16	Overall body size (Bannsch et al., 2020). Standard breed height in pure-bred dogs (Plassais et al., 2019).
GAP43	Upstream	Encodes for neuromodulin and has GO terms related to neuronal and morphological development.	Bio16	Bio18	Local adaptive evolution in North American gray wolves (Schweizer et al., 2016).
GAPDHS	Intron	See table 2.	Elev/ Grass LargeUng/ Medium Ung	Bio8 MediumUng	Under diversifying selection between FRDs and pure-bred dogs (Pilot et al., 2016).
HDAC2	Upstream	Encodes for a histone deacetylase which has GO terms related to gene transcription and cell growth.	LargeUng	Carnivores	Local adaptive evolution in North American gray wolves (Schweizer et al., 2016).
HMGCL	Downstream	Encodes for a mitochondrial enzyme involved in leucine degradation.	Elev	Bio16	Detected as an adaptation to ingest fruit fiber in manned wolves (Chavez et al., 2022).
LRP6	Synonymous	Encodes for low-density lipoprotein receptor-related protein which has multiple GO terms related to Wnt signalling pathway and cellular processes. It has been implicated in the cell entry receptor in canine distemper viral infections	Bio3, Bio8, Bio16, Bio19, Hfpi	Bio9	Local adaptive evolution in North American gray wolves (Schweizer et al., 2016).

		(Gradauskaite et al., 2023).			
NRG1	Intron	Encodes for Neuregulin 1 which has GO terms related to nervous system development.	Bio16	Bio9	Local adaptive evolution in North American gray wolves (Schwizer et al., 2016).
PLD1	Intron	Encodes for a phospholipase which has GO terms related to lipid metabolism and other cellular processes (e.g., response to nutrient, translation and signal transduction).	Bio7	Bio8	Adaptive introgression in Eurasian gray wolves (Pilot et al., 2021)
SOX5	Intron, Upstream	Encodes for SRY-box transcription factor 5 which has GO terms for gene regulation and cell differentiation.	Elev	Bio3	Associated with coat coloration in Chinese dog breeds (Li et al., 2023).
TBX15	Downstream	Encodes for T-box transcription factor 15 which has GO terms related to gene regulation.	Bio16	Bio2	Associated with coat coloration in Chinese dog breeds (Li et al., 2023).
THBS1	Downstream	Encodes for Thrombospondin 1 which has numerous GO terms related to immune system, cellular signalling, response to glucose and xenobiotic stimulus, cellular and muscular development among others.	Elev	Bio16	Adaptive introgression in Eurasian gray wolves (Pilot et al., 2021)
TRPM8	Downstream	Encodes for a transient receptor potential cation channel which has GO terms related to responses to cold temperatures.	Bio2, Bio3, Bio7	Bio19	Local adaptive variation in European gray wolves (Stronen et al., 2015).
TRPV1	Downstream	Encodes for a sensory neuron-specific cation channel capable of integrating various noxious chemical and physical stimuli characterized in dogs (Phelps et al., 2005).	MediumUng	Carnivores	Local adaptive variation in European gray wolves (Stronen et al., 2015).
Free Ranging Dogs					
APOB	Downstream	Encodes for Apolipoprotein B which has GO terms related to lipid metabolism, spermatogenesis, embryonic development, immune response, among others.	Bio13	Bio19	Adaptive correlation with precipitation seasonality and vegetation in North American gray wolves (Schwizer et al., 2016b).
CDH23	Downstream	Encodes for Cadherin related 23 which has GO terms related to sensory perception, stimulus response (e.g., perception of light, sound, equilibrium), locomotory behaviour among others. Associated with coat- colour- related deafness in pure-bred dogs (Abitol et al., 2022).	Bio7	Bio8	Local adaptive evolution in North American gray wolves (Schweizer et al., 2016).
FASN	Downstream	Encodes for fatty acid synthase which has GO terms related to lipid metabolism.	Elev	Bio13	Involved in dog domestication (Wang at el., 2019).

GAA	Downstream	Encodes for α-glucosidase an enzyme involved in degradation of glycogen to glucose in lysosomes. In pure-breed dogs it has been shown to have an essential role in glycogen degradation in cerebral, myocardia, liver and smooth muscle tissue (Seppälä et al., 2013).	Mammals	Mammals	Adaptive introgression in Eurasian gray wolves (Pilot et al., 2021)
GABRA5	Downstream	Encodes for a subunit of the inhibitory neurotransmitter GABBA receptor and been associated with anxiety and fear response (Li et al., 2014).	Bio14, Carnivores	Bio14, LargeUng	Local adaptive evolution in North American gray wolves (Schweizer et al., 2016). Signatures of selection detected in dogs (Li et al., 2014)
GHR	Downstream	Encodes for growth hormone receptor and has GO terms related to cell signalling.	Bio19	Bio13	Pure-bred dog height and weight (Bannasch et al., 2020; Schlamp et al., 2016). Small-breed height in pure-bred dogs (Plassais et al., 2019).
GOLT1A	Upstream	Encodes for a Golgi apparatus protein with annotations of lipase activity	Bio13	Bio13	Signatures of positive selection have been detected in manned wolves (Chavez et al., 2022)
HDAC4	Intron	Encodes for a histone deacetylase which has GO terms related to gene transcription, skeletal and muscle cell growth.	Bio13	Bio13	Local adaptive evolution in North American gray wolves (Schweizer et al., 2016).
IGF2BP2	Intron	Encodes for a binding protein which regulates translation of insulin like growth factor 2.	Bio14, Mammals, Carnivores	Bio19, Mammals	Local adaptive evolution in North American gray wolves (Schweizer et al., 2016). Associated with body weight and size in pure-bred dogs (Bannasch et al., 2020; Wade, 2025).
MITF	Downstream	Encodes for a transcription factor involved in regulation of melanocyte development and cell pigment transcription. It has GO terms related to DNA transcription, WNT signalling and bone development.	Bio14, Mammals, Carnivores	Bio7, Carnivores	Coat-color in breed dogs, white chest and head (Plassais et al., 2019).
MREG	Downstream	Encodes for melanoregulin which has GO terms related to pigmentation.	Bio19	Bio19	Local adaptive evolution in North American gray wolves (Schweizer et al., 2016).
OR13J1	Upstream	Encodes for an olfactory receptor and has GO terms related to olfactory detection.	Bio19	Bio19	Domestication and feralization in dingo (Zhang et al., 2020).

TBCE	Downstream	Encodes for Cofactor E, it is involved in beta-tubulin folding pathway and has GO terms related to locomotory behaviour, growth and cell division.	Bio19	Bio13	Local adaptive evolution in North American gray wolves (Schweizer et al., 2016).
TRHDE	Intron	Encodes for a peptidase involved in degradation of neuropeptide thyrotropin-releasing hormone. Plays a role in metabolism and has been associated with ungulate adaptations to arid environments (Kim et al., 2016).	Bio13, Carnivores	Bio13, MediumUng	Domestication and feralization in dingo (Zhang et al., 2020).
TYR	Downstream	Encodes for tyrosinase involved in the conversion of tyrosine to melanin and it has GO terms related to pigmentation.	Bio13	Bio13	Assocaition with precipitation seasonality and vegetation in North American gray wolves (Schwizer et al., 2016b).
Wolves and Free ranging dogs					
GLI2	Intron	Encodes for a Zinc-finger transcription factor and has GO roles related to gene regulation and bone, cochlear and neuronal cell development.	Bio16/Bio13	Bio19	Local adaptive evolution in North American gray wolves (Schweizer et al., 2018).
RPTOR	Intron	Encodes for a component of a signalling pathway involved in cellular growth in response to nutrient and insulin levels (Sun et al., 2010).	Elev/Bio13	Bio8/Shrubs	Adaptive introgression in Eurasian gray wolves (Pilot et al., 2021) and local adaptive variation in European gray wolves (Stronen et al., 2015).

3.3.2. Genes associated with mammalian species richness

STRING networks for the genes associated with species richness variables comprised 243 nodes in wolves and 313 nodes in wolves and FRDs, respectively. Five genes displayed association with species richness in both species: CD70, HM13, ID1, REM1, DEFB123 at different SNPs in each (Table 2, Tables S5-S6). The wolves' STRING network showcases strong functional enrichment for various catalytic activities (Fig S12). The notable sub-network was enriched for immune system GO terms and chemoattractant activity (Sig S18). For FRDs, the STRING functional enrichment was observed for the GO term "collagen catabolic process" (Fig. S14). The two notable sub-networks were enriched for GO terms related to cardiac functions, solute transport, olfactory receptor activity (Fig. S18), and the canonical Wnt signalling pathway (Fig. S18). A notable missense mutation in Calpain 13 (CAPN13) detected in wolves (Table 2), was associated with species richness of non-ungulate mammals (Fig. S15; Table S5). There was no missense mutation predicted to impact current protein function detected in FRDs.

3.3.3. Genes under both diversifying selection and local adaptation in FRDs

A total of 52 SNPs in FRDs were both detected to be under both diversifying selection in Chapter 2 and with signatures of local adaptation. Based on LFMM 49% of these SNPs were associated with precipitation of driest moth (Bio14), precipitation of the coldest quarter (Bio19) and human footprint. Based on RDA, 45% of these SNPs were associated with precipitation of the wettest month (Bio13). 311 genes located near these 52 SNPs were found to be driving local adaptation in FRDs and their annotated terms were enriched for terms related to sensory perception (i.e., olfaction and taste) and the immune system (Fig. 4, Fig. S19). In wolves there were no SNPs overlapping between the analyses of local adaptive variation in this chapter and diversifying selection in Chapter 2. However, there were 15 genes common in both lists. This number of genes was not sufficient for subsequent GO analysis.

4. Discussion

Categorizing individuals according to biome facilitated cross-taxa comparisons under comparable environmental conditions, acknowledging that their sampling collection distributions do not completely overlap due to the distinct ecological niches each taxa occupies. Our gene ontology enrichment analysis utilized two methods to provide a focused assessment that highlighted similarities between the taxa. The findings support our hypothesis that adaptive genetic variation linked to the biomes affects phenotypic traits related to body size, coat-coloration, skull shape, immunity, and metabolism (Tables 2-3, Tables S5-S6). However, populations of grey wolves and FRDs in similar habitats are subjected to unique selective pressures determined by their different niches, which is reflected in the pattern of taxa specific genes displaying signatures of local adaptation.

4.1. Associations between Genetic Diversity and Environmental Variables

In our study, the number of SNP markers displaying signatures of local adaptation in response to WorldClim variables was similar between grey wolves and FRDs. However,

the minimal overlap of genes associated with these variables between the taxa suggests that the observed patterns of local adaptation are unique to each taxa rather than the result of recent shared evolutionary origin. Environmental variables like precipitation and temperature from WorldClim acted as stronger selective pressures than species richness, with most SNPs showing selection signatures associated with these factors (Fig. 3, Fig. S17). This pattern aligns with findings in North American grey wolves, where similar environmental variables influenced local adaptation (Schweizer et al., 2016).

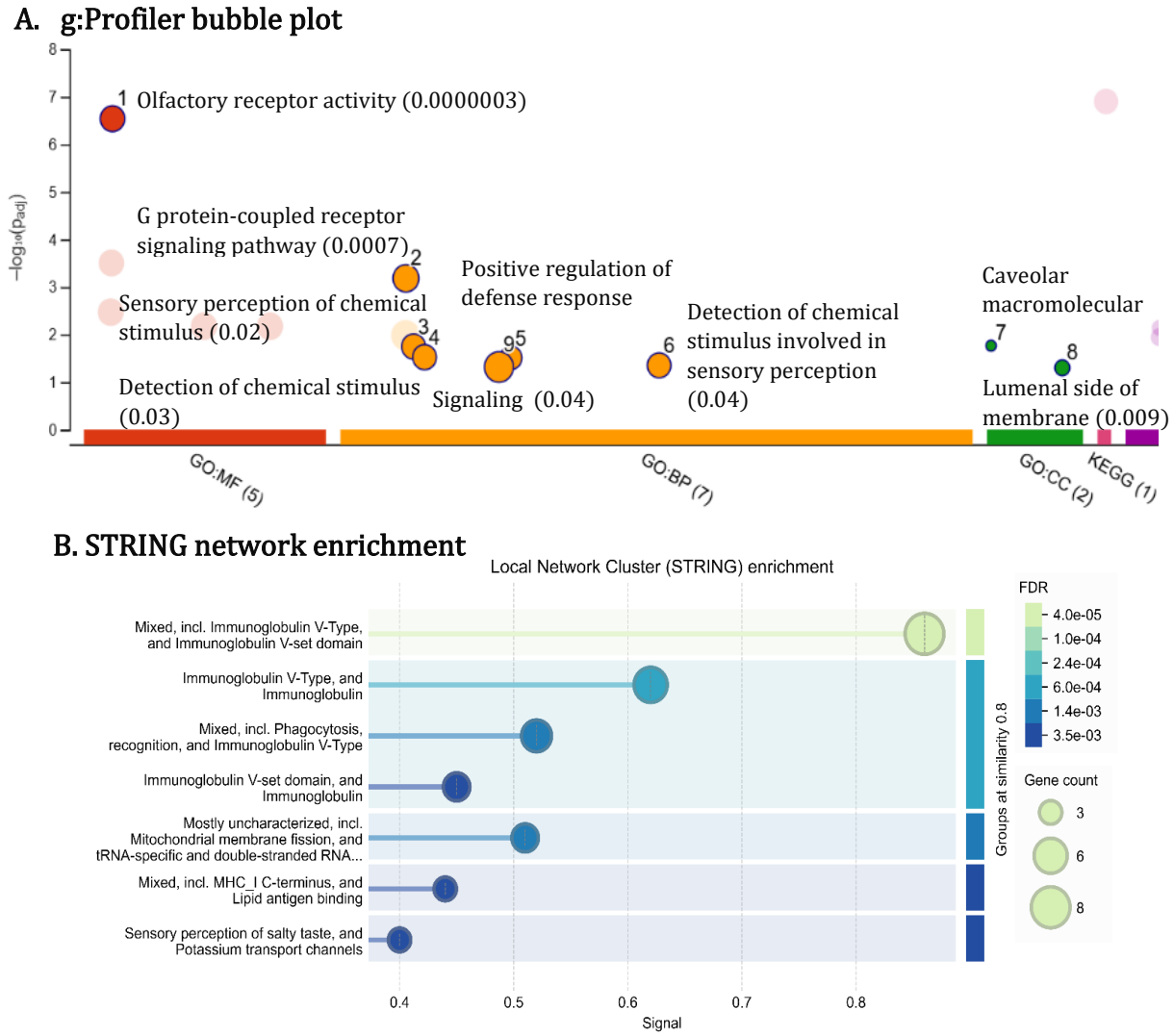


Figure 4. Gene ontology enrichment of genes under both diversifying selection and local adaptation. Gene ontology enrichment bubble plots with the Canis lupus familiaris genome with g:Profiler (A) and STRING protein network (B). The significantly enriched categories were highlighted and numbered based on significance and their respective adjusted p values are included in the parenthesis.

Considering the findings in Chapter 2, adaptive evolution in grey wolves is a multidimensional process exhibiting different selection dynamics across their geographical range, and local populations levels (White & Butlin 2021). The approaches to detect genes with signatures of diversifying selection were based of genetic differentiation, whereas the approaches to detect genes with signatures of local adaptation were based on their associations with environmental variables. Genetically

defined populations identified in the scale of Eurasia typically encompass multiple biomes, which can hinder the detection of climate/biome adaptation signals. The minimal overlap between genes associated with environmental pressures and those under diversifying selection is likely due to the differences in the selective effect of unique environmental pressures each population is experiencing. Another possible explanation for the minimal overlap could be that genes under diversifying selection may be driven by other selective pressures such as human activities e.g., in wolves persecution/hunting, and access to food and measures for population control in dogs,

RDA candidate SNPs correlations suggest adaptations in a climate gradient from cold-wet seasons to warm-wet seasons in both taxa (Fig. S9). Similarly, LFMM associations suggest that local adaptation could be linked to adaptations to precipitation during the wettest period in wolves (Table S5) and seasonal water availability in FRDs (Table S6). Based on the RDA analysis adaptive genetic variation in both taxa is most affected by lower temperatures of the wettest part of the year (Bio8) (Fig. S9). Adaptive genetic variation in wolves was also affected by higher precipitation in the coldest part of the year, lower annual range of temperature and higher elevation. Considering all the associations, it may be suggested that the strongest selective pressures i.e., drivers of adaptation in wolves are environmental conditions related to wet and thermostable areas (i.e., without extreme summers or winters) and higher altitudes. In FRDs the RDA analysis showed that the drivers of adaptation are both lower temperatures of the wettest part of the year (Bio8) and low precipitation in the wettest month (Bio13) (Fig S9). Thus, FRDs adaptive variation detected in this study may be most beneficial in different populations living in either both wet enviroments and semi-arid enviroments, with local adaptations to both these climatic conditions (e.g., drought tolerance). The representation of some biomes was extensive; therefore, it is unlikely that sample size obscured other similarities. This is supported by the detection of associations in less represented biomes like the boreal forests/taiga in wolves.

FRDs exhibited 35% more SNPs with signatures of local adaptation linked to species richness variables than wolves, which suggests that mammalian species richness is acting as a stronger selective force in FRDs than wolves (Fig. 2, Tables S5-S6). Most associations between FRDs genetic diversity and mammalian species richness were with the species richness of non-ungulate mammals, including carnivores. FRDs are often reliant on anthropogenic food sources but also known to scavenge on domestic and wild prey, and experience dietary niche overlaps with other carnivores, leading to interspecific competition (Justa et al., 2023; Reshamwala et al., 2021, Vanak & Gompper, 2009). Such overlaps and the potential for predation by larger carnivores, including wolves, suggest that an increase in carnivore biodiversity may elevate interspecific competition and predation risks, thereby influencing genetic diversity (Justa et al., 2023; Mahajan et al., 2024; Zhong et al., 2024).

Ungulate biodiversity was the strongest selective pressure for wolves, aligning with their ecological role as top predators preying primarily on ungulates (Nowak et al., 2011; Mech & Boitani 2003, Ståhlberg et al., 2017). Genetic diversity in wolves from biomes like boreal forests/tundra and flooded grasslands and savanna showed a negative correlation with ungulate biodiversity (Fig. 3B, Fig. S17), possibly indicating

local adaptation to specific prey species (e.g., olfactory receptors, morphological traits) in biomes with lower prey species diversity.

4.2. Genes associated with environmental variables

The functions of the highest-ranking genes, along with those exhibiting missense mutations, align well with the enriched GO terms observed in our study. Notwithstanding the low percentage of overlap between detected genes under selection in both taxa, there was an overlap between impacted phenotypic traits (e.g., metabolic, dietary, morphological and sensory perception). A similar pattern was observed in big cat species, which showed intraspecific adaptive variation impacting similar phenotypic traits (Figueiró et al., 2017). This suggests that local adaptation affects similar phenotypic traits in carnivores (i.e., morphology, metabolism and sensory perception), however the genes under local adaptation vary between grey wolves and FRDs. Due to the extensive list of genes identified we focused our discussion on adaptations in genes with or near high-ranking SNP variants, missense variations and/or with traits previously described as adaptive.

4.2.1. Morphological, reproductive and cardiovascular adaptations

Morphological adaptations related to body size and weight were identified in both taxa. Three genes with signatures of local adaptation in both, one in grey wolves and two in FRDs have been previously associated with pure-bred dogs' skeletal, height and body weight variation (e.g., FGF4, the fibroblast growth factor 4, growth hormone receptor (GHR) and post-transcriptional regulator of the insulin like growth factor 2 (IGF2BP2)) (Table 3). These genes were linked to elevation, precipitation and temperature in the LFMM and RDA analyses (Tables S5-S6). Additionally, signatures of local adaptation were also detected in FGF19 (Table S5), which has been associated with local adaptation in semi-aquatic carnivores (Wei et al., 2021). In carnivores, habitat shaped the evolution of the fibroblast growth factors, which are involved in regulation of physiological and morphological development (Wei et al., 2021). The FGF4 correlation echoes findings from a study on European wolves, which identified a SNP variant under selection near FGF4 gene (Stronen et al., 2015). Genetic and morphological clinal variation has been previously described in grey wolves (Nowak, 2003; Wheeldon et al., 2012), with increments in size towards northernmost regions exemplified by the Artic grey wolves, which are taller and heavier than their Southern counterparts (O'Keefe et al., 2013). Our findings underscore ongoing selection pressures related to body size in the context of habitat variation (Bannasch et al., 2020; Schlamp et al., 2015). Other genes associated with morphological variation were also detected in our study, including thirty-four genes previously associated with morphological variation in carnivores (Tables S5-S6) (Huang et al., 2021).

We also detected signatures of local adaptation in genes associated with the reproductive system. A notable reproductive gene with signatures of local adaptation is GAPDHS, which is involved in spermatogenesis (Margaryan et al., 2015) and has been previously detected to be under diversifying selection between FRDs and pure-bred dogs (Tables 2-3). Adaptations detected in this gene were associated with elevation, grass cover, temperature and non-carnivore mammalian species richness (Table S5).

Given the role of GAPDHS in spermatogenesis, this result suggests that temperature oscillations and local prey biodiversity are operating upon male grey wolves' reproductive fitness. Prey type and abundance have been shown to impact grey wolves' social dynamics including the number of mating pairs (Ausband, 2024) and group size (Sells, et al., 2021, Tallian et al., 2023). Thus, adaptations in GAPDHS could be molecular adaptations in reproductive fitness in response to changes in prey type and abundance.

Cardiovascular system-related genes displaying signatures of selection were detected in both wolves and FRDs, particularly in association with species richness (Fig. S18, Tables S5-S6). A variant near POLR2M, which has been implicated in cardiomyopathy, a disease which impacts heart muscle strength (Heliö et al., 2021), was identified in both grey wolves and FRDs. In FRDs this variant was associated with precipitation, and in wolves with temperature and day-length variables along with medium ungulate and carnivore species richness (Tables S5-S6). The association in wolves is of particular interest as it suggests local adaptation in response to multiple abiotic factors as well as prey biodiversity and interspecific competition. In FRDs, a variant near MMP17, which also has cardiovascular functions (Table 2), was also associated with ungulate and carnivore species richness, temperature and precipitation (Table S6). This finding is noteworthy as cardiac changes in mammals have been linked to increased workload, such as that induced by exercise, which can be influenced by predator ecology (Abel & Doenst, 2011; Chikachev, 2020; Martin et al., 2024). Investigating anatomical variation in FRD populations in the context of prey availability could provide further insights into how these animals are adapting to shifts between predatory and scavenging behaviours.

4.2.2. Energetic, metabolic and dietary adaptations

In FRDs genes associated with environmental variables and mammalian species richness were enriched for lipid metabolic processes and regulation of lipid biosynthesis, respectively (Table S7). Metabolic flexibility in lipid storage (i.e., energy-rich molecules) in animals can shift during periods of food shortages (Olsen et al., 2021), which suggests that both abiotic (e.g., seasonal changes, droughts) and biotic factors (i.e., local interspecific competition with other mammals) are driving adaptations in these genes. In FRDs A notable adaptation was detected in D2HGDH, a protein that regulates function of another protein essential for generating ATP (Lin et al., 2015) (Table 2). Altogether, this suggests adaptations to food shortages.

We also detected other adaptations related to thermoregulation in both taxa. In addition to the lipid storage adaptations described above, adaptations in HSPA14, a protein involved in cellular response to heat stress (Table 2), were also detected in FRDs. In wolves' molecular adaptation was detected in TRPM8, which has been implicated in regulation of cold-induced thermogenesis and the "wet dog shake", a behaviour to remove water from wet fur (Selescu et al., 2023). This suggests behavioural adaptations in wolves in response to wet environments. These adaptations present an example of intraspecific adaptations in similar traits in these sympatric taxa.

Another complex molecular trait identified to be under selection in both FRDs, and wolves was cell growth in response to nutrient and insulin levels. RPTOR is a regulatory protein for which we detected a SNP variant under selection within its coding region

(Tables S5-S6). Climate adaptations have previously been identified in the human ortholog. Downregulation of this gene has been proposed to impact MTOR pathway impacting thermogenesis and immune response phenotypes, resulting in increased fitness in colder climates (Sun et al., 2010). Thus, it is possible that this gene is involved in cold adaptation in mammals.

Metabolic molecular adaptations in genes involved in digestive processes were observed in both taxa. In wolves, a notable variant was identified near a glucose tolerance gene (MGAT4A), known for its role in carbohydrate metabolism (Tables 2-3). MGAT4A was associated with both temperature and day-length variables along with large ungulate and carnivore species richness (Table S5). Another notable variant was detected near HMGCL gene. In manned wolves (*Chrysocyon brachyurus*) an adaptation in this gene associated with the ability to digest fruit fibre was previously detected (Chavez et al., 2022). This suggests that wolves could be adapted to metabolize carbohydrates, which would enable them to expand their diet to include carbohydrate-based foods. This is consistent with previous studies have found that wolves opportunistically scavenge anthropogenic waste (Ciucci et al., 2020). In wolves, an intriguing pattern emerged with the xenobiotic metabolic process, suggesting adaptations in response to synthetic chemical byproducts from anthropogenic activities (Fig. S18) (Piwowarska & Kiedrzyńska, 2022; Priya et al., 2024). Anthropogenic expansion and development resulted in an increase of xenobiotic compounds (i.e., pollutants) in both terrestrial and aquatic natural habitats (Gałazka et al., 2023; Piwowarska & Kiedrzyńska, 2022). This pattern suggests that grey wolves are adapting to increases of pollutants in natural landscapes, byproducts of anthropogenic activities.

Some other interesting dietary adaptations associated with species richness variables were identified in both wolves and dogs. A missense variant in intestine-expressed protease, CAPN13, was detected in wolves, suggesting its potential role in dietary adaptation (Table 2). Additionally, we identified a missense variant in MRRF (Table 2), which has been previously found to be upregulated in dogs on a raw meat diet (Anturaniemi et al., 2020), suggesting adaptations for consumption of prey or other raw meats in FRDs. Signatures of adaptive evolution in MRRF were also observed in wolves, indicating a shared adaptive mechanism across canids.

4.2.3. Immune system adaptations

We detected signatures of selection in multiple canine beta defensins in wolves (e.g., CBD103, CBD119, and CBD122). Defensins are antimicrobial peptides, and the beta family has various functions in innate immunity, from host protection to cellular signalling to promoting wound healing (Leonard et al., 2012). CBD103 expression has been associated with anti-microbial protein in both skin and nasal cavities in dogs (Leonard et al., 2012; Aono et al., 2021). An allele of CBD103 has been implicated in black coat coloration in both dogs and grey wolves, and its presence in wolves has been proposed to have occurred during an introgression event with indigenous North American dogs 1,598-7,248 years ago (Anderson et al., 2009; Candille et al., 2010; Schweizer et al., 2018). Furthermore, selection upon this gene has been proposed as a

mechanism for enhanced immunity in response to exposure to the distemper virus in wolves (Schweizer et al., 2018). Since we detected signatures of local adaptive evolution in four defensin genes, this suggests that this gene family has an important role in grey wolves' adaptations.

Another interesting immune adaptation was identified in IL12B (Table 3), a gene known to mediate host response to rabies. A SNP variant in IL12B, ~17168bp downstream of the variant detected in wolves, was found to be associated with Artic rabies variants in Arctic foxes (Baecklund et al., 2021). Rabies is a fatal zoonotic disease prevalent in dogs and transmitted to humans, and consequently a public health concern (Marino et al., 2017). Although it has been rarely reported in grey wolves, it is estimated to be a common cause of mortality (2-20%), and it is a serious issue for the endangered Ethiopian wolf (Marino et al., 2017; Gese et al., 2021). Variants in IL12B are likely to represent adaptations to exposure to specific rabies strains in certain environments (e.g., Arctic rabies). Understanding the impact of these genetic variants could be crucial for developing management strategies aimed at reducing the incidence of rabies in canid populations, thereby informing conservation and public health efforts.

4.2.4. Adaptations for prey/food detection

Various genes involved in sensory perception were found to be under selection in both grey wolves and FRDs. Multiple variants under selection were detected in olfactory receptor genes (OR) (Tables S5-S6). In wolves we detected two prey-related variants in TRPV1 and OR5A2. TRPV1 has been previously associated with grey wolves' adaptive variation (Stronen et al., 2015). In the current study, the association was with ungulate species richness, suggesting a role in detection of prey-related stimuli (Table 3). OR5A2 binds to substances present in deer musk (Yoshikawa et al., 2022; Wang et al., 2025) and the detected variant was associated with ungulate species richness, suggesting this adaptation contributed to prey detection by grey wolves. OR5 genes have been previously found to exhibit local adaptation in North American grey wolves (Hendricks et al., 2018) and breed-specific genetic variation in dogs (Kokocińska-Kusiak et al., 2021). In FRDs a notable example is OR2W3, which binds to tridecanaldehyde, a fatty acid found in plant oils, specifically in citrus, evergreens, herbs and other plants (Wang et al., 2025). This is likely an adaptation in FRDs to better locate anthropogenic foods containing these compounds. This not only suggests a consistent involvement of OR genes in local adaptive evolution of both grey wolves and dogs, but it reiterates that OR are essential in canid environmental adaptations.

In FRDs additional sensory genes and behavioural genes were also detected to be associated with precipitation and mammalian species richness. We found that the IMPG1 gene, linked to visual acuity (Table 2), was associated with precipitation of the wettest month in FRDs. This suggests adaptations related to visual acuity possibly improving prey detection and navigation of their enviroments in FRDs. We also found signatures of local adaptation in a gene implicated with anxiety and fear response (i.e., GABRA5) associated with both carnivore and large ungulate species richness. This suggests adaptations to predation risk from carnivores and/or interspecific

competition. This gene has been detected to be under selection in both dogs and wolves (Li et al., 2014; Schweizer et al., 2016); hence it may be involved in behavioural changes associated with domestication.

4.2.5. SLC genes in canid adaptations

The solute carrier (SLC) gene family was particularly represented in this study, with 18 and 16 SLC genes identified with signatures of local adaptive variation in grey wolves and FRDs, respectively. SLC family is the largest group of ubiquitously expressed genes coding transmembrane transporter proteins with key roles in several cellular processes (Casiraghiare et al., 2024). These transporter proteins regulate intake of various nutrients and metabolites (e.g., glucose, neurotransmitters, ions), hence are involved in physiological functions related to metabolism, energy homeostasis, tissue development, immune system and neurological processes (Zhang et al., 2019). SLC genes are also known for their association with carbohydrate absorption and kidney function, especially in the context of a raw meat diet (Arata et al., 2008; Bankir et al., 2024).

Signatures of selection and local adaptation have been previously detected in SLC genes in multiple studies on grey wolves and dogs (Bergström et al., 2022; Pilot et al., 2021; Schweizer et al., 2016; Stronen et al., 2015). SLC5A1, for instance, has been detected to be under selection during domestication, feralization, and local adaptive processes in dogs (Batcherlor et al., 2013; Niego & Benítez-Burraco, 2021; Zhang et al., 2020). We detected two molecular adaptations in wolves in the related SLC5A6 gene. Additionally, we detected some notable variants in SLC6A6 in wolves and SLC6A9 in FRDs (Table 2). Based on earlier studies, these genes were involved in maintaining hepatic function, specifically taurine (Bankir et al., 2024) and glycine uptake (Bankir et al., 2024; Bruchim et al., 2017; Lin et al., 2015). Our study reiterates the adaptive significance of the SLC gene family in canids, highlighting its crucial role in responding to environmental changes and associated dietary shifts.

4.3. Genes in FRDs with signatures of both diversifying selection and local adaptation

The genes under both diversifying selection and local adaptation in FRDs have functions related mostly to the immune and sensory (i.e., olfactory and taste) systems (Fig. 4). Immunoglobulins are a family of proteins which are part of vertebrates' adaptive immunity mechanisms. The variable portion (V) of the protein which determines antigen binding (Flajnik & Kasahara, 2010) is in a constant selective pressure during their coevolution with pathogens. For example, the silver fox displays strong signatures of adaptive evolution and variation in expression of genes coding for the immunoglobulin V subsection (Yi et al., 2025). In FRDs adaptations in these genes likely enabled them to adapt to novel pathogens during their range expansion.

Canids depend on their olfactory system to navigate their environments, acquire resources (i.e., finding food and water), identify threats and even for intraspecific interactions (e.g., selecting mates, establishing hierarchies) (Kokocińska-Kusiak et al., 2021). Comparative studies between wild canids and dogs have shown that dogs' exhibit reductions in the size of anatomical structures involved in olfaction and

loss/pseudogenization of olfactory receptor genes due to their domestication, thus their olfactory acumen was reduced (Bird et al., 2020; Ortiz-Leal et al., 2021; Mouton et al., 2025). Our findings suggest that natural selection is re-shaping olfactory traits in FRDs and maybe even reversing the effects of domestication to some extent.

In addition to olfactory genes, we detected genes involved in taste perception, with signatures for local adaptation and diversifying selection in FRDs. One of the genes detected in this study, SCNN1B (Table S1), has been associated with the perceived taste intensity if sodium chloride (Dias et al., 2013). The molecular mechanisms involved in salty tasty perception are not well understood, still the role of sodium channels in this sensory perception, including subunits encoded by the SCNN1 genes, has been described (Antinucci et al., 2017). Salty taste perception is a protective mechanism for maintaining internal sodium homeostasis (Taruno & Gordon, 2023), which is particularly important for FRDs as human commensals due to the high content of sodium in anthropogenic foods. Overall, genes in FRDs with signatures of local adaptation which are also under diversifying selection, may have facilitated dog's range expansion into multiple habitats and adaptation to their ecological niche as a human commensal.

4.4. Common evolutionary patterns in grey wolves and FRDs

There was an overlap between functional terms of genes involved in local adaptation in wolves and FRDs, notably related to sensory perception/detection of chemical stimulus (Fig. S18). This suggests that local adaptive evolution is occurring in genes with similar functions, but which are distinct in each taxa. The role of OR, SLC and CBD genes in canid adaptive evolution is evident, with adaptive variants of genes from these families detected in both grey wolves and domestic dogs in this and previous studies. This finding indicates a possible convergence in the types of biological processes involved in local adaptation, despite differences at the genetic level. Our analysis also revealed a few genes involved in local adaptive variation in various mammals (e.g., RPTOR) and in both canids (e.g., D2HGDH, POL2RM, GLI2).

Altogether, our findings suggest that FRDs have evolved or retained adaptations associated with expanding their diets to include non-anthropogenic foods, contrasting with their evolutionary trajectory during domestication towards a more carbohydrate-rich diet (Gao et al., 2019; Tazerji et al., 2024). This dietary shift highlights the flexible and adaptive nature of canid metabolism in response to available food resources. Additionally, our results suggest that variations in carnivore and non-ungulate mammal species' richness may elevate interspecific competition and through this drive local adaptations.

5. Conclusion

This comparative study revealed that two closely related sympatric canids, one domestic (FRDs) and the other wild (grey wolves), exhibit distinct adaptations to local environmental conditions despite their shared evolutionary history. Genetic adaptations in individuals from similar habitats are unique to grey wolves and FRDs,

but influence similar phenotypes related to morphology, cardiovascular function, immunity and sensory perception. A particularly novel finding from this study was the identification of molecular adaptations (e.g., cardiovascular, prey detection and metabolism) in both taxa associated with mammalian species richness. Reproductive adaptations in wolves associated with ungulate species richness confirm the role of prey as a selective pressure in the species. Detection of behavioural and sensory adaptations in FRD associated with mammalian species richness suggests that either wild mammals may be important as food sources of FRDs, or mammalian species richness is correlated with another environmental factor driving adaptations in FRDs. Wolves showed adaptations to cope with the effects of anthropogenic expansion from adapting to tolerate xenobiotic substances, and digest carbohydrate-based foods, prominent in both anthropogenic waste and crops. Additionally, wolves showed adaptations to habitat specific pathogens in an extreme environment (i.e., Artic tundra) highlighting their adaptability. Meanwhile, FRDs dietary adaptations suggest either shifts to include non-anthropogenic foods or the retention of ancestral adaptations for carnivorous diet. These associations suggest that canids likely have genetically mediated sensory, metabolic and digestive adaptations that evolved in response to prey diversity and intraspecific competition. Solute carrier proteins (SLC) and canine beta defensins (CBD) emerged as important families of proteins in canids local adaptation. This study revealed parallel adaptations achieved through different genetic pathways, highlighting the significance of examining both domesticated taxa and their wild relatives within shared ecological contexts. By investigating these widespread carnivores, we gained deeper insights into the interplay between natural selection and anthropogenic factors in shaping the genomes of closely related taxa.

References

Abel, E. D., & Doenst, T. (2011). Mitochondrial adaptations to physiological vs. pathological cardiac hypertrophy. *Cardiovascular Research, 90*(2), 234–242. https://doi.org/10.1093/cvr/cvr015

Álvarez, L., Marín-García, P. J., & Llobat, L. (2022). Immunological and genomic characterization of Ibizan Hound dogs in an endemic Leishmania infantum region. *Parasites & vectors*, *15*(1), 445. https://doi.org/10.1186/s13071-022-05504-3

Anderson, T. M., vonHoldt, B. M., Candille, S. I., Musiani, M., Greco, C., Stahler, D. R., Smith, D. W., Padhukasahasram, B., Randi, E., Leonard, J. A., Bustamante, C. D., Ostrander, E. A., Tang, H., Wayne, R. K., & Barsh, G. S. (2009). Molecular and evolutionary history of melanism in North American gray wolves. *Science, 323*(5919), 1339–1343. https://doi.org/10.1126/science.1165448

Anderson, C. M. H., Howard, A., Walters, J. R. F., Ganapathy, V., & Thwaites, D. T. (2009). Taurine uptake across the human intestinal brush-border membrane is via two transporters: H+-coupled PAT1 (SLC36A1) and Na+- and Cl−-dependent TauT (SLC6A6). *The Journal of Physiology, 587*(4), 731–744. https://doi.org/10.1113/jphysiol.2008.164228

Antinucci, M., & Risso, D. (2017). A Matter of Taste: Lineage-Specific Loss of Function of Taste Receptor Genes in Vertebrates. *Frontiers in molecular biosciences*, *4*, 81. https://doi.org/10.3389/fmolb.2017.00081

Anturaniemi, J., Zaldívar-López, S., Savelkoul, H. F. J., Elo, K., & Hielm-Björkman, A. (2020). The effect of atopic dermatitis and diet on the skin transcriptome in Staffordshire Bull Terriers. *Frontiers in Veterinary Science, 7*, 552251. https://doi.org/10.3389/fvets.2020.552251

Aono, S., Dennis, J. C., He, S., Wang, W., Tao, Y. X., & Morrison, E. E. (2021). Exploring pleiotropic functions of canine β-defensin 103: Nasal cavity expression, antimicrobial activity, and melanocortin receptor activity. *Anatomical Record, 304*(1), 210–221. https://doi.org/10.1002/ar.24300

Arata, S., Ogata, N., Shimozuru, M., Takeuchi, Y., & Mori, Y. (2008). Sequences and polymorphisms of the canine monoamine transporter genes SLC6A2, SLC6A3, and SLC6A4 among five dog breeds. *The Journal of Veterinary Medical Science, 70*(9), 971–975. https://doi.org/10.1292/jvms.70.971

Arendt, M., Cairns, K. M., Ballard, J. W., Savolainen, P., & Axelsson, E. (2016). Diet adaptation in dog reflects spread of prehistoric agriculture. *Heredity, 117*(5), 301–306. https://doi.org/10.1038/hdy.2016.48

Ausband, D. E. (2024). Lifetime reproductive characteristics of gray wolves. *Journal of Mammalogy, 105*(6), 1322–1327. https://doi.org/10.1093/jmammal/gyae042

Axelsson, E., Ratnakumar, A., Arendt, M.-L., Maqbool, K., Webster, M. T., et al. (2013). The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature, 495*, 360–364.

Baecklund, T. M., Donaldson, M. E., Hueffer, K., & Kyle, C. J. (2021). Genetic structure of immunologically associated candidate genes suggests arctic rabies variants exert differential selection in arctic fox populations. *PLOS ONE, 16*(10), e0258975. https://doi.org/10.1371/journal.pone.0258975

Bannasch, D. L., Baes, C. F., & Leeb, T. (2020). Genetic variants affecting skeletal morphology in domestic dogs. *Trends in Genetics, 36*(8), 598–609. https://doi.org/10.1016/j.tig.2020.05.005

Bankir, L., Crambert, G., & Vargas-Poussou, R. (2024). The SLC6A18 transporter is most likely a Na-dependent glycine/urea antiporter responsible for urea secretion in the proximal straight tubule: Influence of this urea secretion on glomerular filtration rate. *Nephron, 148*(11-12), 796–822. https://doi.org/10.1159/000539602

Batchelor, D. J., German, A. J., & Shirazi-Beechey, S. P. (2013). Relevance of sodium/glucose cotransporter-1 (SGLT1) to diabetes mellitus and obesity in dogs. *Domestic Animal Endocrinology, 44*(3), 139–144. https://doi.org/10.1016/j.domaniend.2012.12.001

Belo, V. S., Werneck, G. L., da Silva, E. S., Barbosa, D. S., & Struchiner, C. J. (2015). Population estimation methods for free-ranging dogs: A systematic review. *PLOS ONE, 10*(12), e0144830. https://doi.org/10.1371/journal.pone.0144830

Belsare, A., & Vanak, A. T. (2020). Modelling the challenges of managing free-ranging dog populations. *Scientific Reports, 10*, 18874. https://doi.org/10.1038/s41598-020-75828-6

Bergström, A., Stanton, D. W. G., Taron, U. H., et al. (2022). Grey wolf genomic history reveals a dual ancestry of dogs. *Nature, 607*, 313–320. https://doi.org/10.1038/s41586-022-04824-9

Bird, D. J., Jacquemetton, C., Buelow, S. A., Evans, A. W., & Van Valkenburgh, B. (2021). Domesticating olfaction: Dog breeds, including scent hounds, have reduced cribriform plate morphology relative to wolves. *Anatomical Record, 304*(1), 139–153. https://doi.org/10.1002/ar.24518

Boitani, L., Phillips, M., & Jhala, Y. (2023). Canis lupus (amended version of 2018 assessment). *The IUCN Red List of Threatened Species*, e.T3746A247624660. https://dx.doi.org/10.2305/IUCN.UK.2023-1.RLTS.T3746A247624660.en

Bruchim, Y., Horowitz, M., & Aroch, I. (2017). Pathophysiology of heatstroke in dogs - revisited. *Temperature, 4*(4), 356–370. https://doi.org/10.1080/23328940.2017.1367457

Bryce, C. B. (2021). Dogs as pets and pests: Global patterns of canine abundance, activity, and health. *Integrative and Comparative Biology, 61*(1), 154–165. https://doi.org/10.1093/icb/icab046

Candille, S. I., Kaelin, C. B., Cattanach, B. M., Yu, B., Thompson, D. A., Nix, M. A., Kerns, J. A., Schmutz, S. M., Millhauser, G. L., & Barsh, G. S. (2007). A -defensin mutation causes black coat color in domestic dogs. *Science, 318*(5855), 1418–1423. https://doi.org/10.1126/science.1147880

Capitain, S., Cimarelli, G., Blenkuš, U., Range, F., & Marshall-Pescini, S. (2024). Street-wise dog testing: Feasibility and reliability of a behavioural test battery for free-ranging dogs in their natural habitat. *PLOS ONE, 19*(3), e0296509. https://doi.org/10.1371/journal.pone.0296509

Casiraghi, A., Bensimon, A., & Superti-Furga, G. (2021). Recent developments in ligands and chemical probes targeting solute carrier transporters. *Current Opinion in Chemical Biology, 62*, 53–63. https://doi.org/10.1016/j.cbpa.2021.01.012

Chavez, D. E., Gronau, I., Hains, T., Dikow, R. B., Frandsen, P. B., Figueiró, H. V., Garcez, F. S., Tchaicka, L., de Paula, R. C., Rodrigues, F. H. G., Jorge, R. S. P., Lima, E. S., Songsasen, N., Johnson, W. E., Eizirik, E., Koepfli, K. P., & Wayne, R. K. (2022). Comparative genomics uncovers the evolutionary history, demography, and molecular adaptations of South American canids. *Proceedings of the National Academy of Sciences of the United States of America, 119*(34), e2205986119. https://doi.org/10.1073/pnas.2205986119

Chen, C., Granados, A., Brodie, J. F., Kays, R., Davies, T. J., Liu, R., Fisher, J. T., Ahumada, J., McShea, W., Sheil, D., Mohd-Azlan, J., Agwanda, B., Andrianarisoa, M. H., Appleton, R. D., Bitariho, R., Espinosa, S., Grigione, M. M., Helgen, K. M., Hubbard, A., ... Burton, A. C. (2024). Combining camera trap surveys and IUCN range maps to improve knowledge of species distributions. *Conservation Biology, 38*, e14221. https://doi.org/10.1111/cobi.14221

Chen, R., Hou, Y., Connell, M., & Zhu, S. (2021). Homeodomain protein Six4 prevents the generation of supernumerary Drosophila type II neuroblasts and premature differentiation of intermediate neural progenitors. *PLoS genetics*, *17*(2), e1009371. https://doi.org/10.1371/journal.pgen.1009371

Chew, S., Lampinen, R., Saveleva, L., Korhonen, P., Mikhailov, N., Grubman, A., Polo, J. M., Wilson, T., Komppula, M., Rönkkö, T., Gu, C., Mackay-Sim, A., Malm, T., White, A. R., Jalava, P., & Kanninen, K. M. (2020). Urban air particulate matter induces

mitochondrial dysfunction in human olfactory mucosal cells. *Particle and fibre toxicology*, *17*(1), 18. https://doi.org/10.1186/s12989-020-00352-4

Chikachev, R. (2020). The index of the hearts squad Carnivora, as an indicator of predator ecology. *E3S Web of Conferences, 203,* 01010. https://doi.org/10.1051/e3sconf/202020301010

Choi, S. W., Yeon, J. T., Ryu, B. J., Kim, K. J., Moon, S. H., Lee, H., Lee, M. S., Lee, S. Y., Heo, J. C., Park, S. J., & Kim, S. H. (2015). Repositioning Potential of PAK4 to Osteoclastic Bone Resorption. *Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research*, *30*(8), 1494–1507. https://doi.org/10.1002/jbmr.2468

Ciucci, P., Mancinelli, S., Boitani, L., Gallo, O., & Grottoli, L. (2020). Anthropogenic food subsidies hinder the ecological role of wolves: Insights for conservation of apex predators in human-modified landscapes. *Global Ecology and Conservation*, *21*, e00841. https://doi.org/10.1016/j.gecco.2019.e00841

Dauphin, B., Rellstab, C., Wüest, R. O., Karger, D. N., Holderegger, R., Gugerli, F., & Manel, S. (2023). Re-thinking the environment in landscape genomics. *Trends in Ecology & Evolution, 38*(3), 261–274. https://doi.org/10.1016/j.tree.2022.10.010

Dias, A. G., Rousseau, D., Duizer, L., Cockburn, M., Chiu, W., Nielsen, D., & El-Sohemy, A. (2013). Genetic variation in putative salt taste receptors and salt taste perception in humans. *Chemical senses*, *38*(2), 137–145. https://doi.org/10.1093/chemse/bjs090

Duarte, J., García, F. J., & Fa, J. E. (2016). Depredatory impact of free-roaming domestic dogs on Mediterranean deer in southern Spain: Implications for human-wolf conflict. *Folia Zoologica, 65*(2), 135–141. https://doi.org/10.25225/fozo.v65.i2.a8.2016

Espinosa-Soto, C., Martin, O. C., & Wagner, A. (2011). Phenotypic plasticity can facilitate adaptive evolution in gene regulatory circuits. *BMC Evolutionary Biology, 11*, 5. https://doi.org/10.1186/1471-2148-11-5

Figueiró, H. V., Li, G., Trindade, F. J., Assis, J., Pais, F., Fernandes, G., Santos, S. H. D., Hughes, G. M., Komissarov, A., Antunes, A., Trinca, C. S., Rodrigues, M. R., Linderoth, T., Bi, K., Silveira, L., Azevedo, F. C. C., Kantek, D., Ramalho, E., Brassaloti, R. A., Villela, P. M. S., … Eizirik, E. (2017). Genome-wide signatures of complex introgression and adaptive evolution in the big cats. *Science advances*, *3*(7), e1700299. https://doi.org/10.1126/sciadv.170029

Flajnik, M., & Kasahara, M. (2010). Origin and evolution of the adaptive immune system: genetic events and selective pressures. *Nature Reviews Genet* 11, 47–59. https://doi.org/10.1038/nrg2703

Freedman, A. H., Lohmueller, K. E., & Wayne, R. K. (2016). Evolutionary history, selective sweeps, and deleterious variation in the dog. *Annual Review of Ecology, Evolution, and Systematics, 47*, 73–96. https://doi.org/10.1146/annurev-ecolsys-121415-032155

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1km spatial resolution climate surfaces for global land areas. *International Journal of Climatology, 37*(12), 4302–4315.

Gałazka, A., Jankiewicz, U., & Szczepkowski, A. (2023). Biochemical characteristics of laccases and their practical application in the removal of xenobiotics from water. *Applied Sciences, 13*(7), 4394. https://doi.org/10.3390/app13074394

Gallagher, M. D., & Chen-Plotkin, A. S. (2018). The post-GWAS era: From association to function. *American Journal of Human Genetics, 102*(5), 717–730. https://doi.org/10.1016/j.ajhg.2018.04.002

Gao, X., Li, Y., Adetula, A. A., Wu, Y., & Chen, H. (2019). Analysis of new retrogenes provides insight into dog adaptive evolution. *Ecology and Evolution, 9*(20), 11185–11197. https://doi.org/10.1002/ece3.5620

Ge, H., Mu, L., Jin, L., Yang, C., Chang, Y. E., Long, Y., DeLeon, G., Deleyrolle, L., Mitchell, D. A., Kubilis, P. S., Lu, D., Qi, J., Gu, Y., Lin, Z., & Huang, J. (2017). Tumor associated CD70 expression is involved in promoting tumor migration and macrophage infiltration in GBM. *International journal of cancer*, *141*(7), 1434–1444. https://doi.org/10.1002/ijc.30830

Gese, E. M., Hart, J. P., & Terietzky, P. A. (2021). Gray wolves. Wildlife Damage Management Technical Series. USDA, APHIS, WS National Wildlife Research Center. Ft. Collins, Colorado. 20 p.

Gompper, M. E. (2014). Introduction: Outlining the ecological influences of a subsidized, domesticated predator. In Free-Ranging Dogs and Wildlife Conservation (pp. 1–8). Oxford University Press.

He, J., Yang, Y., Zhang, J., Chen, J., Wei, X., He, J., & Luo, L. (2017). Ribosome biogenesis protein Urb1 acts downstream of mTOR complex 1 to modulate digestive organ development in zebrafish. *Journal of genetics and genomics = Yi chuan xue bao*, *44*(12), 567–576. https://doi.org/10.1016/j.jgg.2017.09.013

Hecht, E. E., Barton, S. A., Rogers Flattery, C. N., & Meza Meza, A. (2023). The evolutionary neuroscience of domestication. *Trends in Cognitive Sciences, 27*(6), 553–567. https://doi.org/10.1016/j.tics.2023.03.008

Heliö, K., Mäyränpää, M. I., Saarinen, I., Ahonen, S., Junnila, H., Tommiska, J., Weckström, S., Holmström, M., Toivonen, M., Nikus, K., Hathaway, J., Siivonen, P., Muona, M., Sistonen, J., Salmenperä, P., Gentile, M., Paananen, J., Myllykangas, S., Alastalo, T. P., Heliö, T., ... Koskenvuo, J. (2021). GRINL1A complex transcription unit containing GCOM1, MYZAP, and POLR2M genes associates with fully penetrant

recessive dilated cardiomyopathy. *Frontiers in Genetics, 12*, 786705. https://doi.org/10.3389/fgene.2021.786705

Hendricks, S. A., Schweizer, R. M., & Wayne, R. K. (2019). Conservation genomics illuminates the adaptive uniqueness of North American gray wolves. *Conservation Genetics, 20*(1), 29–43. https://doi.org/10.1007/s10592-018-1118-z

Hijmans, R. (2024). terra: Spatial Data Analysis (Version 1.7-81). https://github.com/rspatial/terra
Hitti-Malin, R. J., Burmeister, L. M., Ricketts, S. L., et al. (2020). A LINE-1 insertion situated in the promoter of IMPG2 is associated with autosomal recessive progressive retinal atrophy in Lhasa Apso dogs. *BMC Genetics, 21*, 100. https://doi.org/10.1186/s12863-020-00911-w

Hrckova Turnova, E., Bielikova, M., Kostal, V., Turna, J., & Dudas, A. (2022). Occurrence of the dominant black KB allele of CBD103 in German Shepherd Dogs. *Animal Genetics, 53*(2), 230–231. https://doi.org/10.1111/age.13171

Hughes, J., & Macdonald, D. W. (2013). A review of the interactions between free-roaming domestic dogs and wildlife. *Biological Conservation, 157*, 341–351. https://doi.org/10.1016/j.biocon.2012.07.005

Huang, X., Sun, D., Wu, T., Liu, X., Xu, S., & Yang, G. (2021). Genomic insights into body size evolution in Carnivora support Peto's paradox. *BMC genomics*, *22*(1), 429. https://doi.org/10.1186/s12864-021-07732-w

Hussain, S. R., Mahar, N., Dirzo, R., & Habib, B. (2021). Successful neighbour: Interactions of the generalist carnivore red fox with dogs, wolves and humans for continued survival in dynamic anthropogenic landscapes. *Global Ecology and Conservation, 25,* e01446. https://doi.org/10.1016/j.gecco.2020.e01446

Justa, P., & Lyngdoh, S. (2023). Understanding carnivore interactions in a cold arid trans-Himalayan landscape: What drives co-existence patterns within predator guild along varying resource gradients? *Ecology and Evolution, 13*(5), e10040. https://doi.org/10.1002/ece3.10040

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., & Haussler, D. (2002). The human genome browser at UCSC. *Genome Research, 12*(6), 996–1006.
Kokocińska-Kusiak, A., Woszczyło, M., Zybala, M., Maciocha, J., Barłowska, K., & Dzięcioł, M. (2021). Canine olfaction: Physiology, behavior, and possibilities for practical applications. *Animals, 11*(8), 2463. https://doi.org/10.3390/ani11082463

Krafft, E., Lybaert, P., Roels, E., Laurila, H. P., Rajamäki, M. M., Farnir, F., Myllärniemi, M., Day, M. J., Mc Entee, K., & Clercx, C. (2014). Transforming growth factor beta 1 activation, storage, and signaling pathways in idiopathic pulmonary fibrosis in

dogs. *Journal of veterinary internal medicine*, *28*(6), 1666–1675. https://doi.org/10.1111/jvim.12432

Kolberg, L., Raudvere, U., Kuzmin, I., Adler, P., Vilo, J., & Peterson, H. (2023). g:Profiler—interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkad347

Kuijper, D. P., Sahlén, E., Elmhagen, B., Chamaillé-Jammes, S., Sand, H., Lone, K., & Cromsigt, J. P. (2016). Paws without claws? Ecological effects of large carnivores in anthropogenic landscapes. Proceedings. *Biological sciences, 283*(1841), 20161625. https://doi.org/10.1098/rspb.2016.1625

Lancto, C. A., Torres, S. M., Hendrickson, J. A., Martins, K. V., & Rutherford, M. S. (2013). Altered expression of antimicrobial peptide genes in the skin of dogs with atopic dermatitis and other inflammatory skin conditions. *Veterinary Dermatology, 24*(4), 414–e90. https://doi.org/10.1111/vde.12034

Lasky, J. R., Josephs, E. B., & Morris, G. P. (2023). Genotype-environment associations to reveal the molecular basis of environmental adaptation. *The Plant Cell, 35*(1), 125–138. https://doi.org/10.1093/plcell/koac267

Leonard, B. C., Marks, S. L., Outerbridge, C. A., Affolter, V. K., Kananurak, A., Young, A., Moore, P. F., Bannasch, D. L., & Bevins, C. L. (2012). Activity, expression and genetic variation of canine β-defensin 103: A multifunctional antimicrobial peptide in the skin of domestic dogs. *Journal of Innate Immunity, 4*(3), 248–259. https://doi.org/10.1159/000334566

Li, Y., Wang, G. D., Wang, M. S., Irwin, D. M., Wu, D. D., & Zhang, Y. P. (2014). Domestication of the dog from the wolf was promoted by enhanced excitatory synaptic plasticity: a hypothesis. *Genome biology and evolution*, *6*(11), 3115–3121. https://doi.org/10.1093/gbe/evu245

Lino, S., Rossa, M., Fernandes, J.M., Barros, T., Hipólito, D., Ferreira, E., Aliácar, S.C., Cadete, D., Fonseca, C., Torres, R.T., Rosalino, L.M. & Carvalho, J. (2023). Dog in sheep's clothing: livestock depredation by free-ranging dogs may pose new challenges to wolf conservation. *European Journal of Wildlife Research,* 69, 107. https://doi.org/10.1007/s10344-023-01740-9

Liu, Y.-H., Wang, L., Xu, T., Guo, X., Li, Y., Yin, T.-T., Yang, H.-C., Hu, Y., Adeola, A. C., Sanke, O. J., Otecko, N. O., Wang, M., Ma, Y., Charles, O. S., Sinding, M.-H. S., Gopalakrishnan, S., Samaniego, J. A., Hansen, A. J., Fernandes, C., Gaubert, P., Budd, J., Dawuda, P. M., Rueness, E. K., Jiang, L., Zhai, W., Gilbert, M. T. P., Peng, M.-S., Qi, X., Wang, G.-D., & Zhang, Y.-P. (2018). Whole-genome sequencing of African dogs provides insights into adaptations against tropical parasites. *Molecular Biology and Evolution, 35*(2), 287–298. https://doi.org/10.1093/molbev/msx258

Liu, Y.-H., Wang, L., Zhang, Z., Otecko, N.O., Khederzadeh, S., Dai, Y., Liang, B., Wang, G., & Zhang, Y.P. (2021). Whole-Genome Sequencing Reveals Lactase Persistence Adaptation in European Dogs. *Molecular biology and evolution, 38*(11), 4884–4890. https://doi.org/10.1093/mlbev/msab214

Losey, R. J., Drake, A. G., Ralrick, P. E., Jass, C. N., Lieverse, A. R., Bieraugle, M., Christenson, R., & Steuber, K. (2022). Dogs and wolves on the northern plains: A look from beyond the site in Alberta. *Journal of Archaeological Science*, *148*, Article 105690. https://doi.org/10.1016/j.jas.2022.105690

Mahajan, P., Khandal, D., Dhakad, M., & Rasal, V. (2024). A dog-eat-dog world: predation and resource acquisition by free-ranging dogs around Ranthambhore Tiger Reserve, Rajasthan, India. *Ethology Ecology & Evolution*, *36*(5), 572–582. https://doi.org/10.1080/03949370.2024.2358767

Margaryan, H., Dorosh, A., Capkova J., Manaskova-Postlerova, P., Philiminenko, A., Hozak, P., & Peknicova, J. (2015). Characterization and possible function of glyceraldehyde-3-phosphate dehydrogenase-spermatogenic protein FAPDHS in mammalian sperm. *Reproductive biology and endocrinology: RB&E, 13* 15. https://doi.org/10.1186/s12958-015-0008-1

McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R., Thormann, A., Flicek, P., & Cunningham, F. (2016). The Ensembl Variant Effect Predictor. *Genome biology*, 17(1), 122. https://doi.org/10.1186/s13059-016-0974-4

Mech, L. D., & Boitani, L. (Eds.). (2003). *Wolves: Behavior, ecology, and conservation*. University of Chicago Press.

Mouton, A., Levine, J. M., Bird, D. J., Li, G., Morselli, M., Craven, B. A., Pellegrini, M., Van Valkenburgh, B., Wayne, R. K., & Murphy, W. J. (2025). Genetic and anatomical determinants of olfaction in dogs and wild canids. *Molecular Biology and Evolution, 42*, 1–14. https://doi.org/10.1093/molbev/msaf035

Newsome, T. M., Fleming, P. J. S., Dickman, C. R., Doherty, T. S., Ripple, W. J., Ritchie, E. G., & Wirsing, A. J. (2017). Making a new dog? *BioScience, 67*(4), 374–381. https://doi.org/10.1093/biosci/bix022

Niego, A., & Benítez-Burraco, A. (2021). Are feralization and domestication truly mirror processes? *Ethology Ecology & Evolution, 34*(5), 557–590. https://doi.org/10.1080/03949370.2021.1975314

Nowak, R. (2003). Wolf evolution and taxonomy. In Wolf behavior, ecology and conservation (pp. 239-258).

Nowak, S., Mysłajek, R. W., & Kłosińska, A. (2011). Diet and prey selection of wolves (*Canis lupus*) recolonising Western and Central Poland. *Mammalian Biology, 76*, 709–715. https://doi.org/10.1016/j.mambio.2011.06.007

O'Keefe, F. R., Meachen, J., Fet, E. V., & Brannick, A. (2013). Ecological determinants of clinal morphological variation in the cranium of the North American gray wolf.

Journal of Mammalogy, 94(6), 1223–1236. https://doi.org/10.1644/13-MAMM-A-069

Oksanen, J., Simpson, G., Blanchet, F., Kindt, R., Legendre, P., Minchin, P., O'Hara, R., Solymos, P., Stevens, M., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., De Caceres, M., Durand, S., Evangelista, H., FitzJohn, R., Friendly, M., Furneaux, B., Hannigan, G., Hill, M., Lahti, L., McGlinn, D., Ouellette, M., Ribeiro Cunha, E., Smith, T., Stier, A., Ter Braak, C., & Weedon, J. (2024). vegan: Community ecology package (Version 2.7-0). https://github.com/vegandevs/vegan

Ollivier, M., Tresset, A., Bastian, F., Lagoutte, L., Axelsson, E., Arendt, M. L., Bălăşescu, A., Marshour, M., Sablin, M. V., Salanova, L., Vigne, J. D., Hitte, C., & Hänni, C. (2016). Amy2B copy number variation reveals starch diet adaptations in ancient European dogs. *Royal Society Open Science, 3*(11), 160449. https://doi.org/10.1098/rsos.160449

Olsen, L., Thum, E., & Rohner, N. (2021). Lipid metabolism in adaptation to extreme nutritional challenges. *Developmental cell*, *56*(10), 1417–1429. https://doi.org/10.1016/j.devcel.2021.02.024

Ortiz-Leal, I., Torres, M. V., López-Callejo, L. N., Fidalgo, L. E., López-Beceiro, A., & Sanchez-Quinteiro, P. (2022). Comparative Neuroanatomical Study of the Main Olfactory Bulb in Domestic and Wild Canids: Dog, Wolf and Red Fox. *Animals*, *12*(9), 1079. https://doi.org/10.3390/ani12091079

Osada, N., Mishra, S. K., Nakano, M., Tokoro, Y., Nagae, M., Doerksen, R. J., & Kizuka, Y. (2024). Self-regulation of MGAT4A and MGAT4B activity toward glycoproteins through interaction of lectin domain with their own N-glycans. *iScience, 27*(11), 111066. https://doi.org/10.1016/j.isci.2024.111066

Phelps, P. T., Anthes, J. C., & Correll, C. C. (2005). Cloning and functional characterization of dog transient receptor potential vanilloid receptor-1 (TRPV1). *European Journal of Pharmacology, 513*(1–2), 57-66. https://doi.org/10.1016/j.ejphar.2005.02.045

Pierotti, R. (2024). Domestication and human/wildlife mutualism. *Humans, 4*(4), 371-384. https://doi.org/10.3390/humans4040024

Pilot, M., Malewski, T., Moura, A. E., Grzybowski, T., Oleński, K., Kamiński, S., Fadel, F. R., Alagaili, A. N., Mohammed, O. B., & Bogdanowicz, W. (2016). Diversifying selection between pure-breed and free-breeding dogs inferred from genome-wide SNP analysis. *G3 Genes|Genomes|Genetics, 6*(8), 2285–2298. https://doi.org/10.1534/g3.116.029678

Pilot, M., Greco, C., vonHoldt, B., & et al. (2014). Genome-wide signatures of population bottlenecks and diversifying selection in European wolves. *Heredity, 112,* 428–442. https://doi.org/10.1038/hdy.2013.122

Pilot, M., Malewski, T., Moura, A. E., Grzybowski, T., & Oleński, K. (2015). On the origin of mongrels: Evolutionary history of free-breeding dogs in Eurasia. *Proceedings of the Royal Society B: Biological Sciences, 282*, 20152189. https://doi.org/10.1098/rspb.2015.2189

Piwowarska, D., & Kiedrzyńska, E. (2022). Xenobiotics as a contemporary threat to surface waters. *Ecohydrology & Hydrobiology, 22*(2), 337-354. https://doi.org/10.1016/j.ecohyd.2021.09.003

Plassais, J., Kim, J., Davis, B. W., & et al. (2019). Whole genome sequencing of canids reveals genomic regions under selection and variants influencing morphology. *Nature Communications, 10*, 1489. https://doi.org/10.1038/s41467-019-09373-w

Priya, A. K., Muruganandam, M., Kumar, A., Senthilkumar, N., Shkir, M., Pandit, B., Imran, M., Prakash, C., & Ubaidullah, M. (2024). Recent advances in microbial-assisted degradation and remediation of xenobiotic contaminants; challenges and future prospects. *Journal of Water Process Engineering, 60,* 105106. https://doi.org/10.1016/j.jwpe.2024.105106

Proust-Lima, C., Philipps, V., Diakite, A., & Liquet, B. (2023). lfmm: Extended mixed models using latent classes and latent processes (Version 2.1.0). https://cran.r-project.org/package=lcmm

Reshamwala, H.S., Mahar, N., Dirzo, R., & Habib, B. (2021). Successful neighbour: Interactions of the generalist carnivore red fox with dogs, wolves and humans for continued survival in dynamic anthropogenic landscapes. *Global Ecology and Conservation, 25*, e01446. https://doi.org/10.1116/j.gecco.2020.e01446

Rocha, J. L., Godinho, R., Brito, J. C., & Nielsen, R. (2021). Life in deserts: The genetic basis of mammalian desert adaptation. *Trends in Ecology & Evolution, 36*(7), 637–650. https://doi.org/10.1016/j.tree.2021.03.007

Ståhlberg, S., Bassi, E., Viviani, V., & et al. (2017). Quantifying prey selection of Northern and Southern European wolves (*Canis lupus*). *Mammalian Biology, 83*, 34–43. https://doi.org/10.1016/j.mambio.2016.11.001

Schlamp, F., van der Made, J., Stambler, R., Chesebrough, L., Boyko, A. R., & Messer, P. W. (2016). Evaluating the performance of selection scans to detect selective sweeps in domestic dogs. *Molecular Ecology, 25*(1), 342–356. https://doi.org/10.1111/mec.13485

Schweizer, R. M., Durvasula, A., Smith, J., Vohr, S. H., Stahler, D. R., Galaverni, M., Thalmann, O., Smith, D. W., Randi, E., Ostrander, E. A., Green, R. E., Lohmueller, K. E., Novembre, J., & Wayne, R. K. (2018). Natural selection and origin of a melanistic allele in North American gray wolves. *Molecular Biology and Evolution, 35*(5), 1190–1209. https://doi.org/10.1093/molbev/msy031

Schweizer, R. M., Robinson, J., Harrigan, R., Silva, P., Galverni, M., Musiani, M., Green, R. E., Novembre, J., & Wayne, R. K. (2016). Targeted capture and resequencing of

1040 genes reveal environmentally driven functional variation in grey wolves. *Molecular Ecology, 25*(1), 357–379. https://doi.org/10.1111/mec.13467

Schweizer, R. M., vonHoldt, B. M., Harrigan, R., Knowles, J. C., Musiani, M., Coltman, D., Novembre, J., & Wayne, R. K. (2016). Genetic subdivision and candidate genes under selection in North American grey wolves. *Molecular Ecology, 25*(1), 380–402. https://doi.org/10.1111/mec.13364

Selescu, T., Bivoleanu, R. A., Iodi Carstens, M., Manolache, A., Caragea, V. M., Hutanu, D. E., Meerupally, R., Wei, E. T., Carstens, E., Zimmermann, K., & Babes, A. (2024). TRPM8-dependent shaking in mammals and birds. *bioRxiv*. https://doi.org/10.1101/2023.12.27.573364

Sells, S. N., M. S. Mitchell, K. M. Podruzny, D. E. Ausband, D. J. Emlen, J. A. Gude, T. D. Smucker, D. K. Boyd, and K. E. Loonam. (2022). Competition, prey, and mortalities influence gray wolf group size. *Journal of Wildlife Management,* 86, e22193. https://doi.org/10.1002/jwmg.22193

Sinding, M.S., Gopalakrishnan, S., Ramos-Madrigal, J., de Manuel, M., Pitulko, V.V., Kuderna, L., Feurborn, T.R., Frantz, L.A.F., Vieira, F.G., Niemann, J., Samaniego Castruita, J.A., Carøe, C., Anderson-Ranberg, E.U., Jordan, P.D., Pavlova, E.Y., Nikolskiy, P.A., Kasparov, A.K., Ivanova, V.V., Willerslev, E., Skoglund, P., … Gilbert, M.T.P. (2020). Artic-adapted dogs emerged at the Pleistocene-Holocene transition. *Science (New York, NY) 368*(6498), 1495–1499. https://doi.org/10.1126/science.aaz8599

Storey, J.D., Bass, A.J., Dabney, A., & Robinson, D. (2025). *qvalue: Q-value estimation for false discovery rate control*. doi:10.18129/B9.bioc.qvalue, R package version 2.40.0, https://bioconductor.org/packages/qvalue

Sun, C., Southard, C., Witonsky, D. B., Kittler, R., & Di Rienzo, A. (2010). Allele-specific down-regulation of RPTOR expression induced by retinoids contributes to climate adaptations. *PLoS Genetics, 6*(10), e1001178. https://doi.org/10.1371/journal.pgen.1001178

Stronen, A. V., Jędrzejewska, B., Pertoldi, C., Demontis, D., Randi, E., Niedziałkowska, M., Borowik, T., Sidorovich, V. E., Kusak, J., Kojola, I., Karamanlidis, A. A., Ozolins, J., Dumenko, V., & Czarnomska, S. D. (2015). Genome-wide analyses suggest parallel selection for universal traits may eclipse local environmental selection in a highly mobile carnivore. *Ecology and Evolution, 5*(19), 4410–4425. https://doi.org/10.1002/ece3.1695

Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., Gable, A. L., Fang, T., Doncheva, N. T., Pyysalo, S., Bork, P., Jensen, L. J., & von Mering, C. (2023). The STRING database in 2023: Protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Research, 51*(D1), D638–D646. https://doi.org/10.1093/nar/gkac1000

Tancredi, D., & Cardinali, I. (2023). Being a dog: A review of the domestication process. *Genes, 14*(5), 992. https://doi.org/10.3390/genes14050992

Takeuchi, H., Takeuchi, T., Gao, J., Cantley, L. C., & Hirata, M. (2010). Characterization of PXK as a Protein Involved in Epidermal Growth Factor Receptor Trafficking. *Molecular and Cellular Biology*, *30*(7), 1689–1702. https://doi.org/10.1128/MCB.01105-09

Tallian, A., Ciucci, P., Milleret, C., Smith, D., Stahler, D., Wikenros, C., & Ordiz, A. (2023). Wolves in a human world: Social dynamics of the Northern Hemisphere's most iconic social carnivore. In A. Tallian & P. Ciucci (Eds.), Social strategies of carnivorous mammalian predators (pp. 89-138). Springer International Publishing

Taruno, A., & Gordon, M. D. (2023). Molecular and Cellular Mechanisms of Salt Taste. *Annual review of physiology*, *85*, 25–45. https://doi.org/10.1146/annurev-physiol-031522-075853

Tazerji, S. S., Elahinia, A., Akhtardanesh, B., Kabir, F., Vazir, B., Duarte, P. M., Haji pour, P., Rehman, A., Ilyas, M. F., & Hassanzadeh, S. (2024). Nutritional risks and consequences of meat-only diets for dogs and cats. *German Journal of Veterinary Research, 4*(1), 62-76. https://doi.org/10.51585/gjvr.2024.1.0076

The UniProt Consortium. (2025). UniProt: The universal protein knowledgebase in 2025. *Nucleic Acids Research, 53*(D1), D609–D617. https://doi.org/10.1093/nar/gkae1010

Uffelmann, E., Huang, Q. Q., Munung, N. S., & et al. (2021). Genome-wide association studies. *Nature Reviews Methods Primers, 1*, 59. https://doi.org/10.1038/s43586-021-00056-9

Vanak, A.T. and Gompper, M.E. (2009). Dogs *Canis familiaris* as carnivores: their role and function in intraguild competition. *Mammal Review, 39*, 265-283. https://doi.org/10.1111/j.1365-2907.2009.00148.x

Wadgymar, S. M., DeMarche, M. L., Josephs, E. B., Sheth, S. N., & Anderson, J. T. (2022). Local adaptation: Causal agents of selection and adaptive trait divergence. *Annual Review of Ecology, Evolution, and Systematics, 53*(1), 87–111. https://doi.org/10.1146/annurev-ecolsys-012722-035231

Wang, G. D., Fan, R. X., Zhai, W., Liu, F., Wang, L., Zhong, L., Wu, H., Yang, H. C., Wu, S. F., Zhu, C. L., Li, Y., Gao, Y., Ge, R. L., Wu, C. I., & Zhang, Y. P. (2014). Genetic convergence in the adaptation of dogs and humans to the high-altitude environment of the Tibetan plateau. *Genome Biology and Evolution, 6*(8), 2122–2128. https://doi.org/10.1093/gbe/evu162

Wang, J., Zhang, Q., Fan, W., Shi, Q., Mao, J., Xie, J., Chai, G., & Zhang, C. (2025). Deciphering olfactory receptor binding mechanisms: A structural and dynamic perspective on olfactory receptors. *Frontiers in Molecular Biosciences, 11*, 1498796. https://doi.org/10.3389/fmolb.2024.1498796

Wei, Q., Dong, Y., Sun, G., Wang, X., Wu, X., Gao, X., Sha, W., Yang, G., & Zhang, H. (2021). *FGF* gene family characterization provides insights into its adaptive evolution in Carnivora. *Ecology and evolution*, *11*(14), 9837–9847. https://doi.org/10.1002/ece3.7814

Wei, T., & Simko, V. (2024). R package 'corrplot': Visualization of a correlation matrix (Version 0.94). https://github.com/taiyun/corrplot

Wheeldon, T. J., & Patterson, B. R. (2012). Genetic and morphological differentiation of wolves (Canis lupus) and coyotes (Canis latrans) in northeastern Ontario. *Canadian Journal of Zoology, 90*(10), 1221–1230. https://doi.org/10.1139/z2012-090

White, N. J., & Butlin, R. K. (2021). Multidimensional divergent selection, local adaption and speciation. *Evolution; international journal of organic evolution, 75*(9), 2167–2178. https://doi.org/10.1111/evo.14312

Wierzbowska, I. A., Hędrzak, M., Popczyk, B., Okarma, H., & Crooks, K. R. (2016). Predation of wildlife by free-ranging domestic dogs in Polish hunting grounds and potential competition with the grey wolf. *Biological Conservation, 201*, 1-9. https://doi.org/10.1016/j.biocon.2016.06.016

Wu, X., Yan, R., Cao, P., Qian, H., & Yan, N. (2022). Structural advances in sterol-sensing domain-containing proteins. *Trends in Biochemical Sciences, 47*(4), 289–300. https://doi.org/10.1016/j.tibs.2021.12.005

Wynne, C. D. L. (2021). Dogs' (Canis lupus familiaris) behavioral adaptations to a human-dominated niche: A review and novel hypothesis, *Advances in the Study of Behavior 53*, 97–162. https://doi.org/10.1016/bs.asb.2021.03.004

Xiao, C., Wang, Y., Cheng, Q., & Fan, Y. (2022). Increased expression of MMP17 predicts poor clinical outcomes in epithelial ovarian cancer patients. *Medicine, 101*(34), e30279. https://doi.org/10.1097/MD.0000000000030279

Yi, X., Qiu, Y., Xie, P., Wang, S., & Sun, X. (2025). Molecular characterization and complexity of the immunoglobulin repertoire in the silver-black fox (*Vulpes vulpes*). *BMC veterinary research*, *21*(1), 214. https://doi.org/10.1186/s12917-025-04676-1

Yoshikawa, K., Deguchi, J., Hu, J., Lu, H. Y., & Matsunami, H. (2022). An odorant receptor that senses four classes of musk compounds. *Current Biology, 32*(23), 5172–5179.e5. https://doi.org/10.1016/j.cub.2022.10.038

Zhang, Z., Wang, Z., Liu, T., et al. (2024). Exploring the role of ITGB6: fibrosis, cancer, and other diseases. *Apoptosis, 29*, 570–585. https://doi.org/10.1007/s10495-023-01921-6

Zhang, S. J., Wang, G. D., Ma, P., Zhang, L. L., Yin, T. T., Liu, Y. H., Otecko, N. O., Wang, M., Ma, Y. P., Wang, L., Mao, B., Savolainen, P., & Zhang, Y. P. (2020). Genomic regions under selection in the feralization of the dingoes. *Nature Communications, 11*(1), 671. https://doi.org/10.1038/s41467-020-14515-6

Zhang, Y., Zhang, Y., Sun, K., Meng, Z., & Chen, L. (2019). The SLC transporter in nutrient and metabolic sensing, regulation, and drug development. *Journal of molecular cell biology, 11*(1), 1–13. https://doi.org/10.1093/jmcb/mjy052

Zhong, H., Atzeni, L., Li, F., Liu, Y., Wang, R., Alemiti, A., & Shi, K. (2024). Trophic interaction and livestock dependence of snow leopard and sympatric carnivores in Tianshan, Northwest China. https://doi.org/10.21203/rs.3.rs-4539773/v1

Chapter 4 | South-north decline in adaptive potential of free-ranging dogs points to dog domestication in southern Eurasia

Contents

Abstract **128**

1. Introduction **129**

2. Material and Methods **131**

2.1. Dataset and gene ontology analysis 131

2.2. Genomic Offset analysis 131

2.3. Species distribution and suitable niche modelling 133

2.4 Unhoused dogs' population density estimation 133

3. Results **134**

3.1. Genomic offset and risk of maladaptation 134

3.2 Suitable ecological habitat modelling 135

4. Discussion **138**

4.1. Climate change and the risk of maladaptiveness 138

4.2. Dogs' suitable habitats are semi-arid, sub-tropical and tropical 143

4.3. Feralization of dogs in semi-arid, sub-tropical and tropical habitats 144

5. Conclusions **145**

References **146**

Abstract

It is well established that domestic dogs (*Canis lupus familiaris*) originate from grey wolves (*Canis lupus*), but the geographic origin of the wolf population(s) ancestral to dogs remains elusive. Given that the expansion of early dogs from the place of their origin was very fast in the evolutionary time scale and was associated with the anthropogenic niche, it may be expected that dogs may retain adaptations to the natural environment where they first emerged. In this study we investigated the adaptive potential under environmental shifts in 370 Eurasian free-ranging dogs (FRDs). Additionally, we modelled dogs' suitable habitats in Eurasia based on the distribution of 79564 dogs. Genomic offset analysis showed that FRDs' adaptive potential is highest in the southernmost parts of Eurasia and declines along a latitudinal gradient. Accordingly, dogs' suitable habitat is predicted to be in southernmost parts of Eurasia. These results imply that domestic dogs have a higher capability for independent survival in the southern regions of the continent. This is consistent with higher densities of FRDs in southern latitudes as well as the occurrence of truly feral populations only in warm climates. Taken together, these results suggest that early dogs originated from southern parts of Eurasia.

Keywords: Free-ranging dogs, adaptive potential, domestication, genomic offset, suitable habitat modelling

1. Introduction

The domestication of dogs (*Canis lupus familiaris*) from grey wolves (*Canis lupus*) has been the focus of numerous studies, yet the geographic origin of the first domesticated population(s) remains elusive (Bergström et al. 2022, Tancredi & Cardinali, 2023). The wolf population ancestral to domestic dogs either became extinct (Freedman et al., 2014; Bergström et al., 2020) or its genetic composition substantially changed over time due to admixture with other wolf populations (Bergström et al. 2022). This hinders the inference of the domestication centre based on genetic data from modern wolves. One solution to this problem is the genomic analysis of ancient wolves from times contemporary with the early domestication period (Bergström et al. 2022). However, the timing of domestication is not precisely known, and potential domestication regions pointed to by different studies encompass nearly entire Eurasia (reviewed in Tancredi & Cardinali, 2023), which makes such study difficult, given the scarcity of ancient wolf samples and the effort required to obtain genomic data from ancient DNA. Archaeological studies, which could provide information required to narrow down both the timing and location of domestication, brought ambiguous results to date (Tancredi & Cardinali, 2023). Therefore, other methods to infer the domestication place and time are needed.

Dog domestication occurred recently in the evolutionary timescale, with recent estimates pointing to 16-40 thousand years ago (Tancredi & Cardinali, 2023), and the global spread of dog populations from the region of origin occurred even later, probably in association with the Neolithic revolution (reviewed in Sykes et al., 2020; Zhang et al., 2020). Therefore, the time to acquire adaptations to a broad range of environmental conditions during the dog expansion was very limited, which can result in the differentiation of the adaptive potential of dog populations depending on how different local climatic conditions are from those in the domestication region.

Although ancient DNA studies demonstrated that genetic analyses of modern dogs cannot provide sufficient information to reconstruct the domestication process (Bergström et al., 2020; Bergström et al.., 2022; Larson et al., 2012; Thalmann et al., 2013), modern domestic dogs can provide valuable data to inform domestication studies. For example, comparative studies on the cognition and behavioural ecology of dogs and grey wolves have suggested that both have the same capacity to learn (Rivas-Blanco et al., 2024), however wolves are more aggressive and vigilant towards humans (Amici et al., 2024; Lazzaroni et al., 2020). This alludes to the behavioural changes dogs underwent during their domestication.

Genomic studies have also highlighted how selection in behavioural genes was important during the domestication of dogs (Pendleton et al., 2018), and consequently their divergence from wolves. While there is a growing number of studies on behavioural ecology of free-ranging dogs (FRDs) (Vanak & Gompper, 2009; Lazarroni et al., 2020; Range & Marshall-Pescini, 2022), knowledge of other aspects of their ecology remains more limited (Wheat & Wynne, 2025; Vanak and Gompper, 2009). Although FRDs' reliance on human waste has been well described (Belo et al., 2015; Belsare & Vanak, 2020; Bryce, 2021; Hussain et al., 2021; Mahajan et al., 2024), their interactions with ecological resources (e.g., wildlife), particularly in the context of natural

landscapes has not been well described yet. Despite limited knowledge of their ecology, their ecological success cannot be understated.

Dogs are distributed nearly worldwide which subjects them to a wide range of environmental conditions. Approximately 75% of dogs globally are free-ranging, i.e. are not confined by humans, relying instead on their abilities to access resources from urban, rural, and natural landscapes (Gompper, 2014). With few exceptions, these FRDs are dependent on anthropogenic food sources, although some rural populations supplement their diet by hunting wildlife (Duarte et al., 2016; Hughes et al., 2013; Hussain et al., 2021; Lino et al., 2023; Wierzbowska et al., 2016). Unlike pure-bred dogs, FRDs are not restricted by humans in their reproduction and therefore can breed freely. Even though they can interact and interbreed with pure-bred dogs (Belo et al., 2015; Belsare & Vanak 2020), this is relatively rare. FRDs are not an admixture of breeds, but a genetically distinct group that preceded the breed formation process (Pilot et al. 2015). Phylogeographic patterns inferred for FRDs (Pilot et al. 2015, Shannon et al., 2015; Zhang et al., 2024) are consistent with those based on ancient DNA (Frantz et al., 2016; Bergström et al. 2020), suggesting the continuity of lineages from early dogs to modern free-ranging populations.

Given the limited reliance on humans (except the dependence on anthropogenic food) and the ability to breed freely, FRDs are exposed to environmental pressures alike those faced by wild canids (Belsare & Vanak, 2020; Bryce, 2021), as well as intraspecific interactions associated with mate competition and mate choice (Cafazzo et al., 2010; Cafazzo et al., 2014; Natoli et al. 2021). As a result, they are subject to natural and sexual selection (Pilot et al. 2016). Their distribution is considerably larger compared with grey wolves, whose original distribution encompassed the entire Eurasia and North America (Wolf & Ripple, 2017). Wolves inhabit diverse biomes—from arid deserts to frozen tundra—which demonstrates their significant adaptive potential (Boitani et al., 2023; Schweizer et al., 2016). However, their southern distribution range reaches the tropical regions only in Asia, where the southern edge of the range occurs in regions with hot, humid climates (Gonzalez Bernal et al., 2022; Wolf & Ripple, 2017), implying the presence of limiting factors such as competition with other large carnivores or lack of physiological adaptations to the tropical climate.

In contrast, domestic dogs are nearly ubiquitous worldwide, implying an ability to adapt to a broader range of climates. Dogs' distribution and population density of dogs are intricately linked to that of humans (Perri et al., 2021; Gompper, 2014) and are direct consequences of their ecological niche as a human commensal. This ecological niche is associated with reduced dependence on food accessibility and quantity in natural landscapes. Nevertheless, dogs are exposed to local environmental conditions, and therefore their survival may rely on specific local adaptations. However, there is not much known about FRDs local adaptations in natural landscapes. Previous studies on dogs living in high altitude environments and African indigenous dogs identified adaptations associated with hypoxia tolerance (Gou et al. 2014; Li et al. 2014) and tropical environments (e.g., tropical parasite resistance, immunity, angiogenesis, ultraviolet protection and insulin metabolism) (Liu et al. 2018). Except for these studies,

focused on individual regional populations, knowledge on the environmental adaptations in dogs is limited.

Domestic dogs are a relatively recent lineage and therefore identifying local differentiation in the adaptive potential of FRD populations could inform on their domestication origin. In this study we aim to better understand how dogs have adapted to their ecological niche and make inferences of both their domestication origin and how they may cope under climate change. To that end, we investigate the adaptive potential of both FRDs and their wild ancestor, grey wolves, under four predicted pathways of environmental shifts from 2021-2100. Additionally, we estimate dogs' suitable ecological habitats under current and future environmental conditions. Based on earlier studies on dogs (Liu et al. 2018), we predict that their adaptive potential would vary between habitats, with the habitats showing the highest adaptive potential representing the environmental conditions most similar to those experienced by early domestic dogs. Also, we expected that most suitable habitats would be similar to those in which dogs were fist domesticated in.

2. Material and Methods

2.1. Dataset

The data used for this analysis originates from the previous study focused on comparative landscape genomics of Eurasian FRDs and grey wolves (Chapter 3). The original dataset comprised 73,360 SNP quality-controlled genotypes for FRDs (N=370) from across Eurasia, as described in Chapter 2. The individuals were grouped by biomes and represented 8 biomes from deserts to tundra (Table 1). For comparative purposes, we also analysed SNP genotypes from grey wolves (N=232) from 10 biomes with the same quality control methods. In this study we analysed a subset of SNP variants, 286 SNP variants of FRD and 288 wolves linked to environmental factors (e.g., temperature and precipitation). The SNPs used for the genomic offset analysis, were detected to be associated with environmental variables (Chapter 3), in a dual-approach analysis, LFMM and RDA, with the R packages lfmm v1.1 (Proust-Lima et al., 2023) and vegan v2.6-6.1 (Oksane et al., 2024).

2.2. Genomic Offset analysis

To investigate adaptive potential of FRDs, this study incorporates genomic offset statistics which blend genomic-environmental association (GEA) data with predicted future environmental scenarios (Gain et al., 2021), under four different socioeconomic pathways (SSPs). These pathways range from a sustainable world scenario (SSP1-2.6) to a fossil fuel-intensive scenario (SSP5-8.5), which aids in modelling the adaptive responses of populations to climate changes (Riahi et al., 2017). Genomic offset analysis was conducted to measure the disparity between the current genomic composition at loci with signatures of local adaptive variation and the genomic composition required under predicted future environmental conditions (i.e., genetic gap) (Rellstab et al., 2021). We utilized the R package LEA v3.20 (Frichot, & François, 2015), specifically the genetic.gap function, to calculate these genomic offset statistics. This function was applied with scaling modifiers and targeted the candidate loci, which were the SNPs

identified as showing signatures of local adaptation (Frichot & François, 2015). The latent factors, based on our previous population structure analysis, were set at K=8 for wolves and K=5 for FRDs (Chapter 2).

Table 1. Distribution of sampled individuals by biome

Biomes	**FRDs**	**Wolves**
Boreal Forests/Taiga	18	7
Deserts & Xeric Shrublands	60	20
Flooded Grasslands & Savannas	0	1
Mediterranean Forests, Woodlands & Scrub	83	29
Montane Grasslands & Shrublands	0	4
Temperate Broadleaf & Mixed Forests	165	129
Temperate Conifer Forests	6	4
Temperate Grasslands, Savannas & Shrublands	17	34
Tropical & Subtropical Moist Broadleaf Forests	17	2
Tundra	4	2

Table 2. List of WorldClim predictors retained for FRDs and grey wolves, respectively.

Abrev	**Predictor**	**Taxa**
BIO2	Mean Diurnal Range (Mean of monthly (max temp - min temp))	Both
BIO3	Isothermality (BIO2/BIO7) (×100)	Wolves
BIO7	Temperature Annual Range (BIO5-BIO6)	Both
BIO8	Mean Temperature of Wettest Quarter	Both
BIO9	Mean Temperature of Driest Quarter	Wolves
BIO13	Precipitation of Wettest Month	FRDs
BIO14	Precipitation of Driest Month	FRDs
BIO16	Precipitation of Wettest Quarter	Wolves
BIO18	Precipitation of Warmest Quarter	Wolves
BIO19	Precipitation of Coldest Quarter	Both

Our analysis also incorporated the concept of the risk of non-adaptiveness (RONA), a measure of the required changes in SNP frequencies under new environmental conditions (i.e., environmental distance), as proposed by Rellstab et al. (2021). RONA calculations in LEA include adjustments for confounding factors and enable multivariate analysis, enhancing the reliability of our findings. BioClim variables (Table 2) were used in this study due to the lack of future projections for the other environmental variables. Environmental variable rasters were processed with the terra v1.7-81 (Hijmans, 2024) R package, including extraction of environmental values from current and future environmental variables. The future environmental variables used in this analysis were for the non-collinear BioClim variables (Table 2) across four Shared Socioeconomic Pathways (SSPs): 126, 245, 370, and 585 from the time frame 2021-2100. These predictions were sourced from the CMIP6 GMC MRI-ESM2-0 scenarios and downloaded as TIFF files from the WorldClim website (Yukimoto et al., 2019).

2.3. Species distribution and suitable niche modelling

To understand how environmental conditions may impact the distribution of FRDs under current conditions and future climate change, we performed species distribution modelling (SDM) and suitable niche modelling using the R package Biomod2 (Guéguen et al., 2025). The presence data used to train our SDM was the occurrence data for dogs (N=79, 564) downloaded from Global Biodiversity Information Facility (GIBF) (Fig. 1). The data formatted for pseudo-absence parameters were based on software manual recommendations, using a random strategy with 4 repeats, 1000 pseudo absence points per repetition and removal of any NA value. The environmental layers used for formatting the data were the same ones used for the genomic offset analysis. To avoid data loss, the presence was not thinned, considering that it has been shown that thinning can negatively affect modelling performance (Caten & Dallas, 2023).

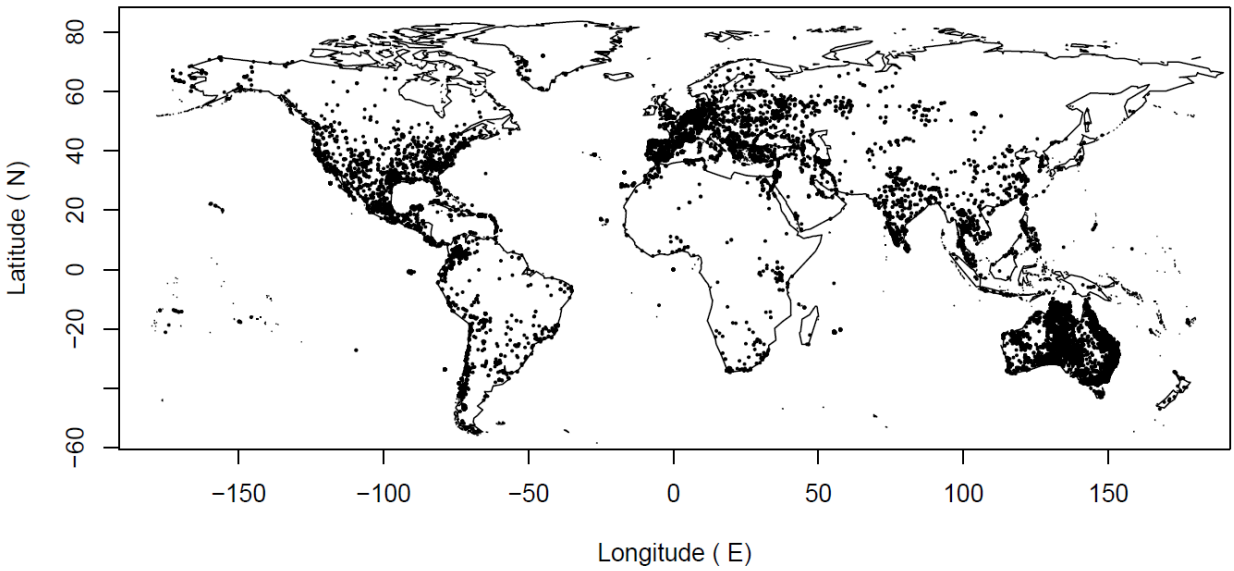


Figure 1. Distribution of dogs' occurrence data. Map with the distribution of occurrence datapoints used for informing the presence data for training the species distribution model.

The suitable ecological niche/habitat models were built based on a random cross-validation strategy, 2 repetitions, data split for training the models at 0.8, 0.5 species prevalence in the training data and the model evaluation metrics area under the receiver operating characteristic (ROC) curve and true skill statistic (TSS). We created ensemble models, i.e., combined predictions, based on the top three model algorithms with the highest ROC and KAPPA ratios (Fig. S3); which were random forest (RF), extreme gradient boosting (XGBOOST) and maximum entropy (MAXENT) (Guéguen et al., 2025; Yackulic et al., 2012). The ensemble modelling was done using the same metric evaluations, with 3 permutations, a 0.05 alpha value for the confidence interval and models which passed a TSS threshold of 0.7; final values ranged from 0.703-0.784. The use of these parameters in par with ensemble modelling, allowed us to reduce the impact of sampling bias (Dubos et al., 2022) and create a more robust prediction.

Ensemble model was used for past and future projections of suitable ecological habitats. Future projections were based on the 2081-2100 timeframe for SSP 585 environmental rasters. This time frame and SSP were informed by the genomic offset analysis, to reduce computing power required to finalize the projections. The past projections were

done with the environmental rasters from mid-Holocene. This timeframe was the furthermost in the past, which was available in WorldClim at the necessary resolution (i.e., 30 arc seconds) for the analysis. Projections for all timeframes, past, current and future climatic conditions (Fig. 4) were done with the same BioClim environmental rasters as used in the genomic offset (see above). The changes in range size estimations (i.e., gain/loss in pixels) were performed with the ensembled projections with binary metric based on TSS.

2.4 Unhoused dogs' population density estimation

Unhoused dog density estimates were calculated to detect patterns related to the relationships between unhoused dog densities and ecological conditions across the world. Densities were calculated using the following formula: D = P/A, where D is the dog population density estimate, P represented the estimated unhoused dog population per country and A = the area in km^2 of the respective countries. The unhoused dog per thousand people estimates were extracted from the WellBeing International Global Dog Campaign (GDC). In the GDC dataset, the human population estimates were attained from the World Health Organization from the 1st of January 2023 human population estimates by country in thousands. These values were used to calculate the unhoused dog populations per country, by multiplying the unhoused dogs per thousand people estimates by the human population estimates in thousands. The areas of the countries were extracted from the World Bank Land area (sq. km) data set from 2022. Estimates of unhoused dog population density were calculated only for the countries for which the data was available in the WellBeing International dataset.

Unhoused dog densities were grouped by land mass (i.e. Eurasia, Africa, Oceania and America) and then by regions informed by the countries' biomes per the Resolve Ecoregions 2017 map (Dinerstein et al., 2017). Grouping of the countries was done by landmass then dominant biomes of each of the countries, i.e., tundra, boreal forests/taiga, temperate broadleaf and mixed forests, temperate grasslands, savannas and shrublands, montane grasslands and shrublands, tropical and subtropical moist broadleaf forests, mediterranean forests, woodlands and scrub, deserts and xeric shrublands, tropical and subtropical grasslands, and savannas and shrublands (Table 3). Geopolitical divisions of countries and regions do not reflect ecological conditions in the encompassed areas, but we could not obtain data on a smaller spatial scale than per country. Defining the regions by dominant biomes allowed for the most efficient representation of the ecological conditions unhoused dogs experience within each country.

3. Results

3.1. Genomic offset and risk of maladaptation

FRDs showed modest increases in genomic offset over time, particularly in individuals from Eastern European populations and two individuals from the Indian Peninsula (Fig, S1, Fig. 2A). Most Southern Eurasian FRDs showed low genomic offset (>0.1) across all time frames and SSPs, low values of genomic offset imply high adaptive potential (Gain et al., 2023). Thus, FRDs showed higher adaptive potential in Southern Eurasia under all

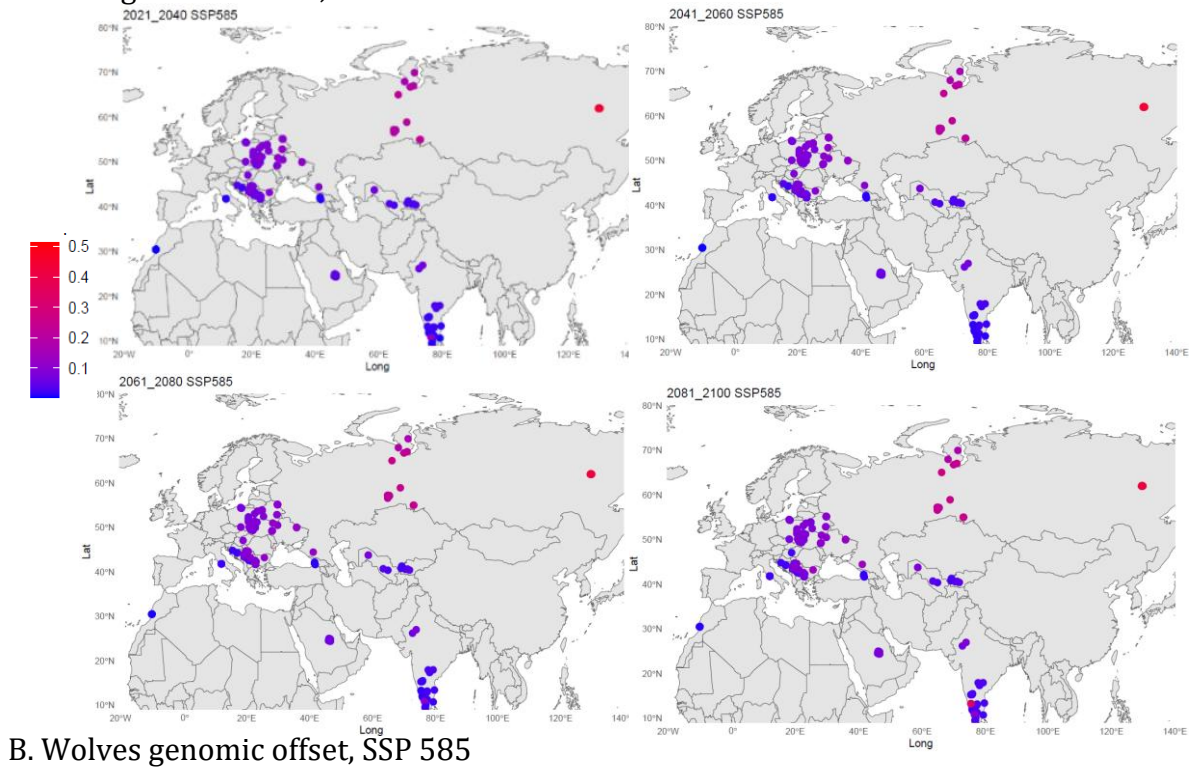
predictions of future climate change. FRDs from Northern Asia consistently displayed higher values of offset (i.e., low adaptive potential) across all timeframes and SSPs. These patterns contrasted with those observed in wolves, which displayed wider increases in genomic offset over time, with the most substantial increase predicted to occur under SSPs 370 and 585, particularly in Northern Eurasia (Fig. S2). Notably, except for a few individuals from the Indian Peninsula, genomic offset increased for most wolves from 2061 onwards under the 370 and 585 SSPs (Fig. S2). The northernmost Eurasian wolves showed the highest values of genomic offset (2-8) within the 2081-2100 period under the 585 SSP (Fig. 2B). Thus, Northern Eurasian wolves showed the lowest adaptive potential, whereas Southern Eurasian wolves showed the highest adaptive potential.

The patterns observed in the RONA analysis paralleled those of genomic offset; regions with increases in genomic offset also experienced rises in RONA (Figs. S1-S2, Figs. 3A, 4B). In FRDs, the range of the risk of maladaptiveness was consistently between 0.1 and 0.4, with some increases noted in individuals from the Indian Peninsula, although the values largely remained stable throughout the timeframes (Fig. S1, Fig. 3B). FRDs with the lowest values of RONA were from Southern Eurasia. In wolves, the range of RONA across most SSPs varied from 0.025 to 0.125, but in the 585 SSP during the 2081-2100 timeframe, the changes were more pronounced (0-2) (Fig. S2, Fig. 3B). RONA increased for nearly all wolves across Eurasia, except for a few from the Indian Peninsula, where it remained in the lower range. Considering that the RONA is a measure for the risk of maladaptiveness, overall FRDs showed higher risk of maladaptiveness than grey wolves but lower change over time.

3.2 Suitable ecological habitat modelling

Dogs' distribution based on the available presence data from GIBF (Fig. 1) had more data points in Europe, and records in Central Asia were sparse. This dataset is based on records uploaded by individuals from various institutions, hence there is an inherent sampling bias. In regions where there is limited access or interest to upload data points the data is sparse (e.g., Central Asia). However, there are presence points for most of Eurasia hence this dataset is better suited to capture dogs' distribution in the continent than the distribution based on the collection of the FRDs samples used in this study. Furthermore, by using ensemble modelling approach we reduced the impact of sampling bias and created a robust model for the analysis without sacrificing presence points (Dubos et al., 2022). Precipitation of coldest quarter (Bio19) was the variable with the highest importance (40-60%) across all the species distribution models from (Fig. S3). We built an ensemble species distribution model based on RF, XGBOOST and MAXENT models. This model had excellent sensitivity (ROC ~0.95) and was accurate in predicting the species distribution (KAPPA ~ 0.7) (Fig. S4).

A. FRDs genomic offset, SSP 585



B. Wolves genomic offset, SSP 585

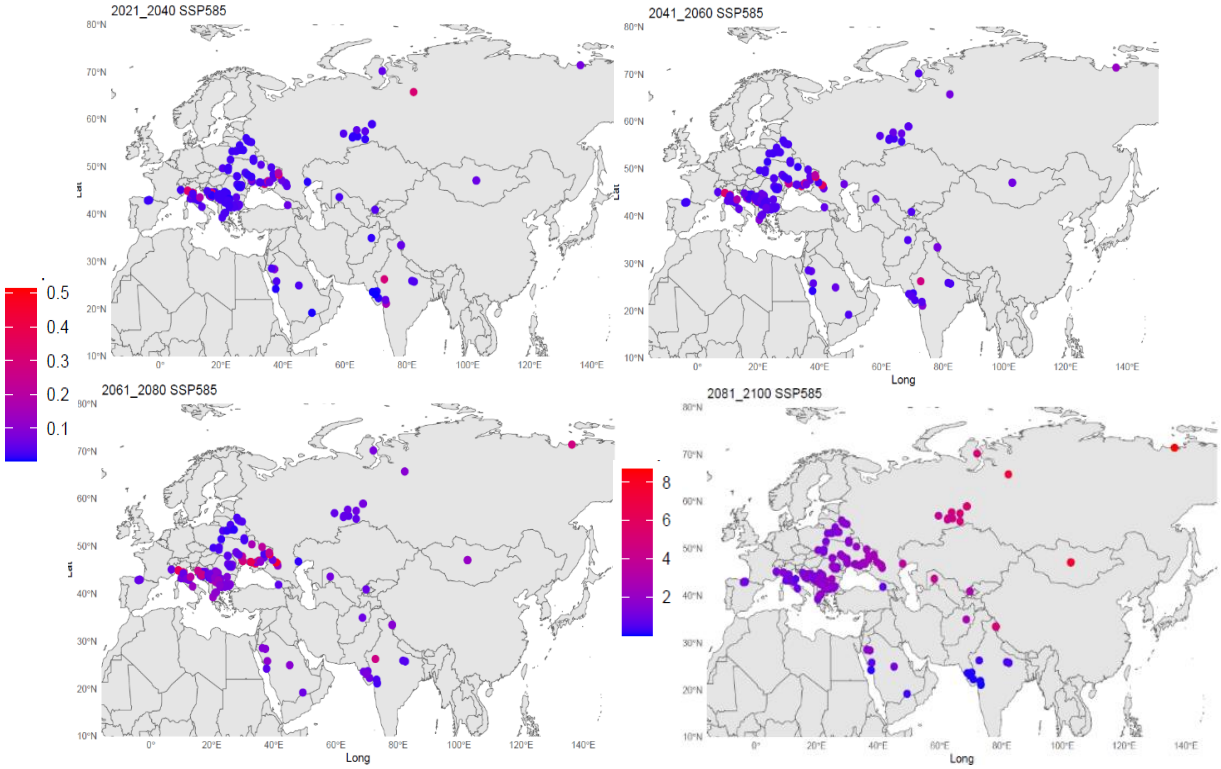
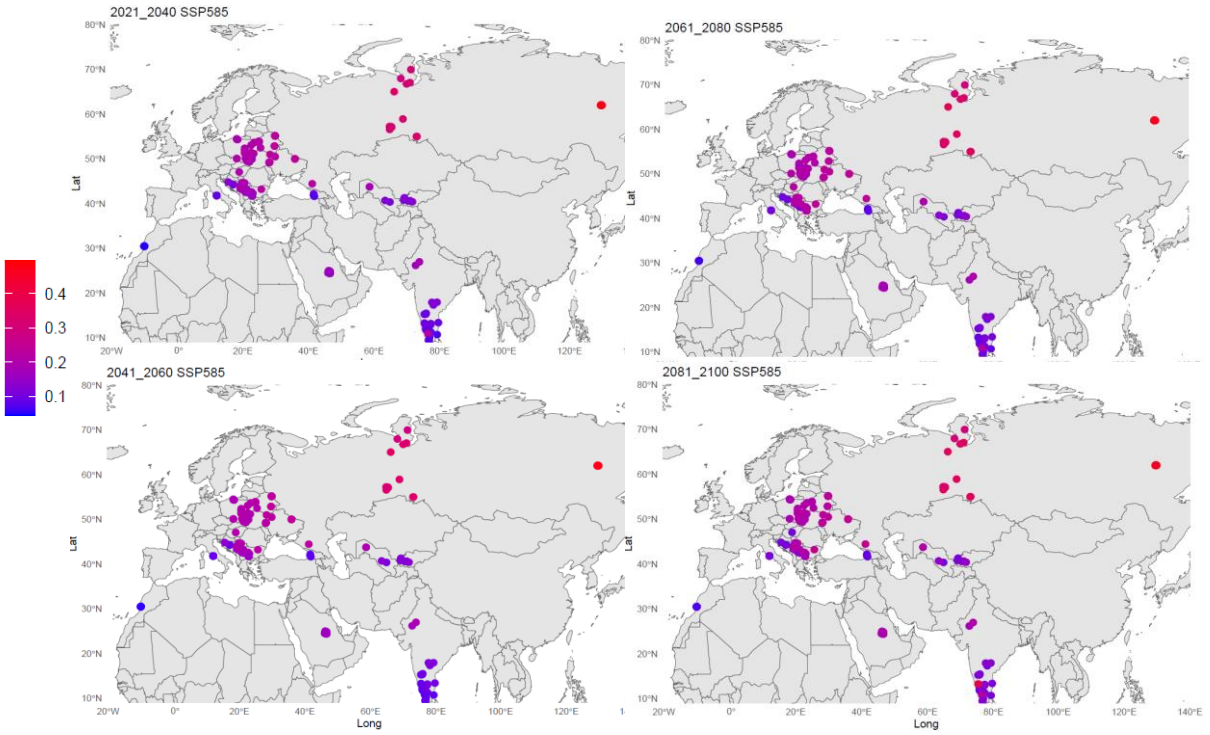


Figure 2. Genomic offset mapped results for SSP 585 in FRDs and wolves. A. FRDs' genomic offset for the SSP 585, time frames from 2021-2100. Genomic offset in FRDs stayed relatively constant,across all SSPs i.e., 126, 245, 370 and 585. B. Wolves' genomic offset for the SSP 585, time frames from 2021-2100. Eurasian grey wolves showed the highest increase in genomic offset (i.e., decrease in adaptive potential) and risk of nondaptiveness at the 2081-2100 time frame under the 585 SSP.

A. FRDs genomic distance (RONA), SSP 585



B. Wolves genomic distance RONA, SSP

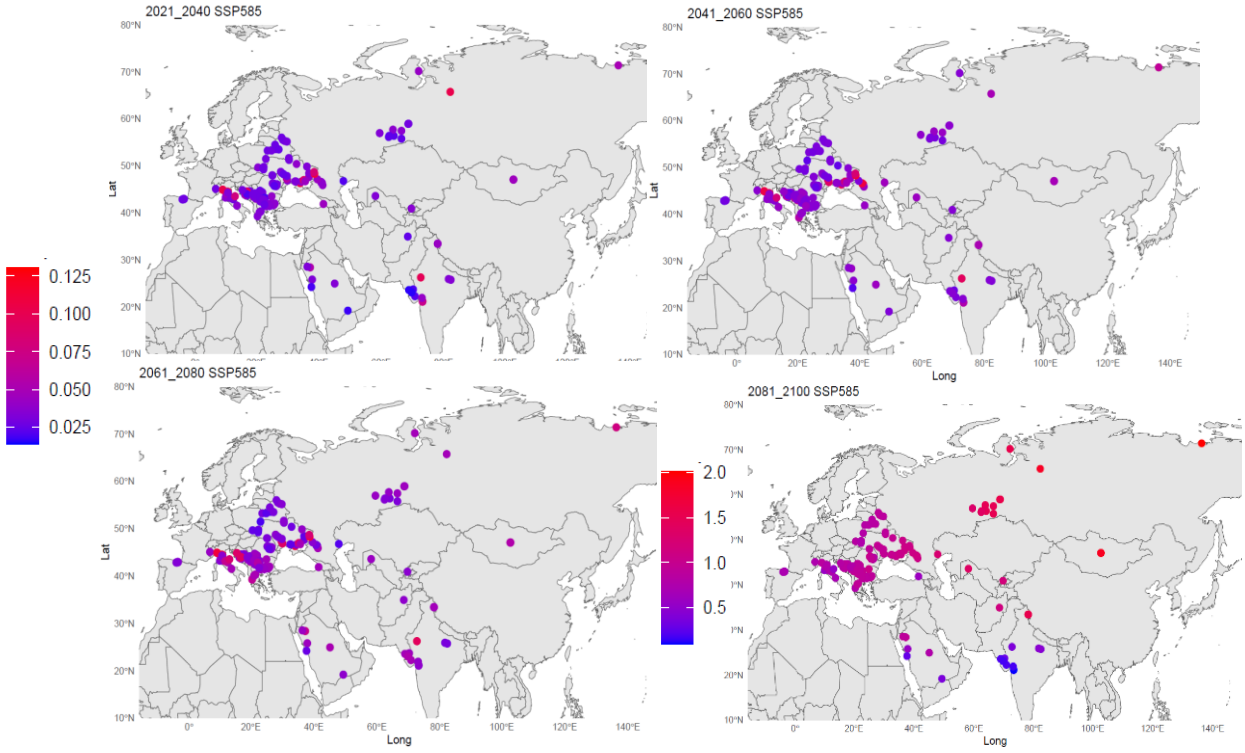


Figure 3. Genomic distance (RONA) mapped results for SSP 585 in FRDs and wolves. A. FRDs' genomic distance (RONA) for the SSP 585, 2021-2100 time frames. Similar to genomic offset patterns, FRDs genomic distance stayed relatively constant across SSPs. B. Wolves' genomic distance (RONA) for the SSP 585,2021-2100 time frames. Eurasian grey wolve showed a sharp increase in genomic distance in the 585 SSP within the 2061-2100 time frames. Wolves showed moderate increase in genomic distance across the remaining SSPs, i.e., 126, 245, 370.

Based on the calibrated species model under the environmental conditions from ~6000 years ago to the future i.e., 2100, Europe and Southern Asia consistently show the most suitable habitats for the species (Fig. S5-S9). Dogs' suitable habitats were sparse ~6000 years ago (mid-Holocene) with moderate suitability values in Europe and higher values in Southern Asia (Fig. 4A, Fig. S6). Under current environmental conditions the highest values of habitat suitability are in Europe and Southern Asia (Fig. 4B, Fig. S5). Furthermore, dogs' most suitable ecological habitat is predicted to expand under environmental shifts based on the fossil-fuelled development (SSP 585) prediction models (Fig. 5A, Fig. S7). The exception in this pattern of expansion was in Northern and Eastern Europe where the habitat suitability for dogs is predicted to decline (Fig 5B). Based on the ensemble models, dogs' suitable ecological habitat has been continuously expanding since mid-Holocene (Figs. S8-S9). Unhoused dog estimates paralleled dogs' suitable ecological niches, with the highest densities predicted in warmer biomes of Eurasia (i.e., tropical and subtropical most broadleaf forests, montane grasslands & shrublands) specifically within Southeast Asia, and Southern Asia, where the dominant biome is deserts and xeric shrublands (Table 3). Countries with the highest densities were in Southern Asia, i.e., Philippines, India and Pakistan.

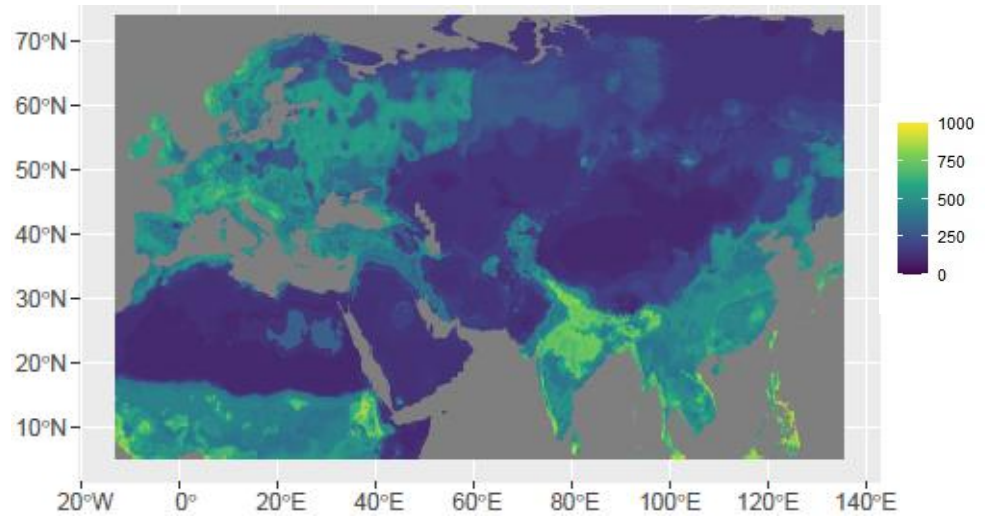
4. Discussion

4.1. Climate change and the risk of maladaptiveness

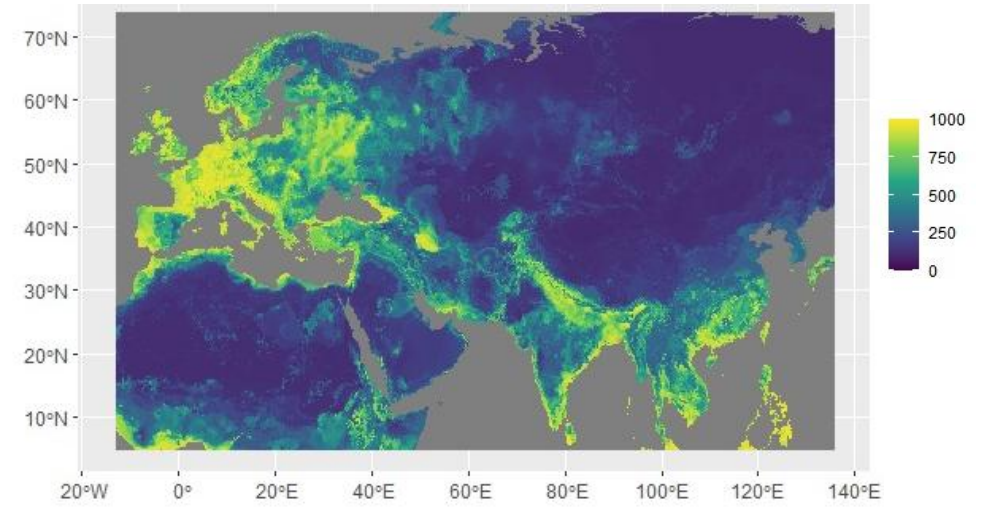
The patterns of adaptive potential and risk of maladaptiveness were different in FRDs and grey wolves. Genomic offset measured the difference between the current genomic makeup and the optimal genetic makeup facilitating the survival in the future environment (Gain et al., 2023). Genomic offset values in FRDs remained nearly constant and consistently low across central to southern Eurasia throughout all Shared Socioeconomic Pathways (SSPs) (Fig. S1). However, RONA values indicated that only the southernmost Eurasian populations are at low risk of maladaptiveness (Fig. 3A). This suggests that FRDs may be better adapted to warmer climates, and more likely to thrive in warmer habitats (e.g., habitats within the biomes tropical and subtropical forests, deserts and xeric shrublands), without relying on anthropogenic habitat modifications that may improve survival. Furthermore, FRDs feralization would also be less likely in colder habitats since dogs are more likely to thrive and have higher reproductive success in warmer climates (Niego & Benítez-Burraco et al., 2020; Porto et al., 2023).

In contrast, grey wolves' patterns change through time and across SSPs, with the highest values of genetic offset observed under the 585 SSP. Thus, the highest fossil fuel development pathway i.e., the 585, would result in the highest vulnerability of wolf populations to climate change (Fig. 2B, Fig. S2). Based on these predictions the least vulnerable populations under climate change are wolves from the Indian Peninsula, which represent two lineages of wolves, lowland Indian and Himalayan, differentiated from the other Eurasian lineages (Chapter 2; Hennelly et al., 2021; Hennelly et al., 2024; Khan et al., 2022; Wang et al., 2022). The Southern European and Northern Eurasian populations are predicted to be the most vulnerable and the first impacted by environmental shifts.

A. mid-Holocene



B. Current environmental conditions



C. Environmental shift (SSP 585, 2081-2100)

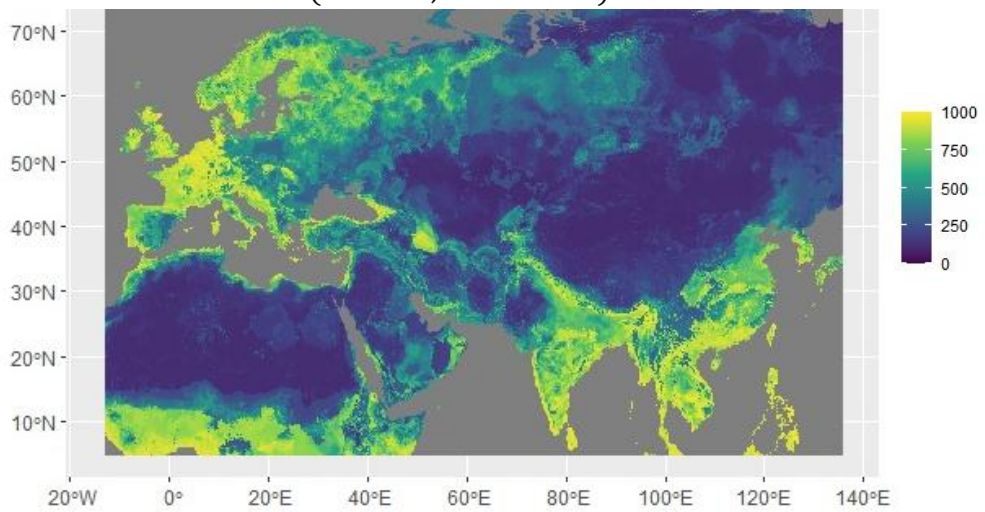


Figure 4. Dogs' suitable ecological niche models projections. Ecological niche models projections under mid-Holocene (A) current (B) and 2081-2100 (SSP 585) (C) environmental conditions. Areas in yellow have the highest niche suitability scores and areas in dark blue have suitability scores close to 0.

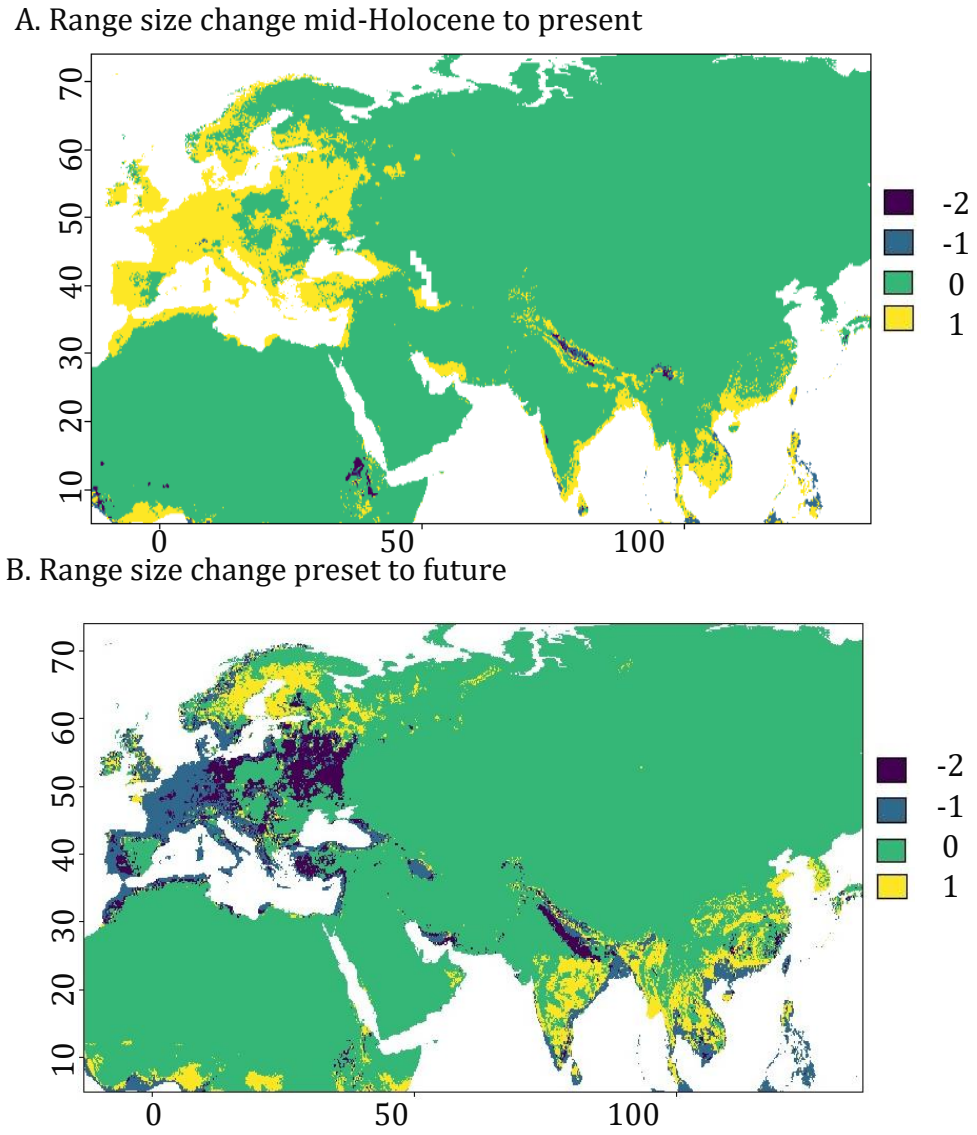


Figure 5. Dogs' suitable ecological niche size change over timeframes. (A) Range size difference based on the mid-Holocene and current projections (A) and (B) based current and future projections. The colours are in reference to the prediction values based on the changes in the pixels, -2 is currently occupied predicted to be lost, -1 predicted to remain occupied, 0 predicted to remain unoccupied and 1 predicted to be gained. The projections are suitable ecological niche modelling ensemble model by weighted means, the range difference was by cross validation and all three by TSS, merged data, merged run and merged algorithms

Altogether the results suggest that the populations of both FRDs and grey wolves are better adapted to warmer climates within their respective ranges. This is not surprising when considering that there are longitudinal gradients in environmental conditions and the closer to the equator the higher the biodiversity as the access to resources (e.g., food, water) increases (Mannion et al., 2013; Chaudhary et al., 2013). Similar patterns have been observed in other species (i.e., birds), with populations in colder climates showing higher genomic offset values (Chen et al., 2021).

Table 3. Estimates of unhoused dog populations densities

	Unhoused Dog/1000 person	Unhoused dog population	Unhoused dog density (dogs/km2)
Eurasia			
Tundra	**20**	2908810.00	**0.18**
Russia	20	2908810.00	0.18
Boreal Forests/Taiga	**4**	**49108.91**	**0.07**
Finland	5	28005.93	0.09
Sweden	2	21102.99	0.05
Temperate Broadleaf and Mixed Forests	**10**	**3281935.37**	**0.83**
France	5	332194.11	0.61
Germany	1	84548.23	0.24
Belgium	1	11712.89	0.38
United Kingdom	1	68682.96	0.28
Denmark	1	5948.14	0.15
Poland	15	581442.66	1.90
Slovakia	5	27590.28	0.57
Hungary	10	96864.63	1.06
Bulgaria	15	101937.05	0.94
Romania	45	860331.56	3.74
Ukraine	15	565992.54	0.98
Lithuania	5	14270.50	0.23
Latvia	5	9411.98	0.15
Estonia	2	2734.39	0.06
Turkiye	3	261811.50	0.34
South Korea	5	132091.02	1.35
Japan	1	124370.95	0.34
Temperate Grasslands, Savannas and Shrublands	**40**	117735.72	4.14
Armenia	40	117735.72	4.14
Montane Grasslands and Shrublands	**43**	**61171152.12**	**6.52**
China	43	61171152.12	6.52
Tropical and Subtropical Moist Broadleaf Forests	**31**	**81569797.78**	**13.62**
Indonesia	12	3374280.80	1.78
Philippines	100	11489119.90	38.53
Thailand	25	1792560.88	3.51
Viet nam	2	200704.38	0.64
India	45	64713131.82	21.77
Mediterranean Forests, Woodlands and Scrub	**19**	**1384215.01**	**1.36**
Portugal	15	156461.07	1.71
Spain	10	479115.79	0.96
Italy	10	594994.53	2.01
Greece	15	153643.62	1.19
Deserts and Xeric Shrublands	**32**	**14103237.96**	**3.95**

Saudi Arabia	1	33264.29	0.02
Afghanistan	17	704730.94	1.08
Pakistan	54	13365242.73	17.34
Africa			
Mediterranean Forests, Woodlands and Scrub	**10**	**377125.05**	**0.85**
Morocco	10	377125.05	0.85
Deserts and Xeric Shrublands	**26**	**2373318.11**	**0.78**
Egypt	3	343607.32	0.35
Namibia	45	133339.28	0.16
South Africa	30	1896371.52	1.56
Tropical and Subtropical Grasslands, Savannas and Shrublands	**76**	**13263204.34**	**6.49**
Uganda	55	2676113.06	13.35
United Republic of Tanzania	76	5062938.06	5.72
Kenya	70	3873730.21	6.81
Zimbabwe	101	1650423.02	4.27
Oceania			
Deserts and Xeric Shrublands	**5**	**132255.62**	**0.02**
Australia	5	132255.62	0.02
Temperate Broadleaf and Mixed Forests	**3**	**15518.51**	**0.06**
New Zealand	3	15518.51	0.06
America			
Boreal Forests/Taiga	**3**	**117897.32**	**0.01**
Canada	3	117897.32	0.01
Temperate Grasslands, Savannas and Shrublands	**28**	**3723003.12**	**0.29**
United States of America	4	1373909.34	0.15
Chile	50	982941.75	1.32
Argentina	30	1366152.03	0.50
Tropical and Subtropical Moist Broadleaf Forests	**29**	**8528267.03**	**0.79**
Puerto Rico	30	97260.69	10.97
Costa Rica	40	204221.00	4.00
Colombia	20	1046423.04	0.94
Perú	25	846140.43	0.66
Brazil	30	6334221.87	0.76
Deserts and Xeric Shrublands	**36**	**4670631.32**	**2.40**
Mexico	36	4670631.32	2.40

Historically, warmer habitats have facilitated radiations in canids, and these evolutionary processes are more likely to occur in habitats where canids retain higher adaptive potential (Porto et al., 2023). It is important to remember, however, that genomic offset calculations do not account for gene flow between populations, which can impact adaptive potential (Rellstab et al., 2021). Nevertheless, genomic offset

statistics emerge as a valuable tool not only for planning management strategies for wild and invasive populations—such as assisted migration or the introduction of captive-bred populations—but also for deepening our understanding of evolutionary processes. Considering that FRDs and grey wolves show different adaptive potential patterns, if FRDs have retained the adaptive potential in the regions which they were initially domesticated in, our results suggest dogs were domesticated in southern parts of Eurasia.

4.2. Dogs' suitable habitats are semi-arid, sub-tropical and tropical

Based on the ensemble model predictions by the mid-Holocene dogs' most suitable habitats were in southern Eurasia. These habitats expanded since that period and are predicted to continue to expand with future environmental shifts (Fig 5). It is possible that the dogs' distribution model used for the modelling of most suitable habitats underrepresented their current distribution due to the assumptions made by the models. This includes the assumption that the training data (i.e., presence points), used for constructing the species distribution model, are representative of species suitable habitats and the probability of detecting individuals across sampled area is relatively constant. Moreover, the models assume that the provided environmental variables are the main factors influencing the species' distribution and biotic factors are not considered (Luan et al., 2020; Yackulic et al., 2013). Dogs are human commensals hence their distribution is inherently impacted by humans (Feuerborn et al., 2021; Perri et al., 2021, reviewed in Sykes et al., 2020). Due to the sampling bias, there are likely missing presence points within the sampled area, although the use of ensemble modelling reduced the impact of this bias upon our results.

Dogs' most suitable habitats were not continuous, but two different areas: Southern Asia—a region with warm climate, dominated by tropical forest, xeric shrubland habitats, and Europe —a region with temperate climate (Fig. 4). The former had the most suitable habitat since the mid-Holocene till the predictions for 2100. This is congruent with local adaptive variation patterns in FRDs, which showed that the main drivers of adaptive variation were wet cool and dry and hot environmental conditions (Chapter 3), and genomic offset patterns with higher adaptive potentials in warmer regions (Fig. 2). Consistency between these approaches suggests a strong alignment between dogs' adaptive variation, suitable habitats and their adaptive potential. To make inferences about the significance of these patterns, it is important to consider the climate conditions in these habitats around the time dogs are predicted to have been domesticated i.e., 16,000-40,000 years ago.

Approximately ~19000-27000 years ago, during the Last Glacial Maximum, the habitats of Southern Asia were tropical and sub-tropical forests and semi-arid deserts. The fluctuations in the monsoons in the region modified the seasonal rainfall resulting in patches of wet and dry habitats (Chabangborn et al., 2010; Hussain et al., 2025; Niu et al., 2025). Climate in Europe during the last glacial maximum was cold with low rainfall (Hughes et al., 2020). By the Holocene, the climate was significantly warmer and with higher precipitation (Popescu et al., 2010; Stivrins et al., 2022; Strandberg et al., 2022). The climate in Southern Asia by the Holocene was warmer and there was a return of the monsoons, meaning there where increases in rainfall (Hussain et al., 2025; Niu et al.,

2025). Thus, the ensemble model predicted that by the mid-Holocene the warmer climates with higher precipitation in Europe and Southern Asia created more suitable habitats for dogs. It also explains why the genomic offset statistics predicted European FRDs populations to retain an adaptive potential albeit lower than the Southern Asian populations.

Local adaptive variation patterns in Chapter 3 suggest that the detected in FRDs is most beneficial for habitats with both heavy rain (e.g., tropical and subtropical forests) and dry conditions (e.g., xeric shrublands), like those predicted to have been present in Southern Eurasia in the past. Modern dogs' most suitable habitat in natural landscapes lies within warmer areas and may include dry and/or rainy seasons. Unhoused dog density estimates patterns echo this, with higher densities of unhoused dogs estimated in regions where, sub-tropical and tropical forests, and desert and xeric shrublands biomes dominate in both Eurasia and America (Table 3). Furthermore, FRDs retain an adaptive potential for regions consistent with dogs' most suitable ecological habitat (Figs. 2, 4).

We obtained consistent patterns based on independent approaches with three different data types (i.e., SNP genotypes of FRDs, presence data of dogs and demographic estimates of unhoused dogs). Altogether, our results suggest a Southern Eurasia domestication centre with habitats like current tropical and subtropical forests, and xeric shrublands. Current distribution of modern dogs into less suitable habitats may be explained by their relationship with humans. Considering that the human population density began to significantly increase with agricultural revolution and urbanization (Ellis, 2021; Ellis et al., 2021, reviewed in Sykes et al., 2020). Both adaptive introgression with wolves (e.g., adaptations to high altitude habitats (Signore et al., 2019; Wang et al., 2020)), and dogs' relationship with humans, may explain how the dogs' range expanded even into less suitable habitats. Due to their niche as a human commensal and possibilities to interbreed with wild canids, dogs have been able to bypass some of their ecological limitations and rapidly expand their range to an almost global scale.

4.3. Feralization of dogs in semi-arid, sub-tropical and tropical habitats

Feralization is a process by which a domesticated animal adapts and "returns" to occupy a niche in natural landscapes (Niego & Benítez-Burraco, 2022). Most FRDs are not feral because, even though they range and reproduce freely, they are dependent on human-derived food sources (Gompper, 2014; Pilot et al., 2015). There are few well-described examples of truly feral dogs, the Australian dingo, New Guinea singing dogs, Carolina dogs and Galapagos dogs. The taxonomy of the Australian dingo and New Guinea singing dogs is strongly debated particularly due to ongoing hybridization with domestic canids, still genetic studies have shown that they descend from domestic dogs (Field et al., 2022; Oskarsson et al., 2012; Shipman, 2020; Surbakti et al., 2020; van Asch et al., 2013). Dingoes are apex predators in Australia and dominate the semi-arid habitats in the island (i.e., desert and xeric shrublands, and tropical and subtropical grasslands, savannas and shrublands biomes) (Shipman, 2020). New Guinea singing dogs once roamed the humid tropical island of New Guinea and have been re-discovered in the highlands (Surbakti et al., 2020). Carolina dogs are found in

southeastern North America (van Asch et al., 2013) in areas dominated by temperate grasslands, savannas and shrublands. The Galapagos dogs inhabited the desert and xeric shrublands biome in the islands ~180 years, until their eradication in 1980s (Reponen et al., 2014). Our results could help explain why these are the only confirmed instances of feralization across the widespread distribution range of domestic dogs.

The successful feralization of Australian dingoes and New Guinea singing dogs may have been facilitated not only by geographic isolation but also because these semi-arid and tropical enviroments, respectively, are the type environments for which domestic dogs retained an adaptive potential. In the case of the Carolina dogs more studies are necessary to elucidate the factors impacting their divergence, still the warmer climate in the region likely facilitated this evolutionary process. Furthermore, these same patterns could help explain the higher FRD density in warmer regions across the globe (Table 3), since their reproductive success is likely higher there compared with colder regions. Thus, free-ranging domestic dogs are likely to achieve highest ecological success (i.e., live longer and show higher reproductive success) in tropical, sub-tropical habitats and other warm habitats (e.g., xeric shrublands).

5. Conclusions

Dogs most suitable habitats for FRDs are in warm (e.g., sub-tropical and tropical forests and, xeric shrublands) and temperate regions (e.g., temperate broadleaf and mixed forests) where they retain a high adaptive potential. The range of dogs' suitable habitats has continued to grow in par with the human population growth, and with climate change increasing the range of warmer habitats coupled with the continued anthropogenic expansion, FRD populations will likely grow unless population control measures are applied. Higher FRDs densities in warmer climates likely result from these regions being better suited for FRDs survival with minimal human assistance. Furthermore, since dogs' most suitable habitat is in warmer climates and FRDs have the largest adaptive potential in such climates, it explains why truly feral dog populations only occur in warmers climates. Overall, our results point to the domestication origin of dogs in a southern region of Eurasia, with Southeast Asia and South Europe being the most likely regions. Although we do not have sufficient data to provide a more precise geographic origin, our findings point to places where it may be worth focusing the sampling effort in future studies on dog domestication.

References

Amici, F., Meacci, S., Caray, E., Oña, L., Liebal, K., & Ciucci, P. (2024). A first exploratory comparison of the behaviour of wolves (*Canis lupus*) and wolf-dog hybrids in captivity. *Animal Cognition*, *27*(1), 9. https://doi.org/10.1007/S9071-024-01849-7

Belo, V. S., Werneck, G. L., da Silva, E. S., Barbosa, D. S., & Struchiner, C. J. (2015). Population estimation methods for free-ranging dogs: A systematic review. PLOS ONE, 10(12), e0144830. https://doi.org/10.1371/journal.pone.0144830
Belsare, A., & Vanak, A. T. (2020). Modelling the challenges of managing free-ranging dog populations. *Scientific Reports*, 10, 18874. https://doi.org/10.1038/S31598-020-75828-6

Bergström, A., Frantz, L., Schmidt, R., Ersmark, E., Lebrasseur, O., Girdland-Flink, L., Lin, A. T., Storå, J., Sjögren, K. G., Anthony, D., Antipina, E., Amiri, S., Bar-Oz, G., Bazaliiskii, V. I., Bulatović, J., Brown, D., Carmagnini, A., Davy, T., Fedorov, S., Fiore, I., … Skoglund, P. (2020). Origins and genetic legacy of prehistoric dogs. *Science (New York, N.Y.)*, *370*(6516), 557–564. https://doi.org/10.1126/science.aba9572

Bergström, A., Stanton, D. W. G., Taron, U. H., et al. (2022). Grey wolf genomic history reveals a dual ancestry of dogs. *Nature*, 607, 313–320. https://doi.org/10.1038/S31586-022-04824-9

Boitani, L., Phillips, M., & Jhala, Y. (2023). *Canis lupus* (amended version of 2018 assessment). The IUCN Red List of Threatened Species, 2023, e.T3746A247624660. https://dx.doi.org/10.2305/IUCN.UK.2023-1.RLTS.T3746A247624660.en

Bryce, C. B. (2021). Dogs as pets and pests: Global patterns of canine abundance, activity, and health. Integrative and Comparative Biology, 61(1), 154–165. https://doi.org/10.1093/icb/icab046

Cafazzo, S., Bonanni, R., Valsecchi, P., & Natoli, E. (2014). Social Variables Affecting Mate Preferences ,Copulation and Reproductive Outcomeina Pack of Free Ranging Dogs. *PLoSONE*, 9(6): e98594. https://doi.org/10.1371/journal.pone.0098594

Cafazzo, S., Valsecchi, P., Bonanni, R., & Natoli, E. (2010). Dominance in relation to age, sex, and competitive contexts in a group of free-ranging domestic dogs. *Behavioral Ecology, 21*(3), 443–455. https://doi.org/10.1093/beheco/arq001

Caten, C. T., & Dallas, T. (2023). Thinning occurrence points does not improve species distribution model performance. *Ecosphere, 14*(12), e4405. https://doi.org/10.1002/ecS1.4405

Chabangborn, A., Brandefelt, J. & Wohlfarth, B. (2014), LGM Asian monsoon climate. *Boreas*, 43: 220-242. https://doi.org/10.1111/bor.12032

Chaudhary, C., Richardson, A. J., Schoeman, D. S., & Costello, M. J. (2021). Global warming is causing a more pronounced dip in marine species richness around the equator. *Proceedings of the National Academy of Sciences of the United States of America*, 118(15), e2015094118. https://doi.org/10.1073/pnas.2015094118

Chen, Y., Jiang, Z., Fan, P., Ericson, P. G. P., Song, G., Luo, X., Lei, F., & Qu, Y. (2022). The combination of genomic offset and niche modelling provides insights into climate change-driven vulnerability. *Nature Communications, 13*, 4821. https://doi.org/10.1038/S31467-022-32379-9

Dinerstein, E., Olson, D., Joshi, A., Vynne, C., Burgess, N. D., Wikramanayake, E., Hahn, N., Palminteri, S., Hedao, P., Noss, R., Hansen, M., Locke, H., Ellis, E. C., Jones, B., Barber, C. V., Hayes, R., Kormos, C., Martin, V., Crist, E., Sechrest, W., et al. (2017). An ecoregion-based approach to protecting half the terrestrial realm. Bioscience, 67(6), 534–545. https://doi.org/10.1093/biosci/bix014

Dubos, N., Préau, C., Lenormand, M., Papuga, G., Monsarrat, S., Denelle, P., Le Louarn, M., Heremans, S., May, R., Roche, P., & Luque, S. (2022). Assessing the effect of sample bias correction in species distribution models. *Ecological Indicators*, *145*, 109487. https://doi.org/10.1016/j.ecolind.2022.109487

Duarte, J., García, F. J., & Fa, J. E. (2016). Depredatory impact of free-roaming domestic dogs on Mediterranean deer in southern Spain: Implications for human-wolf conflict. Folia Zoologica, 65(2), 135–141. https://doi.org/10.25225/fozo.v65.i2.a8.2016

Ellis, E. C. (2021). Land use and ecological change: A 12,000-year history. *Annual Review of Environment and Resources, 46*, 1-33. https://doi.org/10.1146/annurev-environ-012220-010822

Ellis, E. C., Gauthier, N., Klein Goldewijk, K., Bliege Bird, R., Boivin, N., Díaz, S., Fuller, D. Q., Gill, J. L., Kaplan, J. O., Kingston, N., Locke, H., McMichael, C. N. H., Ranco, D., Rick, T. C., Shaw, M. R., Stephens, L., Svenning, J.-C., & Watson, J. E. M. (2021). People have shaped most of terrestrial nature for at least 12,000 years. *Proceedings of the National Academy of Sciences*, *118*(17), e2023483118. https://doi.org/10.1073/pnas.2023483118

Feuerborn, T. R., Carmagnini, A., Losey, R. J., Nomokonova, T., Askeyev, A., Askeyev, I., Askeyev, O., Antipina, E. E., Appelt, M., Bachura, O. P., Beglane, F., Bradley, D. G., Daly, K. G., Gopalakrishnan, S., Murphy Gregersen, K., Guo, C., Gusev, A. V., Jones, C., Kosintsev, P. A., Kuzmin, Y. V., … Frantz, L. (2021). Modern Siberian dog ancestry was shaped by several thousand years of Eurasian-wide trade and human dispersal. *Proceedings of the National Academy of Sciences of the United States of America*, 118(39), e2100338118. https://doi.org/10.1073/pnas.2100338118

Field, M. A., Yadav, S., Dudchenko, O., Esvaran, M., Rosen, B. D., Skvortsova, K., Edwards, R. J., Keilwagen, J., Cochran, B. J., Manandhar, B., Bustamante, S., Rasmussen, J. A., Melvin, R. G., Chernoff, B., Omer, A., Colaric, Z., Chan, E. K. F., Minoche, A. E., Smith, T. P. L., Gilbert, M. T. P., … Ballard, J. W. O. (2022). The

Australian dingo is an early offshoot of modern breed dogs. *Science Advances*, *8*(16), eabm5944. https://doi.org/10.1126/sciadv.abm5944

Frantz, L. A., Mullin, V. E., Pionnier-Capitan, M., Lebrasseur, O., Ollivier, M., Perri, A., Linderholm, A., Mattiangeli, V., Teasdale, M. D., Dimopoulos, E. A., Tresset, A., Duffraisse, M., McCormick, F., Bartosiewicz, L., Gál, E., Nyerges, É. A., Sablin, M. V., Bréhard, S., Mashkour, M., Bălăşescu, A., … Larson, G. (2016). Genomic and archaeological evidence suggest a dual origin of domestic dogs. *Science (New York, N.Y.)*, *352*(6290), 1228–1231. https://doi.org/10.1126/science.aaf3161

Freedman, A. H., Gronau, I., Schweizer, R. M., Ortega-Del Vecchyo, D., Han, E., Silva, P. M., Galaverni, M., Fan, Z., Marx, P., Lorente-Galdos, B., Beale, H., Ramirez, O., Hormozdiari, F., Alkan, C., Vilà, C., Squire, K., Geffen, E., Kusak, J., Boyko, A. R., Parker, H. G., … Novembre, J. (2014). Genome sequencing highlights the dynamic early history of dogs. *PLoS genetics*, *10*(1), e1004016. https://doi.org/10.1371/journal.pgen.1004016

Gain, C., Rhoné, B., Cubry, P., Salazar, I., Forbes, F., Vigouroux, Y., Jay, F., & François, O. (2023). A Quantitative Theory for Genomic Offset Statistics. *Molecular biology and evolution*, *40*(6), msad140. https://doi.org/10.1093/molbev/msad140

GIBF.org (16 May 2025). GIBF Occurrence download. https://doi.org/10.15468/dl.8y3923

Gompper, M. E. (2014). Introduction: Outlining the ecological influences of a subsidized, domesticated predator. In Free-Ranging Dogs and Wildlife Conservation (pp. 1–8). Oxford University Press.

Gonzalez Bernal, A., Martínez-Meyer, E., & Rojas-Soto, O. (2022). Climatic comparison of the gray wolf (*Canis lupus*) subspecies in North America using niche-based distribution models and its implications for conservation programs. *Journal of Mammalogy. 103*(10), 1305–1314. https://doi.org/10.1093/jmammal/gyac066

Gou, X., Zhang, Z., Li, J., & Zhang, Y. (2014). Dogs living in high-altitude environment of the Tibetan Plateau show selection signatures in genes associated with hypoxia adaptation. *Molecular Biology and Evolution, 31*(3), 707–715. https://doi.org/10.1093/molbev/msu026

Guéguen M, Blancheteau H, Thuiller W (2025). biomod2: Ensemble Platform for Species Distribution Modeling. R package version 4.3-4, https://biomodhub.github.io/biomod2/

Hennelly, L. M., Habib, B., Modi, S., Rueness, E. K., Gaubert, P., & Sacks, B. N. (2021). Ancient divergence of Indian and Tibetan wolves revealed by recombination-aware phylogenomics. *Molecular Ecology*, *30*(24), 6687–6700. https://doi.org/10.1111/mec.16127.

Hennelly, L. M., Sarwar, G., Fatima, H., Werhahn, G., Abbas, F. I., Khan, A. M., Mahmood, T., Kachel, S., Kubanychbekov, Z., Waseem, M. T., Zahra Naqvi, R., Hamid,

A., Abbas, Y., Aisha, H., Waseem, M., Farooq, M., & Sacks, B. N. (2024). Genomic analysis of wolves from Pakistan clarifies boundaries among three divergent wolf lineages. *The Journal of Heredity*, *115*(4), 339–348. https://doi.org/10.1093/jhered/esad066
Hijmans, R. (2024). terra: Spatial Data Analysis (Version 1.7-81). https://github.com/rspatial/terra

Hughes, J., & Macdonald, D. W. (2013). A review of the interactions between free-roaming domestic dogs and wildlife. *Biological Conservation*, 157, 341–351. https://doi.org/10.1016/j.biocon.2012.07.005

Hughes, P. D., Palacios, D., García-Ruiz, J. M., & Andrés, N. (2020). The European glacial landscapes from the Last Glacial Maximum - synthesis. In *The Encyclopedia of Quaternary Science* (pp. 43-56). Elsevier. https://doi.org/10.1016/B978-0-12-823498-3.00005-4
Hussain, S. A., Tan, L., Kathayat, G., Hussain, A., & Widory, D. (2025). Climate change during the Holocene in South Asia: A review study of Pakistan. *Quaternary Science Reviews*, 352, 109203. https://doi.org/10.1016/j.quascirev.2025.109203

Hussain, S. R., Mahar, N., Dirzo, R., & Habib, B. (2021). Successful neighbour: Interactions of the generalist carnivore red fox with dogs, wolves and humans for continued survival in dynamic anthropogenic landscapes. Global Ecology and Conservation, 25, e01446. https://doi.org/10.1016/j.gecco.2020.e01446

Khan, S., Shrotriya, S., Sadhukhan, S., Lyngdoh, S., Goya, S. P., & Habib, B. (2022). Comparative Ecological Perspectives of Two Ancient Lineages of Gray Wolves: Wooly Wolf (Canis lupus chanco) and Indian Wolf (Canis lupus pallipes). *Frontiers in Ecology and Evolution, 10*, 775612. https://doi.org/10.3389/fevo.2022.775612

Larson, G., Karlsson, E. K., Perri, A., Webster, M. T., Ho, S. Y., Peters, J., Stahl, P. W., Piper, P. J., Lingaas, F., Fredholm, M., Comstock, K. E., Modiano, J. F., Schelling, C., Agoulnik, A. I., Leegwater, P. A., Dobney, K., Vigne, J. D., Vilà, C., Andersson, L., & Lindblad-Toh, K. (2012). Rethinking dog domestication by integrating genetics, archeology, and biogeography. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(23), 8878–8883. https://doi.org/10.1073/pnas.1203005109

Lazzaroni, M., Range, F., Backes, J., Portele, K., Scheck, K., & Marshall-Pescini, S. (2020). The Effect of Domestication and Experience on the Social Interaction of Dogs and Wolves With a Human Companion. *Frontiers in psychology*, *11*, 785. https://doi.org/10.3389/fpsyg.2020.00785

Li, R., Zhang, Y., Zhang, Q., & Zhao, W. (2014). Whole-genome sequencing of dogs reveals genetic signatures of high-altitude adaptation. *Nature Communications, 5*, 5518. https://doi.org/10.1038/ncommS5518

Lino, S., Rossa, M., Fernandes, J.M., Barros, T., Hipólito, D., Ferreira, E., Aliácar, S.C., Cadete, D., Fonseca, C., Torres, R.T., Rosalino, L.M. & Carvalho, J. (2023). Dog in sheep's clothing: livestock depredation by free-ranging dogs may pose new challenges to wolf conservation. *Eur. J. Wildl. Res.* 69, 107. https://doi.org/10.1007/S9344-023-01740-9

Liu, Y.-H., Wang, L., Xu, T., Guo, X., Li, Y., Yin, T.-T., Yang, H.-C., Hu, Y., Adeola, A. C., Sanke, O. J., Otecko, N. O., Wang, M., Ma, Y., Charles, O. S., Sinding, M.-H. S., Gopalakrishnan, S., Samaniego, J. A., Hansen, A. J., Fernandes, C., Gaubert, P., Budd, J., Dawuda, P. M., Rueness, E. K., Jiang, L., Zhai, W., Gilbert, M. T. P., Peng, M.-S., Qi, X., Wang, G.-D., & Zhang, Y.-P. (2018). Whole-genome sequencing of African dogs provides insights into adaptations against tropical parasites. *Molecular Biology and Evolution, 35*(2), 287–298. https://doi.org/10.1093/molbev/msx258

Luan, J., Zhang, C., Xu, B., Xue, Y., & Ren, Y. (2020). The predictive performances of random forest models with limited sample size and different species traits. *Fisheries Research, 227*, 105534. https://doi.org/10.1016/j.fishres.2020.105534

Mahajan, P., Khandal, D., Dhakad, M., & Rasal, V. (2024). A dog-eat-dog world: predation and resource acquisition by free-ranging dogs around Ranthambhore Tiger Reserve, Rajasthan, India. *Ethology Ecology & Evolution*, *36*(5), 572–582. https://doi.org/10.1080/03949370.2024.2358767

Mannion, P. D., Upchurch, P., Benson, R. B., & Goswami, A. (2014). The latitudinal biodiversity gradient through deep time. *Trends in ecology & evolution*, 29(1), 42–50. https://doi.org/10.1016/j.tree.2013.09.012

Natoli, E., Bonanni, R., Cafazzo, S., Mills, D. S., Pontier, D., & Pilot, M. (2021). Genetic inference of the mating system of free-ranging domestic dogs. Behavioral Ecology, 32(4), 646–656. https://doi.org/10.1093/beheco/arab011

Niego, A., & Benítez-Burraco, A. (2021). Are feralization and domestication truly mirror processes? *Ethology Ecology & Evolution*, 34(5), 557–590.

Niu, X., Wang, J., Kang, L., Li, Y., Zhang, H., Dong, X., Li, H., Sha, L., Yi, L., Sinha, A., Ning, Y., Jia, X., Zong, B., Zhang, F., Cai, Y., Woodhead, J., Liang, F., Chu, Z., Guo, J., Edwards, R. L., & Cheng, H. (2025). Millennial-scale climate variability of the Asian summer monsoon over the last 690,000 years: Insights from cave records. *Science Bulletin*, 70(9), 1513–1522. https://doi.org/10.1016/j.scib.2025.03.022

Oskarsson, M. C., Klütsch, C. F., Boonyaprakob, U., Wilton, A., Tanabe, Y., & Savolainen, P. (2012). Mitochondrial DNA data indicate an introduction through Mainland Southeast Asia for Australian dingoes and Polynesian domestic dogs. *Proceedings. Biological sciences*, *279*(1730), 967–974. https://doi.org/10.1098/rspb.2011.1395

Oksanen, J., Simpson, G., Blanchet, F., Kindt, R., Legendre, P., Minchin, P., O'Hara, R., Solymos, P., Stevens, M., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., De Caceres, M., Durand, S., Evangelista, H.,

FitzJohn, R., Friendly, M., Furneaux, B., Hannigan, G., Hill, M., Lahti, L., McGlinn, D., Ouellette, M., Ribeiro Cunha, E., Smith, T., Stier, A., Ter Braak, C., & Weedon, J. (2024). vegan: Community ecology package (Version 2.7-0). https://github.com/vegandevs/vegan

Perri, A. R., Feuerborn, T. R., Frantz, L. A. F., Larson, G., Malhi, R. S., Meltzer, D. J., & Witt, K. E. (2021). Dog domestication and the dual dispersal of people and dogs into the Americas. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(6), e2010083118. https://doi.org/10.1073/pnas.2010083118

Pendleton, A. L., Shen, F., Taravella, A. M., Emery, S., Veeramah, K.R., Boyko, A.R., & Kidd, J. M. (2018). Comparison of village dog and wolf genomes highlights the role of the neural crest in dog domestication. *BMC Biology, 16*, 64. https://doi.org/10.1186/s12915-018-0535-2

Pilot, M., Malewski, T., Moura, A. E., Grzybowski, T., & Oleński, K. (2015). On the origin of mongrels: Evolutionary history of free-breeding dogs in Eurasia. Proceedings of the Royal Society B: Biological Sciences, 282(20152189). https://doi.org/10.1098/rspb.2015.2189

Pilot, M., Malewski, T., Moura, A. E., Grzybowski, T., Oleński, K., Kamiński, S., Fadel, F. R., Alagaili, A. N., Mohammed, O. B., & Bogdanowicz, W. (2016). Diversifying selection between pure-breed and free-breeding dogs inferred from genome-wide SNP analysis. *G3 Genes|Genomes|Genetics*, 6(8), 2285–2298. https://doi.org/10.1534/g3.116.029678

Popescu, S.-M., Biltekin, D., Winter, H., Suc, J.-P., Melinte-Dobrinescu, M. C., Klotz, S., Rabineau, M., Combourieu-Nebout, N., Clauzon, G., & Deaconu, F. (2010). Pliocene and Lower Pleistocene vegetation and climate changes at the European scale: Long pollen records and climatostratigraphy. *Quaternary International*, 219(1–2), 152–167. https://doi.org/10.1016/j.quaint.2010.03.004

Porto, L. M. V., Etienne, R. S., & Maestri, R. (2023). Evolutionary radiation in canids following continental colonizations. Evolution; international journal of organic evolution, 77(4), 971–979. https://doi.org/10.1093/evolut/qpad015

Proust-Lima, C., Philipps, V., Diakite, A., & Liquet, B. (2023). lfmm: Extended mixed models using latent classes and latent processes (Version 2.1.0). https://cran.r-project.org/package=lcmm

Range, F., & Marshall-Pescini, S. (2022). Comparing wolves and dogs: current status and implications for human 'self-domestication'. *Trends in cognitive sciences*, *26*(4), 337–349. https://doi.org/10.1016/j.tics.2022.01.003

Rellstab, C., Dauphin, B., & Exposito-Alonso, M. (2021). Prospects and limitations of genomic offset in conservation management. *Evolutionary Applications, 14*(5), 1026-1038. https://doi.org/10.1111/eva.13205

Reponen, S. E., Brown, S. K., Barnett, B. D., & Sacks, B. N. (2014). Genetic and morphometric evidence on a Galápagos Island exposes founder effects and

diversification in the first-known (truly) feral western dog population. *Molecular ecology*, *23*(2), 269–283. https://doi.org/10.1111/mec.12595

Riahi, K., van Vuuren, D. P., Kriegler, E., Edmonds, J., O'Neill, B. C., Fujimori, S., & Lutz, W. (2017). The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Global Environmental Change, 42*, 153–168. https://doi.org/10.1016/j.gloenvcha.2016.05.009

Rivas-Blanco, D., Monteiro, T., Virányi, Z., & Range, F. (2024). Going back to "basics": Harlow's learning set task with wolves and dogs. *Learning & behavior*, *52*(4), 315–329. https://doi.org/10.3758/s13420-024-00631-6

Schweizer, R. M., vonHoldt, B. M., Harrigan, R., Knowles, J. C., Musiani, M., Coltman, D., Novembre, J., & Wayne, R. K. (2016). Genetic subdivision and candidate genes under selection in North American grey wolves. Molecular Ecology, 25(1), 380-402. https://doi.org/10.1111/mec.13364

Signore, A. V., Yang, Y. Z., Yang, Q. Y., Qin, G., Moriyama, H., Ge, R. L., & Storz, J. F. (2019). Adaptive Changes in Hemoglobin Function in High-Altitude Tibetan Canids Were Derived via Gene Conversion and Introgression. *Molecular biology and evolution*, *36*(10), 2227–2237. https://doi.org/10.1093/molbev/msz097

Shannon, L. M., Boyko, R. H., Castelhano, M., & Boyko, A. R. (2015). Genetic structure in village dogs reveals a Central Asian domestication origin. *Proceedings of the National Academy of Sciences, 112*(44), 13639–13644. https://doi.org/10.1073/pnas.1516215112

Shipman P. (2020). What the dingo says about dog domestication. *Anatomical record (Hoboken, N.J. : 2007)*, *304*(1), 19–30. https://doi.org/10.1002/ar.24517

Strandberg, G., Lindström, J., Poska, A., Zhang, Q., Fyfe, R., Githumbi, E., Kjellström, E., Mazier, F., Nielsen, A. B., Sugita, S., Trondman, A.-K., Woodbridge, J., & Gaillard, M.-J. (2022). Mid-Holocene European climate revisited: New high-resolution regional climate model simulations using pollen-based land-cover. *Quaternary Science Reviews, 281*, 107431. https://doi.org/10.1016/j.quascirev.2022.107431

Stivrins, N., Kalnina, L., Cerina, A., Reire, E., Kreslina, S., Ozola, I., Soms, J., & Veski, S. (2022). Climate change impact on peatland dynamics during the Holocene in Latvia, Northeastern Europe. *Quaternary Science Reviews, 259*, 106958. https://doi.org/10.1016/j.quascirev.2021.106958

Surbakti, S., Parker, H. G., McIntyre, J. K., Maury, H. K., Cairns, K. M., Selvig, M., Pangau-Adam, M., Safonpo, A., Numberi, L., Runtuboi, D. Y. P., Davis, B. W., & Ostrander, E. A. (2020). New Guinea highland wild dogs are the original New Guinea singing dogs. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(39), 24369–24376. https://doi.org/10.1073/pnas.2007242117

Sykes, N., Beirne, P., Horowitz, A., Jones, I., Kalof, L., Karlsson, E., King, T., Litwak, H., McDonald, R. A., Murphy, L. J., Pemberton, N., Promislow, D., Rowan, A., Stahl, P. W., Tehrani, J., Tourigny, E., Wynne, C. D. L., Strauss, E., & Larson, G. (2020). Humanity's

Best Friend: A Dog-Centric Approach to Addressing Global Challenges. *Animals*, *10*(3), 502. https://doi.org/10.3390/ani10030502

Tancredi, D., & Cardinali, I. (2023). Being a Dog: A Review of the Domestication Process. Genes, 14(5), 992. https://doi.org/10.3390/genes14050992

Thalmann, O., Shapiro, B., Cui, P., Schuenemann, V. J., Sawyer, S. K., Greenfield, D. L., Germonpré, M. B., Sablin, M. V., López-Giráldez, F., Domingo-Roura, X., Napierala, H., Uerpmann, H. P., Loponte, D. M., Acosta, A. A., Giemsch, L., Schmitz, R. W., Worthington, B., Buikstra, J. E., Druzhkova, A., Graphodatsky, A. S., … Wayne, R. K. (2013). Complete mitochondrial genomes of ancient canids suggest a European origin of domestic dogs. *Science (New York, N.Y.)*, *342*(6160), 871–874. https://doi.org/10.1126/science.1243650

van Asch, B., Zhang, A. B., Oskarsson, M. C., Klütsch, C. F., Amorim, A., & Savolainen, P. (2013). Pre-Columbian origins of Native American dog breeds, with only limited replacement by European dogs, confirmed by mtDNA analysis. *Proceedings. Biological sciences*, *280*(1766), 20131142. https://doi.org/10.1098/rspb.2013.1142

Vanak, A.T. and Gompper, M.E. (2009). Dogs Canis familiaris as carnivores: their role and function in intraguild competition. *Mammal Review*, 39, 265-283. https://doi.org/10.1111/j.1365-2907.2009.00148.x

Wang, M. S., Thakur, M., Jhala, Y., Wang, S., Srinivas, Y., Dai, S. S., Liu, Z. X., Chen, H. M., Green, R. E., Koepfli, K. P., & Shapiro, B. (2022). Genome Sequencing of a Gray Wolf from Peninsular India Provides New Insights into the Evolution and Hybridization of Gray Wolves. *Genome biology and evolution*, *14*(2), evac012. https://doi.org/10.1093/gbe/evac012

Wang, M. S., Wang, S., Li, Y., Jhala, Y., Thakur, M., Otecko, N. O., Si, J. F., Chen, H. M., Shapiro, B., Nielsen, R., Zhang, Y. P., & Wu, D. D. (2020). Ancient Hybridization with an Unknown Population Facilitated High-Altitude Adaptation of Canids. *Molecular biology and evolution*, *37*(9), 2616–2629. https://doi.org/10.1093/molbev/msaa113

WellBeing International. (2020). Global dog campaign: About GDC [Data set]. https://wellbeingintl.org/global-dog-campaign/about-gdc/

Wheat, C.H. & Wynne, C.D.L. (2025). The unfulfilled potential of dogs in studying behavioural ecology and evolution during the Anthropocene, *Animal Behaviour*, 219 (123020), ISSN 0003-3472. https://doi.org/10.1016/j.anbehav.2024.10.034

Wierzbowska, I. A., Hędrzak, M., Popczyk, B., Okarma, H., & Crooks, K. R. (2016). Predation of wildlife by free-ranging domestic dogs in Polish hunting grounds and potential competition with the grey wolf. Biological Conservation, 201, 1-9. https://doi.org/10.1016/j.biocon.2016.06.016

Wolf, C., & Ripple, W. J. (2017). Range contractions of the world's large carnivores. *Royal Society open science*, *4*(7), 170052. https://doi.org/10.1098/rsos.170052

World Bank. (2023). World development indicators. The World Bank. https://data.worldbank.org/indicator/AG.LND.TOTL.K2

World Health Organization (WHO). (2023). Global health observatory data repository. World Health Organization. https://www.who.int/data/gho

Yackulic, C.B., Chandler, R., Zipkin, E.F., Royle, J.A., Nichols, J.D., Campbell Grant, E.H. and Veran, S. (2013). Presence-only modelling using MAXENT: when can we trust the inferences?. *Methods. Ecol. Evol.*, 4: 236-243. https://doi.org/10.1111/2041-210x.12004

Yukimoto, Seiji; Koshiro, Tsuyoshi; Kawai, Hideaki; Oshima, Naga; Yoshida, Kohei; Urakawa, Shogo; Tsujino, Hiroyuki; Deushi, Makoto; Tanaka, Taichu; Hosaka, Masahiro; Yoshimura, Hiromasa; Shindo, Eiki; Mizuta, Ryo; Ishii, Masayoshi; Obata, Atsushi; Adachi, Yukimasa **(2019)**. *MRI MRI-ESM2.0 model output prepared for CMIP6 CMIP*. Version *20240410*[1].Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.621

Zhang, Z., Khederzadeh, S., & Li, Y. (2020). Deciphering the puzzles of dog domestication. *Zoological Research*, 41(2): 97-104. https://doi.org/10.24272/j.issn.2095-8137.2020.002

Zhang, M., Song, Y., Wang, C., Sun, G., Zhuang, L., Guo, M., Ren, L., Wangdue, S., Dong, G., Dai, Q., Cao, P., Yang, R., Liu, F., Feng, X., Bennett, E. A., Zhang, X., Chen, X., Wang, F., Luan, F., Dong, W., … Fu, Q. (2024). Ancient Mitogenomes Reveal the Maternal Genetic History of East Asian Dogs. *Molecular biology and evolution*, *41*(4), msae062. https://doi.org/10.1093/molbev/msae062

Summary of Results

A. Molecular mechanisms underlying vertebrate adaptive evolution

The main molecular mechanism underlying adaptive variation in vertebrates is mutations within regulatory elements and functional domains of regulatory proteins involved in gene regulation, cellular signalling or cellular pathways. Notably, in some environmental contexts down regulation of a gene or loss of a gene may yield advantageous phenotypes. Given the important role of gene regulation in adaptive evolution, further research into gene regulation, e.g., transcriptomic analysis, RNA interference and epigenetics would contribute to advancements in the field. Additionally, further research into the role of genome architecture the emergence of adaptive variation is warranted, given that mutations are more likely to occur in some regions within the genome others.

B. Genetic differentiation in Eurasian grey wolves and FRDs

Both grey wolves and FRDs showed large-scale spatial clustering across Eurasia, but genetic differentiation was weaker in FRDs than in grey wolves. Three of the employed approaches detected the genetic clusters for FRDs in Eurasian, Italian, Arabian and Indian Peninsulas. In grey wolves the inferred genetic clusters were from Iberian, Indian and Italian Peninsulas, as well as four genetic clusters from Southeastern Europe, Balkans, Eastern Europe and the rest of Eurasia. Southern Eurasian populations of both FRDs and wolves, specifically in the peninsulas, were genetically distinct. The only exception were grey wolves from the Arabian Peninsula which were not genetically differentiated. Whereas FRDs from the Arabian Peninsula were genetically differentiation. Wolves from the Indian Peninsula were the most genetically distinct population.

C. Genes driving evolutionary trajectories

There was wide overlap between genes with signatures of diversifying selection among FRDs and wolves. Genes involved in stimulus response, particularly those related to olfaction, were overrepresented among the genes under diversifying selection in both grey wolves and FRDs. In FRDs genes involved in morphological variation (e.g., Wnt Family Member 10) also showed signatures of diversifying selection. In grey wolves a notable gene with signatures of diversifying selection was LDL Receptor Related Protein 6, a functional entry receptor for canine distemper virus. A missense mutation predicted to affect protein function was detected in a solute carrier protein SLC6A9 in both taxa. SLC6A9 is a neurotransmitter (i.e., glycine) transporter that may be involved in behavioural or stimuli-processing adaptations. In FRDs an additional variant predicted to affect protein function was detected in the pattern recognition receptor NOD1 in FRDs.

D. Local adaptation

Based on the redundancy analysis approach, temperature during the wettest quarter of the year was detected as the main driver of genetic variation in both Eurasian FRDs and grey wolves. In both the employed approaches (i.e., LFMM and RDA) to detect associations between SNP markers and environmental variables, I detected three SNPs displaying signatures of local adaptation in both taxa. Thus, FRDs and grey wolves showed taxa-specific genes associated with local adaptation, however the genes are

involved with similar phenotypic traits (e.g., metabolism, sensory perception and immune system). Additionally, I detected associations between mammalian species richness and genes involved in olfactory reception and cardiac functions in both taxa.

Some notable adaptations identified in FRDs were in genes associated with thermoregulation (Heat shock protein 4), olfactory detection of plant oils (Olfactory Receptor 2W3), and metabolism in dogs on a raw meat diet (Mitochondrial ribosome recycling factor). Notable adaptations impacting the same phenotypic traits in wolves were detected in genes implicated in; the wet dog shake thermoregulatory behaviour (Transient Receptor Potential Cation Channel Subfamily M Member 8), olfactory receptor involved in detecting deer musk (Olfactory Receptor 5A2) and a mitochondrial gene which plays a role in carbohydrate metabolism (MGAT4A). Although over 1400 genes were detected to have signatures of adaptation, only 92 were common in both grey wolves and FRDs. Furthermore, 311 genes were detected to be under both local adaptation and diversifying selection in FRDs, while only 15 such genes were detected in grey wolves.

I also detected five genes displaying association with mammalian species richness in both FRDs and grey wolves. The genes associated with species richness in FRDs showed functions associated with olfactory system, development, cardiovascular system, and solute transport. In grey wolves the genes associated with species richness had functions related to the immune system, olfaction, metabolism and morphology. Notably, I detected an adaptation in grey wolves associated with mammalian species richness in the form of a missense mutation in an intestine-expressed protease (Calpain 13).

E. Adaptive potential under environmental shifts

Based on differences between current and required genetic makeup to survive under future environmental conditions, FRDs and grey wolves present different patterns of adaptive potential under climate change. FRDs displayed high adaptive potential and low risk of maladaptiveness in Southern Eurasia (warmer habitats), across all future predictions in all the socio-economic pathways. Across timeframes and socioeconomic pathways only subtle changes in adaptive potential and risk of maladaptiveness were detected in FRDs. Most wolf populations showed high adaptive potential and low risk of maladaptiveness in most future environmental predictions, except for the 2061-2100 timeframes under the regional rivalry and fossil fuel development socio-economic pathways. The northernmost populations of Eurasian grey wolves (i.e., in colder climates) are predicted to be the most negatively impacted by climatic change.

F. Dogs' suitable ecological habitats

I used dog presence data from the Global Biodiversity Information Facility to infer dogs' suitable ecological habitat to better capture the dogs' distribution across Eurasia. Precipitation of the coldest quarter of the year was the environmental variable with the most influence in the ensemble modelling of dogs' most suitable habitats across the continent. I inferred dogs' suitable ecological habitat for three timeframes, past (i.e., mid-Holocene), present and future (i.e., 2081-2100 under the fossil fuel development socio-economic pathway). Through all timeframes the most suitable habitat for dogs

was predicted to be in Southern Asia and then Europe. Furthermore, dogs' suitable habitat has been expanding since mid-Holocene and is projected to continue expanding. Although dogs' suitable habitat is predicted to expand overall, there are some areas where it is predicted to be reduced, i.e., Northeastern Europe.

General discussion and conclusions

Genetic differentiation in grey wolves and FRDs

Grey wolves showed stronger genetic differentiation than FRDs. In wolves and other wild canids (e.g., coyotes and jackals) habitat heterogeneity has been suggested to be the driver of genetic differentiation (Hinton et al., 2019; James et al., 2017; Leonard et al. 2014; Pilot et al. 2006; Rutkowski et al., 2015; Sacks et al., 2004; Schweizer et al. 2016). My findings support it; grey wolf genetic clustering patterns were consistent with variations in habitats, with population differentiation in par with habitat differentiation. In FRDs only the southernmost populations were genetically differentiated. Considering that FRDs are genetically distinct from pure-breed dogs (Pilot et al., 2015), their genetic structure is not directly influenced by artificial selection. Southern populations of both FRDs and grey wolves were genetically differentiated. Southern populations of another wild canid, the North American coyote, have also been detected to be genetically differentiated (Heppenheimer et al., 2018; Hinton et al., 2019). This suggests that selective pressures in the habitats of southernmost regions of the Northern Hemisphere operate strongly upon canids. Another factor that likely contributed to the differentiation of Southern Eurasian populations is the historical isolation of wolves in glacial refugia during the Last Glacial Maximum, after which Northern Asian populations recolonized northern Eurasia (Bergström et al. 2022). Overall, habitat heterogeneity affects genetic differentiation in grey wolves more strongly than in domestic dogs.

Genes driving adaptations in FRDs and grey wolves

322 genes that showed signatures of diversifying selection in both grey wolves and FRDs were involved in morphology, olfaction and immune systems (Chapter 2). This suggests that there is some degree of shared adaptive variation among FRDs and grey wolves. However, 311 genes in FRDs showed signatures of both local adaptation and diversifying selection, while wolves had only 15 genes overlapping between these lists (Chapter 3). Consequently, there was little overlap between the genes with signatures of local adaptation among taxa. These patterns suggest that in wolves the genes with signatures of diversifying selection may reflect genetic diversity underlying historical genetic differentiation among populations, which in turn could be older patterns of adaptive variation. Additionally, because of considerably lower sizes of wolf populations compared with FRDs, these patterns could be more strongly influenced by other processes in wolves, e.g., genetic drift and historic bottlenecks.

In FRDs the overlap between the genes with signatures or local adaptation and diversifying selection suggests that environmental pressures have a stronger influence on their genetic differentiation patterns and adaptive evolution across their range. Still, it is important to note that the overlap constitutes only 20% of the genes with signatures of local adaptation (Chapter 3) and 60% of the genes with signatures of diversifying selection (Chapter 2). Thus, our results show local adaptation in FRDs is more strongly influenced by diversifying selection than wolves, but other forms of selection are certainly at play as well. The genes in FRDs with signatures of both diversifying selection and local adaptation had functions related to immunity, olfactory perception and taste perception. This suggests that diverse phenotypes associated with different variants of these genes have been favoured during their evolution. FRDs have

evolved to their niche as a human commensal and within that to varying degrees of human dependence. FRDs are domestic animals but rely on their own abilities to attain most of their resources. Therefore, FRDs may be exhibiting an ongoing widespread adaptation under different natural and anthropogenic environments.

There was little overlap among the genes with signatures of local adaptation among FRDs and grey wolves. Although the genes were different, the phenotypic traits which local selective pressures operated upon where similar, i.e., morphology, sensory perception, immunity and metabolism. Grey wolves and FRDs have different ecological niches and consequently utilize their habitats differently, which could explain why the genes involved in their local adaptation processes were different. This is best exemplified by two adaptations detected in olfactory genes: in wolves the adaptation was related to detection of deer musk, whereas in FRDs the detected adaptation is involved in detection of plant oils (Yoshikawa et al., 2022; Wang et al., 2025). These adaptations both affect the olfactory abilities and reflect the respective ecological niches of these canids as apex predator and human commensal, respectively. Thus, across their range similar selective pressures impact genes involved in maintaining phenotypic diversity in both taxa. However, at the population level local selective pressures operate differently; they impact similar traits but result in unique adaptations.

Genes involved in sensory perception, immune system, morphology and metabolism play an important role in local adaption of vertebrates in general. Adaptations under local environmental pressures in other vertebrate linages has been detected in genes involved in these phenotypic traits (Chapter 1). This suggests that it may be a general pattern within the Chordata phylum transcending families or orders. Furthermore, most of the genes with signatures of local adaptation in both taxa had functions related to gene regulation, cellular or signalling pathways. This is also consistent with the patterns observed other vertebrate lineages (Chapter 1). Thus, our results highlight the role gene regulation and regulation of signalling and cellular pathways in vertebrates' adaptive evolution in response to environmental selective pressures. Rather than drastic phenotypic changes in functional domains of proteins responsible for specific phenotypic traits, the main mechanism underlying adaptive evolution is regulation of molecular processes. This may also be a mechanism by which organisms retain flexibility in their phenotypic traits to be better able to adapt under ever changing environmental conditions., by modifying rather than removing traits completely.

Implications for conservation of the adaptive potential of grey wolves and FRDs

Climate change is a pressing current issue and predictions on the potential of species adapting under climate change can inform population management strategies. Genomic offset statistics have been proved to be a valuable tool for making predictions of how species may adapt under climate change. My predictions on adaptive potential of grey wolves and FRDs can inform population management strategies in both taxa. In Europe and North America there have been numerous endeavours focused on the conservation and re-introduction of grey wolves (Gerber et al., 2024; Musiani et al., 2024; vonHoldt et al., 2024). A better understanding on their adaptive potential under climate change can contribute to efforts to maximize the adaptive potential of these populations.

I found that grey wolves' populations from Southern Eurasia retain the highest adaptive potential under climate change. Thus, it may benefit conservation efforts to focus on increasing genetic connectivity between Southern and Northern Eurasian populations, to introduce genetic variants which may increase their adaptive potential under climate change. Considering that Southern Asian populations live in habitats which have less conservation measures in place (Dinerstein et al., 2017), these are the most vulnerable populations. Thus, grey wolf populations with the highest potential to adapt under climate change are most threatened ones. It is also important to note that the decrease in adaptive potential was predicted in grey wolf populations under the regional rivalry (SSP 370) and fossil fuel development (SSP 585) socioeconomical pathways. Conversely, under the more sustainable socioeconomical pathways most populations are predicted to retain their adaptive potential, confirming that these pathways are the ones which will benefit the recuperation and maintenance of natural landscapes.

Population management efforts focused on FRDs are opposite to those focused on grey wolves. In almost all natural landscapes FRDs are considered as invasive and have a negative impact on ecological communities as competitors, predators and harbourers of diseases (Bryce, 2021; Vanak & Gompper, 2009; Wierzbowska et al., 2016). FRDs also negatively contribute to wildlife-human conflict, e.g. at times they are the perpetrators in the events which cause conflicts, e.g., predation on livestock, for which wild predators are held responsible (Plumer, et al., 2018; Hughes & Macdonald, 2013; Wierzbowska et al., 2016). My investigation suggested that most FRD populations do not retain an adaptive potential to survive without human assistance in temperate and cold habitats. This further supports our conclusion that their strong connectivity is facilitated by their relationship with humans. This investigation also suggested that FRDs' suitable habitats in natural landscapes will continue to expand under climate change into areas where they now have lower densities. Hence, their negative ecological role will be augmented in the future. We suggest that current efforts to reduce the numbers of FRDs may be stronger by targeting Southern Eurasian populations, with higher adaptive potentials.

Dogs' suitable ecological habitat and their domestication origin

There are limitations in the use of modern dogs to discern the dog domestication origin (Bergström et al., 2020; Bergström et al.., 2022; Larson et al., 2012). Thus far, studies on dog domestication centre have not reached a consensus, still some studies on primitive dogs (i.e., early dog lineages that may be genetically closer to grey wolves) (Wang et al., 2016) and ancient dogs (Frantz et al., 2016) propose domestication centres in Southern Asia. Genomic offset statistics showed that FRDs retain their adaptive potential for habitats in Southern Eurasia. Based on dogs' distribution modelling, their most suitable ecological habitat since the mid-Holocene is also predicted to be in Southern Eurasia. The approaches employed were very different; one was based on genetic markers and the other on species distribution modelling, and both showed similar results. If FRDs have retained the adaptive potential of their ancestors which were first domesticated, my results support the theory of a Southern Eurasian domestication centre. My results also showed that individuals from Southern Asia retained the highest adaptive potential and some individuals from Southern Europe came second. Based on species distribution modelling Southern Asia showed the most suitable habitat for dogs with Southern

Europe second. If the translocation of dogs occurred early enough in their domestication history, it could explain the ambiguity in some studies surrounding the domestication centre of dogs. Overall, my results suggest that dogs may have been initially domesticated in Southern Eurasia, and their ancestral population was adapted to warm habitats such as xeric shrublands and tropical forests.

General conclusions

1. Patterns of genetic differentiation in Southern Eurasian grey wolves and FRDs suggest that either barriers are limiting dispersal of both taxa or there is preservation of older patterns of genetic differentiation in these regions echoing a shared evolutionary past.
2. Under similar local selective pressures, FRDs and grey wolves show parallel adaptations achieved through different genetic pathways.
3. Southernmost populations of Eurasian FRDs and grey wolves have the highest potential to adapt under climate change.
4. Dogs' most suitable ecological habitats are in warmer regions and with climate change their suitable habitat will continue to expand further as temperatures continue to rise.
5. FRDs retain an adaptive potential for habitats in Southern Eurasia and dogs most suitable habitat for dogs is also in these regions, suggesting that dogs were first domesticated in Southern Eurasia.

References

Bryce, C. (2021). Dogs as pets and pests: Global patterns of canine abundance, activity, and health. *Integrative and Comparative Biology*, 61, https://doi.org/10.1093/icb/icab046

Chavez, D. E., Gronau, I., Hains, T., Dikow, R. B., Frandsen, P. B., Figueiró, H. V., Garcez, F. S., Tchaicka, L., de Paula, R. C., Rodrigues, F. H. G., Jorge, R. S. P., Lima, E. S., Songsasen, N., Johnson, W. E., Eizirik, E., Koepfli, K. P., & Wayne, R. K. (2022). Comparative genomics uncovers the evolutionary history, demography, and molecular adaptations of South American canids. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(34), e2205986119. https://doi.org/10.1073/pnas.2205986119

Dinerstein, E., Olson, D., Joshi, A., Vynne, C., Burgess, N. D., Wikramanayake, E., Hahn, N., Palminteri, S., Hedao, P., Noss, R., Hansen, M., Locke, H., Ellis, E. C., Jones, B., Barber, C. V., Hayes, R., Kormos, C., Martin, V., Crist, E., Sechrest, W., et al. (2017). An ecoregion-based approach to protecting half the terrestrial realm. *Bioscience, 67*(6), 534–545. https://doi.org/10.1093/biosci/bix014

Fleming, P., Nolan, H., Jackson, S., Ballard, G., Bengsen, A., Brown, W., Meek, P., Mifsud, G., Pal, S., & Sparkes, J. (2017). Roles for the Canidae in food webs reviewed: Where do they fit?. *Food Webs*, *12*, 14-34. https://doi.org/10.1016/j.fooweb.2017.03.001

Freedman, A. H., Gronau, I., Schweizer, R. M., et al. (2014). Genome sequencing highlights the dynamic early history of dogs. *PLoS Genetics, 10*(6), e1004016. https://doi.org/10.1371/journal.pgen.1004016

Frantz, L. A., Mullin, V. E., Pionnier-Capitan, M., Lebrasseur, O., Ollivier, M., Perri, A., Linderholm, A., Mattiangeli, V., Teasdale, M. D., Dimopoulos, E. A., Tresset, A., Duffraisse, M., McCormick, F., Bartosiewicz, L., Gál, E., Nyerges, É. A., Sablin, M. V., Bréhard, S., Mashkour, M., Bălăşescu, A., … Larson, G. (2016). Genomic and archaeological evidence suggest a dual origin of domestic dogs. *Science (New York, N.Y.)*, *352*(6290), 1228–1231. https://doi.org/10.1126/science.aaf3161

Gerber, N., Riesch, F., Bojarska, K., Zetsche, M., Rohwer, N.-K., Signer, J., Isselstein, J., Herzog, S., Okarma, H., Kuijper, D. P. J., & Balkenhol, N. (2024). Do recolonizing wolves trigger non-consumptive effects in European ecosystems? A review of evidence. *Wildlife Biology*, 2024, e01229. https://doi.org/10.1002/wlb3.01229

Gompper, M. E. (2014). Introduction: Outlining the ecological influences of a subsidized, domesticated predator. In Free-Ranging Dogs and Wildlife Conservation (pp. 1–8). Oxford University Press

Gonzalez Bernal, A., Martínez-Meyer, E., & Rojas-Soto, O. (2022). Climatic comparison of the gray wolf (*Canis lupus*) subspecies in North America using niche-based distribution models and its implications for conservation programs. *Journal of Mammalogy, 103*(10), 1305–1314. https://doi.org/10.1093/jmammal/gyac066

Hendricks, S.A., Schweizer, R.M. & Wayne, R.K. (2019). Conservation genomics illuminates the adaptive uniqueness of North American gray wolves. *Conservation Genetics, 20*, 29–43. https://doi.org/10.1007/s10592-018-1118-z

Heppenheimer, E., Cosio, D. S., Brzeski, K. E., Caudill, D., Van Why, K., Chamberlain, M. J., Hinton, J. W., & vonHoldt, B. (2018). Demographic history influences spatial patterns of genetic diversityin recently expanded coyote (*Canis latrans*) populations. *Heredity*, *120*(3), 183–195. https://doi.org/10.1038/s41437-017-0014-5

Hinton, J. W., Heppenheimer, E., West, K. M., Caudill, D., Karlin, M. L., Kilgo, J. C., Mayer, J. J., Miller, K. V., Walch, M., vonHoldt, B., & Chamberlain, M. J. (2019). Geographic patterns in morphometric and genetic variation for coyote populations with emphasis on southeastern coyotes. *Ecology and evolution*, *9*(6), 3389–3404. https://doi.org/10.1002/ece3.4966

Hughes, J. & Macdonald, D. (2013). A review of the interactions between free-roaming domestic dogs and wildlife. *Biological Conservation, 157*, 341–351. https://doi.org/10.1016/j.biocon.2012.07.005

James, R.S., Scott, D.M., Yarnell, R.W. & Overall, A.D.J. (2017). Food availability and population structure: How do clumped and abundant sources of carrion affect the genetic diversity of the black-backed jackal?, *Journal of Zoology*, *301,* 184–192. https://doi.org/10.1111/jzo.12407

Janeiro-Otero, A., Newsome, T., Van Eeden, L., Ripple, W., & Dormann, C. (2020). Grey wolf (*Canis lupus*) predation on livestock in relation to prey availability. *Biological Conservatio*n, *243*, 108433. https://doi.org/10.1016/j.biocon.2020.108433

Larson, G., Karlsson, E. K., Perri, A., Webster, M. T., Ho, S. Y., Peters, J., Stahl, P. W., Piper, P. J., Lingaas, F., Fredholm, M., Comstock, K. E., Modiano, J. F., Schelling, C., Agoulnik, A. I., Leegwater, P. A., Dobney, K., Vigne, J. D., Vilà, C., Andersson, L., & Lindblad-Toh, K. (2012). Rethinking dog domestication by integrating genetics, archeology, and biogeography. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(23), 8878–8883. https://doi.org/10.1073/pnas.1203005109

Lazzaroni, M., Range, F., Backes, J., Portele, K., Scheck, K., & Marshall-Pescini, S. (2020). The Effect of Domestication and Experience on the Social Interaction of Dogs and Wolves With a Human Companion. *Frontiers in psychology*, *11*, 785. https://doi.org/10.3389/fpsyg.2020.00785

Leonard, J. (2014). Ecology drives evolution in grey wolves. *Evolutionary Ecology Research*, 16, 461–473.

Lenormand, T. (2002). Gene flow and the limits to natural selection. *Trends in Ecology & Evolution*, *17*(4), 183-189. https://doi.org/10.1016/S0169-5347(02)02497-7

Losey, R. J., Drake, A. G., Ralrick, P. E., Jass, C. N., Lieverse, A. R., Bieraugle, M., Christenson, R., & Steuber, K. (2022). Dogs and wolves on the northern plains: A look

from beyond the site in Alberta. *Journal of Archaeological Science*, *148*, Article 105690. https://doi.org/10.1016/j.jas.2022.105690

Machado F. A. (2020). Selection and Constraints in the Ecomorphological Adaptive Evolution of the Skull of Living Canidae (Carnivora, Mammalia). *The American naturalist*, *196*(2), 197–215. https://doi.org/10.1086/709610

Mendoza Roldan, J., & Otranto, D. (2023) Zoonotic parasites associated with predation by dogs and cats. *Parasites Vectors, 16*, 55. https://doi.org/10.1186/s13071-023-05670-y

Muñoz-Fuentes, V., Darimont, C.T., Wayne, R.K., Paquet, P.C. & Leonard, J.A. (2009). Ecological factors drive differentiation in wolves from British Columbia. *Journal of Biogeography, 36*, 1516-1531. https://doi.org/10.1111/j.1365-2699.2008.02067.x

Musiani, M., & Randi, E. (2024). Conservation genomics of wolves: The global impact of RK Wayne's research. *The Journal of heredity*, *115*(4), 458–469. https://doi.org/10.1093/jhered/esae007

Newsome, T.M., Boitani, L., Chapron, G., Ciucci, P., Dickman, C.R., Dellinger, J.A., López-Bao, J.V., Peterson, R.O., Shores, C.R., Wirsing, A.J. and Ripple, W.J. (2016), Food habits of the world's grey wolves. *Mammal Review, 46*: 255-269. https://doi.org/10.1111/mam.12067

Perri, A. R., Feuerborn, T. R., Frantz, L. A. F., Larson, G., Malhi, R. S., Meltzer, D. J., & Witt, K. E. (2021). Dog domestication and the dual dispersal of people and dogs into the Americas. *Proceedings of the National Academy of Sciences of the United States of America*, 118(6), e2010083118. https://doi.org/10.1073/pnas.2010083118

Pierotti, R. (2024). Domestication and human/wildlife mutualism. *Humans*, 4(4), 371-384. https://doi.org/10.3390/humans4040024

Pilot, M., Greco, C., vonHoldt, B. M., Jędrzejewska, B., Randi, E., Jędrzejewski, W., Sidorovich, V. E., Ostrander, E. A., & Wayne, R. K. (2014). Genome-wide signatures of population bottlenecks and diversifying selection in European wolves. *Heredity*, 112(4), 428–442. https://doi.org/10.1038/hdy.2013.122

Pilot, M., Jedrzejewski, W., Branicki, W., Sidorovich, V. E., Jedrzejewska, B., Stachura, K., & Funk, S. M. (2006). Ecological factors influence population genetic structure of European grey wolves. *Molecular ecology*, *15*(14), 4533–4553. https://doi.org/10.1111/j.1365-294X.2006.03110.x

Pilot, M., Malewski, T., Moura, A. E., Grzybowski, T., & Oleński, K. (2015). On the origin of mongrels: Evolutionary history of free-breeding dogs in Eurasia. *Proceedings of the Royal Society B: Biological Sciences, 282*(20152189). https://doi.org/10.1098/rspb.2015.2189

Plumer, L., Talvi, T., Männil, P., Urmas, S. (2018). Assessing the roles of wolves and dogs in livestock predation with suggestions for mitigating human–wildlife conflict and conservation of wolves. *Conservation Genetics*, 19, 665–672. https://doi.org/10.1007/s10592-017-1045-4

Porto, L. M. V., Bennett, D., Maestri, R., & Etienne, R. S. (2024). Canids in a changing climate: predicting range shifts and evolutionary rescue under distinct future scenarios. *Biological Journal of the Linnean Society*, *143*(2), blae094. https://doi.org/10.1093/biolinnean/blae094

Porto, L. M. V., Etienne, R. S., & Maestri, R. (2023). Evolutionary radiation in canids following continental colonizations. *Evolution; international journal of organic evolution, 77*(4), 971–979. https://doi.org/10.1093/evolut/qpad015

Randi, E. (2011), Genetics and conservation of wolves *Canis lupus* in Europe. *Mammal Review, 41*, 99-111. https://doi.org/10.1111/j.1365-2907.2010.00176.x

Range, F., & Marshall-Pescini, S. (2022). Comparing wolves and dogs: current status and implications for human 'self-domestication'. *Trends in cognitive sciences*, *26*(4), 337–349. https://doi.org/10.1016/j.tics.2022.01.003

Rutkowski, R., Krofel, M., Giannatos, G., Ćirović, D., Männil, P., Volokh, A. M., Lanszki, J., Heltai, M., Szabó, L., Banea, O. C., Yavruyan, E., Hayrapetyan, V., Kopaliani, N., Miliou, A., Tryfonopoulos, G. A., Lymberakis, P., Penezić, A., Pakeltytė, G., Suchecka, E., & Bogdanowicz, W. (2015). A European Concern? Genetic Structure and Expansion of Golden Jackals (*Canis aureus*) in Europe and the Caucasus. *PloS one*, *10*(11), e0141236. https://doi.org/10.1371/journal.pone.0141236

Sacks, B.N., Brown, S.K. and Ernest, H.B. (2004). Population structure of California coyotes corresponds to habitat-specific breaks and illuminates species history. *Molecular Ecology*, 13: 1265–1275. https://doi.org/10.1111/j.1365-294X.2004.02110.x

Schweizer, R. M., vonHoldt, B. M., Harrigan, R., Knowles, J. C., Musiani, M., Coltman, D., Novembre, J., & Wayne, R. K. (2016). Genetic subdivision and candidate genes under selection in North American grey wolves. *Molecular Ecology, 25*(1), 380–402. https://doi.org/10.1111/mec.13364

Shannon, L. M., Boyko, R. H., Castelhano, M., & Boyko, A. R. (2015). Genetic structure in village dogs reveals a Central Asian domestication origin. *Proceedings of the National Academy of Sciences, 112*(44), 13639–13644. https://doi.org/10.1073/pnas.1516215112

Tancredi, D., & Cardinali, I. (2023). Being a Dog: A Review of the Domestication Process. Genes, 14(5), 992. https://doi.org/10.3390/genes14050992

Tangredi, B. P., & Lawler, D. F. (2024). How does evolutionary evaluation illuminate body size among canids?. *American journal of veterinary research*, *85*(4), ajvr.23.10.0248. https://doi.org/10.2460/ajvr.23.10.0248

Thalmann, O., Shapiro, B., Cui, P., Schuenemann, V. J., Sawyer, S. K., Greenfield, D. L., Germonpré, M. B., Sablin, M. V., López-Giráldez, F., Domingo-Roura, X., Napierala, H., Uerpmann, H. P., Loponte, D. M., Acosta, A. A., Giemsch, L., Schmitz, R. W., Worthington, B., Buikstra, J. E., Druzhkova, A., Graphodatsky, A. S., … Wayne, R. K. (2013). Complete mitochondrial genomes of ancient canids suggest a European origin of domestic dogs. *Science (New York, N.Y.)*, *342*(6160), 871–874. https://doi.org/10.1126/science.1243650

Torres, D.F., Oliveira, E.S., & Alves, R.R.N. (2020). Conflicts Between Humans and Terrestrial Vertebrates: A Global Review. *Tropical Conservation Science 11*(1). https://doi.org/10.1177/1940082918794084

Vanak, A.T. and Gompper, M.E. (2009). Dogs *Canis familiaris* as carnivores: their role and function in intraguild competition. *Mammal Review, 39*, 265-283. https://doi.org/10.1111/j.1365-2907.2009.00148.x

vonHoldt, B. M., Stahler, D. R., Brzeski, K. E., Musiani, M., Peterson, R., Phillips, M., Stephenson, J., Laudon, K., Meredith, E., Vucetich, J. A., Leonard, J. A., & Wayne, R. K. (2024). Demographic history shapes North American gray wolf genomic diversity and informs species' conservation. *Molecular Ecology*, 33(3), e17231. https://doi.org/10.1111/mec.17231

Vychodilova, L., Necesankova, M., Albrechtova, K., Hlavac, J., Modry, D., Janova, E., Vyskocil, M., Mihalca, A. D., Kennedy, L. J., & Horin, P. (2018). Genetic diversity and population structure of African village dogs based on microsatellite and immunity-related molecular markers. *PLOS ONE*, 13(6), e0199506. https://doi.org/10.1371/journal.pone.0199506

Wang, G. D., Zhai, W., Yang, H. C., Wang, L., Zhong, L., Liu, Y. H., Fan, R. X., Yin, T. T., Zhu, C. L., Poyarkov, A. D., Irwin, D. M., Hytönen, M. K., Lohi, H., Wu, C. I., Savolainen, P., & Zhang, Y. P. (2016). Out of southern East Asia: the natural history of domestic dogs across the world. *Cell research*, *26*(1), 21–33. https://doi.org/10.1038/cr.2015.147

Wang, J., Zhang, Q., Fan, W., Shi, Q., Mao, J., Xie, J., Chai, G., & Zhang, C. (2025). Deciphering olfactory receptor binding mechanisms: A structural and dynamic perspective on olfactory receptors. *Frontiers in Molecular Biosciences, 11*, 1498796. https://doi.org/10.3389/fmolb.2024.1498796

Wheat, C.H. & Wynne, C.D.L. (2025). The unfulfilled potential of dogs in studying behavioural ecology and evolution during the Anthropocene, *Animal Behaviour*, 219 (123020), ISSN 0003-3472. https://doi.org/10.1016/j.anbehav.2024.10.034

Wierzbowska, I. A., Hędrzak, M., Popczyk, B., Okarma, H., & Crooks, K. R. (2016). Predation of wildlife by free-ranging domestic dogs in Polish hunting grounds and potential competition with the grey wolf. *Biological Conservation*, 201, 1–9. https://doi.org/10.1016/j.biocon.2016.06.016

Wolf, C., & Ripple, W. J. (2017). Range contractions of the world's large carnivores. *Royal Society open science*, *4*(7), 170052. https://doi.org/10.1098/rsos.170052

Yoshikawa, K., Deguchi, J., Hu, J., Lu, H. Y., & Matsunami, H. (2022). An odorant receptor that senses four classes of musk compounds. *Current Biology, 32*(23), 5172–5179.e5. https://doi.org/10.1016/j.cub.2022.10.038

SUPPLEMENTARY DATA: THE EFFECT OF ENVIRONMENTAL FACTORS ON ADAPTIVE GENETIC VARIATION IN GREY WOLVES AND FREE-RANGING DOGS

Doctoral thesis

Francelly Martínez Sosa

Contents

Supplementary tables and figures: Chapter 2 ..... S4
Supplementary tables S5
Supplementary table S1. Pre-quality controls and post quality controls in FRDs and wolves by country S5
Supplementary table S2. SNPs which failed the CanFam6 Lifting S6
Supplementary table S3. g:Profiler excluded genes for grey wolves, for not having a UnitProt Code S7
Supplementary table S4. g:Profiler excluded genes in FRDs, for hot having a UnitProt Code S12
Supplementary Table S5. Genes with UnitProt codes automatically excluded by g:Profiler in each run S17

Supplementary figures S19
Supplementary figure 1. Principal component plots for addressing any potential taxonomic misidentification. S19
Supplementary figure 2. Cross validation and cross-entropy plots for population structure approaches. S20
Supplementary figure 3. Bayesian index criterion (BIC) for wolves and dogs and a-score optimization for the Discriminant Analysis of Principal Components (DAPC). S21
Supplementary figure 4. Discriminant Analysis of Principal Components (DAPC) density plots. S22
Supplementary figure 5. Population structure histograms wolves. S25
Supplementary figure 6. Population structure histograms FRDs. S28
Supplementary figure 7. Spatial principal component analysis (sPCA) plots. S29
Supplementary figure 8. Histograms of Monte Carlo simulations for global and local structures. S30
Supplementary Figure 9. Density plots and Moran I scatter plots. S32
Supplementary figure 10. Map of the spatial principal component analysis (sPCA). S33
Supplementary figure 11. PCAdapt results. S34
Supplementary Figure 12. Manhattan plots of Bayescan results. S35
Supplementary figure 13. Variant effect predictor (VEP) results summary. S36
Supplementary figure 14. STRING network of genes under diversifying selection wolves S41
Supplementary figure 15. STRING Network for genes under diversifying selection FRDs. S42
Supplementary figure 16. Heat maps of biological processes gene ontology enriched terms from g:Profiler Wolves S43
Supplementary figure 17. Heat maps of biological processes gene ontology enriched terms from g:Profiler FRDs. S44
Supplementary figure 18. STRING Networks and heatmaps of the largest STRING sub-clusters Wolves. S45
Supplementary figure 19. STRING Networks and heatmaps of the largest STRING sub-clusters FRDs. S47

Supplementary figure 20. Mapped genotype frequencies fort the SLC6A9..... S49

Supplementary tables and figures: Chapter 3..... S50
Custom script for species richness rasters S51
Supplementary tables..... S64
Supplementary table 1. Number of individuals per species for each biome S64
Supplementary table 2. SNPs which failed the CanFam6 Lifting S65
Supplementary table 3. Genes without a UnitProt codes, excluded from g:Profiler analysis S66
Supplementary table 4. Genes with UnitProt codes automatically excluded by g:Profiler S88
Supplementary table 5. Variant Effect Predictor (VEP) results summary wolves. S90
Supplementary table 6. Variant Effect Predictor (VEP) results summary FRDs. S124
Supplementary table 7. g:Profiler results summary for wolves and FRDs using the dog and human genomes S164
Supplementary table 8. Genes displaying signatures of both diversifying selection and local adaptation S166
Supplementary table 9. Uncharacterized and genes without UnitProt codes, displaying signatures of both local adaptation and diversifying selection S172
Supplementary table 10. RNA and pseudogenes displaying signatures of both local adaptation and diversifying selection S174

Supplementary figures.....S177
Supplementary figure 1. Corrplot for collinearity for the BioClim environmental predictors wolves. S178
Supplementary figure 2. Corrplot for collinearity for the BioClim environmental predictors FRDs.....S178
Supplementary figure 3. Latent factor mixed model (LFMM) plots wolves. S181
Supplementary figure 4 Latent factor mixed model (LFMM) plots FRDs.....S184
Supplementary figure 5. Manhattan plots for latent factor mixed models (LFMM) SNP candidates Wolves.....S185
Supplementary figure 6. Manhattan plots for latent factor mixed models (LFMM) SNP candidates FRDs. S186
Supplementary figure 7. RDA histograms wolves. S187
Supplementary figure 8. RDA histograms FRDs. S189
Supplementary figure 9. Ordination in space of candidate SNPs as calculated in the RDA.....S192
Supplementary figure 10. g:Profiler bubble plots of gene ontology (GO) enrichment analysis. S196
Supplementary figure 11. Wolves' Complete STRING Network. Complete STRING network for the genes displaying signatures of local adaptation. S197
Supplementary figure 12. Wolves' complete STRING network, species richness analysis. S198
Supplementary figure 13. FRDs' complete STRING network. S199

Supplementary figure 14. FRDs' complete STRING network, species richness analysis. S200
Supplementary figure 15. Summary of the VEP results for the genes predicted to have missense mutations in wolves. S201
Supplementary figure 16. Summary of the VEP results for the genes predicted to have missense mutations in FRDs S202
Supplementary figure 17. RDA ordination in space for candidate SNPs displaying signatures of local adaptive evolution and environmental vectors. S204
Supplementary Figure 18. The STRING functional protein association sub-networks. S209
Supplementary figure 19. STRING protein network genes under diversifying and local adaptation in FRDs. S210

Supplementary figures and tables: Chapter 4 ..... S212
Supplementary figure 1. Genomic offset and risk of nonadaptiveness (RONA) mapped results for FRDs. S216
Supplementary figure 2. Genomic offset and risk of nonadaptiveness (RONA) mapped results for wolves. S222
Supplementary figure 3. Evaluation of single species distribution models. S224
Supplementary figure 4. Evaluations for ensemble model. S226
Supplementary figure 5. Suitable ecological niche modelling ensemble model for current environmental conditions. S228
Supplementary figure 6. Suitable ecological niche modelling ensemble model with mid-Holocene environmental rasters, ... S230
Supplementary figure 7. Suitable ecological niche modelling ensemble model with SSP 585 environmental rasters, ... S232
Supplementary figure 8. Range difference based on the mid-Holocene and current suitable habitats models. S234
Supplementary figure 9. Range difference based on the current and future suitable habitats models. S236

Supplementary tables and figures: Chapter 2

Supplementary tables

Supplementary table S1. Pre-quality controls and post quality controls in FRDs and wolves by country

Country	Pre-QC dogs	Dogs included	Pre-QC wolves	Wolves included
Afghanistan	0	0	1	1
Arabia	45	42	11	6
Belarus	37	31	34	14
Bosnia	3	3	36	23
Bulgaria	21	20	23	22
Georgia	4	4	8	7
Greece	0	0	6	6
Hungary	1	1	0	0
India	46	32	25	15
Italy	47	20	22	21
Mongolia	0	0	7	3
Morocco	199	99	0	0
Poland	76	73	6	6
Romania	0	0	12	12
Russia	23	22	16	12
Serbia	26	24	55	47
Spain	0	0	11	6
Ukraine	55	49	38	28
Uzbekistan	13	12	2	2
Yakutia	17	16	1	1
Totals	613	370	315	232

Supplementary table S2. SNPs which failed the CanFam6 Lifting

Wolves SNPs	***Dog SNPs***
chr27:12221495-12221495	chr15:7966269-7966269

*the error was: "#Deleted in new"

Supplementary table S3. g:Profiler excluded genes for grey wolves, for not having a UnitProt Code

Gene	*Gene classification*	*Consequence*
ENSCAFG00000055243	miRNA	upstream_gene_variant
ENSCAFG00000057961	lncRNA	upstream_gene_variant
ENSCAFG00000054112	lncRNA	upstream_gene_variant
ENSCAFG00000021107	snRNA	upstream_gene_variant
ENSCAFG00000054466	lncRNA	upstream_gene_variant
ENSCAFG00000056528	lncRNA	upstream_gene_variant
ENSCAFG00000055598	lncRNA	upstream_gene_variant
ENSCAFG00000053754	lncRNA	upstream_gene_variant
ENSCAFG00000055554	lncRNA	upstream_gene_variant
ENSCAFG00000058557	lncRNA	upstream_gene_variant
ENSCAFG00000022520	snRNA	upstream_gene_variant
ENSCAFG00000028119	ribozyme	upstream_gene_variant
ENSCAFG00000048801	lncRNA	upstream_gene_variant
ENSCAFG00000053697	snRNA	upstream_gene_variant
ENSCAFG00000054332	snRNA	upstream_gene_variant
ENSCAFG00000052727	lncRNA	upstream_gene_variant
ENSCAFG00000054065	lncRNA	upstream_gene_variant
ENSCAFG00000055787	lncRNA	upstream_gene_variant
ENSCAFG00000051921	lncRNA	upstream_gene_variant
ENSCAFG00000021014	snRNA	upstream_gene_variant
ENSCAFG00000045215	miRNA	upstream_gene_variant
ENSCAFG00000056265	lncRNA	upstream_gene_variant
ENSCAFG00000022253	snRNA	upstream_gene_variant
ENSCAFG00000026082	snoRNA	upstream_gene_variant
ENSCAFG00000058534	lncRNA	upstream_gene_variant
ENSCAFG00000059178	lncRNA	upstream_gene_variant
ENSCAFG00000020547	miRNA	upstream_gene_variant
ENSCAFG00000038294	lncRNA	upstream_gene_variant
ENSCAFG00000056504	lncRNA	upstream_gene_variant
ENSCAFG00000051332	lncRNA	upstream_gene_variant
ENSCAFG00000029549	snRNA	upstream_gene_variant
ENSCAFG00000055501	snRNA	upstream_gene_variant
ENSCAFG00000055132	lncRNA	upstream_gene_variant
ENSCAFG00000056889	lncRNA	upstream_gene_variant
ENSCAFG00000046301	lncRNA	upstream_gene_variant
ENSCAFG00000058868	lncRNA	upstream_gene_variant
ENSCAFG00000052308	lncRNA	upstream_gene_variant
ENSCAFG00000054483	miRNA	upstream_gene_variant
ENSCAFG00000022936	miRNA	upstream_gene_variant
ENSCAFG00000047324	lncRNA	upstream_gene_variant
ENSCAFG00000055031	lncRNA	upstream_gene_variant
ENSCAFG00000056699	lncRNA	upstream_gene_variant
ENSCAFG00000052713	lncRNA	upstream_gene_variant
ENSCAFG00000051637	lncRNA	upstream_gene_variant
ENSCAFG00000057445	lncRNA	upstream_gene_variant

ENSCAFG00000053959	lncRNA	upstream_gene_variant
ENSCAFG00000057071	lncRNA	upstream_gene_variant
ENSCAFG00000051070	lncRNA	upstream_gene_variant
ENSCAFG00000056859	lncRNA	upstream_gene_variant
ENSCAFG00000050366	lncRNA	upstream_gene_variant
ENSCAFG00000053965	snRNA	upstream_gene_variant
ENSCAFG00000059447	lncRNA	upstream_gene_variant
ENSCAFG00000049146	pseudogene	upstream_gene_variant
ENSCAFG00000049284	pseudogene	upstream_gene_variant
ENSCAFG00000043459	pseudogene	upstream_gene_variant
ENSCAFG00000041221	pseudogene	downstream_gene_variant
ENSCAFG00000055685	miRNA	downstream_gene_variant
ENSCAFG00000057785	lncRNA	downstream_gene_variant
ENSCAFG00000059318	lncRNA	downstream_gene_variant
ENSCAFG00000042389	lncRNA	downstream_gene_variant
ENSCAFG00000052163	lncRNA	downstream_gene_variant
ENSCAFG00000058242	lncRNA	downstream_gene_variant
ENSCAFG00000052823	lncRNA	downstream_gene_variant
ENSCAFG00000051086	miRNA	downstream_gene_variant
ENSCAFG00000052519	miRNA	downstream_gene_variant
ENSCAFG00000059058	lncRNA	downstream_gene_variant
ENSCAFG00000055381	snRNA	downstream_gene_variant
ENSCAFG00000052972	lncRNA	downstream_gene_variant
ENSCAFG00000055439	lncRNA	downstream_gene_variant
ENSCAFG00000053648	lncRNA	downstream_gene_variant
ENSCAFG00000057394	lncRNA	downstream_gene_variant
ENSCAFG00000052571	lncRNA	downstream_gene_variant
ENSCAFG00000054467	lncRNA	downstream_gene_variant
ENSCAFG00000058413	lncRNA	downstream_gene_variant
ENSCAFG00000048199	miRNA	downstream_gene_variant
ENSCAFG00000056045	snRNA	downstream_gene_variant
ENSCAFG00000057983	snRNA	downstream_gene_variant
ENSCAFG00000055117	snRNA	downstream_gene_variant
ENSCAFG00000053577	snRNA	downstream_gene_variant
ENSCAFG00000054320	snRNA	downstream_gene_variant
ENSCAFG00000052555	lncRNA	downstream_gene_variant
ENSCAFG00000054099	lncRNA	downstream_gene_variant
ENSCAFG00000056062	lncRNA	downstream_gene_variant
ENSCAFG00000053075	lncRNA	downstream_gene_variant
ENSCAFG00000020981	snoRNA	downstream_gene_variant
ENSCAFG00000026284	snoRNA	downstream_gene_variant
ENSCAFG00000056330	lncRNA	downstream_gene_variant
ENSCAFG00000043782	snRNA	downstream_gene_variant
ENSCAFG00000057075	lncRNA	downstream_gene_variant
ENSCAFG00000059383	lncRNA	downstream_gene_variant
ENSCAFG00000054039	snRNA	downstream_gene_variant
ENSCAFG00000045449	lncRNA	downstream_gene_variant
ENSCAFG00000047122	pseudogene	downstream_gene_variant

ENSCAFG00000041603	lncRNA	downstream_gene_variant
ENSCAFG00000058170	lncRNA	downstream_gene_variant
ENSCAFG00000054389	lncRNA	downstream_gene_variant
ENSCAFG00000056490	pseudogene	downstream_gene_variant
ENSCAFG00000058863	lncRNA	downstream_gene_variant
ENSCAFG00000025934	snRNA	downstream_gene_variant
ENSCAFG00000042252	misc_RNA	downstream_gene_variant
ENSCAFG00000051434	lncRNA	downstream_gene_variant
ENSCAFG00000059098	lncRNA	downstream_gene_variant
ENSCAFG00000057930	lncRNA	downstream_gene_variant
ENSCAFG00000053565	lncRNA	downstream_gene_variant
ENSCAFG00000054596	lncRNA	downstream_gene_variant
ENSCAFG00000057389	lncRNA	downstream_gene_variant
ENSCAFG00000051752	lncRNA	downstream_gene_variant
ENSCAFG00000056979	lncRNA	downstream_gene_variant
ENSCAFG00000057058	miRNA	downstream_gene_variant
ENSCAFG00000055553	miRNA	downstream_gene_variant
ENSCAFG00000043087	snRNA	downstream_gene_variant
ENSCAFG00000050225	lncRNA	downstream_gene_variant
ENSCAFG00000053899	lncRNA	downstream_gene_variant
ENSCAFG00000058374	lncRNA	downstream_gene_variant
ENSCAFG00000050237	lncRNA	downstream_gene_variant
ENSCAFG00000058856	lncRNA	downstream_gene_variant
ENSCAFG00000053060	lncRNA	downstream_gene_variant
ENSCAFG00000052304	lncRNA	downstream_gene_variant
ENSCAFG00000053712	lncRNA	downstream_gene_variant
ENSCAFG00000053492	lncRNA	downstream_gene_variant
ENSCAFG00000059337	lncRNA	downstream_gene_variant
ENSCAFG00000055388	lncRNA	downstream_gene_variant
ENSCAFG00000053745	lncRNA	downstream_gene_variant
ENSCAFG00000056038	lncRNA	downstream_gene_variant
ENSCAFG00000052098	lncRNA	downstream_gene_variant
ENSCAFG00000028401	snRNA	downstream_gene_variant
ENSCAFG00000044330	pseudogene	downstream_gene_variant
ENSCAFG00000043307	pseudogene	downstream_gene_variant
ENSCAFG00000058167	lncRNA	downstream_gene_variant
ENSCAFG00000050358	lncRNA	non_coding_transcript_exon_variant
ENSCAFG00000048801	lncRNA	intron_variant, non_coding_transcript_variant
ENSCAFG00000047324	lncRNA	intron_variant, non_coding_transcript_variant
ENSCAFG00000050237	lncRNA	intron_variant, non_coding_transcript_variant
ENSCAFG00000028550	Zinc finger protein 850-like	upstream_gene_variant
ENSCAFG00000041829	Interferon lambda-4-like	upstream_gene_variant
ENSCAFG00000047369	Uncharacterized protein	upstream_gene_variant
ENSCAFG00000045033	Uncharacterized protein	upstream_gene_variant
ENSCAFG00000058699	Ig-like domain-containing protein	upstream_gene_variant
ENSCAFG00000055775	Ig-like domain-containing protein	upstream_gene_variant
ENSCAFG00000057398	Ig-like domain-containing protein	upstream_gene_variant

ENSCAFG00000050263	Immunoglobulin V-set domain-containing protein	upstream_gene_variant
ENSCAFG00000045242	Ig-like domain-containing protein	upstream_gene_variant
ENSCAFG00000044478	Small ribosomal subunit protein uS15 N-terminal domain-containing protein	upstream_gene_variant
ENSCAFG00000054905	Ig-like domain-containing protein	upstream_gene_variant
ENSCAFG00000047905	Uncharacterized protein	upstream_gene_variant
ENSCAFG00000054560	Cyclin-dependent kinases regulatory subunit	upstream_gene_variant
ENSCAFG00000051647	Uncharacterized protein	upstream_gene_variant
ENSCAFG00000055958	60S ribosomal protein L39-like	upstream_gene_variant
ENSCAFG00000014478	Ig-like domain-containing protein	upstream_gene_variant
ENSCAFG00000031359	Immunoglobulin V-set domain-containing protein	upstream_gene_variant
ENSCAFG00000045360	Ig-like domain-containing protein	upstream_gene_variant
ENSCAFG00000043031	Immunoglobulin V-set domain-containing protein	upstream_gene_variant
ENSCAFG00000056839	Ig-like domain-containing protein	upstream_gene_variant
ENSCAFG00000058265	Immunoglobulin C1-set domain-containing protein	upstream_gene_variant
ENSCAFG00000055010	Olfactory receptor	upstream_gene_variant
ENSCAFG00000041808	Uncharacterized protein	upstream_gene_variant
ENSCAFG00000016551	Prefoldin subunit 4	upstream_gene_variant
ENSCAFG00000049889	Mitochondrial import inner membrane translocase subunit TIM23	downstream_gene_variant
ENSCAFG00000046247	Zinc finger protein 671	downstream_gene_variant
ENSCAFG00000052759	Interferon lambda-4-like	downstream_gene_variant
ENSCAFG00000016344	ADP-ribosylation factor-like protein 2-binding protein (ARF-like 2-binding protein)	downstream_gene_variant
ENSCAFG00000042815	ADP/ATP translocase (ADP,ATP carrier protein)	downstream_gene_variant
ENSCAFG00000049311	MICOS complex subunit	downstream_gene_variant
ENSCAFG00000045372	Transmembrane protein 69	downstream_gene_variant
ENSCAFG00000049482	Ig-like domain-containing protein	downstream_gene_variant
ENSCAFG00000051129	MHC class I-like antigen recognition-like domain-containing protein	downstream_gene_variant
ENSCAFG00000042292	Large ribosomal subunit protein uL10 (60S acidic ribosomal protein P0)	downstream_gene_variant
ENSCAFG00000003811	Ig-like domain-containing protein	downstream_gene_variant
ENSCAFG00000024810	Ig-like domain-containing protein	downstream_gene_variant
ENSCAFG00000029837	Ig-like domain-containing protein	downstream_gene_variant
ENSCAFG00000046664	Ig-like domain-containing protein	downstream_gene_variant
ENSCAFG00000055018	Ig-like domain-containing protein	downstream_gene_variant
ENSCAFG00000057117	Ig-like domain-containing protein	downstream_gene_variant
ENSCAFG00000052806	Ig-like domain-containing protein	downstream_gene_variant
ENSCAFG00000055372	Immunoglobulin V-set domain-containing protein	downstream_gene_variant
ENSCAFG00000057504	Ig-like domain-containing protein	downstream_gene_variant
ENSCAFG00000058902	Cornichon family AMPA receptor auxiliary protein 1	downstream_gene_variant
ENSCAFG00000009787	Pre-mRNA-splicing regulator WTAP (Female-lethal(2)D homolog) (WT1-associated protein) (Wilms tumor 1-associating protein)	downstream_gene_variant
ENSCAFG00000055600	Immunoglobulin V-set domain-containing protein	5_prime_UTR_variant

ENSCAFG00000047151	Olfactory receptor	synonymous_variant
ENSCAFG00000042267	KRAB domain-containing protein	intron_variant
ENSCAFG00000042586	Ig-like domain-containing protein	intron_variant

Supplementary table S4. g:Profiler excluded genes in FRDs, for hot having a UnitProt Code

Genes without a UnitProt Code

ENSMBL ID	UnitProt Gene Name	Consequence
ENSCAFG00000051032	Glyceraldehyde 3-phosphate dehydrogenase NAD(P) binding domain-containing protein	missense_variant
ENSCAFG00000028550	Zinc finger protein 850-like	upstream_gene_variant
ENSCAFG00000037135	KRAB domain-containing protein	upstream_gene_variant
ENSCAFG00000047812	KRAB domain-containing protein	upstream_gene_variant
ENSCAFG00000048373	C2H2-type domain-containing protein	upstream_gene_variant
ENSCAFG00000044781	Uncharacterized protein	upstream_gene_variant
ENSCAFG00000041829	Interferon lambda-4-like	upstream_gene_variant
ENSCAFG00000045033	Uncharacterized protein	upstream_gene_variant
ENSCAFG00000058699	Ig-like domain-containing protein	upstream_gene_variant
ENSCAFG00000055775	Ig-like domain-containing protein	upstream_gene_variant
ENSCAFG00000057398	Ig-like domain-containing protein	upstream_gene_variant
ENSCAFG00000050263	Immunoglobulin V-set domain-containing protein	upstream_gene_variant
ENSCAFG00000045242	Ig-like domain-containing protein	upstream_gene_variant
ENSCAFG00000044478	Small ribosomal subunit protein uS15 N-terminal domain-containing protein	upstream_gene_variant
ENSCAFG00000054905	Ig-like domain-containing protein	upstream_gene_variant
ENSCAFG00000051647	Uncharacterized protein	upstream_gene_variant
ENSCAFG00000014478	Ig-like domain-containing protein	upstream_gene_variant
ENSCAFG00000031359	Immunoglobulin V-set domain-containing protein	upstream_gene_variant
ENSCAFG00000045360	Ig-like domain-containing protein	upstream_gene_variant
ENSCAFG00000043031	Immunoglobulin V-set domain-containing protein	upstream_gene_variant
ENSCAFG00000056839	Ig-like domain-containing protein	upstream_gene_variant
ENSCAFG00000058265	Immunoglobulin C1-set domain-containing protein	upstream_gene_variant
ENSCAFG00000055010	Olfactory receptor	upstream_gene_variant
ENSCAFG00000041808	Uncharacterized protein	upstream_gene_variant
ENSCAFG00000030477	Large ribosomal subunit protein eL37 (60S ribosomal protein L37)	upstream_gene_variant
ENSCAFG00000016551	Prefoldin subunit 4	upstream_gene_variant
ENSCAFG00000047974	Uncharacterized protein	upstream_gene_variant
ENSCAFG00000049889	Mitochondrial import inner membrane translocase subunit TIM23	downstream_gene_variant
ENSCAFG00000046247	Zinc finger protein 671	downstream_gene_variant
ENSCAFG00000000014	Large ribosomal subunit protein uL23 (60S ribosomal protein L23a)	downstream_gene_variant
ENSCAFG00000058280	Uncharacterized protein	downstream_gene_variant
ENSCAFG00000052759	Interferon lambda-4-like	downstream_gene_variant
ENSCAFG00000049482	Ig-like domain-containing protein	downstream_gene_variant
ENSCAFG00000059114	Uncharacterized protein	downstream_gene_variant
ENSCAFG00000051129	MHC class I-like antigen recognition-like domain-containing protein	downstream_gene_variant
ENSCAFG00000023580	EGF-like domain-containing protein	downstream_gene_variant
ENSCAFG00000049091	Protein SDA1	downstream_gene_variant
ENSCAFG00000003804	Coiled-coil domain containing 162	downstream_gene_variant
ENSCAFG00000003811	Ig-like domain-containing protein	downstream_gene_variant
ENSCAFG00000024810	Ig-like domain-containing protein	downstream_gene_variant
ENSCAFG00000029837	Immunoglobulin V-set domain-containing protein	downstream_gene_variant

ENSCAFG00000046664	Ig-like domain-containing protein	downstream_gene_variant
ENSCAFG00000055018	Ig-like domain-containing protein	downstream_gene_variant
ENSCAFG00000057117	Ig-like domain-containing protein	downstream_gene_variant
ENSCAFG00000052806	Ig-like domain-containing protein	downstream_gene_variant
ENSCAFG00000055372	Immunoglobulin V-set domain-containing protein	downstream_gene_variant
ENSCAFG00000057504	Ig-like domain-containing protein	downstream_gene_variant
ENSCAFG00000058902	Cornichon family AMPA receptor auxiliary protein 1	downstream_gene_variant
ENSCAFG00000052503	L1 transposable element RRM domain-containing protein	downstream_gene_variant
ENSCAFG00000053426	Protein kinase domain-containing protein	downstream_gene_variant
ENSCAFG00000055600	Immunoglobulin V-set domain-containing protein	5_prime_UTR_variant
ENSCAFG00000047151	Olfactory receptor	3_prime_UTR_variant
ENSCAFG00000042267	KRAB domain-containing protein	intron_variant
ENSCAFG00000042586	Ig-like domain-containing protein	intron_variant
ENSCAFG00000055243	miRNA	upstream_gene_variant
ENSCAFG00000057961	lncRNA	upstream_gene_variant
ENSCAFG00000056556	lncRNA	upstream_gene_variant
ENSCAFG00000054112	lncRNA	upstream_gene_variant
ENSCAFG00000054466	lncRNA	upstream_gene_variant
ENSCAFG00000028170	snoRNA	upstream_gene_variant
ENSCAFG00000022149	snoRNA	upstream_gene_variant
ENSCAFG00000056528	lncRNA	upstream_gene_variant
ENSCAFG00000055598	lncRNA	upstream_gene_variant
ENSCAFG00000053754	lncRNA	upstream_gene_variant
ENSCAFG00000057554	lncRNA	upstream_gene_variant
ENSCAFG00000053086	lncRNA	upstream_gene_variant
ENSCAFG00000056022	lncRNA	upstream_gene_variant
ENSCAFG00000057631	lncRNA	upstream_gene_variant
ENSCAFG00000055160	lncRNA	upstream_gene_variant
ENSCAFG00000022520	snRNA	upstream_gene_variant
ENSCAFG00000020642	miRNA	upstream_gene_variant
ENSCAFG00000054033	lncRNA	upstream_gene_variant
ENSCAFG00000050476	lncRNA	upstream_gene_variant
ENSCAFG00000052727	lncRNA	upstream_gene_variant
ENSCAFG00000054065	lncRNA	upstream_gene_variant
ENSCAFG00000051245	lncRNA	upstream_gene_variant
ENSCAFG00000055787	lncRNA	upstream_gene_variant
ENSCAFG00000051921	lncRNA	upstream_gene_variant
ENSCAFG00000022225	snRNA	upstream_gene_variant
ENSCAFG00000021014	snRNA	upstream_gene_variant
ENSCAFG00000045215	miRNA	upstream_gene_variant
ENSCAFG00000056265	lncRNA	upstream_gene_variant
ENSCAFG00000022253	snRNA	upstream_gene_variant
ENSCAFG00000026082	snoRNA	upstream_gene_variant
ENSCAFG00000057466	lncRNA	upstream_gene_variant
ENSCAFG00000038294	lncRNA	upstream_gene_variant
ENSCAFG00000056504	lncRNA	upstream_gene_variant
ENSCAFG00000052308	lncRNA	upstream_gene_variant
ENSCAFG00000054483	miRNA	upstream_gene_variant

ENSCAFG00000053893	snRNA	upstream_gene_variant
ENSCAFG00000047324	lncRNA	upstream_gene_variant
ENSCAFG00000055031	lncRNA	upstream_gene_variant
ENSCAFG00000056699	lncRNA	upstream_gene_variant
ENSCAFG00000052713	lncRNA	upstream_gene_variant
ENSCAFG00000051637	lncRNA	upstream_gene_variant
ENSCAFG00000057445	lncRNA	upstream_gene_variant
ENSCAFG00000053959	lncRNA	upstream_gene_variant
ENSCAFG00000057071	lncRNA	upstream_gene_variant
ENSCAFG00000051070	lncRNA	upstream_gene_variant
ENSCAFG00000056859	lncRNA	upstream_gene_variant
ENSCAFG00000050366	lncRNA	upstream_gene_variant
ENSCAFG00000053965	snRNA	upstream_gene_variant
ENSCAFG00000021387	snRNA	upstream_gene_variant
ENSCAFG00000035239	lncRNA	upstream_gene_variant
ENSCAFG00000057488	lncRNA	upstream_gene_variant
ENSCAFG00000049146	pseudogene	upstream_gene_variant
ENSCAFG00000049284	pseudogene	upstream_gene_variant
ENSCAFG00000043459	pseudogene	upstream_gene_variant
ENSCAFG00000020965	snRNA	upstream_gene_variant
ENSCAFG00000054224	lncRNA	upstream_gene_variant
ENSCAFG00000056666	lncRNA	upstream_gene_variant
ENSCAFG00000053720	lncRNA	downstream_gene_variant
ENSCAFG00000041221	pseudogene	downstream_gene_variant
ENSCAFG00000027148	snRNA	downstream_gene_variant
ENSCAFG00000050516	lncRNA	downstream_gene_variant
ENSCAFG00000050153	lncRNA	downstream_gene_variant
ENSCAFG00000057785	lncRNA	downstream_gene_variant
ENSCAFG00000059318	lncRNA	downstream_gene_variant
ENSCAFG00000054532	lncRNA	downstream_gene_variant
ENSCAFG00000042389	lncRNA	downstream_gene_variant
ENSCAFG00000052163	lncRNA	downstream_gene_variant
ENSCAFG00000058242	lncRNA	downstream_gene_variant
ENSCAFG00000052823	lncRNA	downstream_gene_variant
ENSCAFG00000052519	miRNA	downstream_gene_variant
ENSCAFG00000059058	lncRNA	downstream_gene_variant
ENSCAFG00000055381	snRNA	downstream_gene_variant
ENSCAFG00000052972	lncRNA	downstream_gene_variant
ENSCAFG00000055439	lncRNA	downstream_gene_variant
ENSCAFG00000053648	lncRNA	downstream_gene_variant
ENSCAFG00000050998	lncRNA	downstream_gene_variant
ENSCAFG00000050124	lncRNA	downstream_gene_variant
ENSCAFG00000050865	lncRNA	downstream_gene_variant
ENSCAFG00000025692	miRNA	downstream_gene_variant
ENSCAFG00000048471	lncRNA	downstream_gene_variant
ENSCAFG00000052435	lncRNA	downstream_gene_variant
ENSCAFG00000054099	lncRNA	downstream_gene_variant
ENSCAFG00000056062	lncRNA	downstream_gene_variant

ENSCAFG00000054931	lncRNA	downstream_gene_variant
ENSCAFG00000059425	miRNA	downstream_gene_variant
ENSCAFG00000055765	miRNA	downstream_gene_variant
ENSCAFG00000054793	lncRNA	downstream_gene_variant
ENSCAFG00000053075	lncRNA	downstream_gene_variant
ENSCAFG00000020981	snoRNA	downstream_gene_variant
ENSCAFG00000026284	snoRNA	downstream_gene_variant
ENSCAFG00000056330	lncRNA	downstream_gene_variant
ENSCAFG00000043782	snRNA	downstream_gene_variant
ENSCAFG00000057075	lncRNA	downstream_gene_variant
ENSCAFG00000028192	snoRNA	downstream_gene_variant
ENSCAFG00000026117	snRNA	downstream_gene_variant
ENSCAFG00000028313	snRNA	downstream_gene_variant
ENSCAFG00000058342	lncRNA	downstream_gene_variant
ENSCAFG00000053524	lncRNA	downstream_gene_variant
ENSCAFG00000047122	pseudogene	downstream_gene_variant
ENSCAFG00000056490	pseudogene	downstream_gene_variant
ENSCAFG00000058863	lncRNA	downstream_gene_variant
ENSCAFG00000057618	lncRNA	downstream_gene_variant
ENSCAFG00000042252	misc_RNA	downstream_gene_variant
ENSCAFG00000051434	lncRNA	downstream_gene_variant
ENSCAFG00000059098	lncRNA	downstream_gene_variant
ENSCAFG00000057930	lncRNA	downstream_gene_variant
ENSCAFG00000053565	lncRNA	downstream_gene_variant
ENSCAFG00000054596	lncRNA	downstream_gene_variant
ENSCAFG00000057389	lncRNA	downstream_gene_variant
ENSCAFG00000051752	lncRNA	downstream_gene_variant
ENSCAFG00000056979	lncRNA	downstream_gene_variant
ENSCAFG00000044497	lncRNA	downstream_gene_variant
ENSCAFG00000043087	snRNA	downstream_gene_variant
ENSCAFG00000050225	lncRNA	downstream_gene_variant
ENSCAFG00000053899	lncRNA	downstream_gene_variant
ENSCAFG00000058374	lncRNA	downstream_gene_variant
ENSCAFG00000050237	lncRNA	downstream_gene_variant
ENSCAFG00000058856	lncRNA	downstream_gene_variant
ENSCAFG00000053060	lncRNA	downstream_gene_variant
ENSCAFG00000052304	lncRNA	downstream_gene_variant
ENSCAFG00000053712	lncRNA	downstream_gene_variant
ENSCAFG00000053492	lncRNA	downstream_gene_variant
ENSCAFG00000059337	lncRNA	downstream_gene_variant
ENSCAFG00000055388	lncRNA	downstream_gene_variant
ENSCAFG00000053745	lncRNA	downstream_gene_variant
ENSCAFG00000056038	lncRNA	downstream_gene_variant
ENSCAFG00000052098	lncRNA	downstream_gene_variant
ENSCAFG00000021054	snRNA	downstream_gene_variant
ENSCAFG00000055895	lncRNA	downstream_gene_variant
ENSCAFG00000051389	lncRNA	downstream_gene_variant
ENSCAFG00000027465	snoRNA	downstream_gene_variant

ENSCAFG00000046259	lncRNA	downstream_gene_variant
ENSCAFG00000006771	pseudogene	downstream_gene_variant
ENSCAFG00000057251	lncRNA	downstream_gene_variant
ENSCAFG00000044330	pseudogene	downstream_gene_variant
ENSCAFG00000043307	pseudogene	downstream_gene_variant
ENSCAFG00000056650	lncRNA	downstream_gene_variant
ENSCAFG00000059035	lncRNA	downstream_gene_variant
ENSCAFG00000051226	lncRNA	downstream_gene_variant
ENSCAFG00000055100	lncRNA	downstream_gene_variant
ENSCAFG00000059214	lncRNA	downstream_gene_variant
ENSCAFG00000053351	lncRNA	downstream_gene_variant
ENSCAFG00000036065	pseudogene	downstream_gene_variant
ENSCAFG00000014984	processed_pseudogene	downstream_gene_variant
ENSCAFG00000051796	lncRNA	downstream_gene_variant

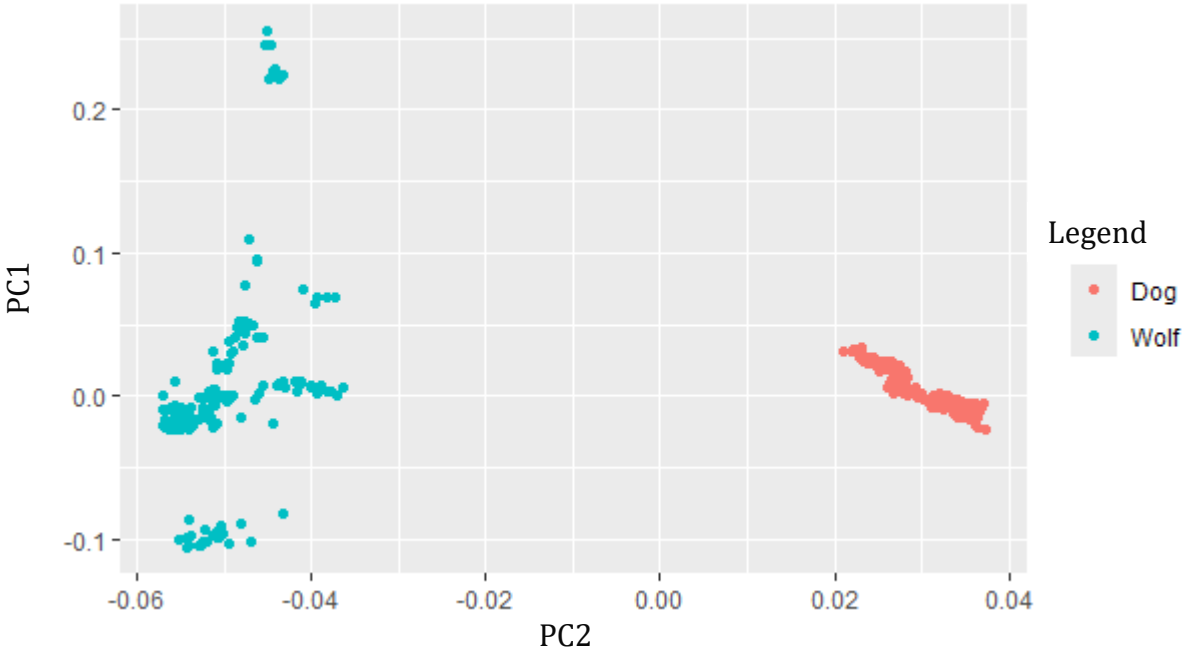
Supplementary table S5. Genes with UnitProt codes automatically excluded by g:Profiler in each run

UnitProt ID	*ENSMBL ID*	*Consequence*	*Genome*	*Grey wolves*	*FRDs*
AK9	ENSCAFG00000003835	downstream_gene_variant	D	N	Y
AKR1D1	ENSCAFG00000029647	downstream_gene_variant	D	Y	Y
ARNTL	ENSCAFG00000008095	downstream_gene_variant	H	N	Y
C12H6ORF141	ENSCAFG00000030374	downstream_gene_variant	H	N	Y
C1H19ORF18	ENSCAFG00000023343	downstream_gene_variant	H	Y	Y
C8H14ORF28	ENSCAFG00000013976	intron_variant	H	Y	N
C8H14ORF93	ENSCAFG00000011284	downstream_gene_variant	H	Y	Y
CDK5R2	ENSCAFG00000014996	downstream_gene_variant	D	N	Y
DLA-12	ENSCAFG00000032222	upstream_gene_variant	H	Y	Y
DLA88	ENSCAFG00000000487	upstream_gene_variant	H	Y	Y
KCNJ5	ENSCAFG00000010255	upstream_gene_variant	D	Y	Y
LOC100687327	ENSCAFG00000025063	upstream_gene_variant	D/H	Y	N
LOC102152102	ENSCAFG00000058042	upstream_gene_variant	H	Y	Y
LOC102152670	ENSCAFG00000058939	upstream_gene_variant	H	Y	Y
LOC106558870	ENSCAFG00000053824	upstream_gene_variant	D/H	Y	Y
LOC111093559	ENSCAFG00000053510	upstream_gene_variant	H	Y	Y
LOC111098992	ENSCAFG00000044740	downstream_gene_variant	H	Y	Y
LOC119863873	ENSCAFG00000019322	upstream_gene_variant	H	Y	N
LOC119863903	ENSCAFG00000029155	upstream_gene_variant	H	Y	Y
LOC119869843	ENSCAFG00000009288	downstream_gene_variant	H	Y	N
LOC119881676	ENSCAFG00000002706	downstream_gene_variant	H	Y	Y
LOC481603	ENSCAFG00000058088	missense_variant	H	Y	Y
LOC482174	ENSCAFG00000002805	upstream_gene_variant	H	N	Y
LOC482175	ENSCAFG00000002808	upstream_gene_variant	H	N	Y
LOC482176	ENSCAFG00000028617	downstream_gene_variant	H	N	Y
LOC482215	ENSCAFG00000048824	upstream_gene_variant	H	Y	Y
LOC482217	ENSCAFG00000055528	upstream_gene_variant	H	Y	Y
LOC483514	ENSCAFG00000008025	downstream_gene_variant	H	Y	Y
LOC483515	ENSCAFG00000056605	downstream_gene_variant	H	Y	Y
LOC490269	ENSCAFG00000046356	downstream_gene_variant	H	Y	N
LOC491385	ENSCAFG00000042771	downstream_gene_variant	D/H	N	Y
LOC609365	ENSCAFG00000025063	upstream_gene_variant	H	Y	N
LOC609690	ENSCAFG00000017369	downstream_gene_variant	H	Y	N
LOC610105	ENSCAFG00000002133	upstream_gene_variant	H	N	Y
LOC610875	ENSCAFG00000048426	downstream_gene_variant	H	N	Y
MYPOP	ENSCAFG00000023848	downstream_gene_variant	D	N	Y
OR13C11C	ENSCAFG00000056577	upstream_gene_variant	H	Y	Y
OR13C7F	ENSCAFG00000053326	upstream_gene_variant	H	Y	Y
OR13C7G	ENSCAFG00000029096	downstream_gene_variant	H	Y	Y
OR13C7H	ENSCAFG00000031867	downstream_gene_variant	H	Y	Y
OR13C7I	ENSCAFG00000032491	upstream_gene_variant	H	Y	Y
OR2AJ7	ENSCAFG00000041843	upstream_gene_variant	H	Y	Y
OR5M12	ENSCAFG00000008035	downstream_gene_variant	H	Y	Y
OR5M13	ENSCAFG00000052559	upstream_gene_variant	H	Y	Y

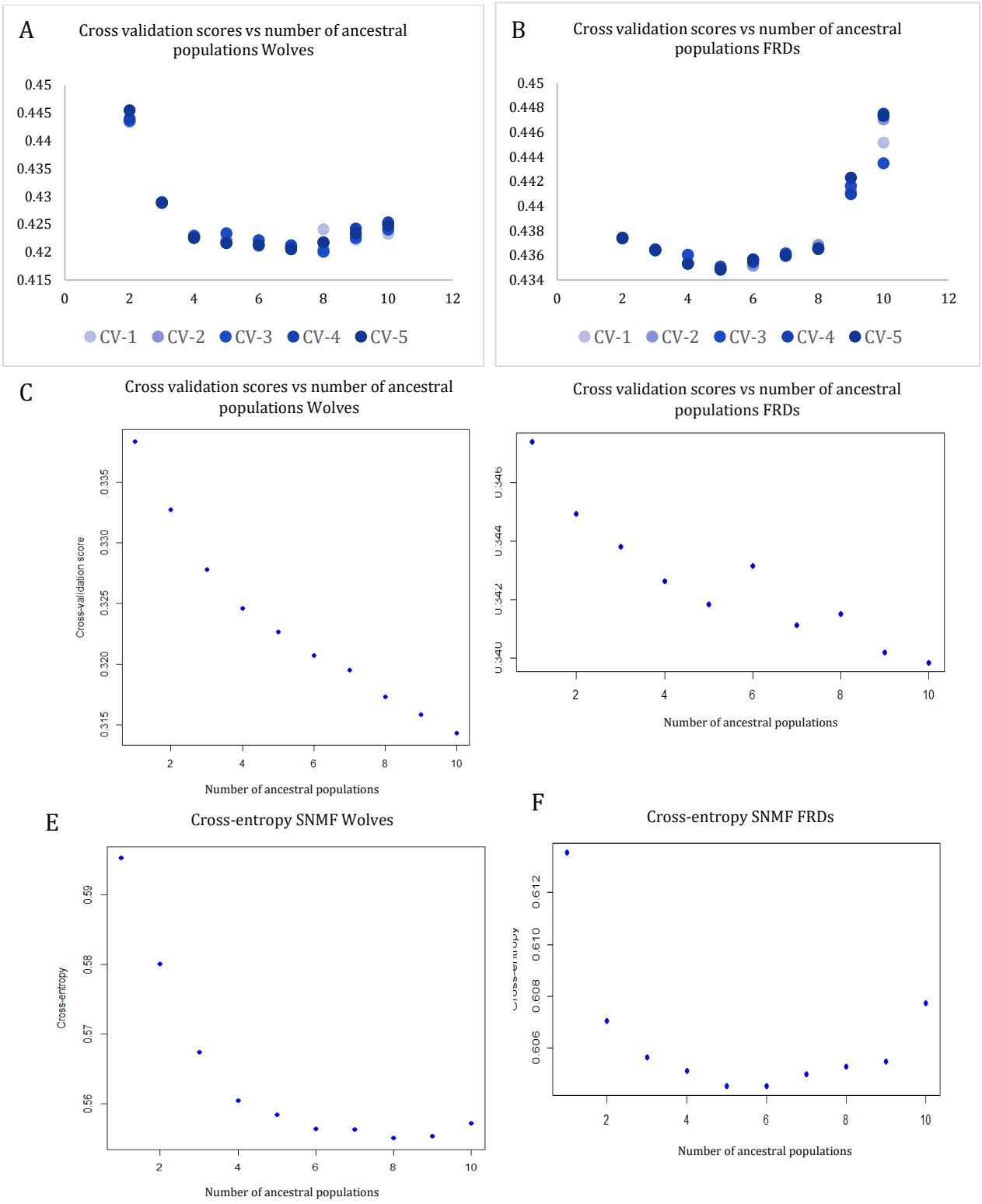
OR5M3C	ENSCAFG00000031836	upstream_gene_variant	H	Y	Y
OR5M5	ENSCAFG00000055301	upstream_gene_variant	H	Y	Y
OR6C38	ENSCAFG00000050743	downstream_gene_variant	H	N	Y
OR6C63	ENSCAFG00000049942	downstream_gene_variant	H	N	Y
OR7A23	ENSCAFG00000054648	downstream_gene_variant	H	Y	Y
OR7A55	ENSCAFG00000058847	downstream_gene_variant	H	Y	Y
OR7A57	ENSCAFG00000057285	upstream_gene_variant	H	Y	Y
OR7A58	ENSCAFG00000018627	upstream_gene_variant	H	Y	N
OR7A60	ENSCAFG00000041284	downstream_gene_variant	D/H	Y	Y
OR7A63	ENSCAFG00000050877	downstream_gene_variant	H	Y	Y
OR7A64	ENSCAFG00000051510	downstream_gene_variant	H	Y	Y
TNSF1	ENSCAFG00000041401	downstream_gene_variant	H	Y	N
ZBTB24	ENSCAFG00000003828	downstream_gene_variant	D	Y	N

D= Dog Genome Run, H= Human Genome Run, Y= Yes, N= No

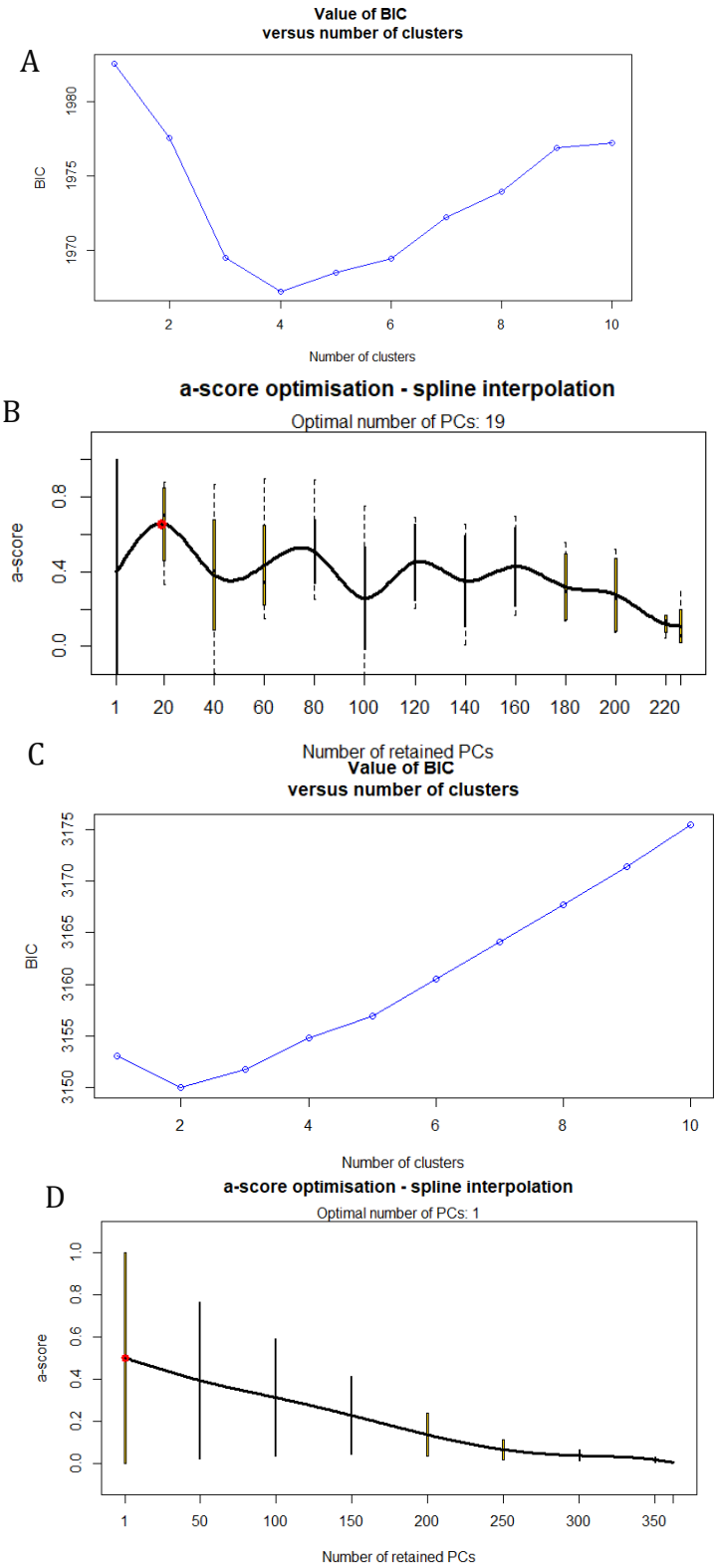
Supplementary figures



Supplementary figure 1. Principal component plots for addressing any potential taxonomic misidentification. Two distinct clusters were detected, one for wolves and another for FRDs. The eigen values for the first 10 principal components were: 39.65, 11.65, 10.28, 7.58, 6.04, 4.31, 4.25, 3.89, 3.56 and 3.36.

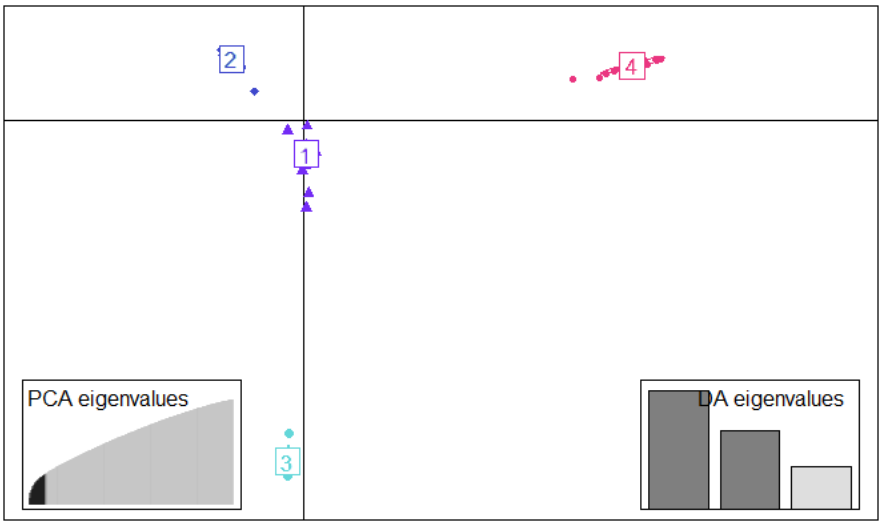


Supplementary figure 2. Cross validation and cross-entropy plots for population structure approaches. A-B. Cross validation (CV) values for all reiterations of population structure analysis in ADMIXTURE for wolves and FRDs. C-D. CV values for all reiterations of population structure analysis in Tess3r. E-F. Cross-entropy values for al reiterations of population structure analysis with sNMF. The optimal K identified was K=8 for grey wolves (A, E) and K=5 for FRDs (B, D, F). No conclusive K value was selected for wolves based on the Tess3r plot.

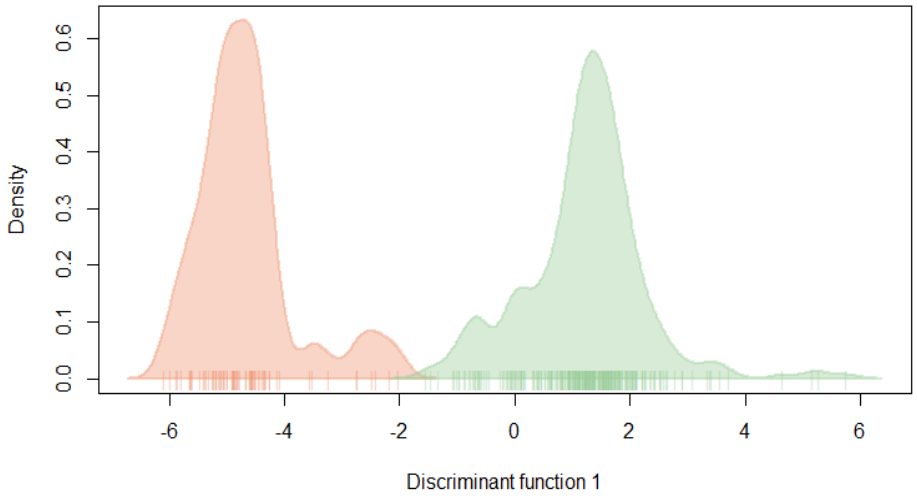


Supplementary figure 3. Bayesian index criterion (BIC) for wolves and FRDs and a-score optimization for the Discriminant Analysis of Principal Components (DAPC). A: BIC plot for wolves, where K=4 was selected. B: A-score optimization plot for wolves, where 19 principal components were retained. C: A-score optimization plot for FRDs, where 1 principal component was retained. D: BIC plot for FRDs K=2 was selected.

A. DAPC density plot grey wolves

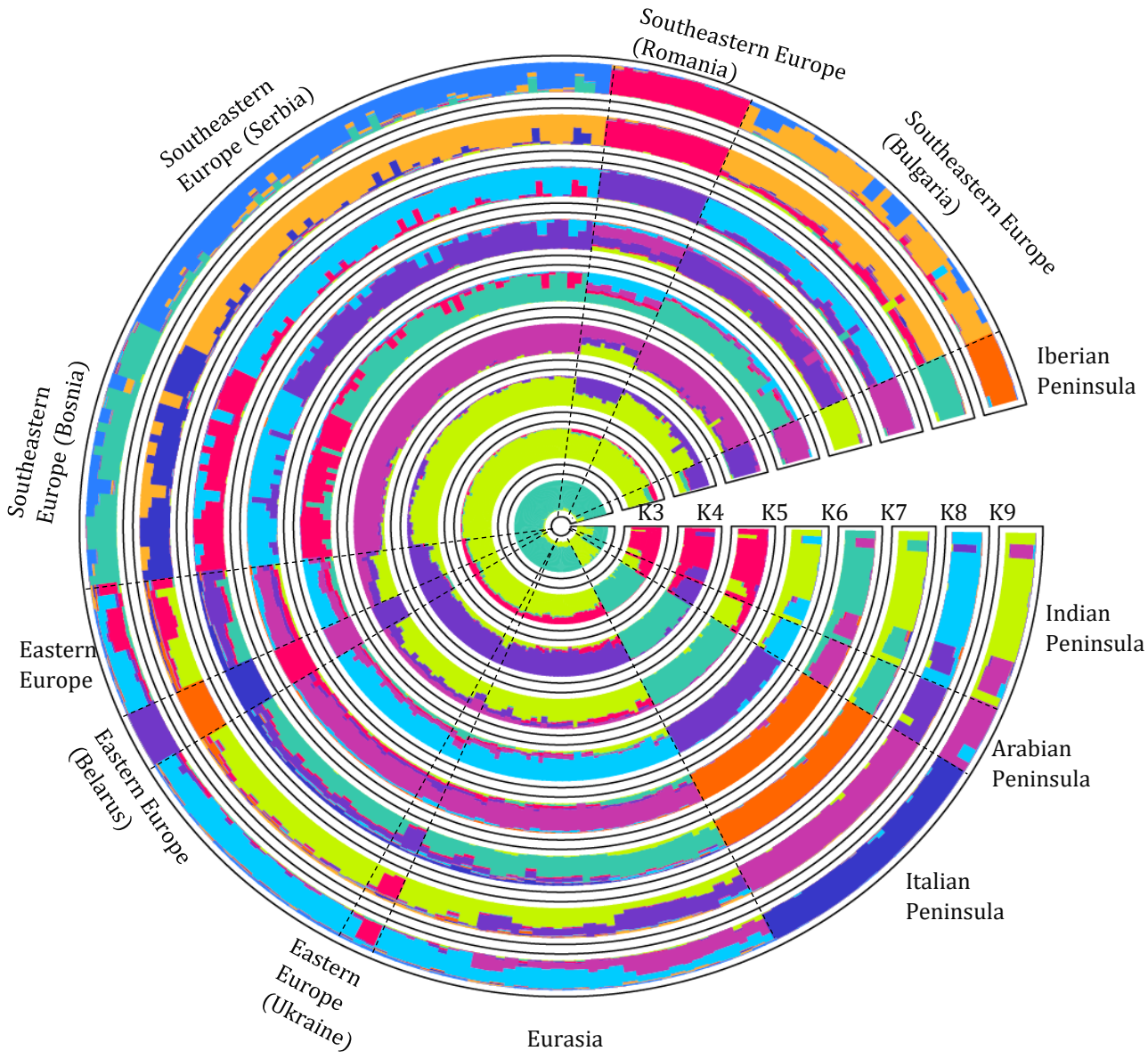


B. DAPC density plot FRDs

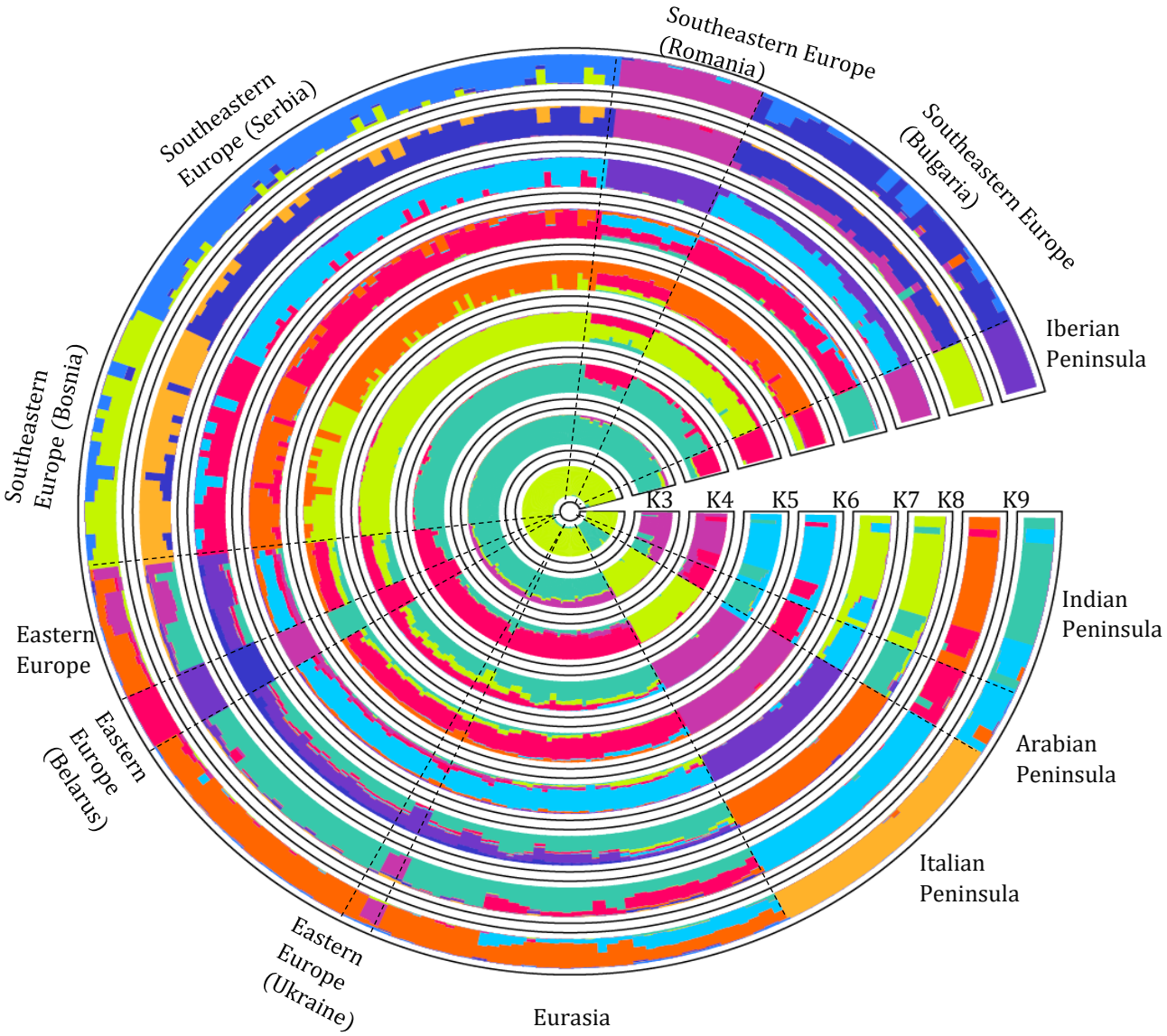


Supplementary figure 4. Discriminant Analysis of Principal Components (DAPC) density plots. A: Wolves' DAPC density plot with two axes (x=axis 1, y=axis 2) based on the BIC (see above), the optimal was K=4. Wolves from the Italian Peninsula (Cluster 4) were from other clusters by both axes. Wolves from the Indian Peninsula (Cluster 3) where the second most distant cluster was separated also by both axes from second cluster and one axis from the first. Most of the wolves from Southeastern Europe (i.e., Bosnia, Serbia and Bulgaria) were assigned to Cluster 2, the remaining Eurasian individuals were assigned to Cluster 1. B. FRDs' DAPC density plot showing two clusters, based on the BIC (see above), the optimal K=2. FRDs from the Indian and Arabian Peninsulas belong to the first genetic cluster (orange) the remaining individuals belong to the second cluster (green). Only 1 principal component was retained for FRDs thus, the plot only has 2 axes for this reason it looks different than that of the wolves.

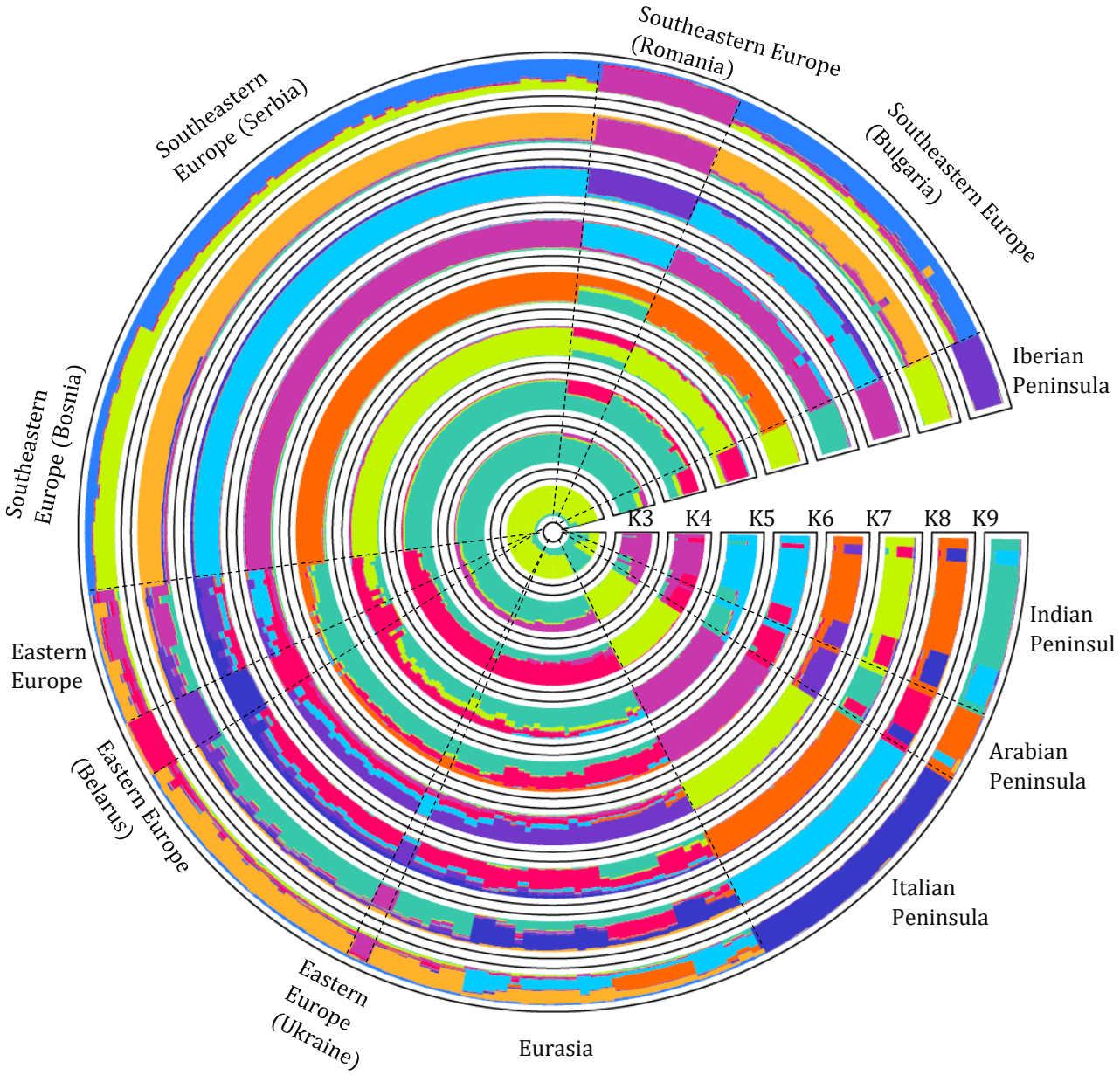
A. Wolves sNMF ancestry histograms K1:K10



B. Wolves Admixture ancestry histograms K1:K10

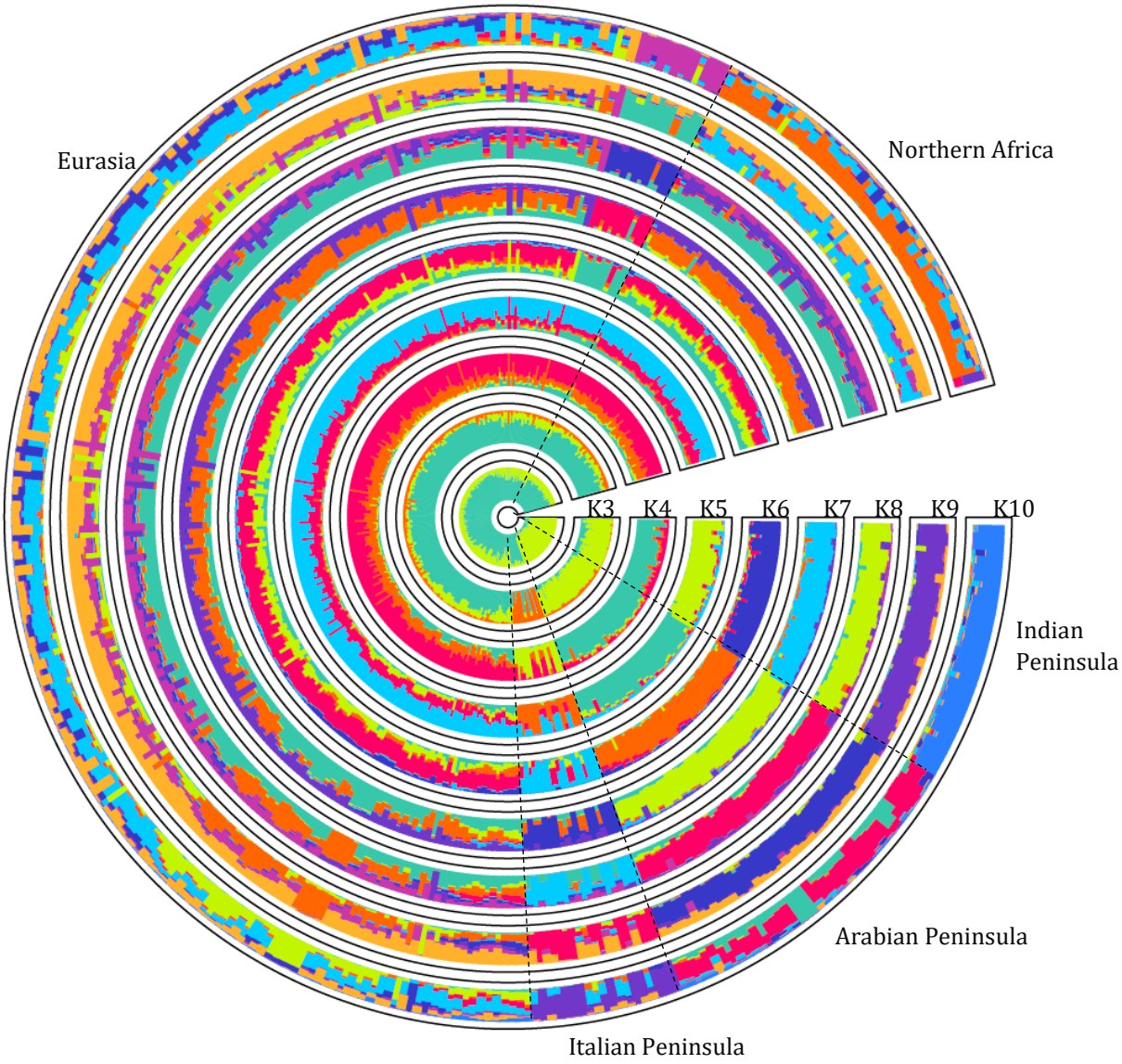


C. Wolves Tess3R ancestry histograms K1:K10

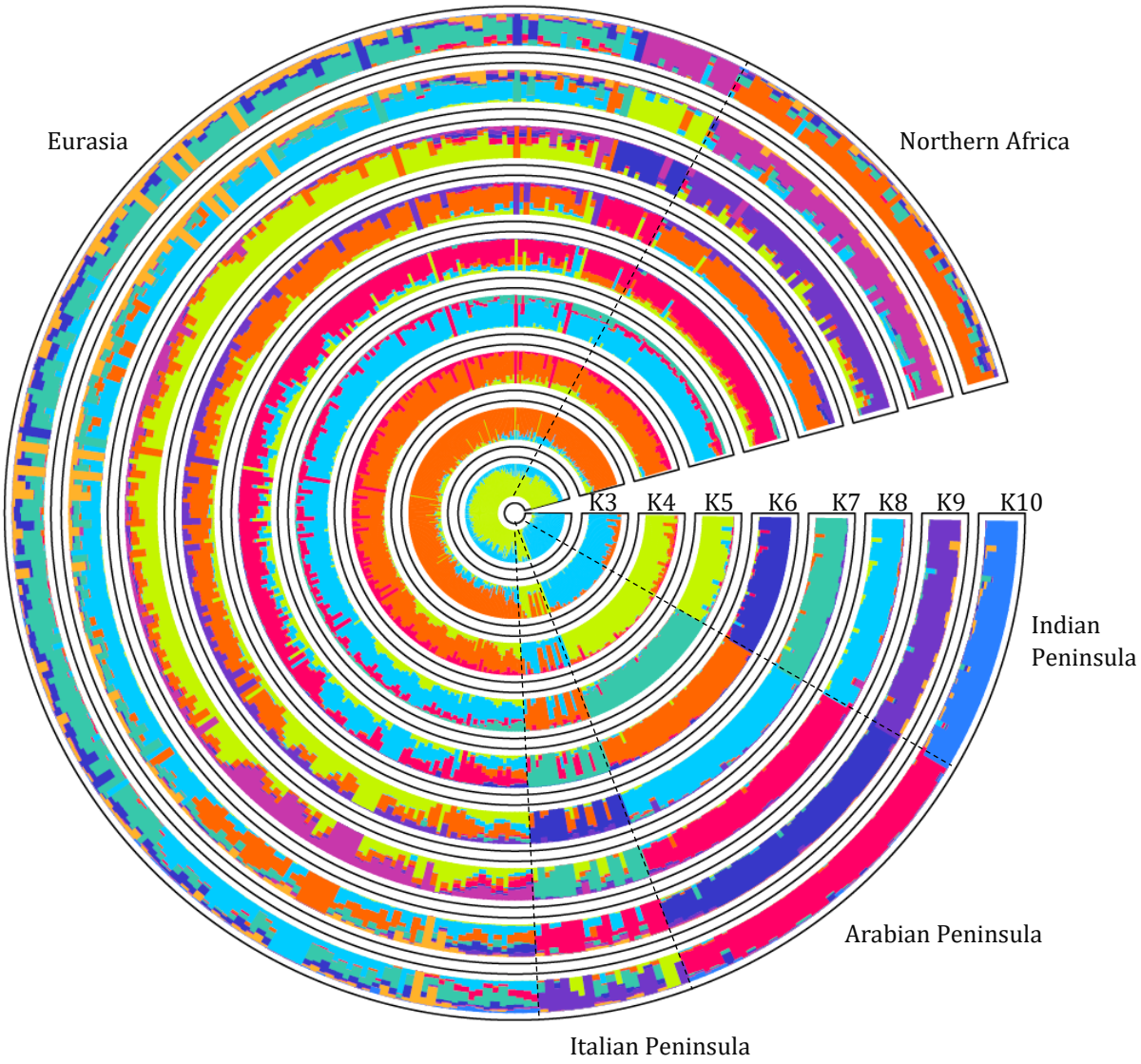


Supplementary figure 5. Population structure histograms wolves. A. Population structure histograms Wolves from K=2 to K-10, based on the cluster prediction coefficients from sNMF. Based on cross validation (CV) values the K=8 was the optimal. B. Population structure histograms Wolves from K=2 to K-10, based on the cluster prediction coefficients from Admixture. Based on CV values, K=8 was the optimal. C. Population structure histograms Wolves from K=2 to K-10, based on the cluster prediction coefficients from Tess3r. Broken lines highlight the clusters based on K=8, and in parentheses are the countries from which clusters originate.

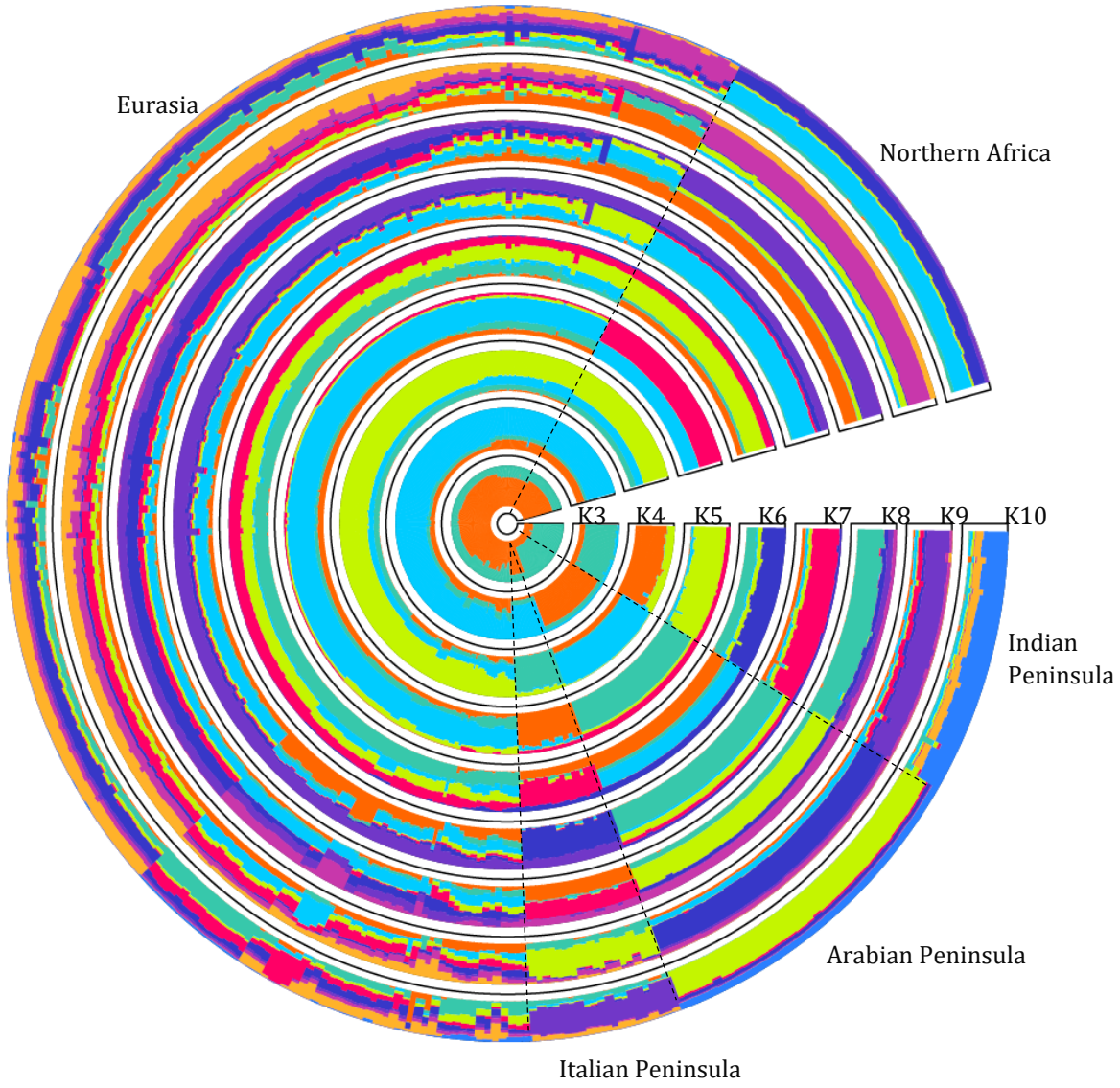
A. FRDs. sNMF ancestry histograms K1:K10



B. FRDs Admixture ancestry histograms K1:K10

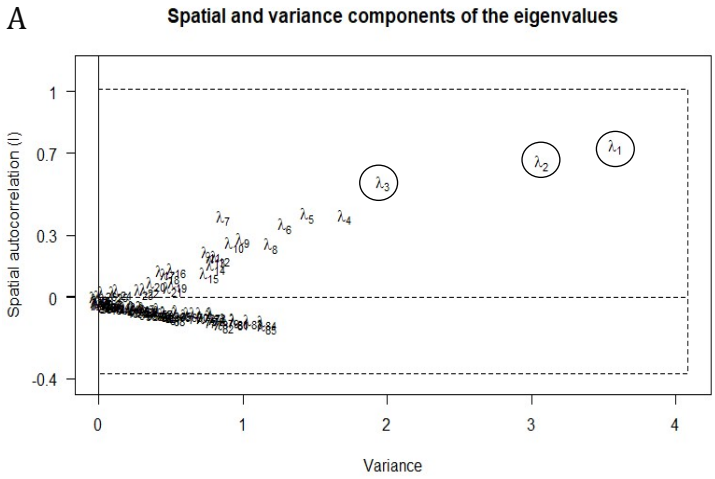


C. FRDs Tess3r ancestry histograms K1:K10

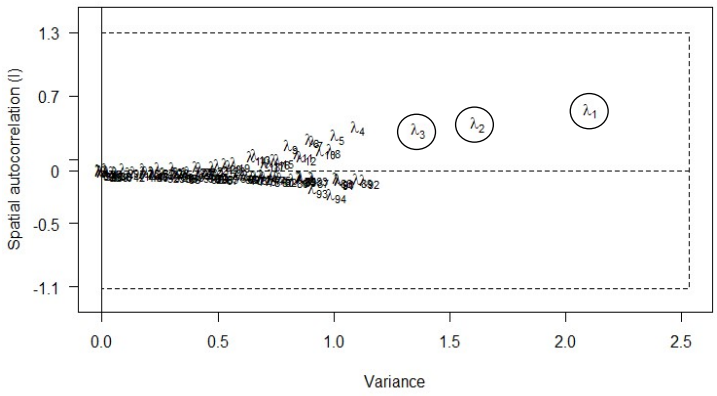


Supplementary figure 6. Population structure histograms FRDs. A. Population structure histograms FRDs from K=2 to K-10, based on the cluster prediction coefficients from sNMF. Based on cross validation (CV) values, K=5 was the optimal. B. Population structure histograms FRDs from K=2 to K-10, based on the cluster prediction coefficients from ADMIXTURE. C. Population structure histograms FRDs from K=2 to K-10, based on the cluster prediction coefficients from Tess3r. Based on CV values the K=5 was the optimal. Broken lines highlight the clusters based on K=5, and in parentheses are the countries from which these clusters originate.

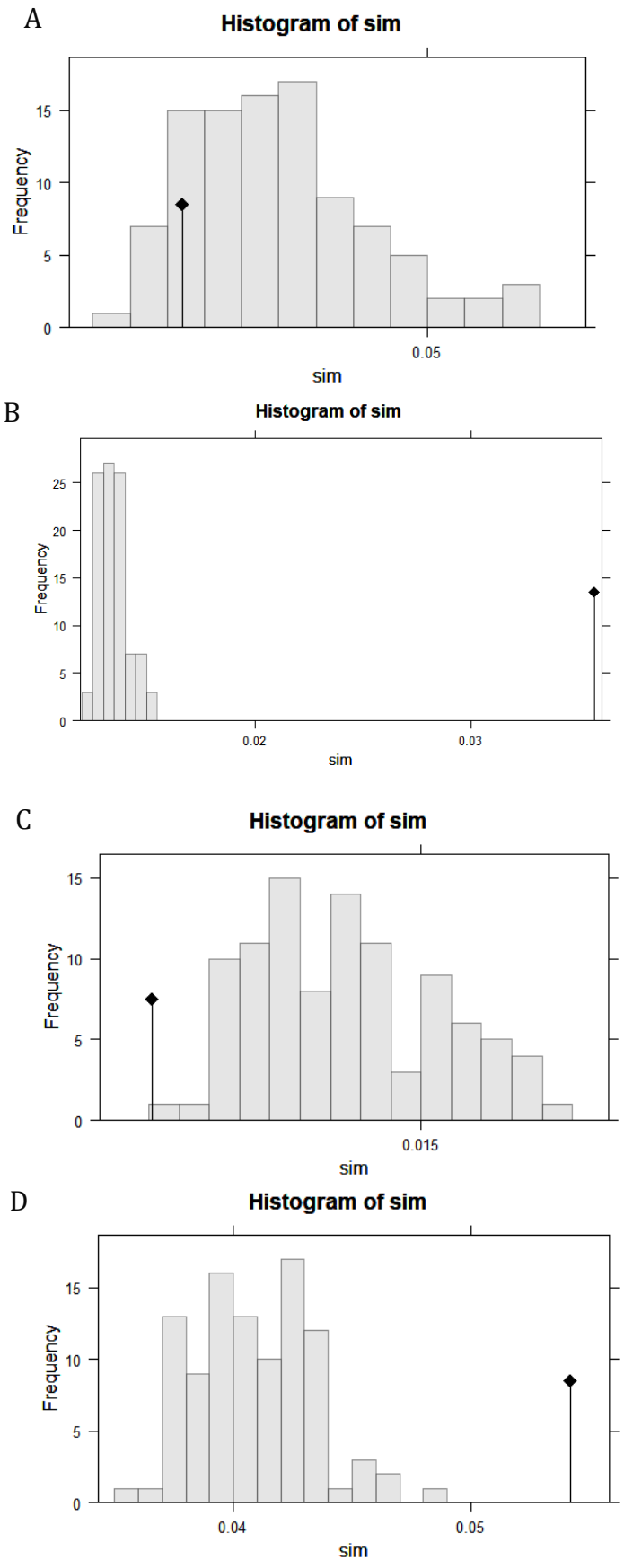
A



B



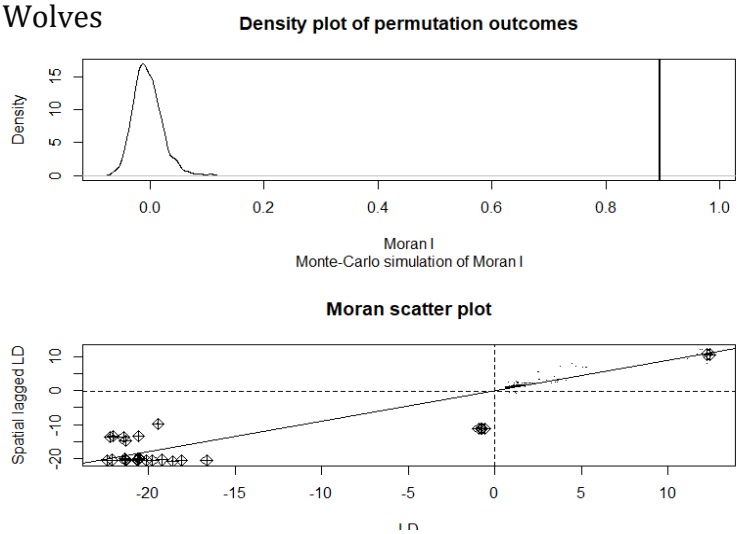
Supplementary figure 7. Spatial principal component analysis (sPCA) plots. A. Eigen values for spatial and variance components wolves. B. Eigen values for spatial and variance components FRDs. The values of the sPCA are composites of both the spatial autocorrelation measurements of Moran's I statistics and genetic variation; thus, they can be used for selection of the principal component to interpret. The largest eigen values were λ1 and λ2 consequently the first and second axes (i.e., global structures) were interpreted. Since sPCA values are a composite which measures both genetic variance and the spatial autocorrelation measured by Moran's I, decomposition of these values can be used to select which principal component (PC) to interpret.



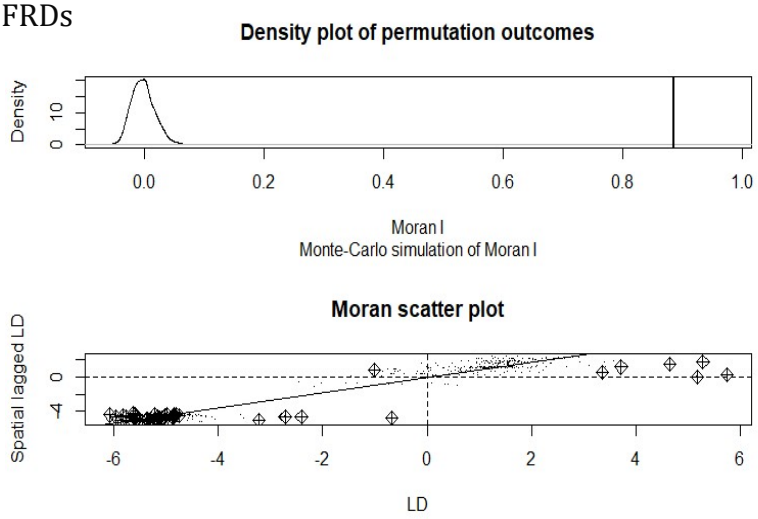
Supplementary figure 8. Histograms of Monte Carlo simulations for global and local structures. A: Monte Carlo simulation for global structure (Gtest) Wolves, calculated

statistics were, p-value: 0.01, std. obs.: 5.35e+00, expectation: 4.09e-02 and variance: 6.21e-06. B: Monte Carlo simulation for local structure (Ltest) Wolves, calculated statistics was, p-value: 0.9, std. obs.: -1.10e+00, expectation: 4.60e-02 and variance: 5.71e-06. C: Monte Carlo simulation for global structure (Gtest) FRDs, simulated p-value: 0.01, std. obs.: 3.36e+01, expectation: 1.35e-02 and variance: 4.37e-07. D: Monte Carlo simulation for local structure (Ltest) FRDs, simulated p-value: 1, std. obs.: -2.13e+00, expectation: 1.45e-02 and variance: 3.619e-07.

A. Wolves

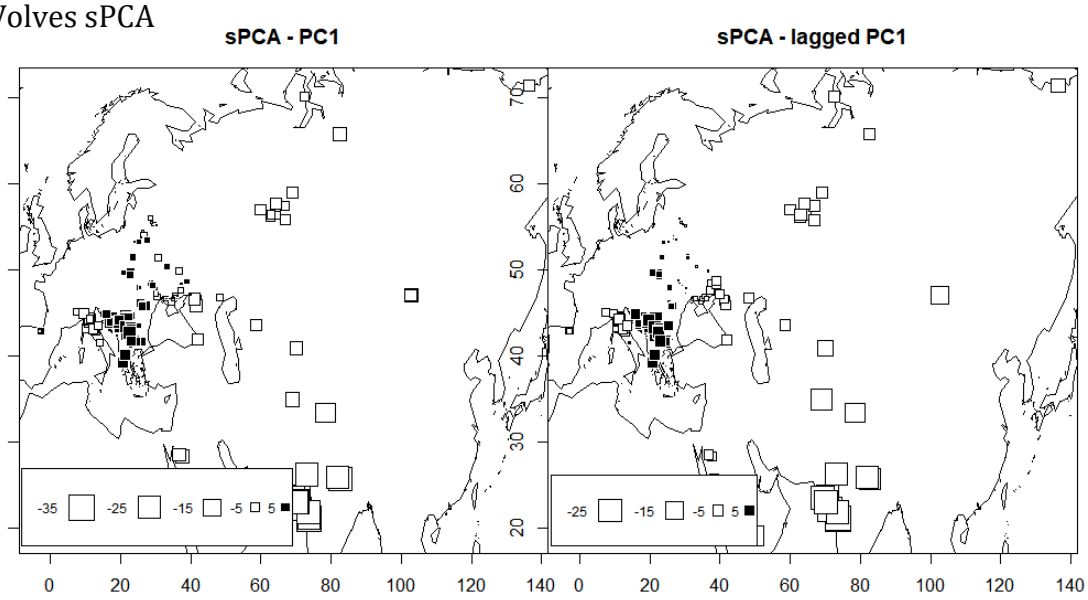


B. FRDs



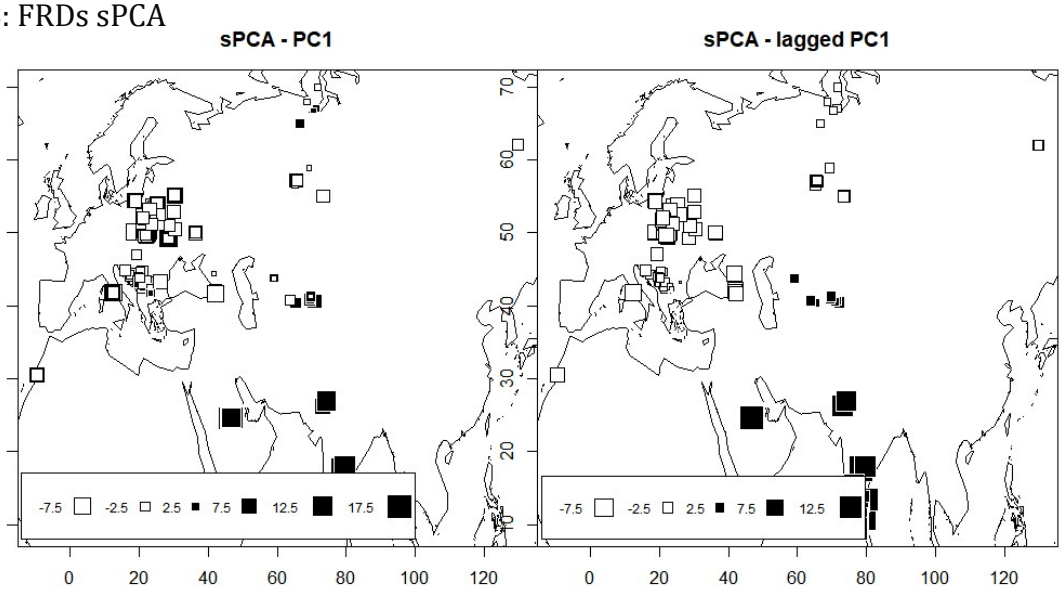
Supplementary Figure 9. Density plots and Moran I scatter plots. A. Density plot of permutation outcomes and Moran I scatter plot wolves. B. Density plot of permutation outcomes and Moran I scatter plot FRDs. Positive autocorrelation was detected in both wolves and FRDs. The calculated Moran I statistics for wolves were z-score=0.90, p=0.001, and Monte Carlo test: std.obs =-7.03e-0 expectation=9.90e-03, variance=2.41e-06). Moran I statistics for FRDs were z-score= 0.89 p=0.001 and Monte Carlo test: std.obs =-1.23e+01, expectation=5.81e-03, variance=5.07e-07.

A: Wolves sPCA



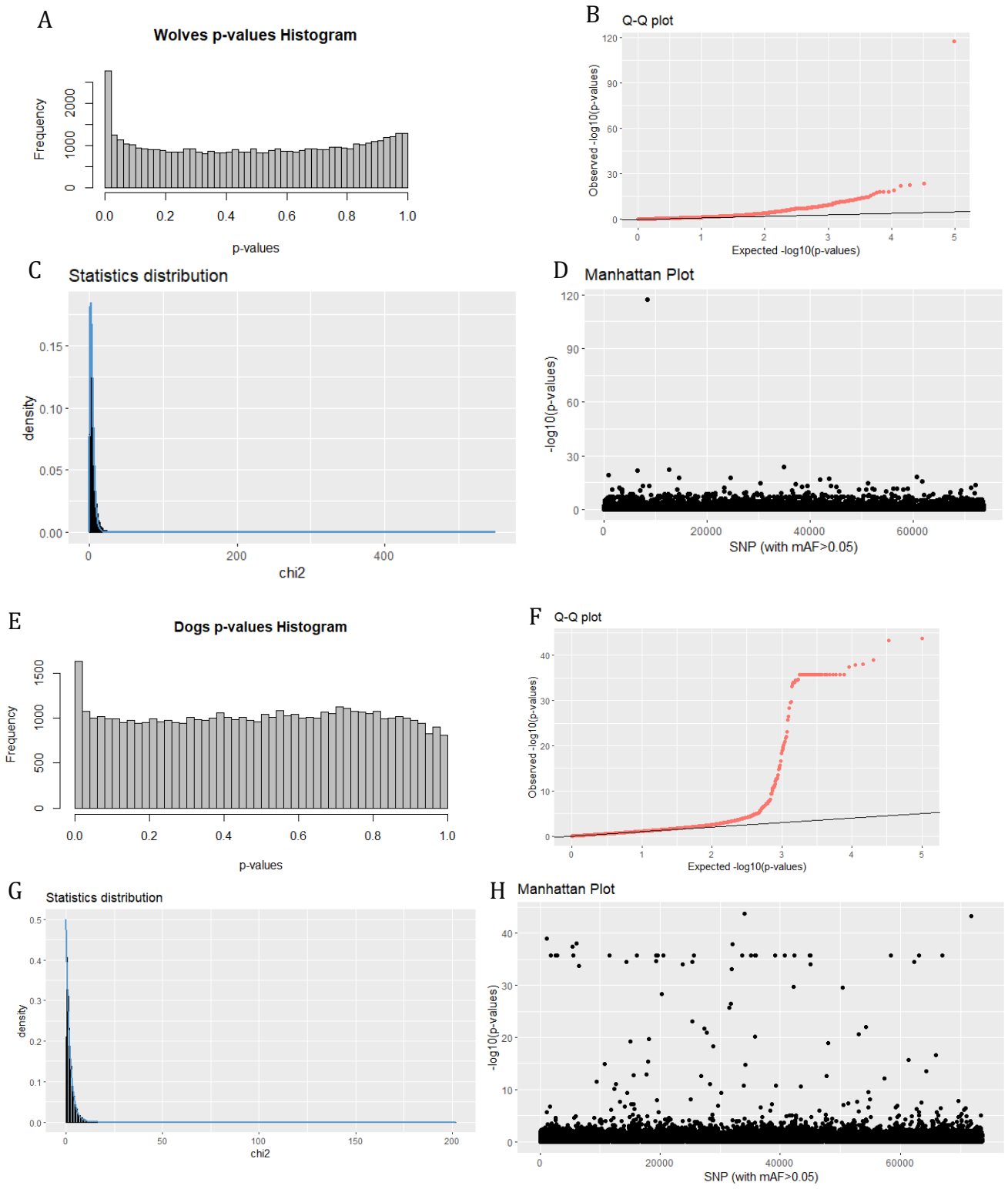
	Axis 1	Axis 243
PCA Axis1	0.983	0.0118
PCA Axis2	-0.073	-0.004

B: FRDs sPCA



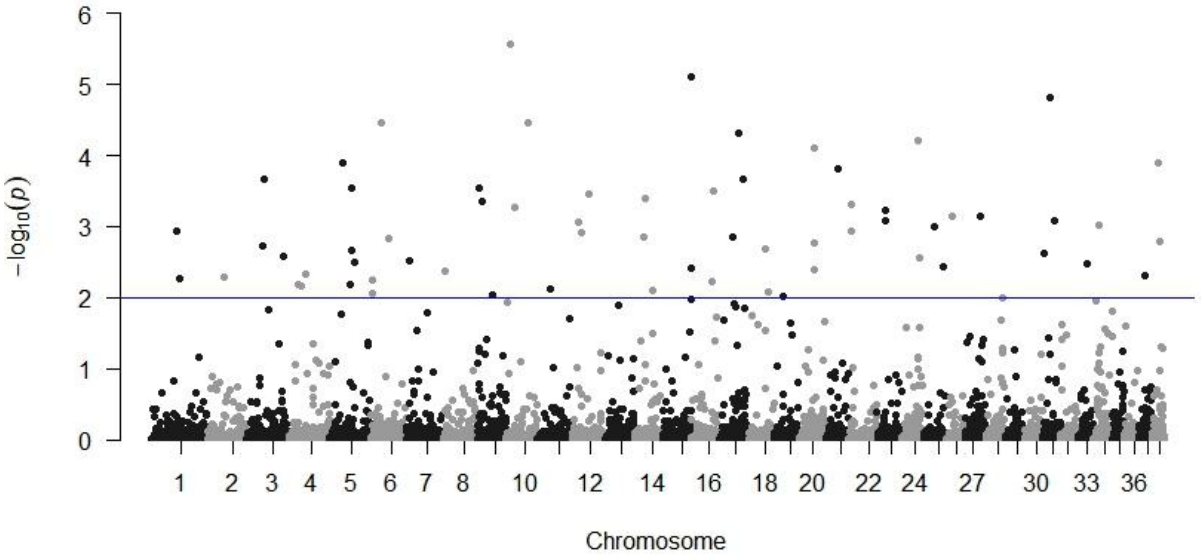
	Axis 1	Axis 299
PCA Axis1	-0.985	0.011
PCA Axis2	0.005	-0.003

Supplementary figure 10. Map of the spatial principal component analysis (sPCA). Map of the first principal component for wolves (A-left) and FRDs (B-left) and lagged reduced data noise for grey wolves (A-right) and for FRDs (B-right). White squares represent positive values while black squares represent negative values, size is proportional to the absolute value of the coordinate. Genetic variation is suggested to be spatially structured for both grey wolves and FRDs.

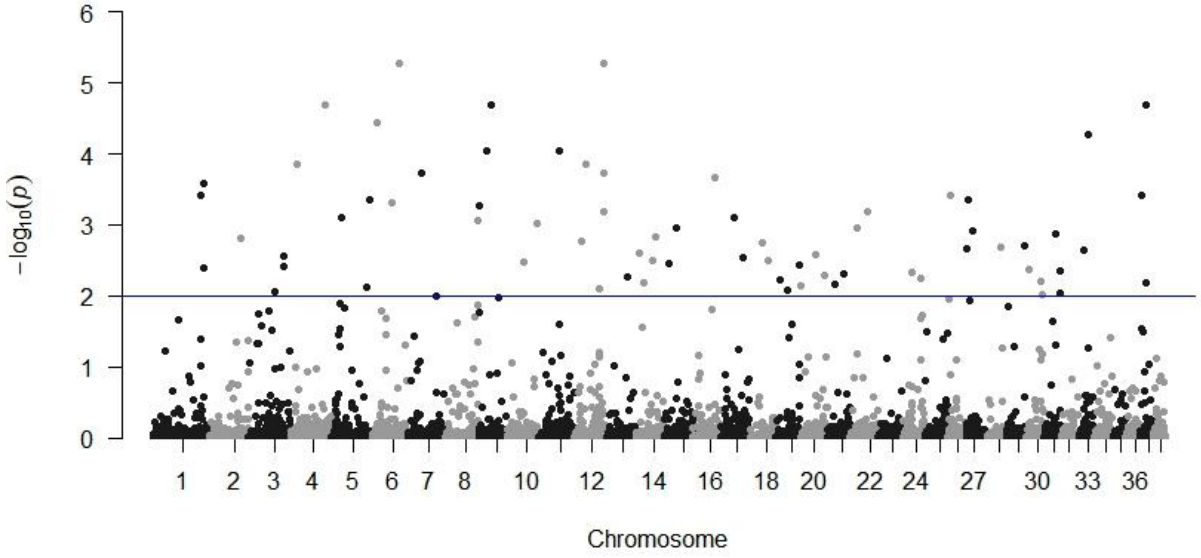


Supplementary figure 11. PCAdapt results. P-value histograms for grey wolves (A) and FRDs (E), Q-Q plots for wolves (B) and FRDs (F), statistics distribution for wolves (C) and FRDs (G) and Manhattan plots for candidate SNPs identified to be under diversifying selection with this approach for wolves (D) and FRDs (H).

A: Wolves Bayesacan Manhattan plot

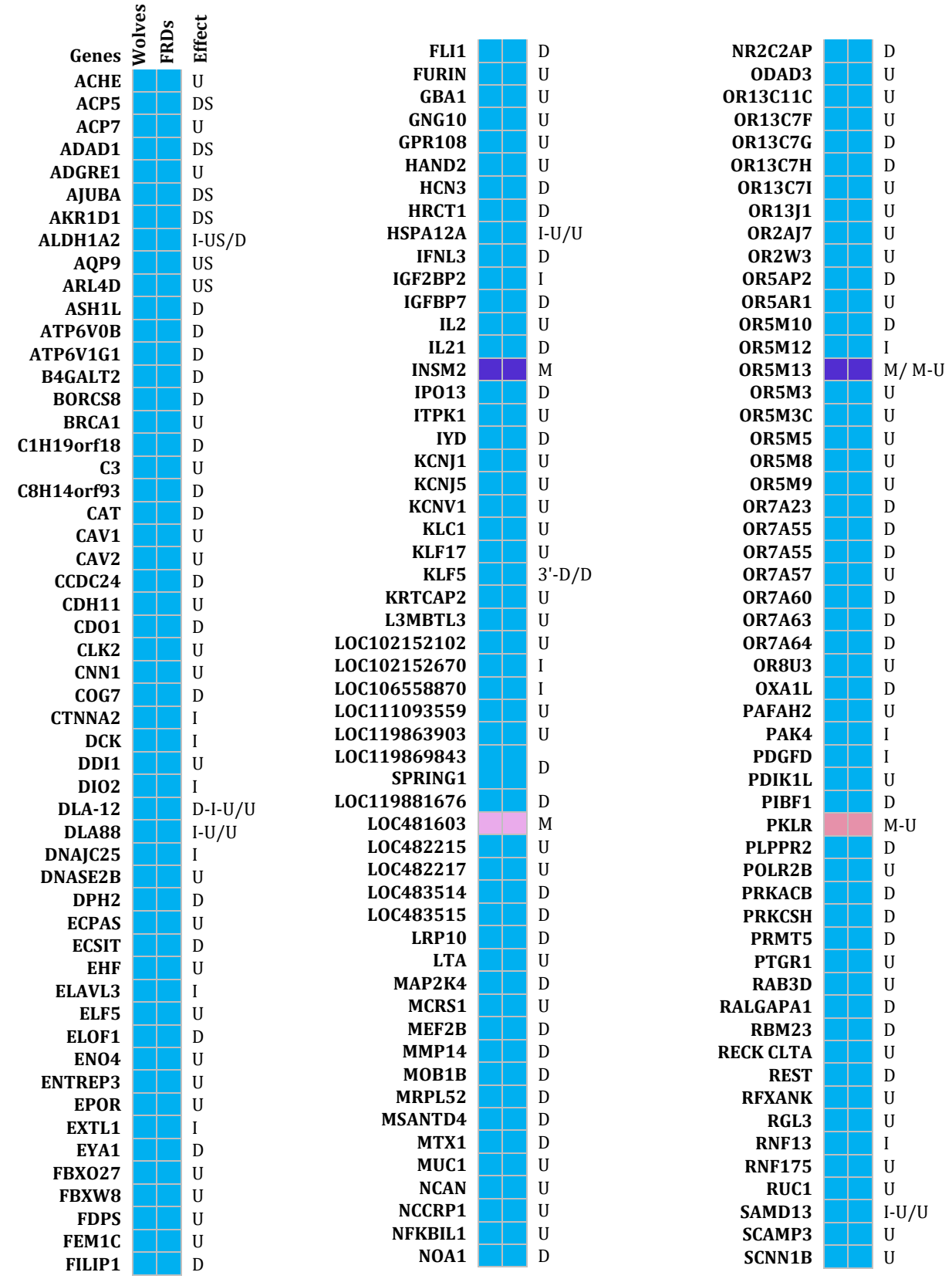


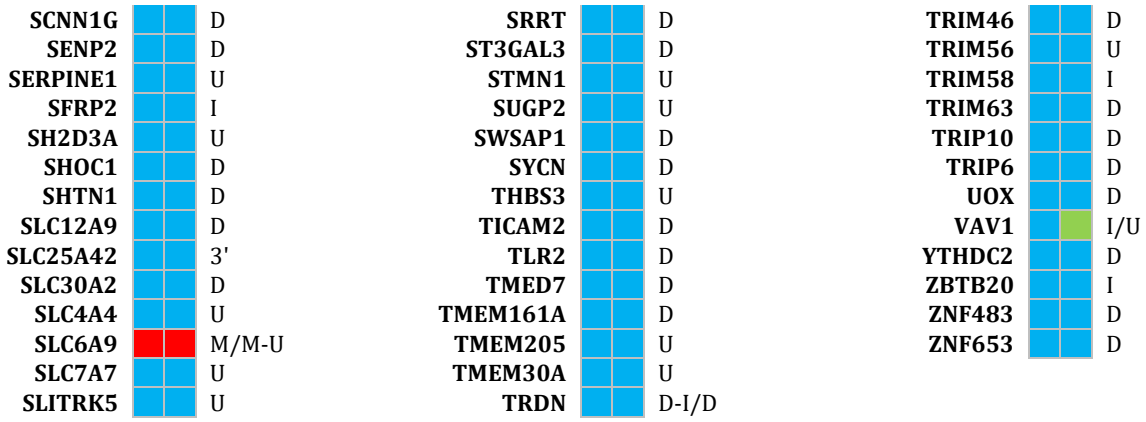
B: FRDs Bayesacan Manhattan plot



Supplementary Figure 12. Manhattan plots of Bayescan results. Manhattan plots for candidate SNPs under diversifying selection (A) wolves and (B) FRDs. Calculated q-values were transformed into $\log_{10}$, outliers under diversifying selection were candidates with positive alpha values and q-values less than 0.05.

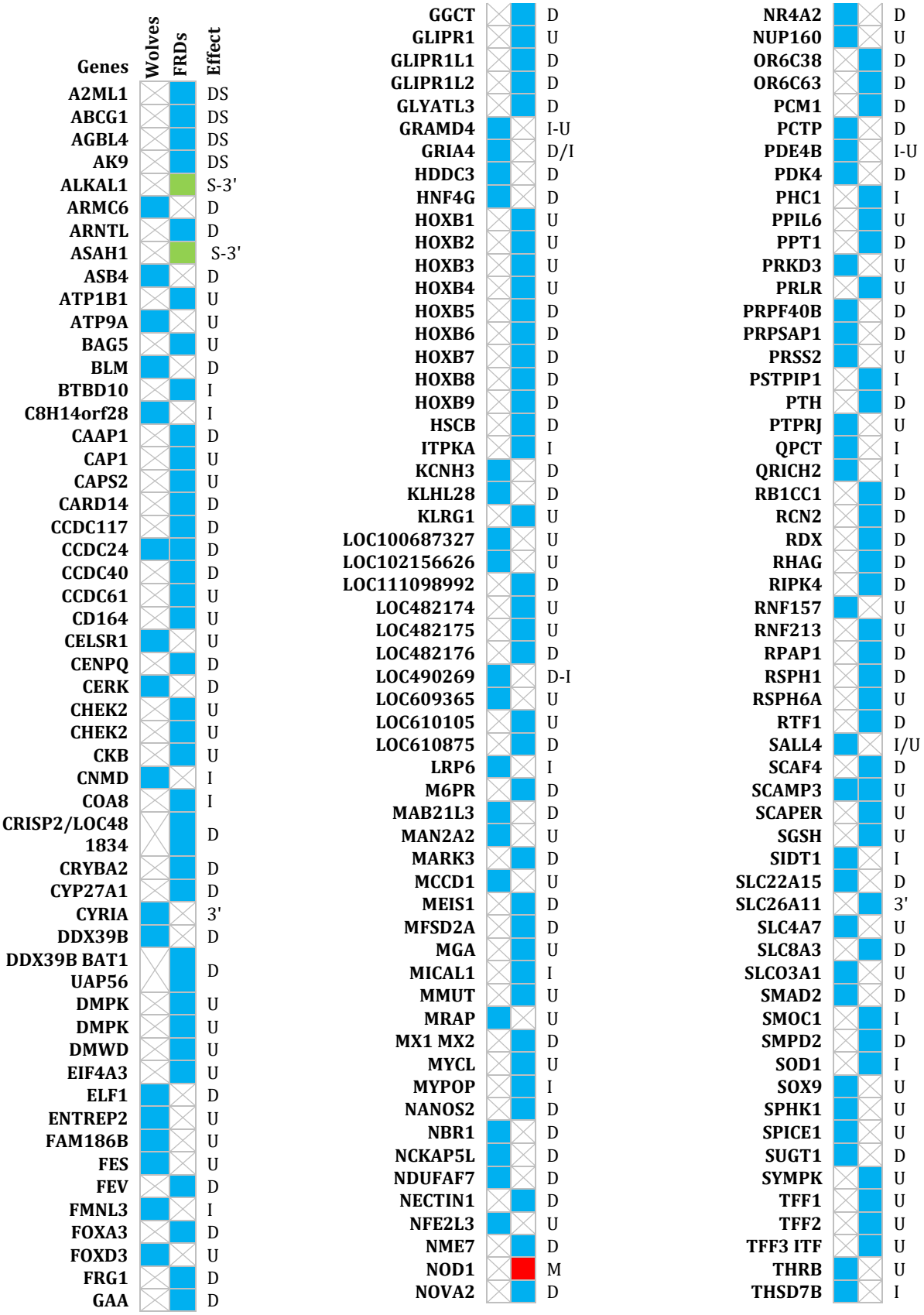
A. VEP results summary overlapping genes among wolves and FRDs

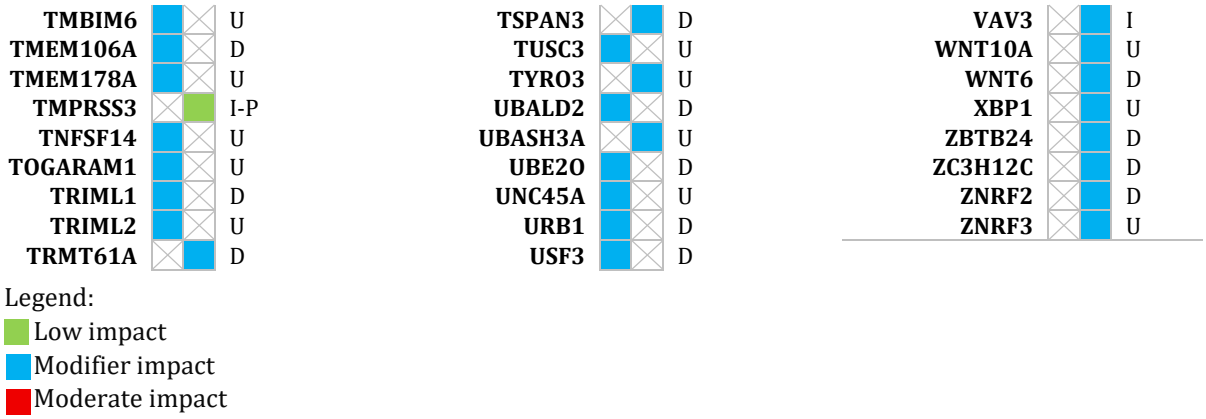




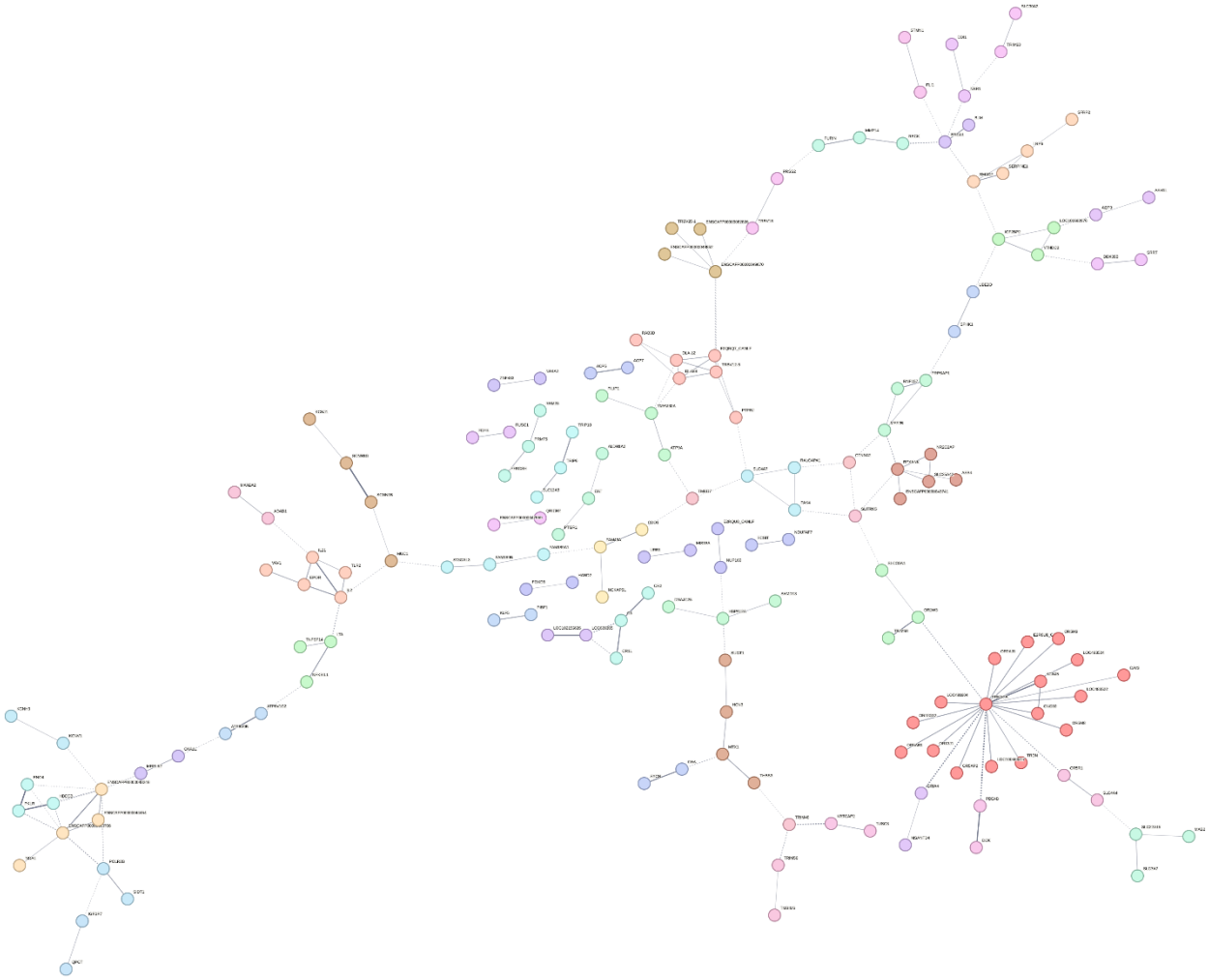
Legend:
Low impact
Modifier impact
Moderate impact

B. VEP results summary non-overlapping genes among wolves and FRDs

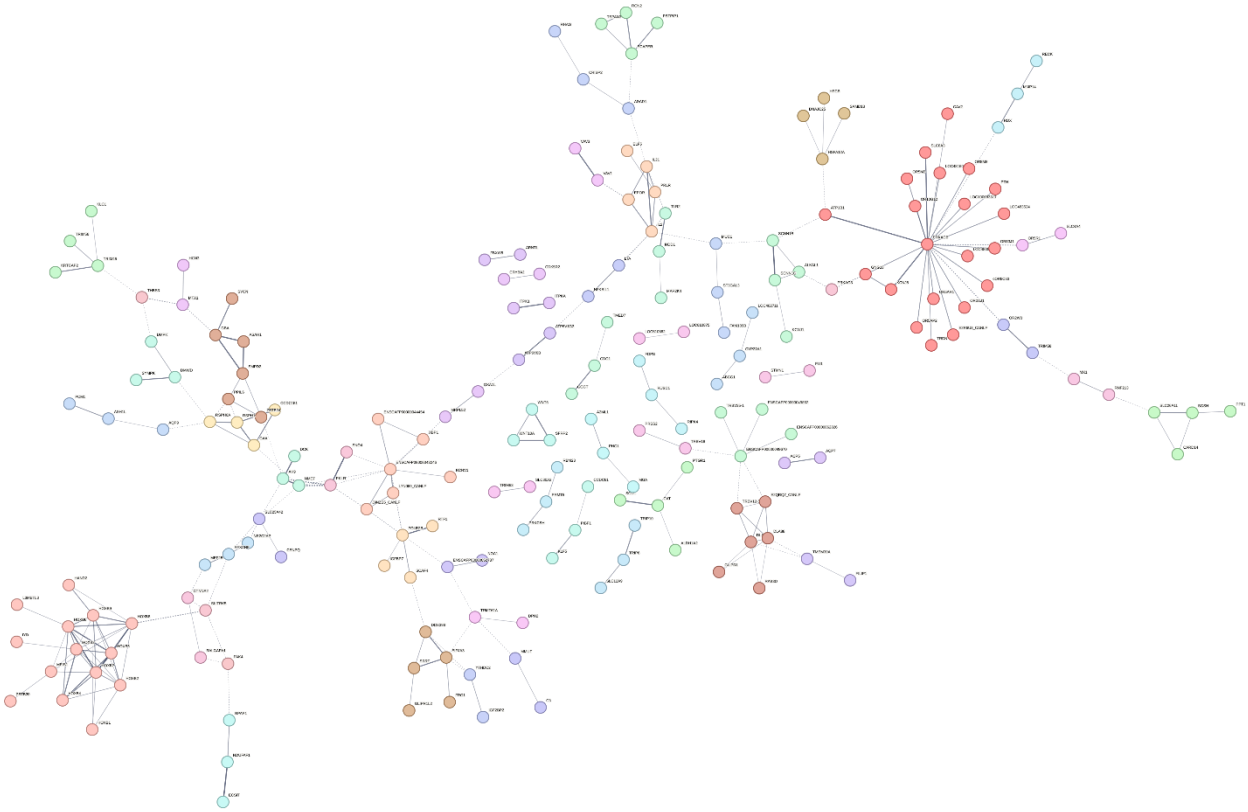



TMBIM6 U
TMEM106A D
TMEM178A U
TMPRSS3 I-P
TNFSF14 U
TOGARAM1 U
TRIML1 D
TRIML2 U
TRMT61A D
TSPAN3 D
TUSC3 U
TYRO3 U
UBALD2 D
UBASH3A U
UBE2O D
UNC45A U
URB1 D
USF3 D
VAV3 I
WNT10A U
WNT6 D
XBP1 U
ZBTB24 D
ZC3H12C D
ZNRF2 D
ZNRF3 U
Legend:
Low impact
Modifier impact
Moderate impact

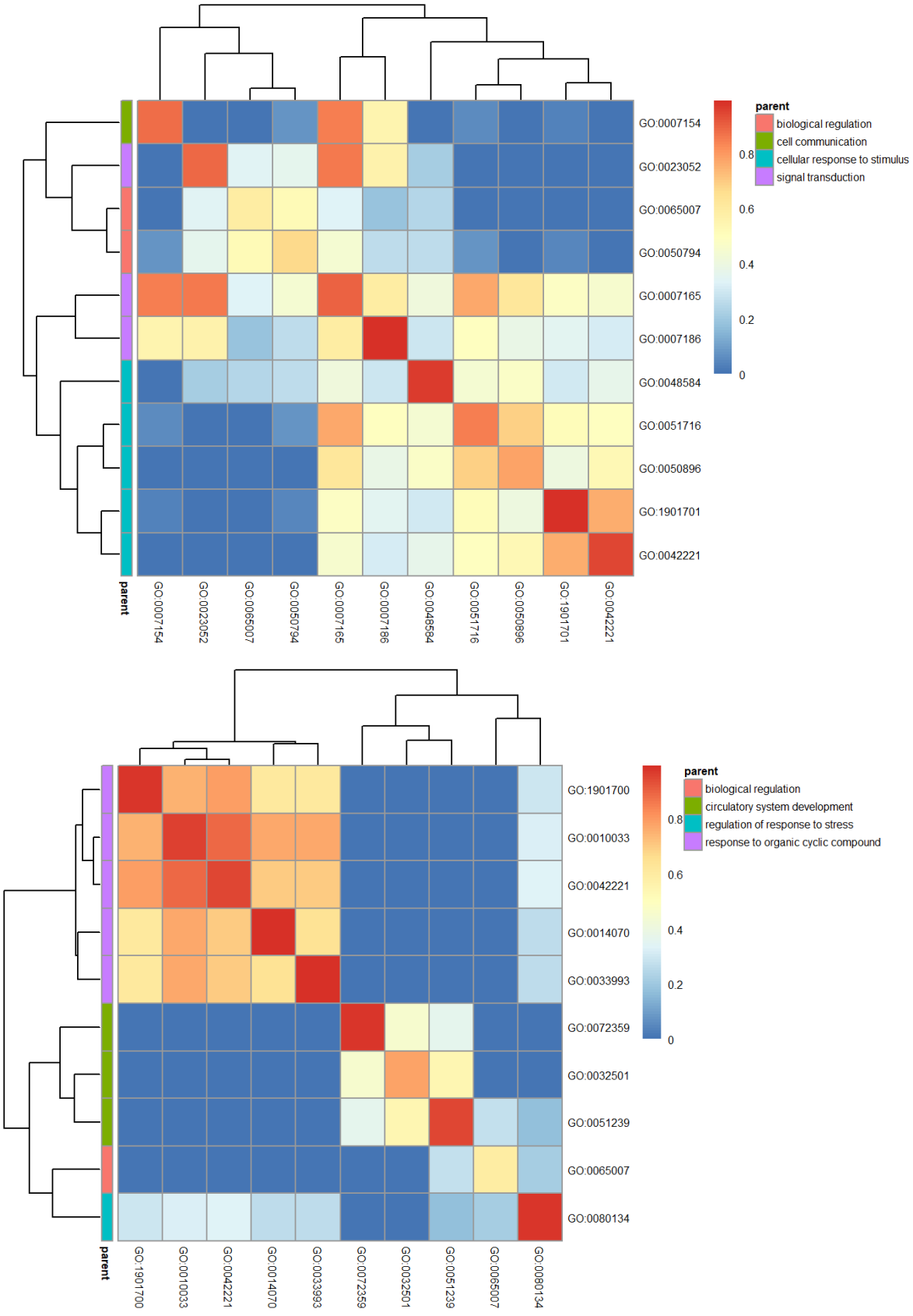
Supplementary figure 13. Variant effect predictor (VEP) results summary. Overlapping (A) and non-overlapping (B) coding genes displaying signatures of selection among wolves and FRDs. A total of 393 genes displaying signatures of selection were detected in both wolves and FRDs. The colours indicate the predicted impact, green for low and blue for modifier, all other colours were moderate impact. Crossed out cells note genes which did not have variants under selection in that species. Consequences are abbreviated as follows: D= Downstream, U= Upstream, I= Intron, M= Missense, 3'= 3'UTR, S= Synonymous and P= Splice Donor. Missense mutations were coloured in shades of red, the intensity of the colour signifies whether it was predicted to be tolerated or not (i.e., predicted to impact function); the more intense the red the less tolerated the mutation. The gene coloured purple was not suited for the SWIF analysis to predict the putative impact of the variant on protein function. Genes with more than one predicted effect/consequence separated with (-), these genes had more than one isoform and the variants were predicted to have different effects in each isoform or had more than one variant under selection. Genes with different predicted consequences in FRDs and wolves also had more than one consequence, and it was divided by forward slash (/).



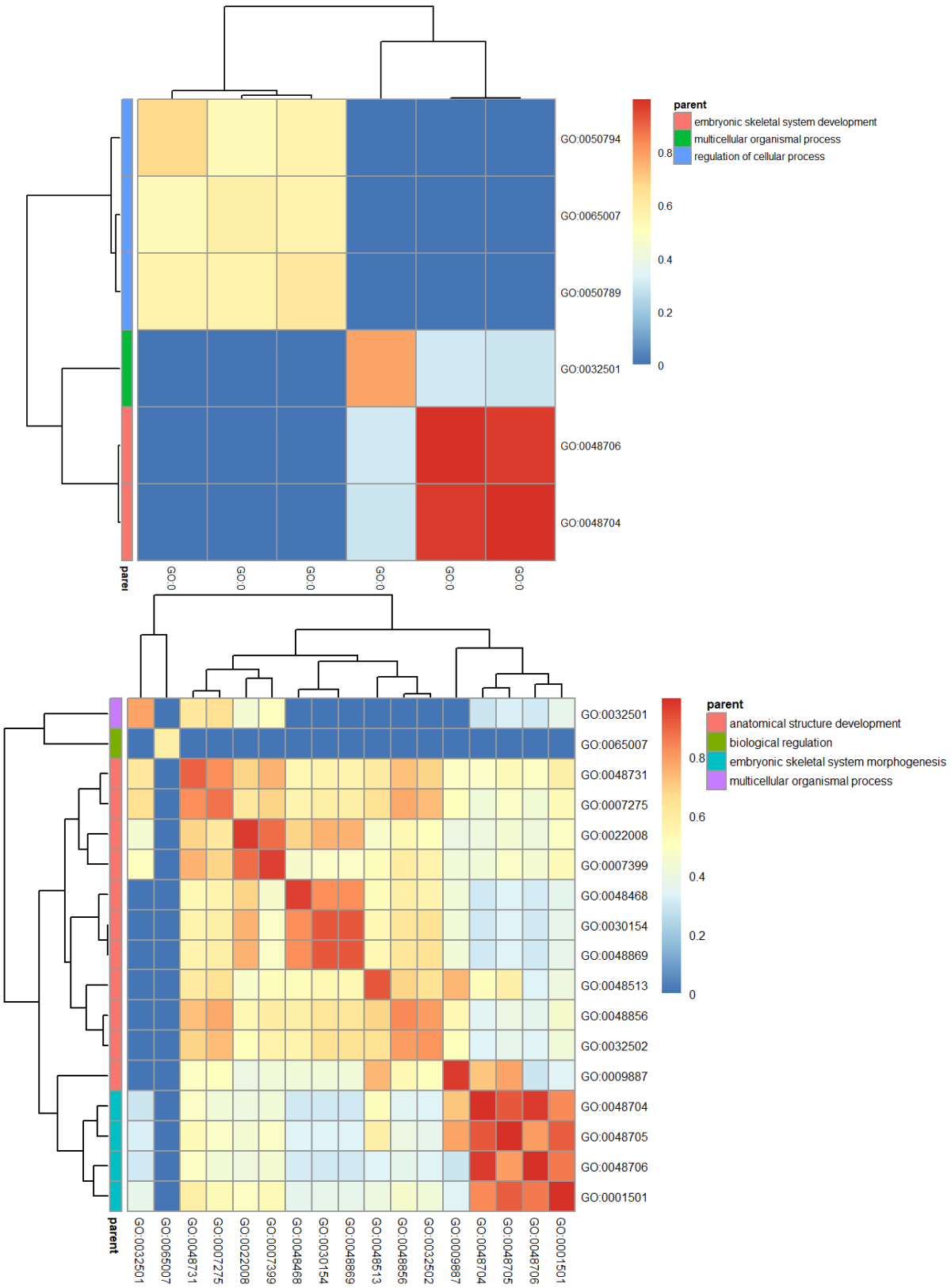
Supplementary figure 14. STRING network of genes under diversifying selection wolves. 55 protein interaction clusters were detected. Non-connected proteins were hidden for clarity. The network is comprised of 273 protein nodes and has an average local clustering coefficient of 0.335 and a PPI enrichment p-value of 5.19e-06. The network did not have functionally enriched gene ontology terms, and two local cluster networks were detected to be enriched for both immunoglobulin V-type and immunoglobulin. With a 6/35, 8/53 counts in network; 1.1 and 1.05 strength and 0.037, 0.07 false discovery rate respectively. The line thickness indicates the strength of data supporting the predicted interactions.



Supplementary figure 15. STRING Network for genes under diversifying selection FRDs. 57 protein interaction clusters were detected. Non-connected proteins were hidden for clarity. The network is comprised of 314 protein nodes with a PPI enrichment p-value of 3.85e-10 and 0.40 average local clustering coefficient. The network was functionally enriched for the biological process gene ontology terms GO:0048704 (embryonic skeletal system morphogenesis) and GO:0048706 (embryonic skeletal system development) with strength of 1.0 and 0.92 respectively.

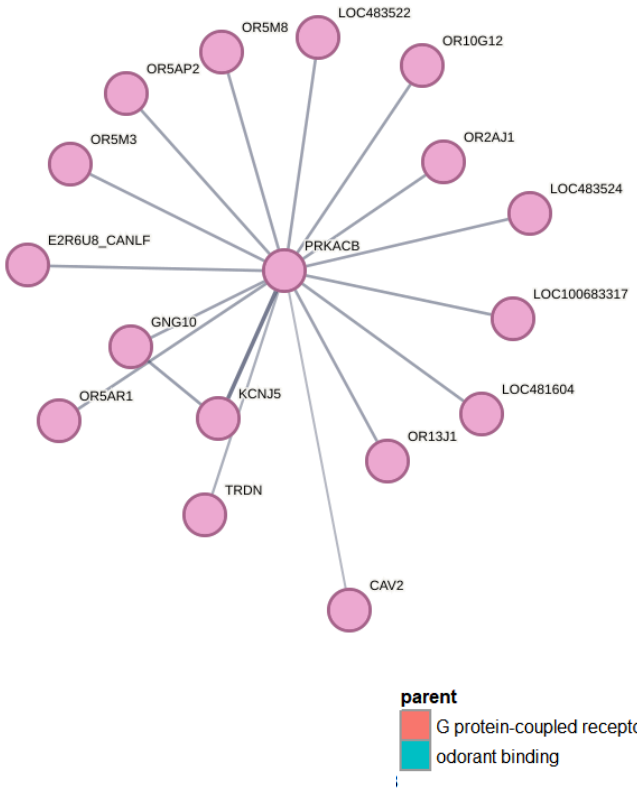


Supplementary figure 16. Heat maps of biological processes gene ontology enriched terms from g:Profiler Wolves. With the *Canis lupus familiaris* genome (top) and *Homo sapiens sapiens* genome (bottom) heatmap colour intensity relates to significance of enrichment the parent terms colour codes are specified in the respect legends.

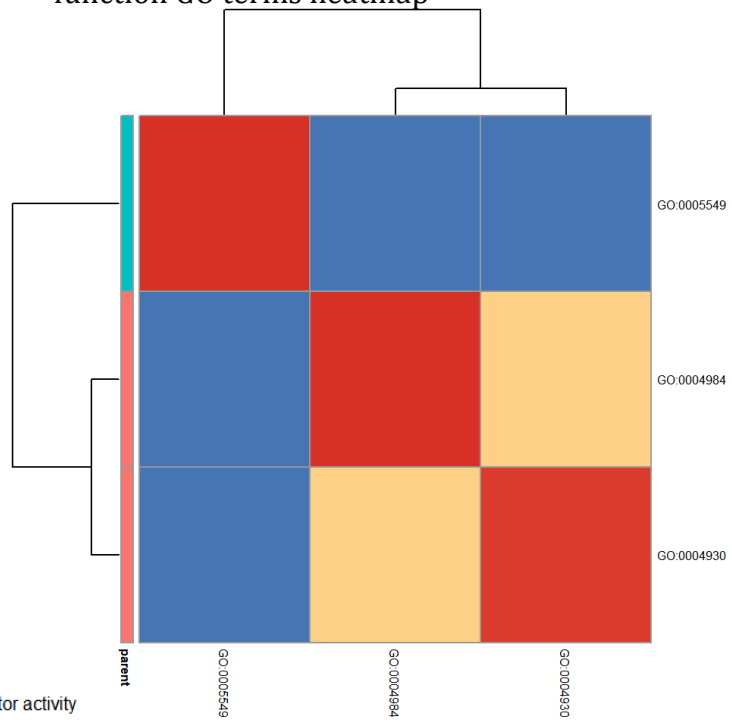


Supplementary figure 17. Heat maps of biological processes gene ontology enriched terms from g:Profiler FRDs. With the *Canis lupus familiaris* genome (top) and *Homo sapiens sapiens* genome (bottom) heatmap colour intensity relates to significance of enrichment the parent terms colour codes are specified in the respect legends.

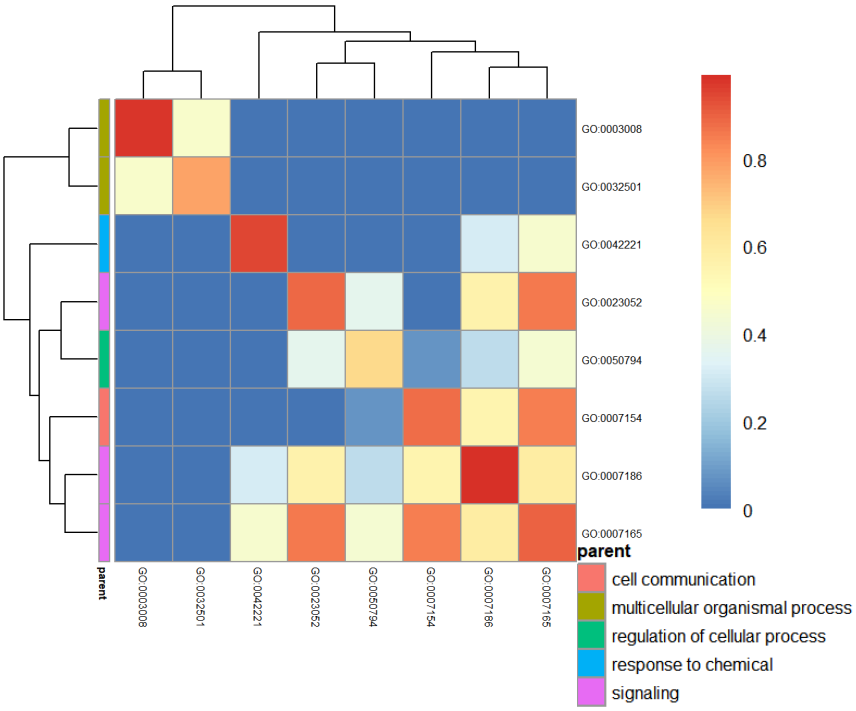
A. Wolves STRING network Cluster 1



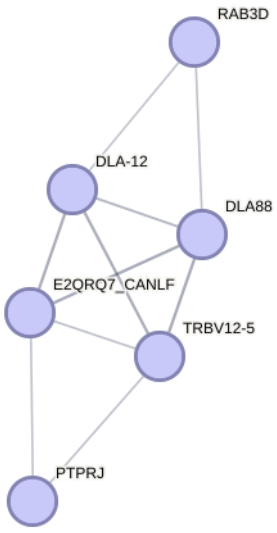
B Wolves STRING network Cluster 1 molecular function GO terms heatmap



C. Wolves STRING network Cluster 1 biological processes GO terms heatmap

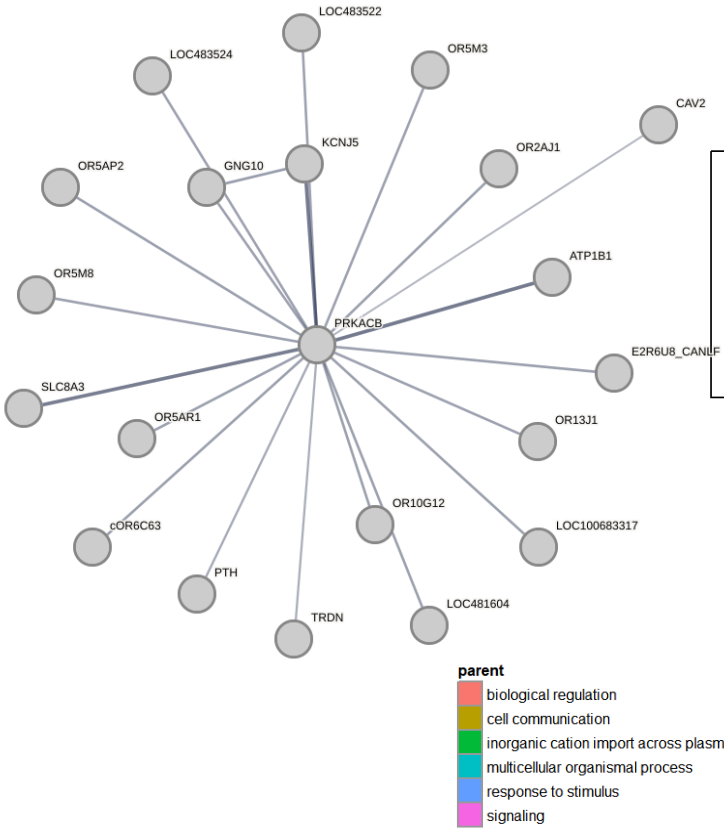


D. Wolves STRING network Cluster 2

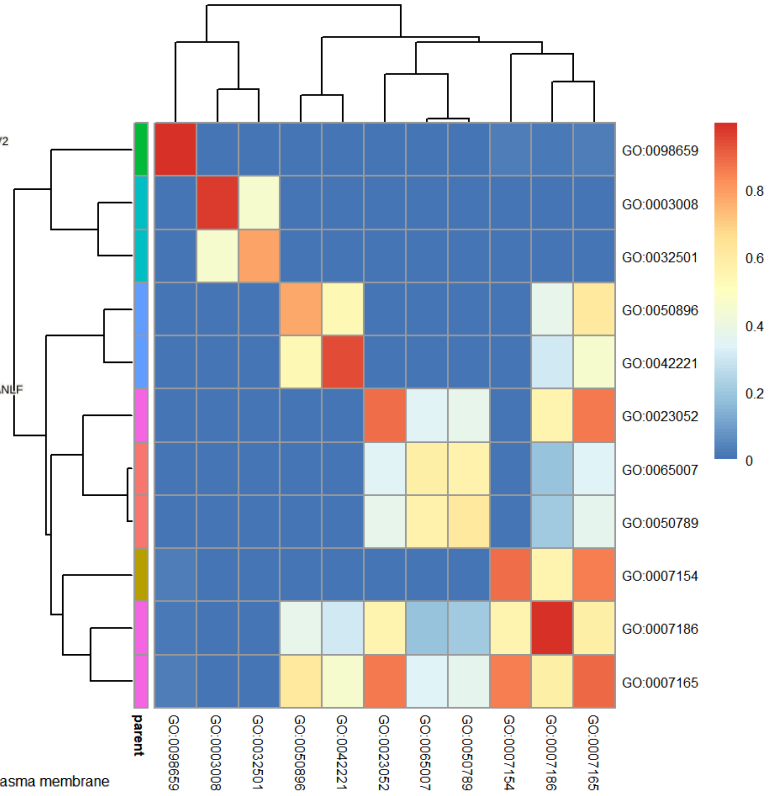


Supplementary figure 18. STRING Networks and heatmaps of the largest STRING sub-clusters Wolves. Cluster 1 (A) had enriched gene ontology (GO) terms for molecular functions (B) and biological (C). Cluster 2 (D) did not have any enriched GO terms.

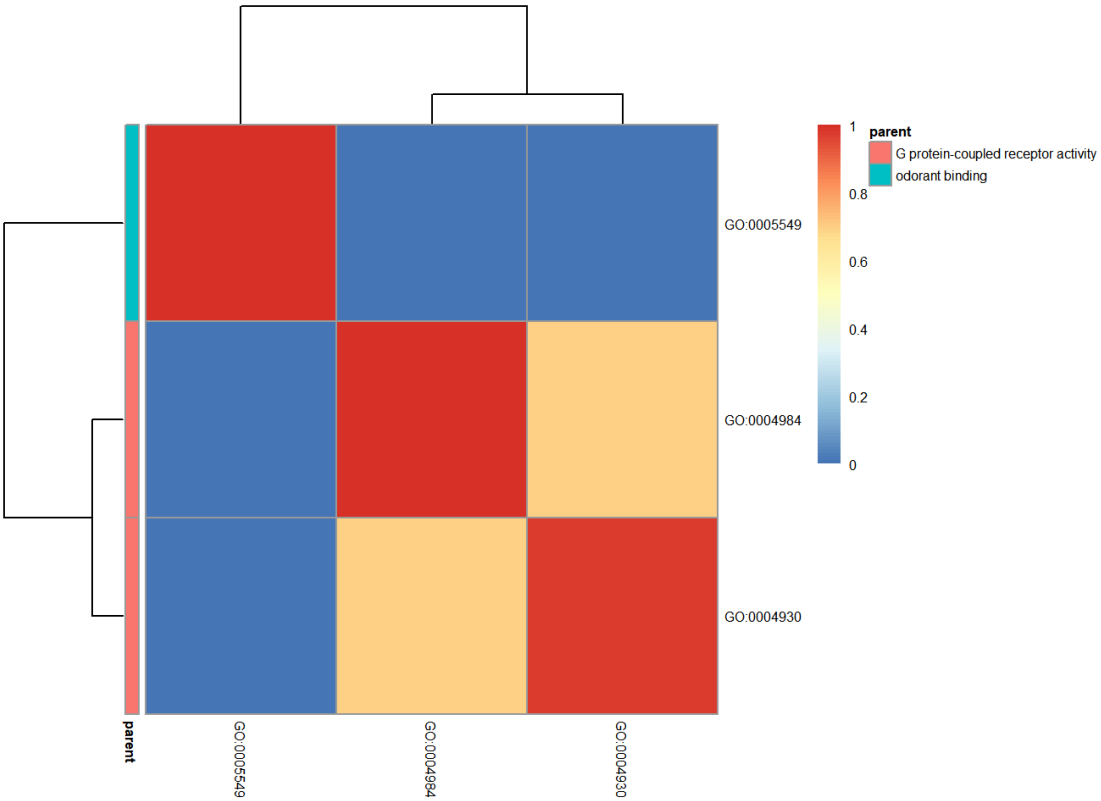
A. FRDs STRING network Cluster 1



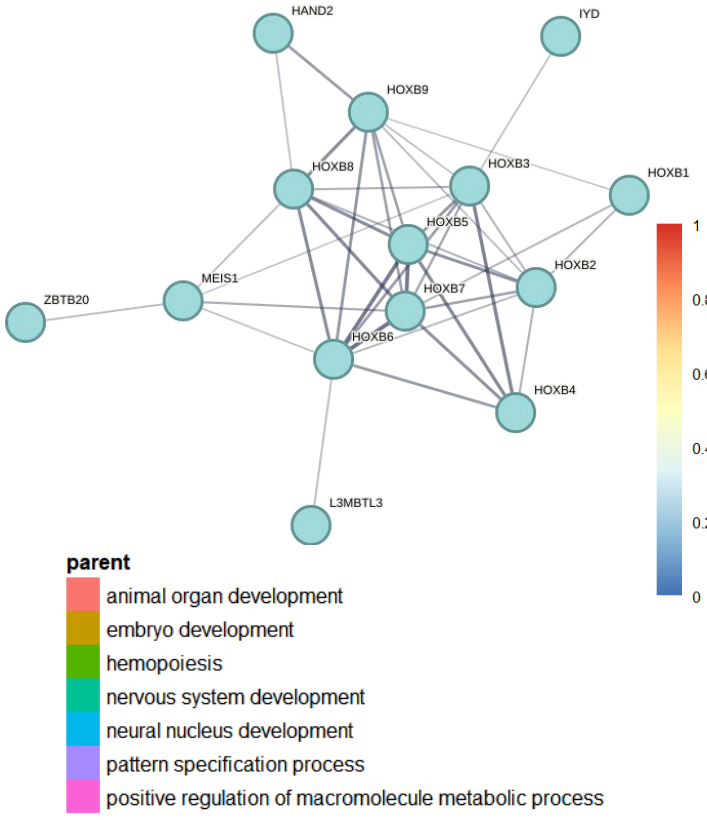
B. Heatmap of GO biological process terms



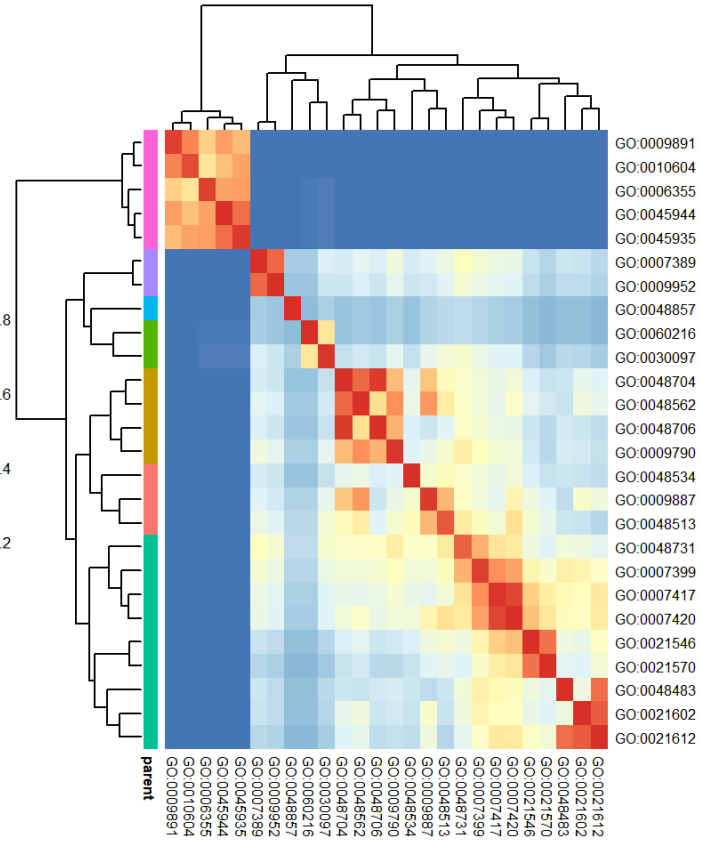
C. FRDs STRING network Cluster 1 heatmap of GO molecular functions terms



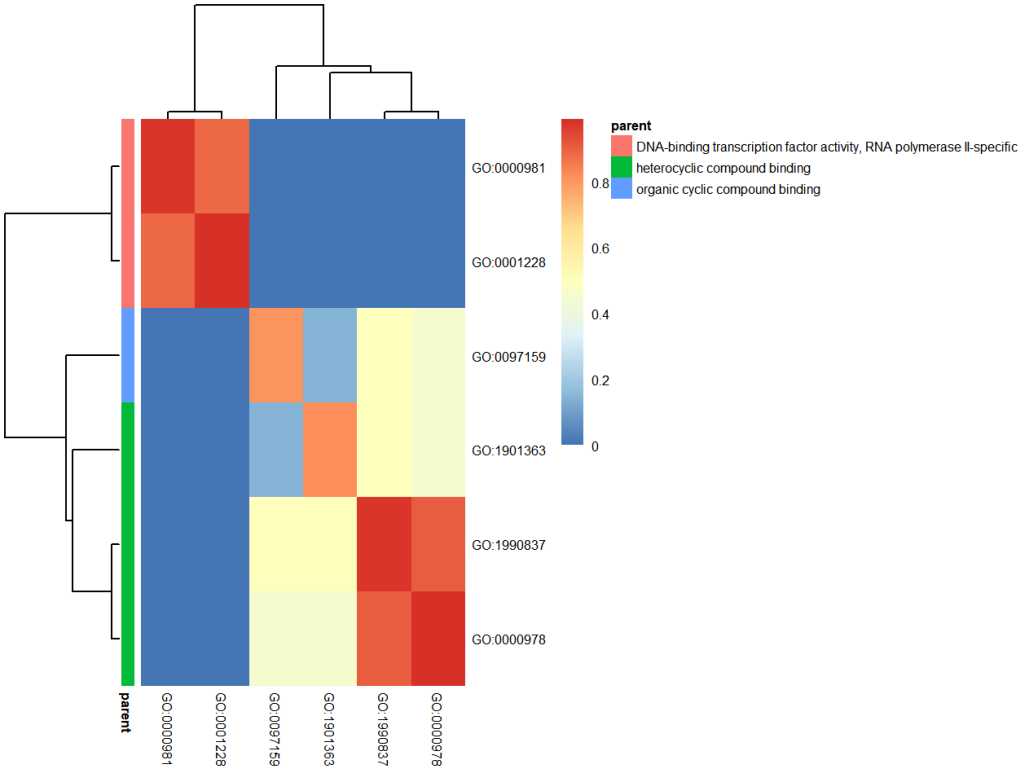
D. FRDs STRING network Cluster 2



E. Heatmap of GO biological process terms



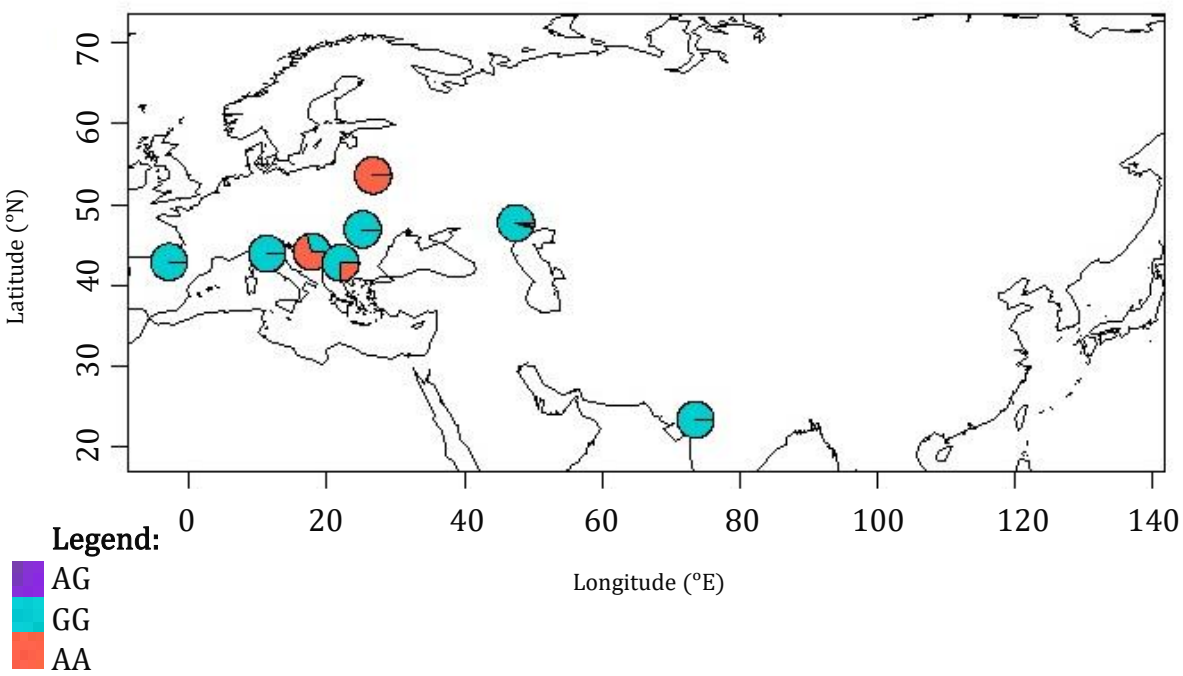
F. FRDs STRING network Cluster 1 heatmap of GO molecular functions terms



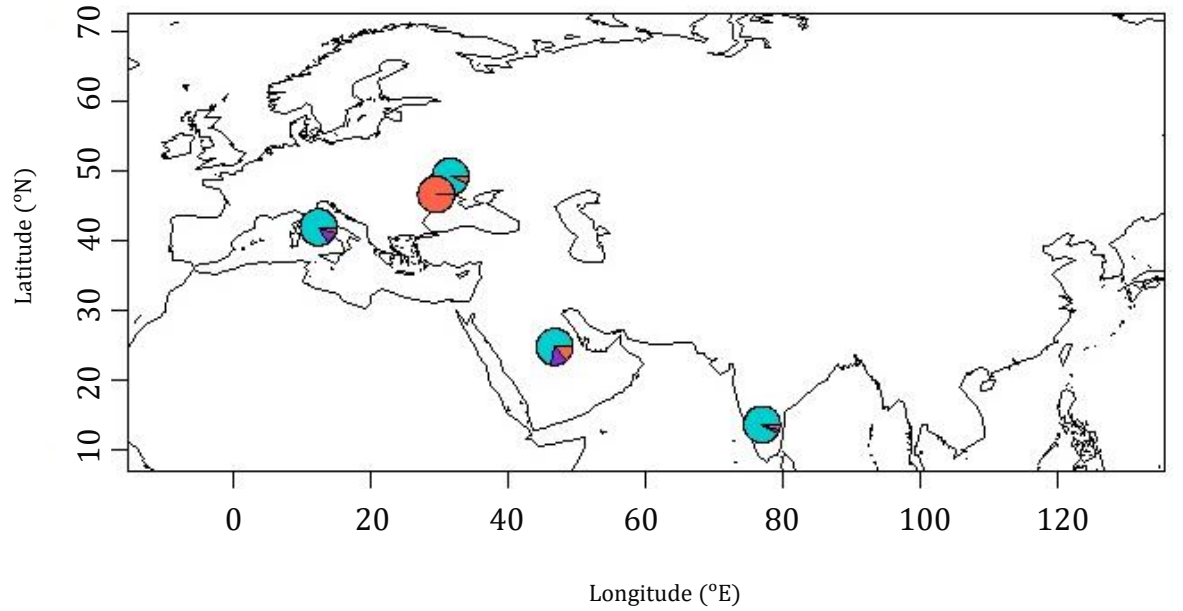
Supplementary figure 19. STRING Networks and heatmaps of the largest STRING sub-clusters FRDs. Cluster 1 (A) was the largest cluster and was enriched for the biological processes gene ontology terms (GO) presented in the heatmap (B) and for the molecular functions GO terms presented in the heatmap (C). Cluster 2 (D) protein interaction

cluster was enriched for the biological process GO terms in the heatmap (E) and for the molecular functions GO terms in the heatmap (F).

A. Genotype frequencies for the SLC6A9 variant Wolves



B. Genotype frequencies for the SLC6A9 variant FRDs



Supplementary figure 20. Mapped genotype frequencies fort the SLC6A9. A mutation (allele A) predicted to impact protein function was detected in both taxa in solute carrier gene SLC6A9. The genotype frequences of this variant were mapped for the populations defined snmf cluster in wolves (A) and FRDs (B).

Supplementary tables and figures: Chapter 3

Custom script for species richness rasters

```
library(maps)
library(terra)
library(ggbiplot)
library(rgdal)
####ENVIRONMENTAL DATA PREP#####
#Loading .shp file for mammals downloaded from IUCN
mammals <- terra::vect('MAMMALS.shp')
#View file specifications
mammals
# Inspect attribute table
terra::head(mammals)
mammals_dist <- read.table('IUCN Mammals.csv',header=T, sep=',', stringsAsFactors = F)
#Raterizing species in bulk
library(tidyverse)
# Make a SpatRaster stack with all species occurrences
# Subset the SpatVector Large Ungulates by Order
#L_Ungulate_Families <- c('BOVIDAE','RAGULIDAE','CERVIDAE',
#                  'GIRAFFIDAE', 'CAMELIDAE', 'HIPPOPOTAMIDAE',
#                 'ANTILOCAPRIDAE', 'EQUIDAE','TAPIRIDAE')
#Raterizing the shp file
#Define an empty SpatRaster of the world at 2 spatial resolution
(r_1deg <- terra::rast())
####LARGE UNGULATES####
#Bovids
Bovid_ranges <- mammals[grep('BOVIDAE',mammals$family),]
Bovid_ranges <- subset(Bovid_ranges, Bovid_ranges$presence==1 &
                       Bovid_ranges$origin==1 & Bovid_ranges$seasonal==1)
```

```
Bovid_terra <- sapply(unique(Bovid_ranges$sci_name),FUN=function(sp) {
  terra::rasterize(subset(Bovid_ranges,Bovid_ranges$sci_name==sp), r_1deg, touches=T)
})
Bovid_terra <- terra::rast(Bovid_terra)
Bovid_richness <- sum(Bovid_terra,na.rm=T)
plot(Bovid_richness, main = "Bovid Richness")
#Cervids
Cervids_ranges <- mammals[grep('CERVIDAE',mammals$family),]
Cervids_ranges <- subset(Cervids_ranges, Cervids_ranges$presence==1 &
                         Cervids_ranges$origin==1 & Cervids_ranges$seasonal==1)
Cervids_terra <- sapply(unique(Cervids_ranges$sci_name),FUN=function(sp) {
  terra::rasterize(subset(Cervids_ranges,Cervids_ranges$sci_name==sp), r_1deg,
touches=T)
})
Cervids_terra <- terra::rast(Cervids_terra)
Cervids_richness <- sum(Cervids_terra,na.rm=T)
plot(Cervids_richness, main = "Cervids Richness")
#Giraffidae
Giraffidae_ranges <- mammals[grep('GIRAFFIDAE',mammals$family),]
Giraffidae_ranges <- subset(Giraffidae_ranges, Giraffidae_ranges$presence==1 &
                            Giraffidae_ranges$origin==1 & Giraffidae_ranges$seasonal==1)
Giraffidae_terra <- sapply(unique(Giraffidae_ranges$sci_name),FUN=function(sp) {
  terra::rasterize(subset(Giraffidae_ranges,Giraffidae_ranges$sci_name==sp), r_1deg,
touches=T)
})
Giraffidae_terra <- terra::rast(Giraffidae_terra)
```

```
Giraffidae_richness <- sum(Giraffidae_terra,na.rm=T)
plot(Giraffidae_richness, main = "Giraffidae Richness")
#Camelids
Camelids_ranges <- mammals[grep('CAMELIDAE',mammals$family),]
Camelids_ranges <- subset(Camelids_ranges, Camelids_ranges$presence==1 &
                    Camelids_ranges$origin==1 & Camelids_ranges$seasonal==1)
Camelids_terra <- sapply(unique(Camelids_ranges$sci_name),FUN=function(sp) {
  terra::rasterize(subset(Camelids_ranges,Camelids_ranges$sci_name==sp), r_1deg,
touches=T)
})
Camelids_terra <- terra::rast(Camelids_terra)
Camelids_richness <- sum(Camelids_terra,na.rm=T)
plot(Camelids_richness, main = "Camelids Richness")
#Hippos
Hippos_ranges <- mammals[grep('HIPPOPOTAMIDAE',mammals$family),]
Hippos_ranges <- subset(Hippos_ranges,Hippos_ranges$presence==1 &
                  Hippos_ranges$origin==1 & Hippos_ranges$seasonal==1)
Hippos_terra <- sapply(unique(Hippos_ranges$sci_name),FUN=function(sp) {
  terra::rasterize(subset(Hippos_ranges,Hippos_ranges$sci_name==sp), r_1deg, touches=T)
})
Hippos_terra <- terra::rast(Hippos_terra)
Hippos_richness <- sum(Hippos_terra,na.rm=T)
plot(Hippos_richness, main = "Hippos Richness")
#Antilope
Antilope_ranges <- mammals[grep('ANTILOCAPRIDAE',mammals$family),]
Antilope_ranges <- subset(Antilope_ranges, Antilope_ranges$presence==1 &
                    Antilope_ranges$origin==1 & Antilope_ranges$seasonal==1)
Antilope_terra <- sapply(unique(Antilope_ranges$sci_name),FUN=function(sp) {
```

```
  terra::rasterize(subset(Antilope_ranges,Antilope_ranges$sci_name==sp), r_1deg,
touches=T)
})
Antilope_terra <- terra::rast(Antilope_terra)
Antilope_richness <- sum(Antilope_terra,na.rm=T)
plot(Antilope_richness, main = "Antilope Richness")
#Equids
Equids_ranges <- mammals[grep('EQUIDAE',mammals$family),]
Equids_ranges <- subset(Equids_ranges, Equids_ranges$presence==1 &
                        Equids_ranges$origin==1 & Equids_ranges$seasonal==1)
Equids_terra <- sapply(unique(Equids_ranges$sci_name),FUN=function(sp) {
  terra::rasterize(subset(Equids_ranges,Equids_ranges$sci_name==sp), r_1deg, touches=T)
})
Equids_terra <- terra::rast(Equids_terra)
Equids_richness <- sum(Equids_terra,na.rm=T)
plot(Equids_richness, main = "Equids Richness")
#Tapirs
Tapirs_ranges <- mammals[grep('TAPIRIDAE',mammals$family),]
Tapirs_ranges <- subset(Tapirs_ranges, Tapirs_ranges$presence==1 &
                        Tapirs_ranges$origin==1 & Tapirs_ranges$seasonal==1)
Tapirs_terra <- sapply(unique(Tapirs_ranges$sci_name),FUN=function(sp) {
  terra::rasterize(subset(Tapirs_ranges,Tapirs_ranges$sci_name==sp), r_1deg, touches=T)
})
Tapirs_terra <- terra::rast(Tapirs_terra)
Tapirs_richness <- sum(Tapirs_terra,na.rm=T)
plot(Tapirs_richness, main = "Tapirs Richness")
### Creating Colection of SpatVectors
L_Ungulates_ranges <- svc(Antilope_ranges, Bovid_ranges,Camelids_ranges,
```

```
                Cervids_ranges, Equids_ranges, Giraffidae_ranges,
                Hippos_ranges, Tapirs_ranges)
length(L_Ungulates_ranges)
L_Ungulate_vect <-vect(L_Ungulates_ranges)
plot (L_Ungulate_vect)
#Species Included
unique(L_Ungulate_vect$sci_name)
#Rasterizing the Large Ungulate Vector
L_Ungulate_terra <- sapply(unique(L_Ungulate_vect$sci_name),FUN=function(sp) {
  terra::rasterize(subset(L_Ungulate_vect,L_Ungulate_vect$sci_name==sp), r_1deg,
touches=T)
})
# Make a SpatRaster stack with all species occurrences
L_Ungulate_terra <- terra::rast(L_Ungulate_terra)
plot(L_Ungulate_terra)
#Calculating Species Richness
L_Ungulate_richness <- sum(L_Ungulate_terra,na.rm=T)
map(databse = "world")
plot(L_Ungulate_richness, add=T, alpha=0.8)
####MEDIUM UNGULATES####
#Tragulidae
Tragulidae_ranges <- mammals[grep('TRAGULIDAE',mammals$family),]
Tragulidae_ranges <- subset(Tragulidae_ranges, Tragulidae_ranges$presence==1 &
                    Tragulidae_ranges$origin==1 & Tragulidae_ranges$seasonal==1)
Tragulidae_terra <- sapply(unique(Tragulidae_ranges$sci_name),FUN=function(sp) {
  terra::rasterize(subset(Tragulidae_ranges,Tragulidae_ranges$sci_name==sp), r_1deg,
touches=T)
})
```

```
Tragulidae_terra <- terra::rast(Tragulidae_terra)
Tragulidae_richness <- sum(Tragulidae_terra,na.rm=T)
plot(Tragulidae_richness, main = "Tragulidae Richness")
#Moschidae
Moschidae_ranges <- mammals[grep('MOSCHIDAE',mammals$family),]
Moschidae_ranges <- subset(Moschidae_ranges, Moschidae_ranges$presence==1 &
                     Moschidae_ranges$origin==1 & Moschidae_ranges$seasonal==1)
Moschidae_terra <- sapply(unique(Moschidae_ranges$sci_name),FUN=function(sp) {
  terra::rasterize(subset(Moschidae_ranges,Moschidae_ranges$sci_name==sp), r_1deg,
touches=T)
})
Moschidae_terra <- terra::rast(Moschidae_terra)
Moschidae_richness <- sum(Moschidae_terra,na.rm=T)
plot(Moschidae_richness, main = "Moschidae Richness")
#Suidae
Suidae_ranges <- mammals[grep('SUIDAE',mammals$family),]
Suidae_ranges <- subset(Suidae_ranges, Suidae_ranges$presence==1 &
                  Suidae_ranges$origin==1 & Suidae_ranges$seasonal==1)
Suidae_terra <- sapply(unique(Suidae_ranges$sci_name),FUN=function(sp) {
  terra::rasterize(subset(Suidae_ranges,Suidae_ranges$sci_name==sp), r_1deg, touches=T)
})
Suidae_terra <- terra::rast(Suidae_terra)
Suidae_richness <- sum(Suidae_terra,na.rm=T)
plot(Suidae_richness, main = "Suidae Richness")
#Tayassuidae
Tayassuidae_ranges <- mammals[grep('TAYASSUIDAE',mammals$family),]
Tayassuidae_ranges <- subset(Tayassuidae_ranges, Tayassuidae_ranges$presence==1 &
```

```
                    Tayassuidae_ranges$origin==1 & Tayassuidae_ranges$seasonal==1)
Tayassuidae_terra <- sapply(unique(Tayassuidae_ranges$sci_name),FUN=function(sp) {
  terra::rasterize(subset(Tayassuidae_ranges,Tayassuidae_ranges$sci_name==sp), r_1deg,
touches=T)
})
Tayassuidae_terra <- terra::rast(Tayassuidae_terra)
Tayassuidae_richness <- sum(Tayassuidae_terra,na.rm=T)
plot(Tayassuidae_richness, main = "Tayassuidae Richness")
#Rhinocerotidae
Rhinocerotidae_ranges <- mammals[grep('RHINOCEROTIDAE',mammals$family),]
Rhinocerotidae_ranges <- subset(Rhinocerotidae_ranges,
Rhinocerotidae_ranges$presence==1 &
                      Rhinocerotidae_ranges$origin==1 &
Rhinocerotidae_ranges$seasonal==1)
Rhinocerotidae_terra <-
sapply(unique(Rhinocerotidae_ranges$sci_name),FUN=function(sp) {
  terra::rasterize(subset(Rhinocerotidae_ranges,Rhinocerotidae_ranges$sci_name==sp),
r_1deg, touches=T)
})
Rhinocerotidae_terra <- terra::rast(Rhinocerotidae_terra)
Rhinocerotidae_richness <- sum(Rhinocerotidae_terra,na.rm=T)
plot(Rhinocerotidae_richness, main = "Rhinocerotidae Richness")
### Creating Colection of SpatVectors
M_Ungulates_ranges <- svc(Tragulidae_ranges, Moschidae_ranges,
                  Suidae_ranges, Tayassuidae_ranges,
                  Rhinocerotidae_ranges)
length(M_Ungulates_ranges)
M_Ungulate_vect <-vect(M_Ungulates_ranges)
plot (M_Ungulate_vect)
```

```
#Species Included

unique(M_Ungulate_vect$sci_name)

#Rasterizing the Large Ungulate Vector

M_Ungulate_terra <- sapply(unique(M_Ungulate_vect$sci_name),FUN=function(sp) {

  terra::rasterize(subset(M_Ungulate_vect,M_Ungulate_vect$sci_name==sp), r_1deg,
touches=T)

})

# Make a SpatRaster stack with all species occurrences

M_Ungulate_terra <- terra::rast(M_Ungulate_terra)

plot(M_Ungulate_terra)

#Calculating Species Richness

M_Ungulate_richness <- sum(M_Ungulate_terra,na.rm=T)

map(databse = "world")

plot(M_Ungulate_richness, add=T, alpha=0.8)

####Other Mammals####

Dermoptera_ranges <- mammals[grep('DERMOPTERA',mammals$order_),]

Dermoptera_ranges <- subset(Dermoptera_ranges, Dermoptera_ranges$presence==1 &

                            Dermoptera_ranges$origin==1 & Dermoptera_ranges$seasonal==1)

Dermoptera_terra <- sapply(unique(Dermoptera_ranges$sci_name),FUN=function(sp) {

  terra::rasterize(subset(Dermoptera_ranges,Dermoptera_ranges$sci_name==sp), r_1deg,
touches=T)

})

Dermoptera_terra <- terra::rast(Dermoptera_terra)

Dermoptera_richness <- sum(Dermoptera_terra,na.rm=T)

plot(Dermoptera_richness, main = "Dermoptera Richness")
```

```
#Eulipotyplha

EULIPOTYPHLA_ranges <- mammals[grep('EULIPOTYPHLA',mammals$order_),]

EULIPOTYPHLA_ranges <- subset(EULIPOTYPHLA_ranges,
EULIPOTYPHLA_ranges$presence==1 &

                        EULIPOTYPHLA_ranges$origin==1 &
EULIPOTYPHLA_ranges$seasonal==1)

EULIPOTYPHLA_terra <-
sapply(unique(EULIPOTYPHLA_ranges$sci_name),FUN=function(sp) {

  terra::rasterize(subset(EULIPOTYPHLA_ranges,EULIPOTYPHLA_ranges$sci_name==sp),
r_1deg, touches=T)

})

EULIPOTYPHLA_terra <- terra::rast(EULIPOTYPHLA_terra)

EULIPOTYPHLA_richness <- sum(EULIPOTYPHLA_terra,na.rm=T)

plot(EULIPOTYPHLA_richness, main = "EULIPOTYPHLA Richness")

#HYRACOIDEA

HYRACOIDEA_ranges <- mammals[grep('DERMOPTERA',mammals$order_),]

HYRACOIDEA_ranges <- subset(HYRACOIDEA_ranges, HYRACOIDEA_ranges$presence==1
&

                    HYRACOIDEA_ranges$origin==1 & HYRACOIDEA_ranges$seasonal==1)

HYRACOIDEA_terra <- sapply(unique(HYRACOIDEA_ranges$sci_name),FUN=function(sp) {

  terra::rasterize(subset(HYRACOIDEA_ranges,HYRACOIDEA_ranges$sci_name==sp),
r_1deg, touches=T)

})

HYRACOIDEA_terra <- terra::rast(HYRACOIDEA_terra)

HYRACOIDEA_richness <- sum(HYRACOIDEA_terra,na.rm=T)

plot(HYRACOIDEA_richness, main = "HYRACOIDEA Richness")

#LAGOMORPHA

LAGOMORPHA_ranges <- mammals[grep('DERMOPTERA',mammals$order_),]

LAGOMORPHA_ranges <- subset(LAGOMORPHA_ranges,
LAGOMORPHA_ranges$presence==1 &
```

```
                    LAGOMORPHA_ranges$origin==1 & LAGOMORPHA_ranges$seasonal==1)
LAGOMORPHA_terra <-
sapply(unique(LAGOMORPHA_ranges$sci_name),FUN=function(sp) {
  terra::rasterize(subset(LAGOMORPHA_ranges,LAGOMORPHA_ranges$sci_name==sp),
r_1deg, touches=T)
})
LAGOMORPHA_terra <- terra::rast(LAGOMORPHA_terra)
LAGOMORPHA_richness <- sum(LAGOMORPHA_terra,na.rm=T)
plot(LAGOMORPHA_richness, main = "LAGOMORPHA Richness")
##MACROSCELIDEA
MACROSCELIDEA_ranges <- mammals[grep('DERMOPTERA',mammals$order_),]
MACROSCELIDEA_ranges <- subset(MACROSCELIDEA_ranges,
MACROSCELIDEA_ranges$presence==1 &
                      MACROSCELIDEA_ranges$origin==1 &
MACROSCELIDEA_ranges$seasonal==1)
MACROSCELIDEA_terra <-
sapply(unique(MACROSCELIDEA_ranges$sci_name),FUN=function(sp) {
terra::rasterize(subset(MACROSCELIDEA_ranges,MACROSCELIDEA_ranges$sci_name==sp)
, r_1deg, touches=T)
})
MACROSCELIDEA_terra <- terra::rast(MACROSCELIDEA_terra)
MACROSCELIDEA_richness <- sum(MACROSCELIDEA_terra,na.rm=T)
plot(MACROSCELIDEA_richness, main = "MACROSCELIDEARichness")
#PHOLIDOTA
PHOLIDOTA_ranges <- mammals[grep('DERMOPTERA',mammals$order_),]
PHOLIDOTA_ranges <- subset(PHOLIDOTA_ranges, PHOLIDOTA_ranges$presence==1 &
                    PHOLIDOTA_ranges$origin==1 & PHOLIDOTA_ranges$seasonal==1)
PHOLIDOTA_terra <- sapply(unique(PHOLIDOTA_ranges$sci_name),FUN=function(sp) {
  terra::rasterize(subset(PHOLIDOTA_ranges,PHOLIDOTA_ranges$sci_name==sp), r_1deg,
touches=T)
```

```
})


PHOLIDOTA_terra <- terra::rast(PHOLIDOTA_terra)

PHOLIDOTA_richness <- sum(PHOLIDOTA_terra,na.rm=T)

plot(PHOLIDOTA_richness, main = "PHOLIDOTA Richness")

#PRIMATES

PRIMATES_ranges <- mammals[grep('DERMOPTERA',mammals$order_),]

PRIMATES_ranges <- subset(PRIMATES_ranges, PRIMATES_ranges$presence==1 &

                    PRIMATES_ranges$origin==1 & PRIMATES_ranges$seasonal==1)

PRIMATES_terra <- sapply(unique(PRIMATES_ranges$sci_name),FUN=function(sp) {

  terra::rasterize(subset(PRIMATES_ranges,PRIMATES_ranges$sci_name==sp), r_1deg,
touches=T)

})

PRIMATES_terra <- terra::rast(PRIMATES_terra)

PRIMATES_richness <- sum(PRIMATES_terra,na.rm=T)

plot(PRIMATES_richness, main = "PRIMATESRichness")

#PROBOSCIDEA

PROBOSCIDEA_ranges <- mammals[grep('DERMOPTERA',mammals$order_),]

PROBOSCIDEA_ranges <- subset(PROBOSCIDEA_ranges,
PROBOSCIDEA_ranges$presence==1 &

                      PROBOSCIDEA_ranges$origin==1 & PROBOSCIDEA_ranges$seasonal==1)

PROBOSCIDEA_terra <-
sapply(unique(PROBOSCIDEA_ranges$sci_name),FUN=function(sp) {

  terra::rasterize(subset(PROBOSCIDEA_ranges,PROBOSCIDEA_ranges$sci_name==sp),
r_1deg, touches=T)

})

PROBOSCIDEA_terra <- terra::rast(PROBOSCIDEA_terra)

PROBOSCIDEA_richness <- sum(PROBOSCIDEA_terra,na.rm=T)
```

```
plot(PROBOSCIDEA_richness, main = "PROBOSCIDEARichness")

#RODENTIA

RODENTIA_ranges <- mammals[grep('DERMOPTERA',mammals$order_),]

RODENTIA_ranges <- subset(RODENTIA_ranges, RODENTIA_ranges$presence==1 &

                  RODENTIA_ranges$origin==1 & RODENTIA_ranges$seasonal==1)

RODENTIA_terra <- sapply(unique(RODENTIA_ranges$sci_name),FUN=function(sp) {

  terra::rasterize(subset(RODENTIA_ranges,RODENTIA_ranges$sci_name==sp), r_1deg,
touches=T)

})

RODENTIA_terra <- terra::rast(RODENTIA_terra)

RODENTIA_richness <- sum(RODENTIA_terra,na.rm=T)

plot(RODENTIA_richness, main = "RODENTIARichness")

### Creating Colection of SpatVectors

Other_Mammals_ranges <- svc(Dermoptera_ranges,EULIPOTYPHLA_ranges,

                  HYRACOIDEA_ranges, LAGOMORPHA_ranges,

                  MACROSCELIDEA_ranges,PHOLIDOTA_ranges,

                  PRIMATES_ranges, PROBOSCIDEA_ranges,

                  RODENTIA_ranges )

length(Other_Mammals_ranges)

Other_Mammals_vect <-vect(Other_Mammals_ranges)

plot (Other_Mammals_vect)

#Species Included

unique(Other_Mammals_vect$sci_name)

#Rasterizing the Large Ungulate Vector

Other_Mammals_terra <-
sapply(unique(Other_Mammals_vect$sci_name),FUN=function(sp) {

  terra::rasterize(subset(Other_Mammals_vect,Other_Mammals_vect$sci_name==sp),
r_1deg, touches=T)

})
```

```
# Make a SpatRaster stack with all species occurrences
Other_Mammals_terra <- terra::rast(Other_Mammals_terra)
plot(Other_Mammals_terra)
#Calculating Species Richness
Other_Mammals_richness <- sum(Other_Mammals_terra,na.rm=T)
map(databse = "world")
plot(Other_Mammals_richness, add=T, alpha=0.8)
####Carnivores####
Carnivore_ranges <- mammals[grep('CARNIVORA',mammals$order_),]
Carnivore_ranges <- subset(Carnivore_ranges, Carnivore_ranges$presence==1 &
                    Carnivore_ranges$origin==1 & Carnivore_ranges$seasonal==1)
Carnivore_terra <- sapply(unique(Carnivore_ranges$sci_name),FUN=function(sp) {
  terra::rasterize(subset(Carnivore_ranges,Carnivore_ranges$sci_name==sp), r_1deg,
touches=T)
})
Carnivore_terra <- terra::rast(Carnivore_terra)
Carnivore_richness <- sum(Carnivore_terra,na.rm=T)
plot(Carnivore_richness, main = "Carnivore Richness")
```

Supplementary tables

Supplementary table 1. Number of individuals per species for each biome

Biomes	FRDs	Wolves
Boreal forests/taiga	18	7
Deserts & xeric shrublands	60	20
Flooded grasslands & savannas	0	1
Mediterranean forests, woodlands & scrub	83	29
Montane grasslands & shrublands	0	4
Temperate broadleaf & mixed forests	165	129
Temperate conifer forests	6	4
Temperate grasslands, savannas & shrublands	17	34
Tropical & subtropical moist broadleaf forests	17	2
Tundra	4	2
	370	232

Supplementary table 2. SNPs which failed the CanFam6 Lifting

Wolves	FRDs
chr1:77331625	chr9:30749295
chr5:30633810	chr15:7966269
chr7:45483670	chr16:10592414
chr8:73866479	chr22:4905063
chr33:21799085	chr24:34976566
chr34:34445210	chr29:32704741
	chr29:40737172
	chr31:21179600

*the error was: "#Deleted in new".

Supplementary table 3. Genes without a UnitProt codes, excluded from g:Profiler analysis

		ENSMBL ID	Biotype or Unitprot Name	Predicted Consequence
Wolves	Non-protein coding	ENSCAFG00000047078	lncRNA	Downstream variant
		ENSCAFG00000053379	lncRNA	Downstream variant
		ENSCAFG00000056657	lncRNA	Downstream variant
		ENSCAFG00000051243	lncRNA	Downstream variant
		ENSCAFG00000055193	lncRNA	Downstream variant
		ENSCAFG00000053903	lncRNA	Downstream variant
		ENSCAFG00000054699	lncRNA	Downstream variant
		ENSCAFG00000055143	lncRNA	Downstream variant
		ENSCAFG00000053664	lncRNA	Downstream variant
		ENSCAFG00000058369	lncRNA	Downstream variant
		ENSCAFG00000050507	lncRNA	Downstream variant
		ENSCAFG00000051710	lncRNA	Downstream variant
		ENSCAFG00000054112	lncRNA	Downstream variant
		ENSCAFG00000046977	lncRNA	Downstream variant
		ENSCAFG00000055008	lncRNA	Downstream variant
		ENSCAFG00000051702	lncRNA	Downstream variant
		ENSCAFG00000054665	lncRNA	Downstream variant
		ENSCAFG00000056311	lncRNA	Downstream variant
		ENSCAFG00000052692	lncRNA	Downstream variant
		ENSCAFG00000057918	lncRNA	Downstream variant
		ENSCAFG00000054624	lncRNA	Downstream variant
		ENSCAFG00000057144	lncRNA	Downstream variant
		ENSCAFG00000050463	lncRNA	Downstream variant
		ENSCAFG00000051133	lncRNA	Downstream variant
		ENSCAFG00000055658	lncRNA	Downstream variant
		ENSCAFG00000053621	lncRNA	Downstream variant
		ENSCAFG00000054794	lncRNA	Downstream variant
		ENSCAFG00000053289	lncRNA	Downstream variant
		ENSCAFG00000054191	lncRNA	Downstream variant
		ENSCAFG00000045810	lncRNA	Downstream variant
		ENSCAFG00000053081	lncRNA	Downstream variant
		ENSCAFG00000058449	lncRNA	Downstream variant
		ENSCAFG00000056638	lncRNA	Downstream variant
		ENSCAFG00000051217	lncRNA	Downstream variant
		ENSCAFG00000059330	lncRNA	Downstream variant
		ENSCAFG00000057793	lncRNA	Downstream variant
		ENSCAFG00000057844	lncRNA	Downstream variant
		ENSCAFG00000052488	lncRNA	Downstream variant
		ENSCAFG00000052910	lncRNA	Downstream variant
		ENSCAFG00000056622	lncRNA	Downstream variant
		ENSCAFG00000057994	lncRNA	Downstream variant

		ENSCAFG00000056459	lncRNA	Downstream variant
		ENSCAFG00000053844	lncRNA	Downstream variant
		ENSCAFG00000059104	lncRNA	Downstream variant
		ENSCAFG00000053622	lncRNA	Downstream variant
		ENSCAFG00000053460	lncRNA	Downstream variant
		ENSCAFG00000033391	lncRNA	Downstream variant
		ENSCAFG00000050831	lncRNA	Downstream variant
		ENSCAFG00000056407	lncRNA	Downstream variant
		ENSCAFG00000053898	lncRNA	Downstream variant
		ENSCAFG00000058425	lncRNA	Downstream variant
		ENSCAFG00000055129	lncRNA	Downstream variant
		ENSCAFG00000056110	lncRNA	Downstream variant
		ENSCAFG00000054255	lncRNA	Downstream variant
		ENSCAFG00000055809	lncRNA	Downstream variant
		ENSCAFG00000051549	lncRNA	Downstream variant
		ENSCAFG00000058162	lncRNA	Downstream variant
		ENSCAFG00000052140	lncRNA	Downstream variant
		ENSCAFG00000055266	lncRNA	Downstream variant
		ENSCAFG00000058841	lncRNA	Downstream variant
		ENSCAFG00000051320	lncRNA	Downstream variant
		ENSCAFG00000057211	lncRNA	Downstream variant
		ENSCAFG00000050221	lncRNA	Downstream variant
		ENSCAFG00000054805	lncRNA	Downstream variant
		ENSCAFG00000057810	lncRNA	Downstream variant
		ENSCAFG00000056063	lncRNA	Downstream variant
		ENSCAFG00000056815	lncRNA	Downstream variant
		ENSCAFG00000057067	lncRNA	Downstream variant
		ENSCAFG00000054579	lncRNA	Downstream variant
		ENSCAFG00000051764	lncRNA	Downstream variant
		ENSCAFG00000051181	lncRNA	Downstream variant
		ENSCAFG00000051649	lncRNA	Downstream variant
		ENSCAFG00000055215	lncRNA	Downstream variant
		ENSCAFG00000051554	lncRNA	Downstream variant
		ENSCAFG00000059258	lncRNA	Downstream variant
		ENSCAFG00000055734	lncRNA	Downstream variant
		ENSCAFG00000050952	lncRNA	Downstream variant
		ENSCAFG00000058729	lncRNA	Downstream variant
		ENSCAFG00000051191	lncRNA	Downstream variant
		ENSCAFG00000042352	lncRNA	Downstream variant
		ENSCAFG00000050641	lncRNA	Downstream variant
		ENSCAFG00000054923	lncRNA	Downstream variant
		ENSCAFG00000055945	lncRNA	Downstream variant
		ENSCAFG00000053964	lncRNA	Downstream variant

		ENSCAFG00000059324	lncRNA	Downstream variant
		ENSCAFG00000051123	lncRNA	Downstream variant
		ENSCAFG00000057649	lncRNA	Downstream variant
		ENSCAFG00000059327	lncRNA	Downstream variant
		ENSCAFG00000054661	lncRNA	Downstream variant
		ENSCAFG00000056410	lncRNA	Downstream variant
		ENSCAFG00000050405	lncRNA	Downstream variant
		ENSCAFG00000056021	lncRNA	Downstream variant
		ENSCAFG00000041070	lncRNA	Downstream variant
		ENSCAFG00000058704	lncRNA	Downstream variant
		ENSCAFG00000057368	lncRNA	Downstream variant
		ENSCAFG00000054764	lncRNA	Downstream variant
		ENSCAFG00000007649	lncRNA	Downstream variant
		ENSCAFG00000033953	lncRNA	Downstream variant
		ENSCAFG00000050338	lncRNA	Downstream variant
		ENSCAFG00000051893	lncRNA	Downstream variant
		ENSCAFG00000054076	lncRNA	Downstream variant
		ENSCAFG00000051996	lncRNA	Downstream variant
		ENSCAFG00000057144	lncRNA	Intron variant, noncoding transcript
		ENSCAFG00000050322	lncRNA	Intron variant, noncoding transcript
		ENSCAFG00000056919	lncRNA	Intron variant, noncoding transcript
		ENSCAFG00000045992	lncRNA	Intron variant, noncoding transcript
		ENSCAFG00000057034	lncRNA	Intron variant, noncoding transcript
		ENSCAFG00000057034	lncRNA	Intron variant, noncoding transcript
		ENSCAFG00000052996	lncRNA	Noncoding transcript exon variant
		ENSCAFG00000052228	lncRNA	Upstream variant
		ENSCAFG00000053211	lncRNA	Upstream variant
		ENSCAFG00000055845	lncRNA	Upstream variant
		ENSCAFG00000054845	lncRNA	Upstream variant
		ENSCAFG00000055302	lncRNA	Upstream variant
		ENSCAFG00000035380	lncRNA	Upstream variant
		ENSCAFG00000055624	lncRNA	Upstream variant
		ENSCAFG00000050514	lncRNA	Upstream variant
		ENSCAFG00000050708	lncRNA	Upstream variant
		ENSCAFG00000050799	lncRNA	Upstream variant
		ENSCAFG00000051090	lncRNA	Upstream variant
		ENSCAFG00000050730	lncRNA	Upstream variant
		ENSCAFG00000056492	lncRNA	Upstream variant

		ENSCAFG00000055531	lncRNA	Upstream variant
		ENSCAFG00000052992	lncRNA	Upstream variant
		ENSCAFG00000058293	lncRNA	Upstream variant
		ENSCAFG00000056293	lncRNA	Upstream variant
		ENSCAFG00000050771	lncRNA	Upstream variant
		ENSCAFG00000053380	lncRNA	Upstream variant
		ENSCAFG00000059400	lncRNA	Upstream variant
		ENSCAFG00000059135	lncRNA	Upstream variant
		ENSCAFG00000050347	lncRNA	Upstream variant
		ENSCAFG00000041664	lncRNA	Upstream variant
		ENSCAFG00000056101	lncRNA	Upstream variant
		ENSCAFG00000055147	lncRNA	Upstream variant
		ENSCAFG00000058619	lncRNA	Upstream variant
		ENSCAFG00000057066	lncRNA	Upstream variant
		ENSCAFG00000058950	lncRNA	Upstream variant
		ENSCAFG00000056879	lncRNA	Upstream variant
		ENSCAFG00000053098	lncRNA	Upstream variant
		ENSCAFG00000053796	lncRNA	Upstream variant
		ENSCAFG00000056786	lncRNA	Upstream variant
		ENSCAFG00000050499	lncRNA	Upstream variant
		ENSCAFG00000058102	lncRNA	Upstream variant
		ENSCAFG00000041550	lncRNA	Upstream variant
		ENSCAFG00000054390	lncRNA	Upstream variant
		ENSCAFG00000055057	lncRNA	Upstream variant
		ENSCAFG00000057012	lncRNA	Upstream variant
		ENSCAFG00000054563	lncRNA	Upstream variant
		ENSCAFG00000051288	lncRNA	Upstream variant
		ENSCAFG00000053047	lncRNA	Upstream variant
		ENSCAFG00000054275	lncRNA	Upstream variant
		ENSCAFG00000058603	lncRNA	Upstream variant
		ENSCAFG00000043007	lncRNA	Upstream variant
		ENSCAFG00000058444	lncRNA	Upstream variant
		ENSCAFG00000054295	lncRNA	Upstream variant
		ENSCAFG00000054631	lncRNA	Upstream variant
		ENSCAFG00000056719	lncRNA	Upstream variant
		ENSCAFG00000050670	lncRNA	Upstream variant
		ENSCAFG00000051106	lncRNA	Upstream variant
		ENSCAFG00000057029	lncRNA	Upstream variant
		ENSCAFG00000057508	lncRNA	Upstream variant
		ENSCAFG00000053685	lncRNA	Upstream variant
		ENSCAFG00000056846	lncRNA	Upstream variant
		ENSCAFG00000037664	lncRNA	Upstream variant
		ENSCAFG00000057607	lncRNA	Upstream variant

		ENSCAFG00000055197	lncRNA	Upstream variant
		ENSCAFG00000055378	lncRNA	Upstream variant
		ENSCAFG00000055056	lncRNA	Upstream variant
		ENSCAFG00000050600	lncRNA	Upstream variant
		ENSCAFG00000055445	lncRNA	Upstream variant
		ENSCAFG00000054149	lncRNA	Upstream variant
		ENSCAFG00000058829	lncRNA	Upstream variant
		ENSCAFG00000050192	lncRNA	Upstream variant
		ENSCAFG00000050393	lncRNA	Upstream variant
		ENSCAFG00000057846	lncRNA	Upstream variant
		ENSCAFG00000056919	lncRNA	Upstream variant
		ENSCAFG00000053256	lncRNA	Upstream variant
		ENSCAFG00000052299	lncRNA	Upstream variant
		ENSCAFG00000044148	lncRNA	Upstream variant
		ENSCAFG00000056456	lncRNA	Upstream variant
		ENSCAFG00000058092	lncRNA	Upstream variant
		ENSCAFG00000050181	lncRNA	Upstream variant
		ENSCAFG00000055777	lncRNA	Upstream variant
		ENSCAFG00000058541	lncRNA	Upstream variant
		ENSCAFG00000053858	lncRNA	Upstream variant
		ENSCAFG00000057176	lncRNA	Upstream variant
		ENSCAFG00000059474	lncRNA	Upstream variant
		ENSCAFG00000051914	lncRNA	Upstream variant
		ENSCAFG00000056685	lncRNA	Upstream variant
		ENSCAFG00000053991	lncRNA	Upstream variant
		ENSCAFG00000049979	lncRNA	Upstream variant
		ENSCAFG00000054485	lncRNA	Upstream variant
		ENSCAFG00000055122	lncRNA	Upstream variant
		ENSCAFG00000052760	lncRNA	Upstream variant
		ENSCAFG00000055507	lncRNA	Upstream variant
		ENSCAFG00000057883	lncRNA	Upstream variant
		ENSCAFG00000059447	lncRNA	Upstream variant
		ENSCAFG00000059088	lncRNA	Upstream variant
		ENSCAFG00000051214	lncRNA	Upstream variant
		ENSCAFG00000058735	lncRNA	Upstream variant
		ENSCAFG00000054601	lncRNA	Upstream variant
		ENSCAFG00000052598	lncRNA	Upstream variant
		ENSCAFG00000053274	lncRNA	Upstream variant
		ENSCAFG00000055291	lncRNA	Upstream variant
		ENSCAFG00000055945	lncRNA	Upstream variant
		ENSCAFG00000055230	lncRNA	Upstream variant
		ENSCAFG00000052452	lncRNA	Upstream variant
		ENSCAFG00000059465	lncRNA	Upstream variant

		ENSCAFG00000050379	lncRNA	Upstream variant
		ENSCAFG00000046487	lncRNA	Upstream variant
		ENSCAFG00000055005	lncRNA	Upstream variant
		ENSCAFG00000055005	lncRNA	Upstream variant
		ENSCAFG00000052933	lncRNA	Upstream variant
		ENSCAFG00000057482	lncRNA	Upstream variant
		ENSCAFG00000051513	lncRNA	Upstream variant
		ENSCAFG00000050210	lncRNA	Upstream variant
		ENSCAFG00000053330	lncRNA	Upstream variant
		ENSCAFG00000056186	lncRNA	Upstream variant
		ENSCAFG00000039143	lncRNA	Upstream variant
		ENSCAFG00000054778	lncRNA	Upstream variant
		ENSCAFG00000058727	lncRNA	Upstream variant
		ENSCAFG00000055258	lncRNA	Upstream variant
		ENSCAFG00000058687	lncRNA	Upstream variant
		ENSCAFG00000058350	lncRNA	Upstream variant
		ENSCAFG00000057759	lncRNA	Upstream variant
		ENSCAFG00000050732	lncRNA	Upstream variant
		ENSCAFG00000059161	lncRNA	Upstream variant
		ENSCAFG00000056680	lncRNA	Upstream variant
		ENSCAFG00000056506	lncRNA	Upstream variant
		ENSCAFG00000055660	lncRNA	Upstream variant
		ENSCAFG00000051464	miRNA	Downstream variant
		ENSCAFG00000059461	miRNA	Downstream variant
		ENSCAFG00000055989	miRNA	Downstream variant
		ENSCAFG00000020620	miRNA	Downstream variant
		ENSCAFG00000057461	miRNA	Downstream variant
		ENSCAFG00000058385	miRNA	Downstream variant
		ENSCAFG00000050653	miRNA	Downstream variant
		ENSCAFG00000054697	miRNA	Downstream variant
		ENSCAFG00000020580	miRNA	Downstream variant
		ENSCAFG00000022793	miRNA	Downstream variant
		ENSCAFG00000054483	miRNA	Downstream variant
		ENSCAFG00000056769	miRNA	Upstream variant
		ENSCAFG00000050992	miRNA	Upstream variant
		ENSCAFG00000057294	miRNA	Upstream variant
		ENSCAFG00000050369	miRNA	Upstream variant
		ENSCAFG00000051885	miRNA	Upstream variant
		ENSCAFG00000020542	miRNA	Upstream variant
		ENSCAFG00000025700	miRNA	Upstream variant
		ENSCAFG00000020452	miRNA	Upstream variant
		ENSCAFG00000020463	miRNA	Upstream variant
		ENSCAFG00000020442	miRNA	Upstream variant

		ENSCAFG00000020489	miRNA	Upstream variant
		ENSCAFG00000020617	miRNA	Upstream variant
		ENSCAFG00000020495	miRNA	Upstream variant
		ENSCAFG00000020606	miRNA	Upstream variant
		ENSCAFG00000022889	miRNA	Upstream variant
		ENSCAFG00000054891	miRNA	Upstream variant
		ENSCAFG00000056418	miRNA	Upstream variant
		ENSCAFG00000022817	miRNA	Upstream variant
		ENSCAFG00000055873	miRNA	Upstream variant
		ENSCAFG00000042237	misc_RNA	Downstream variant
		ENSCAFG00000052332	misc_RNA	Upstream variant
		ENSCAFG00000044002	processed_pseudogene	Downstream variant
		ENSCAFG00000045752	processed_pseudogene	Downstream variant
		ENSCAFG00000001382	protein_coding	Downstream variant
		ENSCAFG00000047701	pseudogene	Downstream variant
		ENSCAFG00000044050	pseudogene	Downstream variant
		ENSCAFG00000049571	pseudogene	Downstream variant
		ENSCAFG00000005045	pseudogene	Downstream variant
		ENSCAFG00000053001	pseudogene	Downstream variant
		ENSCAFG00000051390	pseudogene	Downstream variant
		ENSCAFG00000019324	pseudogene	Downstream variant
		ENSCAFG00000057714	pseudogene	Downstream variant
		ENSCAFG00000047469	pseudogene	Downstream variant
		ENSCAFG00000050167	pseudogene	Upstream variant
		ENSCAFG00000025856	snoRNA	Downstream variant
		ENSCAFG00000025938	snoRNA	Downstream variant
		ENSCAFG00000026458	snoRNA	Downstream variant
		ENSCAFG00000028347	snoRNA	Downstream variant
		ENSCAFG00000046646	snoRNA	Downstream variant
		ENSCAFG00000021670	snoRNA	Downstream variant
		ENSCAFG00000047210	snoRNA	Downstream variant
		ENSCAFG00000026218	snoRNA	Downstream variant
		ENSCAFG00000020883	snoRNA	Downstream variant
		ENSCAFG00000021073	snoRNA	Downstream variant
		ENSCAFG00000022180	snoRNA	Downstream variant
		ENSCAFG00000025801	snoRNA	Downstream variant
		ENSCAFG00000025916	snoRNA	Downstream variant
		ENSCAFG00000026010	snoRNA	Downstream variant
		ENSCAFG00000026225	snoRNA	Downstream variant
		ENSCAFG00000026245	snoRNA	Downstream variant
		ENSCAFG00000026308	snoRNA	Downstream variant
		ENSCAFG00000028095	snoRNA	Downstream variant
		ENSCAFG00000049353	snoRNA	Upstream variant

		ENSCAFG00000025856	snoRNA	Upstream variant
		ENSCAFG00000025938	snoRNA	Upstream variant
		ENSCAFG00000042140	snoRNA	Upstream variant
		ENSCAFG00000052908	snRNA	Downstream variant
		ENSCAFG00000027380	snRNA	Downstream variant
		ENSCAFG00000028074	snRNA	Downstream variant
		ENSCAFG00000048739	snRNA	Downstream variant
		ENSCAFG00000025744	snRNA	Downstream variant
		ENSCAFG00000021082	snRNA	Downstream variant
		ENSCAFG00000025943	snRNA	Downstream variant
		ENSCAFG00000027282	snRNA	Downstream variant
		ENSCAFG00000044904	snRNA	Downstream variant
		ENSCAFG00000028044	snRNA	Downstream variant
		ENSCAFG00000058981	snRNA	Downstream variant
		ENSCAFG00000020917	snRNA	Downstream variant
		ENSCAFG00000025934	snRNA	Downstream variant
		ENSCAFG00000053823	snRNA	Downstream variant
		ENSCAFG00000042682	snRNA	Downstream variant
		ENSCAFG00000025914	snRNA	Downstream variant
		ENSCAFG00000041213	snRNA	Downstream variant
		ENSCAFG00000045207	snRNA	Downstream variant
		ENSCAFG00000049208	snRNA	Downstream variant
		ENSCAFG00000028377	snRNA	Upstream variant
		ENSCAFG00000027615	snRNA	Upstream variant
		ENSCAFG00000048434	snRNA	Upstream variant
		ENSCAFG00000021272	snRNA	Upstream variant
		ENSCAFG00000043295	snRNA	Upstream variant
		ENSCAFG00000056325	snRNA	Upstream variant
		ENSCAFG00000025869	snRNA	Upstream variant
		ENSCAFG00000025866	snRNA	Upstream variant
		ENSCAFG00000044492	snRNA	Upstream variant
		ENSCAFG00000027033	snRNA	Upstream variant
		ENSCAFG00000021299	snRNA	Upstream variant
		ENSCAFG00000044164	snRNA	Upstream variant
		ENSCAFG00000027440	snRNA	Upstream variant
		ENSCAFG00000056082	snRNA	Upstream variant
		ENSCAFG00000027467	snRNA	Upstream variant
		ENSCAFG00000028083	snRNA	Upstream variant
		ENSCAFG00000043191	snRNA	Upstream variant
		ENSCAFG00000044989	snRNA	Upstream variant
		ENSCAFG00000027204	snRNA	Upstream variant
		ENSCAFG00000047640	snRNA	Upstream variant
		ENSCAFG00000020895	snRNA	Upstream variant

		ENSCAFG00000022270	snRNA	Upstream variant
		ENSCAFG00000026147	snRNA	Upstream variant
		ENSCAFG00000022640	snRNA	Upstream variant
		ENSCAFG00000027061	snRNA	Upstream variant
		ENSCAFG00000021248	snRNA	Upstream variant
		ENSCAFG00000021194	snRNA	Upstream variant
		ENSCAFG00000052040	snRNA	Upstream variant
		ENSCAFG00000022587	snRNA	Upstream variant
		ENSCAFG00000043433	snRNA	Upstream variant
		ENSCAFG00000028309	snRNA	Upstream variant
		ENSCAFG00000043542	snRNA	Upstream variant
		ENSCAFG00000020758	snRNA	Upstream variant
		ENSCAFG00000027883	snRNA	Upstream variant
		ENSCAFG00000021286	snRNA	Upstream variant
		ENSCAFG00000027167	snRNA	Upstream variant
		ENSCAFG00000022021	snRNA	Upstream variant
	Protein coding	ENSCAFG00000024322	40S ribosomal protein S27	Upstream variant
		ENSCAFG00000019801	5-hydroxyisourate hydrolase (HIU hydrolase) (HIUHase) (EC 3.5.2.17)	Downstream variant
		ENSCAFG00000058987	AHNAK nucleoprotein 2	Upstream variant
		ENSCAFG00000003445	Ankyrin repeat domain 7	Upstream variant
		ENSCAFG00000042967	ATP synthase-coupling factor 6, mitochondrial (ATP synthase peripheral stalk subunit F6)	Downstream variant
		ENSCAFG00000048656	Beta-defensin	Downstream variant
		ENSCAFG00000051604	Beta-defensin	Downstream variant
		ENSCAFG00000047267	Beta-defensin 130-like	Upstream variant
		ENSCAFG00000029223	BPTI/Kunitz inhibitor domain-containing protein	Downstream variant
		ENSCAFG00000044662	Catsper channel auxiliary subunit zeta	Upstream variant
		ENSCAFG00000016664	Cytochrome c oxidase subunit 6C (Cytochrome c oxidase polypeptide VIc)	Downstream variant
		ENSCAFG00000043376	Deoxyribonuclease TATDN1	Downstream variant
		ENSCAFG00000024733	EF-hand domain-containing protein	Upstream variant
		ENSCAFG00000050264	EF-hand domain-containing protein	Downstream variant
		ENSCAFG00000007015	G patch domain-containing protein 11 (Coiled-coil domain-containing protein 75)	Upstream variant
		ENSCAFG00000049586	Glyceraldehyde 3-phosphate dehydrogenase NAD(P) binding domain-containing protein	Downstream variant
		ENSCAFG00000051423	Heterogeneous nuclear ribonucleoprotein A1/A2 C-terminal domain-containing	Downstream variant

			protein	
		ENSCAFG00000049617	HIG1 domain-containing protein	Downstream variant
		ENSCAFG00000030509	Ig-like domain-containing protein	Downstream variant
		ENSCAFG00000047890	Ig-like domain-containing protein	Downstream variant
		ENSCAFG00000053642	KRAB domain-containing protein	Upstream variant
		ENSCAFG00000058087	KRAB domain-containing protein	Upstream variant
		ENSCAFG00000056057	KRAB domain-containing protein	Downstream variant
		ENSCAFG00000051969	L1 transposable element RRM domain-containing protein	Upstream variant
		ENSCAFG00000050271	L1 transposable element RRM domain-containing protein	Upstream variant
		ENSCAFG00000050271	L1 transposable element RRM domain-containing protein	Upstream variant
		ENSCAFG00000044246	Large ribosomal subunit protein uL15m (39S ribosomal protein L15, mitochondrial)	Downstream variant
		ENSCAFG00000043434	L-lactate dehydrogenase (EC 1.1.1.27)	Upstream variant
		ENSCAFG00000041728	Mago homolog B, exon junction complex subunit	Downstream variant
		ENSCAFG00000002173	Methylosome subunit pICln	Downstream variant
		ENSCAFG00000024105	MHC class I-like antigen recognition-like domain-containing protein	Upstream variant
		ENSCAFG00000043857	Mitochondrial ribosomal protein L20	Upstream variant
		ENSCAFG00000032733	NHL repeat containing 4	Downstream variant
		ENSCAFG00000056600	Olfactory receptor	Upstream variant
		ENSCAFG00000019325	Olfactory receptor	
		ENSCAFG00000019327	Olfactory receptor	
		ENSCAFG00000023121	Olfactory receptor-like protein DTMT	
		ENSCAFG00000050625	PDZ domain-containing protein	Upstream variant
		ENSCAFG00000025190	Peptidase C19 ubiquitin carboxyl-terminal hydrolase domain-containing protein	Upstream variant
		ENSCAFG00000052816	Peptidase S26 domain-containing protein	Downstream variant
		ENSCAFG00000031076	PMIS2 transmembrane protein	Upstream variant
		ENSCAFG00000031477	Proline rich 23A	Upstream variant
		ENSCAFG00000057708	Proline rich 23A	Upstream variant
		ENSCAFG00000008535	Proteasome 20S subunit alpha 6	Downstream variant
		ENSCAFG00000058840	Pyruvate dehydrogenase phosphatase regulatory subunit	
		ENSCAFG00000000883	Reticulon	Upstream variant

		ENSCAFG00000055733	Reverse transcriptase domain-containing protein	Upstream variant
		ENSCAFG00000058629	Reverse transcriptase domain-containing protein	
		ENSCAFG00000009308	Ribosomal protein L32	Upstream variant
		ENSCAFG00000053722	Ribosomal protein L37a	Upstream variant
		ENSCAFG00000058812	RIC8B	Upstream variant
		ENSCAFG00000048509	RNA-directed DNA polymerase (EC 2.7.7.49)	
		ENSCAFG00000048437	RRM domain-containing protein	Downstream variant
		ENSCAFG00000028693	Serine/arginine repetitive matrix 5	Upstream variant
		ENSCAFG00000045246	Serine-threonine kinase receptor-associated protein	Upstream variant
		ENSCAFG00000042570	Small integral membrane protein 27	Upstream variant
		ENSCAFG00000050038	Small ribosomal subunit protein uS10m (28S ribosomal protein S10, mitochondrial)	Upstream variant
		ENSCAFG00000013789	SSD domain-containing protein	Intron variant, noncoding transcript
		ENSCAFG00000043934	Thymosin beta-10	Downstream variant
		ENSCAFG00000019300	TLE family member 6, subcortical maternal complex member	Downstream variant
		ENSCAFG00000024456	Transcription factor BTF3	Upstream variant
		ENSCAFG00000055493	Transcription factor MafB	Downstream variant
		ENSCAFG00000010760	Tubulin/FtsZ GTPase domain-containing protein	Downstream variant
		ENSCAFG00000043862	UBC core domain-containing protein	Upstream variant
		ENSCAFG00000046052	Ubiquitin-like domain-containing protein	Upstream variant
		ENSCAFG00000049866	Uncharacterized protein	Upstream variant
		ENSCAFG00000049094	Uncharacterized protein	Upstream variant
		ENSCAFG00000059138	Uncharacterized protein	Upstream variant
		ENSCAFG00000046775	Uncharacterized protein	Upstream variant
		ENSCAFG00000044869	Uncharacterized protein	Upstream variant
		ENSCAFG00000045797	Uncharacterized protein	Downstream variant
		ENSCAFG00000056327	Uncharacterized protein	Downstream variant
		ENSCAFG00000044731	Uncharacterized protein	Downstream variant
		ENSCAFG00000046775	Uncharacterized protein	Downstream variant
		ENSCAFG00000041772	Uncharacterized protein	Downstream variant
		ENSCAFG00000049014	Uncharacterized protein	Downstream variant
		ENSCAFG00000044462	Uncharacterized protein	Downstream variant
		ENSCAFG00000045878	Uncharacterized protein	Downstream variant
		ENSCAFG00000047764	Uncharacterized protein	Intron variant
		ENSCAFG00000052661	Uncharacterized protein	Upstream variant

		ENSCAFG00000047006	Vacuolar protein sorting-associated protein 29	Upstream variant
		ENSCAFG00000057448	Vomeronasal type-1 receptor	Upstream variant
		ENSCAFG00000050043	Vps72/YL1 N-terminal domain-containing protein	Downstream variant
		ENSCAFG00000042312	V-SNARE coiled-coil homology domain-containing protein	Downstream variant
		ENSCAFG00000009723	WAP four-disulfide core domain 11	Downstream variant
		ENSCAFG00000005986	WD repeat domain 87	Upstream variant
Free ranging dogs	Non-protein coding	ENSCAFG00000034862	lncRNA	Intron variant, noncoding transcript
		ENSCAFG00000035932	lncRNA	Intron variant, noncoding transcript
		ENSCAFG00000036487	lncRNA	Downstream variant
		ENSCAFG00000038294	lncRNA	Upstream variant
		ENSCAFG00000041370	lncRNA	Downstream variant
		ENSCAFG00000045810	lncRNA	Downstream variant
		ENSCAFG00000047324	lncRNA	Upstream variant
		ENSCAFG00000048403	lncRNA	Upstream variant
		ENSCAFG00000049921	lncRNA	Downstream variant
		ENSCAFG00000050153	lncRNA	Intron variant, noncoding transcript
		ENSCAFG00000050161	lncRNA	Downstream variant
		ENSCAFG00000050225	lncRNA	Downstream variant
		ENSCAFG00000050228	lncRNA	Downstream variant
		ENSCAFG00000050237	lncRNA	Upstream variant
		ENSCAFG00000050237	lncRNA	Upstream variant
		ENSCAFG00000050237	lncRNA	Upstream variant
		ENSCAFG00000050273	lncRNA	Upstream variant
		ENSCAFG00000050279	lncRNA	Downstream variant
		ENSCAFG00000050366	lncRNA	Upstream variant
		ENSCAFG00000050366	lncRNA	Upstream variant
		ENSCAFG00000050366	lncRNA	Upstream variant
		ENSCAFG00000050398	lncRNA	Downstream variant
		ENSCAFG00000050470	lncRNA	Upstream variant
		ENSCAFG00000050476	lncRNA	Upstream variant
		ENSCAFG00000050516	lncRNA	Upstream variant
		ENSCAFG00000050539	lncRNA	Downstream variant
		ENSCAFG00000050555	lncRNA	Upstream variant
		ENSCAFG00000050612	lncRNA	Downstream variant
		ENSCAFG00000050699	lncRNA	Downstream variant
		ENSCAFG00000050712	lncRNA	Downstream variant
		ENSCAFG00000050789	lncRNA	Upstream variant
		ENSCAFG00000050817	lncRNA	Downstream variant
		ENSCAFG00000050865	lncRNA	Downstream variant

		ENSCAFG00000050929	lncRNA	Upstream variant
		ENSCAFG00000051028	lncRNA	Downstream variant
		ENSCAFG00000051070	lncRNA	Upstream variant
		ENSCAFG00000051217	lncRNA	Upstream variant
		ENSCAFG00000051272	lncRNA	Upstream variant
		ENSCAFG00000051288	lncRNA	Downstream variant
		ENSCAFG00000051302	lncRNA	Upstream variant
		ENSCAFG00000051335	lncRNA	Upstream variant
		ENSCAFG00000051383	lncRNA	Downstream variant
		ENSCAFG00000051395	lncRNA	Upstream variant
		ENSCAFG00000051434	lncRNA	Downstream variant
		ENSCAFG00000051452	lncRNA	Upstream variant
		ENSCAFG00000051460	lncRNA	Downstream variant
		ENSCAFG00000051555	lncRNA	Downstream variant
		ENSCAFG00000051637	lncRNA	Upstream variant
		ENSCAFG00000051661	lncRNA	Downstream variant
		ENSCAFG00000051678	lncRNA	Upstream variant
		ENSCAFG00000051752	lncRNA	Downstream variant
		ENSCAFG00000051796	lncRNA	Downstream variant
		ENSCAFG00000051860	lncRNA	Noncoding transcript exon variant
		ENSCAFG00000051870	lncRNA	Downstream variant
		ENSCAFG00000051921	lncRNA	Upstream variant
		ENSCAFG00000051945	lncRNA	Upstream variant
		ENSCAFG00000052090	lncRNA	Upstream variant
		ENSCAFG00000052105	lncRNA	Downstream variant
		ENSCAFG00000052160	lncRNA	Upstream variant
		ENSCAFG00000052171	lncRNA	Downstream variant
		ENSCAFG00000052304	lncRNA	Downstream variant
		ENSCAFG00000052308	lncRNA	Upstream variant
		ENSCAFG00000052342	lncRNA	Downstream variant
		ENSCAFG00000052352	lncRNA	Downstream variant
		ENSCAFG00000052536	lncRNA	Upstream variant
		ENSCAFG00000052713	lncRNA	Upstream variant
		ENSCAFG00000052727	lncRNA	Upstream variant
		ENSCAFG00000052762	lncRNA	Downstream variant
		ENSCAFG00000052891	lncRNA	Upstream variant
		ENSCAFG00000052972	lncRNA	Downstream variant
		ENSCAFG00000052983	lncRNA	Upstream variant
		ENSCAFG00000053060	lncRNA	Downstream variant
		ENSCAFG00000053075	lncRNA	Downstream variant
		ENSCAFG00000053092	lncRNA	Downstream variant
		ENSCAFG00000053107	lncRNA	Upstream variant

		ENSCAFG00000053379	lncRNA	Upstream variant
		ENSCAFG00000053455	lncRNA	Downstream variant
		ENSCAFG00000053492	lncRNA	Downstream variant
		ENSCAFG00000053524	lncRNA	Downstream variant
		ENSCAFG00000053565	lncRNA	Downstream variant
		ENSCAFG00000053588	lncRNA	Downstream variant
		ENSCAFG00000053648	lncRNA	Downstream variant
		ENSCAFG00000053712	lncRNA	Downstream variant
		ENSCAFG00000053720	lncRNA	Upstream variant
		ENSCAFG00000053745	lncRNA	Downstream variant
		ENSCAFG00000053899	lncRNA	Downstream variant
		ENSCAFG00000053917	lncRNA	Downstream variant
		ENSCAFG00000053959	lncRNA	Upstream variant
		ENSCAFG00000053959	lncRNA	Upstream variant
		ENSCAFG00000053959	lncRNA	Upstream variant
		ENSCAFG00000054033	lncRNA	Upstream variant
		ENSCAFG00000054065	lncRNA	Upstream variant
		ENSCAFG00000054099	lncRNA	Downstream variant
		ENSCAFG00000054112	lncRNA	Upstream variant
		ENSCAFG00000054204	lncRNA	Upstream variant
		ENSCAFG00000054234	lncRNA	Downstream variant
		ENSCAFG00000054466	lncRNA	Upstream variant
		ENSCAFG00000054544	lncRNA	Upstream variant
		ENSCAFG00000054559	lncRNA	Downstream variant
		ENSCAFG00000054563	lncRNA	Upstream variant
		ENSCAFG00000054575	lncRNA	Noncoding transcript exon variant
		ENSCAFG00000054576	lncRNA	Downstream variant
		ENSCAFG00000054596	lncRNA	Downstream variant
		ENSCAFG00000054671	lncRNA	Downstream variant
		ENSCAFG00000054739	lncRNA	Upstream variant
		ENSCAFG00000054760	lncRNA	Upstream variant
		ENSCAFG00000054782	lncRNA	Downstream variant
		ENSCAFG00000054974	lncRNA	Upstream variant
		ENSCAFG00000054974	lncRNA	Upstream variant
		ENSCAFG00000055031	lncRNA	Upstream variant
		ENSCAFG00000055124	lncRNA	Upstream variant
		ENSCAFG00000055139	lncRNA	Downstream variant
		ENSCAFG00000055236	lncRNA	Upstream variant
		ENSCAFG00000055254	lncRNA	Upstream variant
		ENSCAFG00000055275	lncRNA	Downstream variant
		ENSCAFG00000055289	lncRNA	Noncoding transcript exon variant
		ENSCAFG00000055353	lncRNA	Downstream variant

		ENSCAFG00000055388	lncRNA	Downstream variant
		ENSCAFG00000055439	lncRNA	Downstream variant
		ENSCAFG00000055448	lncRNA	Downstream variant
		ENSCAFG00000055497	lncRNA	Downstream variant
		ENSCAFG00000055787	lncRNA	Upstream variant
		ENSCAFG00000055829	lncRNA	Upstream variant
		ENSCAFG00000055840	lncRNA	Upstream variant
		ENSCAFG00000055862	lncRNA	Downstream variant
		ENSCAFG00000055902	lncRNA	Downstream variant
		ENSCAFG00000055908	lncRNA	Upstream variant
		ENSCAFG00000055909	lncRNA	Noncoding transcript exon variant
		ENSCAFG00000055919	lncRNA	Upstream variant
		ENSCAFG00000055985	lncRNA	Downstream variant
		ENSCAFG00000056038	lncRNA	Downstream variant
		ENSCAFG00000056062	lncRNA	Downstream variant
		ENSCAFG00000056124	lncRNA	Upstream variant
		ENSCAFG00000056169	lncRNA	Upstream variant
		ENSCAFG00000056373	lncRNA	Upstream variant
		ENSCAFG00000056483	lncRNA	Downstream variant
		ENSCAFG00000056504	lncRNA	Upstream variant
		ENSCAFG00000056599	lncRNA	Downstream variant
		ENSCAFG00000056699	lncRNA	Upstream variant
		ENSCAFG00000056751	lncRNA	Downstream variant
		ENSCAFG00000056757	lncRNA	Upstream variant
		ENSCAFG00000056785	lncRNA	Downstream variant
		ENSCAFG00000056829	lncRNA	Upstream variant
		ENSCAFG00000056859	lncRNA	Upstream variant
		ENSCAFG00000056859	lncRNA	Upstream variant
		ENSCAFG00000056859	lncRNA	Upstream variant
		ENSCAFG00000056870	lncRNA	Upstream variant
		ENSCAFG00000056979	lncRNA	Downstream variant
		ENSCAFG00000057012	lncRNA	Upstream variant
		ENSCAFG00000057022	lncRNA	Downstream variant
		ENSCAFG00000057071	lncRNA	Upstream variant
		ENSCAFG00000057075	lncRNA	Downstream variant
		ENSCAFG00000057099	lncRNA	Downstream variant
		ENSCAFG00000057178	lncRNA	Upstream variant
		ENSCAFG00000057187	lncRNA	Downstream variant
		ENSCAFG00000057210	lncRNA	Upstream variant
		ENSCAFG00000057251	lncRNA	Downstream variant
		ENSCAFG00000057280	lncRNA	Downstream variant
		ENSCAFG00000057349	lncRNA	Downstream variant

		ENSCAFG00000057389	lncRNA	Downstream variant
		ENSCAFG00000057424	lncRNA	Upstream variant
		ENSCAFG00000057424	lncRNA	Upstream variant
		ENSCAFG00000057445	lncRNA	Upstream variant
		ENSCAFG00000057488	lncRNA	Upstream variant
		ENSCAFG00000057756	lncRNA	Downstream variant
		ENSCAFG00000057785	lncRNA	Downstream variant
		ENSCAFG00000057876	lncRNA	Upstream variant
		ENSCAFG00000057901	lncRNA	Upstream variant
		ENSCAFG00000057924	lncRNA	Upstream variant
		ENSCAFG00000057930	lncRNA	Downstream variant
		ENSCAFG00000057939	lncRNA	Downstream variant
		ENSCAFG00000057961	lncRNA	Upstream variant
		ENSCAFG00000058081	lncRNA	Upstream variant
		ENSCAFG00000058208	lncRNA	Upstream variant
		ENSCAFG00000058342	lncRNA	Downstream variant
		ENSCAFG00000058350	lncRNA	Downstream variant
		ENSCAFG00000058360	lncRNA	Upstream variant
		ENSCAFG00000058374	lncRNA	Downstream variant
		ENSCAFG00000058374	lncRNA	Downstream variant
		ENSCAFG00000058374	lncRNA	Downstream variant
		ENSCAFG00000058383	lncRNA	Downstream variant
		ENSCAFG00000058384	lncRNA	Upstream variant
		ENSCAFG00000058444	lncRNA	Upstream variant
		ENSCAFG00000058491	lncRNA	Upstream variant
		ENSCAFG00000058674	lncRNA	Upstream variant
		ENSCAFG00000058719	lncRNA	Upstream variant
		ENSCAFG00000058839	lncRNA	Upstream variant
		ENSCAFG00000058854	lncRNA	Downstream variant
		ENSCAFG00000058856	lncRNA	Downstream variant
		ENSCAFG00000058863	lncRNA	Downstream variant
		ENSCAFG00000058900	lncRNA	Upstream variant
		ENSCAFG00000058900	lncRNA	Upstream variant
		ENSCAFG00000058944	lncRNA	Downstream variant
		ENSCAFG00000058986	lncRNA	Downstream variant
		ENSCAFG00000059058	lncRNA	Downstream variant
		ENSCAFG00000059075	lncRNA	Upstream variant
		ENSCAFG00000059085	lncRNA	Downstream variant
		ENSCAFG00000059098	lncRNA	Downstream variant
		ENSCAFG00000059134	lncRNA	Downstream variant
		ENSCAFG00000059316	lncRNA	Upstream variant
		ENSCAFG00000059318	lncRNA	Downstream variant
		ENSCAFG00000059322	lncRNA	Upstream variant

		ENSCAFG00000059330	lncRNA	Upstream variant
		ENSCAFG00000059337	lncRNA	Downstream variant
		ENSCAFG00000020468	miRNA	Upstream variant
		ENSCAFG00000022806	miRNA	Downstream variant
		ENSCAFG00000025651	miRNA	Downstream variant
		ENSCAFG00000025665	miRNA	Upstream variant
		ENSCAFG00000025700	miRNA	Upstream variant
		ENSCAFG00000025715	miRNA	Downstream variant
		ENSCAFG00000031179	miRNA	Upstream variant
		ENSCAFG00000046125	miRNA	Upstream variant
		ENSCAFG00000047537	miRNA	Upstream variant
		ENSCAFG00000048199	miRNA	Downstream variant
		ENSCAFG00000050992	miRNA	Upstream variant
		ENSCAFG00000052222	miRNA	Upstream variant
		ENSCAFG00000052519	miRNA	Downstream variant
		ENSCAFG00000053766	miRNA	Upstream variant
		ENSCAFG00000054270	miRNA	Downstream variant
		ENSCAFG00000054713	miRNA	Upstream variant
		ENSCAFG00000055788	miRNA	Downstream variant
		ENSCAFG00000042252	misc_RNA	Downstream variant
		ENSCAFG00000014984	processed_pseudogene	Downstream variant
		ENSCAFG00000018689	pseudogene	Upstream variant
		ENSCAFG00000036065	pseudogene	Downstream variant
		ENSCAFG00000041221	pseudogene	Downstream variant
		ENSCAFG00000041670	pseudogene	Downstream variant
		ENSCAFG00000043307	pseudogene	Downstream variant
		ENSCAFG00000043459	pseudogene	Upstream variant
		ENSCAFG00000044330	pseudogene	Downstream variant
		ENSCAFG00000045749	pseudogene	Upstream variant
		ENSCAFG00000047122	pseudogene	Downstream variant
		ENSCAFG00000048074	pseudogene	Upstream variant
		ENSCAFG00000048084	pseudogene	Downstream variant
		ENSCAFG00000049146	pseudogene	Upstream variant
		ENSCAFG00000049284	pseudogene	Upstream variant
		ENSCAFG00000056490	pseudogene	Downstream variant
		ENSCAFG00000028119	ribozyme	Upstream variant
		ENSCAFG00000021015	rRNA	Downstream variant
		ENSCAFG00000025969	rRNA	Upstream variant
		ENSCAFG00000020883	snoRNA	Upstream variant
		ENSCAFG00000020981	snoRNA	Downstream variant
		ENSCAFG00000026017	snoRNA	Upstream variant
		ENSCAFG00000026053	snoRNA	Downstream variant
		ENSCAFG00000026082	snoRNA	Upstream variant

		ENSCAFG00000026284	snoRNA	Downstream variant
		ENSCAFG00000026468	snoRNA	Downstream variant
		ENSCAFG00000026833	snoRNA	Downstream variant
		ENSCAFG00000027192	snoRNA	Upstream variant
		ENSCAFG00000027480	snoRNA	Upstream variant
		ENSCAFG00000027533	snoRNA	Upstream variant
		ENSCAFG00000028262	snoRNA	Upstream variant
		ENSCAFG00000042265	snoRNA	Upstream variant
		ENSCAFG00000042603	snoRNA	Upstream variant
		ENSCAFG00000047445	snoRNA	Upstream variant
		ENSCAFG00000050039	snoRNA	Downstream variant
		ENSCAFG00000020750	snRNA	Downstream variant
		ENSCAFG00000020815	snRNA	Downstream variant
		ENSCAFG00000020917	snRNA	Upstream variant
		ENSCAFG00000021082	snRNA	Downstream variant
		ENSCAFG00000021107	snRNA	Downstream variant
		ENSCAFG00000021220	snRNA	Downstream variant
		ENSCAFG00000021521	snRNA	Downstream variant
		ENSCAFG00000021530	snRNA	Upstream variant
		ENSCAFG00000022253	snRNA	Upstream variant
		ENSCAFG00000022254	snRNA	Downstream variant
		ENSCAFG00000022493	snRNA	Downstream variant
		ENSCAFG00000022520	snRNA	Upstream variant
		ENSCAFG00000022541	snRNA	Downstream variant
		ENSCAFG00000022899	snRNA	Upstream variant
		ENSCAFG00000025934	snRNA	Upstream variant
		ENSCAFG00000026117	snRNA	Downstream variant
		ENSCAFG00000026930	snRNA	Downstream variant
		ENSCAFG00000026947	snRNA	Upstream variant
		ENSCAFG00000026988	snRNA	Downstream variant
		ENSCAFG00000027033	snRNA	Upstream variant
		ENSCAFG00000027076	snRNA	Upstream variant
		ENSCAFG00000027148	snRNA	Downstream variant
		ENSCAFG00000027239	snRNA	Upstream variant
		ENSCAFG00000027532	snRNA	Upstream variant
		ENSCAFG00000027853	snRNA	Upstream variant
		ENSCAFG00000028004	snRNA	Downstream variant
		ENSCAFG00000028111	snRNA	Upstream variant
		ENSCAFG00000028313	snRNA	Downstream variant
		ENSCAFG00000029409	snRNA	Upstream variant
		ENSCAFG00000043087	snRNA	Downstream variant
		ENSCAFG00000043782	snRNA	Downstream variant
		ENSCAFG00000044164	snRNA	Downstream variant

		ENSCAFG00000045431	snRNA	Upstream variant
		ENSCAFG00000046396	snRNA	Upstream variant
		ENSCAFG00000046457	snRNA	Downstream variant
		ENSCAFG00000047244	snRNA	Upstream variant
		ENSCAFG00000049973	snRNA	Upstream variant
		ENSCAFG00000051000	snRNA	Upstream variant
		ENSCAFG00000053965	snRNA	Upstream variant
		ENSCAFG00000053965	snRNA	Upstream variant
		ENSCAFG00000055381	snRNA	Downstream variant
		ENSCAFG00000055501	snRNA	Upstream variant
		ENSCAFG00000055792	snRNA	Upstream variant
		ENSCAFG00000055888	snRNA	Downstream variant
		ENSCAFG00000056745	snRNA	Upstream variant
		ENSCAFG00000056745	snRNA	Upstream variant
		ENSCAFG00000057912	snRNA	Downstream variant
		ENSCAFG00000059057	snRNA	Downstream variant
		ENSCAFG00000021643	Y_RNA	Downstream variant
		ENSCAFG00000026073	Y_RNA	Upstream variant
	Protein coding	ENSCAFG00000008658	40S ribosomal protein S3a	Downstream variant
		ENSCAFG00000015933	Acylphosphatase-1 (Acylphosphate phosphohydrolase 1)	Upstream variant
		ENSCAFG00000042815	ADP/ATP translocase (ADP, ATP carrier protein)	Upstream variant
		ENSCAFG00000058987	AHNAK nucleoprotein 2	Upstream variant
		ENSCAFG00000054288	BED-type domain-containing protein	Upstream variant
		ENSCAFG00000044557	BPTI/Kunitz inhibitor domain-containing protein	Downstream variant
		ENSCAFG00000044557	BPTI/Kunitz inhibitor domain-containing protein	Downstream variant
		ENSCAFG00000055440	CBM21 domain-containing protein	Downstream variant
		ENSCAFG00000058902	Cornichon family AMPA receptor auxiliary protein 1	Downstream variant
		ENSCAFG00000054560	Cyclin-dependent kinases regulatory subunit	Upstream variant
		ENSCAFG00000041466	EF-hand domain-containing protein	Upstream variant
		ENSCAFG00000042386	Ferritin	Downstream variant
		ENSCAFG00000051032	Glyceraldehyde 3-phosphate dehydrogenase NAD(P) binding domain-containing protein	Missense
		ENSCAFG00000003797	Heat shock cognate 71 kDa protein (EC 3.6.4.10) (Heat shock 70 kDa protein 8)	Upstream variant
		ENSCAFG00000056470	Homeobox domain-containing protein	Upstream variant

		ENSCAFG00000031862	HTH CENPB-type domain-containing protein	Downstream variant
		ENSCAFG00000058699	Ig-like domain-containing protein	Upstream variant
		ENSCAFG00000055775	Ig-like domain-containing protein	Upstream variant
		ENSCAFG00000057398	Ig-like domain-containing protein	Upstream variant
		ENSCAFG00000045242	Ig-like domain-containing protein	Upstream variant
		ENSCAFG00000054905	Ig-like domain-containing protein	Upstream variant
		ENSCAFG00000014478	Ig-like domain-containing protein	Upstream variant
		ENSCAFG00000056839	Ig-like domain-containing protein	Upstream variant
		ENSCAFG00000049482	Ig-like domain-containing protein	Downstream variant
		ENSCAFG00000003811	Ig-like domain-containing protein	Downstream variant
		ENSCAFG00000024810	Ig-like domain-containing protein	Downstream variant
		ENSCAFG00000046664	Ig-like domain-containing protein	Downstream variant
		ENSCAFG00000055018	Ig-like domain-containing protein	Downstream variant
		ENSCAFG00000057117	Ig-like domain-containing protein	Downstream variant
		ENSCAFG00000052806	Ig-like domain-containing protein	Downstream variant
		ENSCAFG00000057504	Ig-like domain-containing protein	Downstream variant
		ENSCAFG00000042586	Ig-like domain-containing protein	Intron variant
		ENSCAFG00000058265	Immunoglobulin C1-set domain-containing protein	Upstream variant
		ENSCAFG00000023933	Immunoglobulin V-set domain-containing protein	Upstream variant
		ENSCAFG00000050263	Immunoglobulin V-set domain-containing protein	Upstream variant
		ENSCAFG00000031359	Immunoglobulin V-set domain-containing protein	Upstream variant
		ENSCAFG00000045360	Immunoglobulin V-set domain-containing protein	Upstream variant
		ENSCAFG00000043031	Immunoglobulin V-set domain-containing protein	Upstream variant
		ENSCAFG00000029837	Immunoglobulin V-set domain-containing protein	Downstream variant
		ENSCAFG00000055372	Immunoglobulin V-set domain-containing protein	Downstream variant
		ENSCAFG00000055600	Immunoglobulin V-set domain-containing protein	5 Prime UTR variant
		ENSCAFG00000041829	Interferon lambda-4-like	Upstream variant
		ENSCAFG00000052759	Interferon lambda-4-like	Downstream variant
		ENSCAFG00000028492	Kisspeptin	Upstream variant

		ENSCAFG00000048255	KRAB domain-containing protein	Downstream variant
		ENSCAFG00000042267	KRAB domain-containing protein	Intron variant
		ENSCAFG00000031109	LSM domain-containing protein	Upstream variant
		ENSCAFG00000051129	MHC class I-like antigen recognition-like domain-containing protein	Downstream variant
		ENSCAFG00000000033	Mitochondrial carrier 2	Upstream variant
		ENSCAFG00000049889	Mitochondrial import inner membrane translocase subunit TIM23	Downstream variant
		ENSCAFG00000047151	Olfactory receptor	synonymous_variant
		ENSCAFG00000050625	PDZ domain-containing protein	Upstream variant
		ENSCAFG00000002084	Peptidyl-prolyl cis-trans isomerase (PPIase) (EC 5.2.1.8)	Downstream variant
		ENSCAFG00000014968	Peptidyl-prolyl cis-trans isomerase (PPIase) (EC 5.2.1.8)	Downstream variant
		ENSCAFG00000002337	Peroxiredoxin-1 (EC 1.11.1.24) (Thioredoxin-dependent peroxiredoxin 1)	Missense
		ENSCAFG00000016551	Prefoldin subunit 4	Upstream variant
		ENSCAFG00000053426	Protein kinase domain-containing protein	Downstream variant
		ENSCAFG00000007678	Ribose-phosphate pyrophosphokinase N-terminal domain-containing protein	Upstream variant
		ENSCAFG00000053722	Ribosomal protein L37a	Upstream variant
		ENSCAFG00000058812	RIC8B	Downstream variant
		ENSCAFG00000041062	RNA-directed DNA polymerase (EC 2.7.7.49)	Downstream variant
		ENSCAFG00000052621	RRM domain-containing protein	Upstream variant
		ENSCAFG00000041560	RRM domain-containing protein	Downstream variant
		ENSCAFG00000043332	Secreted protein	Upstream variant
		ENSCAFG00000042570	Small integral membrane protein 27	Downstream variant
		ENSCAFG00000042783	Small ribosomal subunit protein eS6 (40S ribosomal protein S6)	Downstream variant
		ENSCAFG00000044478	Small ribosomal subunit protein uS15 N-terminal domain-containing protein	Upstream variant
		ENSCAFG00000043184	Superkiller complex protein 8 (WD repeat-containing protein 61)	Upstream variant
		ENSCAFG00000024456	Transcription factor BTF3	Downstream variant
		ENSCAFG00000049094	Uncharacterized protein	Missense
		ENSCAFG00000045033	Uncharacterized protein	Upstream variant
		ENSCAFG00000047905	Uncharacterized protein	Upstream variant
		ENSCAFG00000051647	Uncharacterized protein	Upstream variant
		ENSCAFG00000048629	Uncharacterized protein	Upstream variant
		ENSCAFG00000046775	Uncharacterized protein	Upstream variant

		ENSCAFG00000041808	Uncharacterized protein	Upstream variant
		ENSCAFG00000041820	Uncharacterized protein	Upstream variant
		ENSCAFG00000043214	Uncharacterized protein	Upstream variant
		ENSCAFG00000047606	Uncharacterized protein	Downstream variant
		ENSCAFG00000051605	Uncharacterized protein	Downstream variant
		ENSCAFG00000048629	Uncharacterized protein	Upstream variant
		ENSCAFG00000049104	Uncharacterized protein	Downstream variant
		ENSCAFG00000043988	Uncharacterized protein	Downstream variant
		ENSCAFG00000055022	Uncharacterized protein	Downstream variant
		ENSCAFG00000058607	Uncharacterized protein	Downstream variant
		ENSCAFG00000016645	WASH complex subunit 3 (Coiled-coil domain-containing protein 53)	Upstream variant
		ENSCAFG00000046247	Zinc finger protein 671	Downstream variant
		ENSCAFG00000028550	Zinc finger protein 850-like	Upstream variant

Supplementary table 4. Genes with UnitProt codes automatically excluded by g:Profiler

UnitProt ID	***Genome run***
Wolves	
ABCC7	D/H
ATP5J2	H
BCAS4	D
C17H2ORF81	H
C18H11ORF98	H
C20H19ORF25	H
C20H19ORF53	H
C28H10ORF95	H
C2H1ORF232	H
C8H14ORF180	H
CATSPERD	D
CBD103	H
CBD108	D/H
CBD138	D/H
CBD139	H
CE10	H
CFAP58	D
CMAH	H
COX1	D
GAL3ST2	D
IL17A	D
KLNL	D
LOC100682872	H
LOC100683099	H
LOC100685852	D/H
LOC100686432	D/H
LOC100686484	H
LOC100686511	H
LOC100686580	D/H
LOC100688619	H
LOC100855683	H
LOC102151492	H
LOC102155960	H
LOC102156679	H
LOC106559612	H
LOC111090112	D/H
LOC111090112	D/H
LOC111096443	H
LOC119867218	H
LOC119868927	H
LOC119870529	D/H
LOC119880303	H
LOC119880472	H
LOC119881386	H
LOC474938	H
LOC476445	H
LOC478994	H
LOC479071	H
LOC481841	H
LOC482938	H
LOC484309	H
LOC484341	H
LOC484492	H
LOC487375	H
LOC608975	H
LOC609023	D/H
LOC610304	H
LOC611215	H
LOC611401	H
LOC611446	H
LOC612036	H
MBL1	H
MGAT4F	H
MIRNA	H
MISC_RNA	H
MTA3	D
OR4B1E	H
OR4B1F	H
OR4S7	H
OR5AN2	D/H
OR5AN6	H
OR5AQ1	H
OR5F2	H
OR5J1	H
OR5T1B	H
OR5T2B	H
OR8J2B	H
OR8J2C	H
OR8K71	H
PLAUR	D
RIGI	D
SLBP2	H
SNRNA	D
SPAG11E	H
TEPP	H
TIMP2	D
TMEM141	D
TUBA5	H
YPEL	H
FRDs	
AKR1D1	D
C12H6ORF141	H
C18H11ORF98	H
C1H19ORF18	H
C1H9ORF40	H
C3H15ORF40	H
C4H10ORF105	H
C5H11ORF87	H
C8H14ORF93	H
CATSPERD	D
CDK5R2	D
CRSP2	H
DLA-12	H
DLA88	H
DMBT1	D
ERAP2	D
GAL3ST2	D
KCNJ5	D
LOC100683951	D/H
LOC100685048	H
LOC100686431	H
LOC100687983	H
LOC102152102	H
LOC102152670	H
LOC102153214	H
LOC106558923	H
LOC111093559	H
LOC119866701	H
LOC119866928	H
LOC119880636	H
LOC119881374	H
LOC475563	D/H

LOC476784	H
LOC477391	H
LOC479047	H
LOC479600	H
LOC481603	H
LOC482174	H
LOC482175	H
LOC482176	H
LOC482176	H
LOC482215	H
LOC482217	H
LOC490972	H
LOC607646	D/H
LOC607806	H
LOC607889	H
LOC608513	H
LOC610105	H
LOC610725	H
LOC611691	H
LOC611724	H
OR13C11C	H
OR13C7F	H
OR13C7G	H
OR13C7H	H
OR13C7I	H
OR2AJ7	H
OR6C38	H
OR6C63	H
OR7A23	H
OR7A55	H
OR7A57	D/H
OR7A58	H
OR7A60	H
OR7A63	H
OR7A64	H
PANX	H
PGB1	H
TMED11	H
TNFSF9	D

*D= dog, H= human

Supplementary table 5. Variant Effect Predictor (VEP) results summary wolves.

Results with WorldClim variables			
ENSMBL ID	UnitProt ID or Name, Biotype for non-protein coding	LFMM-Association	RDA-Association
Missense variants			
ENSCAFG00000014554	ACOX3	Bio16	Bio2
ENSCAFG00000011576	ADNP	Bio3, Bio7	Bio2
ENSCAFG00000016566	CC2D1A	Bio16	Bio9
ENSCAFG00000008431	CDIN1	Bio16	Bio8
ENSCAFG00000019209	CFAP74	Bio16	Elev
ENSCAFG00000013224	DDX47	Bio16	Bio8
ENSCAFG00000008507	DNTT	Elev	Bio3
ENSCAFG00000052337	GPR132	Bio16	Bio8
ENSCAFG00000017287	IL12B*	Elev	Bio8
ENSCAFG00000003594	IRF3	Elev	Bio8
ENSCAFG00000031996	IRGQ	Elev	Bio2
ENSCAFG00000014175	ITGA4	Bio16	Bio2
ENSCAFG00000004723	KCNN4	Elev	Bio9
ENSCAFG00000008788	MSH3	Elev	Bio16
ENSCAFG00000047541	OTOGL	Bio16	Bio19
ENSCAFG00000008279	PTPRJ	Bio16	Bio16
ENSCAFG00000015627	SIX4	Elev	Bio16
ENSCAFG00000013789	SSD domain-containing protein*	Elev	Bio19
ENSCAFG00000008348	TM4SF4	Bio16	Bio16
Intron variants			
ENSCAFG00000001786	ACO1	Elev	Bio16
ENSCAFG00000010368	ADIPOR1 KLHL12	Elev	Bio8
ENSCAFG00000019569	AGPAT2	Elev, Grass	Bio8
ENSCAFG00000023699	ANGPTL5	Bio16	Bio9
ENSCAFG00000016484	ARHGEF11	Bio16	Elev
ENSCAFG00000011897	ARV1	Elev	Bio16
ENSCAFG00000019650	CAMTA1	Bio16	Bio16
ENSCAFG00000004607	CCDC122	Elev	Bio16
ENSCAFG00000010700	CCND1	Elev	Bio16
ENSCAFG00000006401	CD36	Bio16	Elev
ENSCAFG00000015131	CEP126	Bio16	Bio9
ENSCAFG00000057002	CEP170B	Bio16	Bio8
ENSCAFG00000007635	CLSTN2	Bio7	Bio3
ENSCAFG00000016325	CNIH3	Bio16	Bio18
ENSCAFG00000012630	CNKSR1	Bio16	Elev
ENSCAFG00000010426	COL4A3	Bio16	Bio8
ENSCAFG00000019174	CPT2	Bio16	Bio2
ENSCAFG00000005886	CRIM1	Elev	Bio16
ENSCAFG00000009271	CRTAC1	Bio16	Bio18
ENSCAFG00000013221	CTNNA3	Bio16	Bio19
ENSCAFG00000000885	DEPTOR	Bio16	Bio8
ENSCAFG00000058269	DKK2	Bio2, Bio3, Bio7	Bio8

ENSCAFG00000000676	DPYS	Elev	Elev
ENSCAFG00000006607	DPYSL3	Elev	Bio16
ENSCAFG00000013220	ELOA RPL11	Elev	Bio16
ENSCAFG00000004755	ETHE1	Elev	Bio9
ENSCAFG00000001308	EXOC4	Bio16	Bio3
ENSCAFG00000002391	EXOSC3	Elev	Bio9
ENSCAFG00000008736	FSIP1	Elev	Bio16
ENSCAFG00000002487	GABBR2	Elev	Bio2
ENSCAFG00000006986	GAPDHS	Elev, Grass	Bio8
ENSCAFG00000004821	GLI2	Bio16	Bio19
ENSCAFG00000011122	GSK3B	Bio16	Bio3
ENSCAFG00000015595	IGSF21	Elev	Bio16
ENSCAFG00000019919	IRF8	Bio16	Bio16
ENSCAFG00000047466	KAZN	Elev	Bio16
ENSCAFG00000019790	KCNC4	Bio16	Bio3
ENSCAFG00000004723	KCNN4	Elev	Bio9
ENSCAFG00000004673	KHK	Bio16	Bio16
ENSCAFG00000004602	LACC1	Elev	Bio16
ENSCAFG00000019560	LCN15	Elev	Bio9
ENSCAFG00000043059	LOC102155960	Bio16	Bio19
ENSCAFG00000006768	LRRC4C	Bio16	Bio3
ENSCAFG00000052675	MCFD2	Bio16	Bio8
ENSCAFG00000052289	MIGA1	Bio16	Bio16
ENSCAFG00000011492	MRPL46	Bio7	Bio16
ENSCAFG00000008716	MTHFD2	Bio16, Elev	Bio16
ENSCAFG00000016062	MYRF	Bio16	Bio3
ENSCAFG00000009026	NECAB1	Bio16	Bio19
ENSCAFG00000006103	NEK11	Bio16	Bio16
ENSCAFG00000005241	NET1	Elev	Bio8
ENSCAFG00000006371	NRG1	Bio16	Bio9
ENSCAFG00000047543	OR5T1B	Bio16	Bio8
ENSCAFG00000013289	PBX1	Bio16	Bio8
ENSCAFG00000006269	PCID2	Bio2, Bio7	Bio2
ENSCAFG00000023022	PDZD2	Bio16	Bio7
ENSCAFG00000015246	PLD1	Bio7	Bio8
ENSCAFG00000010644	PPP3CA	Bio16	Bio7
ENSCAFG00000019531	PWWP3A	Bio16	Bio19
ENSCAFG00000041318	RAB40C	Bio16	Bio7
ENSCAFG00000006503	RBPMS	Bio7	Bio16
ENSCAFG00000005692	RPTOR	Elev	Bio8
ENSCAFG00000008913	RSRC1	Bio16	Bio7
ENSCAFG00000008519	RTKN	Bio16	Bio16
ENSCAFG00000003957	SF3B6	Elev	Bio18
ENSCAFG00000004818	SLC5A6	Bio16	Bio16
ENSCAFG00000009880	SNCA	Elev	Elev
ENSCAFG00000011571	SOX5	Elev	Bio3

ENSCAFG00000020387	ST6GALNAC3	Bio9	Elev
ENSCAFG00000006573	TAFA4	Bio16	Bio16
ENSCAFG00000019293	TLE2	Bio16	Bio8
ENSCAFG00000008348	TM4SF4	Bio16	Bio16
ENSCAFG00000002694	TMC4	Elev	Bio16
ENSCAFG00000044300	TMEM14A	Bio16	Bio16
ENSCAFG00000005012	TMEM163	Bio16	Elev
ENSCAFG00000030694	TMEM38A	Elev	Bio18
ENSCAFG00000014575	TNS1	Bio16	Bio8
ENSCAFG00000017302	TSHR	Elev	Bio8
ENSCAFG00000010287	TSPAN5	Bio3	Bio7
ENSCAFG00000002641	TTC7A	Bio16	Bio8
ENSCAFG00000032209	TYW3	Bio16	Bio8
ENSCAFG00000047764	Uncharacterized protein	Bio16	Elev
ENSCAFG00000000326	UTRN	Elev	Bio18
ENSCAFG00000020031	WWOX	Bio16	Bio18
ENSCAFG00000007373	ZNF143	Bio16	Bio16
Intron variants non protein coding transcripts			
ENSCAFG00000057144	lncRNA	Bio16	Bio19
ENSCAFG00000050322	lncRNA	Elev	Bio19
ENSCAFG00000056919	lncRNA	Bio16	Bio18
ENSCAFG00000045992	lncRNA	Elev	Bio16
ENSCAFG00000057034	lncRNA	Bio16	Bio19
Splice region variant, 5' UTR variant			
ENSCAFG00000013691	PRG4	Elev	Bio9
Splice region variants			
ENSCAFG00000006825	PRR5L	Bio16	Bio3
ENSCAFG00000005138	RPL14	Bio16	Bio3
Splice region variant, synonymous variant			
ENSCAFG00000006986	GAPDHS	Elev, Grass	Bio8
ENSCAFG00000018761	PIEZO2	Bio16	Bio16
Synonymous variants			
ENSCAFG00000003334	ABCA13	Bio16	Elev
ENSCAFG00000018208	ABCC1	Bio16	Elev
ENSCAFG00000018625	C3	Bio16, Elev	Bio18
ENSCAFG00000006401	CD36	Bio16	Elev
ENSCAFG00000017344	CLINT1	Elev	Bio8
ENSCAFG00000019174	CPT2	Bio16	Bio2
ENSCAFG00000020227	DENND1A	Elev	Bio2
ENSCAFG00000000885	DEPTOR	Bio16	Bio8
ENSCAFG00000010131	GBF1	Elev	Bio2
ENSCAFG00000012477	ITGB5	Elev	Bio8
ENSCAFG00000013324	LRP6	Bio3, Bio8, Bio16, Bio19, Hfpi	Bio9
ENSCAFG00000009820	NOS1	Bio16, Bio19	Bio16
ENSCAFG00000002023	RCAN2	Elev	Bio2
ENSCAFG00000015053	SPINDOC	Elev	Bio2

ENSCAFG00000009943	TANK	Bio16	Bio18
ENSCAFG00000019808	TCF25 TUBB3	Bio16	Bio16
ENSCAFG00000008875	TIPARP	Bio16	Bio2
3' UTR variants			
ENSCAFG00000012954	FANK1	Bio16	Bio16
ENSCAFG00000031506	LOC482938	Bio2, Bio3, Bio7	Bio2
ENSCAFG00000005065	LOC484492	Bio16	Bio7
ENSCAFG00000031195	LOC611401	Elev	Bio16
ENSCAFG00000052289	MIGA1	Bio16	Bio16
ENSCAFG00000013495	OLR1	Elev	Bio8
ENSCAFG00000002067	SLC1A1	Elev	Elev
ENSCAFG00000004524	TTYH2	Elev	Elev
ENSCAFG00000001505	UBE2H	Elev	Bio3
ENSCAFG00000013039	ZNF365	Bio16	Bio8
5' UTR variants			
ENSCAFG00000002255	CILK1	Bio16	Bio3
ENSCAFG00000005981	WDR87	Bio16	Bio7
Downstream variants			
ENSCAFG00000019801	5-hydroxyisourate hydrolase (HIU hydrolase) (HIUHase) (EC 3.5.2.17)	Bio16	Bio16
ENSCAFG00000018197	ABCC6	Bio16	Elev
ENSCAFG00000004678	ABHD1	Bio16	Bio16
ENSCAFG00000011513	ABLIM1	Bio16	Bio3
ENSCAFG00000015708	ACTA2	Bio7	Bio2
ENSCAFG00000018565	ACTL9	Bio16	Bio18
ENSCAFG00000010210	ACTR1A CTRN1	Elev	Bio2
ENSCAFG00000005035	ACVR2B	Bio16	Bio16
ENSCAFG00000012986	ADAM12	Bio16	Bio16
ENSCAFG00000006805	ADRA2B	Bio16	Bio8
ENSCAFG00000004537	AGBL5	Bio16	Bio16
ENSCAFG00000001196	AGO2 EIF2C2	Bio16	Bio8
ENSCAFG00000009721	ALCAM	Elev, Grass	Bio2
ENSCAFG00000051627	ANKRD53	Bio16	Bio16
ENSCAFG00000002492	ANKS6	Elev	Bio2
ENSCAFG00000013241	APOLD1	Bio16	Bio8
ENSCAFG00000046427	ASTL	Bio16	Bio8
ENSCAFG00000006951	ATP4A	Elev, Grass	Bio8
ENSCAFG00000019536	ATP5F1D	Bio16	Bio19
ENSCAFG00000030921	ATP5MK LOC102156136	Bio16	Bio2
ENSCAFG00000003442	ATP6V1B1	Bio16	Bio16
ENSCAFG00000008402	AUP1	Bio16	Bio16
ENSCAFG00000009867	BBX	Bio16	Bio2
ENSCAFG00000023887	BCAS4	Bio3, Bio7	Bio2
ENSCAFG00000013352	BCL2L14	Bio3, Bio8, Bio16, Bio19, Hfpi	Bio9
ENSCAFG00000029223	BPTI/Kunitz inhibitor domain-containing protein	Bio16	Bio19
ENSCAFG00000004553	BTBD17	Elev	Elev

ENSCAFG00000053158	C20H19orf25	Bio16	Bio19
ENSCAFG00000051854	C20H19orf53	Bio16	Bio9
ENSCAFG00000048372	C28H10orf95	Elev	Bio2
ENSCAFG00000031172	CADM4	Elev	Bio9
ENSCAFG00000005235	CALML5	Elev	Bio8
ENSCAFG00000019214	CALML6	Bio16	Elev
ENSCAFG00000013830	CAPN1	Elev, Grass	Bio8
ENSCAFG00000019667	CAPN15	Bio16	Bio7
ENSCAFG00000009793	CBLB	Elev, Grass	Bio2
ENSCAFG00000053580	CBX5	Bio16	Bio8
ENSCAFG00000008471	CCDC142	Bio16	Bio16
ENSCAFG00000019621	CCDC78	Bio16	Bio7
ENSCAFG00000058394	CDC42EP2	Elev, Grass	Bio8
ENSCAFG00000000052	CDH19	Bio16	Bio16
ENSCAFG00000013244	CDKN1B	Bio16	Bio8
ENSCAFG00000049719	CENPBD1P	Bio16	Bio16
ENSCAFG00000015131	CEP126	Bio16	Bio9
ENSCAFG00000057002	CEP170B	Bio16	Bio8
ENSCAFG00000014193	CERKL	Bio16	Bio2
ENSCAFG00000008517	CFAP20	Bio16	Elev
ENSCAFG00000030528	CHRAC1	Bio16	Bio8
ENSCAFG00000019618	CIAO3	Bio16	Bio7
ENSCAFG00000019506	CIRBP	Bio16	Bio19
ENSCAFG00000031672	CLBA1	Bio16	Bio8
ENSCAFG00000058021	CLEC1B	Elev	Bio8
ENSCAFG00000025296	CLEC1B	Elev	Bio8
ENSCAFG00000016325	CNIH3	Bio16	Bio18
ENSCAFG00000002699	CNOT3	Elev	Bio16
ENSCAFG00000013071	CNR2	Elev	Bio16
ENSCAFG00000012226	COL6A3	Bio16	Bio19
ENSCAFG00000030168	COPS8	Bio16	Bio19
ENSCAFG00000019921	COX4I1	Bio16	Bio16
ENSCAFG00000002319	CRACDL	Bio2, Bio3, Bio8, Bio16, Elev	Bio3
ENSCAFG00000011809	CRYBA4	Elev	Bio8
ENSCAFG00000012511	CRYBG2	Bio16	Elev
ENSCAFG00000010189	CUEDC2	Elev	Bio2
ENSCAFG00000010350	CYB5R1	Elev	Bio8
ENSCAFG00000009802	CYLD	Bio16	Bio2
ENSCAFG00000032339	CYP2A13	Bio16	Bio7
ENSCAFG00000005052	CYP2B6	Bio16	Bio7
ENSCAFG00000002026	CYP39A1	Elev	Bio2
ENSCAFG00000016664	Cytochrome c oxidase subunit 6C (Cytochrome c oxidase polypeptide VIc)	Bio16, Elev	Elev
ENSCAFG00000019168	CZIB	Bio16	Bio2
ENSCAFG00000013152	D2HGDH	Bio16	Bio3
ENSCAFG00000016070	DAGLA	Bio16	Bio3

ENSCAFG00000019803	DBNDD1	Bio16	Bio16
ENSCAFG00000000675	DCSTAMP	Elev	Elev
ENSCAFG00000003882	DDO	Bio16	Bio8
ENSCAFG00000030438	DEFB123	Elev	Bio16
ENSCAFG00000011510	DET1	Bio7	Bio16
ENSCAFG00000017701	DICER1	Bio16	Bio2
ENSCAFG00000006766	DLC1	Bio16	Bio7
ENSCAFG00000052441	DNAAF2 KTU	Bio16	Bio18
ENSCAFG00000013654	DPF2	Elev, Grass	Bio8
ENSCAFG00000030079	DPM1	Bio3, Bio7	Bio2
ENSCAFG00000008425	DQX1	Bio16	Bio16
ENSCAFG00000011948	DTX3L	Bio19	Bio2
ENSCAFG00000023317	DUSP2	Bio16	Bio8
ENSCAFG00000013267	EBF3	Bio16	Bio2
ENSCAFG00000019544	EDF1	Elev	Bio9
ENSCAFG00000058691	EFNA2	Bio16	Bio19
ENSCAFG00000019575	EGFL7	Elev, Grass	Bio8
ENSCAFG00000013220	ELOA RPL11	Elev	Bio16
ENSCAFG00000004662	EMILIN1	Bio16	Bio16
ENSCAFG00000005128	ENTPD3	Bio16	Bio3
ENSCAFG00000006563	EOGT AER61	Bio16	Bio16
ENSCAFG00000001282	ETF1	Bio16, Elev	Bio16
ENSCAFG00000016469	ETV3	Bio16	Elev
ENSCAFG00000016474	ETV3L	Bio16	Elev
ENSCAFG00000006258	F10	Bio2, Bio7	Bio2
ENSCAFG00000055338	F7	Bio2, Bio7	Bio2
ENSCAFG00000025083	FADS1	Bio16	Bio3
ENSCAFG00000012651	FAM110D	Bio16	Elev
ENSCAFG00000055859	FAM174C	Bio16	Bio19
ENSCAFG00000030215	FAM228B	Elev	Bio18
ENSCAFG00000057006	FAM53C	Bio16, Elev	Bio16
ENSCAFG00000055802	FAM89A	Elev	Bio16
ENSCAFG00000019623	FBXL16	Bio16	Bio7
ENSCAFG00000019539	FBXW5	Elev	Bio9
ENSCAFG00000005904	FEZ2	Elev	Bio16
ENSCAFG00000007027	FFAR2	Elev, Grass	Bio8
ENSCAFG00000010693	FGF19	Elev	Bio16
ENSCAFG00000029983	FGF4	Elev	Bio16
ENSCAFG00000053636	FPR2	Bio16	Bio19
ENSCAFG00000002386	FRMPD1	Elev	Bio9
ENSCAFG00000020361	FUBP1	Bio16	Bio16
ENSCAFG00000013073	FUCA1	Elev	Bio16
ENSCAFG00000051120	GAL3ST2	Bio16	Bio3
ENSCAFG00000013155	GALE	Elev	Bio16
ENSCAFG00000052535	GAMT	Bio16	Bio19
ENSCAFG00000006986	GAPDHS	Elev, Grass	Bio8

ENSCAFG00000049586	Glyceraldehyde 3-phosphate dehydrogenase NAD(P) binding domain-containing protein	Bio16	Bio2
ENSCAFG00000050155	GOLGA7B	Bio16	Bio18
ENSCAFG00000018620	GPR108	Bio16, Elev	Bio18
ENSCAFG00000016631	GPR179	Bio16	Bio8
ENSCAFG00000013215	GPRC5D	Bio16	Bio8
ENSCAFG00000006334	GRTP1	Bio2, Bio7	Bio2
ENSCAFG00000006489	GTF2E2	Bio7	Bio16
ENSCAFG00000051423	Heterogeneous nuclear ribonucleoprotein A1/A2 C-terminal domain-containing protein	Elev	Bio8
ENSCAFG00000049617	HIG1 domain-containing protein	Elev	Bio16
ENSCAFG00000009791	HIVEP1	Bio16	Bio16
ENSCAFG00000013085	HMGCL	Elev	Bio16
ENSCAFG00000001293	HSPA9	Bio16, Elev	Bio16
ENSCAFG00000011983	HSPBAP1	Bio19	Bio2
ENSCAFG00000014546	HTRA3	Bio16	Bio2
ENSCAFG00000012880	ICOS AILIM	Bio2, Bio3, Bio7	Elev
ENSCAFG00000030509	Ig-like domain-containing protein	Elev	Bio16
ENSCAFG00000047890	Ig-like domain-containing protein	Elev	Bio16
ENSCAFG00000013088	ING5	Bio16	Bio3
ENSCAFG00000007259	IPO7	Bio16	Bio16
ENSCAFG00000010397	IRS1	Bio16	Bio8
ENSCAFG00000004019	ITSN2	Elev	Bio18
ENSCAFG00000018401	JAG2	Bio16	Bio8
ENSCAFG00000019628	JMJD8	Bio16	Bio7
ENSCAFG00000055934	KCNIP3	Bio16	Bio8
ENSCAFG00000006920	KCNIP3 FAHD2A	Bio16	Bio8
ENSCAFG00000001233	KDM3B	Bio16, Elev	Bio16
ENSCAFG00000004673	KHK	Bio16	Bio16
ENSCAFG00000056057	KRAB domain-containing protein	Bio2, Bio3, Bio7	Bio2
ENSCAFG00000044246	Large ribosomal subunit protein uL15m (39S ribosomal protein L15, mitochondrial)	Bio16	Bio3
ENSCAFG00000029526	LBX2	Bio16	Bio16
ENSCAFG00000030312	LCN10	Elev	Bio9
ENSCAFG00000006031	LIG4	Bio16	Bio3
ENSCAFG00000051303	LIPF	Bio7	Bio2
ENSCAFG00000058548	LIPK	Bio7	Bio2
ENSCAFG00000047078	lncRNA	Bio16	Bio16
ENSCAFG00000053379	lncRNA	Bio16	Bio8
ENSCAFG00000056657	lncRNA	Bio16	Bio2
ENSCAFG00000051243	lncRNA	Bio16	Bio9
ENSCAFG00000055193	lncRNA	Bio16	Bio9
ENSCAFG00000053903	lncRNA	Bio16	Bio9
ENSCAFG00000054699	lncRNA	Elev	Bio18
ENSCAFG00000055143	lncRNA	Bio16	Bio7
ENSCAFG00000053664	lncRNA	Bio16	Bio7
ENSCAFG00000058369	lncRNA	Bio16	Elev
ENSCAFG00000050507	lncRNA	Bio16	Bio3

ENSCAFG00000051710	lncRNA	Bio16	Bio3
ENSCAFG00000054112	lncRNA	Bio16	Elev
ENSCAFG00000046977	lncRNA	Bio16	Bio2
ENSCAFG00000055008	lncRNA	Bio16	Bio8
ENSCAFG00000051702	lncRNA	Bio16	Bio8
ENSCAFG00000054665	lncRNA	Bio16	Bio19
ENSCAFG00000056311	lncRNA	Elev	Bio8
ENSCAFG00000052692	lncRNA	Bio16	Bio19
ENSCAFG00000057918	lncRNA	Bio16	Bio19
ENSCAFG00000054624	lncRNA	Bio16	Bio19
ENSCAFG00000057144	lncRNA	Bio16	Bio19
ENSCAFG00000050463	lncRNA	Bio16	Bio19
ENSCAFG00000051133	lncRNA	Bio16	Bio2
ENSCAFG00000055658	lncRNA	Bio16	Bio16
ENSCAFG00000053621	lncRNA	Bio16	Bio3
ENSCAFG00000054794	lncRNA	Bio16	Bio3
ENSCAFG00000053289	lncRNA	Bio16	Bio3
ENSCAFG00000054191	lncRNA	Bio16	Bio3
ENSCAFG00000045810	lncRNA	Elev	Bio16
ENSCAFG00000053081	lncRNA	Elev, Grass	Bio8
ENSCAFG00000058449	lncRNA	Elev	Bio19
ENSCAFG00000056638	lncRNA	Elev	Bio19
ENSCAFG00000051217	lncRNA	Elev	Trees
ENSCAFG00000059330	lncRNA	Elev	Trees
ENSCAFG00000057793	lncRNA	Elev	Trees
ENSCAFG00000057844	lncRNA	Bio2, Bio3, Bio8, Bio16, Elev	Bio3
ENSCAFG00000052488	lncRNA	Bio2, Bio3, Bio8, Bio16, Elev	Bio3
ENSCAFG00000052910	lncRNA	Elev	Bio9
ENSCAFG00000056622	lncRNA	Bio16	Bio3
ENSCAFG00000057994	lncRNA	Bio16	Bio8
ENSCAFG00000056459	lncRNA	Bio16	Bio8
ENSCAFG00000053844	lncRNA	Bio16	Bio8
ENSCAFG00000059104	lncRNA	Bio16	Bio8
ENSCAFG00000053622	lncRNA	Bio16	Bio8
ENSCAFG00000053460	lncRNA	Bio16	Bio7
ENSCAFG00000033391	lncRNA	Elev	Bio16
ENSCAFG00000050831	lncRNA	Bio2, Bio3, Bio7	Bio2
ENSCAFG00000056407	lncRNA	Bio16	Bio8
ENSCAFG00000053898	lncRNA	Bio16	Bio16
ENSCAFG00000058425	lncRNA	Bio16, Elev	Bio16
ENSCAFG00000055129	lncRNA	Bio16	Bio3
ENSCAFG00000056110	lncRNA	Bio16	Bio3
ENSCAFG00000054255	lncRNA	Elev	Bio2
ENSCAFG00000055809	lncRNA	Bio16	Bio3
ENSCAFG00000051549	lncRNA	Elev	Elev

ENSCAFG00000058162	lncRNA	Bio16	Elev
ENSCAFG00000052140	lncRNA	Bio16	Bio18
ENSCAFG00000055266	lncRNA	Bio16	Bio18
ENSCAFG00000058841	lncRNA	Bio16	Bio18
ENSCAFG00000051320	lncRNA	Bio16	Bio9
ENSCAFG00000057211	lncRNA	Bio16	Bio16
ENSCAFG00000050221	lncRNA	Bio16	Bio3
ENSCAFG00000054805	lncRNA	Bio16	Bio3
ENSCAFG00000057810	lncRNA	Bio16	Bio2
ENSCAFG00000056063	lncRNA	Bio16	Bio8
ENSCAFG00000056815	lncRNA	Bio16	Bio8
ENSCAFG00000057067	lncRNA	Bio16	Bio8
ENSCAFG00000054579	lncRNA	Bio16	Bio8
ENSCAFG00000051764	lncRNA	Bio16	Bio8
ENSCAFG00000051181	lncRNA	Elev	Bio8
ENSCAFG00000051649	lncRNA	Bio16	Bio3
ENSCAFG00000055215	lncRNA	Bio16	Bio16
ENSCAFG00000051554	lncRNA	Elev	Bio7
ENSCAFG00000059258	lncRNA	Bio16	Bio8
ENSCAFG00000055734	lncRNA	Bio16	Bio18
ENSCAFG00000050952	lncRNA	Bio16	Bio2
ENSCAFG00000058729	lncRNA	Bio16	Bio2
ENSCAFG00000051191	lncRNA	Bio16	Bio16
ENSCAFG00000042352	lncRNA	Bio16	Bio19
ENSCAFG00000050641	lncRNA	Bio16	Bio19
ENSCAFG00000054923	lncRNA	Bio16	Bio19
ENSCAFG00000055945	lncRNA	Bio16	Bio19
ENSCAFG00000053964	lncRNA	Bio16	Bio2
ENSCAFG00000059324	lncRNA	Bio16	Bio3
ENSCAFG00000051123	lncRNA	Elev	Bio8
ENSCAFG00000057649	lncRNA	Bio16	Bio8
ENSCAFG00000010333	LOC100682872	Elev	Bio8
ENSCAFG00000024605	LOC100685852	Bio16	Bio19
ENSCAFG00000055790	LOC100686580	Elev	Bio16
ENSCAFG00000041880	LOC100855683	Bio16	Bio18
ENSCAFG00000030988	LOC102151492	Elev	Bio8
ENSCAFG00000043059	LOC102155960	Bio16	Bio19
ENSCAFG00000030354	LOC102156679	Bio16	Bio7
ENSCAFG00000006940	LOC111090112	Elev, Grass	Bio8
ENSCAFG00000010648	LOC119867218	Bio16	Bio19
ENSCAFG00000003954	LOC119880472	Elev	Bio18
ENSCAFG00000031682	LOC474938	Bio16	Bio16
ENSCAFG00000004727	LOC476445	Elev	Bio2
ENSCAFG00000056722	LOC481841	Bio16	Bio16
ENSCAFG00000024019	LOC484341	Bio16	Bio19
ENSCAFG00000013249	LOC487375	Elev	Bio16

ENSCAFG00000024271	LOC610304	Bio16	Bio3
ENSCAFG00000028453	LOC611446	Elev	Bio16
ENSCAFG00000029947	LOC612036	Elev, Grass	Bio8
ENSCAFG00000008358	LOXL3	Bio16	Bio16
ENSCAFG00000000677	LRP12	Elev	Elev
ENSCAFG00000019142	LRP8	Bio16	Bio2
ENSCAFG00000016502	LRRC71	Bio16	Elev
ENSCAFG00000045584	LSM11	Elev	Bio8
ENSCAFG00000010698	LTO1	Elev	Bio16
ENSCAFG00000004720	LYPD5	Elev	Bio9
ENSCAFG00000041728	Mago homolog B, exon junction complex subunit	Bio16	Bio18
ENSCAFG00000057662	MAGOH	Bio16	Bio2
ENSCAFG00000003980	MAP3K8	Bio16	Bio7
ENSCAFG00000001378	MAPK14	Bio16	Bio3
ENSCAFG00000002689	MBOAT7	Elev	Bio16
ENSCAFG00000006240	MCF2L	Bio2, Bio7	Bio2
ENSCAFG00000019949	MEAK7	Elev	Bio8
ENSCAFG00000002173	Methylosome subunit pICln	Elev	Bio18
ENSCAFG00000019655	METTL26	Bio16	Bio7
ENSCAFG00000010337	MGAT4F	Elev	Bio8
ENSCAFG00000013263	MGMT	Bio16	Bio2
ENSCAFG00000019534	MIDN	Bio16	Bio19
ENSCAFG00000052289	MIGA1	Bio16	Bio16
ENSCAFG00000051464	miRNA	Bio16	Bio16
ENSCAFG00000059461	miRNA	Bio7	Bio16
ENSCAFG00000055989	miRNA	Bio16	Bio16
ENSCAFG00000020620	miRNA	Elev, Grass	Bio8
ENSCAFG00000057461	miRNA	Bio16, Elev	Bio16
ENSCAFG00000058385	miRNA	Elev	Bio16
ENSCAFG00000050653	miRNA	Bio16	Bio2
ENSCAFG00000054697	miRNA	Bio16	Bio3
ENSCAFG00000020580	miRNA	Bio16	Bio9
ENSCAFG00000022793	miRNA	Bio16	Bio9
ENSCAFG00000054483	miRNA	Bio16, Elev	Bio18
ENSCAFG00000042237	misc_RNA	Bio16	Bio7
ENSCAFG00000005106	MOBP	Bio16	Bio16
ENSCAFG00000008478	MOGS	Bio16	Bio16
ENSCAFG00000016615	MRI1	Bio16	Bio9
ENSCAFG00000011492	MRPL46	Bio7	Bio16
ENSCAFG00000032643	MRRF	Elev	Bio9
ENSCAFG00000012525	MUC13	Elev	Bio8
ENSCAFG00000016285	MYZAP POLR2M	Bio3, Bio8, Bio16, Bio19	Bio2
ENSCAFG00000016592	NANOS3	Bio16	Bio9
ENSCAFG00000018741	NAPG	Bio16	Bio16
ENSCAFG00000002703	NDUFA3	Elev	Bio16
ENSCAFG00000053123	NDUFB6	Elev	Bio16

ENSCAFG00000006103	NEK11	Bio16	Bio16
ENSCAFG00000045576	NEU4	Bio16	Bio3
ENSCAFG00000030927	NFE2	Bio16	Bio8
ENSCAFG00000041922	NFILZ	Bio16	Bio18
ENSCAFG00000032733	NHL repeat containing 4	Bio16	Bio7
ENSCAFG00000011094	NR1I2	Bio16	Bio3
ENSCAFG00000011176	NR5A2	Bio16, Grass	Elev
ENSCAFG00000001507	NRF1	Elev	Bio3
ENSCAFG00000018404	NUDT14	Bio16	Bio8
ENSCAFG00000013495	OLR1	Elev	Bio8
ENSCAFG00000008530	OPALIN	Elev	Bio3
ENSCAFG00000020259	OR1L8 LOC119880303	Elev	Bio9
ENSCAFG00000008267	OR4S7	Bio16	Bio16
ENSCAFG00000008094	OR5AQ1	Bio16	Bio8
ENSCAFG00000008097	OR5AS1	Bio16	Bio8
ENSCAFG00000053286	OR5J1	Bio16	Bio8
ENSCAFG00000046319	OR5T1	Bio16	Bio8
ENSCAFG00000057825	OR5T2B	Bio16	Bio8
ENSCAFG00000008430	PAG1	Bio16	Bio16
ENSCAFG00000056693	PARD6B	Bio3, Bio7	Bio2
ENSCAFG00000011953	PARP15 PARP14	Bio19	Bio2
ENSCAFG00000013289	PBX1	Bio16	Bio8
ENSCAFG00000008457	PCGF1	Bio16	Bio16
ENSCAFG00000000393	PCMT1	Bio16	Bio2
ENSCAFG00000019516	PCSK4	Bio16	Bio19
ENSCAFG00000057857	PDIK1L	Bio16	Elev
ENSCAFG00000016515	PEAR1	Bio16	Elev
ENSCAFG00000052816	Peptidase S26 domain-containing protein	Bio16	Bio18
ENSCAFG00000019662	PIGQ	Bio16	Bio7
ENSCAFG00000009687	PIGT	Bio16	Bio19
ENSCAFG00000003977	PIK3CG	Bio16	Bio9
ENSCAFG00000004742	PINLYP	Elev	Bio2
ENSCAFG00000004730	PLAUR	Elev	Bio9
ENSCAFG00000059374	PLD4	Bio16	Bio8
ENSCAFG00000019183	PODN	Bio16	Bio2
ENSCAFG00000016521	PODNL1	Bio16	Bio9
ENSCAFG00000013677	POLA2	Elev, Grass	Bio8
ENSCAFG00000014201	POLE2	Bio16	Bio18
ENSCAFG00000002377	POLR1E	Elev	Bio9
ENSCAFG00000044002	processed_pseudogene	Bio7	Bio16
ENSCAFG00000045752	processed_pseudogene	Elev	Bio16
ENSCAFG00000008535	Proteasome 20S subunit alpha 6	Bio2, Bio3, Bio7	Bio2
ENSCAFG00000001382	protein_coding	Bio16	Bio3
ENSCAFG00000002700	PRPF31	Elev	Bio16
ENSCAFG00000003611	PRR12	Elev	Bio8
ENSCAFG00000019663	PRR35	Bio16	Bio7

ENSCAFG00000024500	PRRG2	Elev	Bio8
ENSCAFG00000010170	PSD	Elev	Bio2
ENSCAFG00000047701	pseudogene	Bio16	Bio16
ENSCAFG00000044050	pseudogene	Bio16	Bio2
ENSCAFG00000049571	pseudogene	Bio16	Bio8
ENSCAFG00000005045	pseudogene	Bio16	Bio16
ENSCAFG00000019670	RAB11FIP3	Bio16	Bio7
ENSCAFG00000010267	RAP1GDS1	Bio3	Bio7
ENSCAFG00000030228	RBM11	Bio16	Bio9
ENSCAFG00000006503	RBPMS	Bio7	Bio16
ENSCAFG00000003620	RCN3	Elev	Bio8
ENSCAFG00000014777	RCOR2	Elev	Bio2
ENSCAFG00000001244	REEP2	Bio16, Elev	Bio16
ENSCAFG00000006991	REM1	Elev	Bio16
ENSCAFG00000016511	RFX1	Bio16	Bio9
ENSCAFG00000001813	RFX4	Elev	Trees
ENSCAFG00000001807	RIGI	Elev	Bio16
ENSCAFG00000048970	RPL36AL	Bio16	Bio18
ENSCAFG00000056257	RPL38	Elev	Elev
ENSCAFG00000049043	RPS29	Bio16	Bio18
ENSCAFG00000048437	RRM domain-containing protein	Bio16	Bio9
ENSCAFG00000015064	RTN3	Elev	Bio2
ENSCAFG00000013188	RTP5	Bio16	Bio3
ENSCAFG00000050277	SC5D	Bio16	Bio3
ENSCAFG00000005051	SCN5A	Bio16	Bio16
ENSCAFG00000006381	SEMA3C	Bio16	Elev
ENSCAFG00000018602	SH2D3A	Bio16, Elev	Bio18
ENSCAFG00000002550	SIX2	Bio16	Bio3
ENSCAFG00000002547	SIX3	Bio16	Bio3
ENSCAFG00000032438	SLBP2	Bio16, Elev	Bio16
ENSCAFG00000010641	SLC17A2	Bio16	Bio16
ENSCAFG00000024892	SLC17A3	Bio16	Bio16
ENSCAFG00000010583	SLC17A4	Bio16	Bio16
ENSCAFG00000003892	SLC22A16	Bio16	Bio8
ENSCAFG00000028559	SLC25A45	Elev, Grass	Bio8
ENSCAFG00000056205	SLC25A51	Elev	Bio9
ENSCAFG00000004828	SLC30A3	Bio16	Bio16
ENSCAFG00000008684	SLC4A5	Bio16	Bio16
ENSCAFG00000004818	SLC5A6	Bio16	Bio16
ENSCAFG00000019791	SLC6A17	Bio16	Bio3
ENSCAFG00000006603	SMUG1	Bio16	Bio8
ENSCAFG00000025856	snoRNA	Elev	Bio9
ENSCAFG00000025938	snoRNA	Elev	Bio9
ENSCAFG00000026458	snoRNA	Bio16, Elev	Bio16
ENSCAFG00000028347	snoRNA	Bio16, Elev	Bio16
ENSCAFG00000046646	snoRNA	Elev	Bio3

ENSCAFG00000021670	snoRNA	Bio16	Bio16
ENSCAFG00000047210	snoRNA	Bio16	Bio3
ENSCAFG00000052908	snRNA	Elev	Bio18
ENSCAFG00000027380	snRNA	Elev	Bio2
ENSCAFG00000028074	snRNA	Bio16	Bio2
ENSCAFG00000048739	snRNA	Bio16	Elev
ENSCAFG00000025744	snRNA	Bio16	Bio18
ENSCAFG00000021082	snRNA	Elev	Bio16
ENSCAFG00000025943	snRNA	Bio16	Bio3
ENSCAFG00000027282	snRNA	Elev	Elev
ENSCAFG00000044904	snRNA	Bio16	Bio16
ENSCAFG00000028044	snRNA	Bio16	Bio8
ENSCAFG00000058981	snRNA	Elev	Elev
ENSCAFG00000020917	snRNA	Bio16, Elev	Bio18
ENSCAFG00000025934	snRNA	Bio16, Elev	Bio18
ENSCAFG00000053823	snRNA	Bio16	Bio8
ENSCAFG00000042682	snRNA	Bio2, Bio7	Bio2
ENSCAFG00000025914	snRNA	Bio16	Bio8
ENSCAFG00000043372	SNRPC	Elev	Bio7
ENSCAFG00000002820	SPACA6	Bio16	Bio19
ENSCAFG00000002072	SPATA6L	Elev	Elev
ENSCAFG00000019811	SPIRE2	Bio16	Bio16
ENSCAFG00000031082	SPMIP8 TEPP	Bio16	Elev
ENSCAFG00000006790	STARD7	Bio16	Bio8
ENSCAFG00000054378	STK32A	Elev	Bio16
ENSCAFG00000000665	TAGAP	Bio16	Bio9
ENSCAFG00000009976	TBX15	Bio16	Bio2
ENSCAFG00000030192	TCF23	Bio16	Bio16
ENSCAFG00000019808	TCF25 TUBB3	Bio16	Bio16
ENSCAFG00000008728	TET3	Bio16, Elev	Bio16
ENSCAFG00000052969	TEX261	Bio16	Bio16
ENSCAFG00000002701	TFPT RPS9	Elev	Bio16
ENSCAFG00000008704	THBS1	Elev	Bio16
ENSCAFG00000017348	THG1L	Elev	Bio8
ENSCAFG00000043934	Thymosin beta-10	Bio16	Elev
ENSCAFG00000019300	TLE family member 6, subcortical maternal complex member	Bio16	Bio8
ENSCAFG00000057295	TLE2 TLE5	Bio16	Bio8
ENSCAFG00000008621	TLL2	Elev	Bio3
ENSCAFG00000028849	TMEM141	Elev	Bio9
ENSCAFG00000008669	TMEM144	Bio16	Bio7
ENSCAFG00000057782	TMEM179	Bio16	Bio3
ENSCAFG00000048610	TMEM258	Bio16	Bio3
ENSCAFG00000001809	TOPORS	Elev	Bio16
ENSCAFG00000006577	TPM3	Bio16	Bio16
ENSCAFG00000013717	TPR	Elev	Bio9
ENSCAFG00000006813	TRAF6	Bio16	Bio3

ENSCAFG00000055493	Transcription factor MafB	Bio16	Bio7
ENSCAFG00000001591	TREM1	Bio16	Bio7
ENSCAFG00000028527	TRMT10B	Elev	Bio9
ENSCAFG00000011936	TRPM8	Bio2, Bio3, Bio7	Bio19
ENSCAFG00000003555	TSKS	Elev	Bio8
ENSCAFG00000005087	TTC21A	Bio16	Bio8
ENSCAFG00000009579	TUBGCP5	Bio16	Bio16
ENSCAFG00000010760	Tubulin/FtsZ GTPase domain-containing protein	Bio16	Bio16
ENSCAFG00000017611	UACA C3VS	Bio3	Elev
ENSCAFG00000017291	UBLCP1	Bio16	Bio7
ENSCAFG00000012551	UBXN11	Bio16	Elev
ENSCAFG00000005250	UCN3	Elev	Bio8
ENSCAFG00000012401	UMPS	Elev	Bio8
ENSCAFG00000045797	Uncharacterized protein	Bio16	Bio7
ENSCAFG00000056327	Uncharacterized protein	Bio16	Bio7
ENSCAFG00000044731	Uncharacterized protein	Bio16	Bio3
ENSCAFG00000046775	Uncharacterized protein	Elev	Elev
ENSCAFG00000041772	Uncharacterized protein	Bio16	Bio3
ENSCAFG00000049014	Uncharacterized protein	Bio16	Bio16
ENSCAFG00000044462	Uncharacterized protein	Elev	Elev
ENSCAFG00000045878	Uncharacterized protein	Elev	Bio8
ENSCAFG00000032173	UPP1	Bio16	Elev
ENSCAFG00000008556	USB1	Bio16	Elev
ENSCAFG00000050043	Vps72/YL1 N-terminal domain-containing protein	Bio16	Bio8
ENSCAFG00000042312	V-SNARE coiled-coil homology domain-containing protein	Bio16	Bio16
ENSCAFG00000009723	WAP four-disulfide core domain 11	Bio16	Bio19
ENSCAFG00000019626	WDR24	Bio16	Bio7
ENSCAFG00000005066	WDR48	Bio16	Bio8
ENSCAFG00000008561	WDR54	Bio16	Bio16
ENSCAFG00000018675	WDR70	Bio16	Elev
ENSCAFG00000019956	WFDC1	Elev	Bio8
ENSCAFG00000057789	WFDC10B	Bio16	Bio19
ENSCAFG00000009696	WFDC2	Bio16	Bio19
ENSCAFG00000008356	WWTR1	Bio16	Bio16
ENSCAFG00000028862	XIRP1	Bio16	Bio8
ENSCAFG00000005028	XYLB	Bio16	Bio16
ENSCAFG00000016626	YJU2B	Bio16	Bio9
ENSCAFG00000007373	ZNF143	Bio16	Bio16
ENSCAFG00000011151	ZNF281	Bio16, Grass	Elev
ENSCAFG00000004738	ZNF428	Elev	Bio2
ENSCAFG00000053547	ZNF575	Elev	Bio2
ENSCAFG00000016604	ZSWIM4	Bio16	Bio9
Upstream variants			
ENSCAFG00000008026	ABCC13	Bio16	Bio9
ENSCAFG00000051012	ABHD13	Bio16	Bio3
ENSCAFG00000018557	ADAMTS10	Bio16	Bio18

ENSCAFG00000010395	ADAMTS16	Bio16	Bio19
ENSCAFG00000023431	ADGRB3	Bio16	Bio9
ENSCAFG00000010368	ADIPOR1 KLHL12	Elev	Bio8
ENSCAFG00000028797	ADM5	Elev	Bio8
ENSCAFG00000011348	AFAP1L2	Bio16	Bio3
ENSCAFG00000058987	AHNAK nucleoprotein 2	Bio16	Bio8
ENSCAFG00000003445	Ankyrin repeat domain 7	Elev	Bio2
ENSCAFG00000055214	ANTKMT	Bio16	Bio7
ENSCAFG00000003541	AP2A1	Elev	Bio8
ENSCAFG00000019523	APC2	Bio16	Bio19
ENSCAFG00000016484	ARHGEF11	Bio16	Elev
ENSCAFG00000007197	ASPH	Bio16	Bio3
ENSCAFG00000019952	ATP2C2	Elev	Bio8
ENSCAFG00000052091	ATRAID	Bio16	Bio16
ENSCAFG00000053130	BCL2L12	Elev	Bio8
ENSCAFG00000047267	Beta-defensin 130-like	Bio2, Bio3, Bio7	Bio2
ENSCAFG00000008477	BLNK	Elev	Bio3
ENSCAFG00000016077	BMPR1A	Bio16	Bio19
ENSCAFG00000008739	BOLA3	Bio16, Elev	Bio16
ENSCAFG00000042729	BRME1	Bio16	Bio9
ENSCAFG00000001386	BRPF3	Bio16	Bio3
ENSCAFG00000008564	C17H2orf81	Bio16	Bio16
ENSCAFG00000038910	C2H1orf232	Bio16	Elev
ENSCAFG00000018327	C8H14orf180	Bio16	Bio3
ENSCAFG00000031172	CADM4	Elev	Bio9
ENSCAFG00000010536	CARMIL1	Bio16	Bio19
ENSCAFG00000044662	Catsper channel auxiliary subunit zeta	Bio16	Bio9
ENSCAFG00000012615	CATSPER4	Bio16	Elev
ENSCAFG00000056510	CBARP	Bio16	Bio19
ENSCAFG00000043922	CBD108	Bio2, Bio3, Bio7	Bio2
ENSCAFG00000006990	CBD122 DEFB122	Elev	Bio16
ENSCAFG00000030308	CD70	Bio16, Elev	Bio18
ENSCAFG00000001220	CDC25C	Bio16, Elev	Bio16
ENSCAFG00000002081	CDC37L1	Elev	Elev
ENSCAFG00000018398	CDCA4	Bio16	Bio8
ENSCAFG00000009714	CE10	Bio16	Bio19
ENSCAFG00000031326	CEBPB	Elev	Bio16
ENSCAFG00000017293	CEP128	Elev	Bio8
ENSCAFG00000012581	CEP85	Bio16	Elev
ENSCAFG00000009316	CES5A	Bio16	Bio3
ENSCAFG00000015127	CFAP300	Bio16	Bio9
ENSCAFG00000010560	CFAP43	Bio16	Bio2
ENSCAFG00000010608	CFAP58	Bio16	Bio2
ENSCAFG00000004676	CGREF1	Bio16	Bio16
ENSCAFG00000002255	CILK1	Bio16	Bio3
ENSCAFG00000013504	CLEC7A CLEC1A	Elev	Bio8

ENSCAFG00000050981	CMAH	Bio16	Bio19
ENSCAFG00000008588	CNGB1	Bio16	Elev
ENSCAFG00000012630	CNKSR1	Bio16	Elev
ENSCAFG00000010423	COL4A4	Bio16	Bio8
ENSCAFG00000007568	CPA6	Bio3	Bio8
ENSCAFG00000003568	CPT1C	Elev	Bio8
ENSCAFG00000011779	CRYBB1	Elev	Bio8
ENSCAFG00000020404	CRYZ	Bio16	Bio8
ENSCAFG00000005088	CSRNP1	Bio16	Bio8
ENSCAFG00000006290	CUL4A	Bio2, Bio7	Bio2
ENSCAFG00000009720	CYFIP1	Bio16	Bio16
ENSCAFG00000031823	CYP2A7	Bio16	Bio7
ENSCAFG00000023821	CYP2F1	Bio16	Bio7
ENSCAFG00000019528	DAZAP1	Bio16	Bio19
ENSCAFG00000002394	DCAF10	Elev	Bio9
ENSCAFG00000016398	DCAF15	Bio16	Bio9
ENSCAFG00000008608	DCTN1	Bio16	Bio16
ENSCAFG00000019806	DEF8	Bio16	Bio16
ENSCAFG00000048458	DEFB109B	Bio2, Bio3, Bio7	Bio2
ENSCAFG00000052287	DEFB119 CBD119	Elev	Bio16
ENSCAFG00000025021	DEFB121	Elev	Bio16
ENSCAFG00000024871	DEFB124	Elev	Bio16
ENSCAFG00000019566	DIPK1B	Elev, Grass	Bio8
ENSCAFG00000007024	DMKN	Elev, Grass	Bio8
ENSCAFG00000004532	DNAI2	Elev	Elev
ENSCAFG00000006607	DPYSL3	Elev	Bio16
ENSCAFG00000000882	DSCC1	Bio16	Bio8
ENSCAFG00000009794	EDN1	Bio16	Bio16
ENSCAFG00000024733	EF-hand domain-containing protein	Elev	Bio8
ENSCAFG00000001254	EGR1	Bio16, Elev	Bio16
ENSCAFG00000013220	ELOA RPL11	Elev	Bio16
ENSCAFG00000019922	EMC8	Bio16	Bio16
ENSCAFG00000004644	ENOX1	Elev	Bio16
ENSCAFG00000004755	ETHE1	Elev	Bio9
ENSCAFG00000005039	EXOG	Bio16	Bio16
ENSCAFG00000016019	FADS2	Bio16	Bio3
ENSCAFG00000019813	FANCA	Bio16	Bio16
ENSCAFG00000010176	FBXL15	Elev	Bio2
ENSCAFG00000009303	FBXO21	Bio16, Bio19	Bio16
ENSCAFG00000002264	FBXO9	Bio16	Bio3
ENSCAFG00000030542	FEN1	Bio16	Bio3
ENSCAFG00000006521	FRMD4B	Bio16	Bio18
ENSCAFG00000013615	FRMD8	Elev, Grass	Bio8
ENSCAFG00000018405	FST	Elev	Bio3
ENSCAFG00000013479	GABARAPL1	Elev	Bio8
ENSCAFG00000019374	GABRD	Bio16	Elev

ENSCAFG00000002494	GALNT12	Elev	Bio2
ENSCAFG00000010848	GAP43	Bio16	Bio18
ENSCAFG00000019802	GAS8	Bio16	Bio16
ENSCAFG00000058489	GNA11	Bio16	Bio8
ENSCAFG00000019279	GNA15	Bio16	Bio8
ENSCAFG00000019216	GNB1	Bio16	Elev
ENSCAFG00000005080	GORASP1 CX3CR1	Bio16	Bio8
ENSCAFG00000006842	GPAT2	Bio16	Bio8
ENSCAFG00000004554	GPR142	Elev	Elev
ENSCAFG00000013217	GPRC5A	Bio16	Bio8
ENSCAFG00000024066	GSTA4	Bio16	Bio3
ENSCAFG00000010593	GSTO1	Bio16	Bio2
ENSCAFG00000024905	GSTO2	Bio16	Bio2
ENSCAFG00000019606	HAGHL	Bio16	Bio7
ENSCAFG00000002813	HAS1	Bio16	Bio19
ENSCAFG00000007599	HELT	Elev	Bio16
ENSCAFG00000007038	HM13	Elev	Bio16
ENSCAFG00000014598	HMX1	Bio16	Bio2
ENSCAFG00000006580	HNRNPA1	Bio16	Bio8
ENSCAFG00000008388	HTRA2	Bio16	Bio16
ENSCAFG00000007045	ID1	Elev	Bio16
ENSCAFG00000015595	IGSF21	Elev	Bio16
ENSCAFG00000017287	IL12B	Elev	Bio8
ENSCAFG00000028874	IL17A IL-17A	Elev	Bio7
ENSCAFG00000016504	IL27RA	Bio16	Bio9
ENSCAFG00000008503	INO80B	Bio16	Bio16
ENSCAFG00000007842	IQCM	Bio7	Bio16
ENSCAFG00000019957	KCNG4	Elev	Bio8
ENSCAFG00000010452	KDM5B	Elev	Bio8
ENSCAFG00000030139	KIF19	Elev	Elev
ENSCAFG00000053642	KRAB domain-containing protein	Bio16	Bio19
ENSCAFG00000058087	KRAB domain-containing protein	Bio16	Bio3
ENSCAFG00000007025	KRTDAP	Elev, Grass	Bio8
ENSCAFG00000051969	L1 transposable element RRM domain-containing protein	Bio16	Bio16
ENSCAFG00000050271	L1 transposable element RRM domain-containing protein	Elev	Bio18
ENSCAFG00000006323	LAMP1	Bio2, Bio7	Bio2
ENSCAFG00000030312	LCN10	Elev	Bio9
ENSCAFG00000019560	LCN15	Elev	Bio9
ENSCAFG00000019563	LCN6	Elev	Bio9
ENSCAFG00000055476	LCN8	Elev	Bio9
ENSCAFG00000016038	LDB3	Bio16	Bio19
ENSCAFG00000002697	LENG1	Elev	Bio16
ENSCAFG00000008007	LIPI	Bio16	Bio9
ENSCAFG00000055887	LIPN	Bio7	Bio2
ENSCAFG00000043434	L-lactate dehydrogenase (EC 1.1.1.27)	Bio16	Bio19

ENSCAFG00000052228	lncRNA	Bio16	Bio16
ENSCAFG00000053211	lncRNA	Bio16	Bio2
ENSCAFG00000055845	lncRNA	Bio16	Bio9
ENSCAFG00000054845	lncRNA	Bio16	Bio9
ENSCAFG00000055302	lncRNA	Elev	Elev
ENSCAFG00000035380	lncRNA	Bio16	Bio19
ENSCAFG00000055624	lncRNA	Bio16	Bio19
ENSCAFG00000050514	lncRNA	Bio16	Bio19
ENSCAFG00000050708	lncRNA	Bio16	Bio7
ENSCAFG00000050799	lncRNA	Bio16	Bio7
ENSCAFG00000051090	lncRNA	Elev	Bio8
ENSCAFG00000050730	lncRNA	Bio16	Elev
ENSCAFG00000056492	lncRNA	Bio16	Elev
ENSCAFG00000055531	lncRNA	Elev	Bio2
ENSCAFG00000052992	lncRNA	Elev	Bio16
ENSCAFG00000058293	lncRNA	Elev	Bio16
ENSCAFG00000056293	lncRNA	Bio16	Bio8
ENSCAFG00000050771	lncRNA	Bio16	Bio8
ENSCAFG00000053380	lncRNA	Bio16	Bio8
ENSCAFG00000059400	lncRNA	Bio16	Bio8
ENSCAFG00000059135	lncRNA	Bio16	Bio8
ENSCAFG00000050347	lncRNA	Bio16	Bio19
ENSCAFG00000041664	lncRNA	Bio16	Bio3
ENSCAFG00000056101	lncRNA	Bio16	Bio19
ENSCAFG00000055147	lncRNA	Bio16	Bio19
ENSCAFG00000058619	lncRNA	Bio16	Bio19
ENSCAFG00000057066	lncRNA	Bio16	Bio19
ENSCAFG00000058950	lncRNA	Bio16	Bio19
ENSCAFG00000056879	lncRNA	Bio16	Bio19
ENSCAFG00000053098	lncRNA	Bio16	Bio19
ENSCAFG00000053796	lncRNA	Bio16	Bio19
ENSCAFG00000056786	lncRNA	Bio16	Bio19
ENSCAFG00000050499	lncRNA	Bio16	Bio19
ENSCAFG00000058102	lncRNA	Bio16	Bio2
ENSCAFG00000041550	lncRNA	Bio16, Grass	Elev
ENSCAFG00000054390	lncRNA	Bio16, Grass	Elev
ENSCAFG00000055057	lncRNA	Bio16	Bio18
ENSCAFG00000057012	lncRNA	Elev	Bio16
ENSCAFG00000054563	lncRNA	Elev	Bio8
ENSCAFG00000051288	lncRNA	Elev	Trees
ENSCAFG00000053047	lncRNA	Bio2, Bio3, Bio8, Bio16, Elev	Bio3
ENSCAFG00000054275	lncRNA	Bio2, Bio3, Bio8, Bio16, Elev	Bio3
ENSCAFG00000058603	lncRNA	Bio2, Bio3, Bio8, Bio16, Elev	Bio3
ENSCAFG00000043007	lncRNA	Bio16, Elev	Bio16
ENSCAFG00000058444	lncRNA	Elev	Bio7

ENSCAFG00000054295	lncRNA	Bio16	Bio3
ENSCAFG00000054631	lncRNA	Bio16	Bio8
ENSCAFG00000056719	lncRNA	Bio16	Bio8
ENSCAFG00000050670	lncRNA	Bio16	Bio8
ENSCAFG00000051106	lncRNA	Bio16	Bio7
ENSCAFG00000057029	lncRNA	Bio16	Bio7
ENSCAFG00000057508	lncRNA	Elev	Bio16
ENSCAFG00000053685	lncRNA	Bio2, Bio3, Bio7	Bio2
ENSCAFG00000056846	lncRNA	Elev	Elev
ENSCAFG00000037664	lncRNA	Elev	Bio16
ENSCAFG00000057607	lncRNA	Bio16	Bio8
ENSCAFG00000055197	lncRNA	Bio16	Bio16
ENSCAFG00000055378	lncRNA	Bio16	Bio3
ENSCAFG00000055056	lncRNA	Bio16	Bio3
ENSCAFG00000050600	lncRNA	Bio16	Bio3
ENSCAFG00000055445	lncRNA	Elev	Bio16
ENSCAFG00000054149	lncRNA	Elev	Bio16
ENSCAFG00000058829	lncRNA	Elev	Bio2
ENSCAFG00000050192	lncRNA	Bio16	Bio3
ENSCAFG00000050393	lncRNA	Elev	Elev
ENSCAFG00000057846	lncRNA	Elev	Elev
ENSCAFG00000056919	lncRNA	Bio16	Bio18
ENSCAFG00000053256	lncRNA	Bio16	Bio18
ENSCAFG00000052299	lncRNA	Bio16	Bio18
ENSCAFG00000044148	lncRNA	Bio16	Bio9
ENSCAFG00000056456	lncRNA	Bio16	Bio19
ENSCAFG00000058092	lncRNA	Bio16	Bio19
ENSCAFG00000050181	lncRNA	Elev	Bio2
ENSCAFG00000055777	lncRNA	Elev	Bio16
ENSCAFG00000058541	lncRNA	Bio16	Bio3
ENSCAFG00000053858	lncRNA	Bio2, Bio7	Bio2
ENSCAFG00000057176	lncRNA	Bio16	Bio16
ENSCAFG00000059474	lncRNA	Bio16	Bio16
ENSCAFG00000051914	lncRNA	Bio16	Bio3
ENSCAFG00000056685	lncRNA	Bio16	Bio2
ENSCAFG00000053991	lncRNA	Bio3, Bio7	Bio2
ENSCAFG00000049979	lncRNA	Bio16	Bio2
ENSCAFG00000054485	lncRNA	Bio16	Bio2
ENSCAFG00000055122	lncRNA	Bio16	Bio8
ENSCAFG00000052760	lncRNA	Bio16	Bio8
ENSCAFG00000055507	lncRNA	Bio7	Bio2
ENSCAFG00000057883	lncRNA	Elev	Bio8
ENSCAFG00000059447	lncRNA	Bio3, Bio8, Bio16, Bio19, Hfpi	Bio9
ENSCAFG00000059088	lncRNA	Bio16	Bio3
ENSCAFG00000051214	lncRNA	Elev	Bio3
ENSCAFG00000058735	lncRNA	Elev	Bio2

ENSCAFG00000054601	lncRNA	Bio16	Bio2
ENSCAFG00000052598	lncRNA	Elev	Elev
ENSCAFG00000053274	lncRNA	Bio16	Bio3
ENSCAFG00000055291	lncRNA	Bio16	Bio2
ENSCAFG00000055945	lncRNA	Bio16	Bio19
ENSCAFG00000055230	lncRNA	Bio16	Bio19
ENSCAFG00000052452	lncRNA	Bio16	Bio7
ENSCAFG00000059465	lncRNA	Bio16	Bio3
ENSCAFG00000050379	lncRNA	Elev	Bio8
ENSCAFG00000046487	lncRNA	Bio16	Bio19
ENSCAFG00000055005	lncRNA	Bio16	Bio19
ENSCAFG00000000399	LOC100683099	Bio16	Bio2
ENSCAFG00000031469	LOC100686484	Bio16	Bio8
ENSCAFG00000029073	LOC100688619	Bio16	Bio18
ENSCAFG00000024198	LOC106559612	Elev	Bio16
ENSCAFG00000045297	LOC111096443	Bio16	Bio16
ENSCAFG00000003851	LOC119870529	Elev	Bio19
ENSCAFG00000004727	LOC476445	Elev	Bio2
ENSCAFG00000014560	LOC479071	Bio16	Bio2
ENSCAFG00000024425	LOC484309	Elev	Bio16
ENSCAFG00000032614	LOC608975	Bio16	Bio7
ENSCAFG00000001593	LOC609023	Bio16	Bio7
ENSCAFG00000024271	LOC610304	Bio16	Bio3
ENSCAFG00000000311	LRIG3	Elev	Bio19
ENSCAFG00000000398	LRP11	Bio16	Bio2
ENSCAFG00000014177	LRR1	Bio16	Bio18
ENSCAFG00000004761	LYPD3	Elev	Bio2
ENSCAFG00000019548	MAMDC4	Elev	Bio9
ENSCAFG00000014976	MARK2	Elev	Bio2
ENSCAFG00000052675	MCFD2	Bio16	Bio8
ENSCAFG00000052880	MCRIP2	Bio16	Bio7
ENSCAFG00000003873	METTL24	Bio16	Bio8
ENSCAFG00000010201	MFSD13A	Elev	Bio2
ENSCAFG00000014184	MGAT2	Bio16	Bio18
ENSCAFG00000002323	MGAT4A	Bio2, Bio3, Bio8, Bio16, Elev	Bio3
ENSCAFG00000024105	MHC class I-like antigen recognition-like domain-containing protein	Bio16	Bio2
ENSCAFG00000056769	miRNA	Elev	Bio19
ENSCAFG00000050992	miRNA	Elev	Bio16
ENSCAFG00000057294	miRNA	Bio16	Bio8
ENSCAFG00000050369	miRNA	Bio16	Bio18
ENSCAFG00000051885	miRNA	Bio16	Bio18
ENSCAFG00000020542	miRNA	Elev	Bio7
ENSCAFG00000025700	miRNA	Elev	Bio7
ENSCAFG00000020452	miRNA	Elev	Bio8
ENSCAFG00000020463	miRNA	Elev	Bio8

ENSCAFG00000020442	miRNA	Elev	Bio3
ENSCAFG00000020489	miRNA	Elev	Bio3
ENSCAFG00000020617	miRNA	Elev	Bio3
ENSCAFG00000020495	miRNA	Bio16	Bio9
ENSCAFG00000020606	miRNA	Bio16	Bio9
ENSCAFG00000022889	miRNA	Bio16	Bio9
ENSCAFG00000054891	miRNA	Bio16	Bio8
ENSCAFG00000056418	miRNA	Bio16	Bio18
ENSCAFG00000022817	miRNA	Elev	Bio2
ENSCAFG00000052332	misc_RNA	Bio16	Bio18
ENSCAFG00000043857	Mitochondrial ribosomal protein L20	Bio16	Bio8
ENSCAFG00000008544	MMP15	Bio16	Elev
ENSCAFG00000015639	MNAT1	Elev	Bio16
ENSCAFG00000008735	MOB1A	Bio16, Elev	Bio16
ENSCAFG00000011596	MOCS3 UBA4	Bio3, Bio7	Bio2
ENSCAFG00000016639	MRPL45	Bio16	Bio8
ENSCAFG00000011502	MRPS11	Bio7	Bio16
ENSCAFG00000008716	MTHFD2	Bio16, Elev	Bio16
ENSCAFG00000003984	MTPAP	Bio16	Bio7
ENSCAFG00000018540	MYO1F	Bio16	Bio18
ENSCAFG00000014746	NAA40 COX8A	Elev	Bio2
ENSCAFG00000019529	NDUFS7	Bio16	Bio19
ENSCAFG00000020353	NEXN	Bio16	Bio16
ENSCAFG00000010155	NFKB2	Elev	Bio2
ENSCAFG00000015493	NLRP1	Bio16	Bio16
ENSCAFG00000003614	NOSIP	Elev	Bio8
ENSCAFG00000016644	NPEPPS	Bio16	Bio8
ENSCAFG00000059367	NPTX1	Elev	Bio8
ENSCAFG00000007813	NR3C2	Bio16	Bio2
ENSCAFG00000015626	NWD1	Elev	Bio18
ENSCAFG00000013744	ODR4	Elev	Bio9
ENSCAFG00000056600	Olfactory receptor	Bio16	Bio8
ENSCAFG00000015975	OPN4	Bio16	Bio19
ENSCAFG00000051241	OR4B1E	Bio16	Bio16
ENSCAFG00000051952	OR4B1F	Bio16	Bio16
ENSCAFG00000008100	OR5F2	Bio16	Bio8
ENSCAFG00000044917	OR5T2	Bio16	Bio8
ENSCAFG00000058806	OR8J2B	Bio16	Bio8
ENSCAFG00000054730	OR8J2C	Bio16	Bio8
ENSCAFG00000059259	OR8K71	Bio16	Bio8
ENSCAFG00000024209	OSCAR	Elev	Bio16
ENSCAFG00000003488	PAPPA	Bio16	Bio3
ENSCAFG00000011970	PARP14	Bio19	Bio2
ENSCAFG00000011940	PARP9	Bio19	Bio2
ENSCAFG00000013184	PDCD1	Bio16	Bio3
ENSCAFG00000006748	PDYN	Bio16, Elev	Bio2

ENSCAFG00000050625	PDZ domain-containing protein	Bio16	Bio8
ENSCAFG00000025190	Peptidase C19 ubiquitin carboxyl-terminal hydrolase domain-containing protein	Bio2, Bio3, Bio7	Bio2
ENSCAFG00000031185	PFN4 LOC119881789	Elev	Bio18
ENSCAFG00000004757	PHLDB3	Elev	Bio2
ENSCAFG00000019551	PHPT1	Elev	Bio9
ENSCAFG00000052089	PITHD1	Elev	Bio16
ENSCAFG00000013210	PITHD1 LYPLA2	Elev	Bio16
ENSCAFG00000002165	PKHD1	Elev	Bio7
ENSCAFG00000004730	PLAUR	Elev	Bio9
ENSCAFG00000002075	PLPP6	Elev	Elev
ENSCAFG00000031076	PMIS2 transmembrane protein	Elev, Grass	Bio8
ENSCAFG00000001392	PNPLA1	Bio16	Bio3
ENSCAFG00000010321	PPFIA4	Elev	Bio8
ENSCAFG00000000402	PPP1R14C	Bio16	Bio2
ENSCAFG00000004688	PREB	Bio16	Bio16
ENSCAFG00000013691	PRG4	Elev	Bio9
ENSCAFG00000003952	PRKAR2B	Bio16	Bio9
ENSCAFG00000019381	PRKCZ	Bio16	Elev
ENSCAFG00000003589	PRMT1	Elev	Bio8
ENSCAFG00000031477	Proline rich 23A	Elev	Elev
ENSCAFG00000057708	Proline rich 23A	Elev	Elev
ENSCAFG00000056213	PRR30	Bio16	Bio16
ENSCAFG00000050167	pseudogene	Bio16	Bio19
ENSCAFG00000009963	PSMD14	Bio16	Bio18
ENSCAFG00000020263	PTGS1 COX1	Elev	Bio9
ENSCAFG00000053284	RAB18	Elev	Bio19
ENSCAFG00000057801	RABIF	Elev	Bio8
ENSCAFG00000019556	RABL6	Elev	Bio9
ENSCAFG00000008410	RASA1	Bio16, Bio19	Bio16
ENSCAFG00000008744	RASGRF2	Elev	Bio16
ENSCAFG00000030041	RBM18	Elev	Bio9
ENSCAFG00000023645	REEP6	Bio16	Bio19
ENSCAFG00000000883	Reticulon	Bio16	Bio8
ENSCAFG00000055733	Reverse transcriptase domain-containing protein	Bio16	Bio2
ENSCAFG00000010409	RHBDD1	Bio16	Bio8
ENSCAFG00000019630	RHBDL1	Bio16	Bio7
ENSCAFG00000057895	RHOT2	Bio16	Bio7
ENSCAFG00000009308	Ribosomal protein L32	Bio16	Bio3
ENSCAFG00000053722	Ribosomal protein L37a	Elev	Bio3
ENSCAFG00000058812	RIC8B	Elev	Trees
ENSCAFG00000041658	RLN3	Bio16	Bio9
ENSCAFG00000017296	RNF145	Elev	Bio8
ENSCAFG00000023247	RPS15	Bio16	Bio19
ENSCAFG00000003604	RRAS	Elev	Bio8
ENSCAFG00000000664	RSPH3	Bio16	Bio9
ENSCAFG00000013018	RTKN2	Bio16	Bio8

ENSCAFG00000031912	SBSN	Elev, Grass	Bio8
ENSCAFG00000028693	Serine/arginine repetitive matrix 5	Elev	Bio2
ENSCAFG00000045246	Serine-threonine kinase receptor-associated protein	Elev	Bio16
ENSCAFG00000004600	SERP2	Elev	Bio2
ENSCAFG00000003957	SF3B6	Elev	Bio18
ENSCAFG00000012567	SH3BGRL3	Bio16	Elev
ENSCAFG00000005973	SIPA1L3	Bio16	Bio7
ENSCAFG00000029797	SIX1	Elev	Bio16
ENSCAFG00000010606	SLC17A1	Bio16	Bio16
ENSCAFG00000019177	SLC1A7	Bio16	Bio2
ENSCAFG00000015651	SLC35E1	Elev	Bio18
ENSCAFG00000009243	SLC38A4	Elev	Bio7
ENSCAFG00000008684	SLC4A5	Bio16	Bio16
ENSCAFG00000004481	SLC6A6	Bio16	Elev
ENSCAFG00000011241	SLCO3A1	Bio3, Bio7, Elev	Bio19
ENSCAFG00000007611	SLCO6A1	Elev	Bio2
ENSCAFG00000042570	Small integral membrane protein 27	Elev	Bio16
ENSCAFG00000050038	Small ribosomal subunit protein uS10m (28S ribosomal protein S10, mitochondrial)	Bio16	Bio8
ENSCAFG00000004726	SMG9	Elev	Bio9
ENSCAFG00000052668	SMIM18	Bio7	Bio16
ENSCAFG00000015648	SMIM7 MED26	Elev	Bio18
ENSCAFG00000049353	snoRNA	Elev	Bio19
ENSCAFG00000025856	snoRNA	Elev	Bio9
ENSCAFG00000025938	snoRNA	Elev	Bio9
ENSCAFG00000042140	snoRNA	Elev	Bio9
ENSCAFG00000028377	snRNA	Bio16	Bio2
ENSCAFG00000027615	snRNA	Bio16	Bio9
ENSCAFG00000048434	snRNA	Elev	Bio16
ENSCAFG00000021272	snRNA	Elev	Bio8
ENSCAFG00000043295	snRNA	Bio16	Bio2
ENSCAFG00000056325	snRNA	Elev	Bio8
ENSCAFG00000025869	snRNA	Bio16	Bio16
ENSCAFG00000025866	snRNA	Bio16, Grass	Elev
ENSCAFG00000044492	snRNA	Bio16	Elev
ENSCAFG00000027033	snRNA	Elev	Bio8
ENSCAFG00000021299	snRNA	Elev	Bio2
ENSCAFG00000044164	snRNA	Elev	Trees
ENSCAFG00000027440	snRNA	Bio16	Bio8
ENSCAFG00000056082	snRNA	Elev	Bio8
ENSCAFG00000027467	snRNA	Elev	Bio7
ENSCAFG00000028083	snRNA	Bio16	Bio3
ENSCAFG00000043191	snRNA	Bio16	Bio3
ENSCAFG00000044989	snRNA	Elev, Grass	Elev
ENSCAFG00000027204	snRNA	Bio16	Bio8
ENSCAFG00000047640	snRNA	Bio7	Bio16
ENSCAFG00000020895	snRNA	Elev	Bio16

ENSCAFG00000022270	snRNA	Bio16	Bio16
ENSCAFG00000026147	snRNA	Bio2, Bio7	Bio2
ENSCAFG00000022640	snRNA	Bio16	Bio7
ENSCAFG00000027061	snRNA	Bio16, Elev	Bio2
ENSCAFG00000021248	snRNA	Bio16, Bio19	Bio16
ENSCAFG00000021194	snRNA	Elev	Bio3
ENSCAFG00000052040	snRNA	Bio3	Bio8
ENSCAFG00000022587	snRNA	Bio16	Bio8
ENSCAFG00000043433	snRNA	Bio16	Bio19
ENSCAFG00000016629	SOCS7	Bio16	Bio8
ENSCAFG00000011745	SORL1	Bio16	Bio3
ENSCAFG00000017351	SOX30	Elev	Bio8
ENSCAFG00000011571	SOX5	Elev	Bio3
ENSCAFG00000009730	SPINT4	Bio16	Bio19
ENSCAFG00000011984	SPP2	Bio2, Bio3, Bio7	Bio19
ENSCAFG00000015699	STAMBPL1	Bio7	Bio2
ENSCAFG00000031099	STUB1	Bio16	Bio7
ENSCAFG00000000875	TAF2	Bio16	Bio8
ENSCAFG00000053294	TFPT	Elev	Bio16
ENSCAFG00000023828	TGM3 STK35	Bio16, Elev	Bio2
ENSCAFG00000013628	TIGD3	Elev, Grass	Bio8
ENSCAFG00000008445	TLX2	Bio16	Bio16
ENSCAFG00000008343	TM4SF1	Bio16	Bio16
ENSCAFG00000010330	TMEM183A	Elev	Bio8
ENSCAFG00000050707	TMEM263	Elev	Trees
ENSCAFG00000058654	TMEM52	Bio16	Elev
ENSCAFG00000013488	TMEM52B	Elev	Bio8
ENSCAFG00000009022	TMEM64	Bio16	Bio19
ENSCAFG00000031735	TNFSF13B	Bio16	Bio3
ENSCAFG00000018627	TNFSF14	Bio16, Elev	Bio18
ENSCAFG00000014575	TNS1	Bio16	Bio8
ENSCAFG00000003963	TP53I3	Elev	Bio18
ENSCAFG00000011758	TPST2	Elev	Bio8
ENSCAFG00000019541	TRAF2	Elev	Bio9
ENSCAFG00000002214	TRAM2	Bio16	Bio16
ENSCAFG00000024456	Transcription factor BTF3	Elev	Bio7
ENSCAFG00000001181	TRAPPC9	Bio16	Bio8
ENSCAFG00000001586	TREM2	Bio16	Bio7
ENSCAFG00000001589	TREML2	Bio16	Bio7
ENSCAFG00000051698	TRIM54 CAD	Bio16	Bio16
ENSCAFG00000012671	TRIM63	Bio16	Elev
ENSCAFG00000018610	TRIP10	Bio16, Elev	Bio18
ENSCAFG00000014566	TRMT44	Bio16	Bio2
ENSCAFG00000006759	TRMT9B	Bio16	Bio7
ENSCAFG00000002675	TSEN34	Elev	Bio16
ENSCAFG00000011808	TSHZ2	Bio16	Bio2

ENSCAFG00000011931	TTC13	Elev	Bio16
ENSCAFG00000008468	TTC31	Bio16	Bio16
ENSCAFG00000002641	TTC7A	Bio16	Bio8
ENSCAFG00000010412	TUBA5	Elev	Bio8
ENSCAFG00000005248	TUBAL3	Elev	Bio8
ENSCAFG00000043862	UBC core domain-containing protein	Bio16	Bio8
ENSCAFG00000032244	UBE2V1 LOC119881719	Elev	Bio16
ENSCAFG00000046052	Ubiquitin-like domain-containing protein	Elev, Grass	Bio8
ENSCAFG00000049866	Uncharacterized protein	Elev	Bio9
ENSCAFG00000049094	Uncharacterized protein	Elev	Bio9
ENSCAFG00000059138	Uncharacterized protein	Bio16	Bio7
ENSCAFG00000046775	Uncharacterized protein	Elev	Elev
ENSCAFG00000044869	Uncharacterized protein	Bio3, Bio7	Bio2
ENSCAFG00000020373	USP33	Bio16	Bio16
ENSCAFG00000047006	Vacuolar protein sorting-associated protein 29	Elev, Grass	Bio2
ENSCAFG00000018590	VAV1	Bio16, Elev	Bio18
ENSCAFG00000057448	Vomeronasal type-1 receptor	Elev	Bio9
ENSCAFG00000015966	WAPL	Bio16	Bio19
ENSCAFG00000008486	WBP1	Bio16	Bio16
ENSCAFG00000005986	WD repeat domain 87	Bio16	Bio7
ENSCAFG00000003949	WDCP	Elev	Bio18
ENSCAFG00000019654	WDR90	Bio16	Bio7
ENSCAFG00000007416	WEE1	Bio16	Bio16
ENSCAFG00000030127	WFDC10A	Bio16	Bio19
ENSCAFG00000009728	WFDC13	Bio16	Bio19
ENSCAFG00000019660	WFIKKN1	Bio16	Bio7
ENSCAFG00000004746	XRCC1	Elev	Bio2
ENSCAFG00000031728	YAP1	Bio16	Bio9
ENSCAFG00000014297	ZC3H12C	Bio16	Bio18
ENSCAFG00000015061	ZFTA	Elev	Bio2
ENSCAFG00000000967	ZHX2	Bio16	Bio8
ENSCAFG00000032633	ZNF283	Elev	Bio9
ENSCAFG00000008559	ZNF319	Bio16	Elev
ENSCAFG00000044320	ZNF576	Elev	Bio2
ENSCAFG00000029079	ZNF615 ZNF350 ZNF613	Bio16	Bio19
ENSCAFG00000032425	ZNF619	Bio16	Bio3
ENSCAFG00000053180	ZNF620	Bio16	Bio3
ENSCAFG00000052603	ZNF621	Bio16	Bio3
ENSCAFG00000008422	ZNF704	Bio16	Bio16
Results with species richness variables			
Misense variants			
ENSCAFG00000005318	CAPN13*	Mammals	Mammals
ENSCAFG00000008507	DNTT	LargeUng, Mammals	LargeUng
ENSCAFG00000003594	IRF3	LargeUng, MediumUng	Carnivores
ENSCAFG00000030013	RHEX	MediumUng	Carnivores

ENSCAFG00000024886	SPAG11E	MediumUng	LargeUng
Intron variants			
ENSCAFG00000004090	ADCY3	LargeUng	LargeUng
ENSCAFG00000019569	AGPAT2	LargeUng, Mammals	Carnivores
ENSCAFG00000012067	AP1B1	LargeUng	MediumUng
ENSCAFG00000051774	C1QTNF1	Mammals	Mammals
ENSCAFG00000002869	CEACAM18	LargeUng	LargeUng
ENSCAFG00000024608	FAM174A	MediumUng	Mammals
ENSCAFG00000013509	FAM177B	MediumUng	MediumUng
ENSCAFG00000008071	FHIP1A	MediumUng	Mammals
ENSCAFG00000006986	GAPDHS	LargeUng, MediumUng	Mammals
ENSCAFG00000011559	HCLS1	MediumUng	Carnivores
ENSCAFG00000008838	KCNAB1	MediumUng	LargeUng
ENSCAFG00000006520	KCNG3	LargeUng	Mammals
ENSCAFG00000016951	KCNIP1	Mammals	Carnivores
ENSCAFG00000000760	MAP3K4	LargeUng	MediumUng
ENSCAFG00000007867	MICAL2	MediumUng	Carnivores
ENSCAFG00000008716	MTHFD2	Mammals	Mammals
ENSCAFG00000031979	PRKG1	LargeUng	Carnivores
ENSCAFG00000018803	SAFB	Mammals	Mammals
ENSCAFG00000000326	UTRN	Mammals	Mammals
Non protein coding exon variants			
ENSCAFG00000052996	lncRNA	MediumUng	Mammals
ENSCAFG00000054969	lncRNA	Mammals	Carnivores
Splice region/intron variants			
ENSCAFG00000006013	CNTN4	MediumUng	Carnivores
ENSCAFG00000016464	ZFYVE26	LargeUng	Carnivores
Splice region/synonymous variant			
ENSCAFG00000006986	GAPDHS	LargeUng, MediumUng	Mammals
Synonymous variants			
ENSCAFG00000011195	ALPI	MediumUng	Carnivores
ENSCAFG00000019330	ASPA	MediumUng	Carnivores
ENSCAFG00000018625	C3	MediumUng	Carnivores
ENSCAFG00000006665	CLIP3	MediumUng	MediumUng
ENSCAFG00000002717	DNAH11	MediumUng	Carnivores
ENSCAFG00000012477	ITGB5	LargeUng	Carnivores
ENSCAFG00000013502	LOC478994	MediumUng	MediumUng
ENSCAFG00000006979	MTUS1	MediumUng	Mammals
ENSCAFG00000009321	RIPK1	MediumUng	Mammals
ENSCAFG00000003472	TSPAN12	MediumUng	Carnivores
3' UTR variants			
ENSCAFG00000007124	FXYD3	Mammals	LargeUng
ENSCAFG00000015807	MAT1A	MediumUng	Carnivores
5' UTR variant			
ENSCAFG00000011559	HCLS1	MediumUng	Carnivores

Downstream variants			
ENSCAFG00000020132	AARS1 AARS	MediumUng	Mammals
ENSCAFG00000012067	AP1B1	LargeUng	MediumUng
ENSCAFG00000042967	ATP synthase-coupling factor 6, mitochondrial (ATP synthase peripheral stalk subunit F6)	LargeUng	Mammals
ENSCAFG00000006951	ATP4A	LargeUng, MediumUng	Mammals
ENSCAFG00000010214	AVPR1B	MediumUng	Carnivores
ENSCAFG00000048656	Beta-defensin	MediumUng	LargeUng
ENSCAFG00000051604	Beta-defensin	MediumUng	LargeUng
ENSCAFG00000007963	BLK	MediumUng	Carnivores
ENSCAFG00000013516	BROX	MediumUng	MediumUng
ENSCAFG00000005537	CANT1	Mammals	Mammals
ENSCAFG00000046329	CBD139	MediumUng	LargeUng
ENSCAFG00000004085	CENPO	LargeUng	LargeUng
ENSCAFG00000003429	CFTR ABCC7	MediumUng	Carnivores
ENSCAFG00000003363	CHRM2	Mammals	Mammals
ENSCAFG00000002319	CRACDL	LargeUng, Mammals	Carnivores
ENSCAFG00000029831	CRBN	MediumUng	Carnivores
ENSCAFG00000029552	CSKMT	LargeUng	Carnivores
ENSCAFG00000003432	CTTNBP2	MediumUng	Carnivores
ENSCAFG00000030438	DEFB123	LargeUng, Mammals	Carnivores
ENSCAFG00000043376	Deoxyribonuclease TATDN1	MediumUng	MediumUng
ENSCAFG00000011170	DIS3L2	MediumUng	Carnivores
ENSCAFG00000004098	DNAJC27	LargeUng	LargeUng
ENSCAFG00000007268	DNMT3B	MediumUng	Carnivores
ENSCAFG00000050797	DTX4	MediumUng	LargeUng
ENSCAFG00000007651	DTX4 FAM111A	MediumUng	LargeUng
ENSCAFG00000015851	DYDC1	MediumUng	Carnivores
ENSCAFG00000011211	ECEL1	MediumUng	Carnivores
ENSCAFG00000031611	EFCAB8 MAPRE1	MediumUng	Carnivores
ENSCAFG00000050264	EF-hand domain-containing protein	MediumUng	Carnivores
ENSCAFG00000019575	EGFL7	LargeUng, Mammals	Carnivores
ENSCAFG00000012160	EWSR1 EMID1	LargeUng	MediumUng
ENSCAFG00000049011	EXOSC6	MediumUng	Mammals
ENSCAFG00000025213	FAM111B	MediumUng	LargeUng
ENSCAFG00000013509	FAM177B	MediumUng	MediumUng
ENSCAFG00000007093	FAM187B	Mammals	LargeUng
ENSCAFG00000044628	FAM72A	MediumUng	Carnivores
ENSCAFG00000011512	FBXO40	MediumUng	Carnivores
ENSCAFG00000007027	FFAR2	LargeUng, MediumUng	Mammals
ENSCAFG00000007124	FXYD3	Mammals	LargeUng
ENSCAFG00000006986	GAPDHS	LargeUng, MediumUng	Mammals
ENSCAFG00000012219	GAS2L1	LargeUng	MediumUng
ENSCAFG00000017594	GDPD1	MediumUng, Mammals	Carnivores

ENSCAFG00000020113	GLG1	MediumUng	Mammals
ENSCAFG00000009513	GLS	LargeUng	Mammals
ENSCAFG00000051067	GNG10 LOC119863873	MediumUng	Carnivores
ENSCAFG00000014643	GNG3	LargeUng	Carnivores
ENSCAFG00000011615	GOLGB1	MediumUng	Carnivores
ENSCAFG00000018620	GPR108	MediumUng	Carnivores
ENSCAFG00000007157	GRAMD1A	Mammals	LargeUng
ENSCAFG00000032284	HCST	MediumUng	MediumUng
ENSCAFG00000049617	HIG1 domain-containing protein	LargeUng, Mammals	Carnivores
ENSCAFG00000007126	HPN	Mammals	LargeUng
ENSCAFG00000009386	HPSE2	LargeUng	MediumUng
ENSCAFG00000004054	HS3ST5	LargeUng	Carnivores
ENSCAFG00000005991	IL5RA	MediumUng	Carnivores
ENSCAFG00000024216	KCND2	MediumUng	Carnivores
ENSCAFG00000030771	LCLAT1	Mammals	Mammals
ENSCAFG00000030312	LCN10	LargeUng, Mammals	Carnivores
ENSCAFG00000007111	LGI4	Mammals	LargeUng
ENSCAFG00000059327	lncRNA	MediumUng	Mammals
ENSCAFG00000054661	lncRNA	MediumUng	Mammals
ENSCAFG00000056410	lncRNA	MediumUng	Mammals
ENSCAFG00000050405	lncRNA	MediumUng	Mammals
ENSCAFG00000056021	lncRNA	Mammals	Mammals
ENSCAFG00000053081	lncRNA	LargeUng, Mammals	Carnivores
ENSCAFG00000057844	lncRNA	LargeUng, Mammals	Carnivores
ENSCAFG00000052488	lncRNA	LargeUng, Mammals	Carnivores
ENSCAFG00000041070	lncRNA	LargeUng	Carnivores
ENSCAFG00000058704	lncRNA	Mammals	Mammals
ENSCAFG00000057368	lncRNA	Mammals	Mammals
ENSCAFG00000054764	lncRNA	LargeUng	Mammals
ENSCAFG00000058425	lncRNA	Mammals	Mammals
ENSCAFG00000007649	lncRNA	MediumUng	LargeUng
ENSCAFG00000033953	lncRNA	MediumUng	LargeUng
ENSCAFG00000050338	lncRNA	LargeUng	Carnivores
ENSCAFG00000051893	lncRNA	MediumUng	LargeUng
ENSCAFG00000054076	lncRNA	MediumUng	Carnivores
ENSCAFG00000051123	lncRNA	LargeUng	Carnivores
ENSCAFG00000051996	lncRNA	MediumUng	Mammals
ENSCAFG00000024413	LOC102151856 ATP5J2 LOC608048	MediumUng	Mammals
ENSCAFG00000006940	LOC111090112	LargeUng, MediumUng	Mammals
ENSCAFG00000018775	LONP1	Mammals	Mammals
ENSCAFG00000053605	LRFN3	MediumUng	MediumUng
ENSCAFG00000018796	MICOS13	Mammals	Mammals
ENSCAFG00000020620	miRNA	LargeUng, Mammals	Carnivores

ENSCAFG00000054483	miRNA	MediumUng	Carnivores
ENSCAFG00000012525	MUC13	LargeUng	Carnivores
ENSCAFG00000016285	MYZAP POLR2M	Carnivores	MediumUng
ENSCAFG00000004075	NCOA1	LargeUng	LargeUng
ENSCAFG00000009314	NQO2	MediumUng	Mammals
ENSCAFG00000015594	NXF1	LargeUng	Carnivores
ENSCAFG00000008530	OPALIN	LargeUng, Mammals	LargeUng
ENSCAFG00000050258	OR5A2	MediumUng	LargeUng
ENSCAFG00000048104	OR5AN2	MediumUng	LargeUng
ENSCAFG00000047091	OR5AN6 LOC119869077	MediumUng	LargeUng
ENSCAFG00000015797	PADI3 PADI4	MediumUng, Mammals	LargeUng
ENSCAFG00000015759	PADI6	MediumUng, Mammals	LargeUng
ENSCAFG00000006953	PDGFRL	MediumUng	Mammals
ENSCAFG00000006601	POLR2I	MediumUng	MediumUng
ENSCAFG00000003611	PRR12	LargeUng, MediumUng	Carnivores
ENSCAFG00000024500	PRRG2	LargeUng, MediumUng	Carnivores
ENSCAFG00000053001	pseudogene	LargeUng	LargeUng
ENSCAFG00000051390	pseudogene	LargeUng	Carnivores
ENSCAFG00000019324	pseudogene	MediumUng	Carnivores
ENSCAFG00000057714	pseudogene	MediumUng	LargeUng
ENSCAFG00000047469	pseudogene	LargeUng	Carnivores
ENSCAFG00000012230	RASL10A	LargeUng	MediumUng
ENSCAFG00000005547	RBFOX3	Mammals	Mammals
ENSCAFG00000003620	RCN3	LargeUng, MediumUng	Carnivores
ENSCAFG00000016446	RDH12	LargeUng	Carnivores
ENSCAFG00000006991	REM1	LargeUng, Mammals	Carnivores
ENSCAFG00000015786	ROM1	LargeUng	Carnivores
ENSCAFG00000007129	SCN1B	Mammals	LargeUng
ENSCAFG00000054878	SDHAF1	MediumUng	MediumUng
ENSCAFG00000015753	SFTPD	MediumUng	Carnivores
ENSCAFG00000018602	SH2D3A	MediumUng	Carnivores
ENSCAFG00000019344	SHPK	MediumUng	Carnivores
ENSCAFG00000023358	SLC22A3	LargeUng	MediumUng
ENSCAFG00000008684	SLC4A5	Mammals	Mammals
ENSCAFG00000017590	SMG8	MediumUng, Mammals	Carnivores
ENSCAFG00000026218	snoRNA	MediumUng	Carnivores
ENSCAFG00000025856	snoRNA	LargeUng, Mammals	Carnivores
ENSCAFG00000025938	snoRNA	LargeUng, Mammals	Carnivores
ENSCAFG00000020883	snoRNA	LargeUng	Carnivores
ENSCAFG00000021073	snoRNA	LargeUng	Carnivores
ENSCAFG00000022180	snoRNA	LargeUng	Carnivores
ENSCAFG00000025801	snoRNA	LargeUng	Carnivores

ENSCAFG00000025916	snoRNA	LargeUng	Carnivores
ENSCAFG00000026010	snoRNA	LargeUng	Carnivores
ENSCAFG00000026225	snoRNA	LargeUng	Carnivores
ENSCAFG00000026245	snoRNA	LargeUng	Carnivores
ENSCAFG00000026308	snoRNA	LargeUng	Carnivores
ENSCAFG00000028095	snoRNA	LargeUng	MediumUng
ENSCAFG00000052908	snRNA	Mammals	Mammals
ENSCAFG00000041213	snRNA	LargeUng	MediumUng
ENSCAFG00000045207	snRNA	MediumUng	Carnivores
ENSCAFG00000049208	snRNA	Mammals	Mammals
ENSCAFG00000020917	snRNA	MediumUng	Carnivores
ENSCAFG00000025934	snRNA	MediumUng	Carnivores
ENSCAFG00000008558	SPAG11B	MediumUng	LargeUng
ENSCAFG00000009797	STAT4	LargeUng	Mammals
ENSCAFG00000015540	STX5	LargeUng	Carnivores
ENSCAFG00000007323	SUN5	MediumUng	Carnivores
ENSCAFG00000008728	TET3	Mammals	Mammals
ENSCAFG00000006649	THAP8	MediumUng	MediumUng
ENSCAFG00000008621	TLL2	LargeUng, Mammals	LargeUng
ENSCAFG00000015631	TMEM223	LargeUng	Carnivores
ENSCAFG00000019339	TRPV1	MediumUng	Carnivores
ENSCAFG00000019336	TRPV3	MediumUng	Carnivores
ENSCAFG00000003555	TSKS	LargeUng, MediumUng	Carnivores
ENSCAFG00000015659	TTC9C	LargeUng	Carnivores
ENSCAFG00000058032	TUBB2A LOC478702	MediumUng	Mammals
ENSCAFG00000012401	UMPS	LargeUng	Carnivores
ENSCAFG00000045878	Uncharacterized protein	LargeUng	Carnivores
ENSCAFG00000015715	UQCC3	LargeUng	Carnivores
ENSCAFG00000015523	WDR74	LargeUng	Carnivores
ENSCAFG00000015650	ZBTB3	LargeUng	Carnivores
ENSCAFG00000014146	ZNF385B	LargeUng	LargeUng
Upstream variants			
ENSCAFG00000024322	40S ribosomal protein S27	LargeUng	MediumUng
ENSCAFG00000028797	ADM5	LargeUng, MediumUng	Carnivores
ENSCAFG00000013540	AIDA	MediumUng	MediumUng
ENSCAFG00000051505	ALKBH6	MediumUng	MediumUng
ENSCAFG00000006669	ALKBH6 SYNE4	MediumUng	MediumUng
ENSCAFG00000003541	AP2A1	LargeUng, MediumUng	Carnivores
ENSCAFG00000015727	ARHGEF10L	MediumUng, Mammals	LargeUng
ENSCAFG00000006278	ATRN	LargeUng	Mammals
ENSCAFG00000015783	B3GAT3	LargeUng	Carnivores
ENSCAFG00000053130	BCL2L12	LargeUng, MediumUng	Carnivores
ENSCAFG00000008477	BLNK	LargeUng, Mammals	LargeUng

ENSCAFG00000008739	BOLA3	Mammals	Mammals
ENSCAFG00000009326	BPHL	MediumUng	Mammals
ENSCAFG00000007329	BPIFB2	MediumUng	Carnivores
ENSCAFG00000023629	BSCL2	LargeUng	Carnivores
ENSCAFG00000015719	C18H11orf98	LargeUng	Carnivores
ENSCAFG00000051774	C1QTNF1	Mammals	Mammals
ENSCAFG00000018769	CATSPERD	Mammals	Mammals
ENSCAFG00000032247	CBD103	MediumUng	LargeUng
ENSCAFG00000006990	CBD122 DEFB122	LargeUng, Mammals	Carnivores
ENSCAFG00000047383	CBD138	MediumUng	LargeUng
ENSCAFG00000030308	CD70	MediumUng	Carnivores
ENSCAFG00000011283	CHRND	MediumUng	Carnivores
ENSCAFG00000024807	CHRNG	MediumUng	Carnivores
ENSCAFG00000020124	CLEC18C	MediumUng	Mammals
ENSCAFG00000032040	COX7A2L	LargeUng	Mammals
ENSCAFG00000003568	CPT1C	LargeUng, MediumUng	Carnivores
ENSCAFG00000010206	CTSE	MediumUng	Carnivores
ENSCAFG00000002874	CTU1 ATPBD3 NCS6	LargeUng	LargeUng
ENSCAFG00000052287	DEFB119 CBD119	LargeUng, Mammals	Carnivores
ENSCAFG00000025021	DEFB121	LargeUng, Mammals	Carnivores
ENSCAFG00000024871	DEFB124	LargeUng, Mammals	Carnivores
ENSCAFG00000024964	DEFB4A	MediumUng	LargeUng
ENSCAFG00000019566	DIPK1B	LargeUng, Mammals	Carnivores
ENSCAFG00000007024	DMKN	LargeUng, MediumUng	Mammals
ENSCAFG00000006374	DNAAF9	LargeUng	Mammals
ENSCAFG00000015857	DYDC2	MediumUng	Carnivores
ENSCAFG00000005545	ENGASE	Mammals	Mammals
ENSCAFG00000007095	FXYD1	Mammals	LargeUng
ENSCAFG00000024320	FXYD5	Mammals	LargeUng
ENSCAFG00000050302	FXYD7	Mammals	LargeUng
ENSCAFG00000007015	G patch domain-containing protein 11 (Coiled-coil domain-containing protein 75)	MediumUng	Mammals
ENSCAFG00000015767	GANAB	LargeUng	Carnivores
ENSCAFG00000007663	GLYAT LOC483462	MediumUng	LargeUng
ENSCAFG00000004049	HDAC2	LargeUng	Carnivores
ENSCAFG00000007038	HM13	LargeUng, Mammals	Carnivores
ENSCAFG00000015686	HNRNPUL2	LargeUng	Carnivores
ENSCAFG00000009367	HPS1	LargeUng	MediumUng
ENSCAFG00000018793	HSD11B1L	Mammals	Mammals
ENSCAFG00000007045	ID1	LargeUng, Mammals	Carnivores
ENSCAFG00000015724	INTS5	LargeUng	Carnivores
ENSCAFG00000008838	KCNAB1	MediumUng	LargeUng
ENSCAFG00000007939	KCNB2	Mammals	Carnivores

ENSCAFG00000029529	KLK10	LargeUng	LargeUng
ENSCAFG00000002886	KLK11	LargeUng	LargeUng
ENSCAFG00000043606	KLK12 KLK13 KLNL	LargeUng	LargeUng
ENSCAFG00000002880	KLK14	LargeUng	LargeUng
ENSCAFG00000042298	KLK9	LargeUng	LargeUng
ENSCAFG00000007025	KRTDAP	LargeUng, MediumUng	Mammals
ENSCAFG00000052933	lncRNA	MediumUng	Mammals
ENSCAFG00000057482	lncRNA	MediumUng	Mammals
ENSCAFG00000051513	lncRNA	MediumUng	Mammals
ENSCAFG00000050210	lncRNA	MediumUng	Mammals
ENSCAFG00000053047	lncRNA	LargeUng, Mammals	Carnivores
ENSCAFG00000054275	lncRNA	LargeUng, Mammals	Carnivores
ENSCAFG00000058603	lncRNA	LargeUng, Mammals	Carnivores
ENSCAFG00000053330	lncRNA	LargeUng	LargeUng
ENSCAFG00000056186	lncRNA	MediumUng	Carnivores
ENSCAFG00000039143	lncRNA	MediumUng	Mammals
ENSCAFG00000054778	lncRNA	LargeUng	Mammals
ENSCAFG00000058727	lncRNA	LargeUng	Mammals
ENSCAFG00000055258	lncRNA	LargeUng	Mammals
ENSCAFG00000058687	lncRNA	LargeUng	Carnivores
ENSCAFG00000058350	lncRNA	LargeUng	Carnivores
ENSCAFG00000057759	lncRNA	Mammals	Mammals
ENSCAFG00000050732	lncRNA	MediumUng	Carnivores
ENSCAFG00000059161	lncRNA	MediumUng	LargeUng
ENSCAFG00000056680	lncRNA	MediumUng	LargeUng
ENSCAFG00000056506	lncRNA	MediumUng	Carnivores
ENSCAFG00000051214	lncRNA	LargeUng, Mammals	LargeUng
ENSCAFG00000054969	lncRNA	Mammals	Carnivores
ENSCAFG00000055660	lncRNA	MediumUng	Carnivores
ENSCAFG00000050379	lncRNA	LargeUng	Carnivores
ENSCAFG00000001959	LOC100686432	LargeUng	LargeUng
ENSCAFG00000028603	LOC100686511	LargeUng	LargeUng
ENSCAFG00000041785	LOC119868926 MBL1	MediumUng	Carnivores
ENSCAFG00000015754	LOC119868927	MediumUng	Carnivores
ENSCAFG00000002855	LOC119881386	LargeUng	LargeUng
ENSCAFG00000011328	LOC611215	MediumUng	Carnivores
ENSCAFG00000015703	LRRN4CL	LargeUng	Carnivores
ENSCAFG00000042311	LSM3	LargeUng	LargeUng
ENSCAFG00000013740	MAP2	LargeUng	Mammals
ENSCAFG00000000760	MAP3K4	LargeUng	MediumUng
ENSCAFG00000002323	MGAT4A	LargeUng, Mammals	Carnivores
ENSCAFG00000055873	miRNA	LargeUng	Carnivores
ENSCAFG00000008735	MOB1A	Mammals	Mammals
ENSCAFG00000006527	MTA3	LargeUng	Mammals

ENSCAFG00000008716	MTHFD2	Mammals	Mammals
ENSCAFG00000006979	MTUS1	MediumUng	Mammals
ENSCAFG00000003614	NOSIP	LargeUng, MediumUng	Carnivores
ENSCAFG00000019325	Olfactory receptor	MediumUng	Carnivores
ENSCAFG00000019327	Olfactory receptor	MediumUng	Carnivores
ENSCAFG00000023121	OR1E1	MediumUng	Carnivores
ENSCAFG00000006605	OVOL3	MediumUng	MediumUng
ENSCAFG00000000758	PLG	LargeUng	MediumUng
ENSCAFG00000031076	PMIS2 transmembrane protein	LargeUng, MediumUng	Mammals
ENSCAFG00000011479	POLQ	MediumUng	Carnivores
ENSCAFG00000015461	POLR2G	LargeUng	Carnivores
ENSCAFG00000003589	PRMT1	LargeUng, MediumUng	Carnivores
ENSCAFG00000011227	PRSS56	MediumUng	Carnivores
ENSCAFG00000015861	PRXL2A	MediumUng	Carnivores
ENSCAFG00000024865	PTRHD1	LargeUng	LargeUng
ENSCAFG00000009350	PYROXD2	LargeUng	MediumUng
ENSCAFG00000058840	Pyruvate dehydrogenase phosphatase regulatory subunit	MediumUng	Mammals
ENSCAFG00000010198	RAB7B	MediumUng	Carnivores
ENSCAFG00000016485	RAD51B	LargeUng	Carnivores
ENSCAFG00000015752	RCC2	MediumUng, Mammals	LargeUng
ENSCAFG00000057833	RDH11	LargeUng	Carnivores
ENSCAFG00000016442	RDH11 VTI1B	LargeUng	Carnivores
ENSCAFG00000058629	Reverse transcriptase domain-containing protein	MediumUng	Carnivores
ENSCAFG00000012114	RHBDD3	LargeUng	MediumUng
ENSCAFG00000048509	RNA-directed DNA polymerase (EC 2.7.7.49)	MediumUng	Carnivores
ENSCAFG00000052558	RPL36	Mammals	Mammals
ENSCAFG00000003604	RRAS	LargeUng, MediumUng	Carnivores
ENSCAFG00000018814	SAFB2	Mammals	Mammals
ENSCAFG00000031912	SBSN	LargeUng, MediumUng	Mammals
ENSCAFG00000002861	SIGLECL1	LargeUng	LargeUng
ENSCAFG00000028309	snRNA	LargeUng	LargeUng
ENSCAFG00000021272	snRNA	LargeUng, MediumUng	Carnivores
ENSCAFG00000043542	snRNA	MediumUng	Mammals
ENSCAFG00000020758	snRNA	Mammals	Mammals
ENSCAFG00000027883	snRNA	MediumUng	Carnivores
ENSCAFG00000021286	snRNA	LargeUng	Carnivores
ENSCAFG00000027167	snRNA	LargeUng	Carnivores
ENSCAFG00000022021	snRNA	MediumUng	Carnivores
ENSCAFG00000021194	snRNA	LargeUng, Mammals	LargeUng
ENSCAFG00000032628	SPATA22	MediumUng	Carnivores
ENSCAFG00000010224	SRGAP2	MediumUng	Carnivores
ENSCAFG00000048528	STAT1	LargeUng	Mammals

ENSCAFG00000003390	TAF1B	MediumUng	Carnivores
ENSCAFG00000015646	TAF6L	LargeUng	Carnivores
ENSCAFG00000006591	TBCB	MediumUng	MediumUng
ENSCAFG00000058977	TIMP2	Mammals	Mammals
ENSCAFG00000015636	TMEM179B	LargeUng	Carnivores
ENSCAFG00000018627	TNFSF14	MediumUng	Carnivores
ENSCAFG00000018610	TRIP10	MediumUng	Carnivores
ENSCAFG00000005985	TRNT1	MediumUng	Carnivores
ENSCAFG00000054854	TYROBP	MediumUng	MediumUng
ENSCAFG00000015710	UBXN1	LargeUng	Carnivores
ENSCAFG00000052661	Uncharacterized protein	MediumUng	MediumUng
ENSCAFG00000005524	USP36	Mammals	Mammals
ENSCAFG00000018590	VAV1	MediumUng	Carnivores
ENSCAFG00000006621	WDR62	MediumUng	MediumUng
ENSCAFG00000054302	YPEL2 YPEL	MediumUng, Mammals	Carnivores
ENSCAFG00000003386	YWHAQ	MediumUng	Carnivores

Supplementary table 6. Variant Effect Predictor (VEP) results summary FRDs.

Results with WorldClim variables			
ENSMBL ID	UnitProt ID or Name, Biotype for non-protein coding	LFMM-Association	RDA-Association
Missense variants			
ENSCAFG00000025018	ADAM7	Bio13	Bio8
ENSCAFG00000018685	CPLANE1	Bio13	Bio13
ENSCAFG00000013152	D2HGDH*	Bio7	Bio8
ENSCAFG00000017844	DNAH3	Bio13	Bio13
ENSCAFG00000007603	FLNB	Bio13	Bio13
ENSCAFG00000051032	Glyceraldehyde 3-phosphate dehydrogenase NAD(P) binding domain-containing protein	Bio14, Bio19, Hfpi	Bio8
ENSCAFG00000052337	GPR132	Bio13	Shrubs
ENSCAFG00000004702	HSPA14 MSANTD7*	Bio19	Bio7
ENSCAFG00000002147	IL1R1	Bio13	Bio13
ENSCAFG00000059254	INSM2	Bio14	Hfpi
ENSCAFG00000008969	KCTD9	Bio13	Bio13
ENSCAFG00000007770	LNPEP	Bio13	Bio13
ENSCAFG00000017518	LOC119866928	Bio14	Bio13
ENSCAFG00000058088	LOC481603	Bio19	Bio19
ENSCAFG00000032643	MRRF*	Bio13	Bio8
ENSCAFG00000001799	MTERF2	Elev	Bio8
ENSCAFG00000015929	N4BP2	Bio13	Bio13
ENSCAFG00000016959	PKLR	Bio14	Bio8
ENSCAFG00000013360	PPP2R3C	Bio13	Bio13
ENSCAFG00000002337	RCC1 domain containing 1	Bio14, Bio19	Bio13
ENSCAFG00000050337	RCCD1	Bio13	Bio2
ENSCAFG00000005055	RNF157	Elev	Hfpi
ENSCAFG00000015627	SIX4	Bio19	Bio8
ENSCAFG00000004826	SLC6A9*	Bio14	Elev
ENSCAFG00000010461	TM4SF20	Bio19	Crops
ENSCAFG00000009125	TMEM67	Bio13	Bio8
ENSCAFG00000049094	Uncharacterized protein	Bio13	Bio8
ENSCAFG00000008892	URB1	Bio13	Crops
Intron variants			
ENSCAFG00000003935	AGBL4	Bio14	Elev
ENSCAFG00000013146	BAZ1B	Bio19	Bio13
ENSCAFG00000001802	BEND4	Bio14	Bio8
ENSCAFG00000002927	CCDC85A	Bio13	Bio13
ENSCAFG00000019070	CHAF1A	Bio13	Hfpi
ENSCAFG00000020270	COL24A1	Bio13	Bio13
ENSCAFG00000013221	CTNNA3	Bio14	Shrubs
ENSCAFG00000002931	DCK	Bio14	Bio13
ENSCAFG00000017250	DENND4A	Bio13	Bio8
ENSCAFG00000058000	DIO2	Bio14	Bio13
ENSCAFG00000003006	DNAJC25	Bio14, Bio19	Bio13

ENSCAFG00000002413	DST	Bio14	Bio13
ENSCAFG00000018666	EGFLAM	Elev	Bio13
ENSCAFG00000017350	ELAVL3	Bio14	Bio7
ENSCAFG00000012712	EXTL1	Bio14	Bio13
ENSCAFG00000018955	FYB2	Bio13	Bio19
ENSCAFG00000009921	GABRG3	Bio14	Bio13
ENSCAFG00000004821	GLI2	Bio13	Bio19
ENSCAFG00000012129	GRK5	Bio13	Bio13
ENSCAFG00000012546	HDAC4	Bio13	Bio13
ENSCAFG00000000969	HSF2	Bio13	Bio13
ENSCAFG00000013297	IGF2BP2	Bio14	Bio19
ENSCAFG00000042586	Ig-like domain-containing protein	Bio14, Bio19, Hfpi	Shrubs
ENSCAFG00000015357	KAT6B	Bio13	Bio8
ENSCAFG00000047466	KAZN	Bio13	Bio8
ENSCAFG00000015560	KCNMB2	Bio13	Bio13
ENSCAFG00000042267	KRAB domain-containing protein	Bio14, Bio19, Hfpi	Bio19
ENSCAFG00000058939	LOC102152670	Bio14, Bio19, Hfpi	Bio19
ENSCAFG00000005321	LOC475563	Bio14	Grass
ENSCAFG00000023536	LOC607646	Bio14	Bio8
ENSCAFG00000019743	LOC611691	Bio14	Trees
ENSCAFG00000008788	MSH3	Bio13	Bio8
ENSCAFG00000047663	MTNR1A	Bio13	Hfpi
ENSCAFG00000056231	NOX4	Bio13	Bio13
ENSCAFG00000029745	NRTN	Bio19	Bio8
ENSCAFG00000005606	PAK4	Bio14, Bio19, Hfpi	Bio7
ENSCAFG00000018583	PDE4B	Bio13	Bio13
ENSCAFG00000011937	PGBD1	Bio13	Bio7
ENSCAFG00000058801	PID1	Shrubs	Bio19
ENSCAFG00000002165	PKHD1	Bio19	Bio14
ENSCAFG00000004179	PRDM5	Bio13	Bio13
ENSCAFG00000005301	PSD3	Bio14	Grass
ENSCAFG00000018073	PSTPIP1	Bio14	Bio13
ENSCAFG00000008370	RNF13	Bio14	Bio13
ENSCAFG00000005692	RPTOR	Bio13	Shrubs
ENSCAFG00000008353	SFRP2	Bio13	Bio2
ENSCAFG00000016602	SMOC1	Bio14	Bio13
ENSCAFG00000003059	SNX30	Bio13	Bio2
ENSCAFG00000002171	SYK	Bio13	Bio13
ENSCAFG00000003662	TBC1D9	Bio14	Shrubs
ENSCAFG00000010669	TRAPPC10	Bio13	Bio13
ENSCAFG00000000473	TRHDE	Bio13	Bio13
ENSCAFG00000001226	TRIM58	Bio14	Bio13
ENSCAFG00000019101	VPS53	Bio13	Bio13
ENSCAFG00000010823	ZBTB20	Bio14	Bio2

ENSCAFG00000014492	TNP1	Bio13	Bio13
ENSCAFG00000034862	lncRNA	Bio7	Bio8
ENSCAFG00000035932	lncRNA	Bio13	Bio13
ENSCAFG00000050153	lncRNA	Bio13	Bio13
Noncoding trasncripts exon variants			
ENSCAFG00000051860	lncRNA	Bio8	Bio8
ENSCAFG00000054575	lncRNA	Bio8	Bio8
ENSCAFG00000055289	lncRNA	Bio13	Bio2
Splice donor or region, intron variants			
ENSCAFG00000012327	WDR11	Elev	Bio8
ENSCAFG00000007310	PAH	Bio13	Bio19
ENSCAFG00000003591	C5	Bio13	Bio13
ENSCAFG00000001577	CEP78	Bio19, Elev	Bio19
Synonymous variants			
ENSCAFG00000011181	ARFGEF2	Bio13	Bio8
ENSCAFG00000007143	ASAH1	Elev	Bio13
ENSCAFG00000013106	ATF7IP	Bio13	Bio8
ENSCAFG00000011268	B3GALNT2	Bio19	Bio13
ENSCAFG00000002033	CALCR	Bio13	Hfpi
ENSCAFG00000056742	CENPX	Elev	Bio13
ENSCAFG00000009653	DOP1B	Bio13	Bio13
ENSCAFG00000001732	HMGXB4	Elev	Bio2
ENSCAFG00000007564	IMMT	Bio13	Bio2
ENSCAFG00000015724	INTS5	Bio13	Bio13
ENSCAFG00000013836	KANSL1L	Bio13	Bio13
ENSCAFG00000025009	NKAPL ZSCAN26	Bio13	Bio7
ENSCAFG00000047151	Olfactory receptor	Bio14	Bio19
ENSCAFG00000013067	PAK2	Bio13	Hfpi
ENSCAFG00000016089	PARP1	Bio13	Bio13
ENSCAFG00000009649	PLEKHA6	Bio13	Bio13
ENSCAFG00000029252	RND3	Bio13	Bio7
ENSCAFG00000000192	SEMA6A	Bio19	Bio13
3' UTR variants			
ENSCAFG00000017820	CD276	Bio19	Bio13
ENSCAFG00000003717	CYRIA	Bio14	Bio8
ENSCAFG00000004995	DMAC2	Bio13	Bio13
ENSCAFG00000029586	GCHFR	Bio14	Elev
ENSCAFG00000008230	LIMA1	Bio13	Bio14
ENSCAFG00000032563	NKIRAS2	Bio19	Bio2
ENSCAFG00000014388	PSAP	Bio7	Bio8
ENSCAFG00000003103	RRAGD	Bio13	Elev
ENSCAFG00000014386	SLC25A42	Bio14	Bio13
ENSCAFG00000055234	TSNAX	Bio13	Bio8
5' UTR variant			
ENSCAFG00000055600	Immunoglobulin V-set domain-containing protein	Bio14	Bio13
Downstream variants			

ENSCAFG00000008658	40S ribosomal protein S3a	Bio13	Bio19
ENSCAFG00000013856	ACADL	Bio13	Bio13
ENSCAFG00000011534	ACAN	Bio13	Bio8
ENSCAFG00000051884	ACOT8	Bio13	Bio13
ENSCAFG00000009768	ACOT8 TNNC2	Bio13	Bio13
ENSCAFG00000017319	ACP5	Bio14	Bio7
ENSCAFG00000005618	ACP7	Bio14, Bio19, Hfpi	Bio7
ENSCAFG00000004013	ADAD1	Bio14	Bio13
ENSCAFG00000024869	ADAMDEC1	Bio13	Bio8
ENSCAFG00000010654	AGPAT3	Bio13	Bio13
ENSCAFG00000000234	AHI1	Bio13	Bio19
ENSCAFG00000011267	AJUBA	Bio14	Bio19
ENSCAFG00000029647	AKR1D1	Bio14	Bio2
ENSCAFG00000000588	ALDH7A1	Bio13	Bio19
ENSCAFG00000005878	ANAPC11	Elev	Bio13
ENSCAFG00000016691	ANGPTL7	Bio13	Bio19
ENSCAFG00000003887	APOB	Bio13	Bio19
ENSCAFG00000050591	ARL5B	Bio14	Bio8
ENSCAFG00000014404	ARMC6	Bio14	Bio13
ENSCAFG00000016944	ASH1L	Bio14	Bio8
ENSCAFG00000053375	ASIC1	Bio13	Bio14
ENSCAFG00000015597	ASIC4	Bio13	Bio2
ENSCAFG00000005830	ASPSCR1	Elev	Bio13
ENSCAFG00000013054	ATG4B	Bio7	Bio8
ENSCAFG00000004854	ATP6V0B	Bio14	Elev
ENSCAFG00000000513	ATP6V1G1 ATP6G	Bio14	Bio13
ENSCAFG00000011322	ATP6V1G3	Bio13	Bio8
ENSCAFG00000004833	B4GALT2	Bio14	Elev
ENSCAFG00000012199	BAG3	Bio13	Bio8
ENSCAFG00000053144	BORCS8	Bio14	Bio13
ENSCAFG00000044557	BPTI/Kunitz inhibitor domain-containing protein	Bio13	Bio13
ENSCAFG00000023629	BSCL2	Bio13	Bio13
ENSCAFG00000013190	BTBD1	Bio13	Bio7
ENSCAFG00000030374	C12H6orf141	Bio14	Bio13
ENSCAFG00000015719	C18H11orf98	Bio13	Bio13
ENSCAFG00000023343	C1H19orf18	Bio14, Bio19, Hfpi	Bio19
ENSCAFG00000013185	C3H15orf40	Bio13	Bio7
ENSCAFG00000023392	C5H11orf87	Elev	Bio8
ENSCAFG00000011284	C8H14orf93	Bio14	Bio19
ENSCAFG00000006941	CAT	Bio14	Bio13
ENSCAFG00000018769	CATSPERD	Bio19	Bio8
ENSCAFG00000055440	CBM21 domain-containing protein	Bio13	Bio8
ENSCAFG00000009633	CBR3	Bio13	Bio13
ENSCAFG00000023153	CCDC152	Bio19	Bio13
ENSCAFG00000046547	CCDC24	Bio14	Elev

ENSCAFG00000004512	CCDC89	Bio13	Bio19
ENSCAFG00000005015	CCDC97	Bio13	Bio13
ENSCAFG00000014229	CDH23	Bio7	Bio8
ENSCAFG00000000206	CDO1	Bio14	Bio13
ENSCAFG00000002126	CENPQ	Bio14	Bio13
ENSCAFG00000057002	CEP170B	Bio13	Shrubs
ENSCAFG00000019759	CHIA	Bio14	Trees
ENSCAFG00000009105	CIBAR1	Bio13	Bio8
ENSCAFG00000031672	CLBA1	Bio13	Shrubs
ENSCAFG00000015757	CNP	Bio19	Bio2
ENSCAFG00000017682	COG7	Bio14	Bio14
ENSCAFG00000010426	COL4A3	Bio19	Crops
ENSCAFG00000058902	Cornichon family AMPA receptor auxiliary protein 1	Bio14	Bio2
ENSCAFG00000023105	CRISP2 LOC481834	Bio14	Bio13
ENSCAFG00000008550	CRSP2 CRSP-2	Bio14	Bio8
ENSCAFG00000001797	CRY1	Elev	Bio8
ENSCAFG00000014999	CRYBA2	Bio14	Bio13
ENSCAFG00000005109	CXCR4	Bio13	Bio13
ENSCAFG00000014943	CYP27A1	Bio14	Bio13
ENSCAFG00000004692	DCLRE1C	Bio19	Bio7
ENSCAFG00000005927	DCXR	Elev	Bio13
ENSCAFG00000000509	DDX39B BAT1 UAP56	Bio14	Bio13
ENSCAFG00000015720	DHX58	Bio19	Bio2
ENSCAFG00000013305	DNAJC12	Bio13	Bio13
ENSCAFG00000009322	DNAJC17	Bio14	Elev
ENSCAFG00000010548	DNER	Shrubs	Bio19
ENSCAFG00000009733	DNTTIP1	Bio13	Bio13
ENSCAFG00000009001	DOCK5	Bio13	Bio13
ENSCAFG00000014622	DOK7 HGFAC	Bio13	Shrubs
ENSCAFG00000004862	DPH2	Bio14	Elev
ENSCAFG00000005952	DUS1L	Elev	Bio13
ENSCAFG00000009033	DYSF	Bio13	Bio13
ENSCAFG00000017334	ECSIT	Bio14	Bio7
ENSCAFG00000006163	EEA1	Bio13, Elev	Bio13
ENSCAFG00000031111	EFCAB11	Bio7	Bio8
ENSCAFG00000029846	ELMOD2	Bio14	Shrubs
ENSCAFG00000058608	ELOF1	Bio14	Bio7
ENSCAFG00000032716	EMCN	Bio13	Bio8
ENSCAFG00000024646	ERAP2	Bio13	Bio13
ENSCAFG00000018143	ETFA	Bio13	Bio7
ENSCAFG00000000831	EXT1	Bio13	Bio19
ENSCAFG00000013335	FAM177A1	Bio13	Bio13
ENSCAFG00000006006	FASN	Elev	Bio13
ENSCAFG00000007273	FAT1	Bio13	Hfpi
ENSCAFG00000042386	Ferritin	Bio14	Bio2
ENSCAFG00000012298	FES	Bio13	Bio2

ENSCAFG00000041791	FEV	Bio14	Bio13
ENSCAFG00000032412	FLI1	Bio14	Bio19
ENSCAFG00000005385	FNTA	Bio13	Bio13
ENSCAFG00000007135	FRG1	Elev	Bio13
ENSCAFG00000019098	FSD1	Bio13	Hfpi
ENSCAFG00000012353	FURIN	Bio13	Bio2
ENSCAFG00000018754	FUT5	Bio19	Bio8
ENSCAFG00000013159	FZD9	Bio19	Bio13
ENSCAFG00000012582	G2E3	Bio19	Bio19
ENSCAFG00000010035	GABRA5	Bio14	Bio13
ENSCAFG00000023738	GDPGP1	Bio13	Bio2
ENSCAFG00000018579	GHR	Bio19	Bio13
ENSCAFG00000023219	GLIPR1L1	Bio14, Bio19, Hfpi	Bio8
ENSCAFG00000029990	GLIPR1L2	Bio14, Bio19, Hfpi	Bio8
ENSCAFG00000002127	GLYATL3	Bio14	Bio13
ENSCAFG00000050916	GNG4	Bio19	Bio13
ENSCAFG00000008288	GPD1	Bio13	Bio14
ENSCAFG00000055900	GPR17	Bio13	Bio8
ENSCAFG00000016963	HCN3	Bio14	Bio8
ENSCAFG00000005026	HNRNPUL1	Bio13	Bio13
ENSCAFG00000015686	HNRNPUL2	Bio13	Bio13
ENSCAFG00000005407	HOOK3	Bio13	Bio13
ENSCAFG00000029080	HRCT1	Bio19	Bio19
ENSCAFG00000031862	HTH CENPB-type domain-containing protein	Bio13	Bio13
ENSCAFG00000012556	HTRA1	Bio13	Bio14
ENSCAFG00000005588	IFNL3	Bio14, Bio19, Hfpi	Bio7
ENSCAFG00000014490	IGFBP2	Bio13	Bio19
ENSCAFG00000002330	IGFBP7	Bio14	Bio13
ENSCAFG00000003811	Ig-like domain-containing protein	Bio14, Bio19, Hfpi	Shrubs
ENSCAFG00000024810	Ig-like domain-containing protein	Bio14, Bio19, Hfpi	Shrubs
ENSCAFG00000046664	Ig-like domain-containing protein	Bio14, Bio19, Hfpi	Shrubs
ENSCAFG00000049482	Ig-like domain-containing protein	Bio14	Bio13
ENSCAFG00000052806	Ig-like domain-containing protein	Bio14, Bio19, Hfpi	Shrubs
ENSCAFG00000055018	Ig-like domain-containing protein	Bio14, Bio19, Hfpi	Shrubs
ENSCAFG00000057117	Ig-like domain-containing protein	Bio14, Bio19, Hfpi	Shrubs
ENSCAFG00000057504	Ig-like domain-containing protein	Bio14, Bio19, Hfpi	Shrubs
ENSCAFG00000004004	IL21	Bio14	Bio13
ENSCAFG00000029837	Immunoglobulin V-set domain-containing protein	Bio14, Bio19, Hfpi	Shrubs

ENSCAFG00000055372	Immunoglobulin V-set domain-containing protein	Bio14, Bio19, Hfpi	Shrubs
ENSCAFG00000013088	ING5	Bio7	Bio8
ENSCAFG00000015653	INHA	Bio13	Bio2
ENSCAFG00000030990	INSYN1	Bio19	Bio13
ENSCAFG00000052759	Interferon lambda-4-like	Bio14, Bio19, Hfpi	Bio7
ENSCAFG00000004875	IPO13	Bio14	Elev
ENSCAFG00000000404	IYD	Bio14, Bio19, Hfpi	Bio13
ENSCAFG00000018401	JAG2	Bio13	Shrubs
ENSCAFG00000015692	KAT2A	Bio19	Bio2
ENSCAFG00000015621	KCNH4	Bio19	Bio2
ENSCAFG00000003043	KIAA1958	Bio13	Bio2
ENSCAFG00000009202	KNL1	Bio14	Elev
ENSCAFG00000048255	KRAB domain-containing protein	Bio14, Bio19	Bio14
ENSCAFG00000011247	LAPTM5	Bio13	Bio13
ENSCAFG00000005309	LBH	Bio13	Bio14
ENSCAFG00000002482	LGSN	Bio13	Bio13
ENSCAFG00000020226	LHX2	Bio13	Bio2
ENSCAFG00000001792	LIMCH1	Bio13	Bio14
ENSCAFG00000007759	LIX1	Bio13	Bio13
ENSCAFG00000036487	lncRNA	Bio13	Bio8
ENSCAFG00000041370	lncRNA	Bio14	Shrubs
ENSCAFG00000045810	lncRNA	Bio19	Bio8
ENSCAFG00000049921	lncRNA	Bio13	Bio13
ENSCAFG00000050161	lncRNA	Bio13	Bio13
ENSCAFG00000050225	lncRNA	Bio14	Bio13
ENSCAFG00000050228	lncRNA	Bio13	Bio2
ENSCAFG00000050279	lncRNA	Bio13	Bio13
ENSCAFG00000050539	lncRNA	Bio13	Bio19
ENSCAFG00000050612	lncRNA	Bio13	Bio13
ENSCAFG00000050699	lncRNA	Bio14	Bio2
ENSCAFG00000050712	lncRNA	Elev	Bio8
ENSCAFG00000051028	lncRNA	Bio13	Bio13
ENSCAFG00000051288	lncRNA	Elev	Bio8
ENSCAFG00000051460	lncRNA	Bio19	Bio8
ENSCAFG00000051555	lncRNA	Bio13	Bio7
ENSCAFG00000051661	lncRNA	Bio14	Bio7
ENSCAFG00000051752	lncRNA	Bio14	Bio8
ENSCAFG00000051796	lncRNA	Bio14	Bio13
ENSCAFG00000051870	lncRNA	Bio13	Bio13
ENSCAFG00000052105	lncRNA	Bio14	Bio8
ENSCAFG00000052171	lncRNA	Bio13	Bio13
ENSCAFG00000052304	lncRNA	Bio14	Bio13
ENSCAFG00000052342	lncRNA	Bio14	Bio8
ENSCAFG00000052352	lncRNA	Bio19	Bio8

ENSCAFG00000052972	lncRNA	Bio14	Bio19
ENSCAFG00000053060	lncRNA	Bio14	Bio13
ENSCAFG00000053075	lncRNA	Bio14, Bio19	Bio13
ENSCAFG00000053455	lncRNA	Bio13	Bio13
ENSCAFG00000053492	lncRNA	Bio14	Bio13
ENSCAFG00000053565	lncRNA	Bio14	Bio8
ENSCAFG00000053588	lncRNA	Bio19	Bio13
ENSCAFG00000053648	lncRNA	Bio14	Hfpi
ENSCAFG00000053712	lncRNA	Bio14	Bio13
ENSCAFG00000053745	lncRNA	Bio14	Bio13
ENSCAFG00000053899	lncRNA	Bio14	Bio13
ENSCAFG00000053917	lncRNA	Bio13	Bio13
ENSCAFG00000054099	lncRNA	Bio14	Bio13
ENSCAFG00000054234	lncRNA	Bio19	Bio13
ENSCAFG00000054559	lncRNA	Bio13	Bio13
ENSCAFG00000054576	lncRNA	Bio14	Bio13
ENSCAFG00000054596	lncRNA	Bio14	Bio8
ENSCAFG00000054671	lncRNA	Bio13	Bio7
ENSCAFG00000054782	lncRNA	Bio13	Bio7
ENSCAFG00000055275	lncRNA	Bio13	Bio14
ENSCAFG00000055388	lncRNA	Bio14	Bio13
ENSCAFG00000055439	lncRNA	Bio14	Hfpi
ENSCAFG00000055497	lncRNA	Bio13	Bio2
ENSCAFG00000055862	lncRNA	Bio13	Bio13
ENSCAFG00000055902	lncRNA	Bio19	Bio8
ENSCAFG00000056038	lncRNA	Bio14	Bio13
ENSCAFG00000056062	lncRNA	Bio14	Bio13
ENSCAFG00000056483	lncRNA	Bio13	Bio2
ENSCAFG00000056979	lncRNA	Bio14	Bio8
ENSCAFG00000057022	lncRNA	Bio14, Bio19	Bio14
ENSCAFG00000057075	lncRNA	Bio14	Bio13
ENSCAFG00000057099	lncRNA	Elev	Bio13
ENSCAFG00000057187	lncRNA	Bio13	Bio13
ENSCAFG00000057251	lncRNA	Bio14	Bio13
ENSCAFG00000057349	lncRNA	Bio13	Bio13
ENSCAFG00000057389	lncRNA	Bio14	Bio8
ENSCAFG00000057756	lncRNA	Bio19	Bio13
ENSCAFG00000057785	lncRNA	Bio14	Bio19
ENSCAFG00000057930	lncRNA	Bio14	Bio8
ENSCAFG00000057939	lncRNA	Bio14	Bio2
ENSCAFG00000058350	lncRNA	Bio13	Bio13
ENSCAFG00000058374	lncRNA	Bio13	Bio8
ENSCAFG00000058383	lncRNA	Bio13	Bio8
ENSCAFG00000058854	lncRNA	Bio13	Bio19
ENSCAFG00000058856	lncRNA	Bio14	Bio13

ENSCAFG00000058863	lncRNA	Bio14	Bio7
ENSCAFG00000058944	lncRNA	Bio13	Elev
ENSCAFG00000059058	lncRNA	Bio14	Bio13
ENSCAFG00000059085	lncRNA	Bio13	Bio13
ENSCAFG00000059098	lncRNA	Bio14	Bio8
ENSCAFG00000059134	lncRNA	Bio13	Bio8
ENSCAFG00000059318	lncRNA	Bio14	Bio19
ENSCAFG00000059337	lncRNA	Bio14	Bio13
ENSCAFG00000002139	LOC100687983	Bio13	Bio8
ENSCAFG00000045251	LOC106558923	Bio14	Trees
ENSCAFG00000048484	LOC119866701	Bio13	Bio8
ENSCAFG00000028580	LOC476784	Bio13	Bio19
ENSCAFG00000032308	LOC477391	Bio13	Bio2
ENSCAFG00000012035	LOC479047	Bio13	Bio2
ENSCAFG00000019740	LOC479600	Bio7	Bio8
ENSCAFG00000028617	LOC482176	Bio14, Bio19, Trees	Bio13
ENSCAFG00000015731	LOC490972	Bio19	Bio2
ENSCAFG00000031485	LOC490972	Bio19	Bio2
ENSCAFG00000049417	LOC607806	Elev	Bio13
ENSCAFG00000024412	LOC607889	Bio14	Shrubs
ENSCAFG00000019015	LRG1	Bio13	Hfpi
ENSCAFG00000011202	LRP10 REM2	Bio14	Bio19
ENSCAFG00000015703	LRRN4CL	Bio13	Bio13
ENSCAFG00000029018	LYPD4	Bio13	Bio13
ENSCAFG00000050842	LYRM4	Bio13	Bio8
ENSCAFG00000012259	MAN2A2	Bio13	Bio2
ENSCAFG00000054114	MATN1	Bio13	Bio13
ENSCAFG00000000918	MCAT	Bio13	Bio13
ENSCAFG00000012225	MCMBP	Bio13	Bio8
ENSCAFG00000014350	MEF2B	Bio14	Bio13
ENSCAFG00000003233	MEIS1	Bio14	Bio13
ENSCAFG00000047744	MFF	Bio19	Crops
ENSCAFG00000051129	MHC class I-like antigen recognition-like domain-containing protein	Bio14	Bio13
ENSCAFG00000022806	miRNA	Bio13	Hfpi
ENSCAFG00000025651	miRNA	Bio13	Bio8
ENSCAFG00000025715	miRNA	Bio13	Hfpi
ENSCAFG00000048199	miRNA	Elev	Hfpi
ENSCAFG00000052519	miRNA	Bio14	Bio8
ENSCAFG00000055788	miRNA	Bio14	Bio8
ENSCAFG00000006496	MITF	Bio14	Bio7
ENSCAFG00000049889	Mitochondrial import inner membrane translocase subunit TIM23	Bio14, Bio19, Hfpi	Bio19
ENSCAFG00000012970	MLXIPL	Bio19	Bio13
ENSCAFG00000015054	MMP1	Bio14	Bio2
ENSCAFG00000023615	MMP12	Bio14	Bio2

ENSCAFG00000011197	MMP14	Bio14	Bio19
ENSCAFG00000015063	MMP3	Bio14	Bio2
ENSCAFG00000051237	MOB1B	Bio14	Bio13
ENSCAFG00000020280	MORN5	Bio13	Bio8
ENSCAFG00000019091	MPND	Bio13	Hfpi
ENSCAFG00000008888	MRAP	Bio13	Crops
ENSCAFG00000014361	MREG	Bio19	Bio19
ENSCAFG00000011191	MRPL52	Bio14	Bio19
ENSCAFG00000051878	MSANTD7	Bio19	Bio7
ENSCAFG00000016996	MTX1	Bio14	Bio8
ENSCAFG00000004404	MYO7B	Bio13	Bio8
ENSCAFG00000016285	MYZAP POLR2M	Bio13	Bio13
ENSCAFG00000055172	NCBP2	Bio13	Hfpi
ENSCAFG00000013100	NCBP2 PIGZ	Bio13	Hfpi
ENSCAFG00000018753	NDUFA11	Bio19	Bio8
ENSCAFG00000005667	NEK10	Bio13	Bio8
ENSCAFG00000011330	NEK7	Bio13	Bio8
ENSCAFG00000009780	NEURL2	Bio13	Bio13
ENSCAFG00000012103	NGRN	Bio13	Bio2
ENSCAFG00000018695	NIPBL	Bio13	Bio13
ENSCAFG00000058583	NKAPL	Bio13	Bio7
ENSCAFG00000002501	NLRP5	Bio13	Elev
ENSCAFG00000002256	NOA1	Bio14	Bio13
ENSCAFG00000029214	NPB	Elev	Bio13
ENSCAFG00000014331	NR2C2AP	Bio14	Bio13
ENSCAFG00000007130	NSMAF	Bio13	Bio8
ENSCAFG00000047342	NSMCE2	Bio13	Bio19
ENSCAFG00000018404	NUDT14	Bio13	Shrubs
ENSCAFG00000015771	ODAD4	Bio19	Bio2
ENSCAFG00000029096	OR13C7G	Bio19	Bio19
ENSCAFG00000031867	OR13C7H	Bio19	Bio19
ENSCAFG00000050743	OR6C38	Bio14, Bio19, Hfpi	Bio8
ENSCAFG00000049942	OR6C63	Bio14, Bio19, Hfpi	Bio8
ENSCAFG00000054648	OR7A23	Bio14	Bio19
ENSCAFG00000058847	OR7A55	Bio14	Bio19
ENSCAFG00000041401	OR7A55 OR7A58	Bio14	Bio19
ENSCAFG00000041284	OR7A60	Bio14	Bio19
ENSCAFG00000050877	OR7A63	Bio14	Bio19
ENSCAFG00000051510	OR7A64	Bio14	Bio19
ENSCAFG00000011159	OXA1L	Bio14	Bio19
ENSCAFG00000007113	PCM1	Elev	Bio13
ENSCAFG00000013184	PDCD1	Bio7	Bio8
ENSCAFG00000002084	Peptidyl-prolyl cis-trans isomerase (PPIase) (EC 5.2.1.8)	Elev	Bio13
ENSCAFG00000014968	Peptidyl-prolyl cis-trans isomerase (PPIase) (EC 5.2.1.8)	Bio13	Bio13
ENSCAFG00000024344	PGB2 PGB1	Bio14	Trees

ENSCAFG00000013050	PIGX	Bio13	Hfpi
ENSCAFG00000054639	PKD1L1	Bio14	Bio7
ENSCAFG00000013089	PLBD1	Bio13	Bio8
ENSCAFG00000059374	PLD4	Bio13	Shrubs
ENSCAFG00000012547	PLEKHA1	Bio13	Bio14
ENSCAFG00000015323	PLIN4	Bio13	Hfpi
ENSCAFG00000019018	PLIN5	Bio13	Hfpi
ENSCAFG00000017423	PLPPR2	Bio14	Bio7
ENSCAFG00000009837	PLTP	Bio13	Bio13
ENSCAFG00000009351	PPP1R14D	Bio14	Elev
ENSCAFG00000012140	PRC1	Bio13	Bio2
ENSCAFG00000020319	PRKACB	Bio14	Bio19
ENSCAFG00000017392	PRKCSH	Bio14	Bio7
ENSCAFG00000031979	PRKG1	Bio13	Bio19
ENSCAFG00000011243	PRMT5	Bio14	Bio19
ENSCAFG00000014984	processed_pseudogene	Bio14	Bio13
ENSCAFG00000053426	Protein kinase domain-containing protein	Bio14	Bio13
ENSCAFG00000031425	PRPSAP1	Elev	Hfpi
ENSCAFG00000041221	pseudogene	Bio14, Bio19, Hfpi	Bio19
ENSCAFG00000041670	pseudogene	Bio14	Elev
ENSCAFG00000043307	pseudogene	Bio14	Bio8
ENSCAFG00000044330	pseudogene	Bio14	Bio8
ENSCAFG00000047122	pseudogene	Bio14	Bio13
ENSCAFG00000048084	pseudogene	Bio19	Crops
ENSCAFG00000056490	pseudogene	Bio14	Bio19
ENSCAFG00000007584	PTCD3	Bio13	Bio2
ENSCAFG00000051663	PTP4A2	Bio19	Bio8
ENSCAFG00000023742	PTPRR	Bio13	Bio13
ENSCAFG00000002035	PUM3	Bio13	Bio13
ENSCAFG00000005063	QRICH2	Elev	Hfpi
ENSCAFG00000015629	RAB5C	Bio19	Bio2
ENSCAFG00000009220	RAD51	Bio14	Elev
ENSCAFG00000047285	RALGAPA1	Bio14	Hfpi
ENSCAFG00000018739	RANBP3	Bio19	Bio8
ENSCAFG00000008919	RASGEF1B	Bio13	Bio13
ENSCAFG00000011217	RBM23	Bio14	Bio19
ENSCAFG00000009922	RBMS1	Bio8	Bio8
ENSCAFG00000018077	RCN2	Bio14	Bio13
ENSCAFG00000007549	REEP1	Bio13	Bio2
ENSCAFG00000002254	REST	Bio14	Bio13
ENSCAFG00000005933	RFNG	Elev	Bio13
ENSCAFG00000002141	RHAG	Bio14	Bio13
ENSCAFG00000029187	RHOV	Bio14	Elev
ENSCAFG00000058812	RIC8B	Elev	Bio8
ENSCAFG00000041062	RNA-directed DNA polymerase (EC 2.7.7.49)	Bio14	Grass
ENSCAFG00000015786	ROM1	Bio13	Bio13

ENSCAFG00000013811	RPE	Bio13	Bio13
ENSCAFG00000052558	RPL36	Bio19	Bio8
ENSCAFG00000041560	RRM domain-containing protein	Bio19	Bio13
ENSCAFG00000021015	rRNA	Bio13	Bio13
ENSCAFG00000007313	RRS1	Bio14	Bio13
ENSCAFG00000012672	SCFD1 G2E3	Bio19	Bio19
ENSCAFG00000017690	SCNN1G	Bio14	Bio14
ENSCAFG00000000907	SCUBE1	Bio13	Bio13
ENSCAFG00000019014	SEMA6B	Bio13	Hfpi
ENSCAFG00000013266	SENP2	Bio14	Bio19
ENSCAFG00000000989	SERINC1	Bio13	Bio13
ENSCAFG00000054384	SFT2D3	Bio13	Bio8
ENSCAFG00000019109	SHD	Bio13	Hfpi
ENSCAFG00000003008	SHOC1	Bio14, Bio19	Bio13
ENSCAFG00000011908	SHTN1	Bio14	Hfpi
ENSCAFG00000019730	SLC25A33	Bio7	Bio8
ENSCAFG00000012702	SLC30A2	Bio14	Bio13
ENSCAFG00000001798	SLC30A9	Bio14	Bio8
ENSCAFG00000003063	SLC46A2	Bio13	Bio13
ENSCAFG00000015723	SLC4A3	Bio19	Bio13
ENSCAFG00000023919	SLC8A3	Bio14	Bio13
ENSCAFG00000042570	Small integral membrane protein 27	Bio13	Bio19
ENSCAFG00000042783	Small ribosomal subunit protein eS6 (40S ribosomal protein S6)	Bio19	Bio7
ENSCAFG00000008306	SMARCD1	Bio13	Bio14
ENSCAFG00000000522	SNCAIP	Bio19	Bio8
ENSCAFG00000020981	snoRNA	Bio14	Bio13
ENSCAFG00000026284	snoRNA	Bio14	Bio13
ENSCAFG00000026468	snoRNA	Bio13	Bio13
ENSCAFG00000026833	snoRNA	Bio13	Bio2
ENSCAFG00000050039	snoRNA	Bio14	Shrubs
ENSCAFG00000020750	snRNA	Bio13	Bio13
ENSCAFG00000020815	snRNA	Bio13	Elev
ENSCAFG00000021082	snRNA	Bio19	Bio8
ENSCAFG00000021107	snRNA	Bio13	Bio2
ENSCAFG00000021220	snRNA	Bio14	Bio13
ENSCAFG00000021521	snRNA	Bio13	Bio19
ENSCAFG00000022254	snRNA	Elev	Bio8
ENSCAFG00000022493	snRNA	Bio19	Bio13
ENSCAFG00000022541	snRNA	Bio13	Bio13
ENSCAFG00000026930	snRNA	Bio19	Bio19
ENSCAFG00000026988	snRNA	Bio13	Bio2
ENSCAFG00000027148	snRNA	Bio13	Bio13
ENSCAFG00000028004	snRNA	Bio13	Elev
ENSCAFG00000043087	snRNA	Bio14	Bio13
ENSCAFG00000043782	snRNA	Bio14	Bio13

ENSCAFG00000044164	snRNA	Elev	Bio8
ENSCAFG00000046457	snRNA	Bio13	Bio13
ENSCAFG00000055381	snRNA	Bio14	Bio13
ENSCAFG00000055888	snRNA	Bio13	Bio19
ENSCAFG00000057912	snRNA	Bio19	Bio19
ENSCAFG00000059057	snRNA	Bio13	Bio13
ENSCAFG00000009614	SOX13	Bio13	Bio13
ENSCAFG00000009778	SPATA25	Bio13	Bio13
ENSCAFG00000009730	SPINT4	Bio13	Bio13
ENSCAFG00000013331	SRP54	Bio13	Bio13
ENSCAFG00000004887	ST3GAL3	Bio14	Elev
ENSCAFG00000011347	STAU1	Bio13	Bio8
ENSCAFG00000015687	STK11IP	Bio13	Bio2
ENSCAFG00000012838	STX6	Bio13	Bio13
ENSCAFG00000058807	SWSAP1	Bio14	Bio7
ENSCAFG00000005592	SYCN	Bio14, Bio19, Hfpi	Bio7
ENSCAFG00000004500	SYTL2	Bio13	Bio19
ENSCAFG00000011291	TBCE	Bio19	Bio13
ENSCAFG00000013010	TBL2 BCL7B MLXIPL	Bio19	Bio13
ENSCAFG00000053610	TICAM2	Bio14	Bio13
ENSCAFG00000008351	TLR2	Bio13	Bio2
ENSCAFG00000000209	TMED7	Bio14	Bio13
ENSCAFG00000004520	TMEM126B	Bio13	Bio19
ENSCAFG00000014369	TMEM161A	Bio14	Bio13
ENSCAFG00000015617	TMEM198	Bio13	Bio2
ENSCAFG00000019733	TMEM201	Bio7	Bio8
ENSCAFG00000050707	TMEM263	Elev	Bio8
ENSCAFG00000018152	TMEM266	Bio13	Bio7
ENSCAFG00000005006	TMEM91 BCKDHA	Bio13	Bio13
ENSCAFG00000024456	Transcription factor BTF3	Bio19	Bio14
ENSCAFG00000001077	TRIB1	Bio13	Bio19
ENSCAFG00000017047	TRIM46	Bio14	Bio8
ENSCAFG00000012671	TRIM63	Bio14	Bio13
ENSCAFG00000018068	TSPAN3	Bio14	Bio13
ENSCAFG00000000909	TTLL12	Bio13	Bio13
ENSCAFG00000004373	TYR	Bio13	Bio13
ENSCAFG00000009745	UBE2C	Bio13	Bio13
ENSCAFG00000016599	UBIAD1	Bio13	Bio19
ENSCAFG00000015710	UBXN1	Bio13	Bio13
ENSCAFG00000059211	UBXN2B	Bio13	Bio8
ENSCAFG00000019059	UBXN6	Bio13	Hfpi
ENSCAFG00000012179	UNC45A	Bio13	Bio2
ENSCAFG00000013772	UNC80	Bio13	Bio13
ENSCAFG00000043988	Uncharacterized protein	Bio13	Bio19
ENSCAFG00000047606	Uncharacterized protein	Bio13	Bio13
ENSCAFG00000049104	Uncharacterized protein	Bio13	Bio13

ENSCAFG00000051605	Uncharacterized protein	Elev	Bio13
ENSCAFG00000055022	Uncharacterized protein	Bio13	Bio19
ENSCAFG00000058607	Uncharacterized protein	Bio13	Bio13
ENSCAFG00000020315	UOX	Bio14	Bio19
ENSCAFG00000002028	VLDLR	Bio13	Bio13
ENSCAFG00000012113	VPS33B	Bio13	Bio2
ENSCAFG00000002018	VPS50	Bio13	Hfpi
ENSCAFG00000004326	WDR33	Bio13	Bio8
ENSCAFG00000013149	WHAMM	Bio13	Bio7
ENSCAFG00000031321	WNT6	Bio14	Bio13
ENSCAFG00000012781	XPR1	Bio13	Bio13
ENSCAFG00000021643	Y_RNA	Bio13	Elev
ENSCAFG00000000224	YTHDC2	Bio14	Bio13
ENSCAFG00000015650	ZBTB3	Bio13	Bio13
ENSCAFG00000046247	Zinc finger protein 671	Bio14, Bio19, Hfpi	Bio19
ENSCAFG00000056037	ZKSCAN8	Bio13	Bio7
ENSCAFG00000009882	ZNF335	Bio13	Bio13
ENSCAFG00000024225	ZNF444	Bio13	Elev
ENSCAFG00000002996	ZNF483	Bio14, Bio19	Bio13
ENSCAFG00000031712	ZNF653	Bio14	Bio7
ENSCAFG00000017814	ZP2	Bio13	Bio13
ENSCAFG00000056901	ZSCAN31	Bio13	Bio7
Upstream variants			
ENSCAFG00000011583	ABHD2	Bio13	Bio8
ENSCAFG00000007394	ABHD6	Bio13	Bio13
ENSCAFG00000015818	ACLY	Bio19	Bio2
ENSCAFG00000015933	Acylphosphatase-1 (Acylphosphate phosphohydrolase 1)	Bio13	Bio13
ENSCAFG00000010872	ADGRB2	Bio19	Bio8
ENSCAFG00000007326	ADHFE1	Bio14	Bio13
ENSCAFG00000042815	ADP/ATP translocase (ADP,ATP carrier protein)	Bio13	Bio13
ENSCAFG00000010465	AGFG1	Bio19	Crops
ENSCAFG00000055880	AHNAK	Bio13	Bio13
ENSCAFG00000058987	AHNAK nucleoprotein 2	Bio13	Shrubs
ENSCAFG00000016466	ALDH1A2	Bio14	Bio19
ENSCAFG00000005877	ALYREF	Elev	Bio13
ENSCAFG00000008821	ANKRD34B	Bio13	Bio8
ENSCAFG00000003105	ANKRD6	Bio13	Elev
ENSCAFG00000017810	ANKS4B	Bio13	Bio13
ENSCAFG00000016556	AQP9	Bio14	Bio19
ENSCAFG00000013922	ARNT2	Bio13	Bio13
ENSCAFG00000007324	ASCL1	Bio13	Bio19
ENSCAFG00000015783	B3GAT3	Bio13	Bio13
ENSCAFG00000054129	B3GNT8	Bio13	Bio13
ENSCAFG00000054288	BED-type domain-containing protein	Bio13	Bio19
ENSCAFG00000054718	BIK	Bio13	Bio13

ENSCAFG00000059340	C4H10orf105	Bio7	Bio8
ENSCAFG00000018949	C8A	Bio13	Bio19
ENSCAFG00000007104	CAPRIN1	Bio13	Hfpi
ENSCAFG00000018746	CAPS	Bio19	Bio8
ENSCAFG00000000481	CAPS2	Bio14, Bio19, Hfpi	Bio8
ENSCAFG00000055058	CAV1 CAV	Bio14, Bio19, Hfpi	Bio19
ENSCAFG00000003402	CAV2 CAV-2	Bio14, Bio19, Hfpi	Bio19
ENSCAFG00000011882	CBLN4	Bio13	Bio13
ENSCAFG00000002089	CD2AP	Elev	Bio13
ENSCAFG00000008962	CDCA2	Bio13	Bio13
ENSCAFG00000018398	CDCA4	Bio13	Shrubs
ENSCAFG00000014996	CDK5R2	Bio14	Bio13
ENSCAFG00000004706	CDNF	Bio19	Bio7
ENSCAFG00000008262	CERS5	Bio13	Bio14
ENSCAFG00000009674	CHAF1B	Bio13	Bio13
ENSCAFG00000045731	CHPF	Bio13	Bio2
ENSCAFG00000024101	CHST8	Bio13	Bio13
ENSCAFG00000012099	CIB1	Bio13	Bio2
ENSCAFG00000000285	CITED2	Bio13	Bio2
ENSCAFG00000018136	CLEC19A	Bio14	Bio2
ENSCAFG00000016969	CLK2	Bio14	Bio8
ENSCAFG00000017326	CNN1	Bio14	Bio7
ENSCAFG00000011260	CSE1L	Bio13	Bio8
ENSCAFG00000029552	CSKMT	Bio13	Bio13
ENSCAFG00000009819	CTSA	Bio13	Bio13
ENSCAFG00000054560	Cyclin-dependent kinases regulatory subunit	Bio14	Bio14
ENSCAFG00000007098	CYP7A1	Bio13	Bio8
ENSCAFG00000005161	DARS1	Bio13	Bio13
ENSCAFG00000019764	DENND2D	Bio14	Trees
ENSCAFG00000032036	DHFR	Bio13	Bio8
ENSCAFG00000011736	DISC1	Bio13	Bio8
ENSCAFG00000016596	DISP3	Bio13	Bio19
ENSCAFG00000032222	DLA-12	Bio14	Bio13
ENSCAFG00000000487	DLA88	Bio14	Bio13
ENSCAFG00000004544	DLG2	Bio13	Bio19
ENSCAFG00000012561	DMBT1	Bio13	Bio14
ENSCAFG00000015749	DNAJC7	Bio19	Bio2
ENSCAFG00000007419	DNASE1L3	Bio13	Trees
ENSCAFG00000020313	DNASE2B	Bio14	Bio19
ENSCAFG00000055333	DTYMK	Bio7	Bio8
ENSCAFG00000018762	DUS3L	Bio19	Bio8
ENSCAFG00000002990	ECPAS	Bio14, Bio19	Bio13
ENSCAFG00000048121	EDDM13	Bio13	Elev
ENSCAFG00000041466	EF-hand domain-containing protein	Bio13	Bio19

ENSCAFG00000031582	EHF	Bio14	Bio13
ENSCAFG00000006935	ELF5	Bio14	Bio13
ENSCAFG00000015795	EML3	Bio13	Bio13
ENSCAFG00000011891	ENO4	Bio14	Hfpi
ENSCAFG00000016980	ENTREP3	Bio14	Bio8
ENSCAFG00000017411	EPOR	Bio14	Bio7
ENSCAFG00000030737	ERICH4	Bio13	Bio13
ENSCAFG00000009619	ETNK2	Bio13	Bio13
ENSCAFG00000008978	EVA1C	Bio13	Crops
ENSCAFG00000013402	EVX2	Bio13	Bio2
ENSCAFG00000005043	EXOC7	Elev	Hfpi
ENSCAFG00000032654	EXOSC5	Bio13	Bio13
ENSCAFG00000004714	FAM107B	Bio19	Bio7
ENSCAFG00000024310	FBXO27	Bio14, Bio19, Hfpi	Bio7
ENSCAFG00000030749	FDPS	Bio14	Bio8
ENSCAFG00000000210	FEM1C	Bio14	Bio13
ENSCAFG00000005050	FOXJ1	Elev	Hfpi
ENSCAFG00000018405	FST	Bio13	Elev
ENSCAFG00000003099	GABRR2	Bio13	Elev
ENSCAFG00000051120	GAL3ST2	Bio7	Bio8
ENSCAFG00000029589	GALP	Bio13	Elev
ENSCAFG00000015767	GANAB	Bio13	Bio13
ENSCAFG00000041383	GATD3	Bio13	Bio13
ENSCAFG00000016988	GBA1	Bio14	Bio8
ENSCAFG00000017520	GLCE	Bio14	Bio13
ENSCAFG00000000525	GLIPR1	Bio14, Bio19, Hfpi	Bio8
ENSCAFG00000019322	GNG10 LOC119863873	Bio14, Bio19	Bio13
ENSCAFG00000014643	GNG3	Bio13	Bio13
ENSCAFG00000008985	GNRH1	Bio13	Bio13
ENSCAFG00000009628	GOLT1A	Bio13	Bio13
ENSCAFG00000005938	GPS1	Elev	Bio13
ENSCAFG00000000582	GRAMD2B	Bio13	Bio19
ENSCAFG00000019928	GSE1	Bio13	Bio19
ENSCAFG00000007858	HAND2	Bio13	Bio8
ENSCAFG00000011542	HAPLN3 MFGE8	Bio13	Bio8
ENSCAFG00000012239	HDDC3	Bio13	Bio2
ENSCAFG00000019033	HDGFL2	Bio13	Hfpi
ENSCAFG00000005378	HGSNAT	Bio13	Bio13
ENSCAFG00000001711	HMOX1	Elev	Bio2
ENSCAFG00000056470	Homeobox domain-containing protein	Bio13	Bio2
ENSCAFG00000031345	HOXD1	Bio13	Bio2
ENSCAFG00000013447	HOXD10	Bio13	Bio2
ENSCAFG00000013429	HOXD12	Bio13	Bio2
ENSCAFG00000055834	HOXD13	Bio13	Bio2

ENSCAFG00000013464	HOXD3	Bio13	Bio2
ENSCAFG00000013452	HOXD4	Bio13	Bio2
ENSCAFG00000013450	HOXD8	Bio13	Bio2
ENSCAFG00000011876	HSPA12A	Bio14	Hfpi
ENSCAFG00000055109	HTD2	Bio13	Trees
ENSCAFG00000014478	Ig-like domain-containing protein	Bio14, Bio19, Hfpi	Shrubs
ENSCAFG00000045242	Ig-like domain-containing protein	Bio14	Bio13
ENSCAFG00000054905	Ig-like domain-containing protein	Bio14	Bio13
ENSCAFG00000055775	Ig-like domain-containing protein	Bio14	Bio13
ENSCAFG00000056839	Ig-like domain-containing protein	Bio14, Bio19, Hfpi	Shrubs
ENSCAFG00000057398	Ig-like domain-containing protein	Bio14	Bio13
ENSCAFG00000058699	Ig-like domain-containing protein	Bio14	Bio13
ENSCAFG00000002145	IL1RL2	Bio13	Bio13
ENSCAFG00000004008	IL2	Bio14	Bio13
ENSCAFG00000058265	Immunoglobulin C1-set domain-containing protein	Bio14, Bio19, Hfpi	Shrubs
ENSCAFG00000023933	Immunoglobulin V-set domain-containing protein	Bio13	Bio13
ENSCAFG00000031359	Immunoglobulin V-set domain-containing protein	Bio14, Bio19, Hfpi	Shrubs
ENSCAFG00000043031	Immunoglobulin V-set domain-containing protein	Bio14, Bio19, Hfpi	Shrubs
ENSCAFG00000045360	Immunoglobulin V-set domain-containing protein	Bio14, Bio19, Hfpi	Shrubs
ENSCAFG00000050263	Immunoglobulin V-set domain-containing protein	Bio14	Bio13
ENSCAFG00000002768	IMPG1	Bio13	Bio13
ENSCAFG00000003054	INIP	Bio13	Bio2
ENSCAFG00000012205	INPP5F	Bio13	Bio8
ENSCAFG00000041829	Interferon lambda-4-like	Bio14, Bio19, Hfpi	Bio7
ENSCAFG00000009877	ITGB6	Bio8	Bio8
ENSCAFG00000011056	ITPK1	Bio14	Bio13
ENSCAFG00000010259	KCNJ1	Bio14	Bio19
ENSCAFG00000010255	KCNJ5	Bio14	Bio19
ENSCAFG00000002032	KCNV2	Bio13	Bio13
ENSCAFG00000012826	KIAA1614	Bio13	Bio13
ENSCAFG00000028492	Kisspeptin	Bio13	Bio13
ENSCAFG00000004819	KLF17	Bio14	Elev
ENSCAFG00000017055	KRTCAP2 KCP2	Bio14	Bio8
ENSCAFG00000001115	L3MBTL3	Bio14, Bio19, Hfpi	Bio8
ENSCAFG00000030771	LCLAT1	Bio13	Bio14
ENSCAFG00000003855	LDAH	Bio13	Bio19
ENSCAFG00000020277	LHX6	Bio13	Bio8
ENSCAFG00000004355	LIMS2	Bio13	Bio8
ENSCAFG00000016108	LIN9	Bio13	Bio13
ENSCAFG00000007127	LMO2	Bio13	Hfpi
ENSCAFG00000038294	lncRNA	Bio13	Bio2

ENSCAFG00000047324	lncRNA	Bio14	Bio8
ENSCAFG00000048403	lncRNA	Elev	Hfpi
ENSCAFG00000050237	lncRNA	Bio13	Bio8
ENSCAFG00000050273	lncRNA	Bio13	Bio8
ENSCAFG00000050366	lncRNA	Bio14	Bio13
ENSCAFG00000050470	lncRNA	Bio13	Bio8
ENSCAFG00000050476	lncRNA	Bio14, Bio19, Hfpi	Bio8
ENSCAFG00000050516	lncRNA	Bio13	Bio13
ENSCAFG00000050555	lncRNA	Bio13	Bio8
ENSCAFG00000050789	lncRNA	Bio13	Bio13
ENSCAFG00000050929	lncRNA	Bio13	Bio13
ENSCAFG00000051070	lncRNA	Bio14	Bio13
ENSCAFG00000051217	lncRNA	Elev	Bio8
ENSCAFG00000051272	lncRNA	Bio13	Bio13
ENSCAFG00000051302	lncRNA	Bio13	Elev
ENSCAFG00000051395	lncRNA	Bio13	Bio2
ENSCAFG00000051452	lncRNA	Bio13	Bio7
ENSCAFG00000051637	lncRNA	Bio14	Bio13
ENSCAFG00000051678	lncRNA	Bio13	Bio7
ENSCAFG00000051921	lncRNA	Bio14	Bio13
ENSCAFG00000051945	lncRNA	Bio14	Bio2
ENSCAFG00000052308	lncRNA	Bio14	Bio13
ENSCAFG00000052536	lncRNA	Bio13	Bio13
ENSCAFG00000052713	lncRNA	Bio14	Bio8
ENSCAFG00000052727	lncRNA	Bio14	Bio13
ENSCAFG00000052891	lncRNA	Bio13	Bio7
ENSCAFG00000052983	lncRNA	Bio19	Bio13
ENSCAFG00000053107	lncRNA	Bio13	Bio2
ENSCAFG00000053379	lncRNA	Bio13	Bio2
ENSCAFG00000053720	lncRNA	Bio14	Bio13
ENSCAFG00000053959	lncRNA	Bio14	Bio13
ENSCAFG00000054033	lncRNA	Bio14, Bio19, Hfpi	Bio8
ENSCAFG00000054065	lncRNA	Bio14	Bio13
ENSCAFG00000054112	lncRNA	Bio14	Bio13
ENSCAFG00000054204	lncRNA	Bio13	Bio19
ENSCAFG00000054466	lncRNA	Bio14	Bio19
ENSCAFG00000054563	lncRNA	Bio13	Shrubs
ENSCAFG00000054739	lncRNA	Elev	Bio13
ENSCAFG00000054760	lncRNA	Bio13	Bio13
ENSCAFG00000054974	lncRNA	Bio13	Bio19
ENSCAFG00000055031	lncRNA	Bio14	Bio8
ENSCAFG00000055124	lncRNA	Elev	Bio13
ENSCAFG00000055236	lncRNA	Bio19	Crops
ENSCAFG00000055254	lncRNA	Bio13	Bio13

ENSCAFG00000055787	lncRNA	Bio14, Bio19	Bio13
ENSCAFG00000055829	lncRNA	Bio13	Bio19
ENSCAFG00000055840	lncRNA	Bio13	Bio14
ENSCAFG00000055908	lncRNA	Bio19	Bio8
ENSCAFG00000056124	lncRNA	Bio13	Bio19
ENSCAFG00000056373	lncRNA	Bio13	Bio13
ENSCAFG00000056504	lncRNA	Bio14, Bio19, Hfpi	Shrubs
ENSCAFG00000056699	lncRNA	Bio14	Bio8
ENSCAFG00000056829	lncRNA	Bio13	Bio2
ENSCAFG00000056859	lncRNA	Bio14	Bio13
ENSCAFG00000056870	lncRNA	Bio13	Bio13
ENSCAFG00000057012	lncRNA	Bio19	Bio8
ENSCAFG00000057071	lncRNA	Bio14	Bio13
ENSCAFG00000057210	lncRNA	Bio13	Bio19
ENSCAFG00000057424	lncRNA	Bio13	Bio8
ENSCAFG00000057445	lncRNA	Bio14	Bio13
ENSCAFG00000057488	lncRNA	Bio14	Bio13
ENSCAFG00000057876	lncRNA	Bio19	Bio8
ENSCAFG00000057924	lncRNA	Bio13	Bio2
ENSCAFG00000057961	lncRNA	Bio14, Bio19, Hfpi	Bio19
ENSCAFG00000058208	lncRNA	Elev	Bio13
ENSCAFG00000058360	lncRNA	Bio13	Bio14
ENSCAFG00000058444	lncRNA	Bio19	Bio14
ENSCAFG00000058491	lncRNA	Bio19	Bio8
ENSCAFG00000058674	lncRNA	Bio13	Bio19
ENSCAFG00000058719	lncRNA	Bio13	Bio13
ENSCAFG00000058839	lncRNA	Bio13	Bio19
ENSCAFG00000058900	lncRNA	Bio13	Bio2
ENSCAFG00000059316	lncRNA	Bio13	Bio2
ENSCAFG00000059322	lncRNA	Bio13	Bio2
ENSCAFG00000059330	lncRNA	Elev	Bio8
ENSCAFG00000013390	LNPK	Bio13	Bio2
ENSCAFG00000057532	LOC100683951	Bio13	Bio19
ENSCAFG00000024969	LOC100685048	Bio13	Bio7
ENSCAFG00000043938	LOC100686431	Bio13	Elev
ENSCAFG00000058042	LOC102152102	Bio14, Bio19, Hfpi	Bio19
ENSCAFG00000032035	LOC102153214	Bio13	Bio13
ENSCAFG00000053510	LOC111093559	Bio14	Bio8
ENSCAFG00000054991	LOC119880636	Elev	Bio8
ENSCAFG00000053412	LOC119881374	Bio14, Bio19	Bio14
ENSCAFG00000002805	LOC482174	Bio14, Bio19, Trees	Bio13

ENSCAFG00000002808	LOC482175	Bio14, Bio19, Trees	Bio13
ENSCAFG00000048824	LOC482215	Bio14	Bio13
ENSCAFG00000055528	LOC482217	Bio14	Bio13
ENSCAFG00000013321	LOC608513	Bio13	Bio13
ENSCAFG00000002133	LOC610105	Bio14	Bio13
ENSCAFG00000055330	LOC610725	Bio13	Bio2
ENSCAFG00000018775	LONP1	Bio19	Bio8
ENSCAFG00000014607	LRPAP1	Bio13	Shrubs
ENSCAFG00000005915	LRRC45	Elev	Bio13
ENSCAFG00000031109	LSM domain-containing protein	Bio13	Hfpi
ENSCAFG00000000515	LTA TNFB TNFSF1	Bio14	Bio13
ENSCAFG00000005893	MAFG	Elev	Bio13
ENSCAFG00000013740	MAP2	Bio13	Bio2
ENSCAFG00000011887	MC3R	Bio13	Bio13
ENSCAFG00000000506	MCCD1	Bio14	Bio13
ENSCAFG00000001709	MCM5	Elev	Bio2
ENSCAFG00000030738	MEIG1	Bio19	Bio7
ENSCAFG00000020468	miRNA	Bio13	Bio2
ENSCAFG00000025700	miRNA	Bio19	Bio14
ENSCAFG00000046125	miRNA	Bio13	Bio7
ENSCAFG00000050992	miRNA	Bio13	Bio8
ENSCAFG00000053766	miRNA	Bio14	Elev
ENSCAFG00000008878	MIS18A	Bio13	Crops
ENSCAFG00000000033	Mitochondrial carrier 2	Bio14	Bio8
ENSCAFG00000006533	MMP17	Bio13	Bio8
ENSCAFG00000015066	MMP27	Bio14	Bio2
ENSCAFG00000023335	MMP8	Bio14	Bio2
ENSCAFG00000002111	MMUT	Bio14	Bio13
ENSCAFG00000015639	MNAT1	Bio19	Bio8
ENSCAFG00000009660	MORC3	Bio13	Bio13
ENSCAFG00000007553	MRPL35	Bio13	Bio2
ENSCAFG00000015809	MTA2	Bio13	Bio13
ENSCAFG00000016648	MTOR	Bio13	Bio19
ENSCAFG00000017037	MUC1	Bio14	Bio8
ENSCAFG00000005911	MYADML2	Elev	Bio13
ENSCAFG00000014218	NCAN	Bio14	Bio13
ENSCAFG00000005593	NCCRP1	Bio14, Bio19, Hfpi	Bio7
ENSCAFG00000053123	NDUFB6	Bio13	Bio19
ENSCAFG00000052422	NDUFS4	Bio13	Elev
ENSCAFG00000009014	NEFL	Bio13	Bio13
ENSCAFG00000045576	NEU4	Bio7	Bio8
ENSCAFG00000030599	NFKBIL1	Bio14	Bio13
ENSCAFG00000005913	NOTUM	Elev	Bio13
ENSCAFG00000017808	NPTN	Bio19	Bio13

ENSCAFG00000059367	NPTX1	Bio13	Shrubs
ENSCAFG00000015641	OBSL1	Bio13	Bio2
ENSCAFG00000017400	ODAD3	Bio14	Bio7
ENSCAFG00000056577	OR13C11C	Bio19	Bio19
ENSCAFG00000053326	OR13C7F	Bio19	Bio19
ENSCAFG00000032491	OR13C7I	Bio19	Bio19
ENSCAFG00000023939	OR13J1	Bio19	Bio19
ENSCAFG00000041843	OR2AJ7	Bio14	Bio13
ENSCAFG00000024736	OR2W3	Bio14	Bio13
ENSCAFG00000057285	OR7A57	Bio14	Bio19
ENSCAFG00000018648	OSMR	Bio13	Bio8
ENSCAFG00000000936	PACSIN2	Bio13	Bio13
ENSCAFG00000012726	PAFAH2	Bio14	Bio13
ENSCAFG00000009847	PCIF1	Bio13	Bio13
ENSCAFG00000005882	PCYT2	Elev	Bio13
ENSCAFG00000009253	PDE8B	Bio13	Bio19
ENSCAFG00000057857	PDIK1L	Bio14	Bio13
ENSCAFG00000050625	PDZ domain-containing protein	Bio13	Shrubs
ENSCAFG00000007377	PEPD	Bio13	Bio13
ENSCAFG00000041427	PHF19	Bio13	Bio13
ENSCAFG00000001795	PHOX2B	Bio13	Bio14
ENSCAFG00000019735	PIK3CD	Bio7	Bio8
ENSCAFG00000020056	PLPPR4	Bio2	Shrubs
ENSCAFG00000007613	POLR1A	Bio13	Bio2
ENSCAFG00000002296	POLR2B	Bio14	Bio13
ENSCAFG00000015461	POLR2G	Bio13	Bio13
ENSCAFG00000052530	POMK	Bio13	Bio13
ENSCAFG00000016551	Prefoldin subunit 4	Bio14	Bio19
ENSCAFG00000014957	PRKAG3	Bio14	Bio13
ENSCAFG00000008898	PRKG2	Bio13	Bio13
ENSCAFG00000018782	PRLR	Bio13	Bio13
ENSCAFG00000013391	PRORP	Bio13	Bio13
ENSCAFG00000024169	PRR22	Bio19	Bio8
ENSCAFG00000014481	PRSS2	Bio14, Bio19, Hfpi	Shrubs
ENSCAFG00000001570	PSAT1	Bio19, Elev	Bio19
ENSCAFG00000043459	pseudogene	Bio14	Bio8
ENSCAFG00000045749	pseudogene	Bio13	Bio13
ENSCAFG00000048074	pseudogene	Bio14	Bio2
ENSCAFG00000049146	pseudogene	Bio14	Bio8
ENSCAFG00000049284	pseudogene	Bio14	Bio8
ENSCAFG00000003574	PSMD5	Bio13	Bio13
ENSCAFG00000003000	PTGR1	Bio14, Bio19	Bio13
ENSCAFG00000020263	PTGS1 COX1	Bio13	Bio8
ENSCAFG00000000446	PTPRB	Bio13	Bio13
ENSCAFG00000010687	PWP2	Bio13	Bio13

ENSCAFG00000007367	PXK	Bio13	Trees
ENSCAFG00000005906	PYCR1	Elev	Bio13
ENSCAFG00000055546	RAB11A RAB11	Bio13	Bio8
ENSCAFG00000017438	RAB3D	Bio14	Bio7
ENSCAFG00000050761	RAC3	Elev	Bio13
ENSCAFG00000013181	RAMAC	Bio13	Bio7
ENSCAFG00000009109	RBM12B	Bio13	Bio8
ENSCAFG00000030041	RBM18	Bio13	Bio8
ENSCAFG00000002281	RECK CLTA	Bio19	Bio19
ENSCAFG00000009623	REN	Bio13	Bio13
ENSCAFG00000019130	RFLNB	Bio13	Bio13
ENSCAFG00000014341	RFXANK	Bio14	Bio13
ENSCAFG00000017405	RGL3	Bio14	Bio7
ENSCAFG00000050536	RHOH	Bio13	Bio13
ENSCAFG00000053722	Ribosomal protein L37a	Bio13	Elev
ENSCAFG00000028119	ribozyme	Elev	Hfpi
ENSCAFG00000009237	RMDN3	Bio14	Elev
ENSCAFG00000042857	RNF175	Bio13	Bio2
ENSCAFG00000019132	RPH3AL	Bio13	Bio13
ENSCAFG00000052621	RRM domain-containing protein	Bio7	Bio8
ENSCAFG00000025969	rRNA	Bio13	Bio13
ENSCAFG00000013188	RTP5	Bio7	Bio8
ENSCAFG00000016954	RUSC1	Bio14	Bio8
ENSCAFG00000020317	SAMD13	Bio14	Bio19
ENSCAFG00000049822	SCAMP3	Bio14	Bio8
ENSCAFG00000018100	SCAPER	Bio14	Bio13
ENSCAFG00000017687	SCNN1B	Bio14	Bio14
ENSCAFG00000007112	SDCBP	Bio13	Bio8
ENSCAFG00000043332	Secreted protein	Elev	Bio13
ENSCAFG00000028589	SELENOP	Bio19	Bio13
ENSCAFG00000006212	SEMA3A	Bio13	Shrubs
ENSCAFG00000013094	SENP5	Bio13	Hfpi
ENSCAFG00000006625	SFSWAP	Bio13	Bio8
ENSCAFG00000019083	SH3GL1	Bio13	Hfpi
ENSCAFG00000057838	SIRT7	Elev	Bio13
ENSCAFG00000029797	SIX1	Bio19	Bio8
ENSCAFG00000002939	SLC4A4	Bio14	Bio13
ENSCAFG00000011177	SLC7A7	Bio14	Bio19
ENSCAFG00000005293	SLITRK5	Bio14	Bio8
ENSCAFG00000044478	Small ribosomal subunit protein uS15 N-terminal domain-containing protein	Bio14	Bio13
ENSCAFG00000020883	snoRNA	Bio13	Bio13
ENSCAFG00000026017	snoRNA	Bio13	Bio13
ENSCAFG00000026082	snoRNA	Bio14	Bio13
ENSCAFG00000027192	snoRNA	Bio19	Bio19
ENSCAFG00000027480	snoRNA	Bio7	Bio8
ENSCAFG00000027533	snoRNA	Bio13	Bio19

ENSCAFG00000042265	snoRNA	Bio13	Bio8
ENSCAFG00000042603	snoRNA	Bio13	Bio13
ENSCAFG00000047445	snoRNA	Bio7	Shrubs
ENSCAFG00000021530	snRNA	Bio13	Bio13
ENSCAFG00000022253	snRNA	Bio14	Bio13
ENSCAFG00000022520	snRNA	Bio14	Bio19
ENSCAFG00000027033	snRNA	Bio13	Shrubs
ENSCAFG00000027076	snRNA	Bio13	Bio13
ENSCAFG00000027239	snRNA	Bio13	Bio13
ENSCAFG00000027853	snRNA	Bio13, Elev	Bio13
ENSCAFG00000028111	snRNA	Bio7	Bio8
ENSCAFG00000029409	snRNA	Bio13	Bio8
ENSCAFG00000045431	snRNA	Bio13	Bio2
ENSCAFG00000046396	snRNA	Bio13	Bio13
ENSCAFG00000047244	snRNA	Bio13	Bio13
ENSCAFG00000049973	snRNA	Bio14	Elev
ENSCAFG00000051000	snRNA	Bio13	Bio13
ENSCAFG00000053965	snRNA	Bio14	Bio8
ENSCAFG00000055501	snRNA	Bio14	Bio8
ENSCAFG00000055792	snRNA	Bio14	Shrubs
ENSCAFG00000056745	snRNA	Bio13	Bio8
ENSCAFG00000044832	SNRPC	Bio13	Bio13
ENSCAFG00000009759	SNX21	Bio13	Bio13
ENSCAFG00000009365	SPINT1	Bio14	Elev
ENSCAFG00000010825	SPOCD1	Bio19	Bio8
ENSCAFG00000019095	STAP2	Bio13	Hfpi
ENSCAFG00000029191	STMN1	Bio14	Bio13
ENSCAFG00000014431	SUGP2	Bio14	Bio13
ENSCAFG00000043184	Superkiller complex protein 8 (WD repeat-containing protein 61)	Bio14	Bio13
ENSCAFG00000004696	SUV39H2	Bio19	Bio7
ENSCAFG00000015646	TAF6L	Bio13	Bio13
ENSCAFG00000002987	TBX18	Bio13	Bio2
ENSCAFG00000017500	TDP1	Bio7	Shrubs
ENSCAFG00000005014	TGFB1 B9D2	Bio13	Bio13
ENSCAFG00000013012	THAP4	Bio7	Bio8
ENSCAFG00000017017	THBS3	Bio14	Bio8
ENSCAFG00000030913	TMEM126A	Bio13	Bio19
ENSCAFG00000017434	TMEM205	Bio14	Bio7
ENSCAFG00000052645	TMEM215	Bio13	Bio19
ENSCAFG00000012301	TNS3	Bio14	Bio7
ENSCAFG00000001717	TOM1	Elev	Bio2
ENSCAFG00000001809	TOPORS	Bio13	Bio19
ENSCAFG00000009354	TOX2	Bio13	Bio13
ENSCAFG00000000913	TSPO	Bio13	Bio13
ENSCAFG00000015659	TTC9C	Bio13	Bio13
ENSCAFG00000000926	TTLL1	Bio13	Bio13

ENSCAFG00000015822	TUT1 EEF1G	Bio13	Bio13
ENSCAFG00000058205	UBALD2	Elev	Hfpi
ENSCAFG00000052580	UBE2J1	Bio13	Elev
ENSCAFG00000003684	UCP1	Bio14	Shrubs
ENSCAFG00000041808	Uncharacterized protein	Bio14	Bio13
ENSCAFG00000041820	Uncharacterized protein	Bio13	Bio8
ENSCAFG00000043214	Uncharacterized protein	Bio13	Bio13
ENSCAFG00000045033	Uncharacterized protein	Bio14	Bio19
ENSCAFG00000046775	Uncharacterized protein	Bio13	Bio19
ENSCAFG00000047905	Uncharacterized protein	Elev	Hfpi
ENSCAFG00000048629	Uncharacterized protein	Bio13	Bio8
ENSCAFG00000051647	Uncharacterized protein	Bio14	Bio13
ENSCAFG00000015715	UQCC3	Bio13	Bio13
ENSCAFG00000014354	VSIR	Bio7	Bio8
ENSCAFG00000016645	WASH complex subunit 3 (Coiled-coil domain-containing protein 53)	Bio14	Bio8
ENSCAFG00000031732	WFDC3	Bio13	Bio13
ENSCAFG00000014987	WNT10A	Bio14	Bio13
ENSCAFG00000052757	WNT4	Bio13	Elev
ENSCAFG00000026073	Y_RNA	Bio13	Bio13
ENSCAFG00000009345	ZFYVE19	Bio14	Elev
ENSCAFG00000028550	Zinc finger protein 850-like	Bio14, Bio19, Hfpi	Bio19
ENSCAFG00000054522	ZKSCAN4	Bio13	Bio7
ENSCAFG00000018578	ZNF131	Bio19	Bio13
ENSCAFG00000031621	ZNF596	Bio13	Bio2
ENSCAFG00000002481	ZNF787	Bio13	Elev
ENSCAFG00000009776	ZSWIM1	Bio13	Bio13
ENSCAFG00000009774	ZSWIM3 CTSA	Bio13	Bio13
Results with species richness variables			
Missense variants			
ENSCAFG00000015929	N4BP2	Carnivores	LargeUng
ENSCAFG00000011978	POU2F3	MediumUng	Carnivores
ENSCAFG00000023568	SH2D2A	MediumUng	Carnivores
ENSCAFG00000059254	INSM2	Mammals, Carnivores	MediumUng
ENSCAFG00000051032	Glyceraldehyde 3-phosphate dehydrogenase NAD(P) binding domain-containing protein	Mammals	Mammals
ENSCAFG00000002337	Peroxiredoxin-1 (EC 1.11.1.24) (Thioredoxin-dependent peroxiredoxin 1)	Mammals	Carnivores
ENSCAFG00000008647	TKT	LargeUng	Carnivores
Intron variants			
ENSCAFG00000007326	ADHFE1	Mammals, Carnivores	Mammals
ENSCAFG00000003935	AGBL4	Mammals, Carnivores	LargeUng
ENSCAFG00000016466	ALDH1A2	Mammals, Carnivores	MediumUng
ENSCAFG00000013146	BAZ1B	Mammals, Carnivores	MediumUng

ENSCAFG00000018625	C3	Carnivores	MediumUng
ENSCAFG00000007788	CAPG	MediumUng	MediumUng
ENSCAFG00000002927	CCDC85A	LargeUng	Mammals
ENSCAFG00000020270	COL24A1	Carnivores	LargeUng
ENSCAFG00000013221	CTNNA3	Mammals, Carnivores	Carnivores
ENSCAFG00000007521	DCDC1	MediumUng	Carnivores
ENSCAFG00000002931	DCK	Mammals, Carnivores	Mammals
ENSCAFG00000018636	DENND1C	Carnivores	MediumUng
ENSCAFG00000058000	DIO2	Mammals, Carnivores	Mammals
ENSCAFG00000032222	DLA-12	Mammals, Carnivores	Mammals
ENSCAFG00000000487	DLA88	Mammals, Carnivores	Mammals
ENSCAFG00000017350	ELAVL3	Mammals, Carnivores	Carnivores
ENSCAFG00000012712	EXTL1	Mammals, Carnivores	Carnivores
ENSCAFG00000006203	FCHSD1	MediumUng	Mammals
ENSCAFG00000009921	GABRG3	Carnivores	LargeUng
ENSCAFG00000014796	GULP1	MediumUng	Carnivores
ENSCAFG00000006702	HSD17B12	Carnivores	Carnivores
ENSCAFG00000011876	HSPA12A	Mammals, Carnivores	Mammals
ENSCAFG00000016701	IDUA	Carnivores	MediumUng
ENSCAFG00000013297	IGF2BP2	Mammals, Carnivores	Mammals
ENSCAFG00000042586	Ig-like domain-containing protein	Mammals, Carnivores	MediumUng
ENSCAFG00000042267	KRAB domain-containing protein	Mammals	Carnivores
ENSCAFG00000058939	LOC102152670	Mammals	Carnivores
ENSCAFG00000005321	LOC475563	Mammals, Carnivores	MediumUng
ENSCAFG00000023536	LOC607646	Mammals, Carnivores	MediumUng
ENSCAFG00000003022	MYCL	Carnivores	Mammals
ENSCAFG00000005606	PAK4	Mammals, Carnivores	Mammals
ENSCAFG00000011230	PANX	Mammals	Mammals
ENSCAFG00000004944	PCDH9	Carnivores	Carnivores
ENSCAFG00000032342	PDGFD	Carnivores	LargeUng
ENSCAFG00000001646	PRUNE2	LargeUng	LargeUng
ENSCAFG00000005301	PSD3	Mammals, Carnivores	MediumUng
ENSCAFG00000018073	PSTPIP1	Mammals, Carnivores	Mammals
ENSCAFG00000008370	RNF13	Mammals, Carnivores	Mammals
ENSCAFG00000020317	SAMD13	Mammals, Carnivores	Carnivores

ENSCAFG00000008353	SFRP2	Mammals, Carnivores	MediumUng
ENSCAFG00000011184	SIAE	Mammals	Mammals
ENSCAFG00000000932	SLC35F1	Carnivores	Mammals
ENSCAFG00000016602	SMOC1	Mammals, Carnivores	Carnivores
ENSCAFG00000000361	TBC1D30	MediumUng	MediumUng
ENSCAFG00000010669	TRAPPC10	Carnivores	MediumUng
ENSCAFG00000000473	TRHDE	Carnivores	MediumUng
ENSCAFG00000001226	TRIM58	Mammals	Mammals
ENSCAFG00000010823	ZBTB20	Carnivores	Carnivores
Intron variants non protein coding			
ENSCAFG00000033774	lncRNA	MediumUng	LargeUng
ENSCAFG00000047324	lncRNA	Carnivores	MediumUng
ENSCAFG00000050237	lncRNA	Carnivores	MediumUng
Exon variant non protein coding			
ENSCAFG00000055909	lncRNA	LargeUng, MediumUng	MediumUng
Synonymous variants			
ENSCAFG00000015915	CHRNA9	Carnivores	LargeUng
ENSCAFG00000014480	FSIP2	MediumUng	Carnivores
ENSCAFG00000047151	Olfactory receptor	Mammals, Carnivores	Mammals
ENSCAFG00000001646	PRUNE2	LargeUng	LargeUng
ENSCAFG00000007809	RETSAT	MediumUng	MediumUng
ENSCAFG00000057434	SLC26A1	Carnivores	MediumUng
3' UTR variants			
ENSCAFG00000007326	ADHFE1	Mammals, Carnivores	Mammals
ENSCAFG00000003717	CYRIA	Mammals, Carnivores	Mammals
ENSCAFG00000009322	DNAJC17	Mammals, Carnivores	Mammals
ENSCAFG00000029586	GCHFR	Mammals, Carnivores	Mammals
ENSCAFG00000005037	KLF5	Mammals, Carnivores	Mammals
ENSCAFG00000014386	SLC25A42	Mammals, Carnivores	LargeUng
ENSCAFG00000005654	SLC26A11	Mammals	Mammals
5' UTR variants			
ENSCAFG00000055600	Immunoglobulin V-set domain-containing protein	Mammals, Carnivores	Mammals
ENSCAFG00000001646	PRUNE2	LargeUng	LargeUng
Downstream variants			
ENSCAFG00000011534	ACAN	Carnivores	LargeUng
ENSCAFG00000017319	ACP5	Mammals, Carnivores	Carnivores
ENSCAFG00000005618	ACP7	Mammals, Carnivores	Mammals

ENSCAFG00000004013	ADAD1	Mammals, Carnivores	LargeUng
ENSCAFG00000010654	AGPAT3	Carnivores	MediumUng
ENSCAFG00000011267	AJUBA	Mammals, Carnivores	Mammals
ENSCAFG00000029647	AKR1D1	Mammals, Carnivores	Mammals
ENSCAFG00000018693	ALKBH7	Carnivores	MediumUng
ENSCAFG00000006214	ARAP3	MediumUng	Mammals
ENSCAFG00000016484	ARHGEF11	MediumUng	Carnivores
ENSCAFG00000050591	ARL5B	Mammals, Carnivores	MediumUng
ENSCAFG00000014404	ARMC6	Mammals, Carnivores	LargeUng
ENSCAFG00000000513	ATP6V1G1 ATP6G	Mammals, Carnivores	Mammals
ENSCAFG00000007053	BCL2L1	Mammals	LargeUng
ENSCAFG00000053144	BORCS8	Mammals, Carnivores	LargeUng
ENSCAFG00000023343	C1H19orf18	Mammals	Carnivores
ENSCAFG00000011284	C8H14orf93	Mammals, Carnivores	Mammals
ENSCAFG00000007788	CAPG	MediumUng	MediumUng
ENSCAFG00000005636	CARD14	Mammals	Mammals
ENSCAFG00000006941	CAT	Mammals, Carnivores	LargeUng
ENSCAFG00000054694	CCDC40	Mammals	Mammals
ENSCAFG00000030308	CD70	Carnivores	MediumUng
ENSCAFG00000010082	CDK18	Carnivores	MediumUng
ENSCAFG00000000206	CDO1	Mammals, Carnivores	Carnivores
ENSCAFG00000018696	CLPP	Carnivores	MediumUng
ENSCAFG00000017682	COG7	Mammals, Carnivores	LargeUng
ENSCAFG00000058902	Cornichon family AMPA receptor auxiliary protein 1	Mammals, Carnivores	Mammals
ENSCAFG00000049307	CRB3	Carnivores	MediumUng
ENSCAFG00000008550	CRSP2	Mammals, Carnivores	MediumUng
ENSCAFG00000014999	CRYBA2	Mammals, Carnivores	Mammals
ENSCAFG00000014943	CYP27A1	Mammals, Carnivores	Mammals
ENSCAFG00000007521	DCDC1	MediumUng	Carnivores
ENSCAFG00000008637	DCP1A	LargeUng	Carnivores
ENSCAFG00000000509	DDX39B BAT1 UAP56	Mammals, Carnivores	Mammals
ENSCAFG00000030438	DEFB123	Mammals	LargeUng
ENSCAFG00000032222	DLA-12	Mammals, Carnivores	Mammals
ENSCAFG00000009322	DNAJC17	Mammals, Carnivores	Mammals

ENSCAFG00000017334	ECSIT	Mammals, Carnivores	Carnivores
ENSCAFG00000031111	EFCAB11	MediumUng	Mammals
ENSCAFG00000058608	ELOF1	Mammals, Carnivores	Carnivores
ENSCAFG00000006203	FCHSD1	MediumUng	Mammals
ENSCAFG00000041791	FEV	Mammals, Carnivores	Mammals
ENSCAFG00000032412	FLI1	Mammals, Carnivores	Mammals
ENSCAFG00000011172	FSTL1	Carnivores	Mammals
ENSCAFG00000013159	FZD9	Mammals, Carnivores	MediumUng
ENSCAFG00000005589	GAA	Mammals	Mammals
ENSCAFG00000010035	GABRA5	Carnivores	LargeUng
ENSCAFG00000023219	GLIPR1L1	Mammals	Mammals
ENSCAFG00000029990	GLIPR1L2	Mammals	Mammals
ENSCAFG00000007038	HM13	Mammals	LargeUng
ENSCAFG00000005588	IFNL3	Mammals, Carnivores	Mammals
ENSCAFG00000002330	IGFBP7	Mammals, Carnivores	Mammals
ENSCAFG00000049482	Ig-like domain-containing protein	Mammals, Carnivores	Mammals
ENSCAFG00000003811	Ig-like domain-containing protein	Mammals, Carnivores	MediumUng
ENSCAFG00000024810	Ig-like domain-containing protein	Mammals, Carnivores	MediumUng
ENSCAFG00000029837	Ig-like domain-containing protein	Mammals, Carnivores	MediumUng
ENSCAFG00000046664	Ig-like domain-containing protein	Mammals, Carnivores	MediumUng
ENSCAFG00000055018	Ig-like domain-containing protein	Mammals, Carnivores	MediumUng
ENSCAFG00000057117	Ig-like domain-containing protein	Mammals, Carnivores	MediumUng
ENSCAFG00000052806	Ig-like domain-containing protein	Mammals, Carnivores	MediumUng
ENSCAFG00000057504	Ig-like domain-containing protein	Mammals, Carnivores	MediumUng
ENSCAFG00000004004	IL21	Mammals, Carnivores	LargeUng
ENSCAFG00000055372	Immunoglobulin V-set domain-containing protein	Mammals, Carnivores	MediumUng
ENSCAFG00000024151	INSRR	MediumUng	Carnivores
ENSCAFG00000052759	Interferon lambda-4-like	Mammals, Carnivores	Mammals
ENSCAFG00000000404	IYD	Mammals, Carnivores	Carnivores
ENSCAFG00000005037	KLF5	Mammals, Carnivores	Mammals
ENSCAFG00000009202	KNL1	Mammals, Carnivores	Mammals

ENSCAFG00000055448	lncRNA	Mammals	Mammals
ENSCAFG00000052342	lncRNA	Mammals, Carnivores	MediumUng
ENSCAFG00000050398	lncRNA	Carnivores	LargeUng
ENSCAFG00000058986	lncRNA	Carnivores	MediumUng
ENSCAFG00000056751	lncRNA	Carnivores	Carnivores
ENSCAFG00000055139	lncRNA	Carnivores	Carnivores
ENSCAFG00000057785	lncRNA	Mammals, Carnivores	Mammals
ENSCAFG00000059318	lncRNA	Mammals, Carnivores	Mammals
ENSCAFG00000056599	lncRNA	Mammals	Mammals
ENSCAFG00000050699	lncRNA	Mammals, Carnivores	Mammals
ENSCAFG00000057756	lncRNA	Mammals, Carnivores	MediumUng
ENSCAFG00000051383	lncRNA	Mammals	MediumUng
ENSCAFG00000059058	lncRNA	Mammals, Carnivores	Mammals
ENSCAFG00000052972	lncRNA	Mammals, Carnivores	Mammals
ENSCAFG00000055439	lncRNA	Mammals, Carnivores	MediumUng
ENSCAFG00000053648	lncRNA	Mammals, Carnivores	MediumUng
ENSCAFG00000050865	lncRNA	Mammals	Mammals
ENSCAFG00000054099	lncRNA	Mammals, Carnivores	Carnivores
ENSCAFG00000056062	lncRNA	Mammals, Carnivores	Carnivores
ENSCAFG00000057075	lncRNA	Mammals, Carnivores	Mammals
ENSCAFG00000058342	lncRNA	Carnivores	Mammals
ENSCAFG00000053524	lncRNA	Carnivores	Mammals
ENSCAFG00000052762	lncRNA	Carnivores	Carnivores
ENSCAFG00000051661	lncRNA	Mammals, Carnivores	Carnivores
ENSCAFG00000058863	lncRNA	Mammals, Carnivores	Carnivores
ENSCAFG00000033774	lncRNA	MediumUng	LargeUng
ENSCAFG00000051434	lncRNA	Mammals, Carnivores	Mammals
ENSCAFG00000059098	lncRNA	Carnivores	MediumUng
ENSCAFG00000057930	lncRNA	Carnivores	MediumUng
ENSCAFG00000053565	lncRNA	Carnivores	MediumUng
ENSCAFG00000054596	lncRNA	Carnivores	MediumUng
ENSCAFG00000057389	lncRNA	Carnivores	MediumUng
ENSCAFG00000051752	lncRNA	Carnivores	MediumUng
ENSCAFG00000056979	lncRNA	Carnivores	MediumUng
ENSCAFG00000057280	lncRNA	Mammals	LargeUng
ENSCAFG00000055353	lncRNA	MediumUng	MediumUng

ENSCAFG00000050225	lncRNA	Carnivores	MediumUng
ENSCAFG00000053899	lncRNA	Carnivores	MediumUng
ENSCAFG00000058374	lncRNA	Carnivores	MediumUng
ENSCAFG00000050237	lncRNA	Carnivores	MediumUng
ENSCAFG00000058856	lncRNA	Carnivores	MediumUng
ENSCAFG00000053060	lncRNA	Carnivores	MediumUng
ENSCAFG00000052304	lncRNA	Carnivores	MediumUng
ENSCAFG00000053712	lncRNA	Carnivores	MediumUng
ENSCAFG00000053492	lncRNA	Carnivores	MediumUng
ENSCAFG00000059337	lncRNA	Carnivores	MediumUng
ENSCAFG00000055388	lncRNA	Carnivores	MediumUng
ENSCAFG00000053745	lncRNA	Carnivores	MediumUng
ENSCAFG00000056038	lncRNA	Carnivores	MediumUng
ENSCAFG00000058383	lncRNA	Carnivores	Carnivores
ENSCAFG00000054576	lncRNA	Mammals, Carnivores	Mammals
ENSCAFG00000057251	lncRNA	Mammals, Carnivores	Mammals
ENSCAFG00000055985	lncRNA	Carnivores	Mammals
ENSCAFG00000056785	lncRNA	Carnivores	Mammals
ENSCAFG00000050817	lncRNA	Carnivores	Mammals
ENSCAFG00000053092	lncRNA	Carnivores	Mammals
ENSCAFG00000051796	lncRNA	Mammals, Carnivores	Mammals
ENSCAFG00000028617	LOC482176	Mammals, Carnivores	Mammals
ENSCAFG00000024412	LOC607889	Mammals, Carnivores	Carnivores
ENSCAFG00000018676	LOC611724	Carnivores	MediumUng
ENSCAFG00000011202	LRP10 REM2	Mammals, Carnivores	Mammals
ENSCAFG00000014350	MEF2B	Mammals, Carnivores	LargeUng
ENSCAFG00000003233	MEIS1	Mammals, Carnivores	Mammals
ENSCAFG00000016750	METTL25B	MediumUng	Carnivores
ENSCAFG00000003018	MFSD2A	Carnivores	Mammals
ENSCAFG00000010137	MFSD4A	Carnivores	MediumUng
ENSCAFG00000051129	MHC class I-like antigen recognition-like domain-containing protein	Mammals, Carnivores	Mammals
ENSCAFG00000055788	miRNA	Mammals, Carnivores	MediumUng
ENSCAFG00000025651	miRNA	Carnivores	Carnivores
ENSCAFG00000054270	miRNA	Carnivores	MediumUng
ENSCAFG00000042252	misc_RNA	Mammals, Carnivores	Mammals
ENSCAFG00000006496	MITF	Mammals, Carnivores	Carnivores
ENSCAFG00000049889	Mitochondrial import inner membrane translocase subunit TIM23	Mammals	Carnivores

ENSCAFG00000012970	MLXIPL	Mammals, Carnivores	MediumUng
ENSCAFG00000015054	MMP1	Mammals, Carnivores	Mammals
ENSCAFG00000023615	MMP12	Mammals, Carnivores	Mammals
ENSCAFG00000011197	MMP14	Mammals, Carnivores	Mammals
ENSCAFG00000015063	MMP3	Mammals, Carnivores	Mammals
ENSCAFG00000051237	MOB1B	Mammals, Carnivores	Mammals
ENSCAFG00000011191	MRPL52	Mammals, Carnivores	Mammals
ENSCAFG00000053169	NMRK1	LargeUng, MediumUng	MediumUng
ENSCAFG00000002256	NOA1	Mammals, Carnivores	Mammals
ENSCAFG00000014331	NR2C2AP	Mammals, Carnivores	LargeUng
ENSCAFG00000010168	NUCKS1	Carnivores	MediumUng
ENSCAFG00000031210	OAF	MediumUng	Carnivores
ENSCAFG00000050743	OR6C38	Mammals	Mammals
ENSCAFG00000049942	OR6C63	Mammals	Mammals
ENSCAFG00000054648	OR7A23	Mammals, Carnivores	Mammals
ENSCAFG00000058847	OR7A55	Mammals, Carnivores	Mammals
ENSCAFG00000041401	OR7A55 OR7A58	Mammals, Carnivores	Mammals
ENSCAFG00000041284	OR7A60	Mammals, Carnivores	Mammals
ENSCAFG00000050877	OR7A63	Mammals, Carnivores	Mammals
ENSCAFG00000051510	OR7A64	Mammals, Carnivores	Mammals
ENSCAFG00000011237	OR8A1	Mammals	Mammals
ENSCAFG00000011159	OXA1L	Mammals, Carnivores	Mammals
ENSCAFG00000005032	PIBF1 KLF5	Mammals, Carnivores	Mammals
ENSCAFG00000017423	PLPPR2	Mammals, Carnivores	Carnivores
ENSCAFG00000009351	PPP1R14D	Mammals, Carnivores	Mammals
ENSCAFG00000003004	PPT1	Carnivores	Mammals
ENSCAFG00000016737	PRCC	MediumUng	Carnivores
ENSCAFG00000020319	PRKACB	Mammals, Carnivores	Carnivores
ENSCAFG00000008689	PRKCD	LargeUng	Carnivores
ENSCAFG00000017392	PRKCSH	Mammals, Carnivores	Carnivores
ENSCAFG00000011243	PRMT5	Mammals, Carnivores	Mammals

ENSCAFG00000014984	processed_pseudogene	Mammals, Carnivores	Mammals
ENSCAFG00000053426	Protein kinase domain-containing protein	Mammals, Carnivores	Mammals
ENSCAFG00000012445	PROX1	Mammals	MediumUng
ENSCAFG00000041221	pseudogene	Mammals	Carnivores
ENSCAFG00000041670	pseudogene	Mammals, Carnivores	LargeUng
ENSCAFG00000047122	pseudogene	Mammals, Carnivores	Mammals
ENSCAFG00000056490	pseudogene	Mammals, Carnivores	Mammals
ENSCAFG00000044330	pseudogene	Mammals, Carnivores	Carnivores
ENSCAFG00000043307	pseudogene	Mammals, Carnivores	Carnivores
ENSCAFG00000036065	pseudogene	Mammals	Carnivores
ENSCAFG00000009280	PTPRT	MediumUng	MediumUng
ENSCAFG00000017264	PTTG1	Carnivores	Carnivores
ENSCAFG00000003723	QRSL1	Mammals	Carnivores
ENSCAFG00000009220	RAD51	Mammals, Carnivores	Mammals
ENSCAFG00000047285	RALGAPA1	Mammals, Carnivores	MediumUng
ENSCAFG00000011217	RBM23	Mammals, Carnivores	Mammals
ENSCAFG00000015898	RBM47	Carnivores	LargeUng
ENSCAFG00000018077	RCN2	Mammals, Carnivores	Mammals
ENSCAFG00000006185	RELL2	MediumUng	Mammals
ENSCAFG00000006991	REM1	Mammals	LargeUng
ENSCAFG00000002254	REST	Mammals, Carnivores	Mammals
ENSCAFG00000050536	RHOH	Carnivores	LargeUng
ENSCAFG00000029187	RHOV	Mammals, Carnivores	Mammals
ENSCAFG00000043911	RNA-directed DNA polymerase (EC 2.7.7.49)	MediumUng	Mammals
ENSCAFG00000041062	RNA-directed DNA polymerase (EC 2.7.7.49)	Mammals, Carnivores	MediumUng
ENSCAFG00000032095	RNF212	Carnivores	MediumUng
ENSCAFG00000007313	RRS1	Mammals, Carnivores	Mammals
ENSCAFG00000017690	SCNN1G	Mammals, Carnivores	LargeUng
ENSCAFG00000013266	SENP2	Mammals, Carnivores	Mammals
ENSCAFG00000011908	SHTN1	Mammals, Carnivores	Mammals
ENSCAFG00000011184	SIAE	Mammals	Mammals
ENSCAFG00000012702	SLC30A2	Mammals, Carnivores	Carnivores
ENSCAFG00000058658	SLC45A3	Carnivores	MediumUng

ENSCAFG00000023919	SLC8A3	Mammals, Carnivores	Carnivores
ENSCAFG00000000522	SNCAIP	Carnivores	LargeUng
ENSCAFG00000026053	snoRNA	LargeUng, MediumUng	MediumUng
ENSCAFG00000050039	snoRNA	Mammals, Carnivores	Carnivores
ENSCAFG00000020981	snoRNA	Mammals, Carnivores	Mammals
ENSCAFG00000026284	snoRNA	Mammals, Carnivores	Mammals
ENSCAFG00000055381	snRNA	Mammals, Carnivores	Mammals
ENSCAFG00000043782	snRNA	Mammals, Carnivores	Mammals
ENSCAFG00000026117	snRNA	Carnivores	Mammals
ENSCAFG00000028313	snRNA	Carnivores	Mammals
ENSCAFG00000043087	snRNA	Carnivores	MediumUng
ENSCAFG00000053965	snRNA	Carnivores	LargeUng
ENSCAFG00000058807	SWSAP1	Mammals, Carnivores	Carnivores
ENSCAFG00000005592	SYCN	Mammals, Carnivores	Mammals
ENSCAFG00000013010	TBL2 BCL7B MLXIPL	Mammals, Carnivores	MediumUng
ENSCAFG00000007832	TCF7L1	MediumUng	MediumUng
ENSCAFG00000024531	TGOLN2	MediumUng	MediumUng
ENSCAFG00000053610	TICAM2	Mammals, Carnivores	Carnivores
ENSCAFG00000008351	TLR2	Mammals, Carnivores	MediumUng
ENSCAFG00000016696	TMED11	Carnivores	MediumUng
ENSCAFG00000000209	TMED7	Mammals, Carnivores	Carnivores
ENSCAFG00000014369	TMEM161A	Mammals, Carnivores	LargeUng
ENSCAFG00000016710	TMEM175	Carnivores	MediumUng
ENSCAFG00000012671	TRIM63	Mammals, Carnivores	Carnivores
ENSCAFG00000018068	TSPAN3	Mammals, Carnivores	Mammals
ENSCAFG00000018634	TUBB4A	Carnivores	MediumUng
ENSCAFG00000059114	Uncharacterized protein	Mammals	Mammals
ENSCAFG00000058607	Uncharacterized protein	Carnivores	MediumUng
ENSCAFG00000042433	Uncharacterized protein	Carnivores	MediumUng
ENSCAFG00000020315	UOX	Mammals, Carnivores	Carnivores
ENSCAFG00000031321	WNT6	Mammals, Carnivores	Mammals
ENSCAFG00000000224	YTHDC2	Mammals, Carnivores	Carnivores
ENSCAFG00000046247	Zinc finger protein 671	Mammals	Carnivores

ENSCAFG00000031712	ZNF653	Mammals, Carnivores	Carnivores
ENSCAFG00000024833	ZNF654	MediumUng	MediumUng
Upstream variants			
ENSCAFG00000011583	ABHD2	Carnivores	LargeUng
ENSCAFG00000015933	Acylphosphatase-1 (Acylphosphate phosphohydrolase 1)	Carnivores	LargeUng
ENSCAFG00000007326	ADHFE1	Mammals, Carnivores	Mammals
ENSCAFG00000016466	ALDH1A2	Mammals, Carnivores	MediumUng
ENSCAFG00000016556	AQP9	Mammals, Carnivores	MediumUng
ENSCAFG00000011934	ARHGEF12	MediumUng	Carnivores
ENSCAFG00000029729	C1H9orf40	LargeUng, MediumUng	MediumUng
ENSCAFG00000029901	C1QTNF2	Carnivores	Carnivores
ENSCAFG00000008525	CACNA1D	LargeUng	Carnivores
ENSCAFG00000002704	CACNG6	Mammals	Mammals
ENSCAFG00000029738	CACNG7	Mammals	Mammals
ENSCAFG00000032667	CACNG8	Mammals	Mammals
ENSCAFG00000003012	CAP1	Carnivores	Mammals
ENSCAFG00000000481	CAPS2	Mammals	Mammals
ENSCAFG00000001703	CARNMT1	LargeUng, MediumUng	MediumUng
ENSCAFG00000014996	CDK5R2	Mammals, Carnivores	Mammals
ENSCAFG00000017326	CNN1	Mammals, Carnivores	Carnivores
ENSCAFG00000007049	COX4I2	Mammals	LargeUng
ENSCAFG00000059217	CRABP2	MediumUng	Carnivores
ENSCAFG00000014888	DDI1	Carnivores	LargeUng
ENSCAFG00000016708	DGKQ	Carnivores	MediumUng
ENSCAFG00000006149	DIAPH1	MediumUng	Mammals
ENSCAFG00000032222	DLA-12	Mammals, Carnivores	Mammals
ENSCAFG00000000487	DLA88	Mammals, Carnivores	Mammals
ENSCAFG00000020313	DNASE2B	Mammals, Carnivores	Carnivores
ENSCAFG00000031582	EHF	Mammals, Carnivores	LargeUng
ENSCAFG00000005614	EIF4A3	Mammals	Mammals
ENSCAFG00000007666	EIF4G2	MediumUng	LargeUng
ENSCAFG00000006935	ELF5	Mammals, Carnivores	LargeUng
ENSCAFG00000010144	ELK4	Carnivores	MediumUng
ENSCAFG00000007801	ELMOD3	MediumUng	MediumUng
ENSCAFG00000011891	ENO4	Mammals, Carnivores	Mammals
ENSCAFG00000017411	EPOR	Mammals, Carnivores	Carnivores

ENSCAFG00000028571	FAM200C	Carnivores	Carnivores
ENSCAFG00000024310	FBXO27	Mammals, Carnivores	Mammals
ENSCAFG00000000210	FEM1C	Mammals, Carnivores	Carnivores
ENSCAFG00000016699	FGFRL1	Carnivores	MediumUng
ENSCAFG00000016743	GAK	Carnivores	MediumUng
ENSCAFG00000041383	GATD3	Carnivores	MediumUng
ENSCAFG00000000525	GLIPR1	Mammals	Mammals
ENSCAFG00000000359	GNS	MediumUng	MediumUng
ENSCAFG00000018681	GTF2F1	Carnivores	MediumUng
ENSCAFG00000007858	HAND2	Carnivores	MediumUng
ENSCAFG00000011542	HAPLN3 MFGE8	Carnivores	LargeUng
ENSCAFG00000006173	HDAC3	MediumUng	Mammals
ENSCAFG00000016741	HDGF	MediumUng	Carnivores
ENSCAFG00000003797	Heat shock cognate 71 kDa protein (EC 3.6.4.10) (Heat shock 70 kDa protein 8)	Mammals	LargeUng
ENSCAFG00000011876	HSPA12A	Mammals, Carnivores	Mammals
ENSCAFG00000007045	ID1	Mammals	LargeUng
ENSCAFG00000058699	Ig-like domain-containing protein	Mammals, Carnivores	Mammals
ENSCAFG00000055775	Ig-like domain-containing protein	Mammals, Carnivores	Mammals
ENSCAFG00000057398	Ig-like domain-containing protein	Mammals, Carnivores	Mammals
ENSCAFG00000045242	Ig-like domain-containing protein	Mammals, Carnivores	Mammals
ENSCAFG00000054905	Ig-like domain-containing protein	Mammals, Carnivores	Mammals
ENSCAFG00000014478	Ig-like domain-containing protein	Mammals, Carnivores	MediumUng
ENSCAFG00000045360	Ig-like domain-containing protein	Mammals, Carnivores	MediumUng
ENSCAFG00000056839	Ig-like domain-containing protein	Mammals, Carnivores	MediumUng
ENSCAFG00000004008	IL2	Mammals, Carnivores	LargeUng
ENSCAFG00000058265	Immunoglobulin C1-set domain-containing protein	Mammals, Carnivores	MediumUng
ENSCAFG00000050263	Immunoglobulin V-set domain-containing protein	Mammals, Carnivores	Mammals
ENSCAFG00000031359	Immunoglobulin V-set domain-containing protein	Mammals, Carnivores	MediumUng
ENSCAFG00000043031	Immunoglobulin V-set domain-containing protein	Mammals, Carnivores	MediumUng
ENSCAFG00000041829	Interferon lambda-4-like	Mammals, Carnivores	Mammals
ENSCAFG00000016753	ISG20L2	MediumUng	Carnivores
ENSCAFG00000011056	ITPK1	Mammals, Carnivores	Mammals
ENSCAFG00000010259	KCNJ1	Mammals, Carnivores	Mammals

ENSCAFG00000010255	KCNJ5	Mammals, Carnivores	Mammals
ENSCAFG00000018667	KHSRP	Carnivores	MediumUng
ENSCAFG00000001115	L3MBTL3	Mammals, Carnivores	LargeUng
ENSCAFG00000057961	lncRNA	Mammals	Carnivores
ENSCAFG00000054112	lncRNA	Mammals, Carnivores	Carnivores
ENSCAFG00000052090	lncRNA	Carnivores	LargeUng
ENSCAFG00000054544	lncRNA	Carnivores	MediumUng
ENSCAFG00000054466	lncRNA	Mammals, Carnivores	Mammals
ENSCAFG00000051945	lncRNA	Mammals, Carnivores	Mammals
ENSCAFG00000057178	lncRNA	Mammals	MediumUng
ENSCAFG00000056169	lncRNA	MediumUng	MediumUng
ENSCAFG00000054033	lncRNA	Mammals	Mammals
ENSCAFG00000050476	lncRNA	Mammals	Mammals
ENSCAFG00000052727	lncRNA	Mammals, Carnivores	Carnivores
ENSCAFG00000054065	lncRNA	Mammals, Carnivores	Carnivores
ENSCAFG00000051921	lncRNA	Mammals, Carnivores	Mammals
ENSCAFG00000059075	lncRNA	Carnivores	Carnivores
ENSCAFG00000038294	lncRNA	Mammals, Carnivores	MediumUng
ENSCAFG00000056504	lncRNA	Mammals, Carnivores	MediumUng
ENSCAFG00000056757	lncRNA	LargeUng	Carnivores
ENSCAFG00000052308	lncRNA	Mammals, Carnivores	LargeUng
ENSCAFG00000033774	lncRNA	MediumUng	LargeUng
ENSCAFG00000047324	lncRNA	Carnivores	MediumUng
ENSCAFG00000055031	lncRNA	Carnivores	MediumUng
ENSCAFG00000056699	lncRNA	Carnivores	MediumUng
ENSCAFG00000052713	lncRNA	Carnivores	MediumUng
ENSCAFG00000058384	lncRNA	MediumUng	MediumUng
ENSCAFG00000052160	lncRNA	MediumUng	MediumUng
ENSCAFG00000051637	lncRNA	Carnivores	MediumUng
ENSCAFG00000057445	lncRNA	Carnivores	MediumUng
ENSCAFG00000053959	lncRNA	Carnivores	MediumUng
ENSCAFG00000057071	lncRNA	Carnivores	MediumUng
ENSCAFG00000051070	lncRNA	Carnivores	MediumUng
ENSCAFG00000056859	lncRNA	Carnivores	MediumUng
ENSCAFG00000050366	lncRNA	Carnivores	MediumUng
ENSCAFG00000050470	lncRNA	Carnivores	Carnivores
ENSCAFG00000057488	lncRNA	Mammals, Carnivores	Mammals
ENSCAFG00000058081	lncRNA	Carnivores	Mammals

ENSCAFG00000057901	lncRNA	Carnivores	Mammals
ENSCAFG00000051335	lncRNA	Carnivores	Mammals
ENSCAFG00000053720	lncRNA	Mammals, Carnivores	Mammals
ENSCAFG00000055919	lncRNA	Carnivores	MediumUng
ENSCAFG00000058042	LOC102152102	Mammals	Carnivores
ENSCAFG00000053510	LOC111093559	Mammals, Carnivores	Carnivores
ENSCAFG00000002805	LOC482174	Mammals, Carnivores	Mammals
ENSCAFG00000002808	LOC482175	Mammals, Carnivores	Mammals
ENSCAFG00000048824	LOC482215	Mammals	Mammals
ENSCAFG00000055528	LOC482217	Mammals	Mammals
ENSCAFG00000011160	LRRC58	Carnivores	Mammals
ENSCAFG00000016502	LRRC71	MediumUng	Carnivores
ENSCAFG00000000515	LTA TNFB TNFSF1	Mammals, Carnivores	Mammals
ENSCAFG00000000506	MCCD1	Mammals, Carnivores	Mammals
ENSCAFG00000054713	miRNA	LargeUng, MediumUng	MediumUng
ENSCAFG00000031179	miRNA	Mammals	Mammals
ENSCAFG00000052222	miRNA	Mammals	Mammals
ENSCAFG00000025665	miRNA	Carnivores	Carnivores
ENSCAFG00000053766	miRNA	Mammals, Carnivores	LargeUng
ENSCAFG00000047537	miRNA	Carnivores	Carnivores
ENSCAFG00000000033	Mitochondrial carrier 2	Mammals, Carnivores	MediumUng
ENSCAFG00000006533	MMP17	LargeUng, Carnivores	MediumUng
ENSCAFG00000015066	MMP27	Mammals, Carnivores	Mammals
ENSCAFG00000023335	MMP8	Mammals, Carnivores	Mammals
ENSCAFG00000016747	MRPL24	MediumUng	Carnivores
ENSCAFG00000051307	MYADM	Mammals	Mammals
ENSCAFG00000003022	MYCL	Carnivores	Mammals
ENSCAFG00000014218	NCAN	Mammals, Carnivores	LargeUng
ENSCAFG00000005593	NCCRP1	Mammals, Carnivores	Mammals
ENSCAFG00000030599	NFKBIL1	Mammals, Carnivores	Mammals
ENSCAFG00000057237	NLRP12	Mammals	Mammals
ENSCAFG00000016528	NTRK1	MediumUng	Carnivores
ENSCAFG00000017400	ODAD3	Mammals, Carnivores	Carnivores
ENSCAFG00000011242	Olfactory receptor	Mammals	Mammals
ENSCAFG00000041843	OR2AJ7	Mammals	Mammals

ENSCAFG00000024736	OR2W3	Mammals	Mammals
ENSCAFG00000057285	OR7A57	Mammals, Carnivores	Mammals
ENSCAFG00000011240	OR8B12	Mammals	Mammals
ENSCAFG00000001699	OSTF1	LargeUng, MediumUng	MediumUng
ENSCAFG00000012726	PAFAH2	Mammals, Carnivores	Carnivores
ENSCAFG00000057857	PDIK1L	Mammals, Carnivores	Carnivores
ENSCAFG00000016515	PEAR1	MediumUng	Carnivores
ENSCAFG00000002296	POLR2B	Mammals, Carnivores	Mammals
ENSCAFG00000016551	Prefoldin subunit 4	Mammals, Carnivores	MediumUng
ENSCAFG00000014957	PRKAG3	Mammals, Carnivores	Mammals
ENSCAFG00000029928	PRKCG	Mammals	Mammals
ENSCAFG00000014481	PRSS2	Mammals, Carnivores	MediumUng
ENSCAFG00000048074	pseudogene	Mammals, Carnivores	Mammals
ENSCAFG00000018689	pseudogene	Carnivores	MediumUng
ENSCAFG00000049146	pseudogene	Mammals, Carnivores	Carnivores
ENSCAFG00000049284	pseudogene	Mammals, Carnivores	Carnivores
ENSCAFG00000043459	pseudogene	Mammals, Carnivores	Carnivores
ENSCAFG00000010687	PWP2	Carnivores	MediumUng
ENSCAFG00000017438	RAB3D	Mammals, Carnivores	Carnivores
ENSCAFG00000008732	RFT1	LargeUng	Carnivores
ENSCAFG00000014341	RFXANK	Mammals, Carnivores	LargeUng
ENSCAFG00000017405	RGL3	Mammals, Carnivores	Carnivores
ENSCAFG00000050536	RHOH	Carnivores	LargeUng
ENSCAFG00000007678	Ribose-phosphate pyrophosphokinase N-terminal domain-containing protein	MediumUng	LargeUng
ENSCAFG00000009237	RMDN3	Mammals, Carnivores	Mammals
ENSCAFG00000042857	RNF175	Mammals, Carnivores	MediumUng
ENSCAFG00000005675	RNF213	Mammals	Mammals
ENSCAFG00000003719	RTN4IP1	Mammals	Carnivores
ENSCAFG00000020317	SAMD13	Mammals, Carnivores	Carnivores
ENSCAFG00000018100	SCAPER	Mammals, Carnivores	Mammals
ENSCAFG00000017687	SCNN1B	Mammals, Carnivores	LargeUng
ENSCAFG00000006625	SFSWAP	LargeUng, Carnivores	MediumUng

ENSCAFG00000005640	SGSH	Mammals	Mammals
ENSCAFG00000054580	SH2D6	MediumUng	MediumUng
ENSCAFG00000028962	SLC25A21	Mammals	Mammals
ENSCAFG00000018643	SLC25A23 SLC25A41	Carnivores	MediumUng
ENSCAFG00000002939	SLC4A4	Mammals, Carnivores	Mammals
ENSCAFG00000011177	SLC7A7	Mammals, Carnivores	Mammals
ENSCAFG00000005293	SLITRK5	Carnivores	MediumUng
ENSCAFG00000017268	SLU7	Carnivores	Carnivores
ENSCAFG00000044478	Small ribosomal subunit protein uS15 N-terminal domain-containing protein	Mammals, Carnivores	Mammals
ENSCAFG00000026082	snoRNA	Mammals, Carnivores	Mammals
ENSCAFG00000028262	snoRNA	MediumUng	LargeUng
ENSCAFG00000022899	snRNA	LargeUng	LargeUng
ENSCAFG00000026947	snRNA	Carnivores	Carnivores
ENSCAFG00000022520	snRNA	Mammals, Carnivores	Mammals
ENSCAFG00000022253	snRNA	Mammals, Carnivores	Mammals
ENSCAFG00000049973	snRNA	Mammals, Carnivores	LargeUng
ENSCAFG00000055501	snRNA	Mammals, Carnivores	Mammals
ENSCAFG00000027532	snRNA	Carnivores	Carnivores
ENSCAFG00000020917	snRNA	Carnivores	MediumUng
ENSCAFG00000025934	snRNA	Carnivores	MediumUng
ENSCAFG00000053965	snRNA	Carnivores	LargeUng
ENSCAFG00000044832	SNRPC	Carnivores	LargeUng
ENSCAFG00000032261	SPA17	Mammals	Mammals
ENSCAFG00000009365	SPINT1	Mammals, Carnivores	Mammals
ENSCAFG00000029191	STMN1	Mammals, Carnivores	Carnivores
ENSCAFG00000014431	SUGP2	Mammals, Carnivores	LargeUng
ENSCAFG00000000361	TBC1D30	MediumUng	MediumUng
ENSCAFG00000011215	TBRG1	Mammals	Mammals
ENSCAFG00000011963	TLCD5	MediumUng	Carnivores
ENSCAFG00000017434	TMEM205	Mammals, Carnivores	Carnivores
ENSCAFG00000030400	TNFSF9	Carnivores	MediumUng
ENSCAFG00000044061	Uncharacterized protein	Carnivores	MediumUng
ENSCAFG00000045033	Uncharacterized protein	Mammals, Carnivores	Carnivores
ENSCAFG00000013727	Uncharacterized protein	Mammals	Mammals
ENSCAFG00000051647	Uncharacterized protein	Mammals, Carnivores	Carnivores
ENSCAFG00000041808	Uncharacterized protein	Mammals, Carnivores	Mammals

ENSCAFG00000041820	Uncharacterized protein	Carnivores	Carnivores
ENSCAFG00000014987	WNT10A	Mammals, Carnivores	Mammals
ENSCAFG00000028815	ZBED5	MediumUng	LargeUng
ENSCAFG00000009345	ZFYVE19	Mammals, Carnivores	Mammals
ENSCAFG00000028550	Zinc finger protein 850-like	Mammals	Carnivores

Supplementary table 7. g:Profiler results summary for wolves and FRDs using the dog and human genomes

Classification	*Term name*	*Term ID*	*Adjusted p-value*
Wolves' g:Profiler results summary WorldClim variables with Dog genome			
MF	Molecular function inhibitor activity*	GO:0140678	6.36E-04 *(9.97E-03)
MF	Glutathione transferase activity	GO:0004364	1.90E-03
MF	Enzyme binding	GO:0019899	1.91E-02
MF	Oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	GO:0016705	2.54E-02
BP	Multicellular organismal process*	GO:0032501	1.55E-04 *(2.80E-07)
BP	Regulation of nitrogen compound metabolic process*	GO:0051171	1.02E-02 *(2.20E-02)
BP	Regulation of primary metabolic process*	GO:0080090	3.02E-02 *(2.93E-02)
BP	Positive regulation of macromolecule metabolic process	GO:0010604	3.37E-02
BP	Regulation of neuromuscular junction development	GO:1904396	4.85E-02
CC	Cytoplasm*	GO:0005737	1.78E-04 *(2.22E-06)
CC	Cell periphery*	GO:0071944	1.52E-03 *(6.02E-03)
CC	Plasma membrane bounded cell projection	GO:0120025	1.65E-03
CC	Intracellular protein-containing complex*	GO:0140535	4.10E-02 *(1.26E-02)
Wolves' g:Profiler results summary WorldClim variables with human genome			
MF	Protein binding	GO:0005515	9.10E-10
MF	Low-density lipoprotein particle receptor activity	GO:0005041	1.87E-03
MF	Enzyme regulator activity	GO:0030234	3.07E-03
BP	Organonitrogen compound metabolic process	GO:1901564	1.22E-04
BP	Regulation of molecular function	GO:0065009	8.78E-03
BP	Macromolecule modification	GO:0043412	2.03E-02
BP	Positive regulation of osteoclast differentiation	GO:0045672	3.14E-02
CC	Extracellular region	GO:0005576	9.63E-06
CC	Catalytic complex	GO:1902494	9.70E-04
CC	Cell projection	GO:0042995	2.73E-03
CC	Cell junction	GO:0030054	1.60E-02
CC	Neuronal cell body	GO:0043025	4.81E-02
Wolves' g:Profiler results summary species richness variables with Dog genome			
MF	Ion channel regulator activity*	GO:0099106	2.62E-05 *(2.54E-05)
MF	Serine-type endopeptidase activity*	GO:0004252	2.35E-04 *(2.11E-03)
Wolves' g:Profiler results summary species richness variables with human genome			
MF	Protein-arginine deiminase activity	GO:0004668	5.32E-03
BP	Regulation of multicellular organismal process	GO:0051239	1.38E-02
CC	Microtubule plus-end	GO:0035371	4.18E-02
FRDs' g:Profiler results summary WorldClim variables with Dog genome			
MF	Protein binding*	GO:0005515	5.71E-04 *(1.73E-10)
BP	Multicellular organism development	GO:0007275	5.65E-05
BP	Morphogenesis of an epithelium	GO:0002009	4.80E-02
CC	Cytoplasm*	GO:0005737	5.57E-06 *(1.60E-09)
CC	Nucleoplasm	GO:0005654	4.02E-04
CC	Endomembrane system	GO:0012505	1.54E-03

CC	Membrane	GO:0016020	2.50E-02
FRDs' g:Profiler results summary WorldClim variables with human genome			
BP	Developmental process	GO:0032502	7.09E-10
BP	Collagen catabolic process	GO:0030574	9.26E-05
BP	Carbohydrate derivative metabolic process	GO:1901135	2.54E-03
BP	Biological regulation	GO:0065007	3.04E-03
BP	Response to stimulus	GO:0050896	3.31E-03
BP	Small molecule metabolic process	GO:0044281	5.72E-03
BP	Extracellular matrix organization	GO:0030198	6.00E-03
BP	Lipid metabolic process	GO:0006629	1.35E-02
BP	Localization	GO:0051179	1.42E-02
BP	Organonitrogen compound metabolic process	GO:1901564	2.53E-02
BP	Response to organic cyclic compound	GO:0014070	3.82E-02
BP	Growth	GO:0040007	4.85E-02
CC	Microtubule cytoskeleton	GO:0015630	1.10E-02
FRDs' g:Profiler results summary species richness variables with dog genome			
BP	Biological regulation*	GO:0065007	3.81E-02 *(3.11E-03)
CC	Lumenal side of membrane	GO:0098576	8.29E-03
FRDs' g:Profiler results summary species richness variables with human genome			
MF	Inorganic molecular entity transmembrane transporter activity	GO:0015318	2.50E-02
MF	Metalloendopeptidase activity	GO:0004222	4.52E-02
BP	Collagen catabolic process	GO:0030574	1.55E-04
BP	Multicellular organismal process	GO:0032501	3.28E-03
BP	Regulation of lipid biosynthetic process	GO:0046890	2.75E-02
BP	Cell communication	GO:0007154	3.46E-02
CC	Cytoplasm	GO:0005737	1.58E-04
CC	L-type voltage-gated calcium channel complex	GO:1990454	3.38E-03
CC	Vacuolar lumen	GO:0005775	4.39E-03
CC	Extracellular matrix	GO:0031012	6.29E-03
CC	Lysosome	GO:0005764	2.38E-02
All overlapping genes in grey wolves and FRDs g:Profiler results summary			
BP	Regulation of synaptic assembly at neuromuscular junction	GO:0008582	1.10E-02
BP	Olfactory placode formation	GO:0030910	1.10E-02
BP	Regulation of branch elongation involved in ureteric bud branching	GO:0072095	3.29E-02
CC	Non-motile cilium	GO:0097731	1.80E-02

MF: Molecular function, BP: Biological process and CC: Cellular component. * (x.xx) Term was also in the human run within the parenthesis the p-value in the human genome run.

Supplementary table 8. Genes displaying signatures of both diversifying selection and local adaptation

Free-Ranging Dogs					
CanFam3 SNP location	**ENSMBL ID**	**UnitProt Code**	**LFMM**	**RDA**	**Variant consequence**
chr20:49895596	ENSCAFG00000017319	ACP5	Bio14, Bio19, Hfpi	Bio7	Downstream
chr1:114023508	ENSCAFG00000005618	ACP7	Bio14, Bio19, Hfpi	Bio7	Downstream
chr19:17722381	ENSCAFG00000004013	ADAD1	Bio14	Bio13	Downstream
chr15:11194359	ENSCAFG00000003935	AGBL4	Bio14, Bio19, Hfpi	Elev	Intron
chr8:3231105	ENSCAFG00000011267	AJUBA	Bio14, Bio19, Hfpi	Bio19	Downstream
chr16:14048745	ENSCAFG00000029647	AKR1D1	Bio14, Bio19, Hfpi	Bio2	Downstream
chr30:23187554	ENSCAFG00000016466	ALDH1A2	Bio14	Bio19	Upstream
chr30:23187554	ENSCAFG00000016556	AQP9	Bio14	Bio19	Upstream
chr20:44045441	ENSCAFG00000014404	ARMC6	Bio14, Bio19, Hfpi	Bio13	Downstream
chr16:41300147	ENSCAFG00000007143	ASAH1	Elev	Bio13	Synonymous
chr7:42264018	ENSCAFG00000016944	ASH1L	Bio14	Bio8	Downstream
chr15:16254754	ENSCAFG00000004854	ATP6V0B	Bio14, Bio19	Elev	Downstream
chr12:973829	ENSCAFG00000000513	ATP6V1G1 ATP6G	Bio14	Bio13	Downstream
chr15:16254754	ENSCAFG00000004833	B4GALT2	Bio14, Bio19	Elev	Downstream
chr20:44045441	ENSCAFG00000053144	BORCS8	Bio14, Bio19, Hfpi	Bio13	Downstream
chr12:17433935	ENSCAFG00000030374	C12H6orf141	Bio14, Bio19	Bio13	Downstream
chr1:100048794	ENSCAFG00000023343	C1H19orf18	Bio14, Bio19, Hfpi	Bio19	Downstream
chr8:3231105	ENSCAFG00000011284	C8H14orf93	Bio14, Bio19, Hfpi	Bio19	Downstream
chr10:15948439	ENSCAFG00000000481	CAPS2	Bio14, Bio19, Hfpi	Bio8	Upstream
chr18:33302676	ENSCAFG00000006941	CAT	Bio14, Bio19, Hfpi	Bio13	Downstream
chr14:55381209	ENSCAFG00000055058	CAV1 CAV	Bio14, Bio19	Bio19	Upstream
chr14:55381209	ENSCAFG00000003402	CAV2 CAV-2	Bio14, Bio19	Bio19	Upstream
chr15:16254754	ENSCAFG00000046547	CCDC24	Bio14, Bio19	Elev	Downstream
chr37:25461648	ENSCAFG00000014996	CDK5R2	Bio14	Bio13	Upstream
chr11:5292792	ENSCAFG00000000206	CDO1	Bio14, Bio19, Hfpi	Bio13	Downstream
chr12:17433935	ENSCAFG00000002126	CENPQ	Bio14, Bio19	Bio13	Downstream
chr7:42264018	ENSCAFG00000016969	CLK2	Bio14	Bio8	Upstream
chr20:49895596	ENSCAFG00000017326	CNN1	Bio14, Bio19, Hfpi	Bio7	Upstream
chr6:22487777	ENSCAFG00000017682	COG7	Bio14	Bio14	Downstream
chr12:17433935	ENSCAFG00000023105	CRISP2 LOC481834	Bio14, Bio19	Bio13	Downstream
chr37:25461648	ENSCAFG00000014999	CRYBA2	Bio14	Bio13	Downstream
chr37:25461648	ENSCAFG00000014943	CYP27A1	Bio14	Bio13	Downstream
chr13:60225215	ENSCAFG00000002931	DCK	Bio14, Bio19, Hfpi	Bio13	Intron
chr12:973829	ENSCAFG00000000509	DDX39B BAT1 UAP56	Bio14	Bio13	Downstream
chr8:52691027	ENSCAFG00000058000	DIO2	Bio14, Bio19, Hfpi	Bio13	Intron
chr12:973829	ENSCAFG00000032222	DLA-12	Bio14	Bio13	Upstream

chr12:973829	ENSCAFG00000000487	DLA88	Bio14	Bio13	Upstream
chr11:66236758	ENSCAFG00000003006	DNAJC25	Bio14	Bio13	Intron
chr6:63552340	ENSCAFG00000020313	DNASE2B	Bio14, Bio19, Hfpi	Bio19	Upstream
chr15:16254754	ENSCAFG00000004862	DPH2	Bio14, Bio19	Elev	Downstream
chr11:66236758	ENSCAFG00000002990	ECPAS	Bio14	Bio13	Upstream
chr20:49895596	ENSCAFG00000017334	ECSIT	Bio14, Bio19, Hfpi	Bio7	Downstream
chr18:33302676	ENSCAFG00000031582	EHF	Bio14, Bio19, Hfpi	Bio13	Upstream
chr20:49895596	ENSCAFG00000017350	ELAVL3	Bio14, Bio19, Hfpi	Bio7	Intron
chr18:33302676	ENSCAFG00000006935	ELF5	Bio14, Bio19, Hfpi	Bio13	Upstream
chr20:49895596	ENSCAFG00000058608	ELOF1	Bio14, Bio19, Hfpi	Bio7	Downstream
chr28:27378066	ENSCAFG00000011891	ENO4	Bio14, Bio19, Hfpi	Hfpi	Upstream
chr7:42264018	ENSCAFG00000016980	ENTREP3	Bio14	Bio8	Upstream
chr20:49895596	ENSCAFG00000017411	EPOR	Bio14, Bio19, Hfpi	Bio7	Upstream
chr2:73943796	ENSCAFG00000012712	EXTL1	Bio14, Bio19, Hfpi	Bio13	Intron
chr1:114023508	ENSCAFG00000024310	FBXO27	Bio14, Bio19, Hfpi	Bio7	Upstream
chr7:42264018	ENSCAFG00000030749	FDPS	Bio14	Bio8	Upstream
chr11:5292792	ENSCAFG00000000210	FEM1C	Bio14, Bio19, Hfpi	Bio13	Upstream
chr37:25461648	ENSCAFG00000041791	FEV	Bio14	Bio13	Downstream
chr5:5788946	ENSCAFG00000032412	FLI1	Bio14, Bio19, Hfpi	Bio19	Downstream
chr16:41300147	ENSCAFG00000007135	FRG1	Elev	Bio13	Downstream
chr7:42264018	ENSCAFG00000016988	GBA1	Bio14	Bio8	Upstream
chr10:15948439	ENSCAFG00000000525	GLIPR1	Bio14, Bio19, Hfpi	Bio8	Upstream
chr10:15948439	ENSCAFG00000023219	GLIPR1L1	Bio14, Bio19, Hfpi	Bio8	Downstream
chr10:15948439	ENSCAFG00000029990	GLIPR1L2	Bio14, Bio19, Hfpi	Bio8	Downstream
chr12:17433935	ENSCAFG00000002127	GLYATL3	Bio14, Bio19	Bio13	Downstream
chr11:66236758	ENSCAFG00000019322	GNG10 LOC119863873	Bio14	Bio13	Upstream
chr25:24159234	ENSCAFG00000007858	HAND2	Bio13	Bio8	Upstream
chr7:42264018	ENSCAFG00000016963	HCN3	Bio14	Bio8	Downstream
chr11:52471471	ENSCAFG00000029080	HRCT1	Bio19	Bio19	Downstream
chr28:27378066	ENSCAFG00000011876	HSPA12A	Bio14, Bio19, Hfpi	Hfpi	Upstream
chr1:114023508	ENSCAFG00000005588	IFNL3	Bio14, Bio19, Hfpi	Bio7	Downstream
chr34:18428630	ENSCAFG00000013297	IGF2BP2	Bio14	Bio19	Intron
chr13:49074660	ENSCAFG00000002330	IGFBP7	Bio14, Bio19, Hfpi	Bio13	Downstream
chr19:17722381	ENSCAFG00000004008	IL2	Bio14	Bio13	Upstream
chr19:17722381	ENSCAFG00000004004	IL21	Bio14	Bio13	Downstream
chr8:14217867	ENSCAFG00000059254	INSM2	Bio14, Bio19, Hfpi	Hfpi	Missense
chr15:16254754	ENSCAFG00000004875	IPO13	Bio14, Bio19	Elev	Downstream
chr8:2213672	ENSCAFG00000011056	ITPK1	Bio14	Bio13	Upstream
chr1:40988912	ENSCAFG00000000404	IYD	Bio14, Bio19, Hfpi	Bio13	Downstream
chr5:5788946	ENSCAFG00000010259	KCNJ1	Bio14, Bio19, Hfpi	Bio19	Upstream

chr5:5788946	ENSCAFG00000010255	KCNJ5	Bio14, Bio19, Hfpi	Bio19	Upstream
chr15:16254754	ENSCAFG00000004819	KLF17	Bio14, Bio19	Elev	Upstream
chr7:42264018	ENSCAFG00000017055	KRTCAP2 KCP2	Bio14	Bio8	Upstream
chr1:68537105	ENSCAFG00000001115	L3MBTL3	Bio14, Bio19, Hfpi	Bio8	Upstream
chr1:100048794	ENSCAFG00000058042	LOC102152102	Bio14, Bio19, Hfpi	Bio19	Upstream
chr1:100048794	ENSCAFG00000058939	LOC102152670	Bio14, Bio19, Hfpi	Bio19	Intron
chr31:25502664	ENSCAFG00000053510	LOC111093559	Bio14	Bio8	Upstream
chr11:52471471	ENSCAFG00000058088	LOC481603	Bio19	Bio19	Missense
chr13:58679628	ENSCAFG00000002805	LOC482174	Bio14	Bio13	Upstream
chr13:58679628	ENSCAFG00000002808	LOC482175	Bio14	Bio13	Upstream
chr13:58679628	ENSCAFG00000028617	LOC482176	Bio14	Bio13	Downstream
chr14:1591017	ENSCAFG00000048824	LOC482215	Bio14	Bio13	Upstream
chr14:1591017	ENSCAFG00000055528	LOC482217	Bio14	Bio13	Upstream
chr12:17433935	ENSCAFG00000002133	LOC610105	Bio14, Bio19	Bio13	Upstream
chr8:3231105	ENSCAFG00000011202	LRP10 REM2	Bio14, Bio19, Hfpi	Bio19	Downstream
chr12:973829	ENSCAFG00000000515	LTA TNFB TNFSF1	Bio14	Bio13	Upstream
chr12:973829	ENSCAFG00000000506	MCCD1	Bio14	Bio13	Upstream
chr20:44045441	ENSCAFG00000014350	MEF2B	Bio14, Bio19, Hfpi	Bio13	Downstream
chr10:65930498	ENSCAFG00000003233	MEIS1	Bio14	Bio13	Downstream
chr8:3231105	ENSCAFG00000011197	MMP14	Bio14, Bio19, Hfpi	Bio19	Downstream
chr12:17433935	ENSCAFG00000002111	MMUT	Bio14, Bio19	Bio13	Upstream
chr13:60225215	ENSCAFG00000051237	MOB1B	Bio14, Bio19, Hfpi	Bio13	Downstream
chr8:3231105	ENSCAFG00000011191	MRPL52	Bio14, Bio19, Hfpi	Bio19	Downstream
chr7:42264018	ENSCAFG00000016996	MTX1	Bio14	Bio8	Downstream
chr7:42264018	ENSCAFG00000017037	MUC1	Bio14	Bio8	Upstream
chr20:44045441	ENSCAFG00000014218	NCAN	Bio14, Bio19, Hfpi	Bio13	Upstream
chr1:114023508	ENSCAFG00000005593	NCCRP1	Bio14, Bio19, Hfpi	Bio7	Upstream
chr12:973829	ENSCAFG00000030599	NFKBIL1	Bio14	Bio13	Upstream
chr13:49074660	ENSCAFG00000002256	NOA1	Bio14, Bio19, Hfpi	Bio13	Downstream
chr20:44045441	ENSCAFG00000014331	NR2C2AP	Bio14, Bio19, Hfpi	Bio13	Downstream
chr20:49895596	ENSCAFG00000017400	ODAD3	Bio14, Bio19, Hfpi	Bio7	Upstream
chr11:52471471	ENSCAFG00000056577	OR13C11C	Bio19	Bio19	Upstream
chr11:52471471	ENSCAFG00000053326	OR13C7F	Bio19	Bio19	Upstream
chr11:52471471	ENSCAFG00000029096	OR13C7G	Bio19	Bio19	Downstream
chr11:52471471	ENSCAFG00000031867	OR13C7H	Bio19	Bio19	Downstream
chr11:52471471	ENSCAFG00000032491	OR13C7I	Bio19	Bio19	Upstream
chr11:52471471	ENSCAFG00000023939	OR13J1	Bio19	Bio19	Upstream
chr14:1591017	ENSCAFG00000041843	OR2AJ7	Bio14	Bio13	Upstream
chr14:1591017	ENSCAFG00000024736	OR2W3	Bio14	Bio13	Upstream
chr10:15948439	ENSCAFG00000050743	OR6C38	Bio14, Bio19, Hfpi	Bio8	Downstream
chr10:15948439	ENSCAFG00000049942	OR6C63	Bio14, Bio19, Hfpi	Bio8	Downstream

chr20:47633717	ENSCAFG00000054648	OR7A23	Bio14, Bio19, Hfpi	Bio19	Downstream
chr20:47633717	ENSCAFG00000058847	OR7A55	Bio14, Bio19, Hfpi	Bio19	Downstream
chr20:47633717	ENSCAFG00000041401	OR7A55 OR7A58	Bio14, Bio19, Hfpi	Bio19	Downstream
chr20:47633717	ENSCAFG00000057285	OR7A57	Bio14, Bio19, Hfpi	Bio19	Upstream
chr20:47633717	ENSCAFG00000041284	OR7A60	Bio14, Bio19, Hfpi	Bio19	Downstream
chr20:47633717	ENSCAFG00000050877	OR7A63	Bio14, Bio19, Hfpi	Bio19	Downstream
chr20:47633717	ENSCAFG00000051510	OR7A64	Bio14, Bio19, Hfpi	Bio19	Downstream
chr8:3231105	ENSCAFG00000011159	OXA1L	Bio14, Bio19, Hfpi	Bio19	Downstream
chr2:73943796	ENSCAFG00000012726	PAFAH2	Bio14, Bio19, Hfpi	Bio13	Upstream
chr1:114023508	ENSCAFG00000005606	PAK4	Bio14, Bio19, Hfpi	Bio7	Intron
chr16:41300147	ENSCAFG00000007113	PCM1	Elev	Bio13	Downstream
chr2:73943796	ENSCAFG00000057857	PDIK1L	Bio14, Bio19, Hfpi	Bio13	Upstream
chr7:42264018	ENSCAFG00000016959	PKLR	Bio14	Bio8	Missense
chr20:49895596	ENSCAFG00000017423	PLPPR2	Bio14, Bio19, Hfpi	Bio7	Downstream
chr13:49074660	ENSCAFG00000002296	POLR2B	Bio14, Bio19, Hfpi	Bio13	Upstream
chr6:63552340	ENSCAFG00000020319	PRKACB	Bio14, Bio19, Hfpi	Bio19	Downstream
chr37:25461648	ENSCAFG00000014957	PRKAG3	Bio14	Bio13	Upstream
chr20:49895596	ENSCAFG00000017392	PRKCSH	Bio14, Bio19, Hfpi	Bio7	Downstream
chr4:73033961	ENSCAFG00000018782	PRLR	Bio13	Bio13	Upstream
chr8:3231105	ENSCAFG00000011243	PRMT5	Bio14, Bio19, Hfpi	Bio19	Downstream
chr16:6850163	ENSCAFG00000014481	PRSS2	Bio14, Bio19, Hfpi	Shrubs	Upstream
chr30:39285101	ENSCAFG00000018073	PSTPIP1	Bio14	Bio13	Intron
chr11:66236758	ENSCAFG00000003000	PTGR1	Bio14	Bio13	Upstream
chr20:49895596	ENSCAFG00000017438	RAB3D	Bio14, Bio19, Hfpi	Bio7	Upstream
chr8:14217867	ENSCAFG00000047285	RALGAPA1	Bio14, Bio19, Hfpi	Hfpi	Downstream
chr8:3231105	ENSCAFG00000011217	RBM23	Bio14, Bio19, Hfpi	Bio19	Downstream
chr30:39285101	ENSCAFG00000018077	RCN2	Bio14	Bio13	Downstream
chr11:52471471	ENSCAFG00000002281	RECK CLTA	Bio19	Bio19	Upstream
chr13:49074660	ENSCAFG00000002254	REST	Bio14, Bio19, Hfpi	Bio13	Downstream
chr20:44045441	ENSCAFG00000014341	RFXANK	Bio14, Bio19, Hfpi	Bio13	Upstream
chr20:49895596	ENSCAFG00000017405	RGL3	Bio14, Bio19, Hfpi	Bio7	Upstream
chr12:17433935	ENSCAFG00000002141	RHAG	Bio14, Bio19	Bio13	Downstream
chr23:44584325	ENSCAFG00000008370	RNF13	Bio14, Bio19, Hfpi	Bio13	Intron
chr15:51546086	ENSCAFG00000042857	RNF175	Bio14, Bio19, Hfpi	Bio2	Upstream
chr7:42264018	ENSCAFG00000016954	RUSC1	Bio14	Bio8	Upstream
chr6:63552340	ENSCAFG00000020317	SAMD13	Bio14, Bio19, Hfpi	Bio19	Upstream
chr7:42264018	ENSCAFG00000049822	SCAMP3	Bio14	Bio8	Upstream
chr30:39285101	ENSCAFG00000018100	SCAPER	Bio14	Bio13	Upstream
chr6:22487777	ENSCAFG00000017687	SCNN1B	Bio14	Bio14	Upstream

chr6:22487777	ENSCAFG00000017690	SCNN1G	Bio14	Bio14	Downstream
chr34:18428630	ENSCAFG00000013266	SENP2	Bio14	Bio19	Downstream
chr15:51546086	ENSCAFG00000008353	SFRP2	Bio14, Bio19, Hfpi	Bio2	Intron
chr11:66236758	ENSCAFG00000003008	SHOC1	Bio14	Bio13	Downstream
chr28:27378066	ENSCAFG00000011908	SHTN1	Bio14, Bio19, Hfpi	Hfpi	Downstream
chr20:44045441	ENSCAFG00000014386	SLC25A42	Bio14, Bio19, Hfpi	Bio13	3' UTR
chr2:73943796	ENSCAFG00000012702	SLC30A2	Bio14, Bio19, Hfpi	Bio13	Downstream
chr13:60225215	ENSCAFG00000002939	SLC4A4	Bio14, Bio19, Hfpi	Bio13	Upstream
chr15:16254754	ENSCAFG00000004826	SLC6A9	Bio14, Bio19	Elev	Missense
chr8:3231105	ENSCAFG00000011177	SLC7A7	Bio14, Bio19, Hfpi	Bio19	Upstream
chr8:43694299	ENSCAFG00000023919	SLC8A3	Bio14, Bio19	Bio13	Downstream
chr22:39352549	ENSCAFG00000005293	SLITRK5	Bio14, Bio19	Bio8	Upstream
chr8:43694299	ENSCAFG00000016602	SMOC1	Bio14, Bio19	Bio13	Intron
chr15:16254754	ENSCAFG00000004887	ST3GAL3	Bio14, Bio19	Elev	Downstream
chr2:73943796	ENSCAFG00000029191	STMN1	Bio14, Bio19, Hfpi	Bio13	Upstream
chr20:44045441	ENSCAFG00000014431	SUGP2	Bio14, Bio19, Hfpi	Bio13	Upstream
chr20:49895596	ENSCAFG00000058807	SWSAP1	Bio14, Bio19, Hfpi	Bio7	Downstream
chr1:114023508	ENSCAFG00000005592	SYCN	Bio14, Bio19, Hfpi	Bio7	Downstream
chr7:42264018	ENSCAFG00000017017	THBS3	Bio14	Bio8	Upstream
chr11:5292792	ENSCAFG00000053610	TICAM2	Bio14, Bio19, Hfpi	Bio13	Downstream
chr15:51546086	ENSCAFG00000008351	TLR2	Bio14, Bio19, Hfpi	Bio2	Downstream
chr11:5292792	ENSCAFG00000000209	TMED7	Bio14, Bio19, Hfpi	Bio13	Downstream
chr20:44045441	ENSCAFG00000014369	TMEM161A	Bio14, Bio19, Hfpi	Bio13	Downstream
chr20:49895596	ENSCAFG00000017434	TMEM205	Bio14, Bio19, Hfpi	Bio7	Upstream
chr7:42264018	ENSCAFG00000017047	TRIM46	Bio14	Bio8	Downstream
chr14:1591017	ENSCAFG00000001226	TRIM58	Bio14	Bio13	Intron
chr2:73943796	ENSCAFG00000012671	TRIM63	Bio14, Bio19, Hfpi	Bio13	Downstream
chr30:39285101	ENSCAFG00000018068	TSPAN3	Bio14	Bio13	Downstream
chr6:63552340	ENSCAFG00000020315	UOX	Bio14, Bio19, Hfpi	Bio19	Downstream
chr37:25461648	ENSCAFG00000014987	WNT10A	Bio14	Bio13	Upstream
chr37:25461648	ENSCAFG00000031321	WNT6	Bio14	Bio13	Downstream
chr11:3681562	ENSCAFG00000000224	YTHDC2	Bio14, Bio19, Hfpi	Bio13	Downstream
chr33:19263277	ENSCAFG00000010823	ZBTB20	Bio14	Bio2	Intron
chr11:66236758	ENSCAFG00000002996	ZNF483	Bio14	Bio13	Downstream
chr20:49895596	ENSCAFG00000031712	ZNF653	Bio14, Bio19, Hfpi	Bio7	Downstream
Grey wolves					
chr20:53567088	ENSCAFG00000018625	C3	Bio16, Elev	Bio18	Synonymous
chr20:53567088	ENSCAFG00000018620	GPR108	Bio16, Elev	Bio18	Downstream
chr27:34122311	ENSCAFG00000013324	LRP6	Bio3, Bio8, Bio16, Bio19, Hfpi	Bio9	Synonymous
chr2:73819217	ENSCAFG00000057857	PDIK1L	Bio16, Elev	Bio18	Downstream

chr18:41630019	ENSCAFG00000008279	PTPRJ	Bio16	Bio16	Missense
chr3:48323401	ENSCAFG00000011241	SLCO3A1	Bio3, Bio7, Elev	Bio19	Upstream
chr20:53567088	ENSCAFG00000018602	SH2D3A	Bio16, Elev	Bio18	Downstream
chr20:53567088	ENSCAFG00000018627	TNFSF14	Bio16, Elev	Bio18	Upstream
chr2:73819217	ENSCAFG00000012671	TRIM63	Bio16, Elev	Bio18	Upstream
chr20:53567088	ENSCAFG00000018610	TRIP10	Bio16, Elev	Bio18	Upstream
chr20:53567088	ENSCAFG00000018590	VAV1	Bio16, Elev	Bio18	Upstream

Supplementary table 9. Uncharacterized and genes without UnitProt codes, displaying signatures of both local adaptation and diversifying selection

CanFam3 SNP location	ENSMBL ID	Gene name/ type	LFMM	RDA	Variant consequence
chr16:14048745	ENSCAFG00000058902	Cornichon family AMPA receptor auxiliary protein 1	Bio14, Bio19, Hfpi	Bio2	Downstream
chr10:15948439	ENSCAFG00000051032	Glyceraldehyde 3-phosphate dehydrogenase NAD(P) binding domain-containing protein	Bio14, Bio19, Hfpi	Bio8	Missense
chr8:2213672	ENSCAFG00000058699	Ig-like domain-containing protein	Bio14	Bio13	Upstream
chr8:2213672	ENSCAFG00000055775	Ig-like domain-containing protein	Bio14	Bio13	Upstream
chr8:2213672	ENSCAFG00000057398	Ig-like domain-containing protein	Bio14	Bio13	Upstream
chr8:2213672	ENSCAFG00000045242	Ig-like domain-containing protein	Bio14	Bio13	Upstream
chr8:2213672	ENSCAFG00000054905	Ig-like domain-containing protein	Bio14	Bio13	Upstream
chr16:6850163	ENSCAFG00000014478	Ig-like domain-containing protein	Bio14, Bio19, Hfpi	Shrubs	Upstream
chr16:6850163	ENSCAFG00000045360	Ig-like domain-containing protein	Bio14, Bio19, Hfpi	Shrubs	Upstream
chr16:6850163	ENSCAFG00000056839	Ig-like domain-containing protein	Bio14, Bio19, Hfpi	Shrubs	Upstream
chr8:2213672	ENSCAFG00000049482	Ig-like domain-containing protein	Bio14	Bio13	Downstream
chr16:6850163	ENSCAFG00000003811	Ig-like domain-containing protein	Bio14, Bio19, Hfpi	Shrubs	Downstream
chr16:6850163	ENSCAFG00000024810	Ig-like domain-containing protein	Bio14, Bio19, Hfpi	Shrubs	Downstream
chr16:6850163	ENSCAFG00000046664	Ig-like domain-containing protein	Bio14, Bio19, Hfpi	Shrubs	Downstream
chr16:6850163	ENSCAFG00000055018	Ig-like domain-containing protein	Bio14, Bio19, Hfpi	Shrubs	Downstream
chr16:6850163	ENSCAFG00000057117	Ig-like domain-containing protein	Bio14, Bio19, Hfpi	Shrubs	Downstream
chr16:6850163	ENSCAFG00000052806	Ig-like domain-containing protein	Bio14, Bio19, Hfpi	Shrubs	Downstream
chr16:6850163	ENSCAFG00000057504	Ig-like domain-containing protein	Bio14, Bio19, Hfpi	Shrubs	Downstream
chr16:6850163	ENSCAFG00000042586	Ig-like domain-containing protein	Bio14, Bio19, Hfpi	Shrubs	Intron
chr16:6850163	ENSCAFG00000058265	Immunoglobulin C1-set domain-containing protein	Bio14, Bio19, Hfpi	Shrubs	Upstream
chr8:2213672	ENSCAFG00000050263	Immunoglobulin V-set domain-containing protein	Bio14	Bio13	Upstream
chr16:6850163	ENSCAFG00000031359	Immunoglobulin V-set domain-containing protein	Bio14, Bio19, Hfpi	Shrubs	Upstream
chr16:6850163	ENSCAFG00000043031	Immunoglobulin V-set domain-containing protein	Bio14, Bio19, Hfpi	Shrubs	Upstream
chr16:6850163	ENSCAFG00000029837	Immunoglobulin V-set domain-containing protein	Bio14, Bio19, Hfpi	Shrubs	Downstream
chr16:6850163	ENSCAFG00000055372	Immunoglobulin V-set domain-containing protein	Bio14, Bio19, Hfpi	Shrubs	Downstream
chr8:2213672	ENSCAFG00000055600	Immunoglobulin V-set domain-containing protein	Bio14	Bio13	5' UTR
chr1:114023508	ENSCAFG00000041829	Interferon lambda-4-like	Bio14, Bio19, Hfpi	Bio7	Upstream
chr1:114023508	ENSCAFG00000052759	Interferon lambda-4-like	Bio14, Bio19, Hfpi	Bio7	Downstream
chr1:100048794	ENSCAFG00000042267	KRAB domain-containing protein	Bio14, Bio19, Hfpi	Bio19	Intron
chr12:973829	ENSCAFG00000051129	MHC class I-like antigen recognition-like domain-containing protein	Bio14	Bio13	Downstream

chr1:100048794	ENSCAFG00000049889	Mitochondrial import inner membrane translocase subunit TIM23	Bio14, Bio19, Hfpi	Bio19	Downstream
chr20:47633717	ENSCAFG00000047151	Olfactory receptor	Bio14, Bio19, Hfpi	Bio19	Synonymous
chr30:23187554	ENSCAFG00000016551	Prefoldin subunit 4	Bio14	Bio19	Upstream
chr30:39285101	ENSCAFG00000053426	Protein kinase domain-containing protein	Bio14	Bio13	Downstream
chr8:2213672	ENSCAFG00000044478	Small ribosomal subunit protein uS15 N-terminal domain-containing protein	Bio14	Bio13	Upstream
chr6:63552340	ENSCAFG00000045033	Uncharacterized protein	Bio14, Bio19, Hfpi	Bio19	Upstream
chr11:5292792	ENSCAFG00000051647	Uncharacterized protein	Bio14, Bio19, Hfpi	Bio13	Upstream
chr23:44584325	ENSCAFG00000041808	Uncharacterized protein	Bio14, Bio19, Hfpi	Bio13	Upstream
chr1:100048794	ENSCAFG00000046247	Zinc finger protein 671	Bio14, Bio19, Hfpi	Bio19	Downstream
chr1:100048794	ENSCAFG00000028550	Zinc finger protein 850-like	Bio14, Bio19, Hfpi	Bio19	Upstream

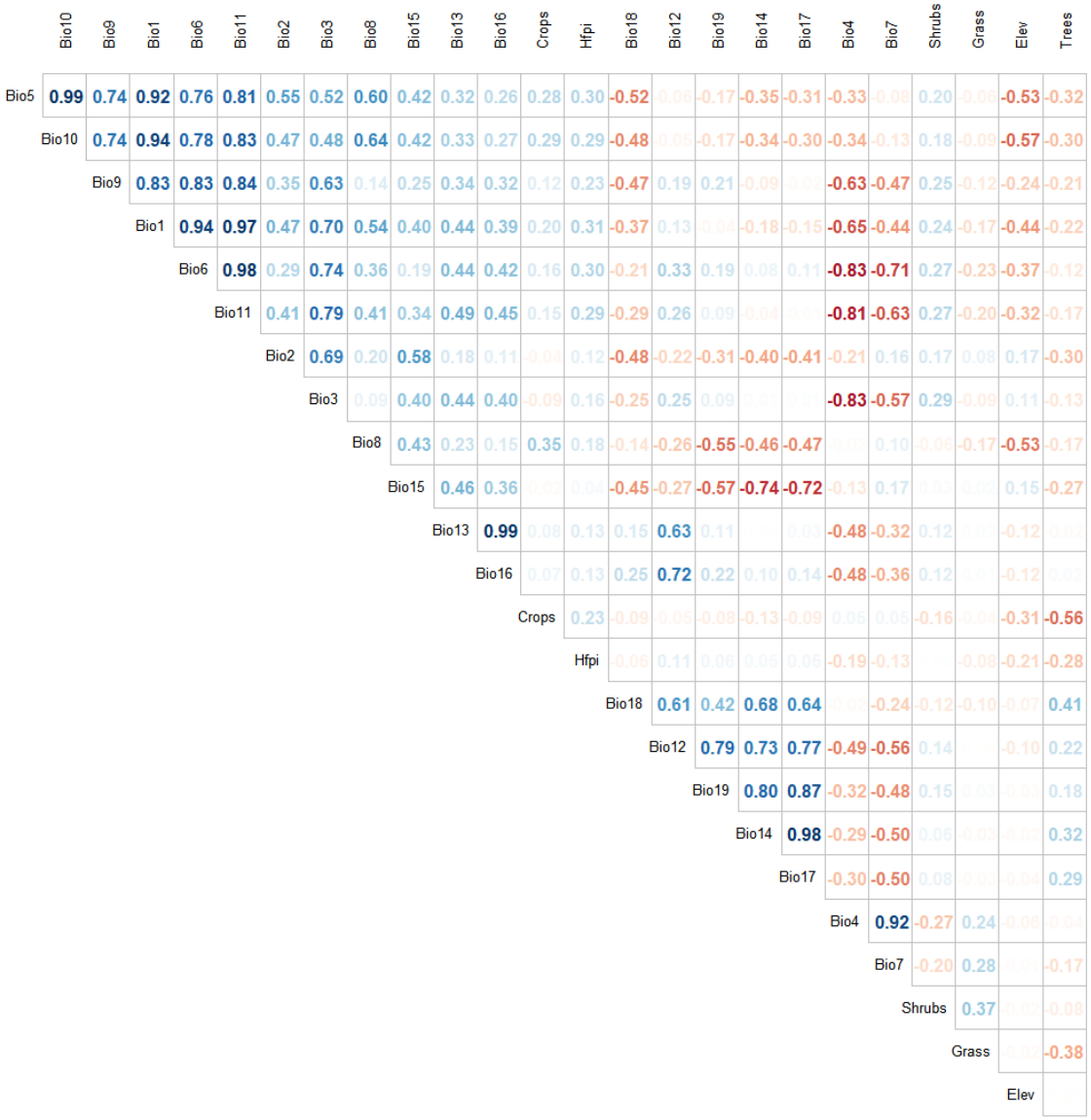
Supplementary table 10. RNA and pseudogenes displaying signatures of both local adaptation and diversifying selection

Free ranging dogs					
CanFam3 SNP location	**ENSMBL ID**	**Gene name/ type**	**LFMM**	**RDA**	**Variant consequence**
chr37:25461648	ENSCAFG00000014984	processed pseudogene	Bio14	Bio13	Downstream
chr31:25502664	ENSCAFG00000049146	pseudogene	Bio14	Bio8	Upstream
chr31:25502664	ENSCAFG00000049284	pseudogene	Bio14	Bio8	Upstream
chr31:25502664	ENSCAFG00000043459	pseudogene	Bio14	Bio8	Upstream
chr1:100048794	ENSCAFG00000041221	pseudogene	Bio14, Bio19, Hfpi	Bio19	Downstream
chr17:41433947	ENSCAFG00000047122	pseudogene	Bio14, Bio19, Hfpi	Bio13	Downstream
chr20:47633717	ENSCAFG00000056490	pseudogene	Bio14, Bio19, Hfpi	Bio19	Downstream
chr31:25502664	ENSCAFG00000044330	pseudogene	Bio14	Bio8	Downstream
chr31:25502664	ENSCAFG00000043307	pseudogene	Bio14	Bio8	Downstream
chr1:100048794	ENSCAFG00000057961	lncRNA	Bio14, Bio19, Hfpi	Bio19	Upstream
chr2:73943796	ENSCAFG00000054112	lncRNA	Bio14, Bio19, Hfpi	Bio13	Upstream
chr5:5788946	ENSCAFG00000054466	lncRNA	Bio14, Bio19, Hfpi	Bio19	Upstream
chr10:15948439	ENSCAFG00000054033	lncRNA	Bio14, Bio19, Hfpi	Bio8	Upstream
chr10:15948439	ENSCAFG00000050476	lncRNA	Bio14, Bio19, Hfpi	Bio8	Upstream
chr11:5292792	ENSCAFG00000052727	lncRNA	Bio14, Bio19, Hfpi	Bio13	Upstream
chr11:5292792	ENSCAFG00000054065	lncRNA	Bio14, Bio19, Hfpi	Bio13	Upstream
chr11:66236758	ENSCAFG00000055787	lncRNA	Bio14	Bio13	Upstream
chr12:973829	ENSCAFG00000051921	lncRNA	Bio14	Bio13	Upstream
chr15:51546086	ENSCAFG00000038294	lncRNA	Bio14, Bio19, Hfpi	Bio2	Upstream
chr16:6850163	ENSCAFG00000056504	lncRNA	Bio14, Bio19, Hfpi	Shrubs	Upstream
chr20:44045441	ENSCAFG00000052308	lncRNA	Bio14, Bio19, Hfpi	Bio13	Upstream
chr22:39352549	ENSCAFG00000047324	lncRNA	Bio14, Bio19	Bio8	Upstream
chr22:39352549	ENSCAFG00000055031	lncRNA	Bio14, Bio19	Bio8	Upstream
chr22:39352549	ENSCAFG00000056699	lncRNA	Bio14, Bio19	Bio8	Upstream
chr22:39352549	ENSCAFG00000052713	lncRNA	Bio14, Bio19	Bio8	Upstream
chr25:24242325	ENSCAFG00000051637	lncRNA	Bio14, Bio19	Bio13	Upstream
chr25:24242325	ENSCAFG00000057445	lncRNA	Bio14, Bio19	Bio13	Upstream
chr25:24242325	ENSCAFG00000053959	lncRNA	Bio14, Bio19	Bio13	Upstream
chr25:24242325	ENSCAFG00000057071	lncRNA	Bio14, Bio19	Bio13	Upstream
chr25:24242325	ENSCAFG00000051070	lncRNA	Bio14, Bio19	Bio13	Upstream
chr25:24242325	ENSCAFG00000056859	lncRNA	Bio14, Bio19	Bio13	Upstream
chr25:24242325	ENSCAFG00000050366	lncRNA	Bio14, Bio19	Bio13	Upstream
chr30:39285101	ENSCAFG00000057488	lncRNA	Bio14	Bio13	Upstream
chr37:25461648	ENSCAFG00000053720	lncRNA	Bio14	Bio13	Upstream
chr4:73033961	ENSCAFG00000050516	lncRNA	Bio13	Bio13	Upstream
chr4:73033961	ENSCAFG00000050153	lncRNA	Bio13	Bio13	Intron, non-coding transcript
chr5:5788946	ENSCAFG00000057785	lncRNA	Bio14, Bio19, Hfpi	Bio19	Downstream

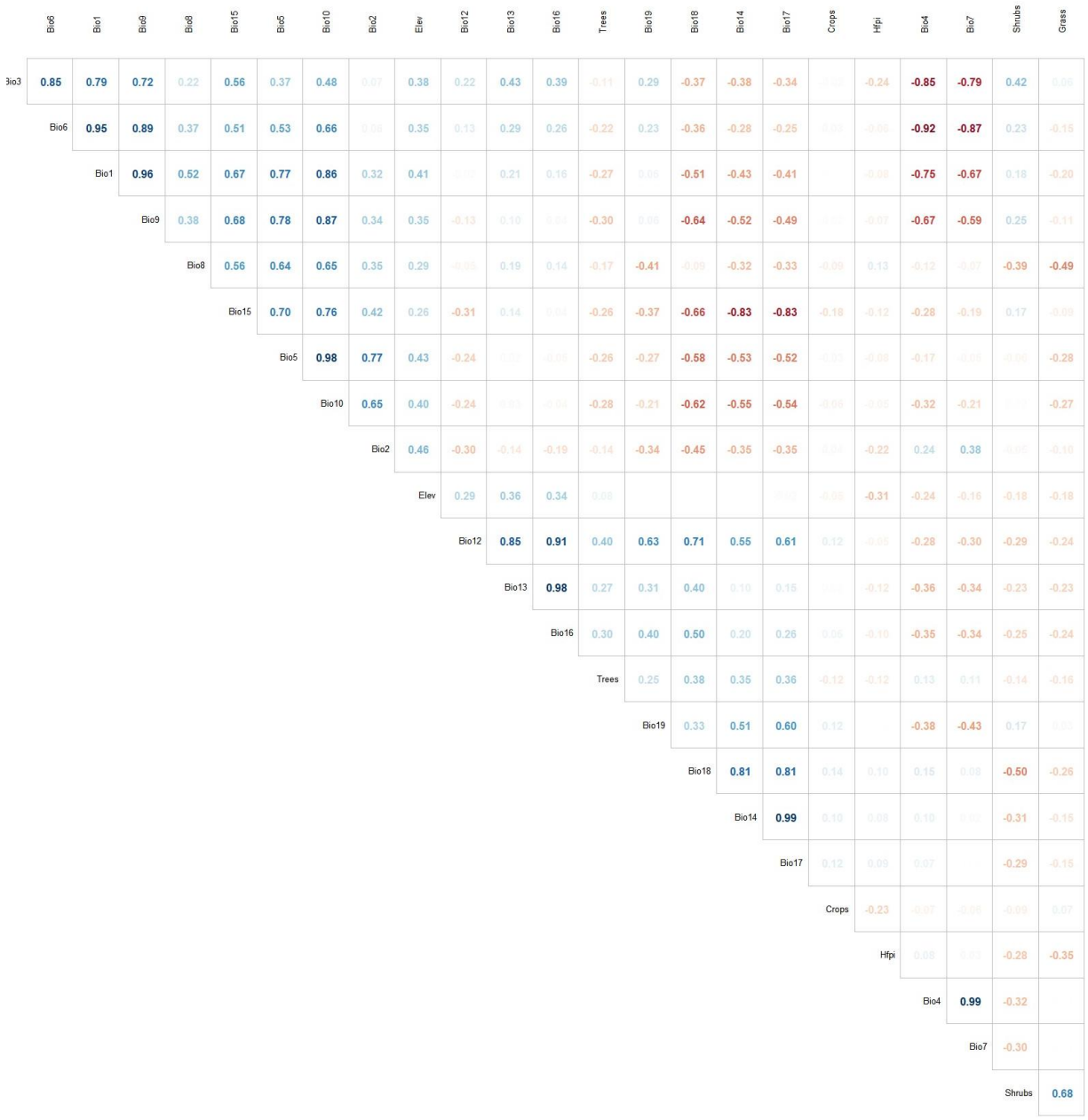
chr5:5788946	ENSCAFG00000059318	lncRNA	Bio14, Bio19, Hfpi	Bio19	Downstream
chr8:2213672	ENSCAFG00000059058	lncRNA	Bio14	Bio13	Downstream
chr8:3231105	ENSCAFG00000052972	lncRNA	Bio14, Bio19, Hfpi	Bio19	Downstream
chr8:14217867	ENSCAFG00000055439	lncRNA	Bio14, Bio19, Hfpi	Hfpi	Downstream
chr8:14217867	ENSCAFG00000053648	lncRNA	Bio14, Bio19, Hfpi	Hfpi	Downstream
chr11:5292792	ENSCAFG00000054099	lncRNA	Bio14, Bio19, Hfpi	Bio13	Downstream
chr11:5292792	ENSCAFG00000056062	lncRNA	Bio14, Bio19, Hfpi	Bio13	Downstream
chr11:66236758	ENSCAFG00000053075	lncRNA	Bio14	Bio13	Downstream
chr13:60225215	ENSCAFG00000057075	lncRNA	Bio14, Bio19, Hfpi	Bio13	Downstream
chr20:49895596	ENSCAFG00000058863	lncRNA	Bio14, Bio19, Hfpi	Bio7	Downstream
chr22:39352549	ENSCAFG00000059098	lncRNA	Bio14, Bio19	Bio8	Downstream
chr22:39352549	ENSCAFG00000057930	lncRNA	Bio14, Bio19	Bio8	Downstream
chr22:39352549	ENSCAFG00000053565	lncRNA	Bio14, Bio19	Bio8	Downstream
chr22:39352549	ENSCAFG00000054596	lncRNA	Bio14, Bio19	Bio8	Downstream
chr22:39352549	ENSCAFG00000057389	lncRNA	Bio14, Bio19	Bio8	Downstream
chr22:39352549	ENSCAFG00000051752	lncRNA	Bio14, Bio19	Bio8	Downstream
chr22:39352549	ENSCAFG00000056979	lncRNA	Bio14, Bio19	Bio8	Downstream
chr25:24242325	ENSCAFG00000050225	lncRNA	Bio14, Bio19	Bio13	Downstream
chr25:24242325	ENSCAFG00000053899	lncRNA	Bio14, Bio19	Bio13	Downstream
chr25:24159234	ENSCAFG00000058374	lncRNA	Bio13	Bio8	Downstream
chr25:24159234	ENSCAFG00000050237	lncRNA	Bio13	Bio8	Upstream
chr25:24242325	ENSCAFG00000058856	lncRNA	Bio14, Bio19	Bio13	Downstream
chr25:24242325	ENSCAFG00000053060	lncRNA	Bio14, Bio19	Bio13	Downstream
chr25:24242325	ENSCAFG00000052304	lncRNA	Bio14, Bio19	Bio13	Downstream
chr25:24242325	ENSCAFG00000053712	lncRNA	Bio14, Bio19	Bio13	Downstream
chr25:24242325	ENSCAFG00000053492	lncRNA	Bio14, Bio19	Bio13	Downstream
chr25:24242325	ENSCAFG00000059337	lncRNA	Bio14, Bio19	Bio13	Downstream
chr25:24242325	ENSCAFG00000055388	lncRNA	Bio14, Bio19	Bio13	Downstream
chr25:24242325	ENSCAFG00000053745	lncRNA	Bio14, Bio19	Bio13	Downstream
chr25:24242325	ENSCAFG00000056038	lncRNA	Bio14, Bio19	Bio13	Downstream
chr30:39285101	ENSCAFG00000057251	lncRNA	Bio14	Bio13	Downstream
chr37:25461648	ENSCAFG00000051796	lncRNA	Bio14	Bio13	Downstream
chr7:42264018	ENSCAFG00000052519	miRNA	Bio14	Bio8	Downstream
chr13:60225215	ENSCAFG00000026082	snoRNA	Bio14, Bio19, Hfpi	Bio13	Upstream
chr12:973829	ENSCAFG00000020981	snoRNA	Bio14	Bio13	Downstream
chr12:973829	ENSCAFG00000026284	snoRNA	Bio14	Bio13	Downstream
chr8:3231105	ENSCAFG00000022520	snRNA	Bio14, Bio19, Hfpi	Bio19	Upstream
chr13:49074660	ENSCAFG00000022253	snRNA	Bio14, Bio19, Hfpi	Bio13	Upstream
chr26:1215156	ENSCAFG00000053965	snRNA	Bio14	Bio8	Upstream
chr4:73033961	ENSCAFG00000027148	snRNA	Bio13	Bio13	Downstream
chr8:2213672	ENSCAFG00000055381	snRNA	Bio14	Bio13	Downstream
chr13:49074660	ENSCAFG00000043782	snRNA	Bio14, Bio19, Hfpi	Bio13	Downstream
chr25:24242325	ENSCAFG00000043087	snRNA	Bio14, Bio19	Bio13	Downstream

Grey wolves					
chr2:73819217	ENSCAFG00000054112	lncRNA	Bio16, Elev	Bio18	Downstream/Upstream
chr27:34122311	ENSCAFG00000059447	lncRNA	Bio3, Bio8, Bio16, Bio19, Hfpi	Bio9	Upstream
chr20:53567088	ENSCAFG00000054483	miRNA	Bio16, Elev	Bio18	Upstream
chr20:53567088	ENSCAFG00000025934	snRNA	Bio16, Elev	Bio18	Downstream

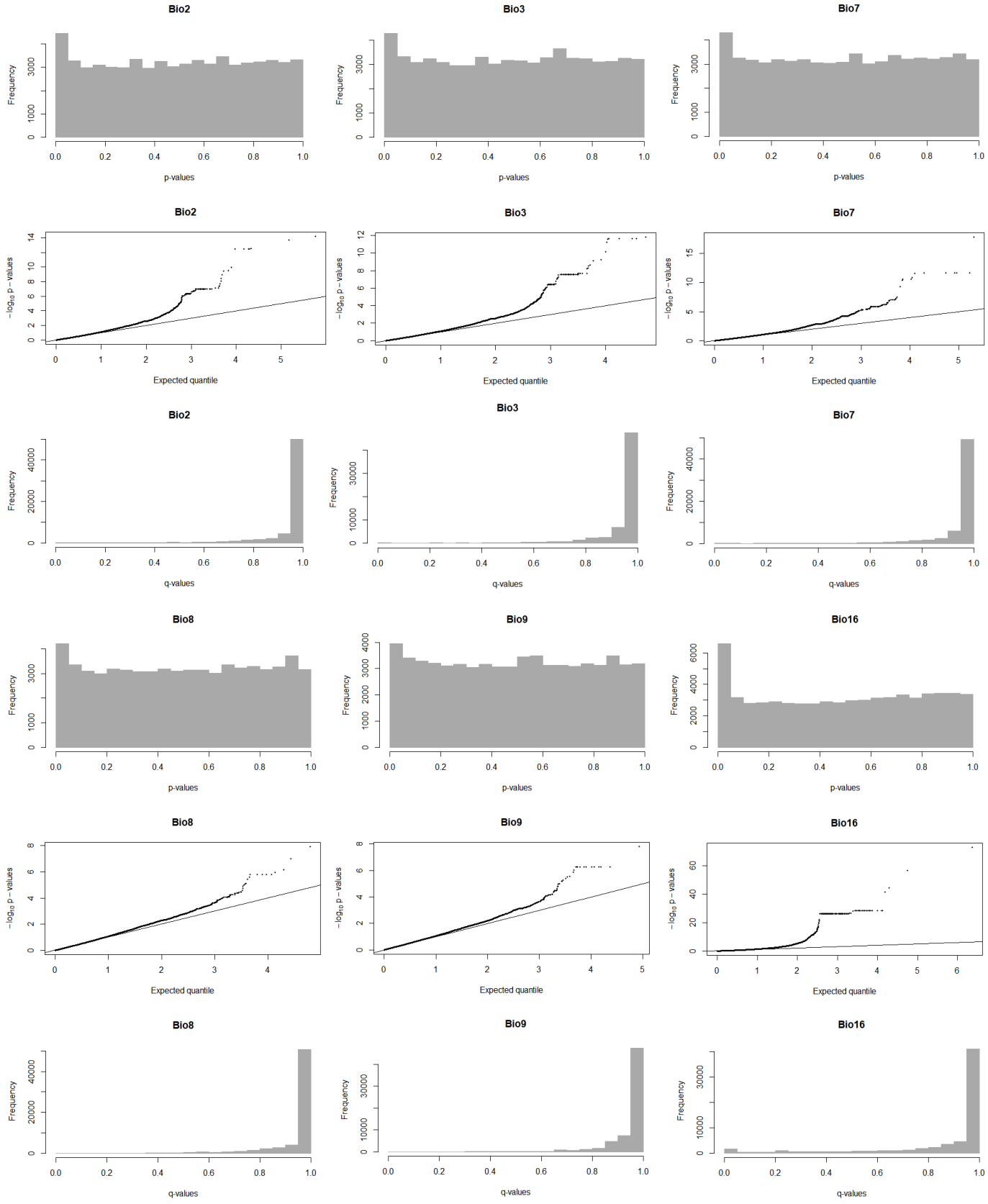
Supplementary figures

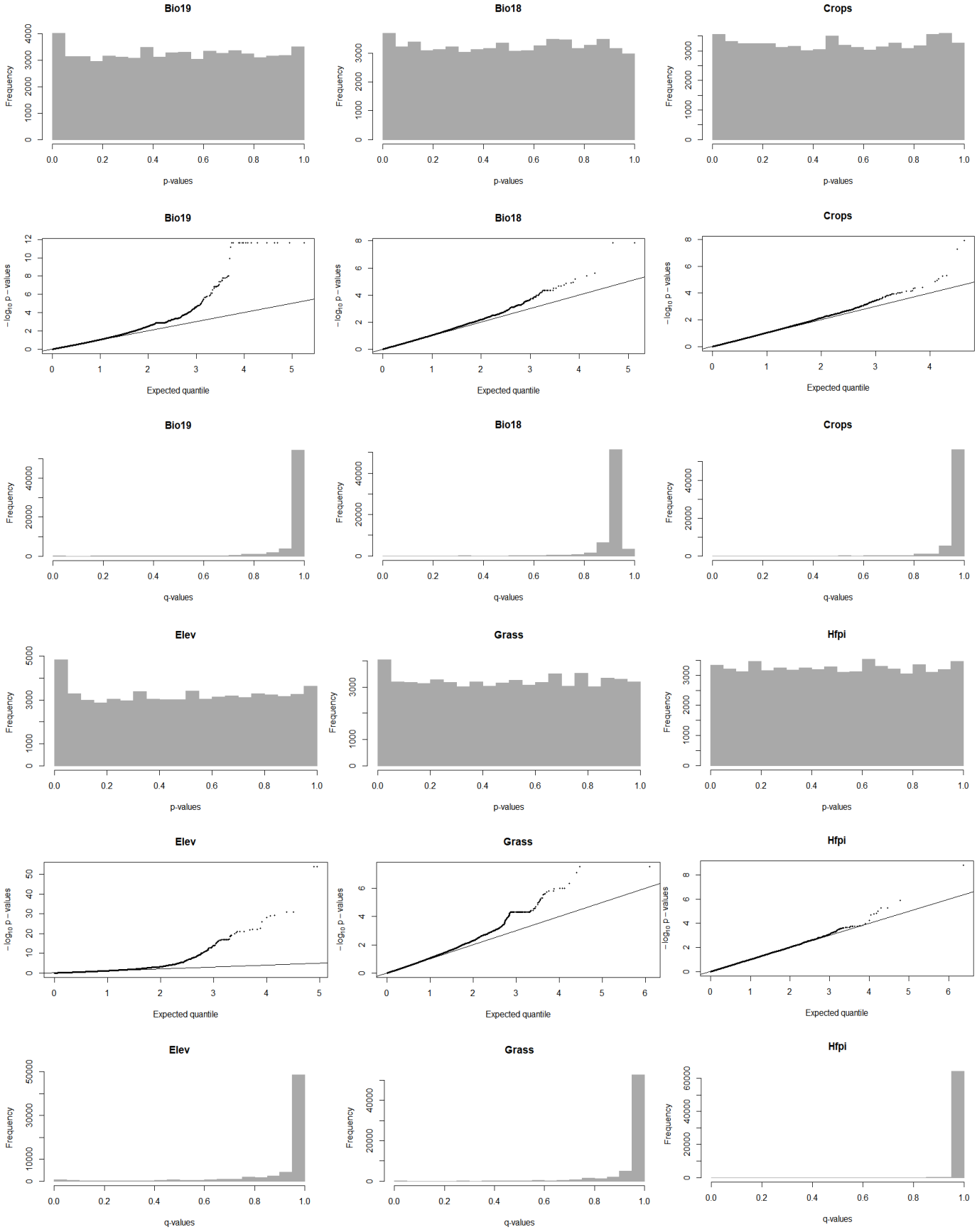


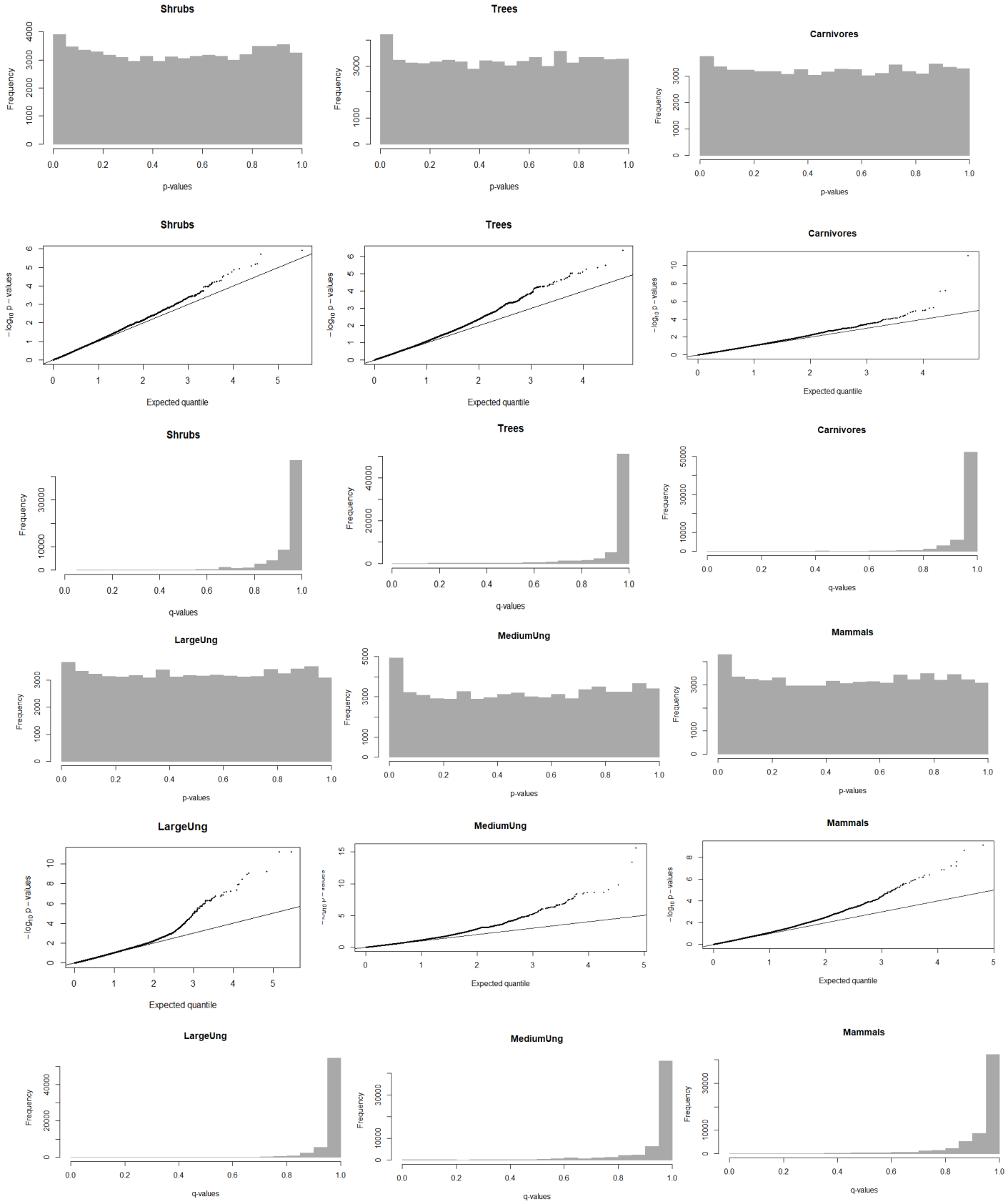
Supplementary figure 1. Corrplot for collinearity for the BioClim environmental predictors wolves. A threshold of 70% was used for removing environmental layers. The colour gradient informs on the collinearity prediction blue (collinear) to red (not collinear). Variables with collinearity values over the threshold which were removed, i.e., Bio1, Bio4, Bio5, Bio6, Bio10, Bio11, Bio12, Bio13, Bio14, Bio15 and Bio17.



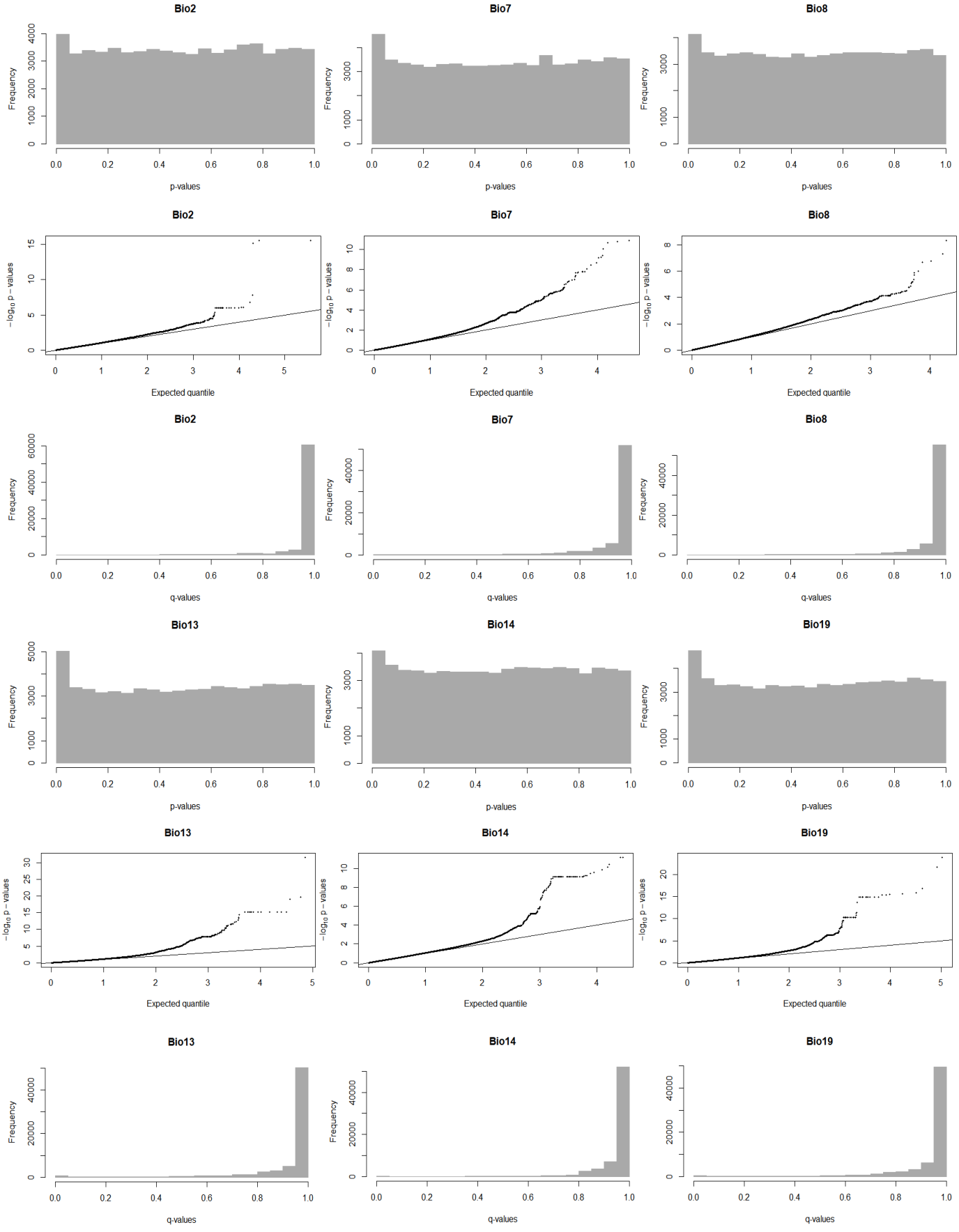
Supplementary figure 2. Corrplot for collinearity for the BioClim environmental predictors FRDs. A threshold of 70% was used for removing environmental layers. The colour gradient informs on the collinearity prediction blue (collinear) to red (not collinear). Variables with collinearity values over the threshold were removed i.e., Bio1, Bio3, Bio4, Bio5, Bio6, Bio9, Bio10, Bio11, Bio12, Bio15, Bio16, Bio17 and Bio18.

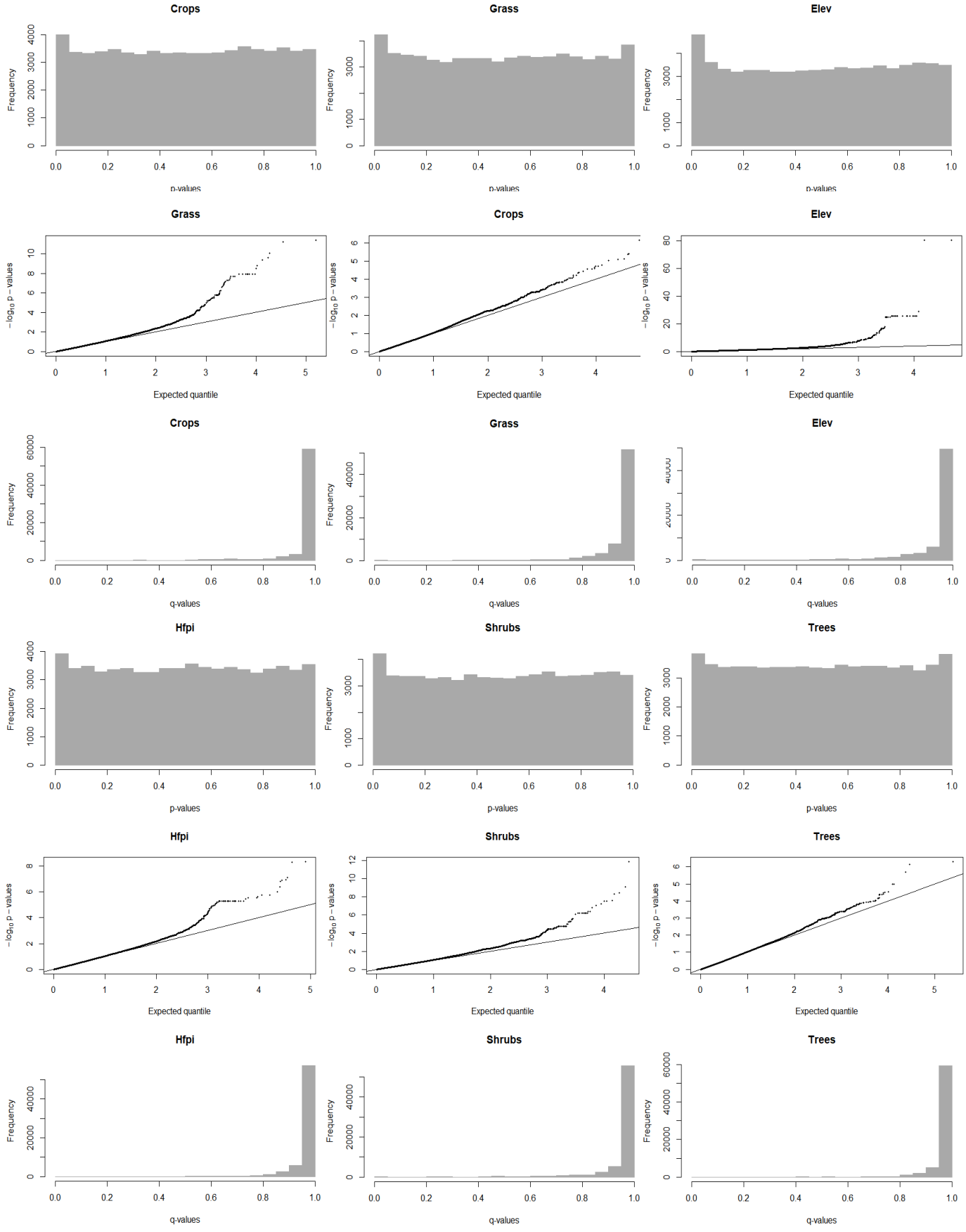

Bio2
Frequency
p-values
Bio3
Frequency
p-values
Bio7
Frequency
p-values
Bio2
−log10 p−values
Expected quantile
Bio3
−log10 p−values
Expected quantile
Bio7
−log10 p−values
Expected quantile
Bio2
Frequency
q-values
Bio3
Frequency
q-values
Bio7
Frequency
q-values
Bio8
Frequency
p-values
Bio9
Frequency
p-values
Bio16
Frequency
p-values
Bio8
−log10 p−values
Expected quantile
Bio9
−log10 p−values
Expected quantile
Bio16
−log10 p−values
Expected quantile
Bio8
Frequency
q-values
Bio9
Frequency
q-values
Bio16
Frequency
q-values

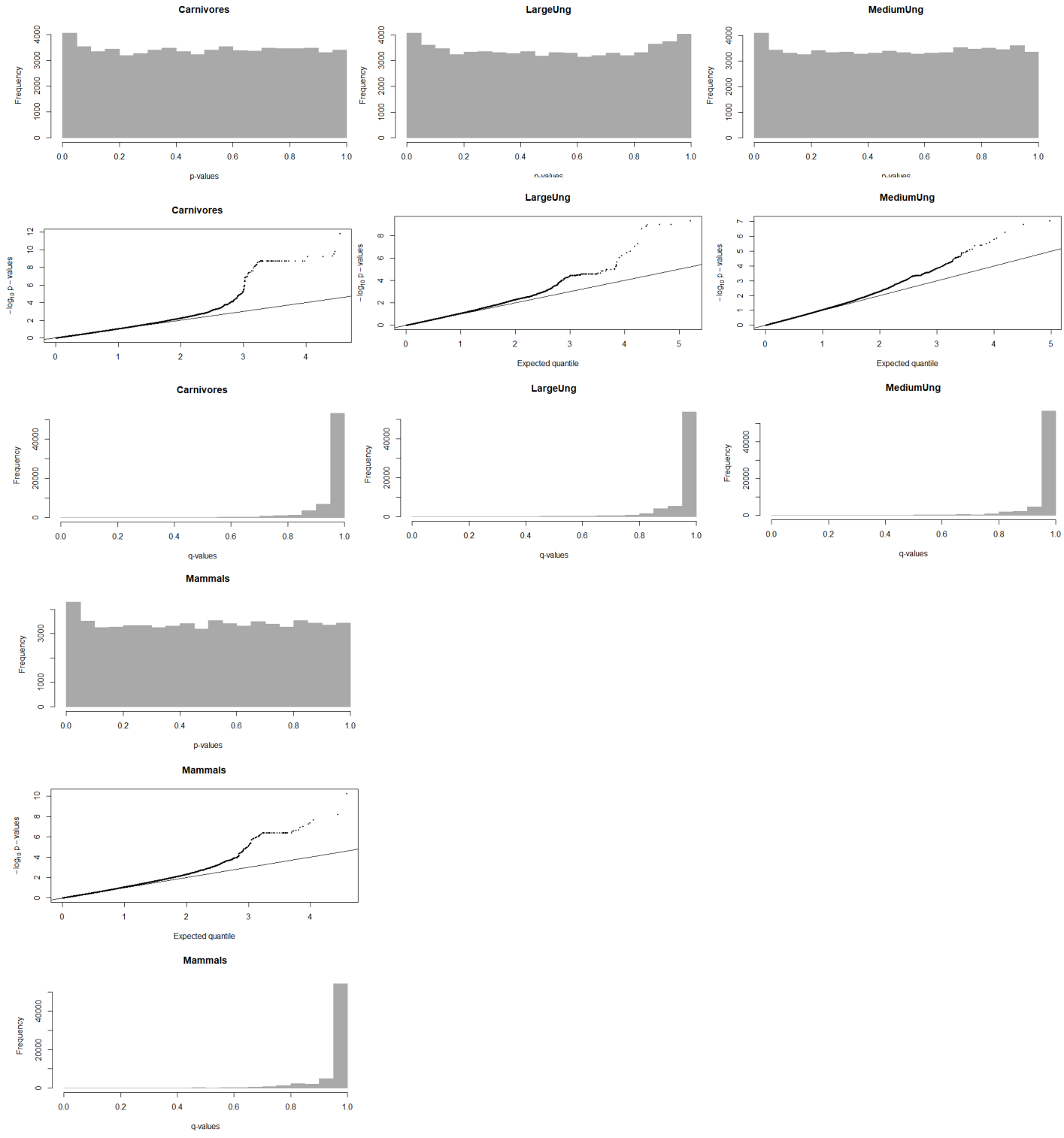

Bio19
Frequency
p-values
Bio18
Frequency
p-values
Crops
Frequency
p-values
Bio19
−log10 p−values
Expected quantile
Bio18
−log10 p−values
Expected quantile
Crops
−log10 p−values
Expected quantile
Bio19
Frequency
q-values
Bio18
Frequency
q-values
Crops
Frequency
q-values
Elev
Frequency
p-values
Grass
Frequency
p-values
Hfpi
Frequency
p-values
Elev
−log10 p−values
Expected quantile
Grass
−log10 p−values
Expected quantile
Hfpi
−log10 p−values
Expected quantile
Elev
Frequency
q-values
Grass
Frequency
q-values
Hfpi
Frequency
q-values



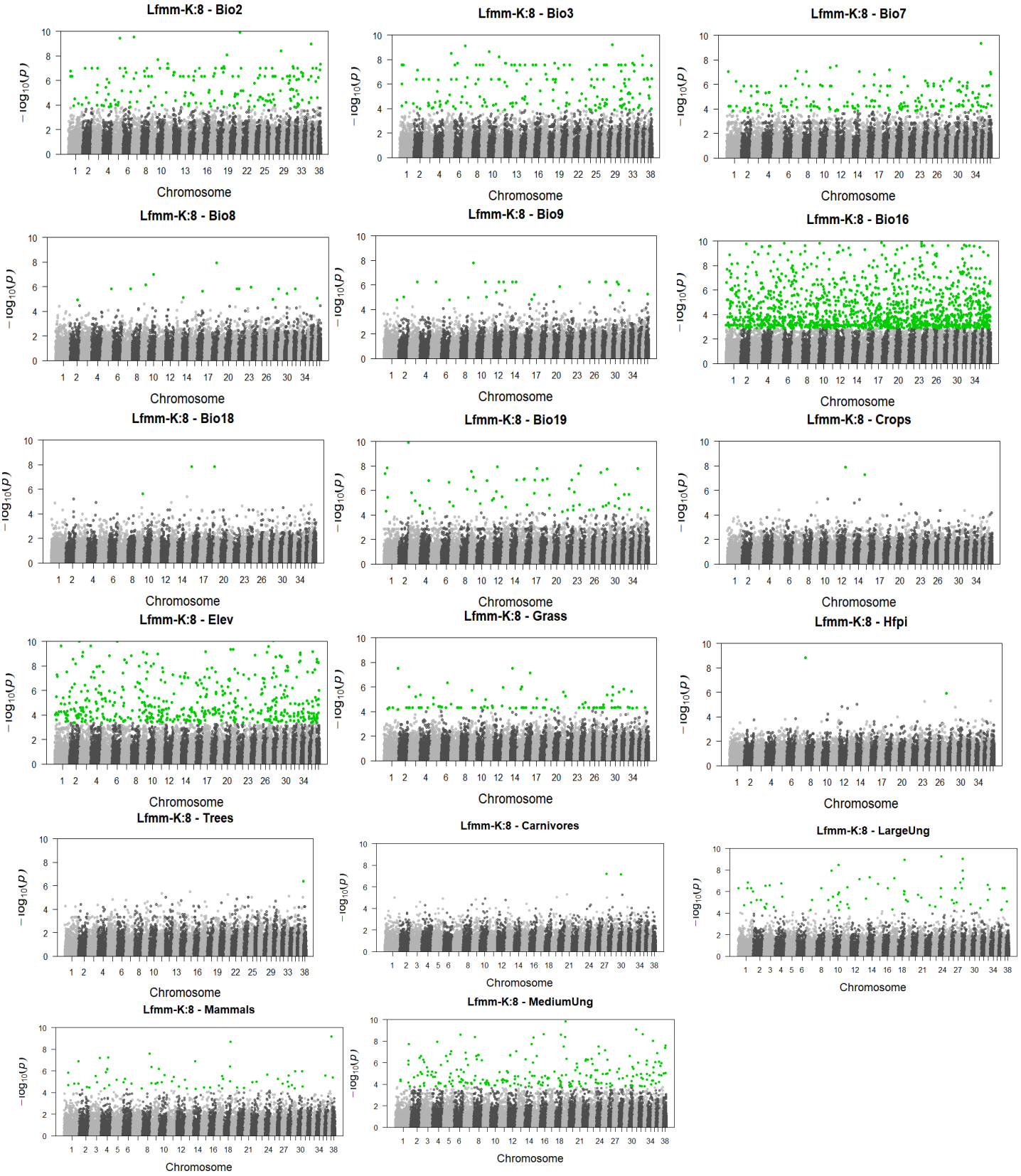
Supplementary figure 3. Latent factor mixed model (LFMM) plots wolves. Histogram plots for p-values as calculated with lfmm (top), quantile-quantile plots (middle) and q-value histograms (bottom) for each of the environmental predictors included from WorldClim and mammalian species richness from both LFMM analysis.


Bio2
Bio7
Bio8
Frequency
p-values
Bio2
Bio7
Bio8
−log10 p−values
Expected quantile
Bio2
Bio7
Bio8
Frequency
q-values
Bio13
Bio14
Bio19
Frequency
p-values
Bio13
Bio14
Bio19
−log10 p−values
Expected quantile
Bio13
Bio14
Bio19
Frequency
q-values

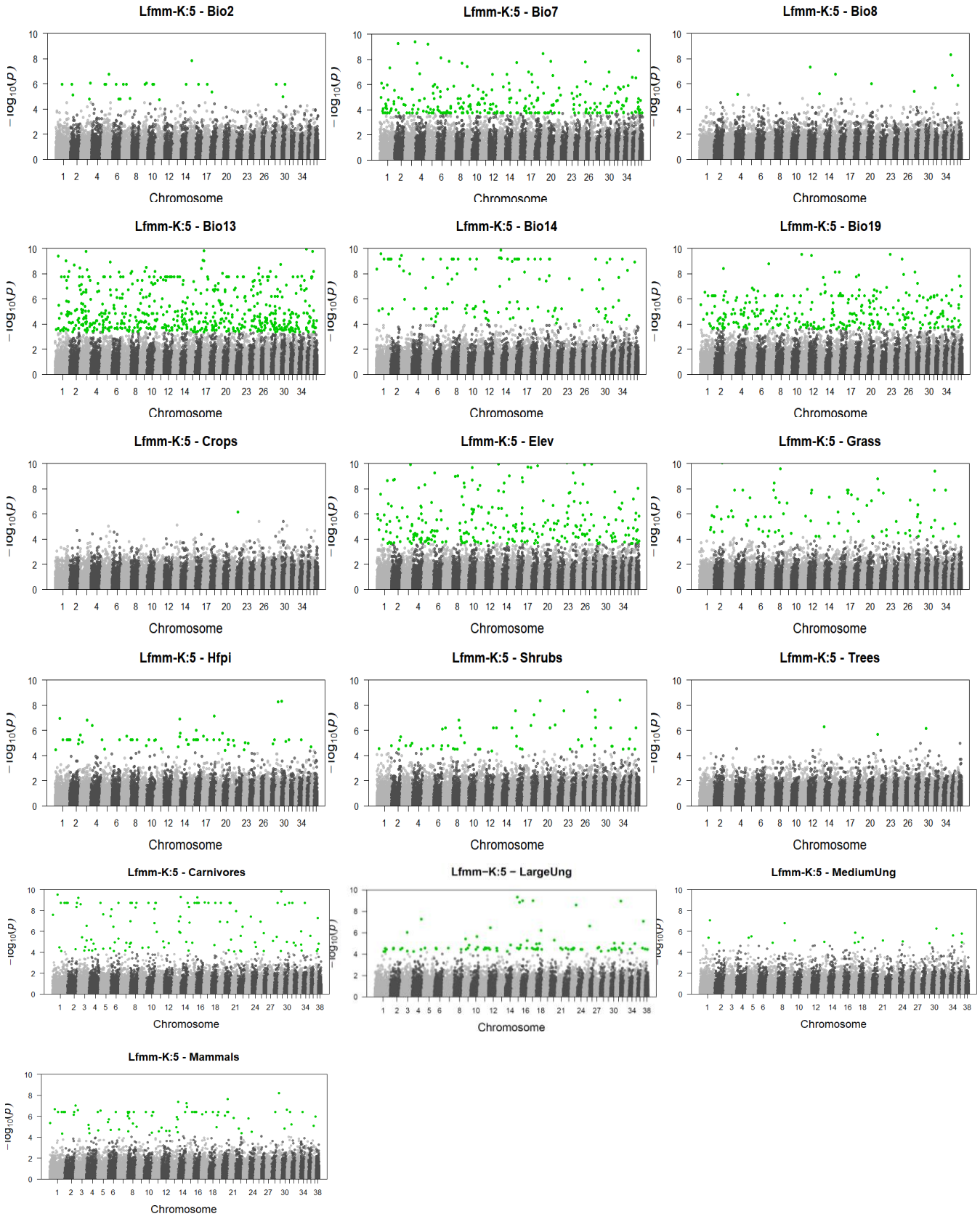

Crops
Grass
Elev
Frequency
p-values
Grass
Crops
Elev
−log10 p−values
Expected quantile
Crops
Grass
Elev
Frequency
q-values
Hfpi
Shrubs
Trees
Frequency
p-values
Hfpi
Shrubs
Trees
−log10 p−values
Expected quantile
Hfpi
Shrubs
Trees
Frequency
q-values



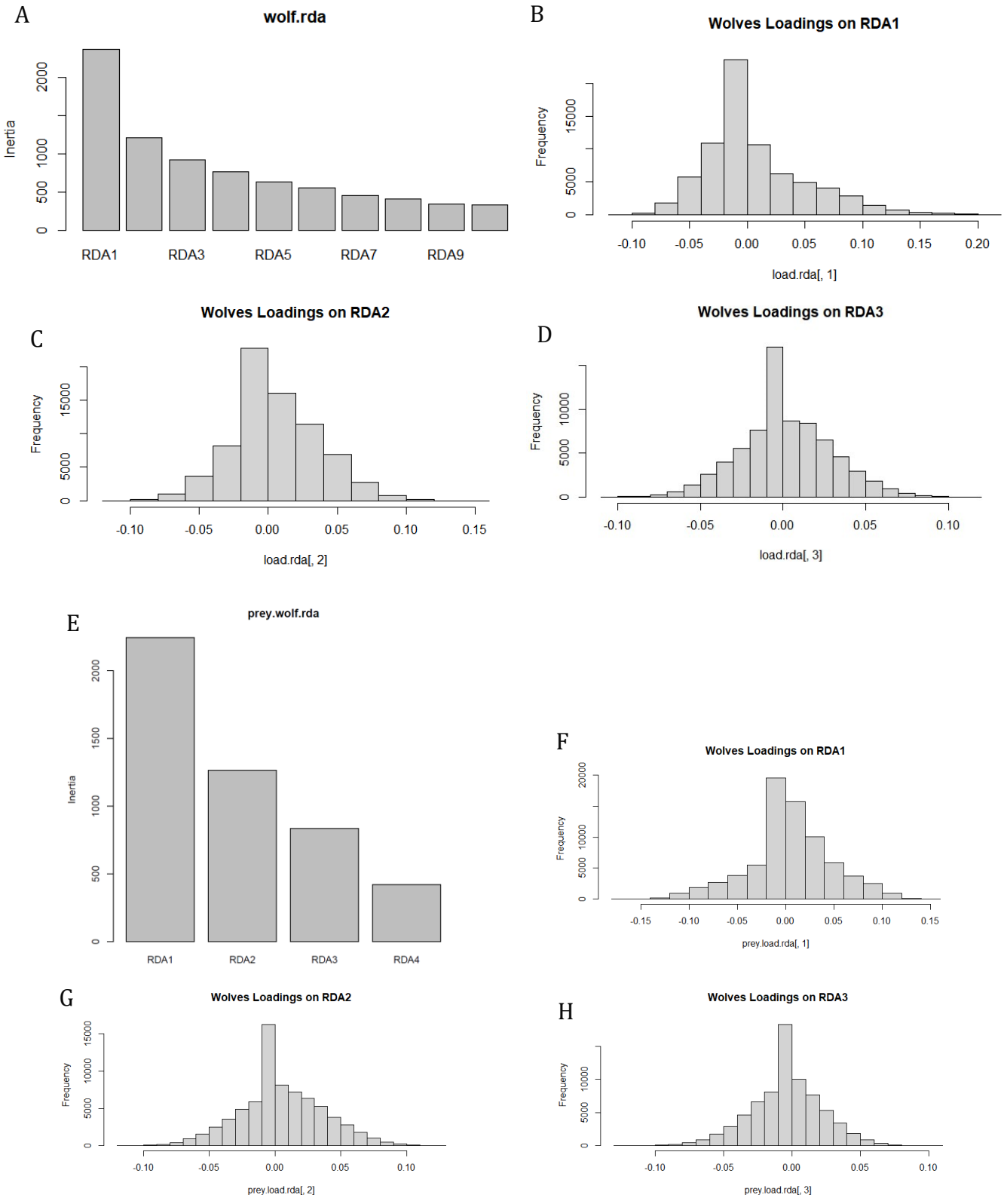
Supplementary figure 4 Latent factor mixed model (LFMM) plots FRDs. Histogram plots for p-values as calculated with lfmm (top), quantile-quantile plots (middle) and q-value histograms (bottom) for each of the environmental predictors included from WorldClim and mammalian species richness from both LFMM analysis.



Supplementary figure 5. Manhattan plots for latent factor mixed models (LFMM) SNP candidates Wolves. Outlier SNPs with significant p-values (0.05), were highlighted in green for each chromosome. A total of 2,653 outlier SNP LFMM candidates were detected with the WorldClim environmental variables and 348 LFMM candidates were detected with mammalian species richness variables.

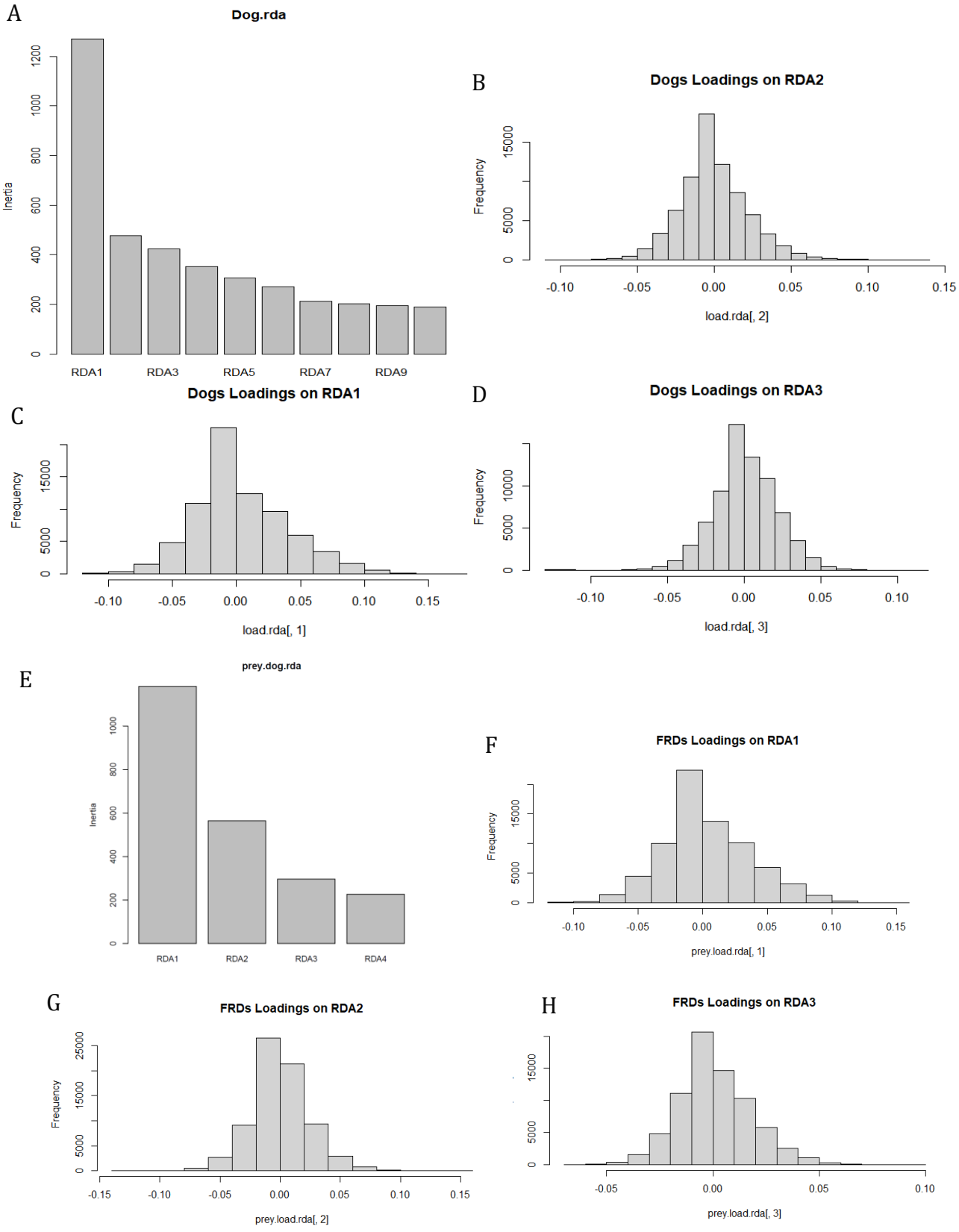


Supplementary figure 6. Manhattan plots for latent factor mixed models (LFMM) SNP candidates FRDs. Outlier SNPs with significant p-values (0.05), were highlighted in green for each chromosome. A total of 2,653 outlier SNP LFMM candidates were detected with the WorldClim environmental variables and 348 LFMM candidates were detected with mammalian species richness variables.



Supplementary figure 7. RDA histograms wolves. A. Scree plot histogram for the first 10 constrained axes of the RDA using WorldClim variables, their respective eigen values were: 2371 (0.263), 1217 (0.135), 917 (0.102), 761 (0.084), 629 (0.070), 552 (0.061), 448 (0.050), 408 (0.0453), 342 (0.038) and 330 (0.037). B. Histograms for loadings of

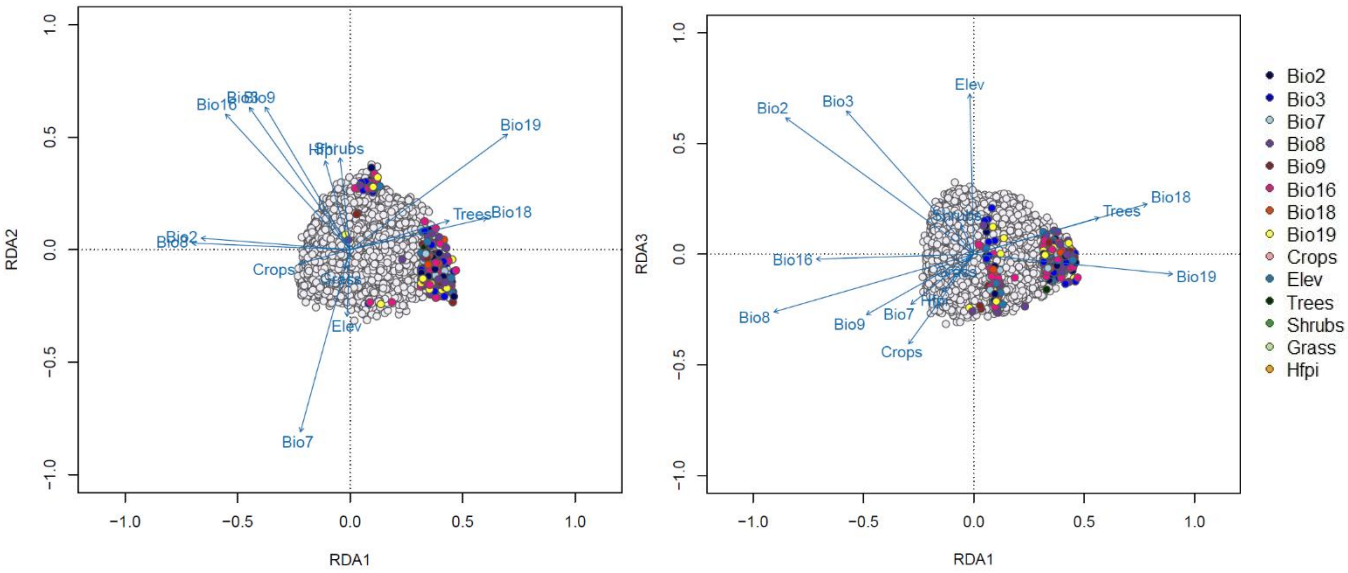
the RDA1 axis, this one accounts for most of the variance, RDA2 axis (C) and RDA3 axis (D). E. Scree plot histogram for the first 4 constrained axes of the species richness variables RDA, their respective eigen values and proportion explained by each were: 2243 (0.471), 1264 (0.266), 834 (0.175) and 419 (0.088). Histograms for loadings of the species richness analysis RDA1 axis (F), this one accounts for most of the variance, RDA2 axis (G) and RDA3 axis (H). In both analysis the first three axes accounted for most of the variance and were the ones selected for plotting.



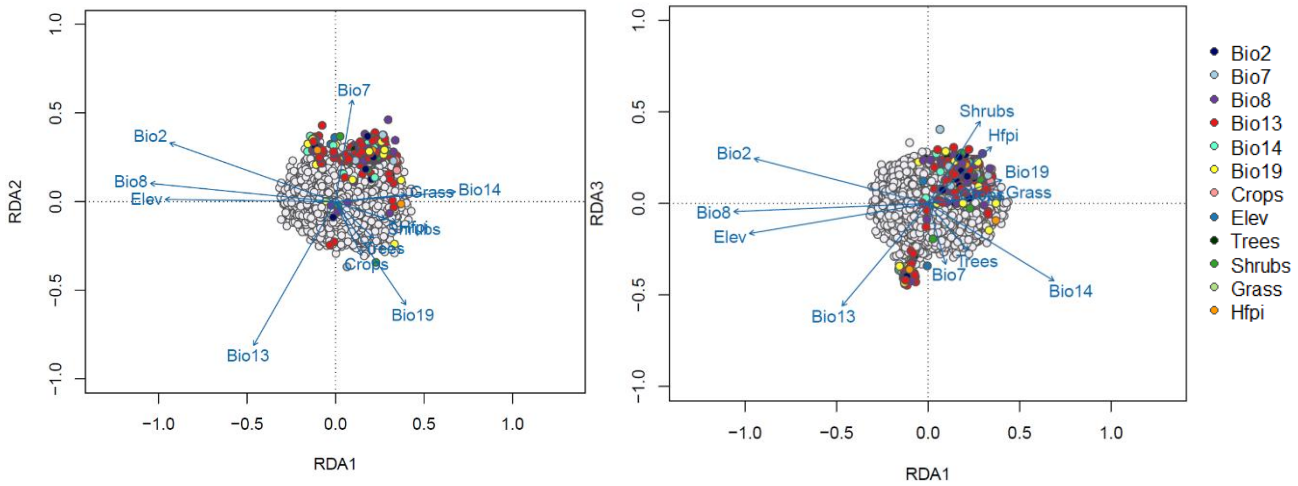
Supplementary figure 8. RDA histograms FRDs. A. Scree plot histogram for the first 10 constrained axes of the RDA using WorldClim variables, their respective eigen values and proportion explained by each were: 1269 (0.298), 476 (0.112), 423 (0.010), 352 (0.083), 306 (0.072), 272 (0.0637), 213 (0.050), 201 (0.047), 194 (0.046) and 191

(0.045). Histograms for loadings of the RDA1 axis (B), calculated explain the highest proportion of variance (0.298) RDA2 axis (C) and RDA3 axis (D). E. Scree plot histogram for the first 4 constrained axes of the species richness variables RDA, their respective eigen values and proportion explained by each were: 1182 (0.521), 564 (0.249), 295 (0.130) and 226 (0.010). Histograms for loadings of the species richness analysis RDA1 axis (F) calculated explain the highest proportion of variance (0.522), RDA2 axis (G) and RDA3 axis (H). In both analysis the first three axes accounted for most of the variance and were the ones selected for plotting.

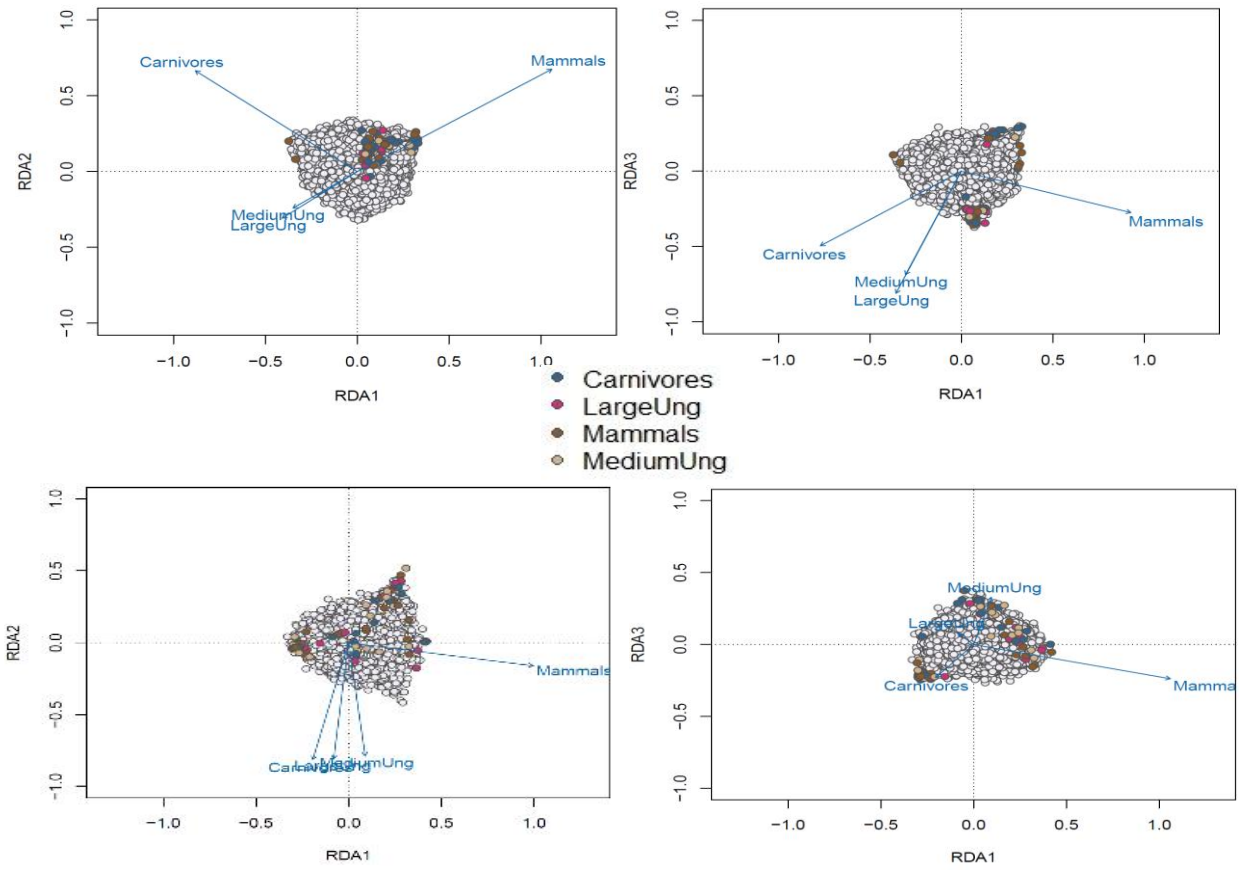
A: Wolves WorldClim variables associated SNPs' ordination space RDA first 3 Axes



B. FRDs WorldClim variables associated SNPs' ordination space RDA first 3 Axes

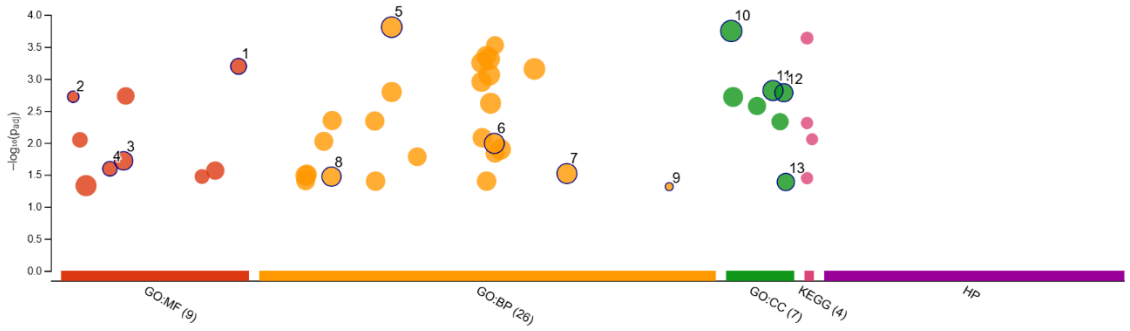


C. Wolves (left) and FRDs (right) species richness variables associated SNPs' ordination space RDA first 3 axes



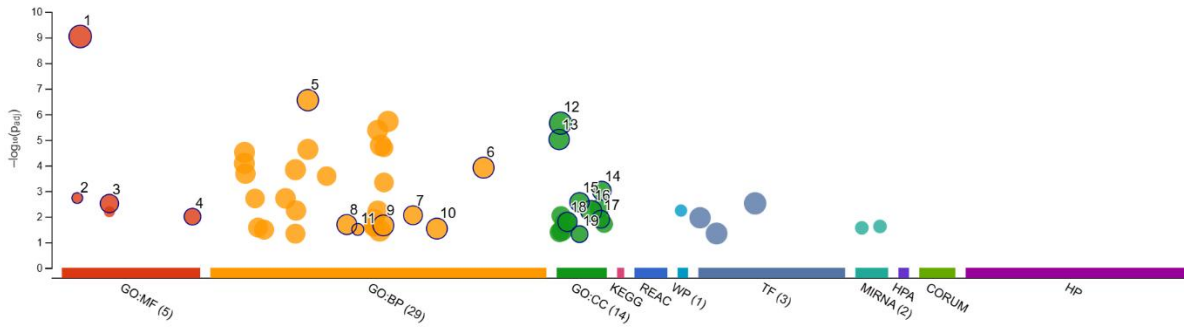
Supplementary figure 9. Ordination in space of candidate SNPs as calculated in the RDA. A. Wolves' candidate SNPs associated with the WorldClim variable ordination in space on the first three axes. Based on canonical correlation calculations, candidate SNPs on RDA axes represent multilocus sets of SNP genotypes associated most strongly with Bio8 (-0.54920746) and Bio19 (0.54614189), Bio7 (-0.71258955) and elevation (0.46860213) on the first, second and third axis respectively. B. FRDs' candidate SNPs associated with the WorldClim variable ordination in space on the first three axes. Based on canonical correlation calculations, candidate SNPs on RDA axes represent multilocus sets of SNP genotypes associated most strongly with Bio8 (-0.71257324) on the first axis and Bio13 (-0.68739140, -0.46716150) on the second and third axes. C. Wolves and FRDs' ordination in space of candidate SNPs associated with species richness variables. Based on canonical correlation calculations, wolf candidate SNPs on RDA axes represent multilocus sets of SNP genotypes associated most strongly with other mammal species richness (0.6977609, 0.4472421) on the first and second axes and large ungulates (-0.7389739) on third axis. Based on canonical correlation calculations, FRDs candidate SNPs on RDA axes represent multilocus sets of SNP genotypes associated most strongly with other mammal species richness (0.86569905), carnivores (-0.8844500) and medium ungulates (0.36168090), on the first, second and third axis respectively.

A. Genes associated with WorldClim variables wolves; g:Profiler gene ontology bubble plot with the Dog genome



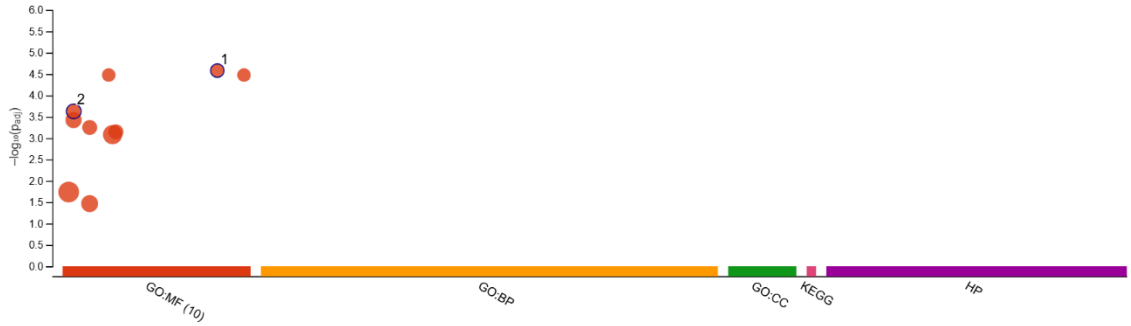
ID	Source	Term ID	Term Name	p_{adj} (query_1)
1	GO:MF	GO:0140678	molecular function inhibitor activity	6.363×10^{-4}
2	GO:MF	GO:0004364	glutathione transferase activity	1.902×10^{-3}
3	GO:MF	GO:0019899	enzyme binding	1.913×10^{-2}
4	GO:MF	GO:0016705	oxidoreductase activity, acting on paired donors, ...	2.544×10^{-2}
5	GO:BP	GO:0032501	multicellular organismal process	1.551×10^{-4}
6	GO:BP	GO:0051171	regulation of nitrogen compound metabolic proce...	1.021×10^{-2}
7	GO:BP	GO:0080090	regulation of primary metabolic process	3.021×10^{-2}
8	GO:BP	GO:0010604	positive regulation of macromolecule metabolic pr...	3.372×10^{-2}
9	GO:BP	GO:1904396	regulation of neuromuscular junction development	4.849×10^{-2}
10	GO:CC	GO:0005737	cytoplasm	1.782×10^{-4}
11	GO:CC	GO:0071944	cell periphery	1.519×10^{-3}
12	GO:CC	GO:0120025	plasma membrane bounded cell projection	1.649×10^{-3}
13	GO:CC	GO:0140535	intracellular protein-containing complex	4.095×10^{-2}

B. Genes associated with WorldClim variables wolves; g:Profiler gene ontology bubble plot with the Human genome



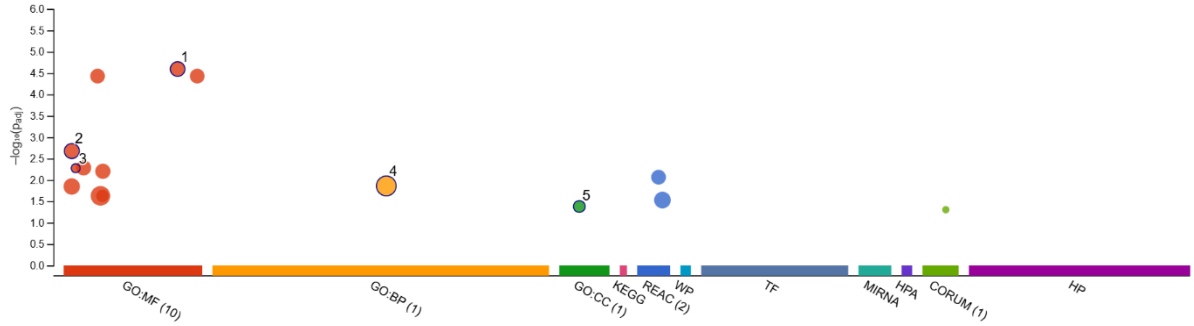
ID	Source	Term ID	Term Name	p_{adj} (query_1)
1	GO:MF	GO:0005515	protein binding	9.098×10^{-10}
2	GO:MF	GO:0005041	low-density lipoprotein particle receptor activity	1.865×10^{-3}
3	GO:MF	GO:0030234	enzyme regulator activity	3.067×10^{-3}
4	GO:MF	GO:0140678	molecular function inhibitor activity	9.966×10^{-3}
5	GO:BP	GO:0032501	multicellular organismal process	2.799×10^{-7}
6	GO:BP	GO:1901564	organonitrogen compound metabolic process	1.220×10^{-4}
7	GO:BP	GO:0065009	regulation of molecular function	8.784×10^{-3}
8	GO:BP	GO:0043412	macromolecule modification	2.033×10^{-2}
9	GO:BP	GO:0051171	regulation of nitrogen compound metabolic proce...	2.204×10^{-2}
10	GO:BP	GO:0080090	regulation of primary metabolic process	2.931×10^{-2}
11	GO:BP	GO:0045672	positive regulation of osteoclast differentiation	3.145×10^{-2}
12	GO:CC	GO:0005737	cytoplasm	2.221×10^{-6}
13	GO:CC	GO:0005576	extracellular region	9.632×10^{-6}
14	GO:CC	GO:1902494	catalytic complex	9.696×10^{-4}
15	GO:CC	GO:0042995	cell projection	2.733×10^{-3}
16	GO:CC	GO:0071944	cell periphery	6.023×10^{-3}
17	GO:CC	GO:0140535	intracellular protein-containing complex	1.261×10^{-2}
18	GO:CC	GO:0030054	cell junction	1.599×10^{-2}
19	GO:CC	GO:0043025	neuronal cell body	4.810×10^{-2}

C. Wolves' genes associated with mammalian species richness g:Profiler gene ontology bubble plot with the Dog genome



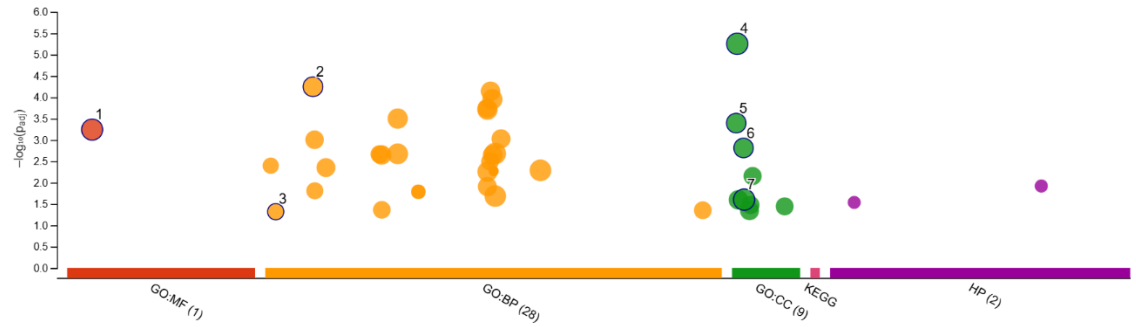
ID	Source	Term ID	Term Name	p_{adj} (query_1)
1	GO:MF	GO:0099106	ion channel regulator activity	2.621×10^{-5}
2	GO:MF	GO:0004252	serine-type endopeptidase activity	2.355×10^{-4}

D. Wolves' genes associated with mammalian species richness g:Profiler gene ontology bubble plot with the Human genome



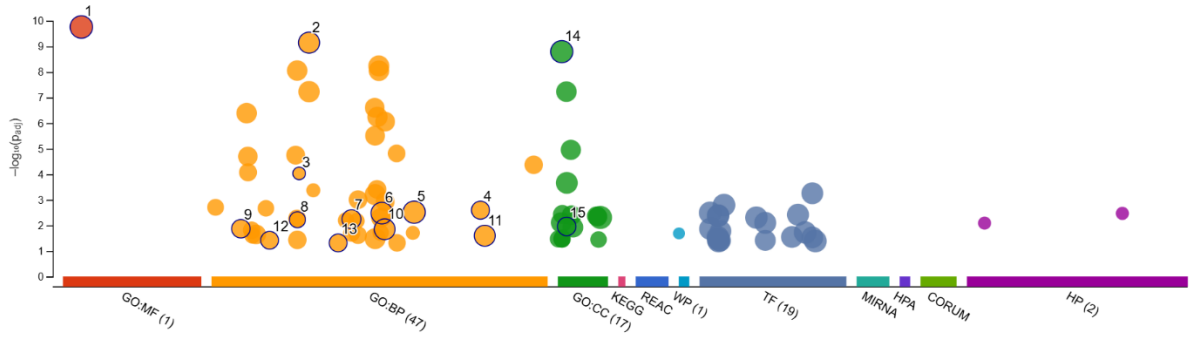
ID	Source	Term ID	Term Name	p_{adj} (query_1)
1	GO:MF	GO:0099106	ion channel regulator activity	2.539×10^{-5}
2	GO:MF	GO:0004252	serine-type endopeptidase activity	2.105×10^{-3}
3	GO:MF	GO:0004668	protein-arginine deiminase activity	5.315×10^{-3}
4	GO:BP	GO:0051239	regulation of multicellular organismal process	1.382×10^{-2}
5	GO:CC	GO:0035371	microtubule plus-end	4.179×10^{-2}

E. Genes associated with WorldClim variables FRDs; g:Profiler gene ontology bubble plot with the Dog genome



ID	Source	Term ID	Term Name	p_{adj} (query_1)
1	GO:MF	GO:0005515	protein binding	5.711×10^{-4}
2	GO:BP	GO:0007275	multicellular organism development	5.650×10^{-5}
3	GO:BP	GO:0002009	morphogenesis of an epithelium	4.796×10^{-2}
4	GO:CC	GO:0005737	cytoplasm	5.575×10^{-6}
5	GO:CC	GO:0005654	nucleoplasm	4.022×10^{-4}
6	GO:CC	GO:0012505	endomembrane system	1.539×10^{-3}
7	GO:CC	GO:0016020	membrane	2.499×10^{-2}

F. Genes associated with WorldClim variables FRDs; g:Profiler gene ontology bubble plot with the Human genome

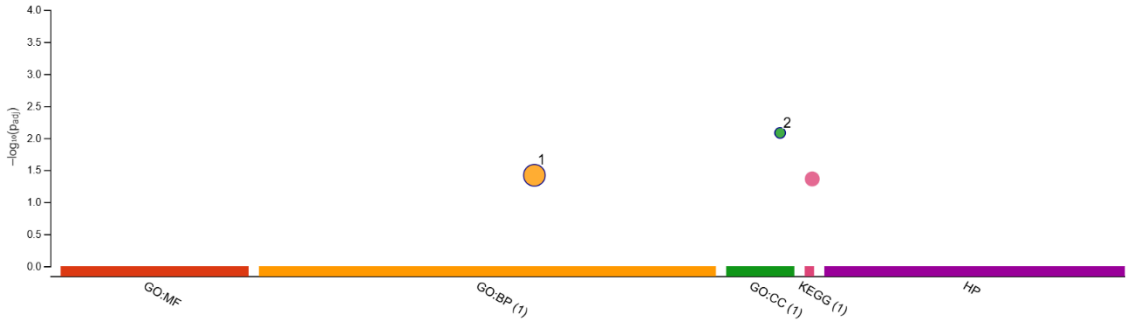


ID	Source	Term ID	Term Name	p_{adj} (query_1)
1	GO:MF	GO:0005515	protein binding	1.733×10^{-10}
2	GO:BP	GO:0032502	developmental process	7.094×10^{-10}
3	GO:BP	GO:0030574	collagen catabolic process	9.256×10^{-5}
4	GO:BP	GO:1901135	carbohydrate derivative metabolic process	2.538×10^{-3}
5	GO:BP	GO:0065007	biological regulation	3.037×10^{-3}
6	GO:BP	GO:0050896	response to stimulus	3.314×10^{-3}
7	GO:BP	GO:0044281	small molecule metabolic process	5.724×10^{-3}
8	GO:BP	GO:0030198	extracellular matrix organization	6.001×10^{-3}
9	GO:BP	GO:0006629	lipid metabolic process	1.349×10^{-2}
10	GO:BP	GO:0051179	localization	1.418×10^{-2}
11	GO:BP	GO:1901564	organonitrogen compound metabolic process	2.534×10^{-2}
12	GO:BP	GO:0014070	response to organic cyclic compound	3.821×10^{-2}
13	GO:BP	GO:0040007	growth	4.852×10^{-2}
14	GO:CC	GO:0005737	cytoplasm	1.596×10^{-9}
15	GO:CC	GO:0015630	microtubule cytoskeleton	1.104×10^{-2}

version	e111_eg58_p18_f463989d
date	9/10/2024, 4:03:27 PM
organism	hsapiens

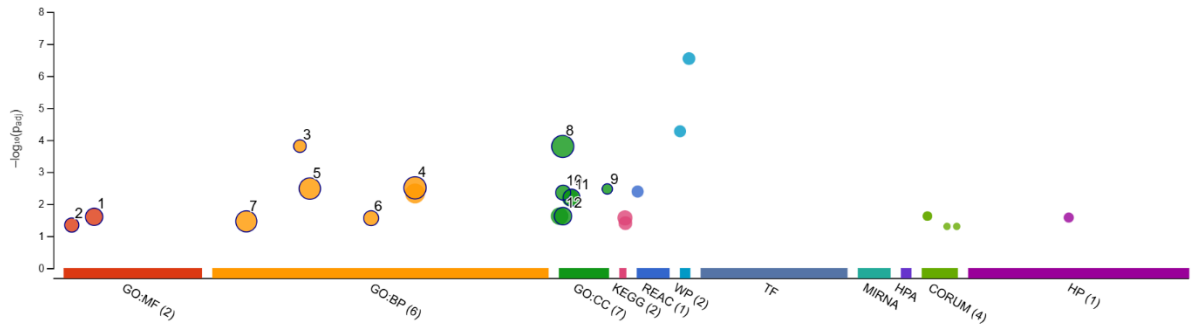
g:Profiler

G. FRDs' genes associated with mammalian species richness g:Profiler gene ontology bubble plot with the Dog genome



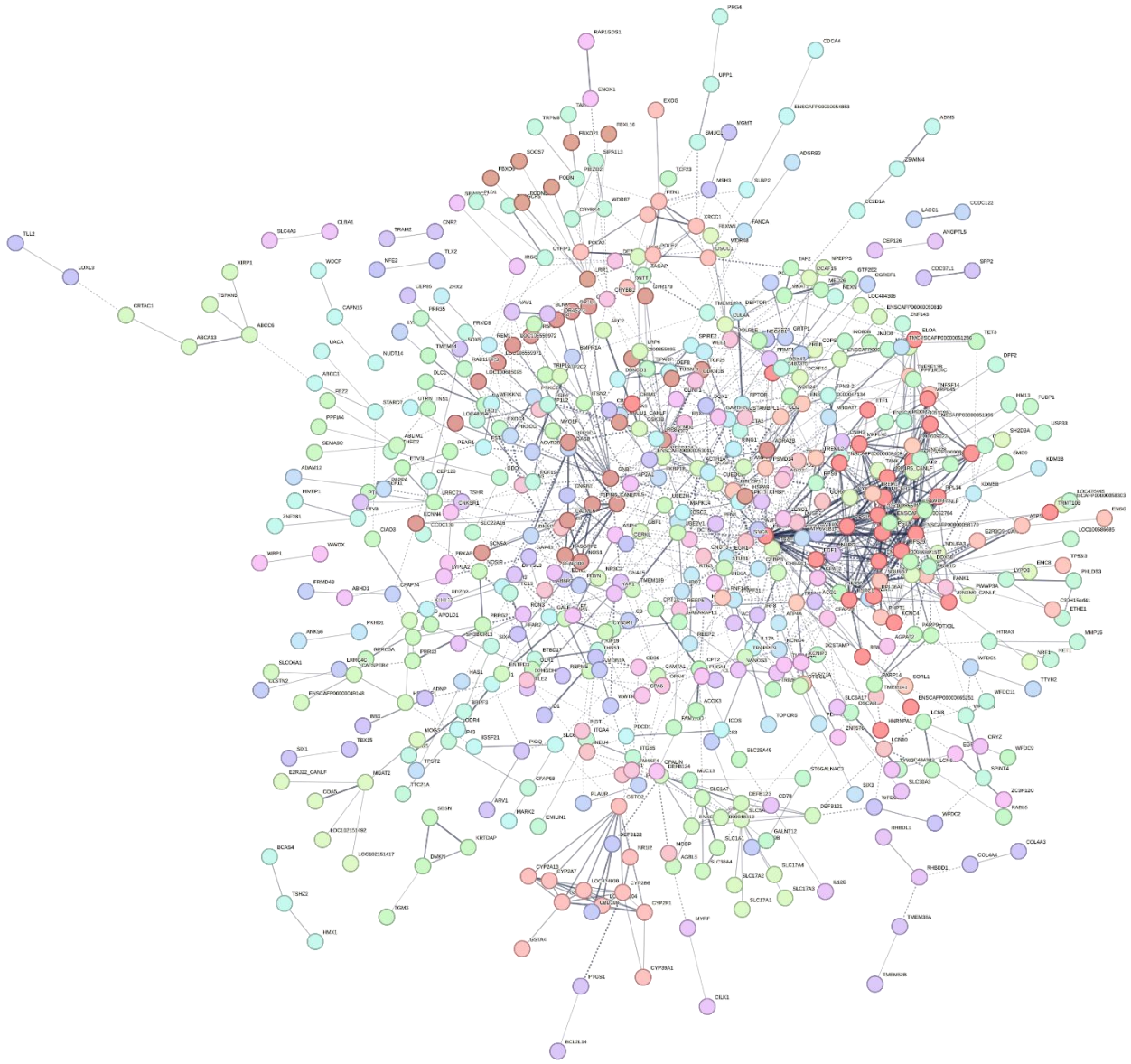
ID	Source	Term ID	Term Name	p_{adj} (query_1)
1	GO:BP	GO:0065007	biological regulation	3.807×10^{-2}
2	GO:CC	GO:0098576	lumenal side of membrane	8.291×10^{-3}

H. FRDs' genes associated with mammalian species richness g:Profiler gene ontology bubble plot with the Human genome

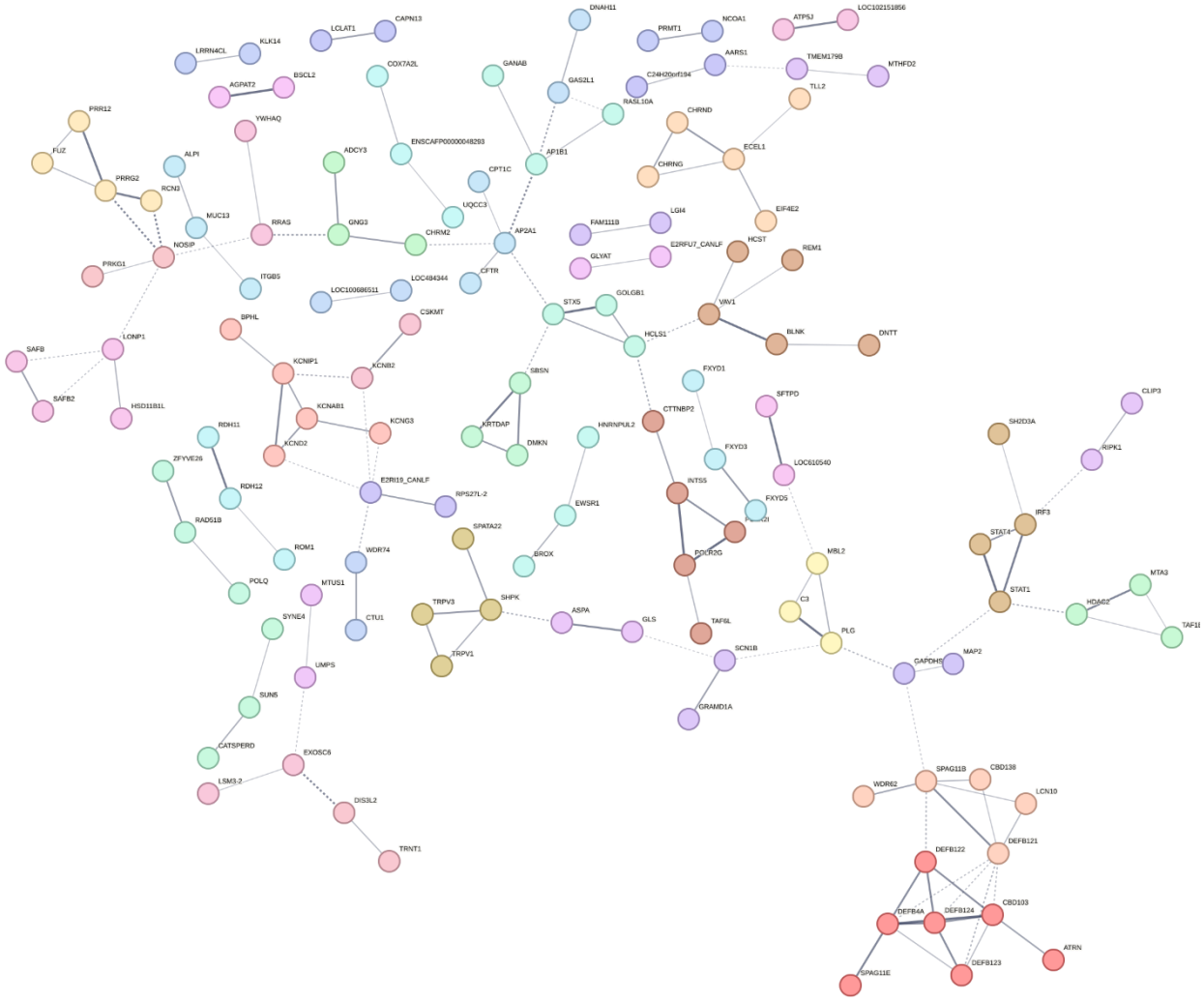


ID	Source	Term ID	Term Name	p_{adj} (query_1)
1	GO:MF	GO:0015318	inorganic molecular entity transmembrane transp...	2.496×10^{-2}
2	GO:MF	GO:0004222	metalloendopeptidase activity	4.518×10^{-2}
3	GO:BP	GO:0030574	collagen catabolic process	1.546×10^{-4}
4	GO:BP	GO:0065007	biological regulation	3.109×10^{-3}
5	GO:BP	GO:0032501	multicellular organismal process	3.281×10^{-3}
6	GO:BP	GO:0046890	regulation of lipid biosynthetic process	2.752×10^{-2}
7	GO:BP	GO:0007154	cell communication	3.458×10^{-2}
8	GO:CC	GO:0005737	cytoplasm	1.579×10^{-4}
9	GO:CC	GO:1990454	L-type voltage-gated calcium channel complex	3.379×10^{-3}
10	GO:CC	GO:0005775	vacuolar lumen	4.386×10^{-3}
11	GO:CC	GO:0031012	extracellular matrix	6.285×10^{-3}
12	GO:CC	GO:0005764	lysosome	2.383×10^{-2}

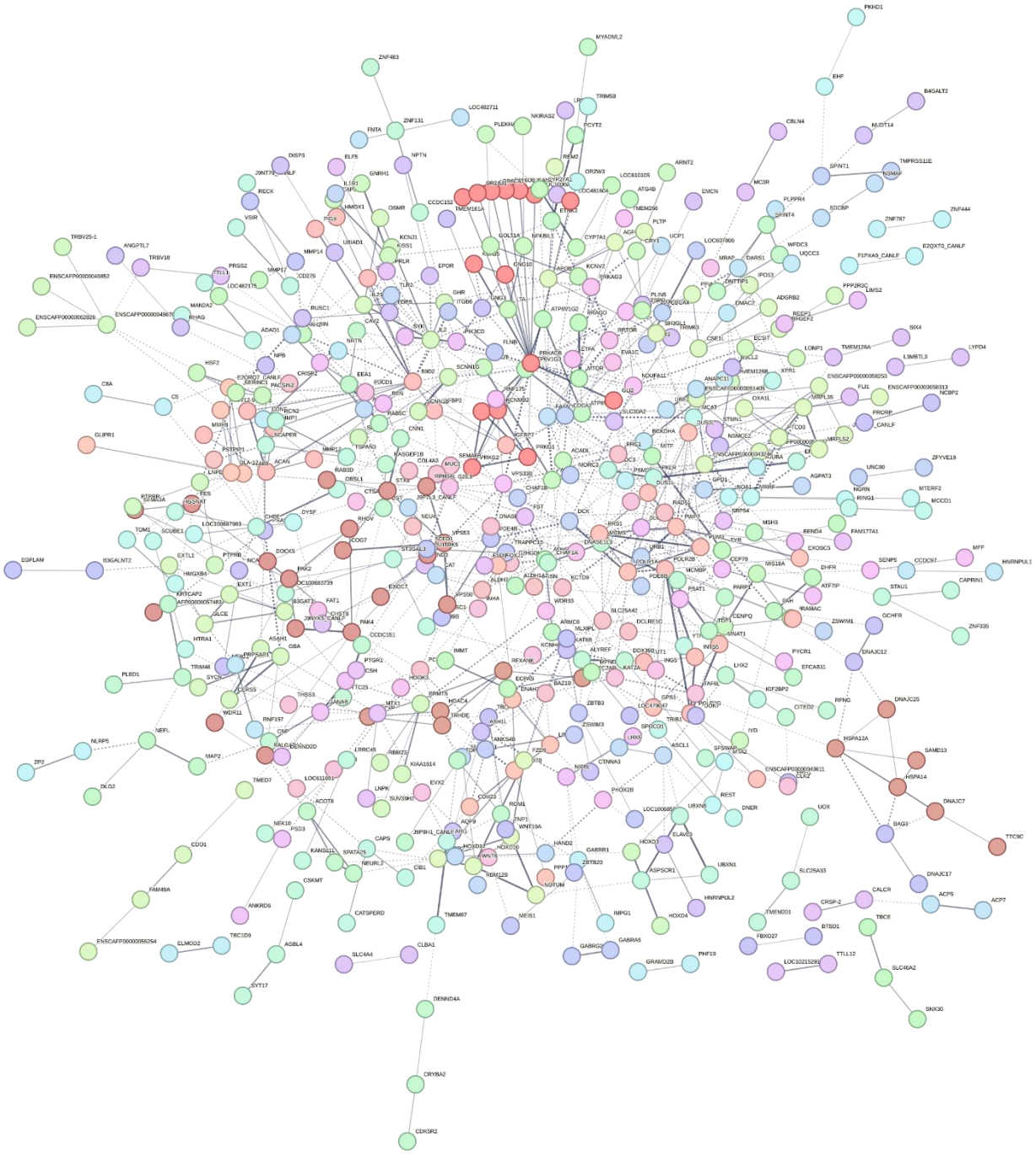
Supplementary figure 10. g:Profiler bubble plots of gene ontology (GO) enrichment analysis. Significantly enriched GOs are highlighted, numbers in the bubbles correspond to the ID columns for the significantly enriched GO categories in the table below.



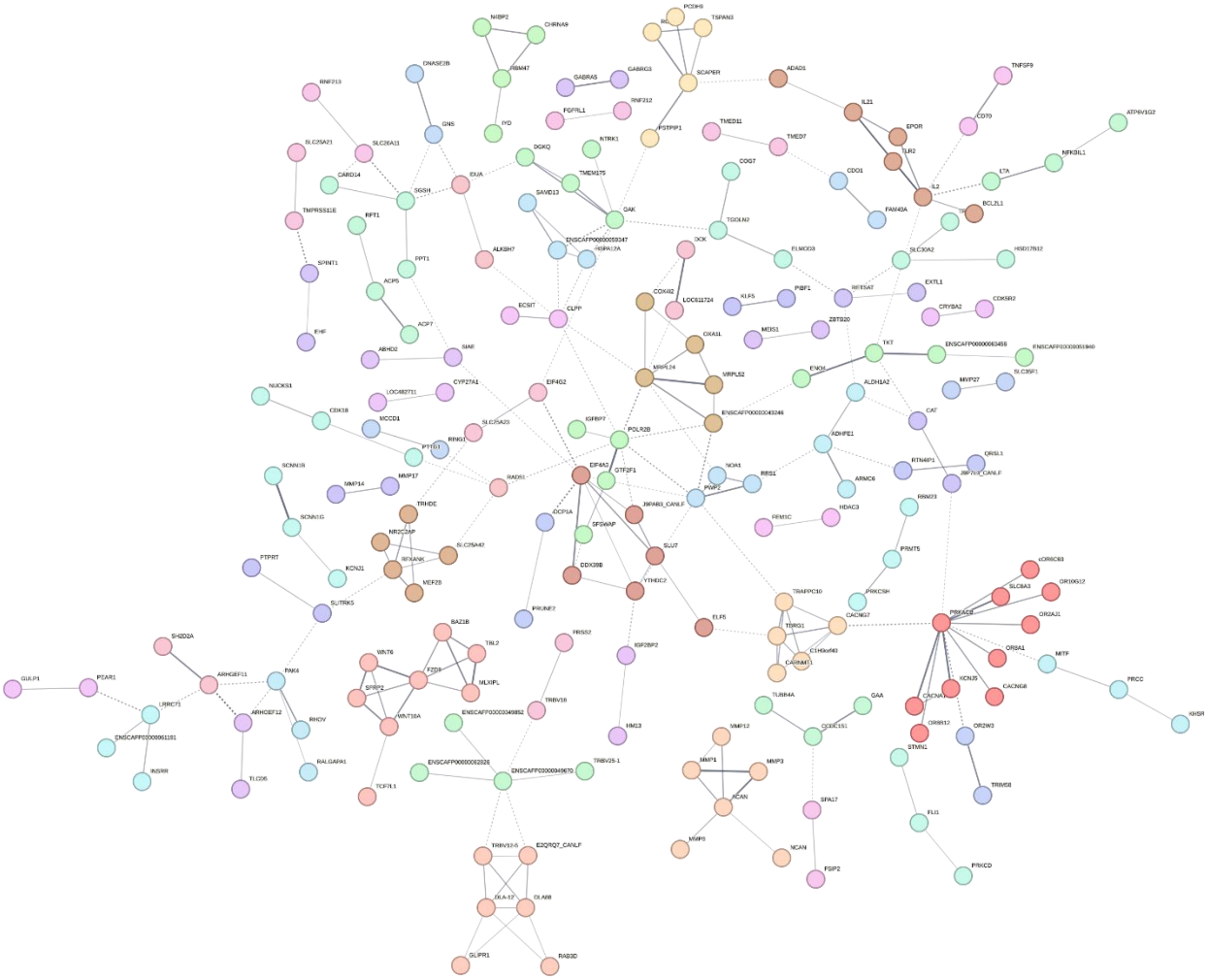
Supplementary figure 11. Wolves' Complete STRING Network. Complete STRING network for the genes displaying signatures of local adaptation. Non-connected proteins were hidden for clarity. The network includes 693 nodes (i.e., gene products), the average local clustering coefficient was 0.35 and a PPI enrichment p-value of 6.85e-05. The line thickness indicates the strength of data supporting the predicted interactions. The network had functional enrichment for the molecular functions; arachidonic acid epoxygenase activity and oxireductase activity acting on paired donors, with incorporation or reduction of molecular oxygen and for the KEGG pathways metabolism of xenobiotics by cytochrome P450, chemical carcinogenesis, drug metabolism- other enzymes, Pertussis, RIG-I-like receptor signalling pathway, C-type lectin receptor signalling pathway and fluid shear stress and atherosclerosis.



Supplementary figure 12. Wolves' complete STRING network, species richness analysis. Complete STRING network for the genes found to be associated with species richness in wolves. Non-connected proteins were hidden for clarity. The network includes 243 nodes (i.e., gene products), the average local clustering coefficient was 0.37 and a PPI enrichment p-value of 6.47e-08. The line thickness indicates the strength of data supporting the predicted interactions. The network had functional enrichment for the biological process defence response to bacterium and for the molecular function GOs CCR6 chemokine receptor binding and chemoattractant activity.



Supplementary figure 13. FRDs' complete STRING network. Complete STRING network for the genes displaying signatures of local adaptation in the FRDs. Non-connected proteins were hidden for clarity. The network includes 684 nodes (i.e., gene products), the average local clustering coefficient was 0.34 and a PPI enrichment p-value of 2.29e-05. The line thickness indicates the strength of data supporting the predicted interactions.



Supplementary figure 14. FRDs' complete STRING network, species richness analysis. Complete STRING network for the genes found to be associated with species richness in the FRDs. Non-connected proteins were hidden for clarity. The network includes 313 nodes (i.e., gene products), the average local clustering coefficient was 0.38 and a PPI enrichment p-value of 1.08e-05. The line thickness indicates the strength of data supporting the predicted interactions. The network had functional enrichment for the biological process collagen catabolic processes.

Ensembl ID	UnitProt ID or name	Amino acid change	
WorldClim variables			
ENSCAFG00000014554	ACOX3	86 (M>I)	
ENSCAFG00000011576	ADNP	603 (I>L)	
ENSCAFG00000016566	CC2D1A	96 (D>N)	
ENSCAFG00000008431	CDIN1	112/12/114 (E>D)	
ENSCAFG00000019209	CFAP74	724 (V>M)	
ENSCAFG00000013224	DDX47	464 (F>V)	
ENSCAFG00000008507*	DNTT	145 (A>T)	
ENSCAFG00000052337	GPR132	258 (A>T)	
ENSCAFG00000017287	IL12B	18 (S>C)	
ENSCAFG00000003594*	IRF3	63 (T>I)	
ENSCAFG00000031996	IRGQ	403 (P>L)	
ENSCAFG00000014175	ITGA4	112/78 (N>S)	
ENSCAFG00000004723	KCNN4	79 (R>H)	
ENSCAFG00000008788	MSH3	944 (M>T)	
ENSCAFG00000047541	OTOGL	711 (A>T)	
ENSCAFG00000008279	PTPRJ	181 (G>D)	
ENSCAFG00000015627	SIX4	387/379 (V>M)	
ENSCAFG00000013789	SSD domain-containing protein	408 (T>I)	
ENSCAFG00000008348	TM4SF4	18 (A>T)	
Species richness			
ENSCAFG00000005318	CAPN13	518 (A>S)	
ENSCAFG00000030013	RHEX	147/104 (H>Y)	
ENSCAFG00000024886	SPAG11E	16 (G>R)	

* Appeared in both VEP results
/ Proteins with multiple isoforms hence the change occurred at a different position

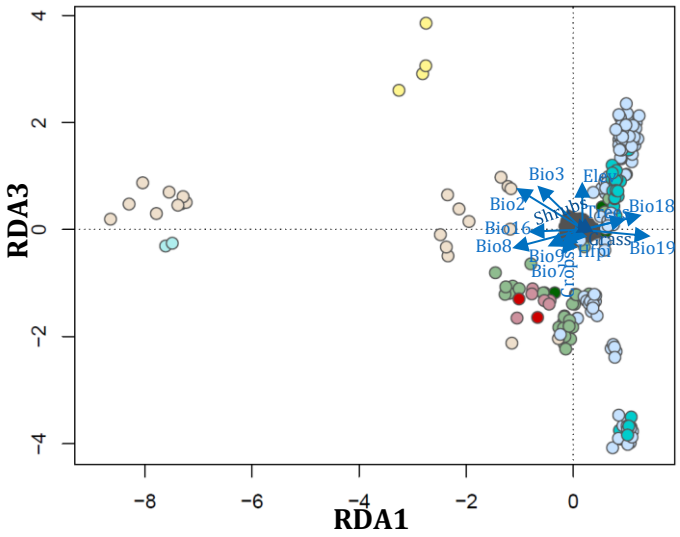
Supplementary figure 15. Summary of the VEP results for the genes predicted to have missense mutations in wolves. The last column of heatmap is based on the predicted tolerance. Intensity of the colour is related to the prediction of tolerance with the light pink being predicted to be tolerated with high confidence, followed by dark pink predicted to be tolerated low confidence, and the red predicted to not be tolerated. Grey signifies no prediction is available for tolerance of the variant identified.

Ensembl ID	UnitProt ID or name	Position (Amino acid change)	
WorldClim Variables			
ENSCAFG00000025018	ADAM7	406/177 (K>Q)	
ENSCAFG00000018685	CPLANE1	2,376 (I>V)	
ENSCAFG00000013152	D2HGDH	301 (P>S)	
ENSCAFG00000017844	DNAH3	850/28/845/45/113/74 (Y>H)	
ENSCAFG00000007603	FLNB	2,189/2,158/2,010 (V>M)	
ENSCAFG00000051032*	Glyceraldehyde 3-phosphate dehydrogenase NAD(P) binding domain-containing protein	69 (A>D)	
ENSCAFG00000052337	GPR132	81 (L>F)	
ENSCAFG00000009859	GPRIN3	435 (R>K)	
ENSCAFG00000004702	HSPA14 MSANTD7	325 (V>I)	
ENSCAFG00000002147	IL1R1	254/242/197/226 (G>R)	
ENSCAFG00000059254*	INSM2	28 (R>L)	
ENSCAFG00000008969	KCTD9	82 (G>E)	
ENSCAFG00000007770	LNPEP	803/792/791 (V>I)	
ENSCAFG00000017518	LOC119866928	56 (F>S)	
ENSCAFG00000058088	LOC481603	38 (V>M)	
ENSCAFG00000032643	MRRF	145/158 (A>V)/(A>T)	
ENSCAFG00000001799	MTERF2	90 (D>N)	
ENSCAFG00000015929*	N4BP2	825/746 (R>H)	
ENSCAFG00000002337*	Peroxiredoxin-1 (EC 1.11.1.24) (Thioredoxin-dependent peroxiredoxin 1)	12 (A>V)	
ENSCAFG00000016959	PKLR	25 (K>N)	
ENSCAFG00000013360	PPP2R3C	144 (A>V)	
ENSCAFG00000050337	RCCD1	220 (V>M)	
ENSCAFG00000005055	RNF157	31/69 (A>E)	
ENSCAFG00000015627	SIX4	387/379 (V>M)	
ENSCAFG00000004826	SLC6A9	111/131/203 (G>S)	
ENSCAFG00000010461	TM4SF20	68 (M>V)	
ENSCAFG00000009125	TMEM67	179/98 (T>S)	
ENSCAFG00000049094	Uncharacterized protein	145 (A>V)	
ENSCAFG00000008892	URB1	1,468/1,281 (E>K)	
Species richness			
ENSCAFG00000018636	DENND1C	56 (V>I)	
ENSCAFG00000011978	POU2F3	355 (R>H)	
ENSCAFG00000023568	SH2D2A	140 (C>S)	
ENSCAFG00000008647	TKT	256/332/223/209 (G>A)	

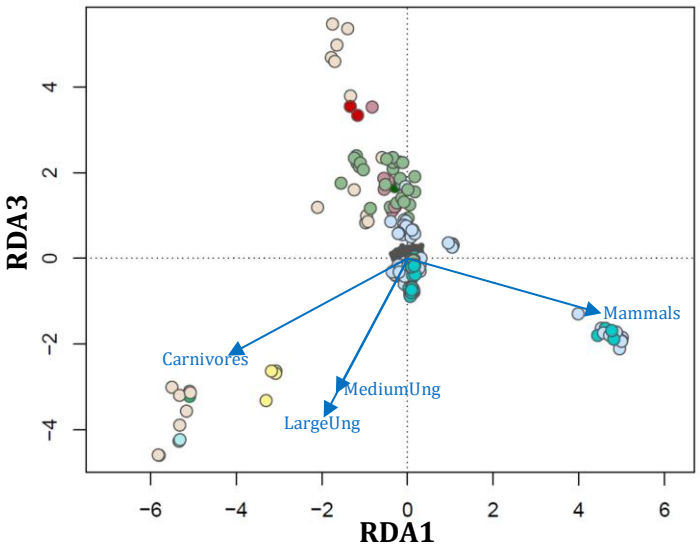
* Appeared in both VEP results
/ Proteins with multiple isoforms hence the change occurred at a different position

Supplementary figure 16. Summary of the VEP results for the genes predicted to have missense mutations in FRDs. Last column heatmap is based on the predicted tolerance. Intensity of the colour is related to the prediction of tolerance with the pink being predicted to be tolerated with high confidence, followed by medium pink predicted to be tolerated low confidence, dark pink predicted to be deleterious with low confidence and the red predicted to not be tolerated, grey signifies no prediction is available for tolerance of the variant identified.

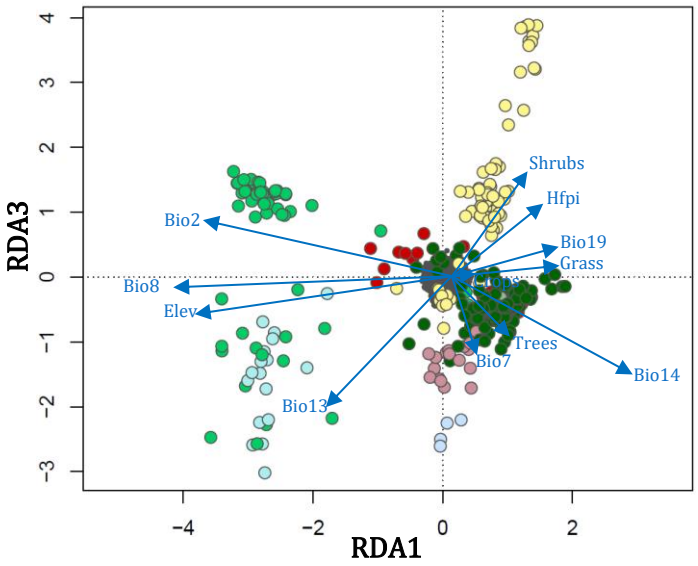
A. Wolves WorldClim variables RDA axes 1 and 3



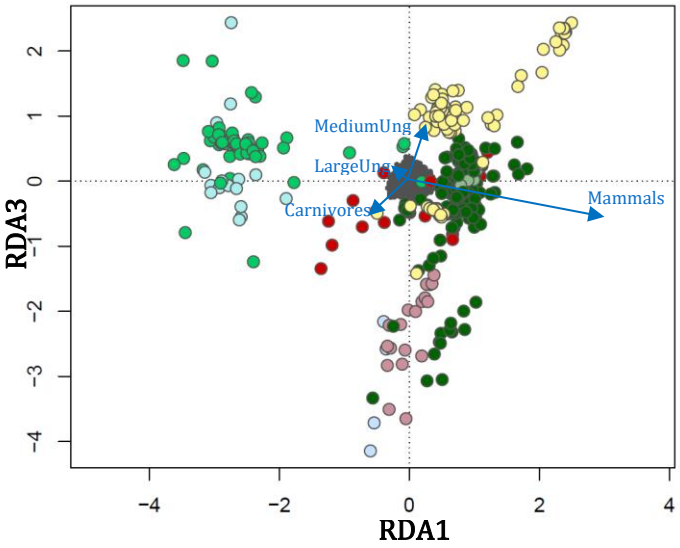
B. Wolves species richness RDA axes 1 and 3



C. FRDs WorldClim variables RDA axes 1 and 3



D. FRDs species richness RDA axes 2 and 3



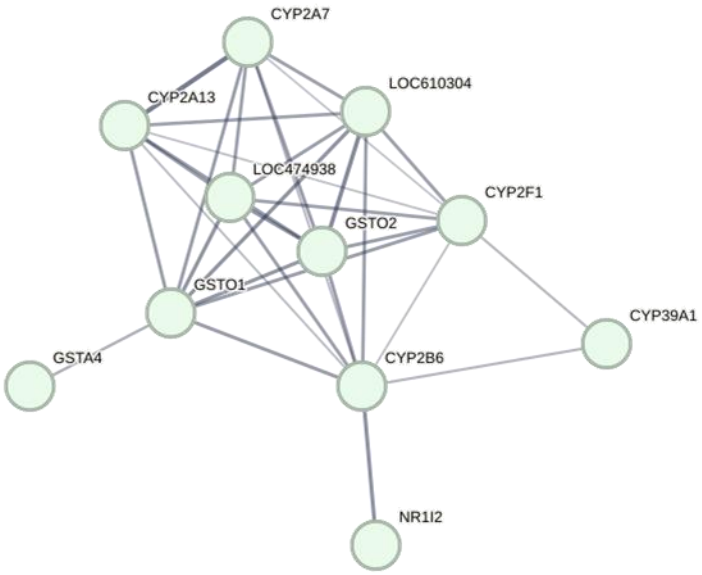
- Deserts & Xeric Shrublands
- Montane Grasslands & Shrublands
- Tropical & Subtropical Moist Broadleaf Forests
- Flooded Grasslands & Savannas
- Temperate Grasslands, Savannas & Shrublands
- Boreal Forests/Taiga
- Temperate Broadleaf & Mixed Forests
- Temperate Conifer Forests
- Tundra
- Mediterranean Forests, Woodlands & Scrub

Supplementary figure 17. RDA ordination in space for candidate SNPs displaying signatures of local adaptive evolution and environmental vectors. Output for RDA analysis, ordination in space of individual genotypes from (A-B) wolves and (C-D) FRDs for the first and third axes for both the WorldClim (A, C) and species richness variables (B, D). Individuals are coloured by biome, SNPs are in grey, and the blue vectors are the environmental predictors vectors. A. Individual genotypes of wolves from flooded grasslands and savanna were corelated with high cover of trees, precipitation of warmest quarter (Bio18) and low shrub cover. B. Individual genotypes of wolves from flooded grasslands and savanna were corelated with high species richness of non-ungulate mammals. C. Individual genotypes of FRDs from boreal forests/taiga and tropical and subtropical moist broadleaf forests and tundra exhibited positive correlations with high mean diurnal range (Bio2), mean temperature of the wettest quarter (Bio8), elevation, and high precipitation of the wettest month. Individual genotypes of FRDs from temperate grasslands, savannas, and shrublands were positively correlated with high cover of shrubs and human footprint, trees, high precipitation of the coldest quarter (Bio19). Individual genotypes of FRDs from desert and xeric shrublands and temperate broadleaf forests and mixed forests were associated with high annual temperature range (Bio7) and tree cover. The latter were also associated with high cover of grass, crops and Bio14, precipitation of driest month. D. Individual genotypes of FRDs from most biomes, except for tundra, tropical and subtropical moist leaf forests and temperate grasslands, savannas and shrubland, were associated with high species richness of non-ungulate mammals.

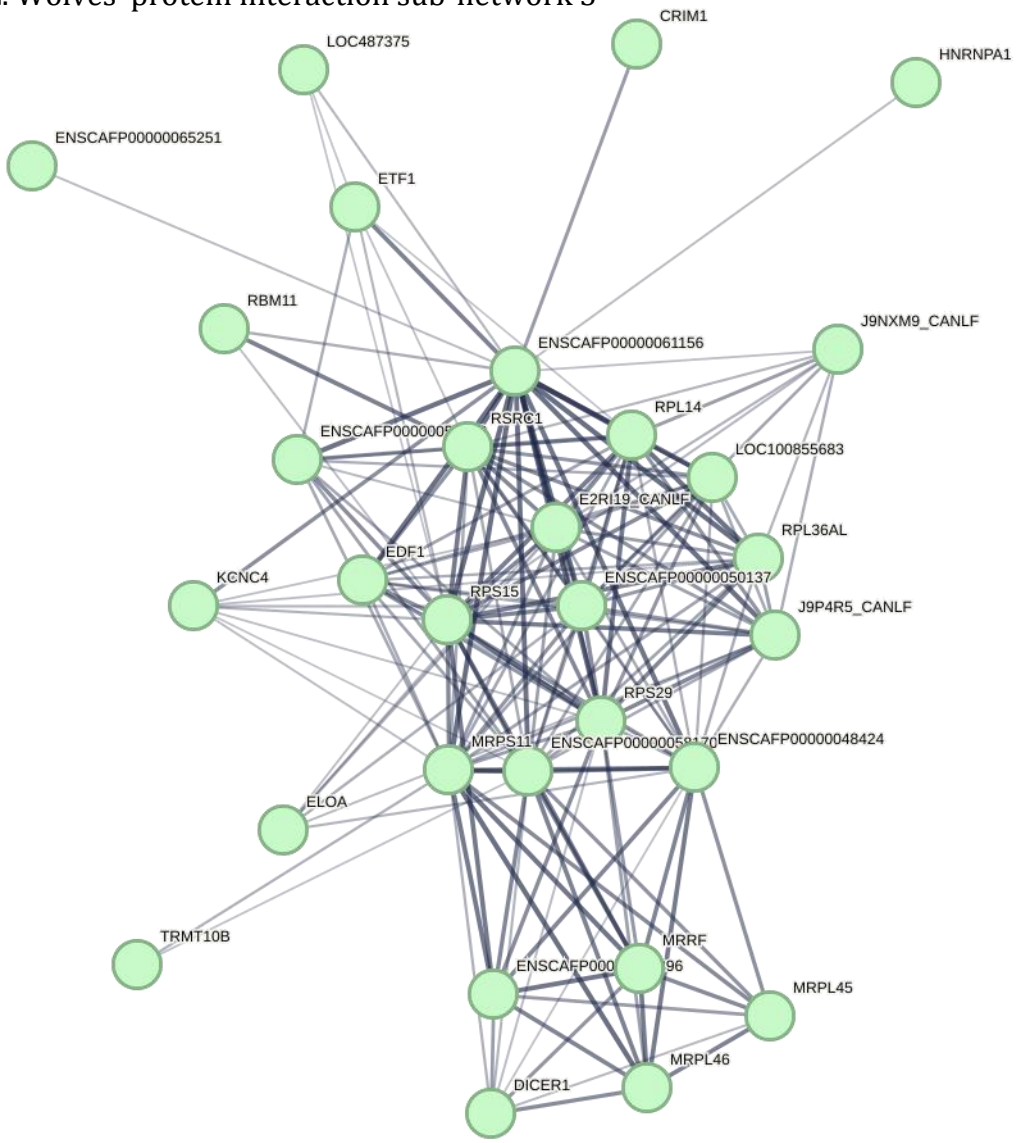
A. Wolves' protein interaction sub-network 1



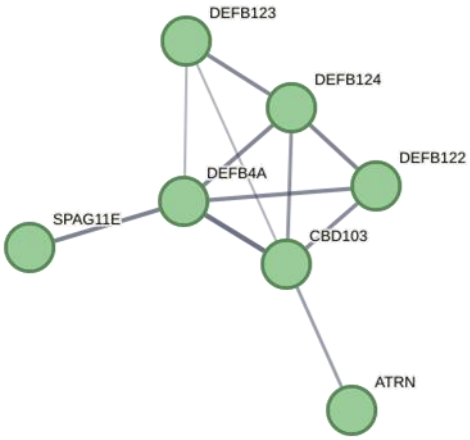
B. Wolves' protein interaction sub-network 2



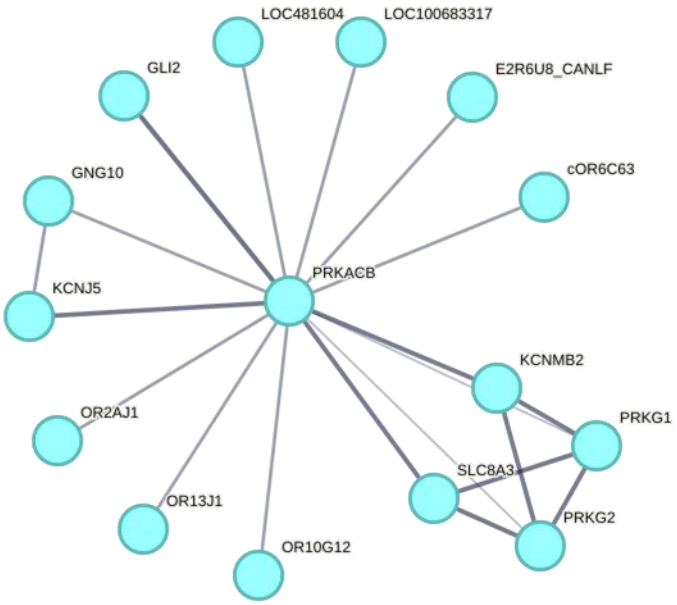
C. Wolves' protein interaction sub-network 3



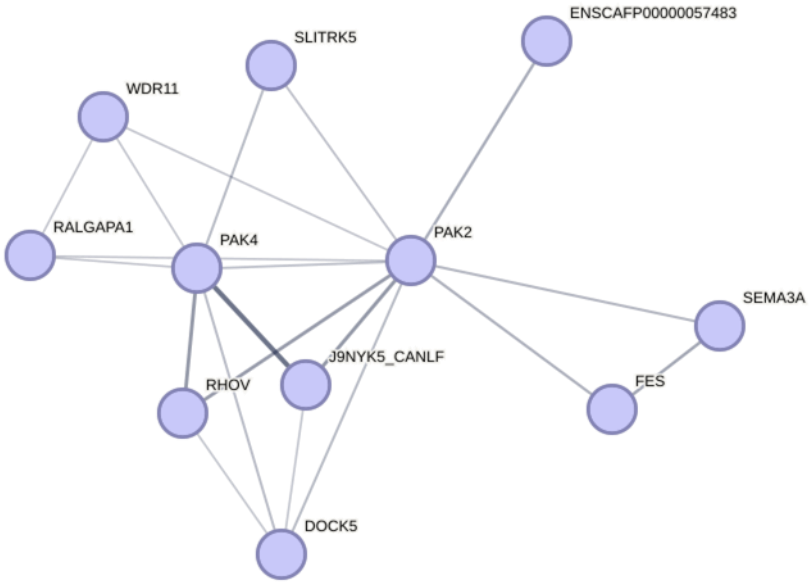
D. Wolves' species richness protein interaction sub-network



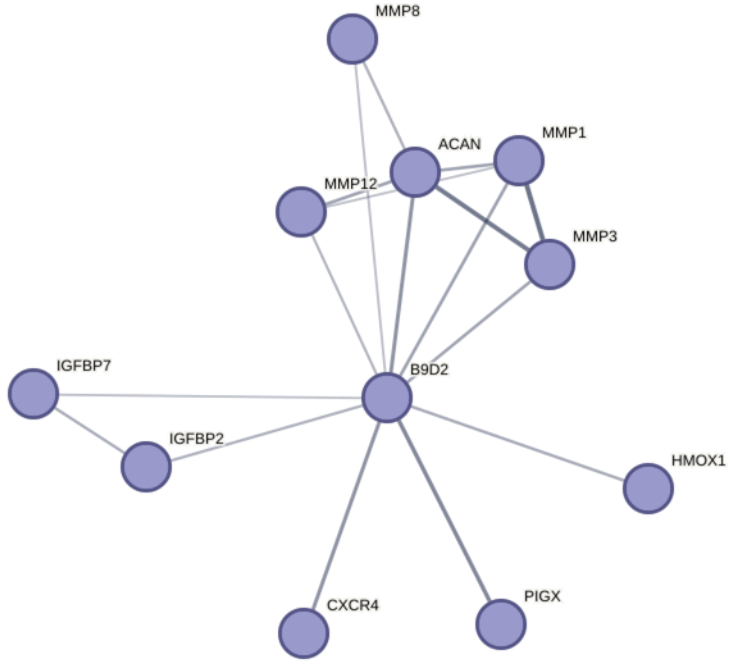
E. FRDs' protein interaction sub-network 1



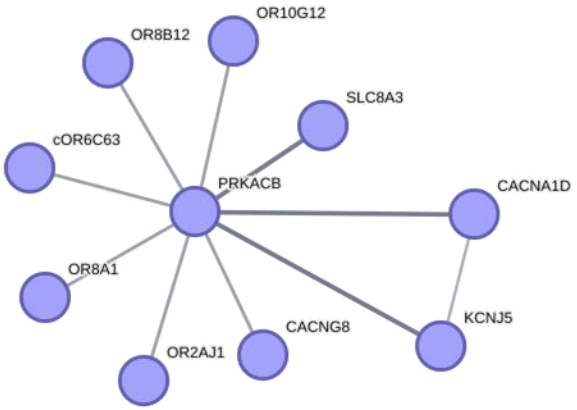
F. FRDs' protein interaction sub-network 2



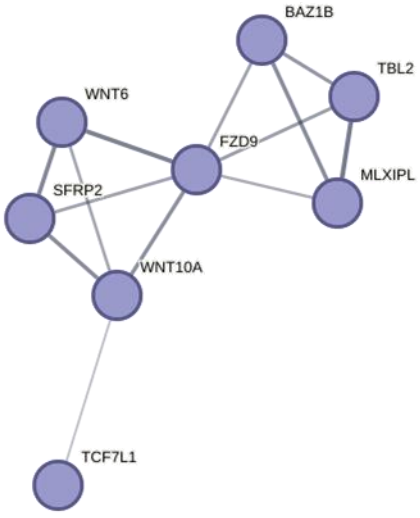
G. FRDs' protein interaction sub-network 3



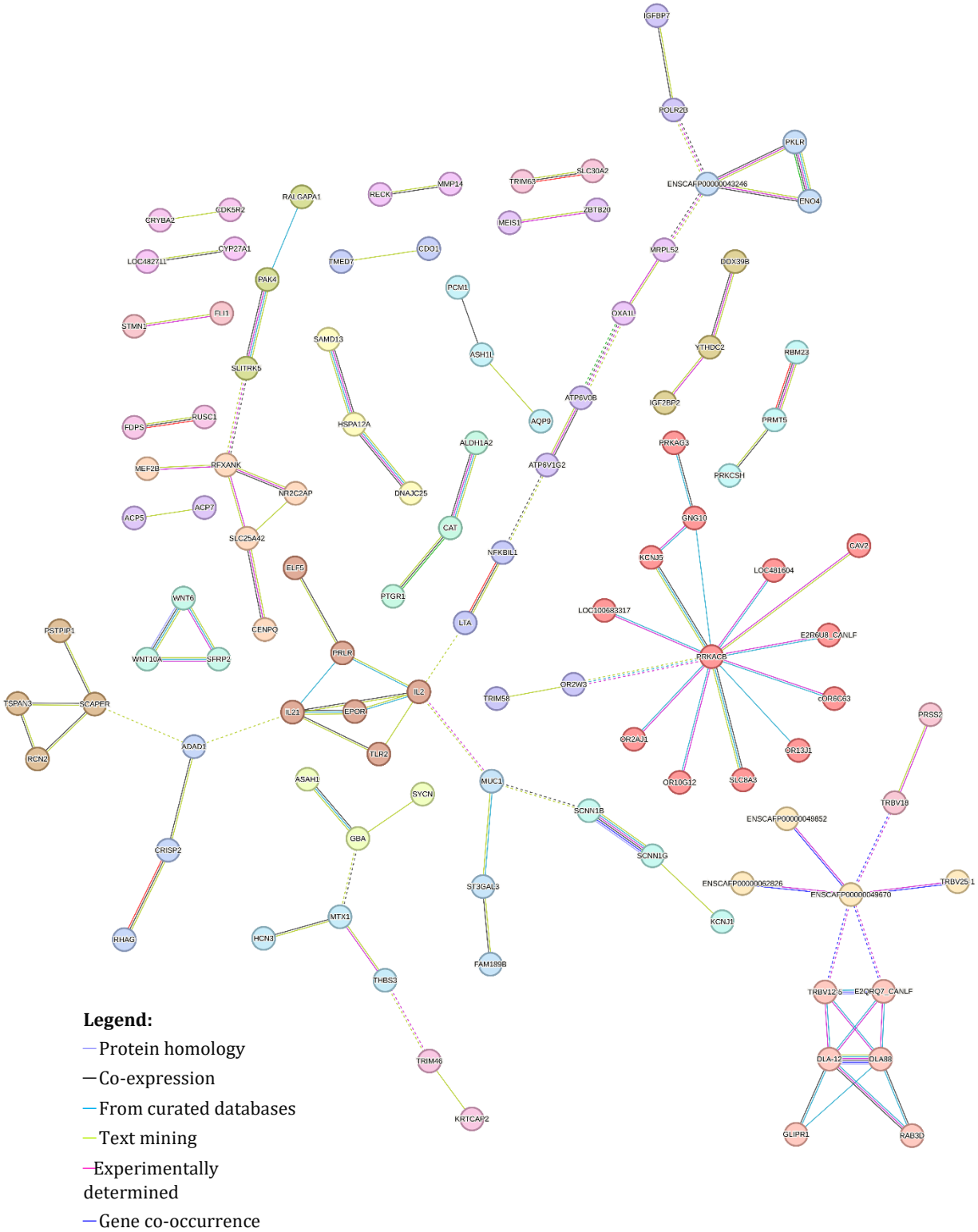
H. FRDs' species richness protein interaction sub-network 1



I. FRDs' species richness protein interaction sub-network 2



Supplementary Figure 18. The STRING functional protein association sub-networks. Wolves' subnetworks of the STRING network of the WorldClim variables associated genes (A-C). The networks were functionally enriched for the biological process gene ontology terms: (A) Ribosomal small subunit assembly, mitochondrial translation, ribosome assembly, translation and ribonucleoprotein complex assembly; (B) Epoxygenase P450 pathway, L-ascorbic acid metabolic process, xenobiotic metabolic process, glutathione metabolic process and monocarboxylic acid metabolic process; and (C) Sensory perception of chemical stimulus, G-protein coupled receptor signalling pathway and system process. Wolves' subnetwork of the STRING network of the species richness associated genes (D) which was functionally enriched for the biological process gene ontology term defense response to bacterium. FRDs' subnetworks of the STRING network of the WorldClim variables associated genes (E-G). The networks were functionally enriched for the biological process gene ontology terms; (E) detection of chemical stimulus, G-protein coupled receptor signalling pathway, system process, nervous system process and cell communication; (F) negative regulation of hydrogen peroxide metabolic process, cellular response to UV-A, collagen catabolic process, regulation of reactive oxygen species metabolic process and extracellular matrix organization; (G) Cdc42 protein signal transduction, dendritic spine development, dendrite development, neuron projection development and regulation of anatomical structure morphogenesis. FRDs' subnetwork of the STRING networks of the species richness associated genes were functionally enriched for the biological processes gene ontology terms; (H) cell-cell signalling involved in cardiac conduction, calcium ion imports across plasma membrane, regulation of heart rate by cardiac conduction, cardiac muscle cell action potentiation involved in contraction and multicellular organismal signalling; (I) the canonical Wnt signaling pathway



Legend:

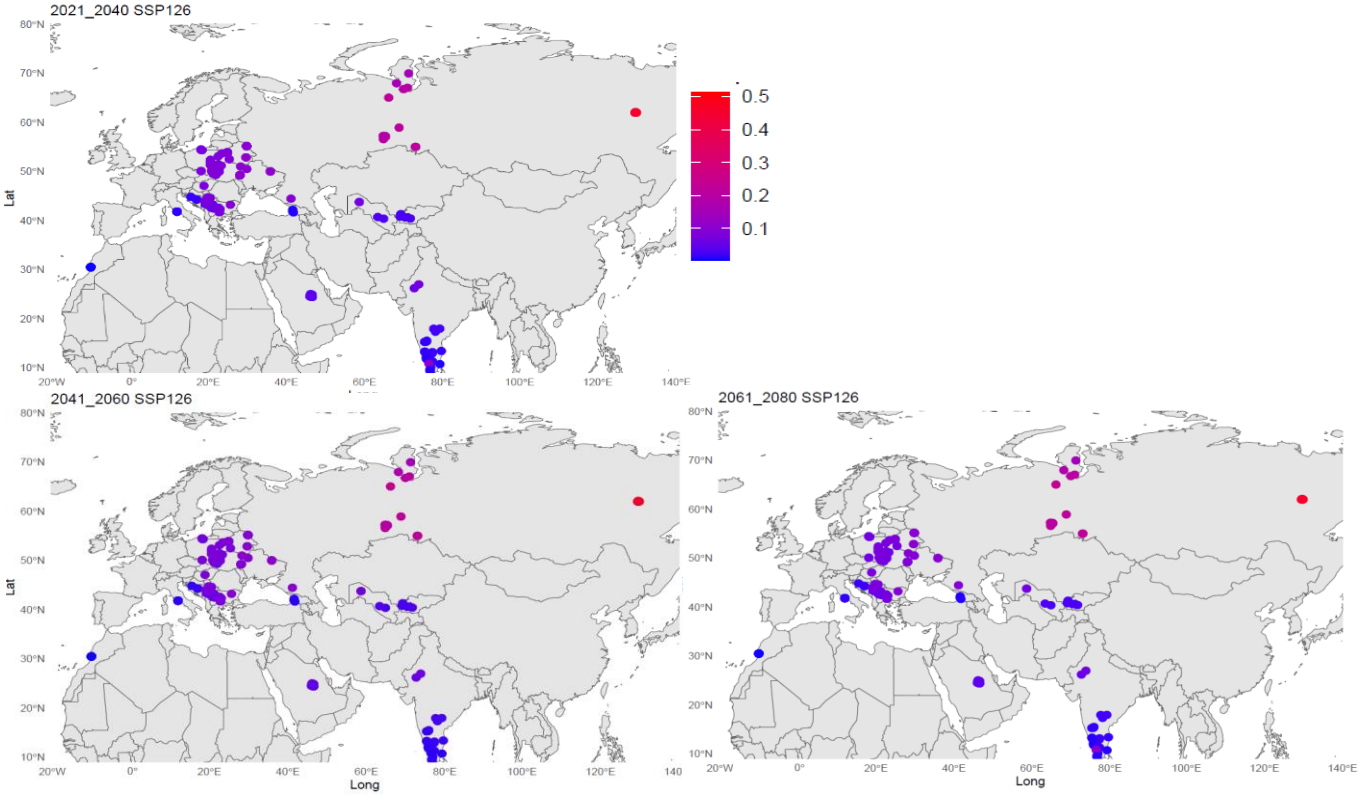
- Protein homology
- Co-expression
- From curated databases
- Text mining
- Experimentally determined
- Gene co-occurrence

Supplementary figure 19. STRING protein network genes under diversifying and local adaptation in FRDs. Protein association STRING network for the genes with signatures of both diversifying and local adaptation. Non-connected proteins were hidden for clarity. The network includes 189 nodes (i.e., gene products), the average local clustering coefficient was 0.39 and a PPI enrichment p-value of 1.06e-08. The dashed lines indicate interaction with low confidence or evidence. The network had functional enrichment for the gene ontology terms subcellular localization, immunoglobulin

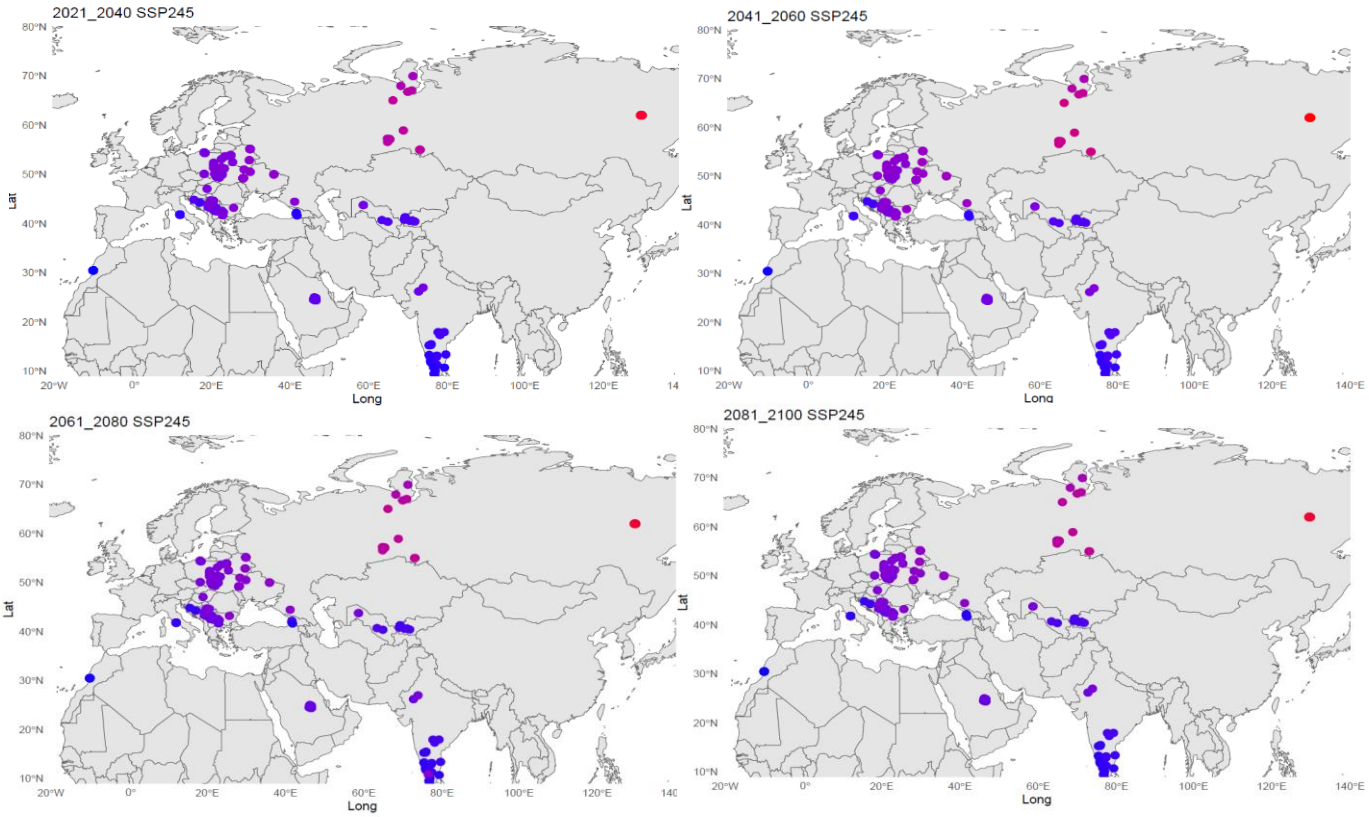
complex and circulation. The largest 3 local network clusters were enriched for the gene ontology terms immunoglobulin V-type, immunoglobulin V-set domain, and phagocytosis, pathogen recognition and immunoglobulin v-type with false discovery rates of 0.0004, and 0.0120 and strength of 1.21 and 1.3, respectively.

Supplementary figures: Chapter 4

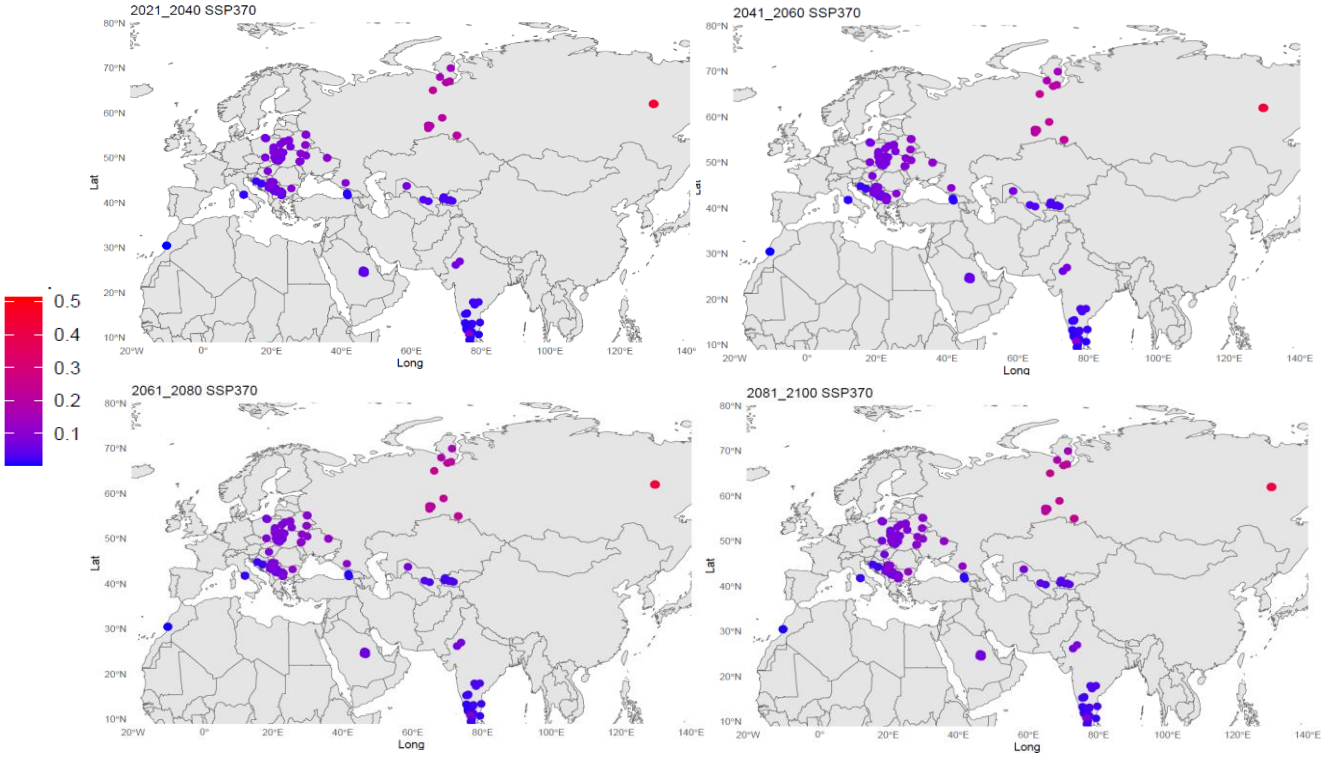
A. FRDs Genomic offset SSP 126



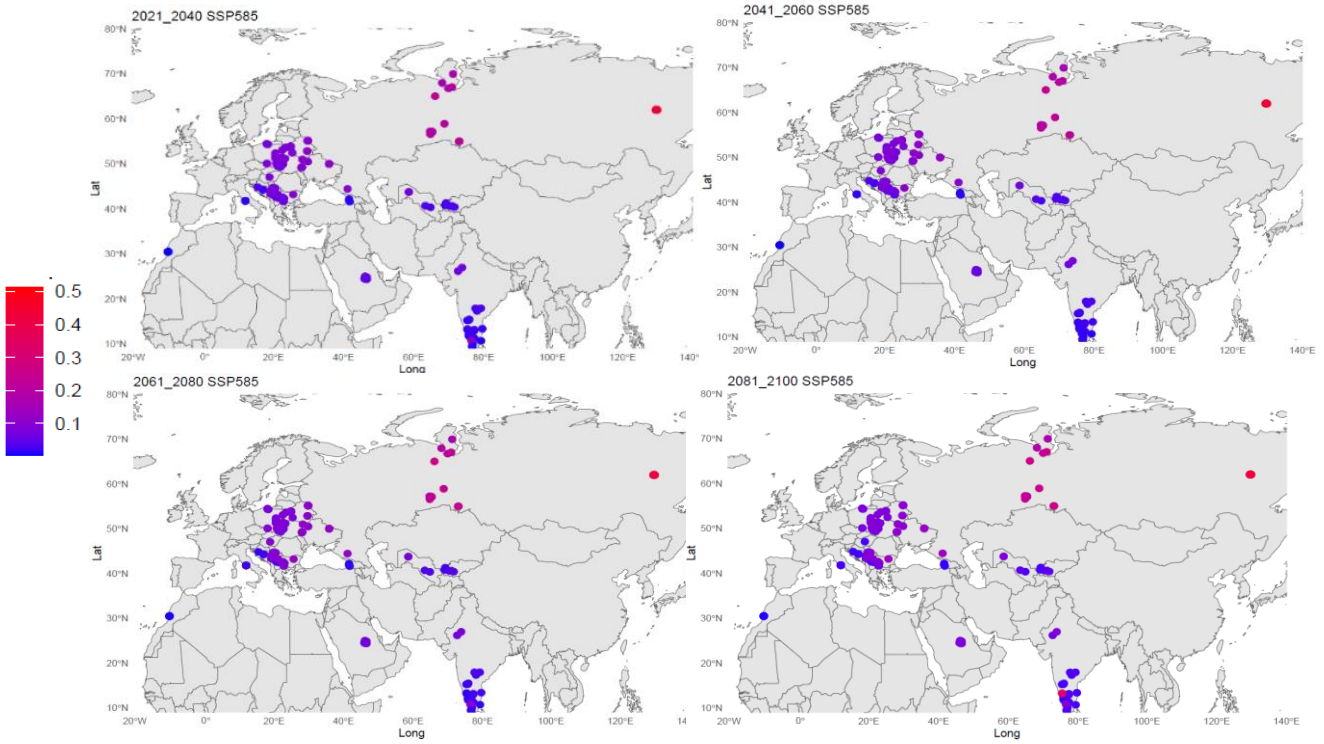
B. FRDs Genomic offset SSP 245



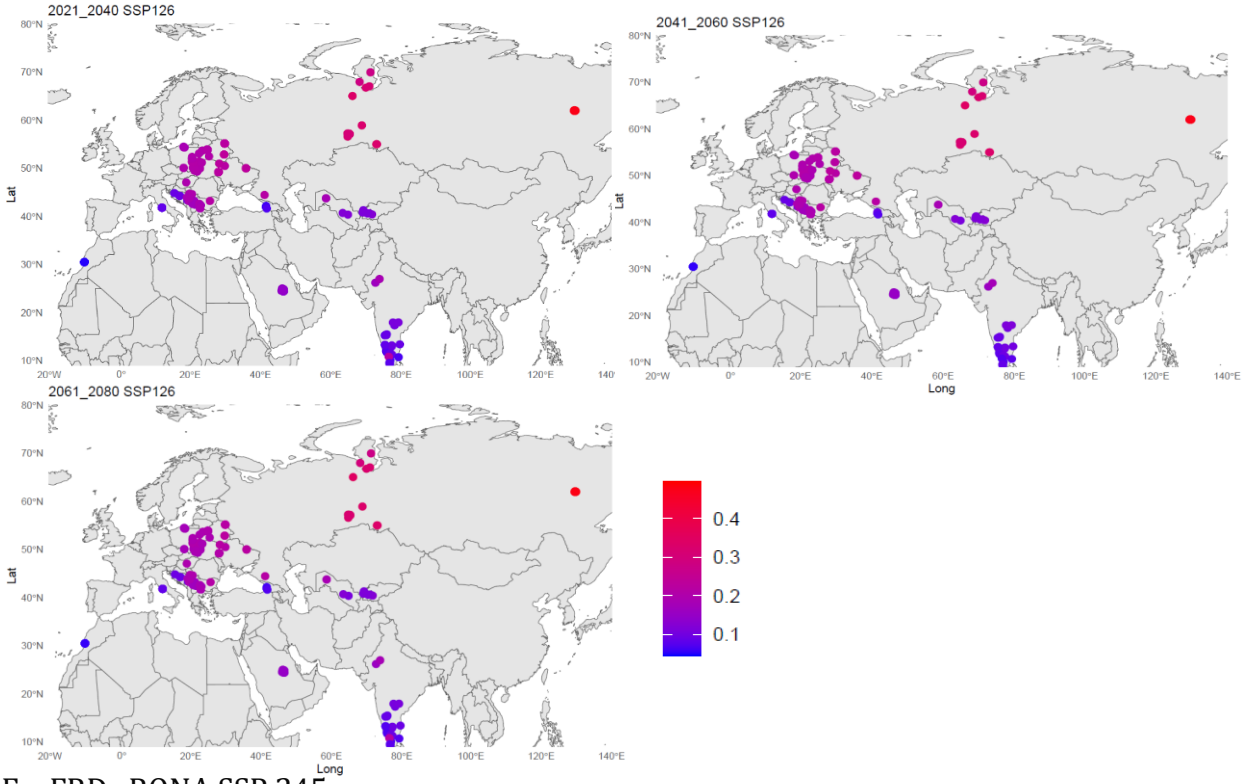
C. FRDs Genomic offset SSP 370



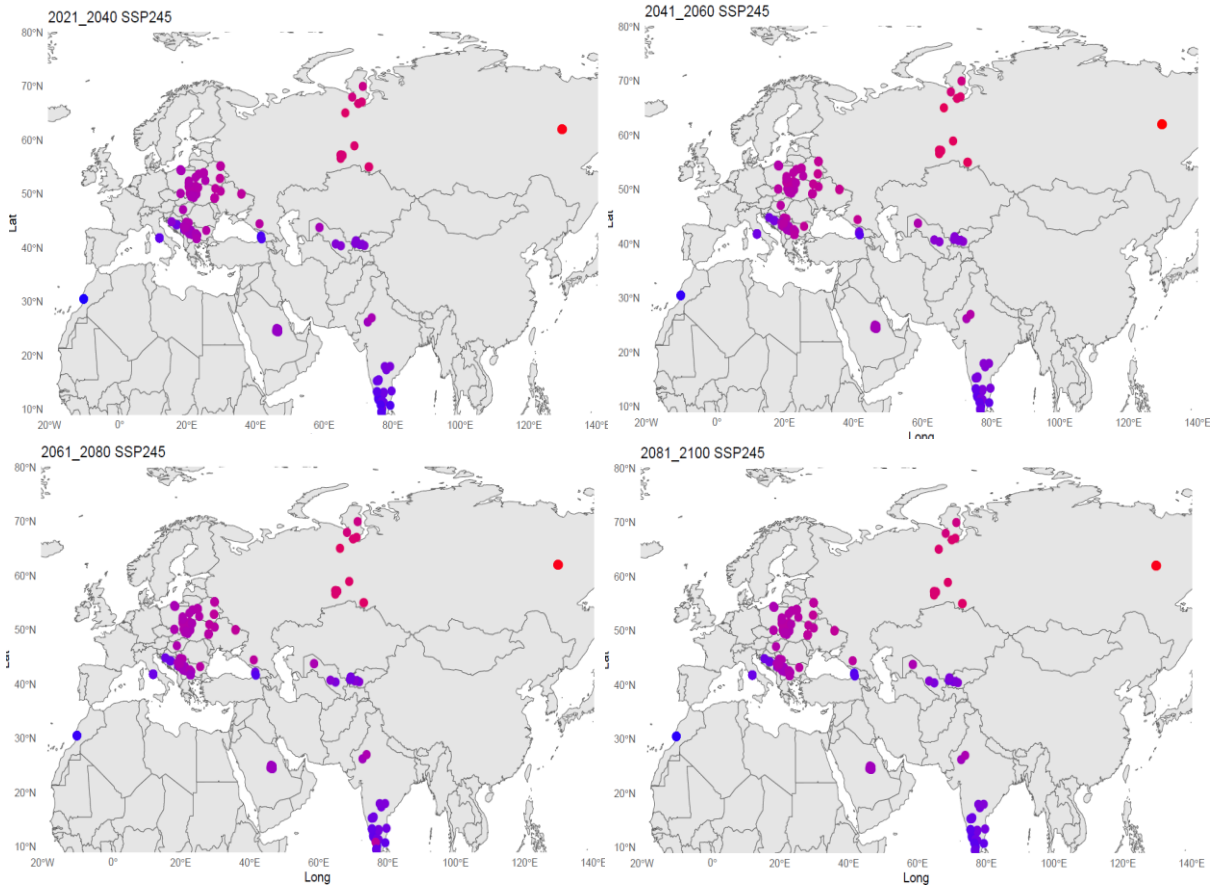
D. FRDs Genomic offset SSP 585



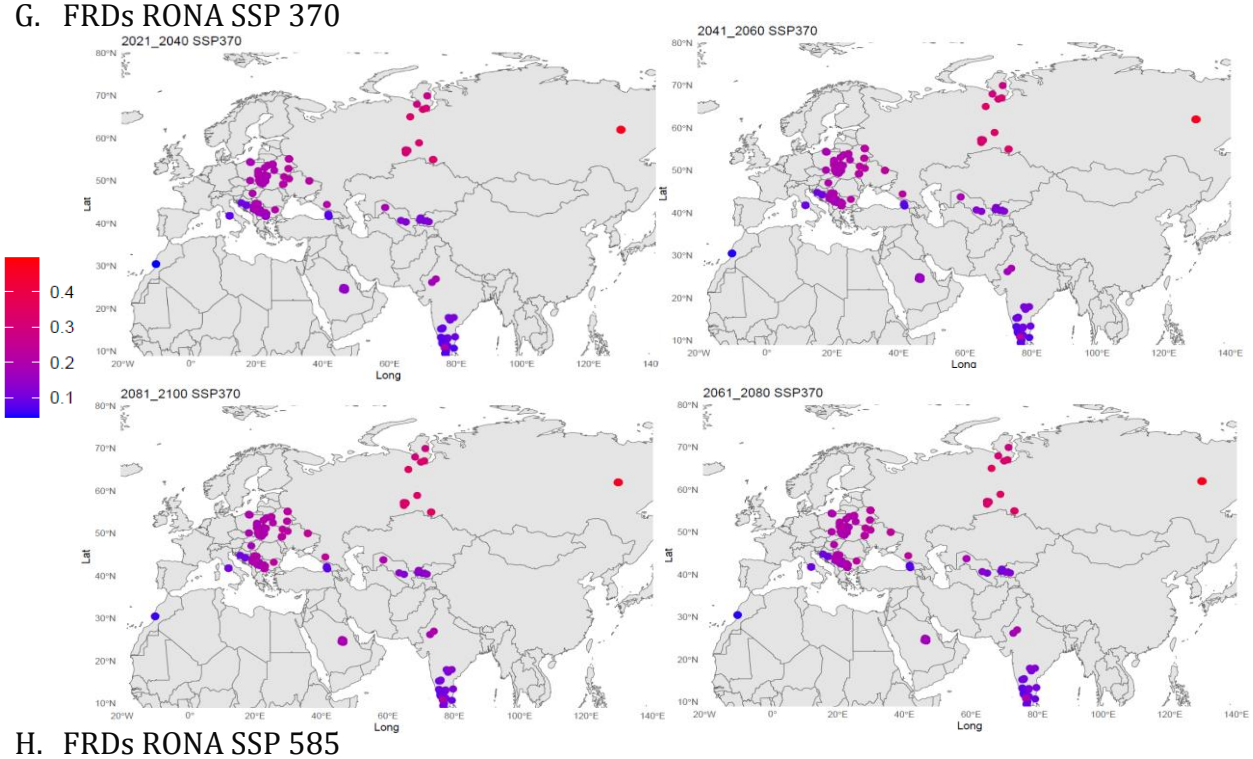
E. FRDs RONA SSP 126



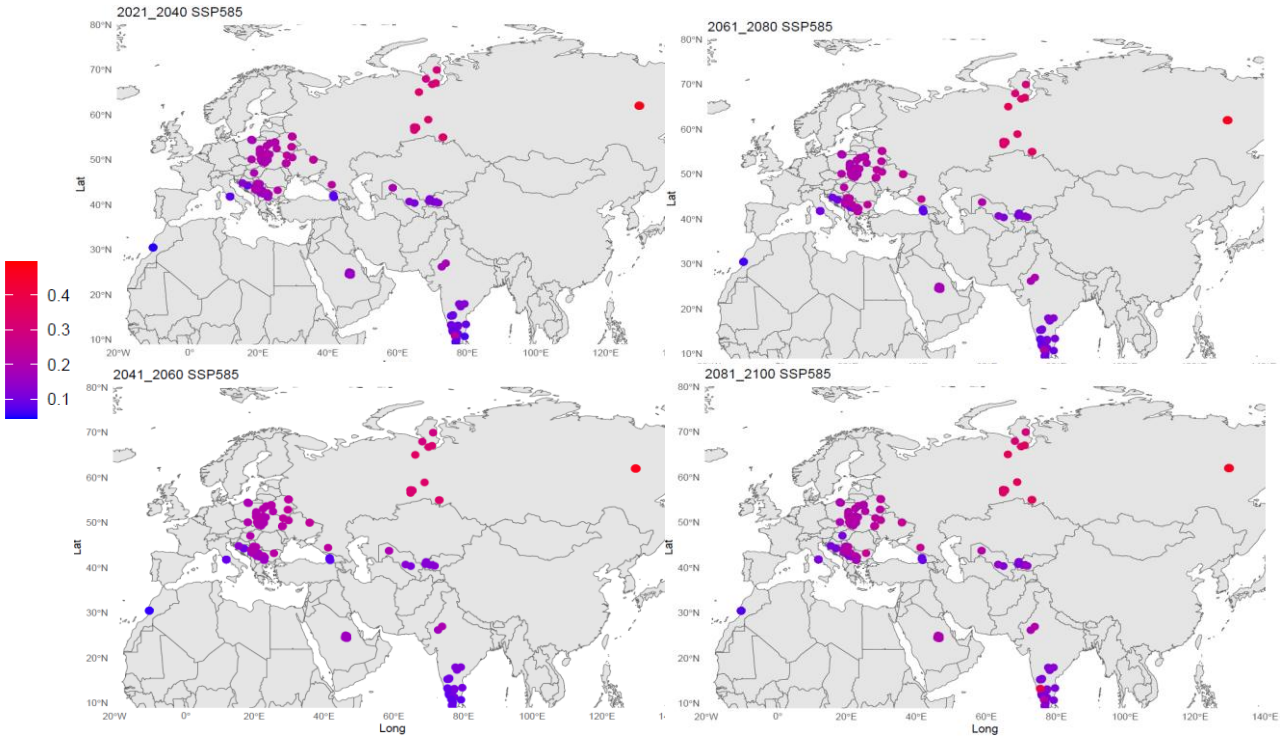
F. FRDs RONA SSP 245



G. FRDs RONA SSP 370



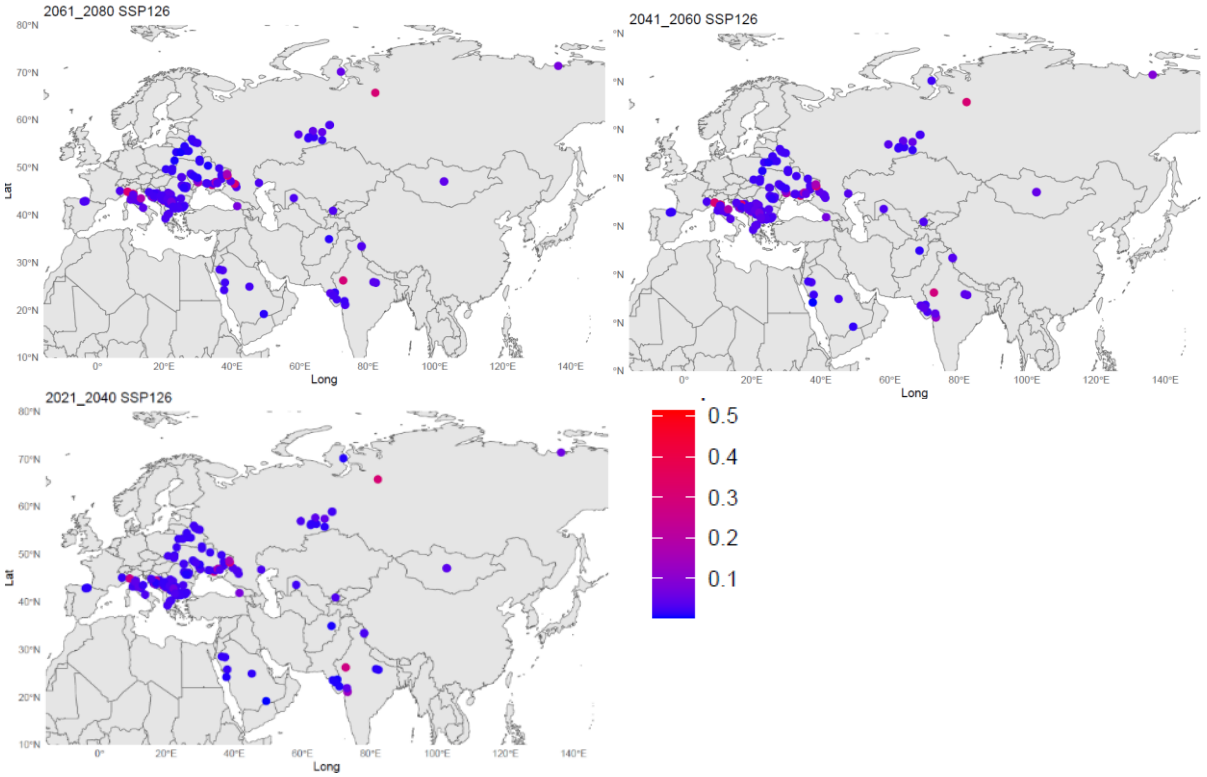
H. FRDs RONA SSP 585



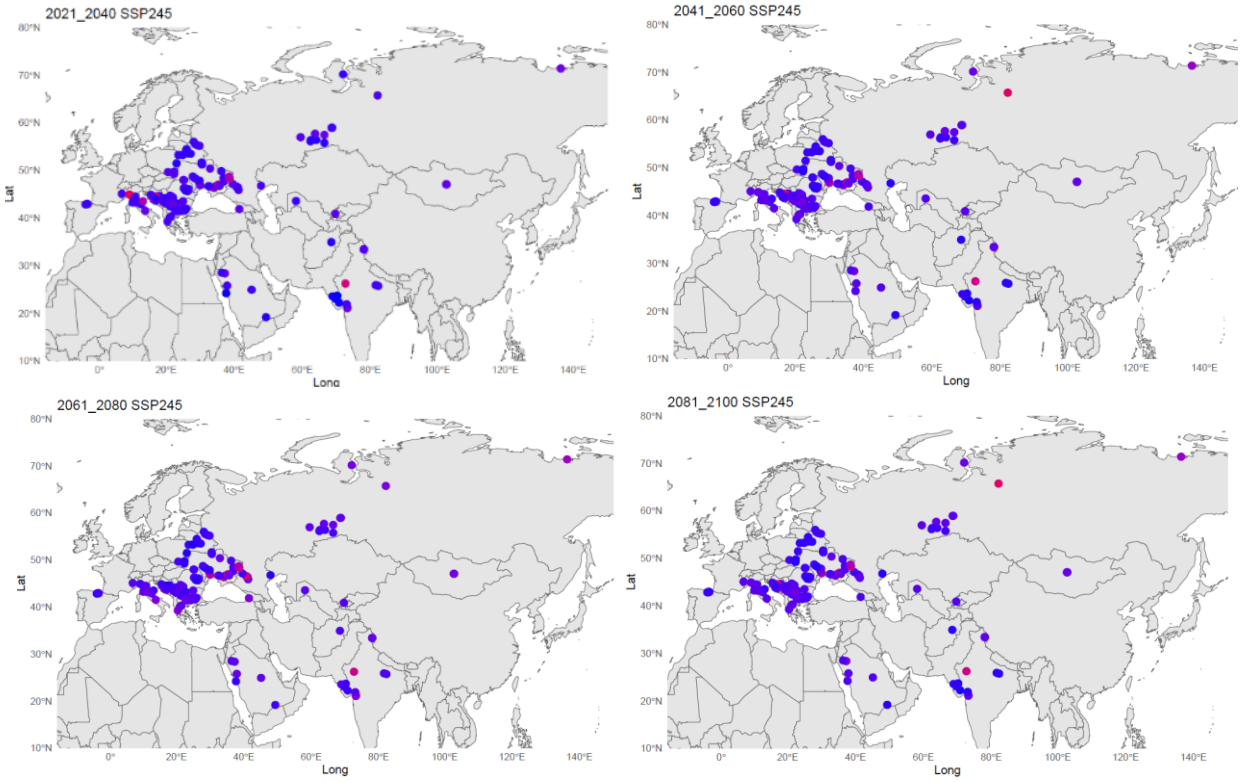
Supplementary figure 1. Genomic offset and risk of nonadaptiveness (RONA) mapped results for FRDs. A-D. FRDs genomic offset results for the socioeconomic pathways (SSPs) 126, 245, 370 and 585, time frames from 2021-2100, respectively. E-H. FRDs RONA results for the SSPs 126, 245, 370 and 585, time frames from 2021-2100. For the 126 SSP the time

frame does not include 2081-2100 as the environmental variables values were not available in WorldClim, thus the analysis for this SSP in this timeframe was not completed. Both genomic offset and RONA values in FRDs showed only minor changes through time frames and SSPs. Overall, only FRDs from Southern Eurasia showed lower values of genomic offset and RONA thorough time frames and SSPs. Individuals from Northern Eurasia showed the highest values of genomic offset and RONA, thus the least retention of adaptive potential.

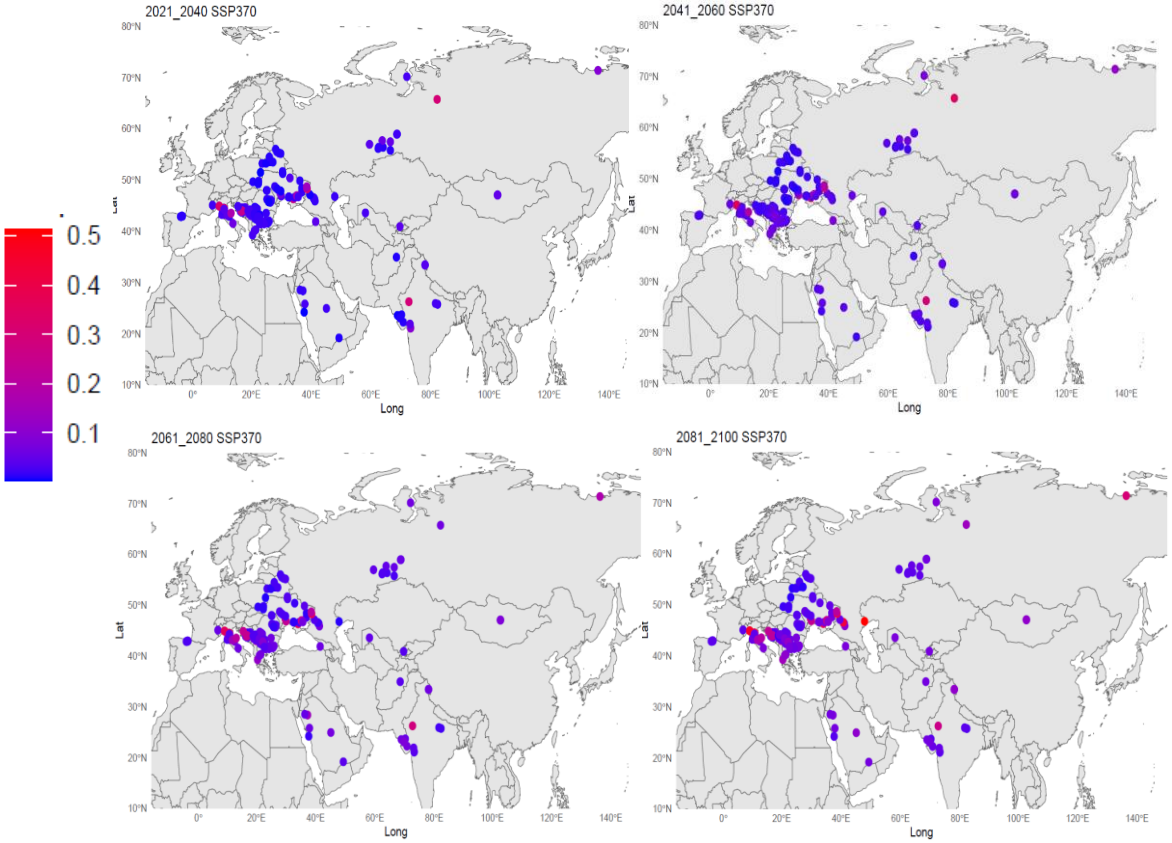
A. Wolves genomic offset SSP 126



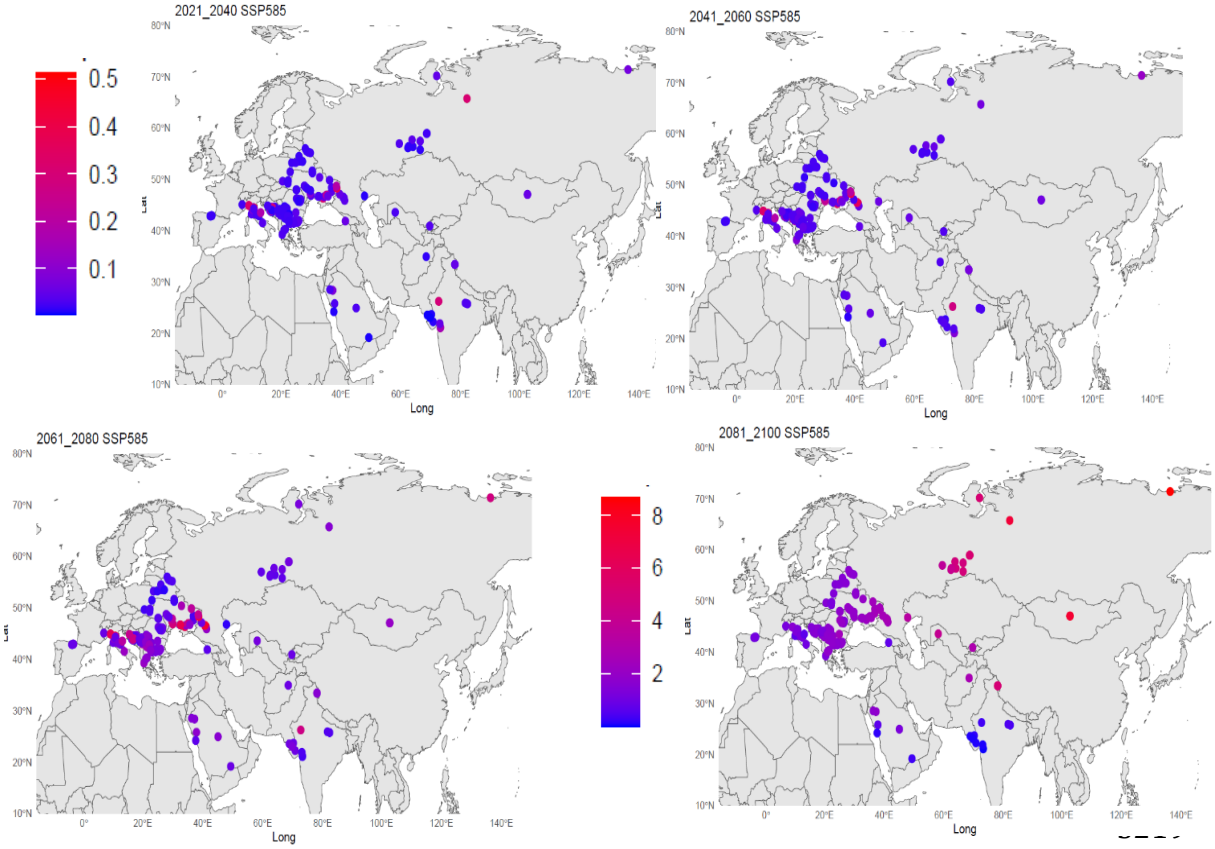
B. Wolves genomic offset SSP 245



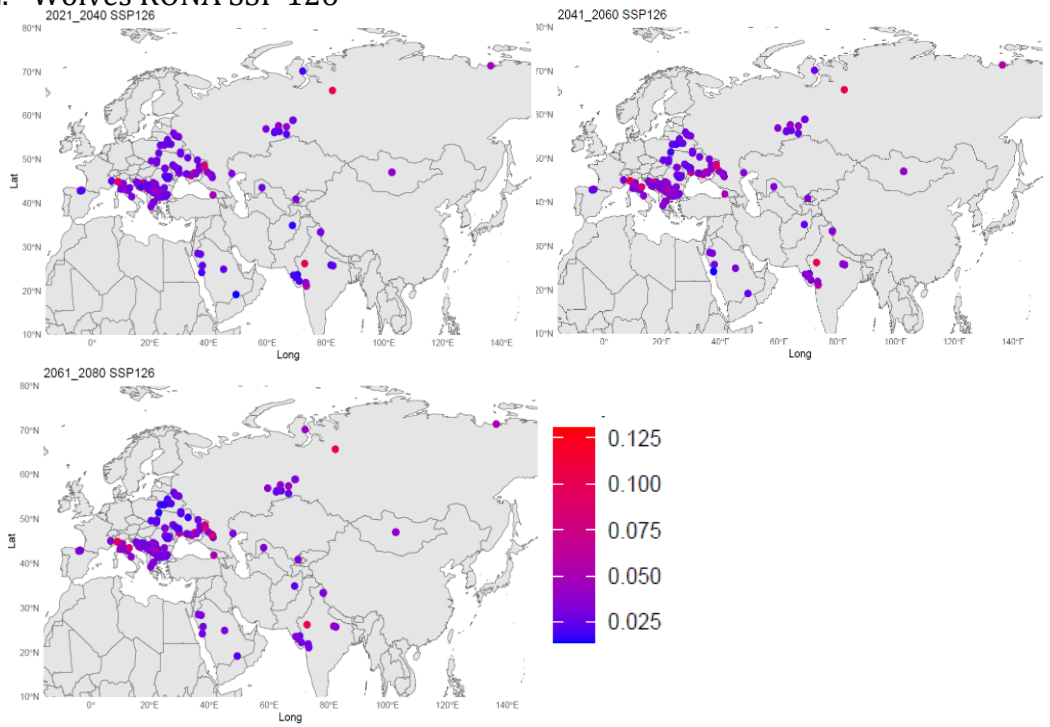
C. Wolves genomic offset SSP 370



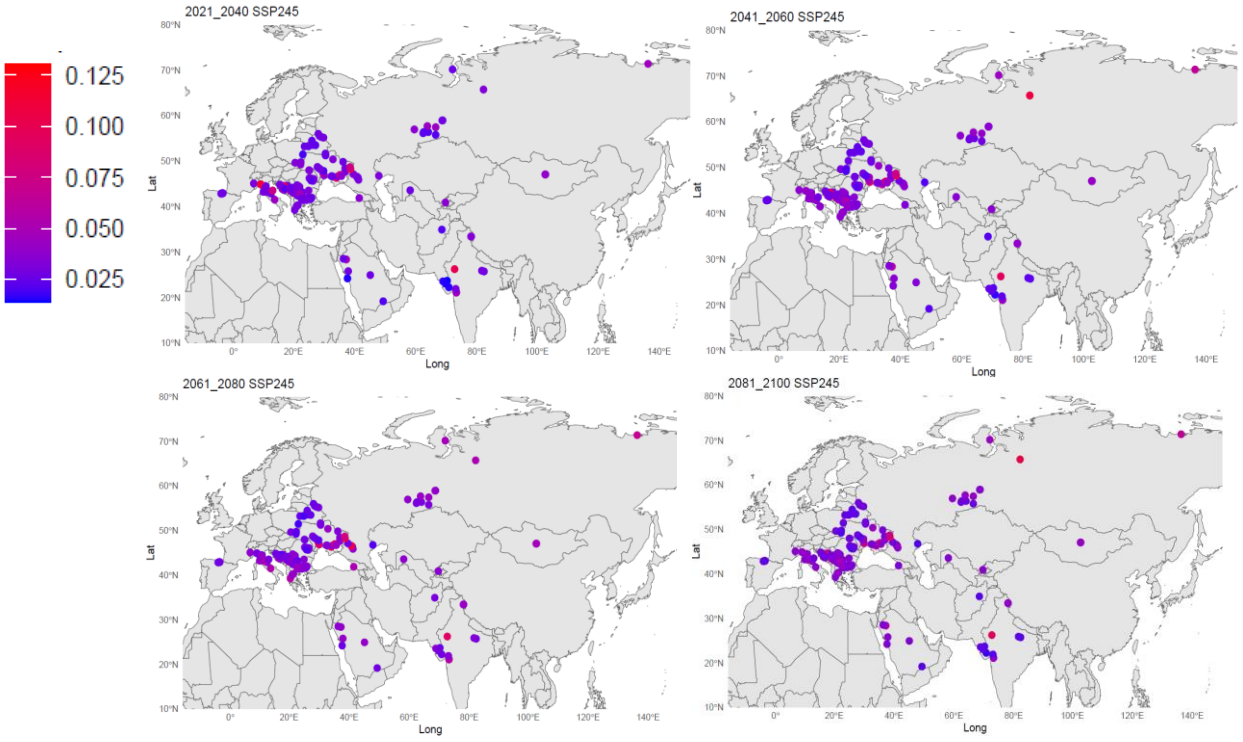
D. Wolves genomic offset SSP 585



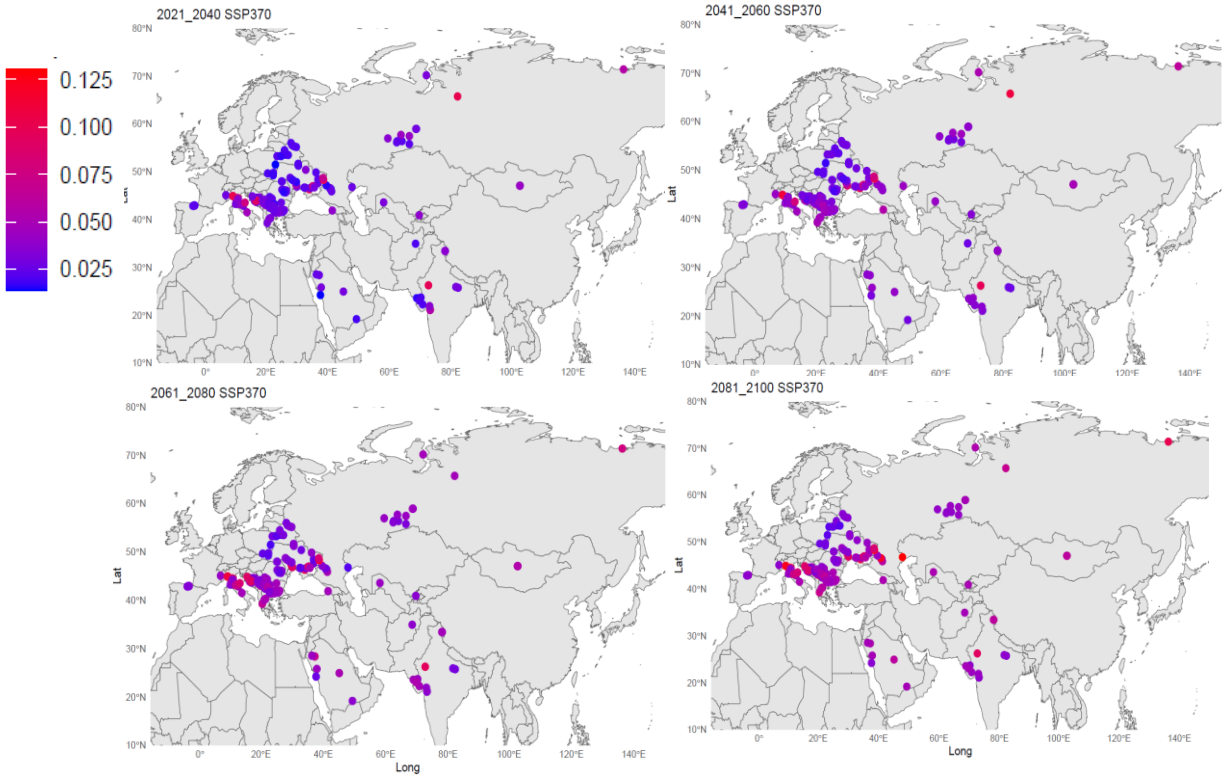
E. Wolves RONA SSP 126



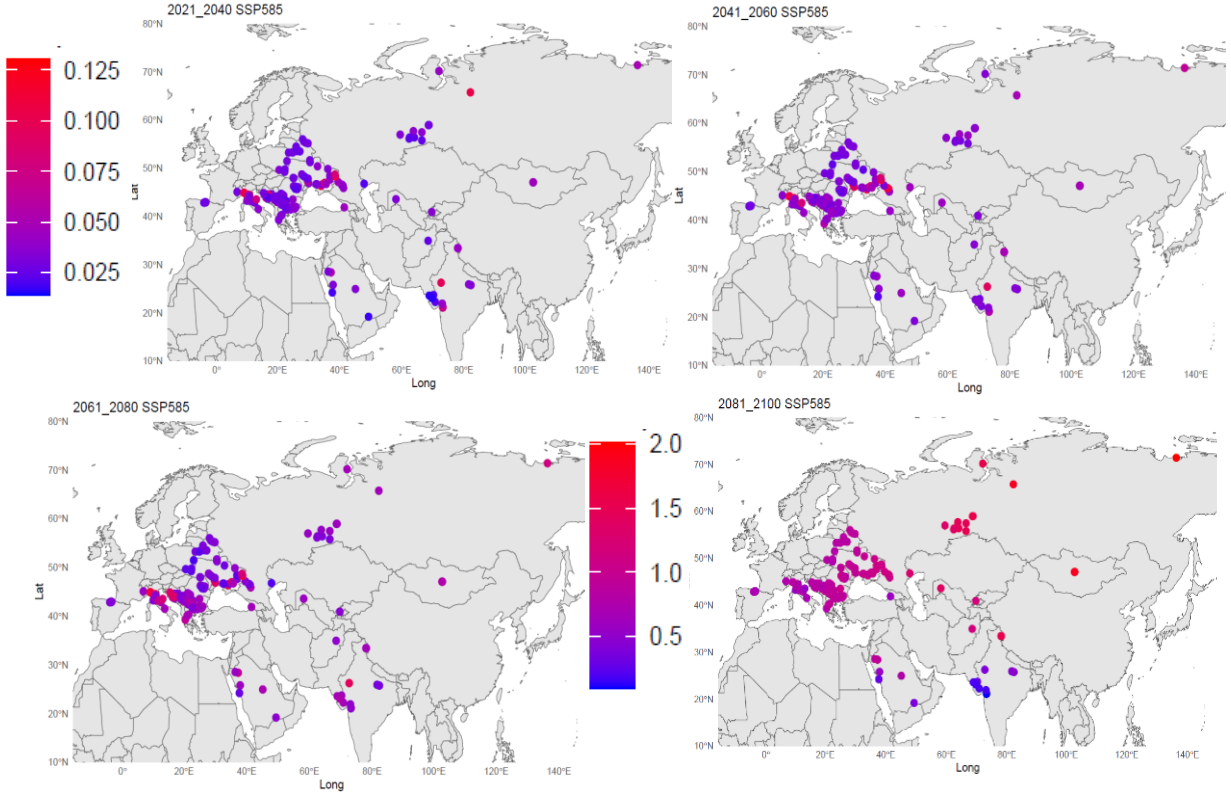
F. Wolves RONA SSP 245



G. Wolves RONA SSP 370

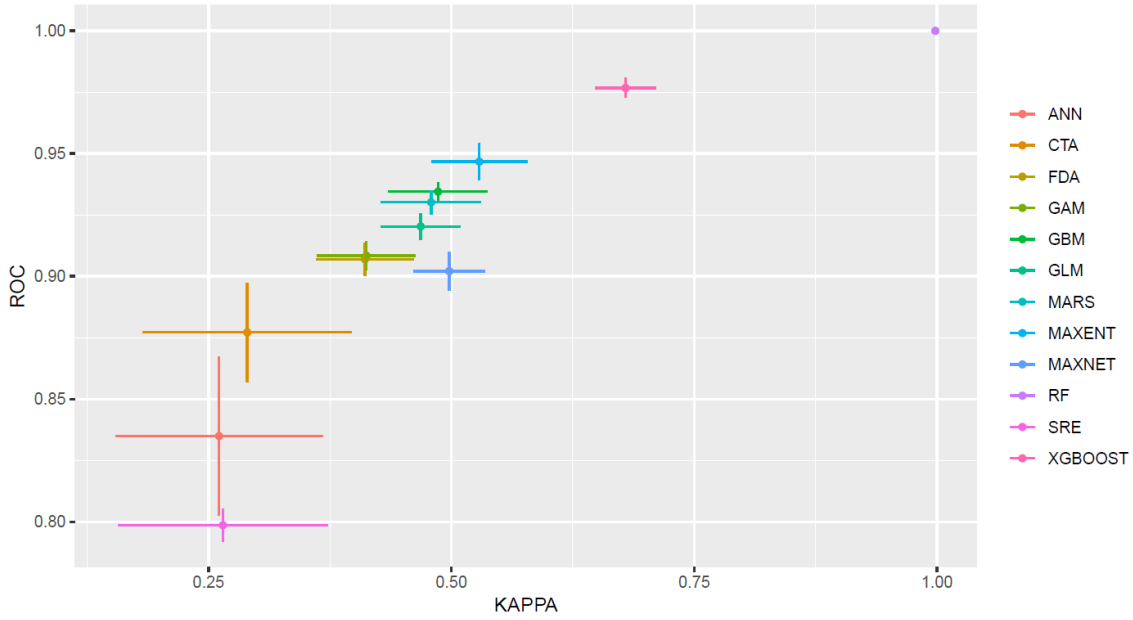


H. Wolves RONA SSP 585

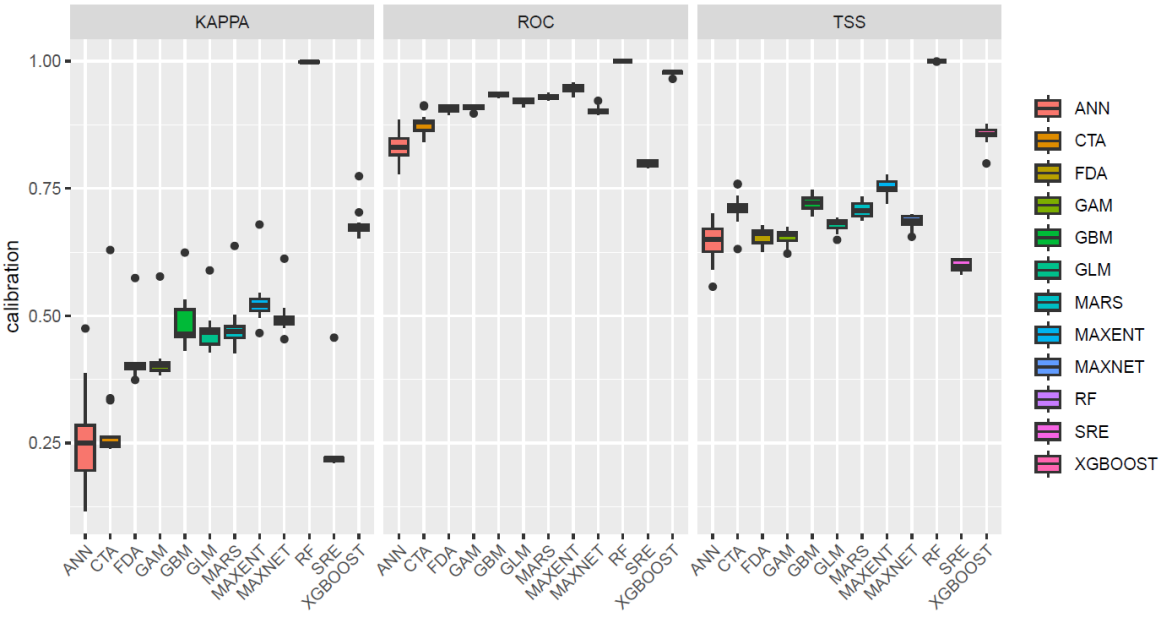


Supplementary figure 2. Genomic offset and risk of nonadaptiveness (RONA) mapped results for wolves. A-D. Wolves genomic offset results for the socioeconomic pathways (SSPs) 126, 245, 370 and 585, time frames from 2021-2100, respectively. E-H. Wolves RONA results for the SSPs 126, 245, 370 and 585, time frames from 2021-2100. For the 126 SSP the time frame does not include 2081-2100 as the environmental variables values were not available in WorldClim, thus the analysis for this SSP in this timeframe was not completed. Both genomic offset and RONA values in wolves showed only slight changes throughout most time frames and SSPs. Increase in genomic offset and RONA values were detected at the time frame 2061-2080 in the 585 SSP and the 2081-2100 time frame the 370 and 585 SSPs. Eurasian grey wolves showed the highest increase in genomic offset (i.e., decrease in adaptive potential) and risk of nondaptiveness at the 2081-2100 time frame under the 585 SSP.

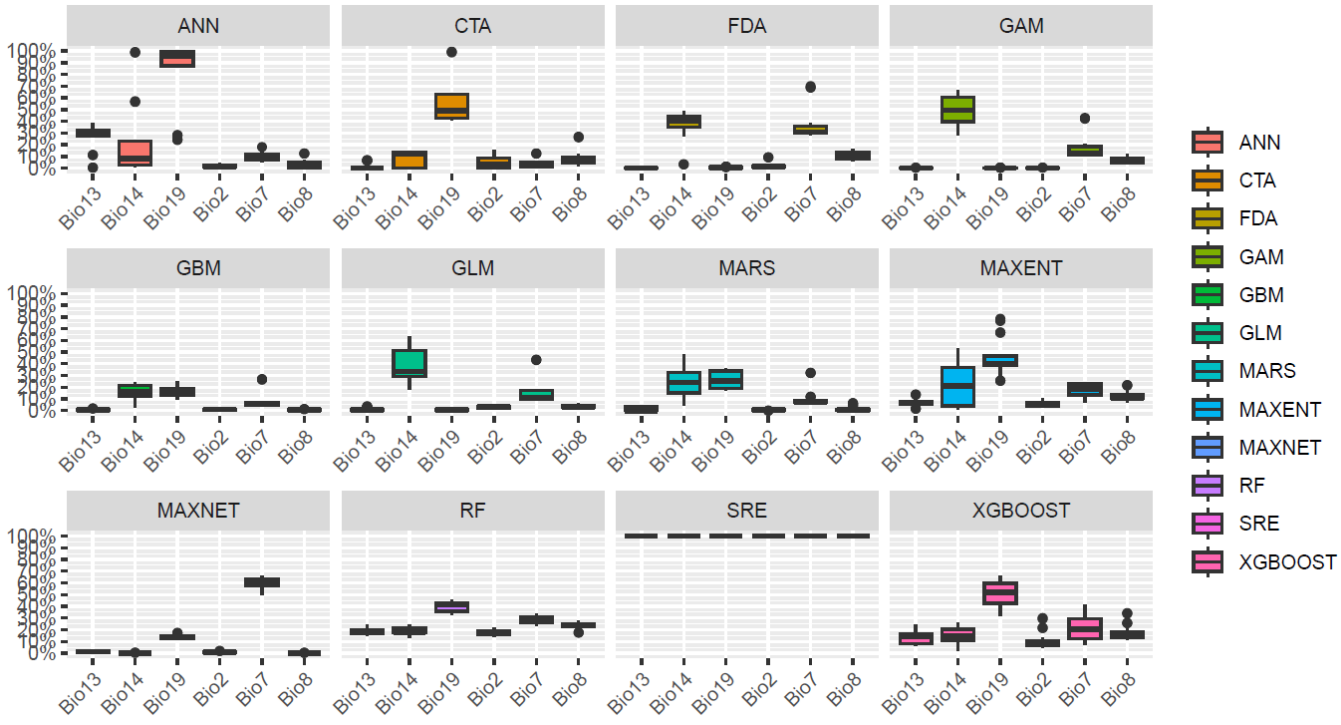
A. Mean evaluation scores



B. Calibration scores for single models

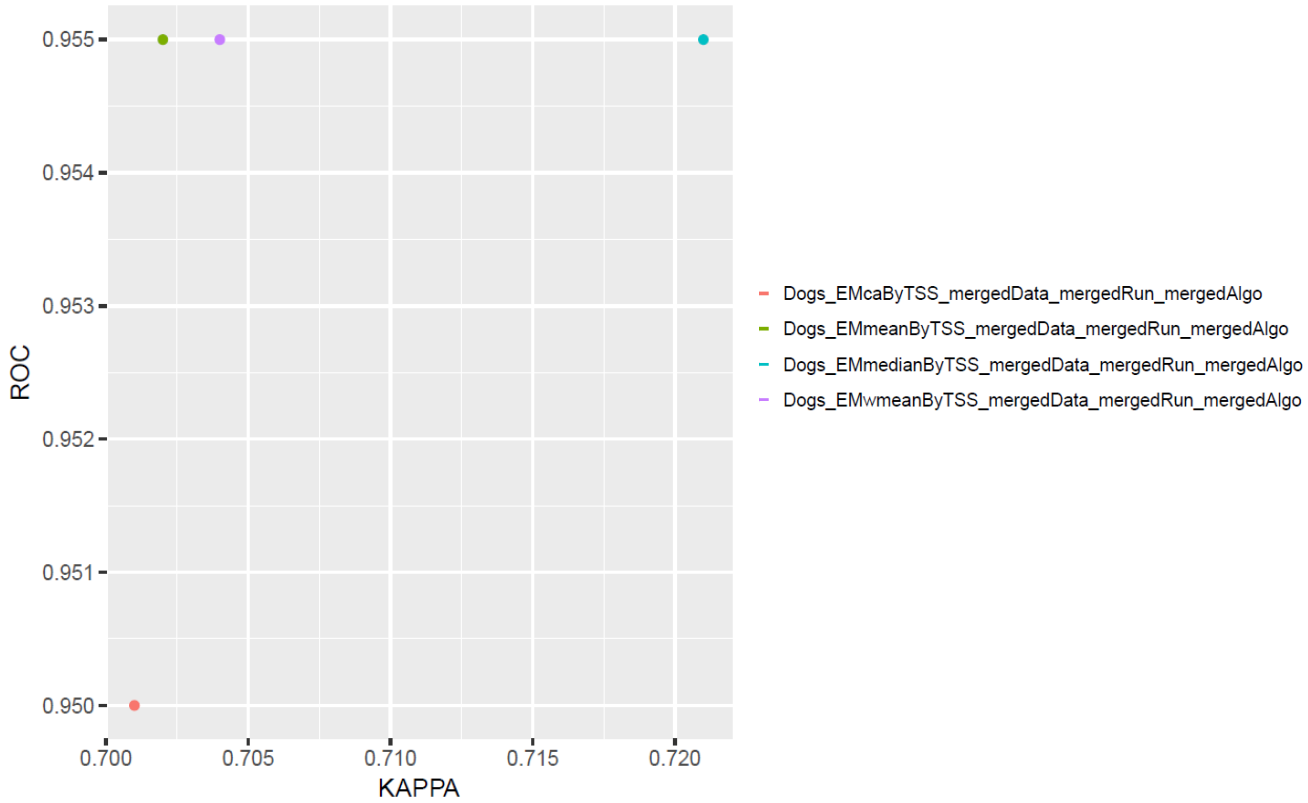


C. Variable importance

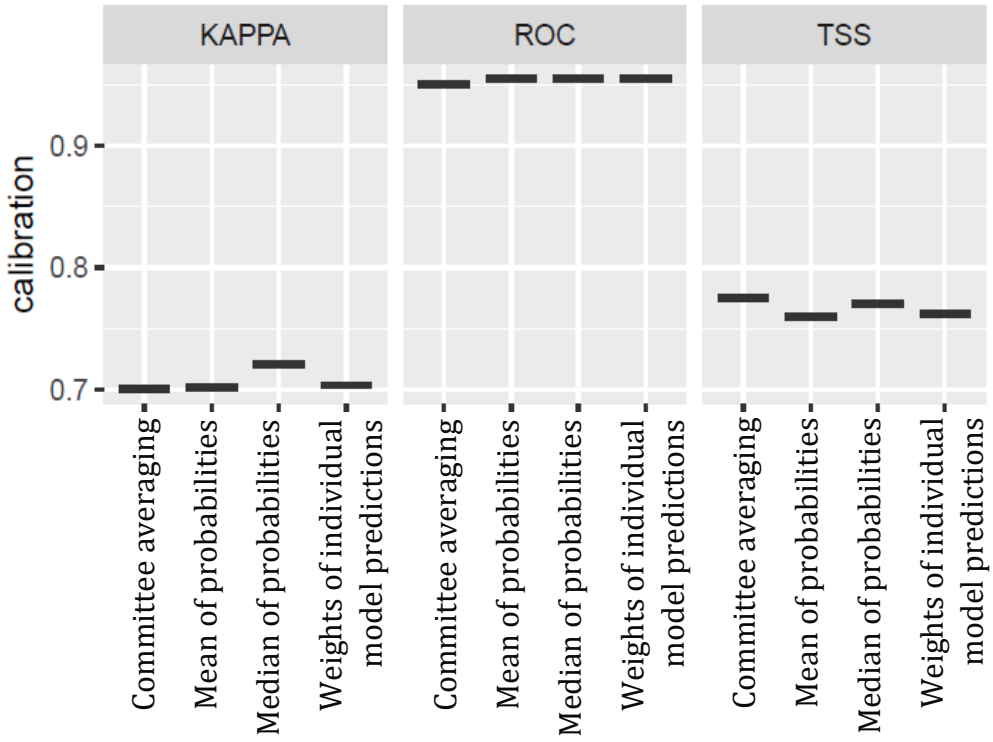


Supplementary figure 3. Evaluation of single species distribution models. A. Mean evaluation scores with their standard deviation of the single species distribution models. The x-axis represents area under the ROC curve and the y-axis the Cohen's Kappa which is a measure of the models. B. Boxplot of calibration scores for the species distribution single models KAPPA, ROC and TSS grouped model. ROC values are high across the board thus the models are accurate in predicting presence and absence location. Only the three models with KAPPA values over 0.5 were selected for suitable niche modelling. Models with lower values were less reliable, i.e., classification of presence and absence at the individual locations are not accurate. C. Boxplot of climatic variable importance by species distribution model. The single models evaluated were artificial neural networks (ANN), classification tree analysis (CTA), flexible discriminant analysis (FDA), generalized additive model (GAM), generalized boosting model (GBM), generalized linear model (GLM), multivariate adaptive regression splines (MARS), maximum entropy (MAXENT), alternative maximum entropy algorithm (MAXNET), random forest (RF), surface range envelope and extreme gradient boosting (XGBOOST).

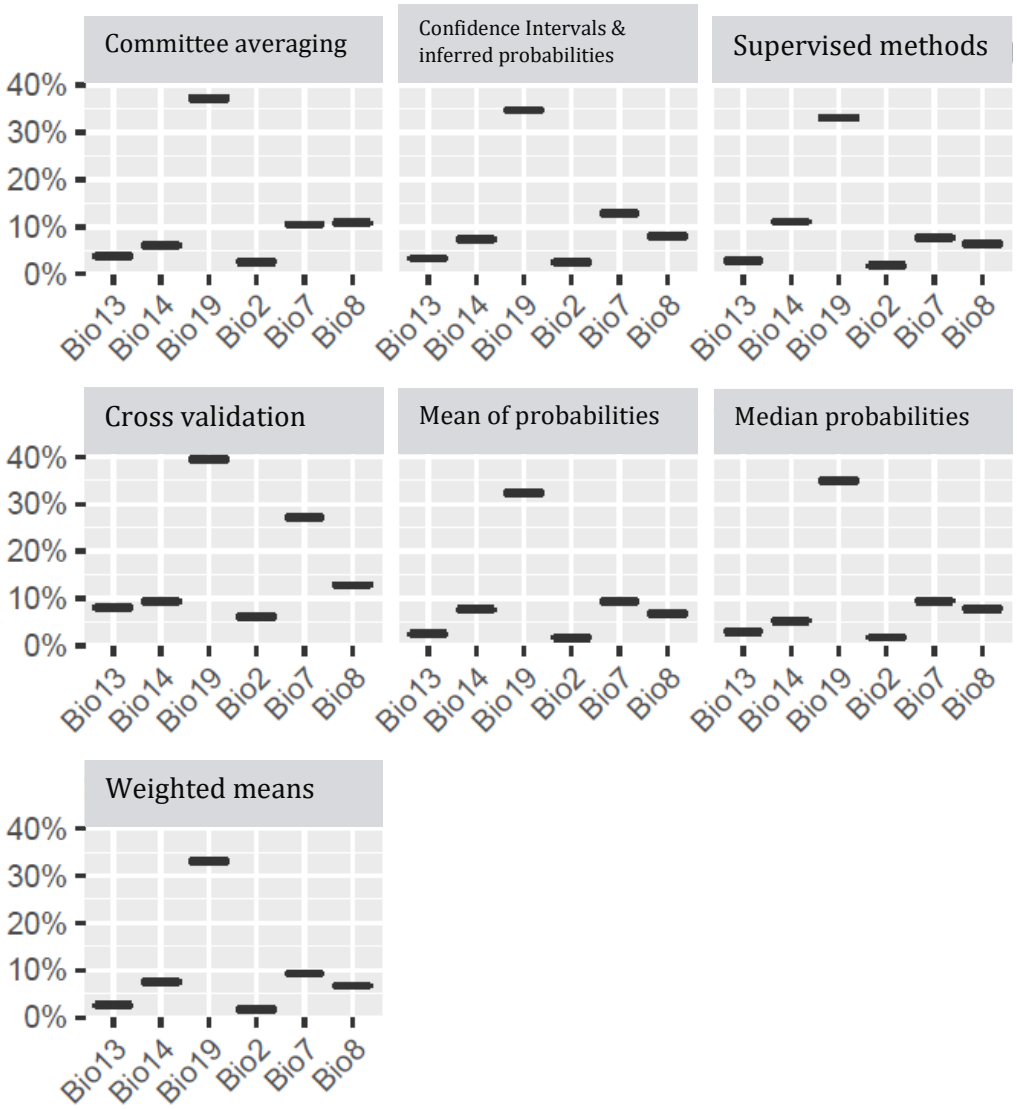
A. Mean evaluation scores for ensemble model



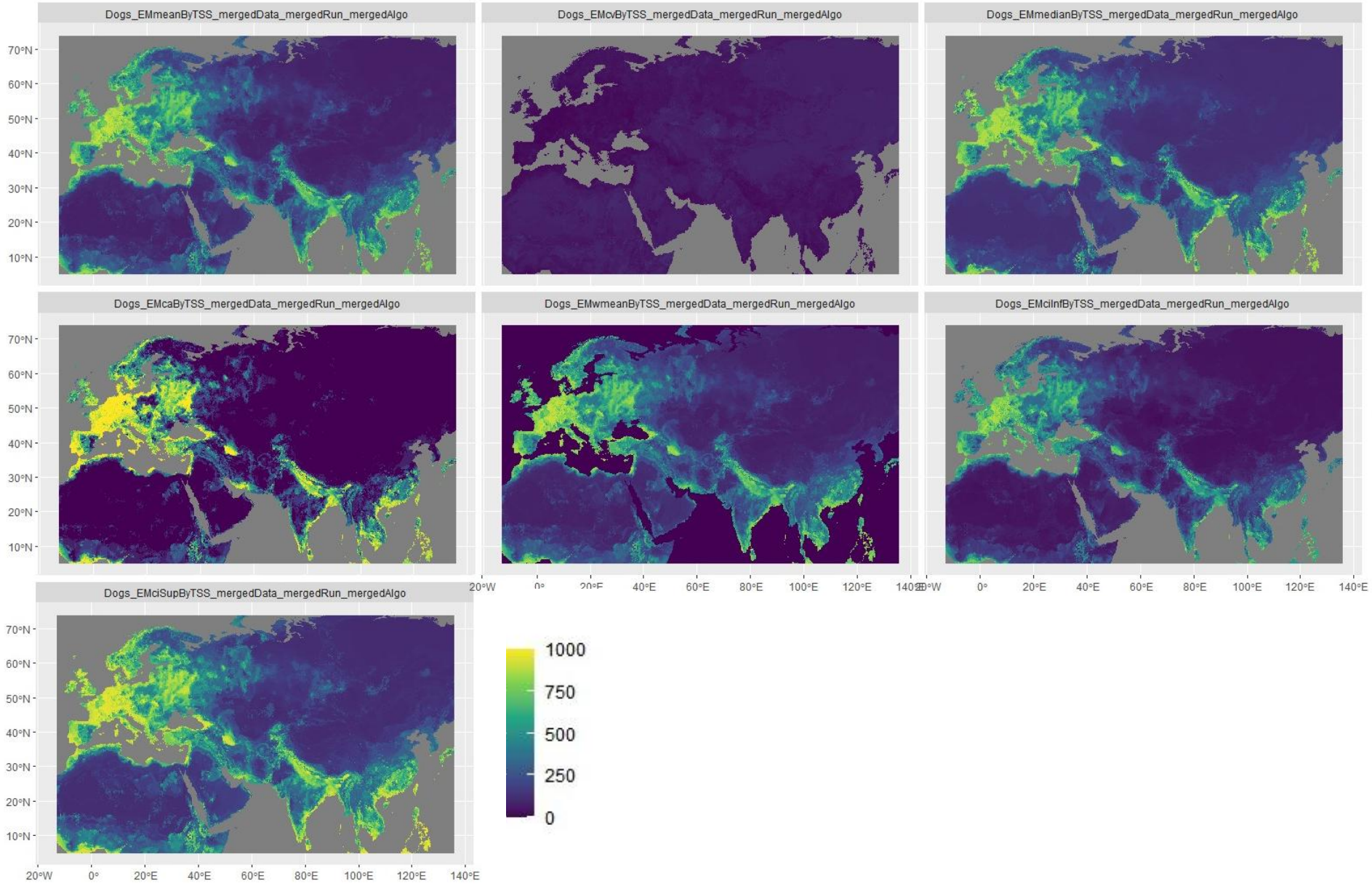
B. Calibration scores for ensemble model



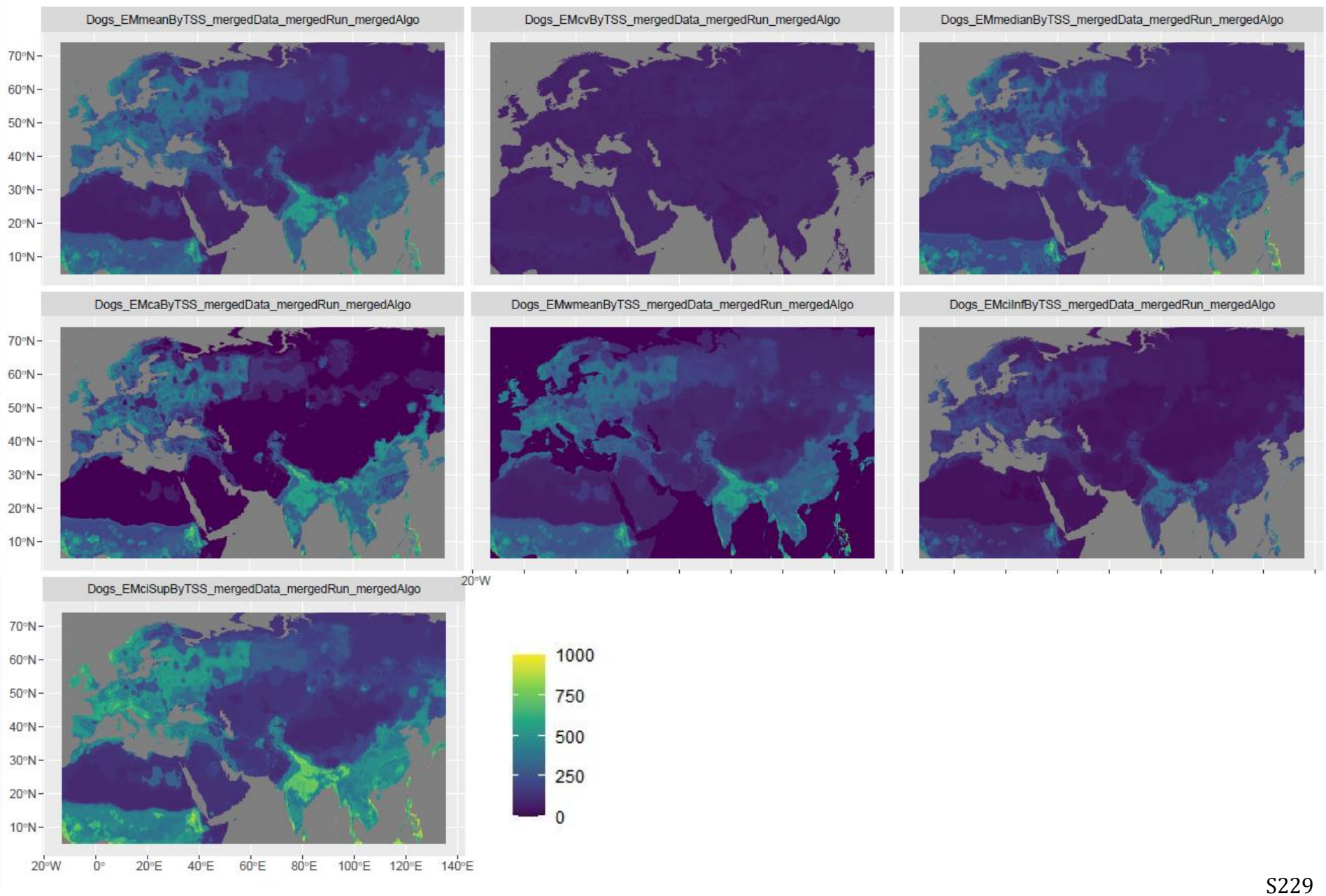
C. Variable importance for ensemble model



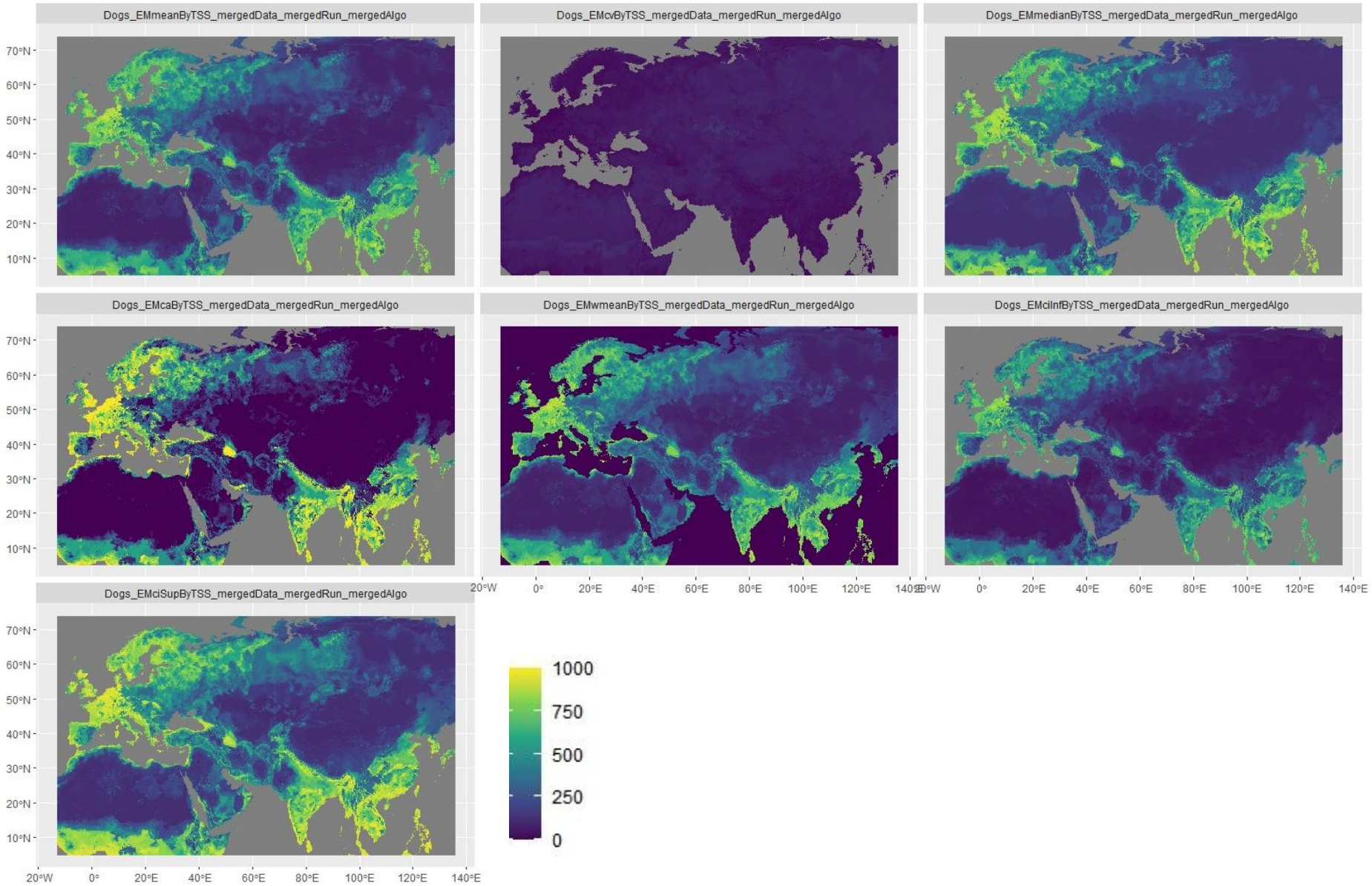
Supplementary figure 4. Evaluations for ensemble model. A. Mean evaluation scores with their standard deviation of the ensembled suitable ecological niche models. B. Boxplot of calibration scores in KAPPA, ROC and TSS, for the ensembled models by committee averaging i.e., runs merged by each model algorithm, mean of probabilities over the selected models, median of probabilities over the selected models and weights of individual model predictions according to their performance all by TSS, merged data, merged run and merged algorithms. C. Boxplot of variable importance for the ensemble model by committee averaging, with confidence intervals and inferred probabilities, supervised methods, cross validation, mean of probabilities, median probabilities and weighted means by TSS, merged data, merged run and merged algorithms. The ensemble model was built with models RF, XGBOOST and MAXENT models, see above.


Dogs_EMmeanByTSS_mergedData_mergedRun_mergedAlgo
Dogs_EMcvByTSS_mergedData_mergedRun_mergedAlgo
Dogs_EMmedianByTSS_mergedData_mergedRun_mergedAlgo
Dogs_EMcaByTSS_mergedData_mergedRun_mergedAlgo
Dogs_EMwmeanByTSS_mergedData_mergedRun_mergedAlgo
Dogs_EMciInfByTSS_mergedData_mergedRun_mergedAlgo
Dogs_EMciSupByTSS_mergedData_mergedRun_mergedAlgo
70°N
60°N
50°N
40°N
30°N
20°N
10°N
20°W
0°
20°E
40°E
60°E
80°E
100°E
120°E
140°E
1000
750
500
250
0

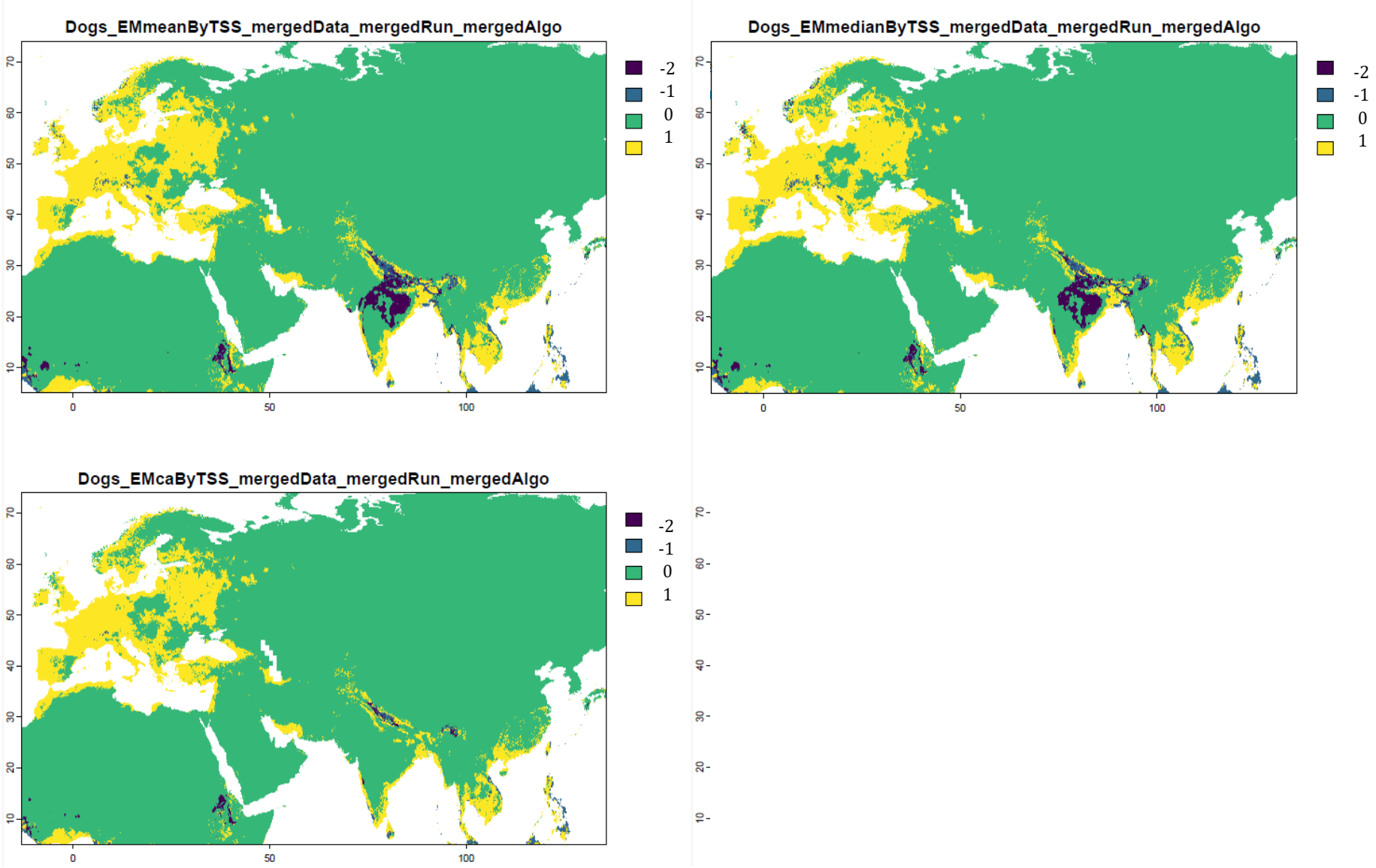
Supplementary figure 5. Suitable ecological niche modelling ensemble model for current environmental conditions. Each of the panels represents the suitable ecological niche model by committee averaging i.e., runs merged by each model algorithm (top left), with confidence intervals and inferred probabilities (top-middle), supervised methods (top-right), cross validation (middle-left), mean of probabilities (middle-middle), median probabilities (middle-right) and weighted means by TSS, merged data, merged run and merged algorithms (bottom-left). Based on ensemble model and all the employed visualizations, dogs' most suitable current ecological niche is in Southern Asia and Europe a close second.


Dogs_EMmeanByTSS_mergedData_mergedRun_mergedAlgo
Dogs_EMcvByTSS_mergedData_mergedRun_mergedAlgo
Dogs_EMmedianByTSS_mergedData_mergedRun_mergedAlgo
Dogs_EMcaByTSS_mergedData_mergedRun_mergedAlgo
Dogs_EMwmeanByTSS_mergedData_mergedRun_mergedAlgo
Dogs_EMciInfByTSS_mergedData_mergedRun_mergedAlgo
Dogs_EMciSupByTSS_mergedData_mergedRun_mergedAlgo
70°N
60°N
50°N
40°N
30°N
20°N
10°N
20°W
0°
20°E
40°E
60°E
80°E
100°E
120°E
140°E
1000
750
500
250
0

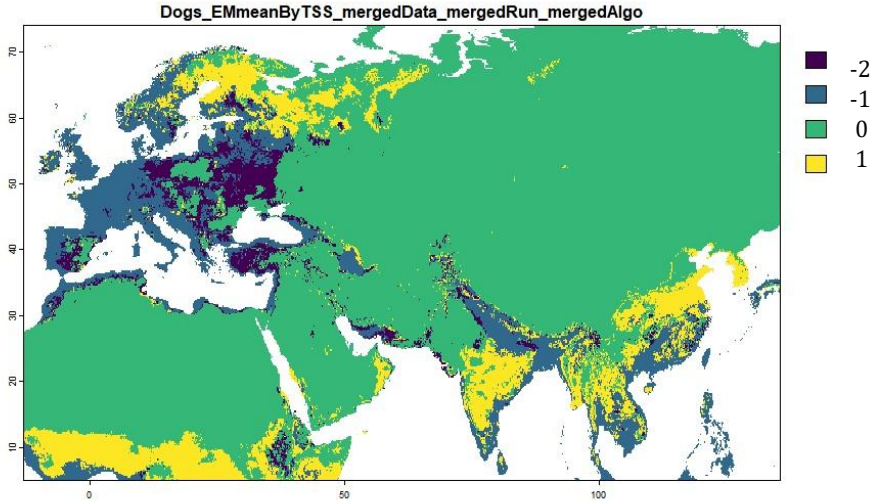
Supplementary figure 6. Suitable ecological niche modelling ensemble model with mid-Holocene environmental rasters. Each of the panels represents the suitable ecological niche model by committee averaging i.e., runs merged by each model algorithm (top left), with confidence intervals and inferred probabilities (top-middle), supervised methods (top-right), cross validation (middle-left), mean of probabilities (middle-middle), median probabilities (middle-right) by TSS, merged data, merged run and merged algorithms (bottom-left). Based on ensemble model and all the employed visualizations, dogs' most suitable ecological niche ~6,000 years ago was in Southern Asia with habitats in Europe a close second.

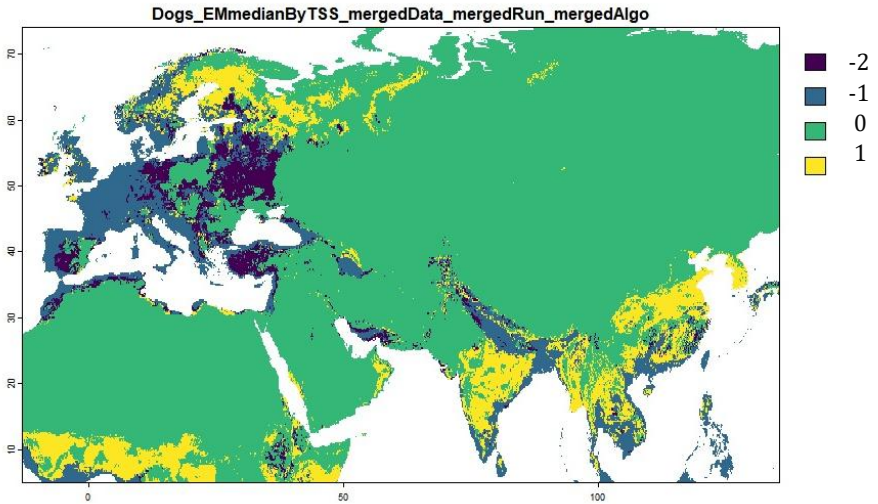

Dogs_EMmeanByTSS_mergedData_mergedRun_mergedAlgo
Dogs_EMcvByTSS_mergedData_mergedRun_mergedAlgo
Dogs_EMmedianByTSS_mergedData_mergedRun_mergedAlgo
Dogs_EMcaByTSS_mergedData_mergedRun_mergedAlgo
Dogs_EMwmeanByTSS_mergedData_mergedRun_mergedAlgo
Dogs_EMciInfByTSS_mergedData_mergedRun_mergedAlgo
Dogs_EMciSupByTSS_mergedData_mergedRun_mergedAlgo
70°N
60°N
50°N
40°N
30°N
20°N
10°N
20°W
0°
20°E
40°E
60°E
80°E
100°E
120°E
140°E
1000
750
500
250
0

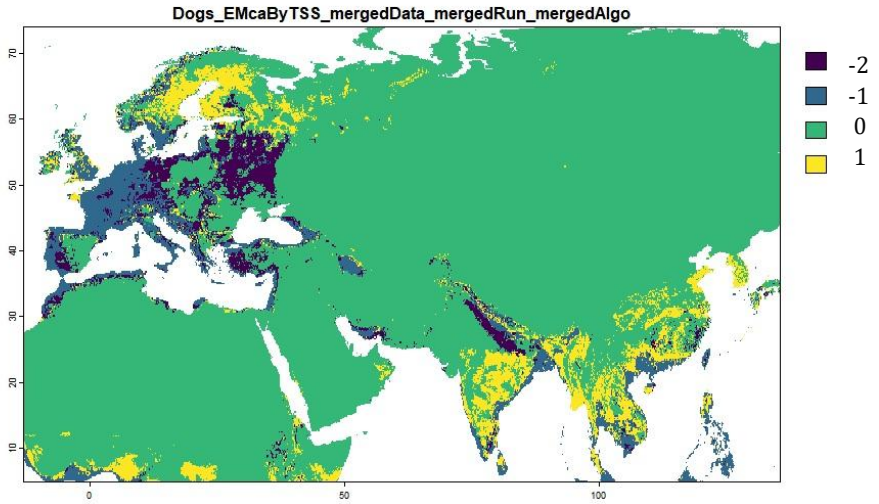
Supplementary figure 7. Suitable ecological niche modelling ensemble model with socioeconomic pathway (SSP) 585 environmental rasters. Each of the panels represents the suitable ecological niche model by committee averaging i.e., runs merged by each model algorithm (top left), with confidence intervals and inferred probabilities (top-middle), supervised methods (top-right), cross validation (middle-left), mean of probabilities (middle-middle), median probabilities (middle-right) by TSS, merged data, merged run and merged algorithms (bottom-left). Based on ensemble models and all the employed visualizations, dogs' most suitable ecological niche under the 585 SSP will be in Southern Asia, Europe and part of Northern Eurasia.


Dogs_EMmeanByTSS_mergedData_mergedRun_mergedAlgo
Dogs_EMmedianByTSS_mergedData_mergedRun_mergedAlgo
Dogs_EMcaByTSS_mergedData_mergedRun_mergedAlgo
-2
-1
0
1
10
20
30
40
50
60
70
0
50
100

Supplementary figure 8. Range difference based on the mid-Holocene and current suitable habitats models. The range difference forecasting was done by committee averaging i.e., runs merged by each model algorithm (top-left), supervised methods (top-right) and cross validation by TSS (bottom). Based on the ensemble model forecasting from the mid-Holocene and current environmental conditions, dogs' suitable habitat has expanded from the one they had ~6,000 years ago.


Dogs_EMmeanByTSS_mergedData_mergedRun_mergedAlgo
-2
-1
0
1
0
50
100
10
20
30
40
50
60
70


Dogs_EMmedianByTSS_mergedData_mergedRun_mergedAlgo
-2
-1
0
1
0
50
100
10
20
30
40
50
60
70


Dogs_EMcaByTSS_mergedData_mergedRun_mergedAlgo
-2
-1
0
1
0
50
100
10
20
30
40
50
60
70

Supplementary figure 9. Range difference based on the current and future suitable habitats models. The range difference forecasting was done by committee averaging i.e., runs merged by each model algorithm (top-left), supervised methods (top-right) and cross validation by TSS (bottom). Based on the ensemble model forecasting of current environmental conditions and future predicted environmental conditions under the 585 socioeconomic pathway, dogs are predicted to lose some of their range in Europe, however their range is predicted to increase across Northern and Southern Eurasia with a net increase of suitable habitat overall.

