

Population structure and local adaptation in red foxes across an environmental gradient

By

Blessing Sekhwela

19000296

Research Dissertation for Master of Science in Biological Sciences: Zoology

Department of Biological Sciences

Faculty of Science, Engineering and Agriculture

University of Venda

Thohoyandou, Limpopo Province, South Africa

Supervisor: Professor Yoshan Moodley

Co-supervisor: Assistant Professor Stefan Prost

Abstract

The red fox (*Vulpes vulpes*) is one of the most successful species across the globe, found on five continents and in diverse habitats across various climatic and environmental conditions. Increasing Arctic warming has allowed the red fox to expand and occupy the Arctic environment. This new environment presents unique selective pressures, characterised by long, dark winters and short, continuous light summers. Arctic animals must endure extended periods of food scarcity, extremely cold conditions, and very short periods of resource abundance. Although red foxes do not exhibit physical adaptations specialized for surviving in the Arctic conditions, they continue to thrive despite these challenges. To investigate whether there are genetic adaptations in the red fox populations in the Arctic region, I used whole genome sequencing to detect signatures of positive selection across the whole genome. To achieve this, I used population branch statistics (PBS). PBS is a three-population test that identifies loci under selection by comparing the pairwise F_{ST} between the focal group (here the Arctic population) and a sister population (here Central Europe), controlling for ancestral polymorphisms using an outgroup population (here Israel). The regions found to have unusually high F_{ST} in the focal population are potential candidates for selection. Because only low coverage genomes were available for this project, the test was conducted at two different depths (-setMinDepthInd 1 and -setMinDepthInd 3) to test whether the results were reliable with more or fewer SNPs. The two tests combined identified 123 candidate genes. Among them were genes involved in lipid metabolism, fertility, olfaction, circadian rhythm, response to hypoxia, thermogenesis and some were disease-related. I found that the Arctic population of red foxes exhibited a high genetic diversity and there was clear genetic structuring among populations. However, significant genetic contact remained between the Arctic and Central European populations. Gene-flow analysis also indicated potential signatures of ancient gene-flow between Central Europe and Israel. Despite the relatively low coverage of the available genomes, this study has identified several genes that could play a role in cold adaptation in red foxes. These genes and some of their orthologs, such as those involved in fatty acid metabolism, thermogenesis, circadian rhythm and response to hypoxia, have been identified in other arctic species like polar bears, reindeer, muskox and marmots as adaptive.

Keywords

Cold adaptation; Population branch statistic; Genetic diversity; Population structure; Positive selection; *Vulpes vulpes*; whole genome sequencing

Declaration

I, Blessing Sekhwela 19000296, declare that this research dissertation is my original work and has not been submitted for any degree at any other university or institution. The dissertation does not contain other persons' writing unless specifically acknowledged and referenced accordingly.

B. Sekhwela

23.04.2025

Acknowledgements

I would like to express my sincere gratitude to my supervisors, Prof. Yoshan Moodley and Assistant Prof. Stefan Prost, for granting me a chance to further my studies under their supervision, most notably their guidance and patience with me at every stage through this project. Sincere gratitude to Øystein Flagstad (Norwegian Institute for Nature Research, Norway) for providing Arctic samples, Dr Franz Suchentrunk, (Vetmeduni Vienna, Austria) for providing Central European samples; and Prof Eli Geffen (Tel Aviv University, Israel) for providing Israeli samples. Special thanks to Prof. Rasmus Nielsen (University of California, Berkeley, USA) for funding genome sequencing and Alexis Johnson and Lydia Smith (University of California, Berkeley, USA) for the laboratory work. Special thanks to Prof JF Masello for the scientific writing course, Dr Tiina Mattila for the help and advice with R, my lab mates for their unwavering support and advice through this project, and my sincere gratitude to my family for their support through this work. Sincere thanks to the Southern African-Nordic Centre for granting me a travel scholarship to visit my co-supervisor's laboratory, the Biodiversity Genomics laboratory, at the University of Oulu, Finland (Cost centre: O267). Lastly, I want to thank the National Research Foundation for funding this project (grant reference: PMDS22063029741).

Table of Contents

Abstract	i
Declaration.....	ii
Acknowledgements.....	iii
Introduction	1
The origin of the true foxes.....	1
The red fox (<i>Vulpes vulpes</i>)	3
Genetic diversity	3
Phylogeographic and Genetic structure.....	4
Gene-flow	6
Range Expansion and Evolution	7
Adaptation in red foxes	8
Detecting Signatures of Selection	9
Justification of study.....	11
Aims.....	12
Hypotheses.....	12
Methods.....	12
Samples.....	12
Genome Sequencing data preprocessing and quality control.....	13
Mapping raw reads.....	14
Quality checking and Filtering	15
Relatedness.....	17
Heterozygosity estimations	18
Admixture analyses.....	18
Model-based inferences.....	19
Phylogenetic Analyses	19
Gene-flow	21
Selection Analyses.....	22
Results.....	24
Sequence base quality.....	24
Mapping Quality	25
Relatedness estimations	26
<i>Population Structure</i>	29
Principal component analysis.....	29

Admixture Proportions	30
Identifying Best K-means clusters	30
Phylogenetic analysis	33
Genetic distance tree	34
UPGMA tree.....	34
Phylogenetic network.....	35
Gene-flow	36
Selection Analysis	38
Candidate genes for selection.....	40
Discussion	51
Diversity	51
Filtering	52
Population structure	53
Phylogeography of the red fox.....	54
Signatures of gene-flow.....	55
Signatures of local adaptation	56
<i>Homeostasis</i>	58
<i>Metabolism</i>	58
<i>Immunity</i>	59
<i>Olfaction</i>	59
<i>Hypoxia</i>	59
<i>Gene expression</i>	59
<i>Circadian rhythm</i>	60
<i>Fertility</i>	60
Conclusion	60
References	62
Appendix 1.....	78

Introduction

Climate has been strongly associated with species range shifts. Increasing global temperatures have resulted in more species moving towards cooler environments, namely the poleward latitudes, higher elevations, and protected microhabitats such as deeper underground, forest floors and deep waters (PattersonPecl et al., 2017; Spence and Tingley, 2020). Range expansion, or rather species range shifts, is thus defined in this context as the phenomenon whereby species colonise new regions outside their past range (Pacifi et al., 2020). Whether the climate is the only or the primary driver of species range shifts has caught the attention of many researchers. For example, Howard et al. (2023), investigated the extent of climate-driven range shifts of European birds relative to other drivers. They found that species range shifts, although primarily influenced by climate, other factors such as reduced persecution, landcover and species traits were among the drivers of this phenomena (Howard et al., 2023). Another factor worth noting is whether or not species survive and reproduce in the new environment into which they have expanded because these new spaces bring with them novel conditions not known to the new migrants (Spence and Tingley, 2020). Spence and Tingley (2020); Allendorf et al. (2022), outlined the primary responses of species colonising new ranges, beginning with phenotypic plasticity, which triggered the species to change or modify their morphology, behaviour and physiology to withstand the conditions of the new environment. Secondly, adaptation, wherein the species evolved traits that allowed them to survive through selection. Thirdly, migration, wherein species moved and tracked climatic niches geographically, and lastly, extinction, where populations became inviable and went extinct.

The origin of the true foxes

The origin of species is linked to their ability to adapt to rapid climatic and environmental fluctuations. True mammalian carnivores belong to the order Carnivora and are distinguished by their carnassial molar teeth, which are designed for shearing meat (Wang et al., 2004). Figueirido et al. (2015) found that canids, one of the oldest families within the Carnivora, evolved with climate changes, developing specialised cursorial abilities, leading to changes in their predatory behaviour as habitats were changing from forests to open areas. These results indicated climate change and its impact on vegetation and habitat structure was crucial for the emergence of ecological innovations and altered the direction of lineage evolution as observed in canids when the environment changed (Figueirido et al., 2015). Fossil evidence dates the origin of canids back to the late Eocene (Wang et al., 2004). The success of the family and all the decedents thereof are a result of multiple evolutionary radiations, which can be defined as rapid increases in the number of species, often as a result of the

availability of ecological space with less competition for resources (Wang et al., 2004; Porto et al., 2023). Canids evolved from the archaic Miacidae, which was confined to North America (Wang et al., 2004). Successive radiations gave rise to the subfamilies Hesperocyoninae, Borophaginae, and Caninae. The Hesperocyoninae and Borophaginae subfamilies, which are now extinct, remained confined in North America, and Caninae radiated into Eurasia by 7-8 million years ago through the Beringia Strait and gave rise to all modern canids in Africa and Eurasia (Wang et al., 2004; Porto et al., 2023). Caninae is the subfamily which unites all extant canids, and this is further divided into tribes Leptocyon (extinct), Canini and Vulpini (Sillero-Zubiri et al., 2004; Wang et al., 2004). The tribe Canini represents all wolf-like canids, and all fox-like canids are grouped into the Vulpini tribe (Sillero-Zubiri et al., 2004; Wang et al., 2004). The tribe Canini can be distinguished from the tribe Vulpini by its expanded paroccipital process, enlarged mastoid process, and the presence of a frontal sinus (Sillero-Zubiri et al., 2004). The fox-like canids are small canids that can be recognised by their long, low body, with relatively short legs, a long-pointed muzzle, large triangular ears, and a bushy tail at least half as long as the head and body. They also have black, triangular face marks between the eyes and nose, long black vibrissae, oval-shaped eyes with elliptical pupils, and the tip of the tail is often different from the rest of the coat (Castelló, 2018). All species in this group, which belong to the monophyletic genus *Vulpes*, are referred to as true foxes (Sillero-Zubiri et al., 2004).

The true fox clade's earliest appearance in the fossil record dates to the Late Miocene between 10-8 Ma in North America (Wang et al., 2004; Castelló, 2018; Lucenti and Madurell-Malapeira, 2020). This appearance is attributed to the extinction of the Borophaginae, another subfamily of canids distributed across North America since the late Oligocene (Wang et al., 2004). The true fox clade is distributed across Africa, Eurasia, and North America, and some were introduced in Australia (Castelló, 2018). Its members include the Arctic dwelling Arctic fox (*Vulpes lagopus*), the Blanford's fox (*Vulpes cana*), which inhabits semi-arid regions of the Middle East, the corsac fox (*Vulpes corsac*) that inhabits the arid regions of central Asia, the Cape fox (*Vulpes chama*), the Pale fox (*Vulpes pallida*) and fennec fox (*Vulpes zerda*) that inhabit the semi-arid regions of southern Africa, central Africa and North Africa respectively, the Rüppell's fox (*Vulpes rueppellii*) which is found in Africa and Asia, the Kit fox (*Vulpes macrotis*) and Swift fox (*Vulpes velox*) found in arid regions of North America, the Indian Fox (*Vulpes bengalensis*) endemic to India, the Tibetan fox (*Vulpes ferillata*) found in the plateau regions of Tibet and Ladakh, and finally the Eurasian red fox (*Vulpes vulpes*). There have been debates over whether the North American red foxes were a separate species or subspecies of the Eurasian red fox. Larivière and Pasitschniak-Arts (1996) outlined that the North American red fox was once considered a species, but they were redesignated as a subspecies

of the *V. vulpes*. On the other hand, Statham et al. (2014) suggest that North American red foxes are a distinct species (*Vulpes fulva*). This designation is not generally accepted, and the North American red fox is still considered a subspecies of *V. vulpes*. The true fox clade thus comprises 12 to 13 species distributed across the globe. The present study is focusing on the red fox (*V. vulpes*).

The red fox (*Vulpes vulpes*)

The red foxes are possibly the most successful members of the order Carnivora. Larivière and Pasitschniak-Arts (1996) showed that *V. vulpes* could comprise up to forty-five subspecies, determined by their locality and morphometric data. Castelló (2018) outlined over fifty subspecies. However, most of these subspecies have not been confirmed by genetic analysis. Statham et al. (2018), analysed nuclear microsatellites and mitochondrial (mt)DNA and supported three red fox subspecies in Europe: *V. v. vulpes*, which is distributed across Fennoscandia, covering Norway, Sweden and likely Finland and Western Russia, *V. v. silacea* is distributed across Iberia and *V. v. crucigera* distributed across the remainder of Europe. However, they did not sample *V. v. induta* distributed in Cyprus and *V. v. ichnusae* distributed in Sardinia and Corsica. Inoue et al. (2007) found evidence that the designation of *V. v. schrencki* and *V. v. japonica* was supported by mtDNA, which showed that the two subspecies belong to two distinct maternal haplogroups in Russia and Japan. The present study includes three potential subspecies: *V. v. vulpes* from Norway, *V. v. crucigera* from Central Europe and *V. v. palaestina* from Israel. However, for simplicity, I will use *V. vulpes* to refer to all samples regardless of their designated subspecies, and will refer to the populations as Arctic, Central Europe and Israel, respectively.

The red fox is found in diverse habitats, from open farmlands, tundra, deserts, wetlands, and grasslands to urban areas (Hoffmann and Sillero-Zubiri, 2016; Castelló, 2018; Kukekova et al., 2018). They also feed on a wide range of foods, from insects, berries, and rodents to medium-sized mammals; they also scavenge big game such as reindeer (Castelló, 2018) or have been found to feed on human garbage, including potatoes (Franz Suchentrunk personal communications). The red fox is listed as a species of *least concern* by the International Union for Conservation of Nature (IUCN) as of 2016, owing to its wide geographical distribution (Hoffmann and Sillero-Zubiri, 2016). The red fox appears in the fossil record of Europe during the Middle Pleistocene (0.781-0.126 mya) (Lucenti and Madurell-Malapeira, 2020; Bartolini Lucenti and Madurell-Malapeira, 2021).

Genetic diversity

Genetic diversity refers to the variation in DNA within and between distinct individuals of a species or population gene-flow (Ellegren and Galtier, 2016). It can be measured by estimating

the number of heterozygous alleles within a population, the number of different alleles present at a locus, copy number variation, haplotype diversity and nucleotide diversity. Species distribution has been associated with high genetic diversity (Miko et al., 2009). Red foxes have a broad geographic distribution, which means that different populations of red foxes will encounter different environmental conditions across their range, leading to variations in relative allele frequencies. When individuals from different areas across the red fox distribution meet and interbreed, the exchange of alleles further increases genetic diversity. Red foxes are also habitat generalists, that is, they can survive in different habitats. This way, they can maintain relatively large population sizes. Large population sizes are generally associated with high genetic diversity, while small population sizes are associated with lower genetic diversity as a result of genetic drift, bottlenecks, and fixation of deleterious alleles (Hague and Routman, 2016; Lonsinger et al., 2018). This is further supported by the neutral theory, which states that genetic diversity increases with a larger effective population size and reduces the potential for genetic drift to reduce population diversity. Studies across the red fox distribution have shown that this species exhibits high genetic diversity. Gachot-Neveu et al. (2009) found high genetic diversity in red foxes (*V. v. crucigera*) from France using random amplified polymorphic DNA markers. Magory Cohen et al. (2013) found high genetic diversity in red foxes (*V. v. palaestina*) across Israel using microsatellite markers. Galov et al. (2014) found high genetic diversity in red foxes (*V. v. crucigera*) from Croatia using mtDNA. Statham et al. (2018), also found generally high genetic diversity in red foxes across Europe.

Phylogeographic and Genetic structure

Phylogeography refers to the study of the spatial distribution of genealogies using genetic information to provide a better understanding of biodiversity, dispersal modes, diversification times, extinctions, and refugial areas of different species (López and Bonasora, 2017). The last glacial maximum, 27 -18 thousand years ago (Sommer, 2020), was a time when most of Europe was uninhabitable. In response, species shifted/retreated their ranges into parts of the continent that still had habitable conditions, such as the southern refugial regions of the Iberian Peninsula, Italy and the Balkans (Taberlet et al., 1998; Morales-Barbero et al., 2018). Species are thought to have been restricted to these southern range limits during the ice ages, from which they expanded north after the glaciers retreated (Hewitt, 1999; Aubry et al., 2009). Several genetic studies have been carried out to study the origin and the phylogeography of red foxes and, especially looking at the events that led to their current distribution. Teacher et al. (2011) and Edwards et al. (2012) found that there was no phylogeographic structuring in red foxes using mtDNA, and they concluded that this species may have existed as a single continuous population during the last glacial maximum in Europe. However, Statham et al. (2014) used mtDNA and showed that all extant foxes originated from

the Middle East. This was supported by Sacks et al. (2018), who asserted that the red fox was of Afro-Eurasian origin, specifically in the Middle East, where evidence of the most ancient mtDNA genetic lineages was found. This suggests that red foxes evolved in the arid climates of the Middle East before colonising and adapting to the temperate climates of Europe. McDevitt et al. (2022) analysed genotyping-by-sequencing (GBS) data, and found contradicting evidence that a continuous population, as previously suggested by Teacher et al. (2011) and Edwards et al. (2012), did not correspond with fossil evidence as fossil data were found in no other places except for the refugial regions. They also identified population structure across Europe, and in addition, they simulated post-glacial colonisation events which showed that red foxes in Scandinavia had ancestors from Central and Eastern Europe. This study by McDevitt et al. (2022) showed that the variation observed in the European red foxes was shaped by an evident refugial population followed by postglacial admixture and isolation when they expanded to their current distribution. Rocha et al. (2023) used whole genomic data from Europe, the Middle East and North Africa. They suggest that red foxes may have originated from Eurasia during the Middle Pleistocene and then radiated into other regions. They only made it into North Africa by 78kya. This study provided a map to uncover how these historical movements of red foxes have shaped the current diversity and population structure. Population structure is the clustering of small breeding groups or subpopulations within a population, they are the actual evolving units and have a finite size, thus prone to inbreeding. The extent of structure among different populations can be measured by F_{st} , a statistic that measures the probability that randomly selected alleles in a subpopulation are identical by descent in relation to an ancestral population (Wright, 1949). This can be crucial in studying population structure because it helps us understand genetic admixture and migration patterns. The population structure of red foxes in Europe has been extensively investigated using various genetic markers. Frati et al. (1998) used allozymes and mtDNA data from the red foxes in the Mediterranean basin. They found evidence of genetic structuring in all populations from this region with both markers. Genetic structuring was attributed to population contraction and expansion during the quaternary glaciations, especially during the last glacial period (115kya-11.7kya). Gachot-Neveu et al. (2009) analysed random amplified polymorphic DNA markers to study red fox populations from France, where they did not find any evidence of genetic structuring. This result was attributed to high levels of gene flow between the populations in this area. Mullins et al. (2014) used microsatellites and found that genetic differentiation was low among subpopulations of north-eastern Poland because of migration and high connectedness. In a survey on Fennoscandian red foxes using microsatellites, Norén et al. (2015) revealed genetic differentiation, but the subpopulations showed variable rates of gene flow and high connectedness. Statham et al. (2018) analysed mtDNA and microsatellite data from red foxes across Europe: Ireland, Britain, Spain, Italy, Serbia, France, Netherlands,

Germany, Denmark, Poland, Estonia, Norway, Sweden, and Russia, including three subspecies as previously stated. Their mtDNA data revealed that Northern red fox populations resulted from migration through Denmark across the land bridge to Sweden from Central Europe. They also identified a Scandinavian Peninsula population with an ancestral relationship with red foxes to the south and a more recent relationship with those to the east in Russia. Roberts (2019) found evidence of population structure across Europe: Norway, Sweden, Finland, European Russia (*V. v. vulpes*), England, Ireland, France, Spain, Portugal, Belgium, Germany, Switzerland, Italy, Croatia, Slovenia, Poland, Serbia, Ukraine (*V. v. crucigera*) using genotyping-by-sequencing data. They found that red foxes from Spain and Portugal were isolated from other populations (Roberts, 2019). This isolation was caused by the Pyrenees mountains (Statham et al., 2018). There was also evidence of structuring in Sweden, where Swedish foxes clustered with eastern European and Norway with Central European foxes. This phylogeographic pattern is also found in other Scandinavian mammal species, such as the Eurasian beaver (Horn et al., 2014). McDevitt et al. (2022) also found evidence of structuring in populations from distinct refugia, which they attributed to isolation after post-glacial recolonisation, and that the Scandinavian population had genetic contact with the eastern European population. This could be a recent relationship identified by (Statham et al., 2018). Magory Cohen et al. (2013) studied the structure and diversity of red foxes (*V. v. palaestina*) across Israel using microsatellites. This is the only study in this region that has used nuclear DNA to investigate red fox populations. They found high genetic variability and structure into four genetic clusters, showing evidence of division in the fox populations of this region.

Gene-flow

Gene-flow is the movement of alleles from one population to the next through migration (Allendorf et al., 2022). Gene-flow is crucial because it maintains population genetic diversity and increases the likelihood of evolutionary adaptation, whereas demographic isolation triggers population genetic structuring and may lead to speciation or extinction (Slatkin, 1985). Species movements or dispersal are influenced by a variety of factors, including inbreeding avoidance, social status, development and maturation stages, resource availability, and environmental changes. In social mammals, inbreeding was avoided by permitting juveniles who were closely related to resident females to disperse (Whiteside et al., 2011). Red fox dispersal patterns have received extensive attention, and it has been found that they can travel considerable distances (Storm et al., 1976; Allen and Sargeant, 1993; Elmhagen et al., 2017; Walton et al., 2018). They travel for varying distances ranging from zero to 300 km (Allen and Sargeant, 1993). However, they have also been found to exhibit capabilities for long-distance dispersal. Walton et al. (2018) recorded that a red fox had moved 1,036 kilometres from its

European home range. Red foxes are also not restricted by large water bodies. Storm et al. (1976) showed that this species were able to cross the Mississippi and Rum Rivers.

Range Expansion and Evolution

Red fox range expansion has been documented since the 20th century (1926–1950) in the Canadian Arctic as a response to changing climate (Gallant et al., 2014; Elmhagen et al., 2017). Rodnikova et al. (2011) reported that red foxes had first been observed in the Yamal Peninsula, Russia, in 1913 and three carcasses were collected in the same region by one of their co-authors, Viktor Shtro, between 1974–1986. Within Northern Europe range expansion of Red foxes have likely started in the 19th century (Chautan et al., 2000; Selås and Vik, 2006). These range shifts have been attributed to environmental changes due to climate warming (Gallant et al., 2020). The Arctic regions are warming four times faster than the rest of the planet (Elmhagen et al., 2017; Warret Rodrigues, 2023). Besides Arctic warming, other factors such as human-induced landscape changes and increased availability of anthropogenic subsidies such as garbage dumps or ungulate carcasses are among the factors driving red fox expansions into the Arctic (McCarty, 2001; Selås and Vik, 2006; Elmhagen et al., 2017; Gallant et al., 2020; Gomo et al., 2021; Wilkinson et al., 2024). Elmhagen et al. (2017) showed that red fox expansion into the Canadian Arctic resulted from food availability. Furthermore, high population densities of reindeer in the sub and low arctic Norway have been identified as the primary cause of the influx of mesocarnivores into the area (Sokolov et al., 2016).

Evolution is a result of cumulative changes in the genetic composition of populations (Hartl, 2020). A population is a group of individuals of the same species living within the same sufficiently restricted geographical area so that any member can potentially mate with any other of the opposite sex. It is within this population that alterations in allele frequency take place that eventually result in the development of adaptive characteristics (Hartl, 2020). The changes in allele frequencies among populations are driven by gene-flow, genetic drift, mutations, and natural selection. The interplay among these processes will ultimately increase genetic variation and enhance the survival and transfer of favourable alleles across generations. The main driver of evolution, as articulated by Darwin (1859), would favour the acquisition of traits that enable the species to survive and reproduce in their natural environment, and that is natural selection, the process that enhances the response of different species to different environments. The frequency of the favoured allele increases in the population, and neighbouring linked variation diminishes due to genetic hitchhiking, creating selective sweeps, these processes leave traces in genomes that can be detected in future (Dutheil, 2020). Studies on adaptation to high altitude, cold and hot temperatures have been undertaken for different species from different taxa, including humans (Macdonald et al., 1999; Geffen and Girard, 2003; Yi et al., 2010; Welch et al., 2014; Kumar et al., 2015; Yelmen, 2021;

Rocha et al., 2023; Zhang et al., 2023). The fluctuating climate of the Pleistocene played a crucial role in causing the species to adjust to the changing environment (Roberts, 2019). The red fox's early appearance in the fossil record during the Middle Pleistocene between 0.781 and 0.126 million years ago (Lucenti and Madurell-Malapeira, 2020) could have had a significant role in the adaptability of the red fox to varied environments and climates (Kutschera et al., 2013). Red foxes occupy an extensive geographical range, rendering them an ideal species for evolutionary genomic research (Sacks et al., 2018). During the last glacial maximum, harsher conditions caused many species, including the red fox, to retreat to areas with more habitable conditions. The red fox population was isolated in different refugial regions, and these include the Italian, the Balkan and Iberian peninsulas (Hewitt, 1999; McDevitt et al., 2022) and further North in the Carpathian Mountains, and the Dordogne region in France (Sommer and Nadachowski, 2006; McDevitt et al., 2022). These populations must have been exposed to distinct selective pressures such as resource availability and diseases, which caused the emergence of new adaptive traits we see today (Roberts, 2019).

Adaptation in red foxes

The only fox species whose adaptations have been well-documented in the Arctic ecosystem is the Arctic fox (*Vulpes lagopus*) (Hersteinsson and Macdonald, 1992; Kumar et al., 2015; Gallant et al., 2020; Lai et al., 2022). The red fox is the only fox species occurring sympatrically with the Arctic fox in parts of its circumpolar distribution range (Gallant et al., 2012; Elmhagen et al., 2017; Gomo et al., 2021) and it is threatening Arctic foxes through encroachment, pushing them further into the arctic edge (Wilkinson et al., 2024). However, the red fox has fewer physical adaptations to cold climates compared to the Arctic fox (Hersteinsson and Macdonald, 1992; Elmhagen et al., 2017; Lai et al., 2022; Warret Rodrigues, 2023). They are large compared to their small competitors, and their size means they have higher food requirements (Elmhagen et al., 2017). The red fox's large size means that they are not well built to move across the snow because their weight will cause them to sink deeper into the ice thus they require more energy to move around to forage for food (Gallant et al., 2012; Warret Rodrigues, 2023). To survive the cold winter temperatures, the Arctic foxes have 50% more underfur than the red fox, which provides better insulation and they also have a low critical temperature, the temperature below which the individual will increase metabolic heat production to maintain the average core temperature (Hersteinsson and Macdonald, 1992). Warret Rodrigues (2023) showed that red foxes survive the Arctic conditions through their high mass-adjusted basal metabolic rate, always accompanied by increased food requirements. During winter, food becomes scarce in the Arctic, making survival even more difficult for the red fox (Hersteinsson and Macdonald, 1992; Lai et al., 2022). Therefore, to survive, the red fox relies on its generalist dietary pattern, feeding on

rodents, voles, birds, flies, chipmunks, poultry, rabbits, arthropods, and fruits and scavenging on large mammals (Castelló, 2018). This shows that red foxes have had to adapt to different food resources; thus, they survive better in all habitats throughout their range. In arid habitats such as deserts (e.g. southern Israel) and the Arctic, the survival of red foxes or any species is difficult because food availability fluctuates from one season to the next, as does their diet. In such habitats, red foxes display opportunistic feeding behaviour demonstrated by the presence of fruits in their diet. They also rely on caching, storing surplus food for use in periods of food scarcity and scavenging behaviour, feeding on seal and reindeer carcasses (Lai et al., 2022; Warret Rodrigues, 2023). Some behavioural modifications, such as adopting a nomadic lifestyle, may reflect the need to satisfy relatively high energy requirements in these resource-deprived regions (Macdonald et al., 1999). Red foxes have successfully colonised different parts of the world, and as such, they were faced with different selective pressures, such as diseases. The red fox is host to many pathogens, such as rabies and sarcoptic mange that may be transmitted to humans and other animals, significant climate fluctuations of the Pleistocene, resource availability, pollution resulting from anthropogenic activities exposing them to impaired immune responses and neurological damage (Roberts, 2019). The different selective pressures faced by red foxes would have left signatures within its genome that can then be used to identify the loci subjected to selection. Red fox adaptation has been studied with different markers. Kumar et al. (2015) used transcriptome sequencing, and Roberts (2019) used genotyping-by-sequencing data to study red foxes' genetic signatures of adaptations across Europe and identified candidate genes associated with energy metabolism, olfaction, sight and development. On the other hand, Rocha et al. (2023) studied climate adaptation in red foxes from North Africa using whole-genome data. They identified candidate genes responsible for temperature perception, non-renal water loss and heat production.

Detecting Signatures of Selection

Positive selection refers to the history of a locus in which advantageous mutations have arisen and fixed or are in the process of fixing, these mutations can rapidly increase in frequency (Hahn, 2019). Patterns generated by the rapid increase in the frequency of advantageous mutation can be used to detect signatures of positive selection, this is done by searching for regions of reduced variation, divergence shifts in the site frequency spectrum and linkage disequilibrium patterns at a locus (Vitti et al., 2013; Hahn, 2019; Dutheil, 2020). ***DN/DS ratio (ω) tests*** measure divergence that is due to nonsynonymous (nucleotide changes that alter the amino acid sequence of a protein) and synonymous (that do not alter amino acid sequences) changes that are derived by counting the number of nonsynonymous differences per nonsynonymous site denoted as dN and the number of synonymous

differences per synonymous site d_S . Synonymous mutations are those mutations that do not change the amino acid sequence and are thus ignored by natural selection; nonsynonymous mutations change the amino acid sequences, and an excess of these changes indicate signatures of positive selection (Vitti et al., 2013; Bailey et al., 2021). The ratio between d_S and d_N is the ratio of nonsynonymous to synonymous substitutions ω (Kimura, 1977). Therefore, $\omega > 1$ is an indication of positive selection where the majority of nonsynonymous mutations are advantageous; $\omega < 1$ indicates purifying or background selection, where the majority of nonsynonymous mutations are deleterious and $\omega = 1$ can mean no selection and that all nonsynonymous mutations are neutral (Hahn, 2019). Although this statistic is robust, Hahn (2019) warned that d_N/d_S must be interpreted cautiously because small d_S values can bias the estimate of d_N/d_S upwards, resulting in misleading values if comparisons are made between species pairs with very different divergence times. It also assumes that all synonymous mutations are neutral; this is not always true. lastly, if there are too many advantageous synonymous substitutions, a lower-than-expected value of d_N/d_S will be observed, which suggests that purifying selection is stronger when it is not.

The McDonald–Kreitman test combines both polymorphism and divergence to test predictions of Kimura’s neutral model, assuming that these quantities should be equal for both synonymous and nonsynonymous substitutions (Kern and Hahn, 2018; Hahn, 2019). The McDonald–Kreitman test is a statistical test that compares the number of nonsynonymous substitutions to synonymous substitutions in the coding region of a locus (McDonald and Kreitman, 1991). In the case of neutral substitutions, the ratio of nonsynonymous substitutions to synonymous fixed differences between species will be the same as the ratio of nonsynonymous substitutions to synonymous polymorphisms within species (McDonald and Kreitman, 1991). The test compares the amount of variation within species divergence between two types of sites where one of the sites is neutral and is used as a reference to detect selection on the sites (Egea et al., 2008). In the presence of selection, the frequency of polymorphisms is reduced (Kern and Hahn, 2018), it is the case when the rate of nonsynonymous substitution differs significantly from the rate of synonymous substitution (Vitti et al., 2013). F_{ST} based selection tests such as the population branch statistics (PBS), identifies loci under selection by estimating the amount of divergence in the focal population relative to two outgroups (Yi et al., 2010). This test detects regions with a higher than expected genetic differentiation under neutrality. These regions will have a high population branch statistic suggesting that the target population has experienced a strong selection since it’s divergence from the outgroup (Shpak et al., 2024)

Hitchhiking and linkage disequilibrium: Positive selection can also act on linked neutral sites across the genome (Stern and Nielsen, 2019). When a newly arising advantageous

mutation increases in frequency, the linked neutral allele's frequency will also increase, and this is called genetic hitchhiking (Hahn, 2019; Stern and Nielsen, 2019). Hitchhiking causes selective sweeps where the regions where new mutations arose have reduced variability as a result of increased identity by descent and haplotype homozygosity, which decay with time (Hahn, 2019; Stern and Nielsen, 2019). Incomplete selective sweeps have uniform haplotypes due to the rapid increase in frequency, leaving no time for recombination and mutation to act on them (Stern and Nielsen, 2019). Neutral linked alleles will be inherited together with the favoured allele, and with time, these get fixed, and the allele frequencies around these regions are low due only to recent mutations occurring in the regions close to the selected alleles. When an adaptive allele has only a single origin, it is referred to as a hard sweep, and the opposite occurs when multiple copies of the advantageous allele are fixed, either because selection acts on standing variation that was previously neutral or deleterious or because mutations towards the same advantageous state occur on different backgrounds before the original copy is fixed (Hahn, 2019). Another consequence of hitchhiking is increased levels of linkage disequilibrium patterns. Linkage disequilibrium occurs when alleles that are close together on a chromosome are inherited together across the generations (Kanaka et al., 2023). In the presence of positive selection, the frequency of an advantageous allele increases rapidly, causing the linked loci to remain in unusually strong LD with that allele (Slatkin, 2008). This reduces genetic variation around the advantageous allele because the linked loci are inherited together (Stern and Nielsen, 2019; Koropoulis et al., 2020). In this study, signatures of selective sweeps were identified through sliding window analysis across the genome.

Justification of study

Red foxes have adapted to numerous environmental and climatic conditions and expanded their range to areas where they did not occur historically. Influenced by resource availability and a warming climate, red foxes have shifted their range northwards, venturing into the Arctic. The new environmental conditions pose selective pressures that might be detected on a genomic level. The present study used whole genome data to detect whether the red fox had evolved beneficial mutations that allowed it to be suitably adapted to the harsh conditions of the Arctic. This study used data from red fox populations sampled at three different latitudinal zones in a north-south transect from Arctic Norway through Central Europe to the Israeli desert. The study focused mainly on regions of the genome that are under selection for advantageous adaptations. While genome-wide studies of the red fox, analysing the signatures of adaptations to different environments, have been performed using transcriptome sequencing, GBS and whole genome data, respectively (Kumar et al., 2015; Roberts, 2019; Rocha et al., 2023). No study has been carried out that addressed the question of whether red foxes showed signatures of adaptation to life in the Arctic ecosystem.

Aims

This study aims to analyse the red fox whole genome sequence data from three populations in a north-south transect through Europe to detect signatures of adaptation to the Arctic environmental conditions.

Hypotheses

Hypothesis 1: In light of previous studies which showed connectedness between Scandinavia and Central and Eastern Europe, I expect the Arctic population to have high genetic diversity. However, they also went through a bottleneck, which could result in low genetic diversity. The Middle East is considered the cradle of modern red fox genetic diversity. Therefore, I would also expect Israel to have high genetic diversity since it has an older population.

Hypothesis 2: Similar to previously described phylogeography studies on red foxes, significant genetic structure could be expected between the three sampled populations. However, given the recent expansion of red foxes into the Arctic, weaker structuring is expected between Arctic and Central Europe, compared to Israel.

Hypothesis 3: Due to the red fox's long-range dispersal capabilities, I hypothesise that populations show evidence of gene-flow.

Hypothesis 4: Given the harsh Arctic environment into which the red fox is dispersing, I hypothesize to be able to detect genomic signatures of adaptations in genes associated with cold environmental conditions.

Methods

Samples

This study was conducted using unpublished red fox genomic data generated by Dr Stefan Prost (co-supervisor to the present study) and Rasmus Nielsen (University of California, Berkeley, USA). The ninety-three red fox samples provided represent three populations: the Arctic (provided by Øystein Flagstad, Norwegian Institute for Nature Research, Norway), Central Europe (provided by Franz Suchentrunk, Vetmeduni Vienna, Austria), and Israel (Eli Geffen, Tel Aviv University, Israel). These samples make a roughly geographical transect from Arctic Norway through Central Europe to the Judean and Negev deserts of Israel. To reduce sampling bias, the project was designed so that each population had at least 20 samples or more; the Arctic, $n=37$; Central Europe, $n=36$; and Israel, $n=20$. In this study, the Arctic was used to define or represent all samples from the northern-most parts of mainland Norway, while Central Europe includes Austria, Bulgaria, Hungary, Turkey, Bosnia,

the Czech Republic, Germany, Poland, Spain and Ukraine and including one captive individual, and Israel represents samples from the Judean and Negev deserts. Figure 1 represents the global distribution of red foxes, while Figure 2 represents the extent of the sampling distribution of the samples used in the present study.

Redfox global distribution

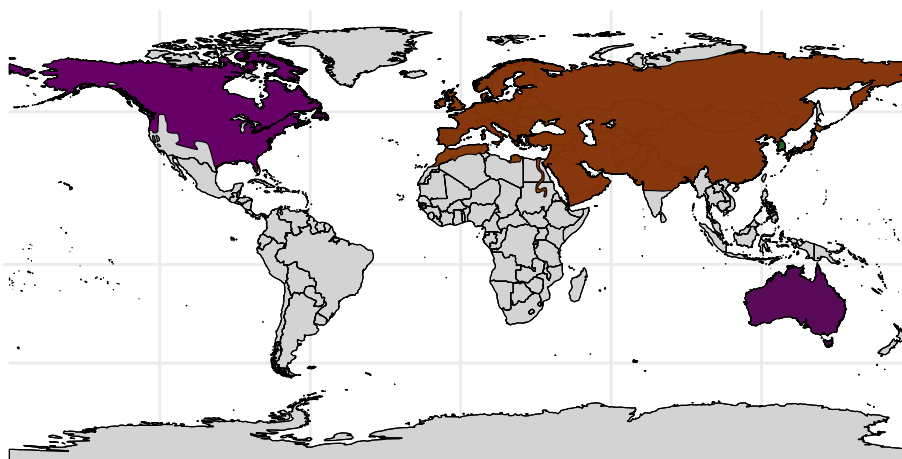


Figure 1 The sampling distribution of red foxes in the present study and the global distribution from IUCN. Australian and North American populations are given as extant and introduced (purple), Eurasia as extant and South Korea as extant and vagrant (green) (<https://www.iucnredlist.org/search/map?query=vulpes%20vulpes&searchType=species>).

DNA extraction and library preparation

DNA was extracted from the tissue samples using the Qiagen Blood and Tissue kit (Qiagen, NL) following the manufacturer's recommendations. The extracted genomic DNA was then sonicated with the qSonica Q800R, and libraries were prepared using the Kapa Hyper Prep kit with dual indexing (Roche sequencing). Sequencing was performed on an Illumina HiSeq X platform. The laboratory work was carried out by Alexis Johnson and Lydia Smith at the University of California Berkeley, USA.

Genome Sequencing data preprocessing and quality control

Short-read next-generation sequencing (also referred to as second-generation sequencing) refers to the next advancement of sequencing technologies after traditional first-generation Sanger sequencing (Sanger et al., 1977; Margulies et al., 2005). These sequencing technologies involve massively parallel sequencing of short reads, where millions of individual sequencing reactions occur in parallel (Margulies et al., 2005). These technologies require that sequences are amplified during the sequencing, which can cause GC-bias, sow problems with adapter contamination, overrepresented sequences, and errors in library preparation (Hu

et al., 2021; Hao et al., 2022). As a result, sequencing quality control and preprocessing are required before any downstream analysis can begin, to ensure the cleanness and accuracy of the data (He et al., 2020). Multiple versatile programs have been designed to carry out sequencing quality analysis. FASTQC, generates multiple statistics from per-base-sequence-quality, per-sequence-quality-scores, per-base-sequence-content, per-sequence-GC-content, sequence-duplication-levels, over-represented-sequences, to adapter content (Andrews, 2010). FASTQC reads a set of sequence files and produces a quality control report consisting of several modules, each of which will help identify a different potential type of problem in the data. These statistics can be summarised in MULTIQC, which creates visual graphics from the FASTQC results (Ewels et al., 2016). FASTQC can detect problems arising both in the sequencer and the library material. The raw reads in this study were run through FASTQC to generate the sequence quality statistics and trimmed as part of the Paleomix pipeline using default settings (see section below).

Mapping raw reads

Second-generation sequencing produces huge amounts of data, and analysing them one at a time is time-consuming and sometimes runs on multiple software that requires different environments, parameters, and formats of the input data (Leipzig, 2017; Park et al., 2024). Different tools and pipelines have been developed that can handle analyses of second-generation sequencing data, one such pipeline is the Paleomix, a set of pipelines and tools designed to rapidly process high-throughput sequencing data from modern to ancient samples (Schubert et al., 2014). Paleomix has four pipelines: bam, phylogenetic, zonkey pipelines and a metagenomic protocol (Schubert et al., 2014). Sequence alignment in this study was done through the Paleomix bam pipeline; this is the primary pipeline that runs sequence alignment against a reference genome, as well as read trimming, filtering, and format conversion (Schubert et al., 2014). As stated above, sequence alignment was done through Paleomix; this pipeline was designed to work with modern and ancient DNA samples. These samples are, in most cases, very low coverage as a result of postmortem DNA degradation and microbial colonisation (Sousa da Mota et al., 2023). Although the data in this study are modern, they are low coverage, making Paleomix the most suitable pipeline for this dataset. Before short-read data can be mapped, all the issues flagged by FASTQC need to be addressed, which means trimming for adapter sequences and low-quality and ambiguous bases. Paleomix does this using adapter removal (Schubert et al., 2016).

The initial step for read mapping is to get a reference sequence, which is an existing genomic assembly that is representative of a species and is used to align short-read sequences (Manel et al., 2016). The scaffold-level red fox reference sequence was downloaded from GenBank Project PRJNA378561 as a compressed file (Kukekova et al.,

2018). The command line `tar xf` decompresses the compressed reference sequence file into the working directory, and the reference was named `GCF_003160815.1_VulVul2.2_genomic.fna`. The reference is indexed before it can be useful, indexing is done to achieve fast random retrieval of alignments overlapping a specified chromosomal region (Li et al., 2009b). Indexing was carried out using the freely available software SAMTOOLS (Li et al., 2009b) using the command *faidx*.

Processed reads resulting from the adapter-trimming step above were mapped using a Burrow-Wheeler short-read aligner, BWA-MEM (Li, 2013) as part of the Paleomix pipeline. BWA-MEM is a fast, high-performing and accurate sequencing alignment program which efficiently aligns short sequences up to several megabases against a large reference genome (Li, 2013). The resulting BAMs are tagged using the information specified in the makefile (sample, library or lane). BAM files are used as input files in many downstream programs in bioinformatics and population and phylogenomics. The records of the BAM resulting from mapping are updated using SAMTOOLS (Li et al., 2009b) *fixmate* to ensure that paired-end reads contain the correct information about the mate read, they are then sorted using SAMTOOLS *sort*, indexed using SAMTOOLS *index*, and validated using PICARD *ValidateSamFile* (PicardToolkit, 2019). Finally, the records were updated using SAMTOOLS *calmd* to ensure consistent reporting of the number of mismatches relative to the reference genome (BAM tag 'NM') (Li et al., 2009a). The following step after updating BAM records is filtering for duplicates, this was done using Picard *filter*, to remove the duplicates, this is followed by an indexing step again and PICARD *ValidateSamFile* from the PicardToolkit (2019) was used to validate the resulting bam files after which the mapped reads were compared using the query name, sequence and qualities to check for the presence of data that has been included multiple times, if any match was found it was assumed to represent a duplication in the input data. The final step is the generation of the final bam files, which are then indexed and can be used for downstream analyses.

Quality checking and Filtering

The depth of sequenced reads is crucial for genomic analyses, and especially when calculating the genotype likelihood of low-coverage genomes (Pedersen and Quinlan, 2018). Depth also gives an indication of how conservative or relaxed downstream filters should be to get the most out of the data without compromising the reliability of the inferences. Mapping quality describes the probability that the sequence reads are aligned to the correct position in the reference genome (Lou et al., 2021). In contrast, the base quality indicates the probability that a given base in a sequence has been called correctly (Lou et al., 2021). Several software packages are used to compute quality scores of the sequence reads. These include SAMTOOLS depth (Li et al., 2009a), BEDTOOLS (Quinlan and Hall, 2010), Analysis of Next-

generation sequencing data (Korneliussen et al., 2014), QUALIMAP (Okonechnikov et al., 2015) and MOSDEPTH (Pedersen and Quinlan, 2018). The quality of mapped files in this study was calculated using qualimap multiple bamqc; this option allows for streamlined multiple-sample analyses.

Filtering is a crucial step in genomic data analysis. Datasets often contain errors and have regions with missing data, and making inferences on these data would produce unreliable and inaccurate results (Allendorf et al., 2022; Hemstrom et al., 2024). Hemstrom et al. (2024), posits that to improve the quality of downstream analysis, some sequencing bases, reads, genetic variants, and individuals have to be removed from the data. In low-coverage whole genome sequencing, where the number of reads covering most of the genomic loci is already low, the choice of filter and thresholds applied will determine whether the data that is left will be enough to make inferences at all. Allendorf et al. (2022) suggested that insufficient data filtering can fail to remove errors or add errors to a dataset, however, filtering can be difficult because filters must be stringent to identify and genotype high-quality loci from second-generation sequencing data. On the other hand, filtering too stringently will remove high-quality loci or loci under selection that are of interest in studying adaptation and fitness. Allendorf et al. (2022) gave an example of minor allele frequency, citing that stringent MAF filters could skew metrics based on the site frequency spectrum, potentially filtering out useful loci, including those under selection. Therefore, to ensure that we kept as many useful loci as possible, minor allele frequencies were estimated based on the size of the dataset, taking into account the diploid nature of data and the amount of missing data allowed. In addition, carrying out tests which rely on the correct estimation of allele frequencies, the sample sizes need to be relatively large (Simonsen et al., 1995). This is required to increase the increase in the statistical power of the said tests (Simonsen et al., 1995). The data in this present study contains at least 93 samples, with equal to or greater than 20 samples per population, this means data are likely to give correct estimates of the population parameters studied. Simonsen et al. (1995) also showed that a sample size of 20 was enough to make statistically powerful inferences. Given that the data used in this study is low coverage, all analyses were run first through the analysis of next-generation sequencing data software (Korneliussen et al., 2014). ANGSD is an efficient software that is robust for low and medium-coverage genome analysis, it allows for multiple error models when calculating genotype likelihoods (Korneliussen et al., 2014). This offers a statistical advantage because genotype likelihoods account for statistical uncertainty that arise from using the low-coverage data (Korneliussen et al., 2014). These could be a result of sequencing errors or random sampling of haploid reads from a diploid genotype (Korneliussen et al., 2014). The following options and filters were applied to the data in ANGSD: -GL 1 -domajminor 1 -snp_pval 1e-6 -domaf 2 -minmaf

0.007 -doGlf 3 -minQ 20 -minMapQ 20 -minInd 73 -doCounts 1 -only_proper_pairs 0 -baq 1 -setMinDepthInd 1 and -setMinDepthInd 3. GL 1, tells the program to estimate genotype likelihoods using an algorithm designed for Samtools ; domajorminor 1 invokes the estimation of major and minor alleles from genotype likelihoods; -snp_pval 1e-6 directs the program which sites to use and in this case means to only work with polymorphic sites with a p-value less than 1e-6; -domaf 2, calculates allele frequency using a fixed allele and an unknown minor allele; -minmaf 0.007, minor allele frequency calculated as $\frac{1}{(73 \times 2)}$, minmaf is estimated from the minimum number of individuals 73, for analysis which uses the whole dataset which speaks to the amount of missing data allowed in this case only a maximum of 20% missing data is allowed because the whole data set has 91 samples. The sample size was reduced from 93 samples to 91 because two samples, A-3394_S1 and I-10-78_S44, had to be excluded because of suspected mislabelling in the lab during sequencing. The value of minmaf changes after relatedness estimations and even more so during selection analysis, where the site frequency spectrum is estimated for each population independently, minInd 73 then tells the program to keep a site if it is found in 80% or more of the individuals in the dataset; -doGlf 3, output format, this outputs a glf file which is the format required by NGSRELATE; -minQ 20, discard bases with a base quality score below 20; minMapQ 20, discard sites with minimum mapping quality below 20; -doCounts 1, do allele counts required for setMinDepthInd to work; -only_proper_pairs 0, means consider regions where only one of the pair is properly mapped; -baq 1, adjusts quality scores around indels, it reduces false positives caused by misalignments; -setMinDepthInd 1, discard an individual if the sequencing depth is below 1 and -setMinDepthInd 3 tells the programs to discard individual if the sequencing depth is below 3. This last filter was set to both 1 and 3, to test for reliability against the number of single nucleotide polymorphisms (SNPs) used for analysis.

Relatedness

Relatedness estimations are helpful in the conservation of both wild and captive populations. They allow us to understand animal social systems, detect avoidance of inbreeding, and optimise conservation breeding and translocation strategies (Allendorf et al., 2022). Genetic relationships between individuals are determined by their degree of genetic similarity, individuals with more shared alleles that are identical by descent share the most recent common ancestor and have a closer relationship (Hanghøj et al., 2019). All individuals in a population are related to some extent; therefore, to avoid skewing downstream analysis, individuals with a sibling and parent-offspring relationship need to be excluded (Coleman et al., 2016; Hahn, 2019). The probability that two individuals inherited a gene or a certain proportion of genes from a common ancestor is called the coefficient of relatedness (r) (Wright, 1922; Widdig, 2017). Different programs have been designed to estimate the coefficient of

relatedness, and they include COANCESTRY (Wang, 2011), NGSREMIK (Nøhr et al., 2021), Relateadmik (Moltke and Albrechtsen, 2014) and NGSRELATE (Hanghøj et al., 2019). The present study utilised NGSRELATE to estimate the degree of relatedness, beginning with calculating genotype likelihoods with ANGSD. Relatedness was estimated through NGSRELATE (Hanghøj et al., 2019). Finally, we used awk to extract all closely related individuals (Aho et al., 1979).

Heterozygosity estimations

Heterozygosity refers to the proportion of individuals that are heterozygous (Allendorf et al., 2022). Average heterozygosity is typically calculated across many loci in a population to estimate genetic variation (Allendorf et al., 2022). Using ANGSD, heterozygosity was computed as the frequency of heterozygous loci of each individual by first calculating the distribution of alleles herein referred to as the site frequency spectrum. The following step was to estimate heterozygosity. The output has two numbers. The first represented the homozygous positions, while the second represented the heterozygous positions. An individual's heterozygosity estimate was calculated in R by dividing the second value in vector *a*, that is, the number of heterozygous positions in the *A_08_06_S58.rmdup.normal.ml* file, by the total number of positions.

Admixture analyses

Populations become structured into genetically distinct populations because of restricted gene-flow as a result of e.g. geographical barriers, distance or reproductive isolation, genetic drift acting on founder populations, and local adaptation (Hartl and Jones, 1998; Allendorf et al., 2022). It is inferred in two different ways: model-based and non-model-based.

Non-model-based inference: Principal component analysis (PCA)

Elhaik (2022) defined the Principal component analysis as a multivariate analysis that reduces the data's dimensionality while preserving their covariance. It reduces dimensionality by constructing a covariance matrix of the allele frequencies after finding the eigenvalues and eigenvectors and then transforms this data into a small number of dimensions termed principal components, where each principal component describes a decreased proportion of the genomic variation (Elhaik, 2022). Principal component analysis can be used to find hidden structure within a sample without defined assumptions or prior knowledge, resulting in the identification of clusters of individuals belonging to a particular group (Hahn, 2019; Bourgeois and Warren, 2021). PCA analysis was carried out using PCANGSD, (Meisner and Albrechtsen (2018) a method of inferring population structure through PCA in an iterative heuristic approach of estimating individual allele frequencies. The estimated principal components were visualised in R.

Model-based inferences

As previously introduced, red fox populations across Europe are structured, possibly because of restricted gene-flow, genetic drift and local adaptation. However, when these population ranges come into contact new immigrants could exchange genes and this will result in the production of admixed offspring (Bertorelle and Excoffier, 1998) whose genetic makeup represents a mixture of alleles from different ancestral populations (Skotte et al., 2013). Model-based clustering methods for population structure inference have been developed over the years. For example, Structure (Pritchard et al., 2000) infers population structure and assigns individuals to populations using Bayesian inferences. There are also maximum likelihood-based methods such as Admixture (Alexander et al., 2009) and NGSADMIX (Skotte et al., 2013). These methods assume that individuals originated from K clusters. In this study population admixture proportions were estimated using NGSADMIX (Skotte et al., 2013), a computationally efficient tool for low and medium-coverage data based on genotype likelihoods. It is a multithreaded c++ program that makes it more efficient for analysing large datasets (Skotte et al., 2013). NGSADMIX is suitable for this study because it considers the uncertainty in NGS sequencing data when deducing an individual's ancestry (Skotte et al., 2013). ANGSD (Korneliussen et al., 2014) was used to calculate genotype likelihoods and output in beagle format. To analyse the genotype likelihoods, NGSADMIX assumed a different number of ancestral populations (K) from $K=2$ to $K=6$, all of which were repeated for 50 independent runs. The results of the estimated admixture proportions for all individuals were first zipped and uploaded onto CLUMPAK (Cluster Markov Packager Across K) (Kopelman et al., 2015) for postprocessing, after which the results were downloaded, and INKSCAPE (InkscapeProject, 2020) was used to clean the plots for better visualisation. The Evanno method Evanno et al. (2005) implemented in CLUMPAK was used to get the best K clusters. This method uses a log probability text file as input. This file contains two columns wherein the first represents the K cluster, and the second represents the best likelihood inferred by NGSADMIX for each replicate. The Evanno method outputs a graph of delta K versus K , assuming the K with the biggest change in delta K is the best K .

Phylogenetic Analyses

Unlike PCA and admixture analyses, phylogenetic analysis enabled us to decipher which groups or individuals are most closely or distantly related to each other. Given extensive sequencing data and the advanced computational ability of programs being developed, we can reconstruct and make meaningful and more accurate phylogenies for many different species using whole genomes. Phylogenies can be reconstructed from simple distance methods such as Neighbour Joining (Saitou and Nei, 1987) or an unweighted pair group with arithmetic mean (UPGMA) (Sokal and Michener, 1958) which requires the calculation of a

distance matrix. These methods are incorporated in PHYLIP (Felsenstein, 1993) and NGSDIST (Vieira et al., 2016). that the latter estimates pairwise genetic distances directly, taking the uncertainty of genotype's assignment into account by using genotype likelihoods or posterior probabilities instead of direct genotype calling (Vieira et al., 2016). Phylogenetic trees can also be constructed using more complex methods, such as those that use maximum likelihoods (e.g. IQTREE (Nguyen et al., 2015) and Bayesian inference such as PHYCAS (Lewis et al., 2015) and BEAST v1.7 (Drummond et al., 2012). However, these usually require reliably called genotypes generated from high-coverage data. NGSDIST uses a Bayesian approach to estimate pairwise genetic distances (Vieira et al., 2016). This approach assumes an infinite sites model, which does not account for recurrent mutations (Vieira et al., 2016). This means that it assumes that each mutation occurs at a unique site which simplifies that evolutionary model. To get the distances, NGSDIST takes the average of all the posterior probability distributions of possible genotypes instead of using fixed genotype calls (Vieira et al., 2016). Branch support values were generated through bootstrapping within NGSDIST (100 replicates) for the distance matrix. Similar to the tests discussed thus far, NGSDIST also requires ANGSD for the initial step to compute the posterior probabilities using the option doGeno 8, which tells ANGSD to compute the posterior probabilities of all possible genotypes. This is followed by the computation of pairwise genetic distances in NGSDIST, after which FASTME(Lefort et al., 2015) was used to search for all possible tree topologies, and RAXML-NG (Kozlov et al., 2019) was then used to place bootstrap on the main tree estimated from the previous step. The bootstrap values are used to place confidence intervals on phylogenies(Felsenstein, 1985).

It is important to note that phylogeny finds a tree that best explains the observed patterns in the data with the fewest assumptions (Huson and Bryant, 2005; Gross et al., 2021). Therefore, phylogenetic trees perform poorly in complex evolutionary scenarios where hybridisation, recombination, horizontal gene transfer, gene loss and duplication exist (Huson and Bryant, 2005; Young and Gillung, 2020; Gross et al., 2021). One solution to deciphering these complex scenarios is to use phylogenetic networks (Huson and Bryant, 2005); graphs used to depict reticulate evolutionary relationships among sequences, genes, chromosomes, or genomes (Allendorf et al., 2022). Huson and Bryant (2005) describe several types of phylogenetic networks: split networks derived from character sequences. These are used to represent incompatible and ambiguous signals in a data set. Reticulate networks are networks that occur when two evolutionary lineages merge partially to create a new lineage, and they are derived from recombination and hybridisation events. Median networks are generated from distance matrices using split decomposition and neighbour net methods. They can also be derived from phylogenetic trees, i.e. consensus and super networks. Lastly, haplotype

networks are designed to depict haplotype relationships. Huson and Bryant (2005) showed that split networks represented incompatible and ambiguous signals in large datasets where phylogenetic tree bootstrap resampling was ineffective. The replicates were similar no matter how poorly the data fit the inferred tree; therefore, the phylogenetic network took care of this systemic error by allowing additional splits with weights, citing that the extra parameters allowed split networks to fit the data better than individual trees. In this study, a phylogenetic network and an Unweighted Pair Group with an Arithmetic Mean tree were computed in SPLITSTREE (Huson and Bryant, 2024), an interactive program for analysing and visualising evolutionary data, which supports several distance transformations, the computation of parsimony splits, spectral analysis and bootstrapping using genotype calls with doGeno 2 in ANGSD. The UPGMA tree was used here to determine the broader grouping of the populations. And so, after genotype computation, ADEGENET (Jombart and Collins, 2015) was used to convert the genotypes into a distance matrix and the matrix was then used to reconstruct the different types of phylogenetic networks and distance-based phylogenetic trees such as the UPGMA tree in SPLITSTREE.

Gene-flow

Gene-flow may either restrict evolution by preventing adaptation to local conditions or contribute to evolution by spreading new genes and combinations of genes throughout a species' range (Allendorf et al., 2022). Patterson's D or ABBA-BABA statistics and the F_4 ratios are among the widely used tests for introgression. The ABBA-BABA statistic is based on the di-allelic patterns of alleles between four individuals or groups of individuals. It provides a way to test the correctness of a hypothetical genetic relationship in a quartet where H1 and H2 positions on the tree represent the ingroup, that is they are sister taxa that share a derived state 'B'; the third taxon H3 represents a more distantly related ingroup taxon, but with the ancestral state 'A' as defined by the outgroup (O) (Soraggi et al., 2018). Gene-flow is therefore estimated by counting sites in the genome that match ABBA and BABA genotype patterns, where ABBA means an ancestral allele in individuals/populations H1 and O and a derived allele for the individuals/populations in positions H2 and H3, and BABA means that individuals sharing the derived allele are in positions H1 and H3 and the ancestral allele shared by positions H2 and O (Soraggi et al., 2018). F_4 -ratio estimation allows inference of the amount of gene-flow during an admixture event even without access to accurate surrogates of the ancestral populations. Gene-flow was estimated using the ANGSD multi-pop ABBA-BABA test, which, instead of applying a single base from the reads of one individual per population, considers all reads from multiple individuals in each population (Soraggi et al., 2018). It was estimated in four steps, beginning with preparing a reference file for error rate estimation. This is the sample with the highest coverage in the dataset. This is followed by creating an error

estimation bamlist which is followed by the actual error estimation and then ABBA-BABA test in ANGSD.

, and lastly R was used create tables for visualisation. The filter setMinDepthInd 1 was chosen in this test to leverage the high number of SNPs, as this, in addition to the large sample size, would increase the chances of getting reliable allele frequencies from the data as well as increasing sensitivity to detect gene-flow even if it weak, as this would be missed when using a lower number of SNPs.

Selection Analyses

Different selective pressures such as diseases, parasites, pollution, competition, predation and climate change all leave signatures in the genome that are indicative of positive selection. Several programs have been designed to detect recent selective sweeps in populations. They include SweeD (Pavlidis et al., 2013), a program designed specifically for whole genomic data, it is a likelihood-based method which allows users to specify demographic models when analysing the distribution of the expected polymorphic sites, the site frequency spectrum. SweepFinder2 (DeGiorgio et al., 2016) is another likelihood-based program which performs a composite likelihood ratio test to detect recent positive selection. This program allows the user to specify test sites and set the distance between test sites (DeGiorgio et al., 2016). Another program designed for the same test is RAiSD (Alachiotis and Pavlidis, 2018), a program that uses a parameter-free detection mechanism for multiple signatures of selective sweeps through the enumeration of SNP vectors. In this study, we used Population branch statistics (PBS) (Yi et al., 2010), implemented in ANGSD (Korneliussen et al., 2014). Unlike previously mentioned tests, the PBS is a three-population test that identifies loci under selection by comparing the pairwise F_{ST} between a first target population that is sister to the second population, with ancestral polymorphisms controlled for by an outgroup population. The population branch statistic is calculated based on the estimation of the branch length of the target population as follows: the pairwise distance between each pair of populations is calculated as a log-transformed function of F_{ST} $T = -\log(1 - F_{ST})$ and to estimate the branch length of the target population, in this case, the Arctic population, we add the pairwise distances between the target population and its sister population here Central Europe, and the target population and the outgroup population then, we subtract the pairwise distances between the sister population and the outgroup population and divide everything by two (Foote et al., 2016; Shpak et al., 2024). In the present study this formula will be as follows:

$$PBS_{Arctic} = \frac{T(Arctic\ CentralEurope) + T(Arctic\ Israel) - T(CentralEurope\ Israel)}{2}$$

This test searches for a population in the triplet that has the fastest allele frequency changes compared to the other two populations. It does this by searching for regions with unusually high F_{st} values across the genome, which means regions with substantial derived allele frequency differences between populations would contain genes that are potential candidates for selection (Yi et al., 2010). This method has also been found to be robust in detecting selection in admixed populations (Yelmen et al., 2021) as well as in high-altitude adaptation in livestock (Wiener et al., 2021). Our data is particularly suitable for this test because the samples were strategically sampled to cover three different climatic and environmental conditions. The population branch statistic was calculated using ANGSD (Korneliussen et al., 2014). This is a four-step method: Step 1 involves an estimation of the site allele frequency spectrum for each population separately calculated using ANGSD -doSaf . The outputs for doSaf are three files saf.gz, saf.pos.gz and saf.idx. In step 2, the saf.idx files are used to get a folded pairwise two-dimensional site frequency spectrum for each pair , folded minor allele frequency spectrum, describes the frequency distribution of the minor alleles (Hemstrom et al., 2024). In step 3, the pairwise F_{st} for all populations, also known as the global estimate, . Lastly, step 4 uses the global estimate to create a sliding window. The sliding window divides the genome into 50 kb windows and the population branch statistic is calculated for each window using the SNPs in that window. The window slides across the genome by a step size of 10 kb, that is the distance from one window to the next and recalculates the PBS for each window. All windows with high PBS scores suggest potential signals for selection. R was then used to plot windows with high PBS using the qqman library (Turner, 2014). To deal with issues arising from multiple tests, the distribution of the p-values was plotted. The Bonferroni correction method (Weisstein, 2004) requires a uniform distribution of the p-values, while the false discovery rate (Benjamini and Hochberg, 1995) does not. The former method also assumes a false positive in the data, which can be too stringent, while the latter carries out multiple tests for the data (Whitlock and Schluter, 2015; Hahn, 2019). As a result, the false discovery rate was chosen because it does not require a normal distribution, and it is not too strict.

BEDTOOLS, Quinlan and Hall (2010), was used to query for genes with adjusted significant p-values flagged in the sliding window from the genomic annotation files. BEDTOOLS *closest* is a program that searches for overlapping features in A and B (Quinlan and Hall, 2010). To run the BEDTOOLS *closest*, the input files must first be sorted by scaffold, and then by start position with . The sort command is part of the GNU core utilities in Linux (MacKenzie, 1994). To get the candidate genes, three tests were done in BEDTOOLS *closest* to extract the default overlapping features in the query file and the gff file; the features or genes downstream in the gff file were extracted using the -iu option and lastly, features/ genes that

were upstream using -id option. Additional options were added to account for features with equal distances. Options -t all and -t first were used; t all tells BEDTOOLS closet to report all features when the features with equal distance and first tells the program to report the first tie. Finally, to link genes to their functions and/or their respective biological pathways, functional analyses were performed on the identified genes on the DAVID web server (Huang et al., 2009; Sherman et al., 2022), NCBI GENE (Brown et al., 2015) and GENECARDS (Safran et al., 2010). This was done using *Vulpes vulpes* and *Canis lupus familiaris* as a reference where *Vulpes vulpes* was not found in the list of published genes.

Results

Sequence base quality

The sequence quality plot (Figure 2) shows that almost all sequences in the data are good quality, with the majority having a Phred score between 30 and 40; this shows that bases can be called with confidence from this dataset, as the probability of error is ≤ 0.001 . The following sequence short read files have a Phred score of <30 , I_13891_S87_R2 and I_14789_S88_R1 and R2.

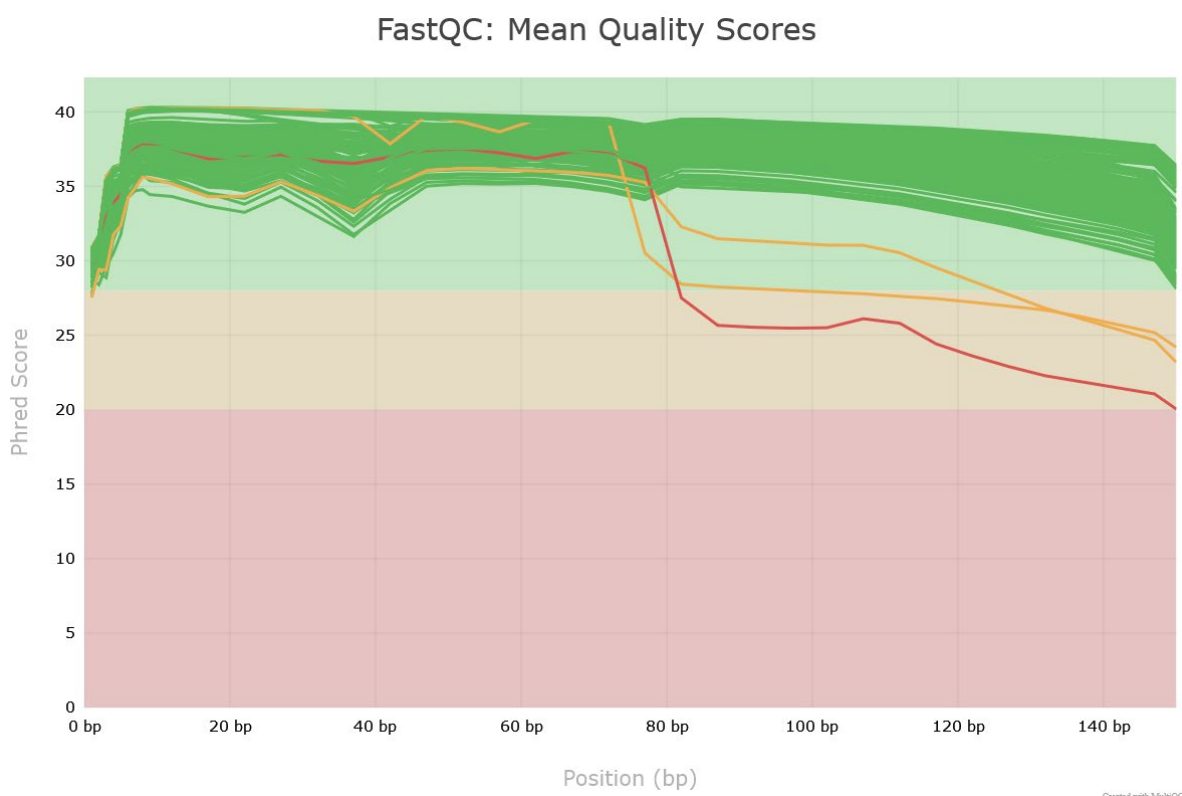


Figure 2 Depicts the number of reads with average quality scores across each base position.

Mapping Quality

Table 1 shows the mapping quality of the mapped sequences in the dataset. The highest coverage genome belongs to a captive sample from Central Europe, while the rest are wild samples. Within the wild populations, the Arctic had samples with coverage ranging from 0.1788X to 0.8626X, and the average for the population was 0.3541X; Central Europe had coverage ranging from 0.0921X to 0.7286X with an average of 0.2830X while Israel had an average of 0.4512X with coverage ranging from 0.0135X to 0.667X and one individual (I_13886_S78) in this population had coverage of 3X.

Table 1 Sample location and sample depth or coverage of the samples analysed in the present study.

Genome Coverage								
Central Europe			Norway(Arctic)		Israel		Outgroup	
Hungary	CE_10_S83	0.3937	A_08_06_S58	0.6216	I_11590_S74	0.3777	Arctic_Fox	18.8156
Hungary	CE_11_S15	0.3102	A_08_100_S37	0.2202	I_11594_S60	0.4786		
Germany	CE_12_S16	0.2466	A_08_107_S57	0.2639	I_11595_S25	0.1442		
Austria	CE_13_S17	0.1948	A_08_108_S10	0.3309	I_12653_S68	0.3239		
Austria	CE_14_S28	0.2577	A_08_113_S9	0.2622	I_12691_S75	0.2324		
Austria	CE_17_S70	0.247	A_08_114_S80	0.2229	I_12696_S61	0.2784		
Austria	CE_18_S29	0.2324	A_08_117_S56	0.5145	I_13407_S76	0.303		
Austria	CE_19_S30	0.3184	A_08_118_S36	0.2721	I_13588_S77	0.4387		
Bulgaria	CE_1_S11	0.4234	A_08_136_S55	0.2887	I_13602_S69	0.2872		
Austria	CE_20_S31	0.1834	A_08_153_S4	0.2086	I_13619_S86	0.2885		
Austria	CE_21_S18	0.1933	A_08_154_S62	0.4051	I_13763_S46	0.667		
Austria	CE_22_S39	0.371	A_08_160_S54	0.5787	I_13884_S35	0.2744		
Austria	CE_23_S71	0.3722	A_08_169_S63	0.2999	I_13886_S78	3		
Austria	CE_24_S32	0.1469	A_08_203_S59	0.3591	I_13891_S87	0.0326		
Austria	CE_25_S40	0.4943	A_08_209_S53	0.416	I_13958_S20	0.3089		
Austria	CE_26_S33	0.3726	A_08_60_S52	0.1788	I_14789_S88	0.0135		
Austria	CE_27_S19	0.2301	A_08_71_S51	0.2969	I_14874_S89	0.3636		
Austria	CE_28_S82	0.3119	A_08_88_S50	0.268	I_14875_S43	0.3403		
Austria	CE_29_S23	0.1831	A_09_03_S64	0.261	I_15068_S79	0.2691		
Ukraine	CE_2_S12	0.2971	A_09_15_S5	0.8626				
Austria	CE_30_S24	0.2581	A_09_50_S3	0.4925				
Austria	CE_31_S41	0.2847	A_09_55_S49	0.3714				
Austria	CE_32_S72	0.7286	A_09_56_S65	0.2264				
Austria	CE_34_S73	0.2795	A_09_57_S2	0.3367				
Austria	CE_35_S42	0.3624	A_09_58_S81	0.3864				
Austria	CE_37_S34	0.2169	A_09_60_S92	0.3152				
Austria	CE_38_S84	0.273	A_09_95_S66	0.2429				
Turkey	CE_3_S22	0.1504	A_10_19_S8	0.2327				
Austria	CE_40_S85	0.2388	A_10_41_S90	0.2311				
Czech Republic	CE_4_S26	0.1927	A_10_47_S48	0.4343				
Turkey	CE_5_S67	0.0921	A_10_4_S91	0.2731				
Spain	CE_6_S13	0.3224	A_10_50_S47	0.4334				
Poland	CE_7_S38	0.301	A_10_51_S7	0.3536				
Bosnia	CE_8_S14	0.1937	A_10_52_S45	0.4213				
Hungary	CE_9_S27	0.2321	A_10_65_S21	0.499				
Captive	redfox	29.4451	A_10_77_S6	0.3668				

Relatedness estimations

Sampling closely related individuals can have confounding effects on population structure and demographic history inferences. All individuals in a population are related to some extent. The coefficient of relatedness ranges between 0 and 1. One means the individuals are identical twins or they are highly inbred, and 0.5 means the two individuals are full siblings or have a parent-offspring relationship, 0.25 means that the individuals are half-siblings, half-siblings meaning share one parent, or they have a grandparent-grandchild relationship, or it is an aunt/uncle-niece/nephew where both parents shared and 0.125 means the individuals have an aunt/uncle-niece/nephew where one parent shared or they are cousins whose parents are full siblings, 0.0625 where the individuals are cousins whose parents are half-siblings and zero which means that the individuals are unrelated or the relationship between the individuals is negligibly small (Wright, 1922; Widdig, 2017). To avoid skewing downstream analysis, individuals with a sibling and parent-offspring relationship need to be excluded from downstream analysis (Coleman et al., 2016; Hahn, 2019). Relatedness estimations address this problem by computing pairwise relatedness where r_{ab} value ≥ 0.3 , which ensures that all individuals with sibling and or parent-offspring relationships are excluded, first by excluding any individual with a high number of close relatives until there was no pair of close relatives left. And in the case where two individuals were related to each other, the one with the lower coverage was excluded. In Table 2, red fox, our captive sample, I_13886_S78, A_08_100_S37, I_14789_S88, were related to multiple individuals, and that warrants their exclusion from the downstream analyses, while in the case of I_11594_S60 and I_13763_S46, the sample with lowest coverage between the two was excluded.

Table 2 High ($r_{ab} \geq 0.3$) pairwise relatedness proportions between sample a (ida) to sample b (idb).

sample a (ida)	sample b (idb)	rab
A_08_06_S58	I_13886_S78	0.334443
A_08_06_S58	redfox	0.307792
A_08_100_S37	A_08_136_S55	0.334585
A_08_100_S37	A_08_60_S52	0.445485
A_08_100_S37	A_08_71_S51	0.30948
A_08_100_S37	A_10_77_S6	0.458679
A_08_100_S37	CE_11_S15	0.366752
A_08_100_S37	CE_40_S85	0.304459
A_08_100_S37	CE_5_S67	0.732821
A_08_100_S37	CE_7_S38	0.405217
A_08_100_S37	CE_8_S14	0.388873
A_08_100_S37	I_13891_S87	0.822893
I_11594_S60	I_13763_S46	0.304064
I_11594_S60	I_13886_S78	0.337025
I_12696_S61	I_14789_S88	0.335905
I_13588_S77	I_13886_S78	0.306969
I_13763_S46	I_13886_S78	0.352511
I_13763_S46	redfox	0.306242
I_13886_S78	I_14874_S89	0.305481
I_13886_S78	redfox	0.481233
I_13891_S87	I_14789_S88	0.525146

Genetic diversity

Table 3 shows the individual heterozygosity estimates in the Arctic, Central Europe and Israeli populations. The estimates for the Arctic range between 0.000772 – 0.000068, while that of Central Europe ranges between 0.001252-0.000028, and Israel ranges between 0.000040 - 0.00376.

Table 3. Individual heterozygosity estimations across the populations

Heterozygosity estimations					
Arctic(Norway)		Central Europe		Israel	
A_10_41_S90	0.000772	CE_8_S14	0.001252	I_12691_S75	0.000376
A_08_60_S52	0.000678	CE_6_S13	0.000526	I_13619_S86	0.000274
A_10_4_S91	0.000463	CE_3_S22	0.000493	I_13763_S46	0.000237
A_08_153_S4	0.000459	CE_11_S15	0.000327	I_14875_S43	0.000211
A_09_60_S92	0.000423	CE_29_S23	0.000265	I_13602_S69	0.000183
A_08_114_S80	0.000420	CE_27_S19	0.000239	I_14874_S89	0.000178
A_08_88_S50	0.000377	CE_13_S17	0.000235	I_13891_S87	0.000139
A_09_57_S2	0.000330	CE_21_S18	0.000209	I_12696_S61	0.000138
A_08_71_S51	0.000295	CE_1_S11	0.000202	I_12653_S68	0.000128
A_09_15_S5	0.000277	CE_24_S32	0.000194	I_13588_S77	0.000125
A_08_113_S9	0.000269	CE_32_S72	0.000170	I_11590_S74	0.000120
A_09_55_S49	0.000244	CE_2_S12	0.000157	I_15068_S79	0.000119
A_08_06_S58	0.000236	CE_30_S24	0.000128	I_13958_S20	0.000117
A_08_203_S59	0.000230	CE_7_S38	0.000116	I_13407_S76	0.000095
A_08_107_S57	0.000224	CE_12_S16	0.000114	I_13884_S35	0.000067
A_09_50_S3	0.000217	CE_23_S71	0.000111	I_11595_S25	0.000040
A_08_160_S54	0.000213	CE_25_S40	0.000110		
A_10_77_S6	0.000209	CE_22_S39	0.000103		
A_08_117_S56	0.000198	CE_17_S70	0.000099		
A_10_52_S45	0.000174	CE_26_S33	0.000098		
A_08_209_S53	0.000168	CE_10_S83	0.000095		
A_10_65_S21	0.000168	CE_14_S28	0.000094		
A_08_154_S62	0.000168	CE_35_S42	0.000090		
A_09_58_S81	0.000158	CE_31_S41	0.000085		
A_09_95_S66	0.000152	CE_19_S30	0.000083		
A_10_47_S48	0.000151	CE_37_S34	0.000076		
A_10_50_S47	0.000139	CE_34_S73	0.000076		
A_08_136_S55	0.000131	CE_28_S82	0.000074		
A_08_118_S36	0.000130	CE_9_S27	0.000072		
A_10_19_S8	0.000123	CE_18_S29	0.000064		
A_09_03_S64	0.000110	CE_38_S84	0.000058		
A_08_108_S10	0.000105	CE_4_S26	0.000055		
A_10_51_S7	0.000103	CE_40_S85	0.000055		
A_08_169_S63	0.000097	CE_20_S31	0.000051		
A_09_56_S65	0.000068	CE_5_S67	0.000028		

In Figure 3, the graph shows each population's average individual heterozygosity levels. The graph further shows that the most diverse population was the Arctic population, followed by Central Europe and then Israel, which had the least genetic diversity.

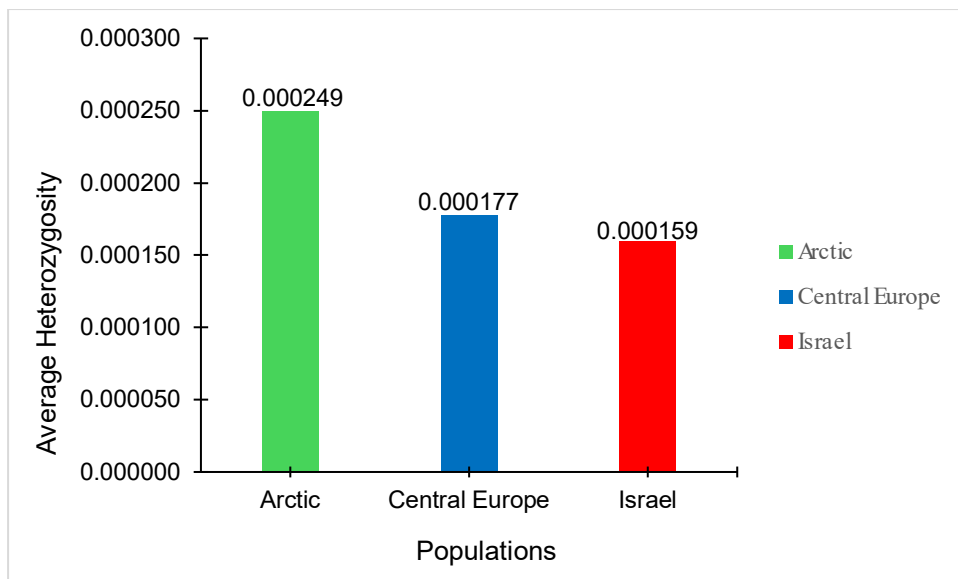


Figure 3 The average individual heterozygosity levels per population show that the Arctic population appears more diverse than that of Central Europe and Israel

Population Structure

Principal component analysis

The principal component analysis plots (Figure 4) show that the sampled populations are genetically structured. In the PCA plot on the left, the plot for -setMinDepthInd 1, PC1 separates two groups of mostly individuals from Israel and a mixed group of Central Europe and Arctic individuals. Most of the Arctic separate from the Central Europe individuals on PC2. The right plot, which is the plot for -setMinDepthInd 3, shows a similar pattern, but with a clearer separation of most Arctic and Central European individuals along PC2. PC1 explains most of the variation between the populations, which is 26.17% in the left plot and 47.32% in the right plot.

In both cases A_08_06_S58 form a separate group with I_12691_S75, I_12696_S61, I_13602_S69, I_13619_S86, I_13891_S87, I_14874_S89 and CE_3_S22, I_11595_S25, I_12653_S68, I_13407_S76, I_13588_S77, I_13884_S35, I_13958_S20, I_14875_S43, and I_15068_S79. The left PCA was generated from 31,849 single nucleotide polymorphisms, whereas the one on the right was generated from 7,667 single nucleotide polymorphisms. The red individual clustering with the Central European population is I_11590_S74, while the blue individual clustering with Israel is the Turkey sample CE-3-S22, the green sample clustering with Israel on the far right is A_08_06_S58, and the five individuals from Arctic (green), clustered within Central Europe: A_08_209_S53, A_08_60_S52, A_09_56_S65, A_09_60_S92, A_10_19_S8.

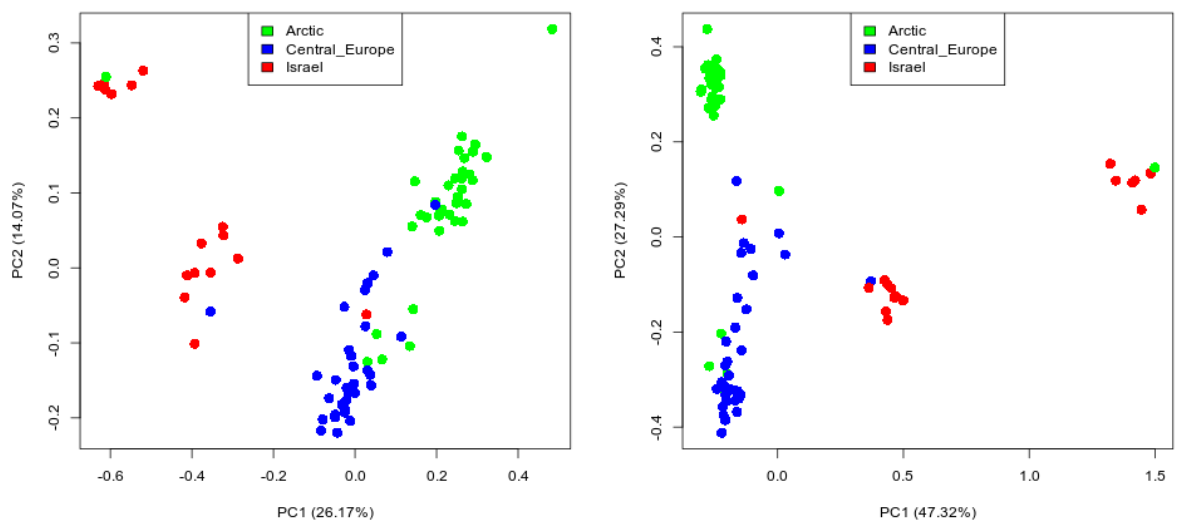


Figure 4 Principal component analyses with -setMinDepthInd 1 (left) and -setMinDepthInd 3 (right). Both plots show evidence of structuring: Green represents the Arctic population, Blue Central Europe, and red Israel.

Admixture Proportions

Identifying Best K-means clusters

The (delta) ΔK (Evanno et al., 2005) computation was used to identify the different clusters where individual red foxes were assigned by the admixture analysis. This is a statistic based on the rate of change in the log probability of data shown as the best likelihood from NGSADMIX between successive K values. Both tests -setMinDepthInd 1 and -setMinDepthInd 3 support the same best k clustering K=3. See Figure 5.

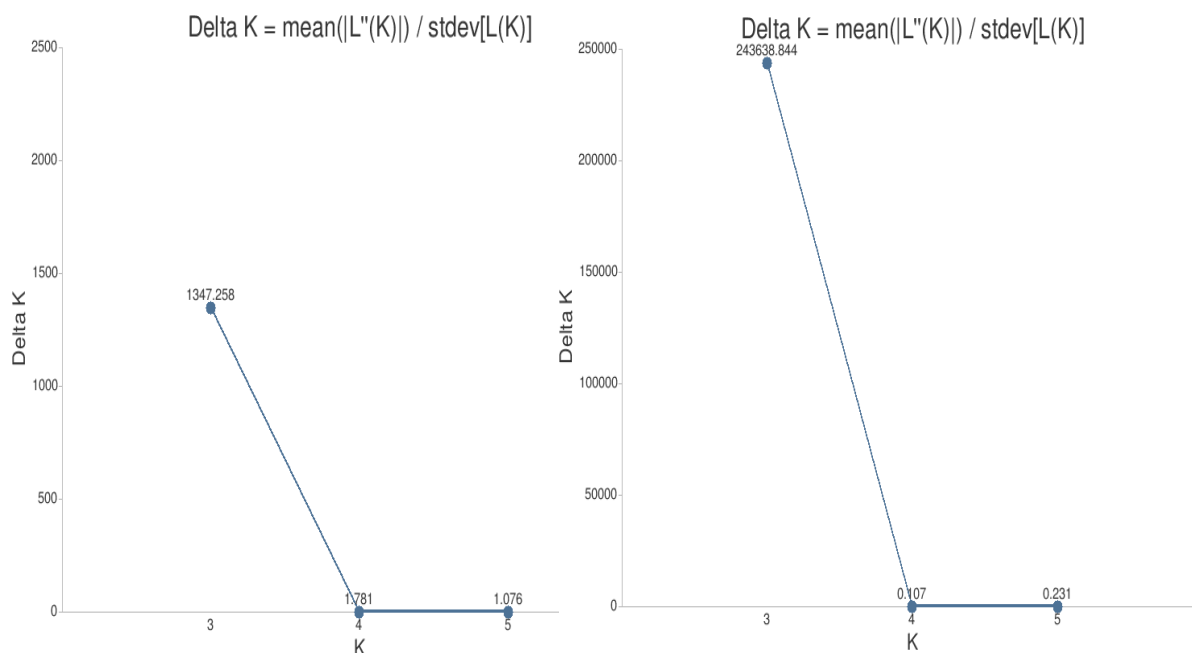


Figure 5 Depicts the best K-means clustering for red fox populations ($K=3$). Delta K is a statistic that is based on the rate of change in the log probability of data between consecutive K values. Delta K is estimated as a mean of absolute values of $L''(K)$ divided by the standard deviation of $L(K)$, and the highest change in delta K will show the best K cluster.

The population admixture proportions (Figure 6), were modelled for five different K-means clusters, $K=2$ to $K=6$, with each K run repeated 50 times to ascertain confidence of the most probable K grouping. $K=3$ was the only K value for which 50 of 50 replicates supported the same grouping without minor clustering. However, there was considerable structuring at all other K values. At $K=2$, Central Europe and Arctic populations fall within the same group, separated from Israel (red) in both plots. Central European individuals have tiny proportions of admixture with Israel these may not be admixture but shared ancestral polymorphism between two populations. One individual CE_3_S22 sampled in Turkey stands out. This individual is completely red and separates from Europe also at $K=2$. The Arctic sample, A_08_06_S58, also shows traces of red at $K=2$. At $K=3$, all three populations were separated from each other. This grouping also showed the same genetic contact between Israel and Central Europe from $K=2$. In addition, there seems to be more gene-flow between Central Europe and the Arctic than there is between Central Europe and Israel. $K=3$ also revealed what could be evidence of recent introgression into the Arctic from Central Europe, Individuals A_08_209_S53, A_08_60_S52, A_09_56_S65, A_09_60_S92, and A_10_19_S8 are still genetically similar to the Central European individuals. The population clusters modelled by $K=4$, $K=5$ and $K=6$ did not appear to add more biologically relevant information in the top plots (-setMinDepthInd 1). On the other hand, the bottom plot (-setMinDepthInd 3), only $K=4$ and

K=5, did not reveal any biologically relevant information. However, at K=6, the top admixture plot (-setMinDepthInd 1) failed to make the separation of the two clusters found in the Principal component analysis. The bottom admixture plot (-setMinDepthInd 3), separated the two groups within Israel with I_13884_S35, I_14875_S43 I_12653_S68, I_13602_S69, I_12691_S75, I_13407_S76, I_13588_S77, and I_15068_S79 separating from the rest.

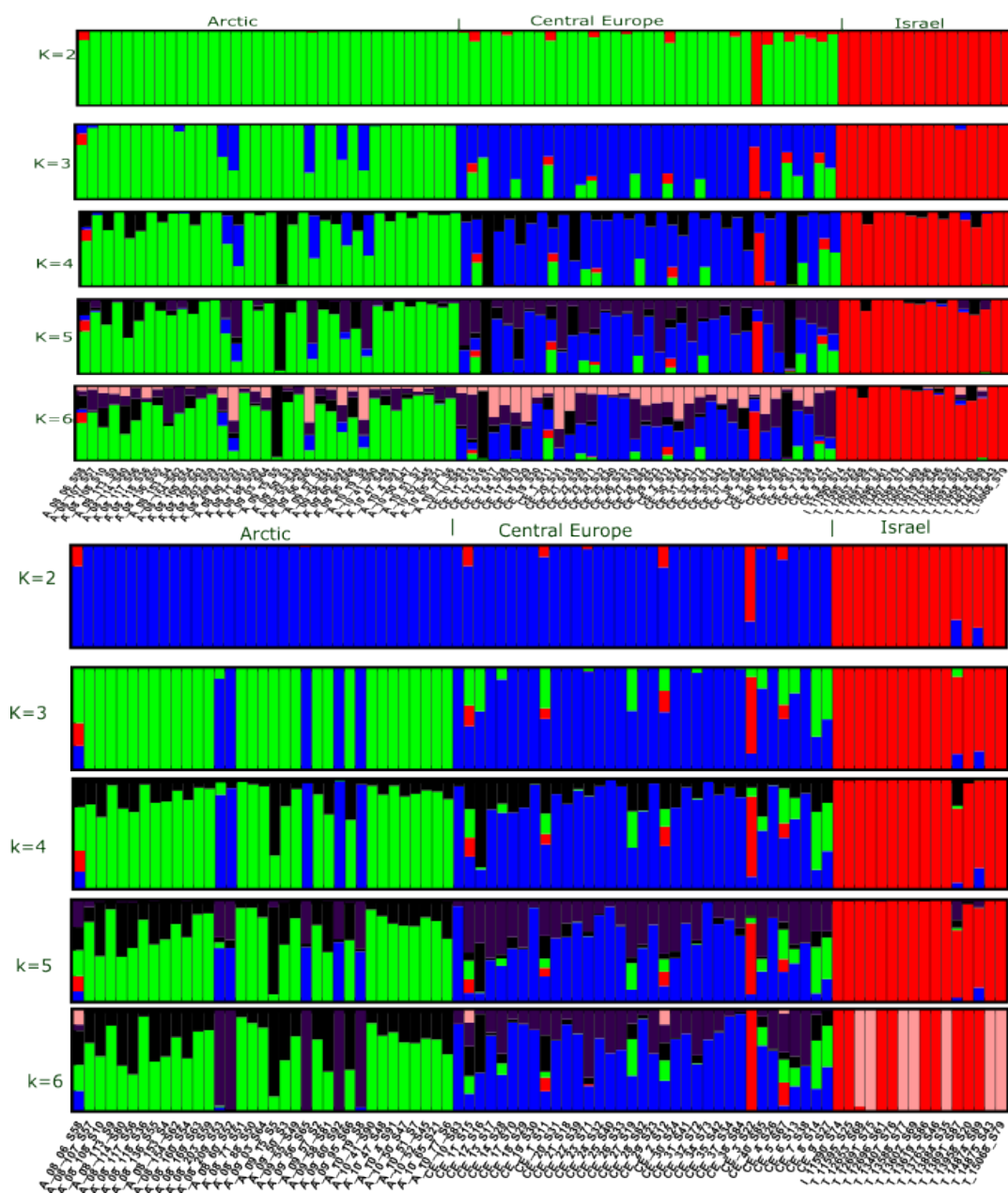


Figure 6 Admixture proportions inferred for K2 to K6 were replicated 50 times to find the best-supported clusters. Optimal K clustering by Evanno delta K supported K=3. These

admixture proportions were computed with -setMinDepthInd 1 using 31849 SNPs (top) and -setMinDepthInd 3 with 7667 SNPs (bottom), and they show similar structuring. The different coloured bars represent individuals in each population moving from the Arctic(green), Central Europe (Blue) and Israel(red); the populations are shown by the small lines on top of the bar charts depicting where one population ends and where the next population begins.

Phylogenetic analysis

The phylogenetic trees (Figure 7 and Figure 8) were reconstructed from the pairwise genetic distance matrix computed using the average of the posterior probabilities' distribution of all possible genotypes in NGSDIST and the genetic distance matrix computed from genotype calls. The tree (Figure 7) does not show any tangible evidence of structure, rather, it shows unusual evidence of a continuous panmictic population between the groups. However, the tree shows to some degree that Arctic and Central European populations have the highest genetic contact. Bootstrap support values range between 57 and 100. All branches with bootstrap support of less than 70 show moderate support for the phylogenetic relationships depicted. All values equal to or greater than 70 indicate reliable relationships, and those greater than 80 indicate strong phylogenetic relationships. There are several places in the phylogenetic tree (Figure 7) where a bootstrap value of 70 -100 suggests strong support for the evolutionary relationship between the Arctic and Central Europe; there are also places where it shows the same relationship between Central Europe and Israel such as the grouping at the bottom which shows CE_11_S16 with Israel samples.

Genetic distance tree

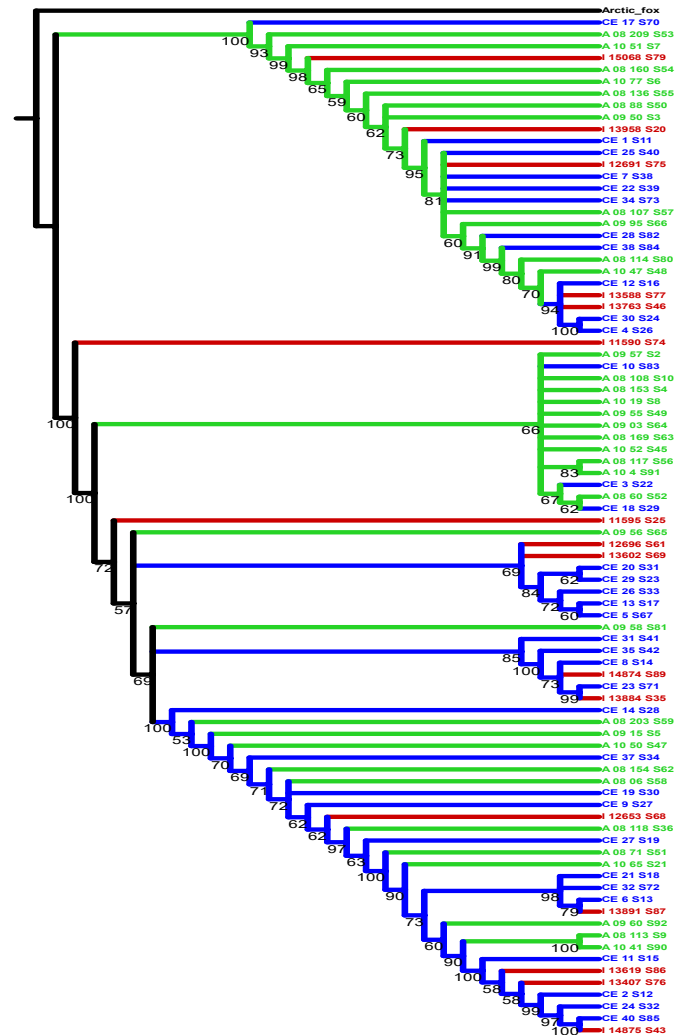


Figure 7 The phylogenetic tree of the red foxes was constructed from the genetic distance matrix generated using posterior probabilities in NGSDIST and FASTME using 100 bootstraps and 11,586 SNPs; green represents the Arctic population, blue represents Central Europe, and red represents Israel. The individual black is the Arctic fox; the outgroup genome used to provide a reference of the ancestral character state.

UPGMA tree

In SPLITSTREE, a variety of phylogenetic trees can be output. This feature was used to output a UPGMA tree (Figure 8) without running another tree computation. The UPGMA tree and the tree in Figure 7 both use genetic distance to infer phylogenetic relationships. This tree (Figure 8) differs from the tree in (Figure 7) in the method used to compute the genetic distance. The former tree from NGSDIST used log-transformed P distances computed based on posterior

probabilities, while the latter used pairwise distances from ADEGENET (Jombart and Collins, 2015), which were computed based on genotype calls in ANGSD. Both trees showed evidence of genetic contact between the Arctic and Central European populations and between Central Europe and Israel. Although Figure 7 depicts this to a lesser extent than that depicted in Figure 8. The UPGMA tree topology illustrates some of the observed grouping similar to the admixture analyses CE-3-S22, clustered in the same group with the Israel population, while A_08_06_S58, groups with I_13958_S20, I_13884_S35, I_13763_S46.

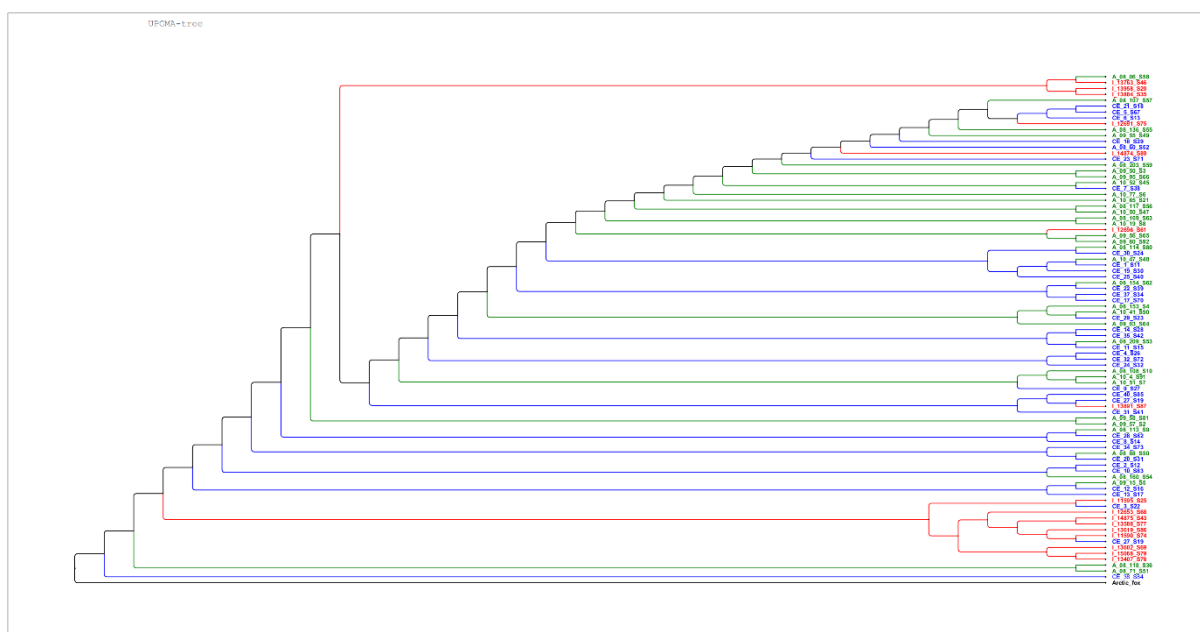


Figure 8 Phylogenetic tree depicting evolutionary relationships of red fox generated from genetic distance computed from genotype calls (using ADEGENET and SPLITSTREE). This tree gives a general view of the tree topology given by the red fox samples using 11,586 SNPs

Phylogenetic network

The phylogenetic network (Figure 10) illustrates that there is higher genetic contact between Arctic and Central European individuals, as well as between Central Europe and Israel. The following Israel individuals (red) I_11595_S25, I_14875_S43, I_13602_S69, I_11590_S74 I_13407_S76, I_13588_S77, and I_14874_S89 forms a separate group. However, genetic relationships exist between Israel, I_13958_S20, I_13619_S86, I_13891_S87, I_12653_S68, and I_12691_S75 and Central Europe. This is not consistent with previous results. In the Principal component analysis, I_11595_S25, I_14875_S43, I_13407_S76 and I_13588_S77 form a separate cluster with CE_3_S22 while I_13602_S69 and I_14874_S89 clusters with A_08_06_S58 which is not consistent with Figure 9. In the

Admixture plot (Figure 6), I_14875_S43, I_13602_S69 and I_13588_S77 are separated from the rest of Israel at K=6 and the remaining are red in the same cluster. Apart from these, the network revealed the admixture between Central Europe and Arctic A_08_209_S53, A_08_60_S52, A_09_56_S65, and A_09_60_S92 are clustered with Central Europe. The clustering of Arctic and most Central European individuals from Spain, Turkey, Austria, and Germany to the right could be evidence of northward expansion from Central Europe to the Arctic.

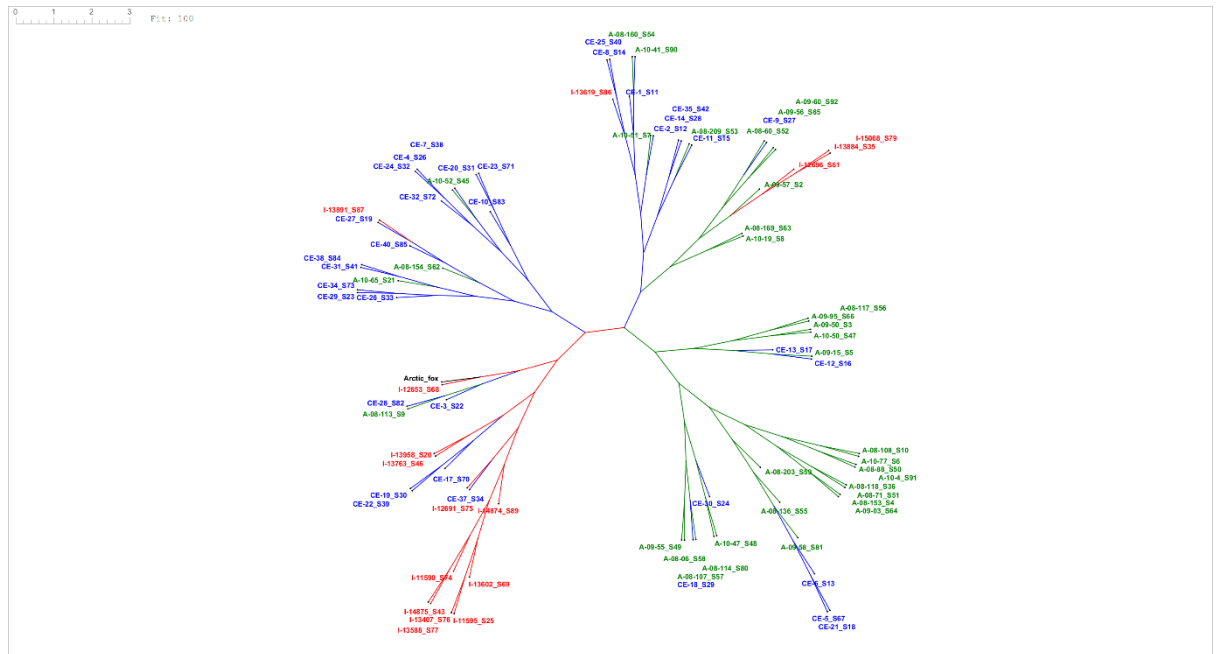


Figure 9 Depicts a consensus phylogenetic network constructed using Neighbour-Joining (Saitou and Nei, 1987) in SPLITSTREE (11,586 SNPs). The network shows distinct clustering of Israel, Central Europe and the Arctic red fox populations. It also depicts evidence which suggests the presence of gene-flow between Central Europe (blue) and Israel (red) and shared ancestry between Arctic and Central Europe.

Gene-flow

Gene-flow analysis was carried out using ABBA-BABA D-statistics (Four-taxon model). This test requires a predefined topology, which defines the evolutionary relationship between the populations ([[[H1, H2], H3], Outgroup]) (Westbury et al., 2021). The admixture analysis above show that the Arctic and Central Europe have the highest genetic contact. This means that these two populations share the most recent common ancestor, making them sister taxa, which means Israel would be a distantly related taxon. In the admixture analyses (Figure 7), Israel is completely separated from the rest of the populations at K=2. Central Europe

clustered with the Arctic, forming the second grouping for $K=2$, which could also be evidence of shared ancestry between the two populations. At $K=3$ in the admixture analyses, there is high genetic contact between Central Europe and the Arctic, and this observation is also true for all the results discussed above. With this in mind, the expected tree topology for the gene-flow analysis would be ([[[Arctic, Central Europe], Israel], Arctic fox]) (Figure 11). Given this, if the test shows equal ABBA and BABA sites, then the D statistic will equal zero, meaning there is no gene-flow. Therefore, only the first row of Table 4 represents a test for gene-flow, the two successive rows do not follow the known evolutionary topology, and any significant values indicate shared ancestry and not gene-flow. The first row of Table 4 shows a positive D-statistic, meaning H2 shares more alleles with H3 than does H1. In the context of our populations, this is a signature of gene-flow between Central Europe and Israel, consistent with phylogenetic and admixture analyses. The significant Z-score value, $Z=8.553$, also supports this result. This measure of significance was suggested by, Reich et al. (2009), who put forward that $zscore>3$ was to be considered significant.

Table 4 Signatures of gene-flow between all combinations of populations regardless of tree topology. A negative Dstat value means that H1 is closer to H3 than H2. Likewise, a positive value means H2 is closer to H3 than H1.

D	JK-D	V(JK-D)	Z	pvalue	nABBA	nBABA	nBlocks	H1	H2	H3	H4
0.0181	0.0181	0.000004	8.5553	0	1225.5	1181.9	1573	Arctic	Central_Europe	Israel	Arctic_fox
0.0135	0.0135	0.000014	3.635	0.000278	1225.5	1192.7	1591	Arctic	Israel	Central_Europe	Arctic_fox
-0.0046	-0.0046	0.000012	-1.2985	0.194111	1181.9	1192.7	1593	Central	Israel	Arctic	Arctic_fox

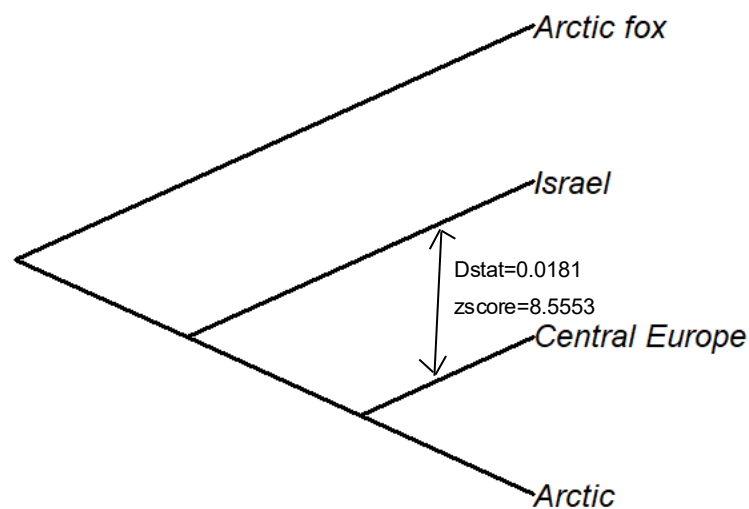


Figure 10 Cladogram of the expected tree topology among the sampled redfox populations. The high score, $Z=8.5553$, observed between Central Europe and Israel suggests that there could be gene-flow between the two populations.

Selection Analysis

To investigate signatures of positive selection in the red foxes in the Arctic, the population branch statistic (PBS) was computed, which flags regions with rapid changes in allele frequency as potential candidates for selection. Since this was a frequency-based test, it was computed using the previously discussed depths, setMinDepthInd 1 and 3. However, although setMinDepthInd 3 yielded high-quality SNPs, it filtered out a lot of SNPs. Hence, the Arctic was left with 9,533 SNPs, Central Europe with 5,323 while Israel was left with 5,755. This raises the concern of filtering out potentially informative SNPs for selection inference. The alternative -setMinDepthInd 1 retained more SNPs, with the Arctic retaining 37,024, Central Europe, 23,035 and Israel, 20,486. Using the setMinDepthInd 3 setting resulted in the flagging of only 17 genes (Table 5). In contrast setMinDepthInd 1 flagged of 106 candidate loci (Table 6). Sliding window results were plotted using the qqman package (Turner, 2014) in R to produce the Manhattan plot in (Figure 11). The red fox only has 34 chromosomes; however, the Manhattan plot shows more than 60 values on the x-axis due to a lack of chromosomal level reference. The only available red fox reference was the scaffold-level assembly by Kukekova et al. (2018). Therefore, the chromosomes are represented by scaffolds. Larger scaffolds or contigs take up the area in the plot, with the smaller ones packed towards the right end of the plot. All scaffolds that showed high values were included in the test when candidate genes were extracted using BEDTOOLS, regardless of size. The plot shows several outliers, representing the locations in the scaffold where candidate genes for selection would likely be found.

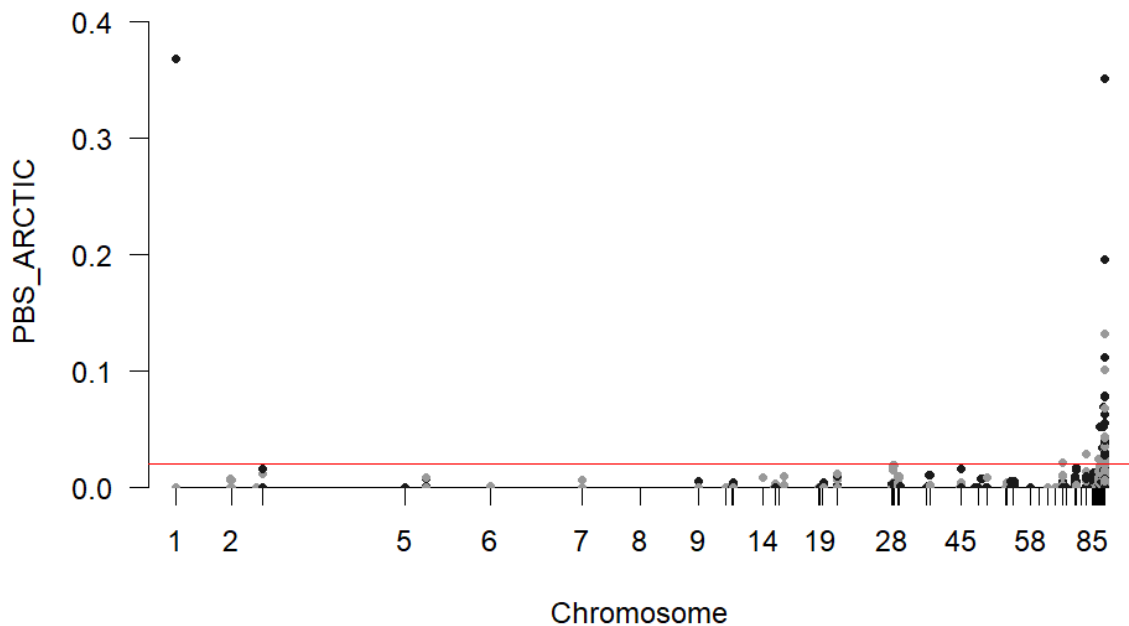


Figure 11 Shows a Manhattan plot of sliding window genome-wide selection scan of red fox from the Arctic. Significant p-values are shown above the red horizontal line ($P=0.01$) representing the average level of genetic drift across the genomes. All SNPs with p-values above the red line were flagged for selection. The red fox data was aligned to scaffold level reference genome, hence the many numbers on the chromosome axis.

Genome-wide scans that test each of thousands of sliding windows across the genome individually raises the issue of multiple testing (Mieth et al., 2016). To assess this, the distribution of the p-values was plotted to determine whether they showed a uniform distribution. Under the null hypothesis of no association, p-values are expected to be uniformly distributed between 0 and 1 (Benjamini and Hochberg, 1995). The plot in Figure 12 shows an exponential distribution, which means the p-values in this data are not behaving uniformly as required under the null hypothesis. Therefore, they need to be adjusted. The false discovery rate was used to correct these data since the distribution does not impact it significantly. This means that the false discovery rate remains effective regardless of the statistical distribution of the p-values. This was done to minimise the chance of getting a false positive without being too conservative.

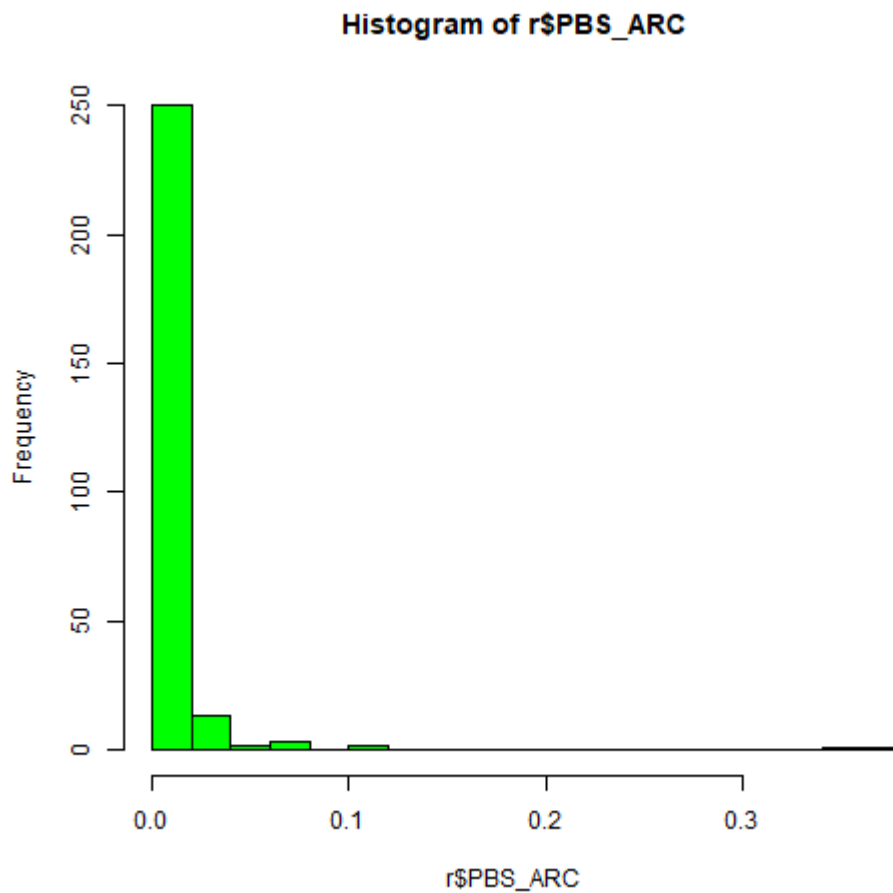


Figure 12 Depicts the distribution of pvalues across sliding windows. The plot shows an exponential distribution, which means the pvalues in this data are not behaving uniformly as required under the null hypothesis therefore, FDR was used to adjust them and control multiple testing less conservatively.

Candidate genes for selection

Since their expansion into the Arctic, red foxes must have faced unique selective pressures, which would have encouraged the development of unique adaptive strategies to survive in the new environment. Red foxes have been found to have inferior physical adaptations compared to the arctic fox, a well-adapted *Vulpes* species in the Arctic environment (Hersteinsson and Macdonald, 1992). To investigate whether the continued persistence of red foxes in the Arctic had to do with this species' evolving mutations that helped them endure the harsh conditions of their new habitats, a population branch statistic test (PBS) was carried out using two Depths -setMinDepthInd 1 and -setMinDepthInd 3. These depths were used to check how reliable the analysis was when the number of single nucleotide polymorphisms (SNPs) changed. This test searched for regions along the whole genome with rapid changes in the allele frequencies, and these would be shown as regions with very high PBS in the sliding window analysis,

revealing signatures of positive selection in the red foxes from the Arctic. All windows with high PBS were extracted from the sliding window and were run through BEDTOOLS (Quinlan and Hall, 2010). All features that overlapped or that were closest to a gene, upstream and downstream in the red fox gff file, a file which describes the locations of annotated genes in the genome were extracted based on the query file. The query file contained the scaffold names, and their locations of SNPs previously identified as having high PBS. Tables 5 and 6 outline all genes with related functions together. All blank spaces in the function column are genes which could not be linked to any function. The functions of the identified genes were found by running searches in NCBI GENE using *Vulpes vulpes* and *Canis lupus familiaris* as reference or target species. *C.l.familiaris* was used for the genes whose function was not found in NCBI using *V.vulpes*. GENECARDS was also used on genes where nothing was found on NCBI GENE (Tables 6 and 7).

The results revealed that red foxes in the Arctic may have evolved unique adaptations which allow them to thrive here. Table 5 summarises some of these unique adaptations. In addition, Tables 6 and 7, account for all genes found in this study and they also show at least seven genes flagged using both -setMinDepthInd 1 and 3 namely ANKRD13A (protein binding), CH25H (enables iron ion binding), CUNH8orf33 (unknown), LIPA (enables sterol ester esterase activity), MRPS24 (enables structural constituents of ribosomes), PDE5A (enables 3',5'-cyclic-GMP phosphodiesterase activity) and LOC112911681 (unknown).

Table 5. Candidate genes for some crucial processes required for species to survive the Arctic conditions that have been identified in the red foxes.

Functions and/Processes	Genes
Metabolism	GALNT13, GBE1, MOGAT1, HSD17B4, ADCY1, ACSL3, LOC112912510, SGPP2, MTM1, PIK3CA, PSAT1, LOC112927225 and PDE5A
Olfaction	LOC112912474, LOC112912081 and LOC112912082
Hypoxia	NR4A2, VEGFC and LOC112927226
Circadian rhythm	ADCY1, KCNJ3
Male fertility	ADAM20, ADAM21, LOC112927152, and NPAP1
Nervous system development	DBNL, NR4A2, GIT2 and GPRIN1
Metal ion binding	GIT2, RNF44, LOC112911121, ADAM21, ADAM20, ADCY1, LOC112912697,

	LOC112911116, ZNF16, LOC112927152, ZNF713, PDE5A and ZMYND11
Androgen metabolic process	HSD17B4 , LOC112912510
Progesterone-mediated oocyte maturation	LOC112911081, PIK3CA and ADCY1
Immunity	LOC112914295, LOC112911121 and LOC112911116

Table 6. All candidate genes (17 genes) and their functions that are potential targets of selection, along with their names and inferred functions based on the Population branch statistics calculated from site allele frequencies computed with -setMinDepthInd 3.

ID	Gene Name	Function
CSMD1	CUB and Sushi multiple domains 1	Regulate brain development complements activity and circuit development
ANKRD13A	ankyrin repeat domain 13A	enables protein binding
LOC112912624	ankyrin repeat domain-containing protein 26-like	enables protein binding
CH25H	cholesterol 25-hydroxylase	enables C-4 methyl sterol oxidase and sterol hydroxylase activities, enables iron ion binding
CUNH8orf33	chromosome unknown C8orf33 homolog	
LOC112912623	coiled-coil domain-containing protein 144B-like	
LIPA	lipase A, lysosomal acid type	enables sterol esterase activity
MRPS24	mitochondrial ribosomal protein S24	enables structural constituents of ribosomes.
		Enables RNA binding
MYOM2	myomesin 2	enables protein binding
LOC112912081	olfactory receptor 52R1-like	enables G-protein-coupled receptor and Olfactory receptor activities

Table 6 continued.

ID	Gene Name	Function
LOC112912082	olfactory receptor 52R1-like	enables G-protein-coupled receptor and Olfactory receptor activities
LOC112912367	paraneoplastic antigen-like protein 6B	
PDE5A	phosphodiesterase 5A	enables protein binding, enables 3',5'-cyclic-GMP phosphodiesterase activity
LOC112912083	ubiquitin carboxyl-terminal hydrolase 17-like protein 6	enables cysteine-type endopeptidase activity
		Enables cysteine-type deubiquitinase activity
LOC112912592	ubiquitin carboxyl-terminal hydrolase 17-like protein 6	enables cysteine-type endopeptidase and deubiquitinase activity
LOC112911681	uncharacterised LOC112911681	
ZNF131	zinc finger protein 131	enables DNA-binding transcription factor activity, RNA polymerase II-specific,
		enables RNA polymerase II cis-regulatory region sequence-specific DNA binding, enables protein binding

Table 7 All candidate genes (106 genes) and their functions that are potential targets of selection, along with their names and inferred functions, based on the Population branch statistics calculated from site allele frequencies computed with -setMinDepthInd 1

ID	Gene Name	Function
ACSL3	acyl-CoA synthetase long-chain family member 3	enables long-chain fatty acid-CoA ligase activity
ACTA2	actin alpha 2, smooth muscle	protein kinase binding
ADAM20	ADAM metallopeptidase domain 20	enables metalloendopeptidase activity
ADAM21	ADAM metallopeptidase domain 21	enables metalloendopeptidase activity
ADCY1	adenylate cyclase 1	enables adenylate cyclase activity

AGA	Aspartylglucosaminidase	enables N--L-asparaginase activity
ANKRD13A	ankyrin repeat domain 13A	enables protein binding
CCDC7	coiled-coil domain containing 7	enables protein binding
CDHR2	cadherin related family member 2	enables cadherin binding calcium ion binding
CGGBP1	CGG triplet repeat binding protein 1	enables double-stranded DNA binding
CH25H	cholesterol 25-hydroxylase	enables C-4 methylsterol oxidase activity, enables iron ion binding, and enables steroid hydroxylase activity.
CLTB	clathrin light chain B	enables clathrin heavy chain binding and structural molecule activity
CMKLR1	chemerin chemokine-like receptor 1	enables G protein-coupled receptor activity and complement receptor activity
CUNH8orf33	Chromosome unknown C8orf33 homolog	
DBNL	drebrin like	enables actin filament binding
DDX56	DEAD-box helicase 56	enables ATP binding, RNA helicase activity, and nucleic acid binding

Table 7 continued.

ID	Gene Name	Function
EIF4E1B	eukaryotic translation initiation factor 4E family member 1B	enables RNA 7-methylguanosine cap binding and translation initiation factor activity
FAF2	Fas-associated factor family member 2	enables ubiquitin binding
FAM170A	family with sequence similarity 170 member A	enables DNA binding activity and metal ion binding activity
FARSB	phenylalanyl-tRNA synthetase subunit beta	enables ATP binding, RNA binding, magnesium ion binding, and phenylalanine-tRNA ligase activity
FAS	Fas cell surface death receptor	enables protein binding and transcription corepressor activity
FCGR2B	Fc fragment of IgG receptor lib	enables transmembrane signalling receptor activity
FICD	FIC domain protein adenylyltransferase	enables protein binding

GALNT13	polypeptide N-acetylgalactosaminyltransferase 13	catalyses the initial reaction in O-linked oligosaccharide biosynthesis
GBE1	1,4-alpha-glucan branching enzyme 1	enables 1,4-alpha-glucan branching enzyme activity, cation binding, and hydrolase activity, hydrolysing O-glycosyl compounds
GIT2	GIT ArfGAP 2	enables GTPase activator activity and protein binding
GPRIN1	G protein regulated inducer of neurite outgrowth 1	enables phosphoprotein binding activity
HSD17B4	hydroxysteroid 17-beta dehydrogenase 4	enables oxidoreductase activity
HTR1F	5-hydroxytryptamine receptor 1F	enables G protein-coupled serotonin receptor activity and neurotransmitter receptor activity

Table 7 continued

ID	Gene Name	Function
IGFBP1	Insulin-like growth factor binding protein 1	enables insulin-like growth factor I binding and insulin-like growth factor II binding
KCNJ3	potassium inwardly-rectifying channel subfamily J member 3	enables G-protein-activated inward rectifier potassium channel activity
KCNMB3	potassium calcium-activated channel subfamily M regulatory beta subunit 3	enables calcium-activated potassium channel activity and potassium channel regulator activity
LIPA	lipase A, lysosomal acid type	enables sterol ester esterase activity
LOC112909788	heat shock transcription factor, X-linked member 3-like	enables DNA-binding transcription factor activity and sequence-specific DNA binding
LOC112909789	melanoma-associated antigen 8-like	negative regulation of transcription by RNA polymerase II
LOC112909799	basic proline-rich protein-like	

LOC112911066	membrane-spanning 4-domains subfamily A member 13-like	involved_in cell surface receptor signalling pathway
LOC112911081	mitotic spindle assembly checkpoint protein MAD2A	mitotic spindle assembly checkpoint signalling

Table 7 continued.

ID	Gene Name	Function
LOC112911116	zinc finger protein 252	enables DNA-binding transcription repressor activity, RNA polymerase II-specific
LOC112911121	zinc finger protein 250-like	enables DNA-binding transcription activator activity, RNA polymerase II-specific, enables DNA-binding transcription repressor activity, RNA polymerase II-specific, enables RNA polymerase II cis-regulatory region sequence-specific DNA binding
LOC112911681	uncharacterised LOC112911681	
LOC112912474	olfactory receptor 4K13	enables G protein-coupled receptor activity and olfactory receptor activity
LOC112912508	ubiquitin carboxyl-terminal hydrolase 17-like protein 17	enables cysteine-type deubiquitinase activity and cysteine-type endopeptidase activity
LOC112912510	3-oxo-5-beta-steroid 4-dehydrogenase-like	enables Delta4-3-oxosteroid 5beta-reductase activity, aldose reductase activity, ketosteroid monooxygenase activity, and steroid dehydrogenase activity
LOC112912696	basic proline-rich protein-like	
LOC112912697	adenylate cyclase type 1-like	enables adenylate cyclase activity

Table 7 continued.

ID	Gene Name	Function
LOC112912790	ankyrin repeat domain-containing protein 26-like	enables protein binding
LOC112914295	Low-affinity immunoglobulin gamma Fc region receptor III-like	enables transmembrane signalling receptor activity
LOC112917682	calcium uniporter regulatory subunit MCUb, mitochondrial-like	enables calcium channel inhibitor activity
LOC112917702	putative glycosyltransferase ALG1-like	enables mannosyltransferase activity
LOC112921570	ankyrin repeat domain-containing protein 26-like	enables protein binding
LOC112927152	disintegrin and metalloproteinase domain-containing protein 20-like	enables metalloendopeptidase activity
LOC112927225	L-lactate dehydrogenase B chain-like	enables L-lactate dehydrogenase activity
LOC112927226	peptidyl-prolyl cis-trans isomerase A-like	enables peptidyl-prolyl cis-trans isomerase activity
LOC112932189	uncharacterised LOC112932189	
LOC112932861	dnaJ homolog subfamily C member 8-like	

Table 7 continued

ID	Gene Name	Function
MAMLD1	mastermind-like domain containing 1	transcriptional co-activator
MED13L	mediator complex subunit 13L	enables transcription coregulator activity
MOGAT1	monoacylglycerol O-acyltransferase 1	enables diacylglycerol O-acyltransferase activity
MRPS17	mitochondrial ribosomal protein S17	enables structural constituent of ribosome
MRPS24	mitochondrial ribosomal protein S24	enables structural constituent of ribosome, enables RNA binding
MS4A15	membrane-spanning 4-domains A15	regulates lipid saturation and length
MTM1	myotubularin 1	enables phosphatidylinositol-3,5-bisphosphate 3-phosphatase activity
MYT1	myelin transcription factor 1	enables DNA-binding transcription factor activity, RNA polymerase II-specific
MZB1	marginal zone B and B1 cell-specific protein	regulates cellular proliferation, mitochondrial dysfunction, and inflammation

NEIL3	nei like DNA glycosylase 3	enables DNA N-glycosylase activity, DNA-endonuclease activity, damaged DNA binding, and zinc ion binding
-------	----------------------------	--

Table 7 continued.

ID	Gene Name	Function
NIPSNAP2	Nip snap homolog 2	vesicular transport
NPAP1	nuclear pore-associated protein 1	Spermatogenesis
NPC1L1	NPC1, like intracellular cholesterol transporter 1	enables lipid transporter activity
NR4A2	nuclear receptor subfamily 4 group A member 2	enables RNA polymerase II cis-regulatory region sequence-specific DNA binding, nuclear glucocorticoid receptor binding, nuclear receptor activity, and zinc ion binding
NUDCD3	NudC domain containing 3	enables unfolded protein binding
OPRL1	opioid-related nociceptin receptor 1	enables neuropeptide binding and nociceptin receptor activity
PAIP2	poly binding protein interacting protein 2	enables mRNA regulatory element binding translation repressor activity
PCMTD2	protein-L-isoaspartate O-methyltransferase domain containing 2	enables protein-L-isoaspartate O-methyltransferase activity
PDE5A	phosphodiesterase 5A	enables 3',5'-cyclic-GMP phosphodiesterase activity and protein binding

Table 7 continued

Gene	Gene Name	Function
PGBD2	piggyBac transposable element derived 2	enables sequence-specific DNA binding
PIK3CA	phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha	enables 1-phosphatidylinositol-3-kinase activity and enables 1-phosphatidylinositol-4-phosphate 3-kinase activity
PROB1	proline rich basic protein 1	
PSAT1	phosphoserine aminotransferase 1	enables O-phospho-L-serine:2-oxoglutarate aminotransferase activity

		and enables pyridoxal phosphate binding
PTCHD3	a patched domain containing 3	hedgehog receptor activity
RAB18	RAB18, member RAS oncogene family	enables GTPase activity, enables GTP binding
RNF44	ring finger protein 44	involved in protein-protein and protein-DNA interactions
SERPINB10	serpin family B member 10	enables serine-type endopeptidase inhibitor activity
SERPINB11	serpin family B member 11	enables serine-type endopeptidase inhibitor activity
SERPINB13	serpin family B member 13	enables serine-type endopeptidase inhibitor activity
SERPINB2	serpin family B member 2	enables serine-type endopeptidase inhibitor activity
SERPINB3	serpin family B member 3	enables serine-type endopeptidase inhibitor activity
SERPINB7	serpin family B member 7	enables serine-type endopeptidase inhibitor activity

Table 7 continued

ID	Gene Name	Function
SGPP2	sphingosine-1-phosphate phosphatase 2	enables sphingosine-1-phosphate phosphatase activity
SLC23A1	solute carrier family 23 member 1	enables transmembrane transporter activity
SNCB	synuclein beta	enables cuprous ion binding
SPCS3	signal peptidase complex subunit 3	enables peptidase activity
SYNJ2BP	synaptojanin 2 binding protein	enables protein binding
TCHP	trichoplein keratin filament binding	enables protein binding
TLE4	TLE family member 4, transcriptional corepressor	enables protein binding and transcription corepressor activity
TMED4	transmembrane p24 trafficking protein 4	enable cargo receptor activity
TNFAIP8	TNF alpha-induced protein 8	enables cysteine-type endopeptidase inhibitor activity involved in the apoptotic process

URGCP	upregulator of cell proliferation	enables GTP binding
VEGFC	vascular endothelial growth factor C	enables chemoattractant activity, growth factor activity, and vascular endothelial growth factor receptor 3 binding

Table 7 continued.

WSCD2	WSC domain containing 2	enable sulfotransferase activity
ZMAT3	zinc finger matrin-type 3	enables nucleic acid binding enables zinc ion binding
ZMYND11	zinc finger MYND-type containing 11	enables DNA binding enables protein binding enables transcription corepressor activity
ZNF16	zinc finger protein 16	enables DNA-binding transcription factor activity RNA polymerase II-specific enables RNA polymerase II cis-regulatory region sequence-specific DNA binding
ZNF654	zinc finger protein 654	enables DNA binding and DNA-binding transcription factor activity, RNA polymerase II-specific
ZNF692	zinc finger protein 692	enables chromatin insulator sequence binding
ZNF713	zinc finger protein 713	enables DNA-binding transcription repressor activity, RNA polymerase II-specific

Discussion

The aim of the study was to use shallow sequenced whole genomes to identify adaptive traits or genes that may be responsible for the adaptation of the red foxes to the cold conditions in the Arctic, Norway. In addition, the study examined whether the data showed consistent patterns of genetic structuring and diversity as observed in previous studies. The data revealed high genetic diversity in the Arctic, which decreased with the environmental gradient to Israel which showed the least diversity. Weak genetic structuring was observed between the Arctic and Central Europe which could be a result of connectivity between the two areas. The genome data also revealed signatures of gene flow between Israel and Central Europe, confirming connectivity previously revealed by Rocha et al. (2023). Most importantly, the study identified numerous candidate genes potentially responsible for the continued survival of red foxes in the Arctic. A comprehensive account of the findings of this study is provided in the subsequent sections.

Diversity

Carnivores are among the most extensively studied species groups for a variety of reasons, including domestication, diseases, conservation and management concerns (Vonholdt and Driscoll, 2016; Hasselgren et al., 2018; Kukekova et al., 2018). Many of these species are threatened by habitat loss and fragmentation, leading to declining populations and a need for genetic rescue. For example, the Arctic fox (*V. lagopus*) is experiencing severe declines in Scandinavia as a result of the competition with larger competitor, the red foxes (*V. vulpes*) for food resources and habitat (Norén et al., 2017; Godoy et al., 2018; Hasselgren et al., 2018). Genetic diversity estimations are required to assess levels of genetic variation between and among populations. These estimations help to determine whether a species is at risk of inbreeding, which can lead to reduced fitness and inbreeding depression. Therefore, monitoring genetic diversity helps make decisions that help maintain genetic variation, ultimately preventing further population declines and extinction. Genetic diversity in red foxes has been widely studied across this species range. Magory Cohen et al. (2013), found high genetic diversity in red foxes from all over Israel, a trend also observed in European red foxes (Gachot-Neveu et al., 2009; Teacher et al., 2011; Galov et al., 2014; Mullins et al., 2014; Norén et al., 2015; Hagenlund et al., 2019). This high genetic diversity has been attributed to the species' high dispersal ability and gene-flow across populations. The red fox boasts a wide distribution worldwide with a recent colonisation into the Arctic. McDevitt et al. (2022), showed using evidence from GBS data that the foxes in Scandinavia had ancestry from Central and Eastern Europe. If the red foxes can still disperse into Scandinavia, then red foxes from the Arctic would be expected to have a high genetic diversity compared to Central Europe and Israel. Genome-wide heterozygosity estimations confirmed that red foxes from the Arctic have

the highest average heterozygosity. McDevitt et al. (2022), suggested several sources for the colonising population. However, we do know that foxes move considerably, so gene-flow could also be a possibility. The average heterozygosity of the red foxes from the Arctic was higher than that found in other carnivores. For instance Gopalakrishnan et al. (2018) found that canids of the genus *Canis* had heterozygosity estimates which ranged between 0.0021 and 0.0003. Similarly, the African wild dog (*Lycaon pictus*) showed values ranging from 0.0008 to 0.0012 (Armstrong et al., 2019). Within red fox populations, the individual heterozygosity estimates varied, with some Central European individuals exhibiting significantly lower heterozygosity than those from Israel and the Arctic (Table 3). The lower levels of heterozygosity observed in Israel and Central Europe as compared to Arctic could be the result of the north-ward range expansion in response to climate change, habitat fragmentation and resource scarcity (Sternberg et al., 2015; Gallant et al., 2020). That means fewer red foxes move back to Israel once they move north, reducing the diversity in the southern parts of their distribution while increasing the diversity northwards. This ultimately leads to a significant reduction in the population size southwards, allowing other forces of evolution, such as drift, to act significantly in these populations, thereby reducing diversity.

Filtering

As discussed in the method section, the `-setMinDepthInd` filter was used to discard individuals with a sequencing depth below a set threshold. This depth is usually calculated as the average depth of coverage of the population and depending on the quality of the data it can be set to 1 or more than 1. Prost et al. (2022) and Cavill et al. (2022) used low coverage whole genome sequences where the same filter was set to 3, whereas Sánchez-Barreiro et al. (2023) who studied museum specimens of black rhinoceros set this filter to 5. Similarly, in this study, this filter was applied at the thresholds of 1 and 3. This meant that the program would exclude all individuals with a sequencing depth below these values. Since the study utilised low-coverage genomic data, filters were applied to ensure that enough quality SNPs were retained after filtering so that the analyses could produce reliable and statistically significant results. In all cases, `-setMinDepthInd 3` filtered out a lot of the loci, which meant that it kept only a few. However, this strict filtering was potentially removing other loci that could be under selection. These would be relevant in studying adaptation. As such, `-setMinDepthInd 1` was also used, and this retained more loci. In the test for selection, the test with `-setMinDepthInd 3` retained 9,533 SNPs in the Arctic, 5,323 in Central Europe and 5,755 in Israel, whereas `-setMinDepthInd 1` retained 37,024 SNPs in the Arctic, 23,035 in Central Europe and 20,486 in Israel. The latter led to the detection of many genes, suggesting increasing sensitivity to detect weaker selection signals. Similarly, gene-flow analyses would also benefit from using more loci as these could detect even the weakest signature of gene-

flow. A total of 35,145 SNPs were used to detect gene flow, which also increased sensitivity to detect even weak signatures of gene flow. In admixture analyses, `-setMinDepthInd 1` retained 31,849 SNPs, whereas `-setMinDepthInd 3` retained 7,667 SNPs. Both PCA and admixture analyses revealed significant genetic structuring, dividing populations into three distinct populations and also indicated admixed individuals. Although `setMinDepthInd 3` provided a clearer subdivision of populations in the admixture plots, suggesting that the stricter filter could improve the resolution, by retaining more true SNPs and comprising less noise due to sequencing errors. In the phylogenetic analysis, `-setMinDepthInd 1` retained 28,731 SNPs but could not resolve the distinct population structure observed in PCAs and admixture analyses. This could be a result of gene-flow between the populations. A phylogenetic network was therefore used in this case with 11,586 SNPs and it produced results consistent with those observed in the PCA and admixture analyses.

Population structure

The genetic structure of red foxes has been intensively investigated across this species range. Even more so, its phylogeography has been studied using mitochondrial DNA, microsatellites, Y-chromosome, GBS and whole genome sequencing data. The present study was carried out with samples representing an environmental gradient from North (Norway), herein referred to as the Arctic, through Central Europe and to the Middle East, represented by Israel covering the distribution of up to three subspecies *V.v.vulpes* from Norway, *V.v.crucigera* from Central Europe and *V.v.palaestina* from Israel. Genetic structure analyses were carried out using model-based and model-free analysis. Both these analyses revealed a clear population subdivision where the populations were divided into three major groups. In the PCA, Israel was further broken down into two populations or clusters, which could suggest a substructure within Israel. On the other hand, the model-based admixture analysis (Figure 6), showed numerous individuals in Central Europe with traces of genetic composition similar to that of the Israeli samples. This could be evidence of gene-flow or shared ancestry between Central Europe and the populations from the Middle East. Rocha et al. (2023), used whole genome sequencing data, which showed high levels of admixture between Middle Eastern and European populations which supports gene-flow between Israel and Europe. The results revealed extensive genetic contact between Central Europe and the Arctic consistent with the findings from GBS data by Roberts (2019) and those by McDevitt et al. (2022) which revealed the Scandinavian red foxes, defined in their study as foxes from Norway and Sweden, originated from an admixture event from Russian and Central European populations around 5,360 years ago. The results observed in the present study then means that the genetic contact between Central European and Arctic populations could be evidence of shared common ancestry between Central Europe and Scandinavian populations.

PCAs at both depths provided evidence of substructuring within Israel, but only the admixture analysis at setMinDepthInd 3 confirmed this structure ($K=6$, lower panel, Figure 6). Substructure in Israel is consistent with Magory Cohen et al. (2013) who found that red foxes were divided into four subpopulations using microsatellite data. Given the conditions in Israel and the Negev and Judean deserts in particular, it is not unusual that the two regions would be structured. Human settlements and Highway 31 between the deserts could be potential causes of limited genetic contact between the two regions.

Both PCA and admixture results revealed patterns of genetic contact between the three populations, which could be historical or recent, as observed in other studies (Statham et al., 2018; Roberts, 2019; McDevitt et al., 2022; Rocha et al., 2023). In contrast, the phylogenetic trees were less informative about genetic structuring. However, the UPGMA tree (Figure 8) did group Central Europe and the Arctic as sister taxa, relative to Israel justifying the experimental setup of the PBS tests for selection. Furthermore, the UPGMA tree revealed substructuring in Israel, consistent with PCA and some admixture results (see above), and the ancient position of one of the Israel branches could also be a result of ancestral diversity (Statham et al., 2014; Sacks et al., 2018). Phylogenetic trees often fail to resolve population structuring in intraspecific data where there is extensive gene-flow (Slatkin and Maddison, 1989; Leaché et al., 2014). Given the long-range dispersal capabilities of red foxes (Walton et al., 2018), it is possible that gene-flow could have played a role in the lack of distinct topology. The results of the admixture analyses also revealed possible gene-flow between the populations. To get around this, a phylogenetic network was used because these were specially designed to reconstruct phylogenetic relationships in the presence of factors such as gene-flow, which would result in a non-tree-like or reticulate structure.

Phylogeography of the red fox

As discussed previously, red foxes in this study covered a wide range, and the results revealed population structure. PCA, admixture analyses and the phylogenetic network revealed two groups in Israel but limited structuring within Central Europe or the Arctic. These results also revealed extensive genetic contact between the Arctic and . This could potentially support evidence of two subspecies, with gene-flow due to factors such as recent divergence between the populations and lack of geographic isolation allowing the species to interbreed in the Arctic and Central Europe. However, high-coverage genomic data are still required to ascertain whether or not the red foxes in the Arctic are a different subspecies. Phylogeographic studies in Europe showed evidence of weak phylogeographic structuring attributed to high dispersal ability and adaptability across the populations (Teacher et al., 2011; Kutschera et al.,

2013; Statham et al., 2018; McDevitt et al., 2022). Teacher et al. (2011) and Edwards et al. (2012) previously suggested that red foxes existed as a single continuous population using mt(DNA). The evidence from GBS data supported the existence of a single continuous population (McDevitt et al., 2022). It is possible that individuals from Israel retained ancestral diversity (Statham et al., 2014; Sacks et al., 2018) and would then hybridise with other red foxes northwards during postglacial recolonisation, as seen in the traces of Israeli genetic contribution within the Central European individuals in the admixture graph Figure 6. The Israeli red foxes belong to a separate subspecies the *Vulpes v. palaestina* (Castelló, 2018). Indeed, our analyses showed a clear separation of the red foxes from Israel and individuals from Central Europe and the Arctic, with very few exceptions. For example, the Turkey individual could belong to the same subspecies given its close geographic distance to Israel. However, the results from this study, revealed separate populations within Israel. This could be due to differing conditions in the sampled areas of the Negev and Judean deserts. As previously discussed, the movement of red foxes between these two localities could be restricted by human settlements and Highway 31 between the deserts.

Signatures of gene-flow

To test whether these signals of shared ancestry and gene-flow indeed existed, an ABBA-BABA statistic test was conducted following the tree topology revealed in the UPGMA tree and in the phylogenetic network. It has been established that red foxes are capable of long-range dispersal, and although dispersal does not translate to genetic exchange or gene-flow, the populations should show evidence of gene-flow even if it is weak. To do this, there was a need to leverage as many SNPs as possible because the data used in this study is very low coverage (between 0.0135X to 0.8626X). Low coverage can result in significant amounts of missing data that can affect the test's statistical power. To ensure this does not occur, a large sample size was required, and our sample size was 86 samples after the exclusion of closely related individuals, which means the sample was still large enough to get reliable estimates of gene-flow (Simonsen et al., 1995; Li et al., 2011). The choice of filters can still affect the test and having seen that the depth filter removed many SNPs when set to 3, this analysis was computed using -setMinDepthInd 1 to keep more SNPs, which could be instrumental in detecting signatures of gene-flow even if they are weak. It was not possible to detect gene-flow, even though it was probably extensive, between the Arctic and Central Europe, because these were sister populations, and the ABBA-BABA test is not set up to resolve this. However, the ABBA-BABA tests revealed evidence of gene-flow between Israel and Central Europe, shown in the first row in Table 3, as a result of the red foxes' extensive dispersal ability. A similar finding was observed between Europe (Iberian Peninsula) and the

Middle East (Rocha et al., 2023). This suggests that our data were able to capture reliable estimates of gene-flow, despite the inclusion of potentially larger numbers of sequencing errors.

Signatures of local adaptation

The Arctic presents novel conditions for red foxes, characterised by long, dark, cold winters and short summers with continuous sunlight and resource scarcity (Blix, 2016). Red foxes likely dispersed into northern latitudes during the 19th or 20th century due to rapid changes in Arctic climate (Chautan et al., 2000; Selås and Vik, 2006; Elmhagen et al., 2017; Rantanen et al., 2022; Warret Rodrigues, 2023). They were first recorded in the Canadian Arctic between 1926 and 1950, in the Yamal Peninsula, Russia in 1913 and again between 1974–1986 (Chautan et al., 2000; Selås and Vik, 2006; Rodnikova et al., 2011; Gallant et al., 2012; Elmhagen et al., 2017). The new Arctic environment would present unique selective pressures that may require red foxes or any species colonising a new environment to evolve mutations to help them survive. To survive the Arctic conditions, species must exhibit behavioural, physiological and physical adaptations. These would include balling up, huddling, seasonal fur changes for insulation, plumage and blubber/fat deposition and mechanisms that help preserve core temperature by keeping very low temperatures in the extremities (Blix, 2016). Red foxes inhabiting the Arctic do not appear to have these physical adaptations. They are relatively large animals, causing them to have high energy requirements compared to the Arctic fox. They also have less fur, meaning poorer insulation (Hersteinsson and Macdonald, 1992). The Arctic fox has 50% more underfur than red foxes, they also have a low critical temperature, which is the temperature below which the Arctic fox would have to increase metabolic heat production to maintain the average core temperature (Hersteinsson and Macdonald, 1992). PBS tests for local adaptation were conducted to investigate whether the continued persistence of red foxes in the Arctic, despite the lack of physical adaptations, was influenced by positive selection. This test searched for regions in the Arctic populations with rapid allele frequency changes compared to sister and outgroup populations, which would be the regions where positive selection is likely to occur. Genes involved in metabolism, olfaction, hypoxia, thermogenesis, nervous system development, disease, metal ion binding, cell surface receptor signalling pathways, fertility, and immune response were identified (Tables 5 and 6). Given the low coverage of our genome data, it is possible that some SNPs are attributable to sequencing errors, especially at -setMinDepthInd 1, resulting in false positive signatures of selection. To get around this, I only discuss the biological significances for detected genes, when they were also found in other studies. Previous studies have found genes responsible for cold adaptations in Yakutian horses (Librado et al., 2015), polar bears

(Liu et al., 2014) woolly mammoths (Lynch et al., 2015), minke whale (Yim et al., 2014; Park et al., 2015) and reindeer and muskox (Li et al., 2023). Yudin et al. (2017) made a compilation and functional characterisation of all mammalian genes involved in adaptation to the Arctic and Antarctic environments. Of all compiled genes identified to be under selection, the present study found three genes under positive selection: AGA, FAS, and NEIL3 (minke whales). Likewise, six genes found to be under positive selection in polar bears were also candidates for positive selection in the red foxes in the present study: ADCY6 is an ortholog to ADCY1. In the present study, ADCY 1 was identified as a candidate for selection, and other genes such as ACTA2, KCNMB3, MOGAT1, MED13 and SERPINB13 were under selection in polar bears. Lastly, nine genes were found to have fixed, derived amino acids and loss of function substitutions in woolly mammoths compared to Asian elephants, and these included ACSL3, ADAM20, ADCY6, an ortholog to ADCY1, DBNL, HSD17B4, KCNMB3, MYT1, PTCHD3 and SLC23A1. Here, the genes identified above will be discussed in more detail. *Candidate genes linked to disease*

AGA has been found to play a key role in the last stage of glycoproteins degradation, mutation of this gene causes a heritable lysosome storage disorder, aspartylglucosaminuria and is associated with changes in the facial skin, dental arches and occlusion (Park et al., 2015). Neil 3, involved in the repair of mutagenic bases, overexpression of this gene indicates the development and progression of cancer (Chen et al., 2022; Holm et al., 2023). FAS plays a central role in the physiological regulation of cell apoptosis and has been implicated in the pathogenesis of various malignancies and diseases of the immune system (Strasser et al., 2009). Another gene with similar implication is the TLE4, this gene has been associated with development and progression of human colorectal cancers (Wang et al., 2015). MS4A15 acts as an oncogene in ovarian cancer (Fang et al., 2022). Cancer causes up to 75% of deaths in juvenile red foxes (Roberts, 2019). Cadherin genes such as CDHR2 act as tumour suppressors (Xia et al., 2019). Roberts (2019) showed that faulty cadherin genes could lose their cell-cell adhesive properties, which may lead to the formation of tumours as cells detach from other surrounding cells. These genes may play important roles in epigenetics as these suggest that foxes living close to human settlements may have been exposed to chemicals such as pesticides which could have exerted selective pressure on DNA repair genes (Roberts, 2019).

Signal transduction

Likewise, six genes found to be under positive selection in polar bears were also candidates for positive selection in the red foxes in this study, including ADCY6 which is an ortholog to ADCY1. In the present study, ADCY1 was identified as a candidate for selection, this gene is

expressed in the brain. It is involved in cAMP biosynthetic process and intracellular signal transduction. Cyclic AMP production is crucial mechanotransduction within the inner ear and is catalysed by adenylate cyclases (Santos-Cortez et al., 2014). Red foxes can hear lower frequencies, and ADCY1 has been identified as a key hearing gene in bats (Li et al., 2024). Roberts (2019) showed that red foxes hear a lower frequency this is crucial for the red fox since it has to hunt for food under ice sheets to survive in the Arctic.

Homeostasis

ACTA2 is involved in vasoconstriction and blood pressure homeostasis. This suggests that red foxes in the Arctic need to minimise heat loss by diverting blood flow from the periphery of the body to maintain a stable internal temperature. KCNMB3 is involved in calcium-activated potassium channel activity and potassium channel regulator activity.

Metabolism

Lipids are essential in cold adaptation, so finding genes involved in lipid metabolism in the red foxes from the Arctic suggests that red foxes can oxidise lipids to generate energy during periods of food scarcity. The Arctic fox is a well-adapted species that developed a unique lipid metabolism (Zhu et al., 2024) that allows it to eat a high-fat diet to increase its body mass through the accumulation of adipose tissue when food is abundant. Amongst other genes NPC1L1, facilitates the uptake of sterols into the enterocyte (Valasek et al., 2008; Xiao et al., 2023). Since this gene, and those discussed below, were identified as being under positive selection, it suggests that for red foxes in the Arctic are adapting to a fatty-acids rich diet in this region. Increased intake of fatty-acid rich foods means more cholesterol is available to maintain membrane fluidity in then cells during winter and also some of the stored fats could provide insulation. MOGAT1, catalyses the synthesis of diacylglycerols. Diacylglycerols are precursors of physiologically crucial lipids such as triacylglycerols and phospholipids (Yen et al., 2002). This gene has been associated with the absorption of dietary fat because the resynthesis of triacylglycerol is required for the assembly of lipoproteins that transport absorbed fat to other tissues (Yen et al., 2002). LIPA is involved in sterol metabolic process, where its primary functions is to regulate cholesterol levels (Li and Zhang, 2019). Another gene ACSL3 plays in synthesis of fatty acyl-CoA esters and regulation of lipid droplet location and fatty acid uptake (Quan et al., 2021). CMKLR1, has been associated with the synthesis adipose tissue and metabolism (Goralski et al., 2007; Ernst et al., 2012). Adipose tissues are primarily fat storage tissues however they also have regulate fatty acid metabolism for energy provision, release of hormones to control food intake (Coelho et al., 2013). HSD17B4, a multifunctional peroxisomal enzyme involved in fatty acid beta-oxidation and steroid metabolism, mutation on this gene can cause D-bifunctional protein deficiency causes an autosomal-recessive disorder of peroxisomal fatty acid beta-oxidation (Pierce et al., 2010).

MS4A15 play an important role in regulating anti-ferroptotic lipid reservoirs which provides a crucial resistance mechanism that is distinct from antioxidant and lipid detoxification pathways (Xin et al., 2022). This means in Arctic where red foxes would face extremely cold conditions, genes such as the MS4A15 would be crucial in preventing cell death which allows continued survival of red foxes in the Arctic.

Immunity

Red foxes are reservoirs for many pathogens such as rabies, canine distemper, *Babesia vulpes* and *Echinococcus multilocularis* (Kukekova et al., 2018; Mierzejewska et al., 2021; Celva, 2022). Several genes for immunity have been identified, these include RAB18, CH25H, LOC112914295, LOC112911121, LOC112911116 and CMKLR1 a chemerin receptor with both innate and adaptive immunity. RAB18 plays a crucial role in modulating autophagy and maintaining cellular proteostasis (Feldmann et al., 2017; Dejgaard and Presley, 2019). Serpins also play a huge role in vertebrate immunity, SERPINB10,11,13,2,3 and 7 were identified as candidates of selection. Mangan et al. (2008), outlined the roles of these genes in the immune system, and they include migration, phagocytosis and elimination of cells that are cancerous or infected by viruses.

Olfaction

Red foxes depend on their olfactory receptors for hunting, evading predators and also communication (McLean et al., 2021). Redfox have complex and unique olfactory receptors (Ortiz-Leal et al., 2023). Olfactory receptor genes identified in red foxes included LOC112912474, LOC112912081 and LOC112912082.

Hypoxia

Several genes for hypoxia were identified including NR4A2, which regulates genes involved in oxidative stress, and inflammation. VEGFC stimulates the growth of blood vessels to help transport blood to cells and tissues experiencing low oxygen levels (Tamura et al., 2020), and LOC112927226 is L-lactate dehydrogenase B chain-like, this gene facilitates the conversion of pyruvate into lactate during periods low oxygen (Farhana and Lappin, 2023).

Gene expression

Transcription factors required to regulate gene expression, genes associated with DNA binding transcription factor activity were identified as candidates for selection LOC112911121, LOC112909788, MYT1, ZNF16 among these were genes associated with transcription repressor activity LOC112911116, ZNF713, encode for protein which contain C2H2 zinc finger domains, which have been associated with regulation of cold stress response in plants (Zhang et al., 2022). Several protein genes were also identified these included CCDC7, FICD, LOC112912790, LOC112921570, SYNJ2BP, TCHP, FAS, TLE4.

Circadian rhythm

Changes in the length of days in the Arctic could also affect circadian rhythm in red foxes in the Arctic, candidate gene for circadian rhythm were identified as circadian ADCY1 and KCNJ3. ADCY1 gene was enriched in circadian entertainment in Tibetan black chicken (Feng et al., 2023).

Fertility

The long cold periods could also influence reproduction in the arctic FAM170A, LOC112911081, PIK3CA, ADAM20, ADAM21, LOC112927152, and NPAP1 were gene associated with fertility identified as candidates of selection. FAM170A is crucial in sperm development, absence of this gene has been linked to severe male subfertility characterized by abnormal sperm head morphology, defective spermiation, and reduced sperm progressive motility (Devlin et al., 2020).

Conclusion

This study initially predicted that red foxes in the Arctic and those from Israel would show high genetic diversity due to increased gene flow into the Arctic and Israel being the cradle of genetic diversity of red foxes (Statham et al., 2014; Rocha et al., 2023). In addition, the Arctic show a low genetic diversity because it experienced a bottleneck. The results showed that Arctic has high genetic diversity and Israel showed lower genetic diversity than all populations under study. It was further predicted that there would be gene flow between the populations and weak signatures of genetic structuring between Arctic and Central Europe. Considering that the red fox has been linked with considerable damage, a vector for zoonotic diseases, which could result in a health risk for humans. Proper management of this species is required, which does not threaten its genetic diversity because the same species is also linked to beneficial ecosystem services as a game species, fur-bearer, and natural pest controller. Moritz (1994) defined an evolutionary significant unit as a population or populations which are reciprocally monophyletic and have significant divergence at nuclear loci. This means that populations have to show significant distinctiveness with no genetic exchange. The results in this study showed that the populations had extensive genetic contact, such as gene-flow, and phylogenetic trees failed to identify monophyletic groups. Selection test showed that red foxes have evolved unique adaptations to survive the Arctic environment, as initially predicted. Genes involved in important process such as lipid metabolism and response to hypoxia were identified in this study which showed that red foxes may be adapting to their Arctic environments.

Low-coverage genomic data have a number of challenges, primarily due to huge amounts missing data. As such the use of tools designed to analyse such data was imperative to getting

the most reliable results. The data was analysed using the analysis of next generation sequencing data (ANGSD) software which instead of calling genotypes it uses genotype likelihood to account for uncertainty (Korneliussen et al., 2014). Using tools like ANGSD does not get rid of all issues, filtering options have to be carefully chosen to maximise chances of getting meaningful results. In the present study, most tests that were run using two depth filters -setMinDepthInd 1 and 3 which produced similar results, but with the latter showing higher resolution in some cases (Figures 4 and 6). PCA and admixture analyses both separated populations and showed admixed individuals clustering closer to their parent populations. Gene-flow patterns were consistent among analyses and with previous studies. Importantly, genes influencing adaptation to the Arctic were also identified showing that the genome data used in the present study was reasonably sufficient to test the hypotheses set out in the introduction. However, high coverage genomic data are still required to test and corroborate these findings. High coverage data would result in a larger number of more reliable SNPs, which would lead to fewer false negatives and a greater resolution of evolutionary processes. High-coverage data also have high sequencing quality and have way less missing data, which is crucial in distinguishing true variants from sequencing errors.

References

- Aho, A.V., Kernighan, B.W., Weinberger, P.J., 1979. Awk—a pattern scanning and processing language. *Software: Practice and Experience* 9, 267-279.
- Alachiotis, N., Pavlidis, P., 2018. RAiSD detects positive selection based on multiple signatures of a selective sweep and SNP vectors. *Communications biology* 1, 79.
- Alexander, D.H., Novembre, J., Lange, K., 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome research* 19, 1655-1664.
- Allen, S.H., Sargeant, A.B., 1993. Dispersal patterns of red foxes relative to population density. *The Journal of Wildlife Management*, 526-533.
- Allendorf, F.W., Funk, W.C., Aitken, S.N., Byrne, M., Luikart, G., Antunes, A., 2022. *Conservation and the genomics of populations*. Third edition. ed. Oxford University Press, Oxford.
- Armstrong, E.E., Taylor, R.W., Prost, S., Blinston, P., van der Meer, E., Madzikanda, H., Mufute, O., Mandisodza-Chikerema, R., Stuelpnagel, J., Sillero-Zubiri, C., 2019. Cost-effective assembly of the African wild dog (*Lycaon pictus*) genome using linked reads. *Gigascience* 8, giy124.
- Aubry, K.B., Statham, M.J., Sacks, B.N., Perrine, J.D., Wisely, S.M., 2009. Phylogeography of the North American red fox: vicariance in Pleistocene forest refugia. *Molecular Ecology* 18, 2668-2686.
- Bailey, S.F., Alonso Morales, L.A., Kassen, R., 2021. Effects of Synonymous Mutations beyond Codon Bias: The Evidence for Adaptive Synonymous Substitutions from Microbial Evolution Experiments. *Genome Biol Evol* 13.
- Bartolini Lucenti, S., Madurell-Malapeira, J., 2021. Ancestral foxes at the gates of Europe: the Pliocene fox from Çalta-1 (Turkey) and their relationships with Asian and European Plio-Pleistocene foxes. *Comptes Rendus Palevol* 20, 619-626.
- Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)* 57, 289-300.
- Bertorelle, G., Excoffier, L., 1998. Inferring admixture proportions from molecular data. *Molecular Biology and Evolution* 15, 1298-1311.
- Blix, A.S., 2016. Adaptations to polar life in mammals and birds. *Journal of Experimental Biology* 219, 1093-1105.
- Bourgeois, Y.X., Warren, B.H., 2021. An overview of current population genomics methods for the analysis of whole-genome resequencing data in eukaryotes. *Molecular Ecology* 30, 6036-6071.

- Brown, G.R., Hem, V., Katz, K.S., Ovetsky, M., Wallin, C., Ermolaeva, O., Tolstoy, I., Tatusova, T., Pruitt, K.D., Maglott, D.R., 2015. Gene: a gene-centered information resource at NCBI. *Nucleic acids research* 43, D36-D42.
- Castelló, J.R., 2018. *Canids of the world: wolves, wild dogs, foxes, jackals, coyotes, and their relatives*. Princeton University Press.
- Cavill, E.L., Gopalakrishnan, S., Puetz, L.C., Ribeiro, Â.M., Mak, S.S., da Fonseca, R.R., Pacheco, G., Dunlop, B., Accouche, W., Shah, N., 2022. Conservation genomics of the endangered Seychelles Magpie-Robin (*Copsychus sechellarum*): A unique insight into the history of a precious endemic bird. *Ibis* 164, 396-410.
- Celva, R., 2022. Assessing red fox (*Vulpes vulpes*) demographics to monitor wildlife diseases: A spotlight on *Echinococcus multilocularis*. *Pathogens*. 12 (1), 60.
- Chautan, M., Pontier, D., Artois, M., 2000. Role of rabies in recent demographic changes in red fox (*Vulpes vulpes*) populations in Europe. *Mammalia* 64, 391-410.
- Chen, L., Huan, X., Gao, X.-D., Yu, W.-H., Xiao, G.-H., Li, T.-F., Wang, Z.-Y., Zhang, Y.-C., 2022. Biological functions of the DNA glycosylase NEIL3 and its role in disease progression including cancer. *Cancers* 14, 5722.
- Coelho, M., Oliveira, T., Fernandes, R., 2013. State of the art paper Biochemistry of adipose tissue: an endocrine organ. *Archives of medical science* 9, 191-200.
- Coleman, J.R., Euesden, J., Patel, H., Folarin, A.A., Newhouse, S., Breen, G., 2016. Quality control, imputation and analysis of genome-wide genotyping data from the Illumina HumanCoreExome microarray. *Briefings in functional genomics* 15, 298-304.
- Darwin, C., 1859. *The origin of species by means of natural selection*. EA Weeks.
- DeGiorgio, M., Huber, C.D., Hubisz, M.J., Hellmann, I., Nielsen, R., 2016. SweepFinder2: increased sensitivity, robustness and flexibility. *Bioinformatics* 32, 1895-1897.
- Dejgaard, S.Y., Presley, J.F., 2019. Rab18: new insights into the function of an essential protein. *Cellular and Molecular Life Sciences* 76, 1935-1945.
- Devlin, D.J., Nozawa, K., Ikawa, M., Matzuk, M.M., 2020. Knockout of family with sequence similarity 170 member A (Fam170a) causes male subfertility, while Fam170b is dispensable in mice. *Biology of Reproduction* 103, 205-222.
- Drummond, A.J., Xie, W., Heled, J., 2012. Bayesian inference of species trees from multilocus data using* BEAST. *Molecular Biology and Evolution* 29, 1969-1973.
- Dutheil, J.Y., 2020. *Statistical population genomics*. Springer Nature.
- Edwards, C.J., Soulsbury, C.D., Statham, M.J., Ho, S.Y., Wall, D., Dolf, G., Iossa, G., Baker, P.J., Harris, S., Sacks, B.N., 2012. Temporal genetic variation of the red fox, *Vulpes vulpes*, across western Europe and the British Isles. *Quaternary science reviews* 57, 95-104.

- Egea, R., Casillas, S., Barbadilla, A., 2008. Standard and generalized McDonald-Kreitman test: a website to detect selection by comparing different classes of DNA sites. *Nucleic Acids Res* 36, W157-162.
- Elhaik, E., 2022. Principal Component Analyses (PCA)-based findings in population genetic studies are highly biased and must be reevaluated. *Scientific Reports* 12, 14683.
- Ellegren, H., Galtier, N., 2016. Determinants of genetic diversity. *Nature Reviews Genetics* 17, 422-433.
- Elmhagen, B., Berteaux, D., Burgess, R.M., Ehrich, D., Gallant, D., Henttonen, H., Ims, R.A., Killengreen, S.T., Niemimaa, J., Norén, K., 2017. Homage to Hersteinsson and Macdonald: climate warming and resource subsidies cause red fox range expansion and Arctic fox decline. *Polar Research* 36, 3.
- Ernst, M.C., Haidl, I.D., Zúñiga, L.A., Dranse, H.J., Rourke, J.L., Zabel, B.A., Butcher, E.C., Sinal, C.J., 2012. Disruption of the chemokine-like receptor-1 (CMKLR1) gene is associated with reduced adiposity and glucose intolerance. *Endocrinology* 153, 672-682.
- Evanno, G., Regnaut, S., Goudet, J., 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular ecology* 14, 2611-2620.
- Ewels, P., Magnusson, M., Lundin, S., Käller, M., 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32, 3047-3048.
- Fang, Y., Yu, H., Zhou, H., 2022. MS4A15 acts as an oncogene in ovarian cancer through reprogramming energy metabolism. *Biochemical and Biophysical Research Communications* 598, 47-54.
- Farhana, A., Lappin, S.L., 2023. Biochemistry, lactate dehydrogenase, StatPearls [internet], StatPearls Publishing.
- Feldmann, A., Bekbulat, F., Huesmann, H., Ulbrich, S., Tatzelt, J., Behl, C., Kern, A., 2017. The RAB GTPase RAB18 modulates macroautophagy and proteostasis. *Biochemical and biophysical research communications* 486, 738-743.
- Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39, 783-791.
- Felsenstein, J., 1993. PHYLIP (phylogeny inference package), version 3.5 c. Joseph Felsenstein.
- Feng, J., Zhu, W., Shi, H., Peng, D., Zang, L., Wang, Y., ZhaXi, L., BaiMa, J., Amevor, F.K., Wang, X., 2023. Analysis of the Selection Signal of the Tibetan Black Chicken Genome Based on Whole-Genome Sequencing. *Genes* 14, 1672.
- Figueirido, B., Martín-Serra, A., Tseng, Z., Janis, C., 2015. Habitat changes and changing predatory habits in North American fossil canids. *Nature Communications* 6, 7976.

- Footo, A.D., Vijay, N., Ávila-Arcos, M.C., Baird, R.W., Durban, J.W., Fumagalli, M., Gibbs, R.A., Hanson, M.B., Komeliussen, T.S., Martin, M.D., 2016. Genome-culture coevolution promotes rapid divergence of killer whale ecotypes. *Nature communications* 7, 11693.
- Frati, F., Hartl, G., Lovari, S., Delibes, M., Markov, G., 1998. Quaternary radiation and genetic structure of the red fox *Vulpes vulpes* in the Mediterranean Basin, as revealed by allozymes and mitochondrial DNA. *Journal of Zoology* 245, 43-51.
- Gachot-Neveu, H., Lefevre, P., Roeder, J.-J., Henry, C., Poulle, M.-L., 2009. Genetic detection of sex-biased and age-biased dispersal in a population of wild carnivore, the red fox, (*Vulpes vulpes*). *Zoological science* 26, 145-152.
- Gallant, D., Lecomte, N., Berteaux, D., 2020. Disentangling the relative influences of global drivers of change in biodiversity: A study of the twentieth-century red fox expansion into the Canadian Arctic. *Journal of Animal Ecology* 89, 565-576.
- Gallant, D., Reid, D.G., Slough, B.G., Berteaux, D., 2014. Natal den selection by sympatric arctic and red foxes on Herschel Island, Yukon, Canada. *Polar Biology* 37, 333-345.
- Gallant, D., Slough, B.G., Reid, D.G., Berteaux, D., 2012. Arctic fox versus red fox in the warming Arctic: four decades of den surveys in north Yukon. *Polar biology* 35, 1421-1431.
- Galov, A., Sindičić, M., Andreanszky, T., Čurković, S., Deždek, D., Slavica, A., Hartl, G.B., Krueger, B., 2014. High genetic diversity and low population structure in red foxes (*Vulpes vulpes*) from Croatia. *Mammalian Biology* 79, 77-80.
- Geffen, E., Girard, I., 2003. Behavioral and physiological adaptations of foxes to hot arid environments: comparing Saharo-Arabian and north American species, in: Marshan A. Sovad, L.C. (Ed.), *The swift fox: Ecology and conservation of swift foxes in a changing world*, Canadian Plains Research Center, pp. 223-229.
- Godoy, E., Norén, K., Angerbjörn, A., 2018. Mating patterns in an inbred Arctic carnivore. *Polar Biology* 41, 945-951.
- Gomo, G., Mattisson, J., Rød-Eriksen, L., Eide, N.E., Odden, M., 2021. Spatiotemporal patterns of red fox scavenging in forest and tundra: the influence of prey fluctuations and winter conditions. *Mammal Research* 66, 257-265.
- Gopalakrishnan, S., Sinding, M.-H.S., Ramos-Madrugal, J., Niemann, J., Castruita, J.A.S., Vieira, F.G., Carøe, C., de Manuel Montero, M., Kuderna, L., Serres, A., 2018. Interspecific gene flow shaped the evolution of the genus *Canis*. *Current Biology* 28, 3441-3449. e3445.
- Goralski, K.B., McCarthy, T.C., Hanniman, E.A., Zabel, B.A., Butcher, E.C., Parlee, S.D., Muruganandan, S., Sinal, C.J., 2007. Chemerin, a novel adipokine that regulates adipogenesis and adipocyte metabolism. *Journal of Biological Chemistry* 282, 28175-28188.
- Gross, E., van Iersel, L., Janssen, R., Jones, M., Long, C., Murakami, Y., 2021. Distinguishing level-1 phylogenetic networks on the basis of data generated by Markov processes. *Journal of Mathematical Biology* 83, 1-24.

- Hagenlund, M., Linløkken, A., Østbye, K., Walton, Z., Odden, M., Samelius, G., Willebrand, T., Wilson, R., 2019. Genetic structure and gene flow in red foxes (*Vulpes vulpes*) in Scandinavia: Implications for the potential future spread of *Echinococcus multilocularis* tapeworm. *Applied Sciences* 9, 5289.
- Hague, M.T., Routman, E.J., 2016. Does population size affect genetic diversity? A test with sympatric lizard species. *Heredity* 116, 92-98.
- Hahn, M.W., 2019. Molecular population genetics. Sinauer Associates : Oxford University Press, New York.
- Hanghøj, K., Moltke, I., Andersen, P.A., Manica, A., Korneliussen, T.S., 2019. Fast and accurate relatedness estimation from high-throughput sequencing data in the presence of inbreeding. *Gigascience* 8, giz034.
- Hao, Z., Liang, X., Li, G., 2022. Quality control and preprocessing of sequencing reads. *BIO-PROTOCOL* 12, 13.
- Hartl, D.L., 2020. A primer of population genetics and genomics. Oxford University Press.
- Hartl, D.L., Jones, E.W., 1998. Genetics: principles and analysis. Jones and Barlett Publishers.
- Hasselgren, M., Angerbjörn, A., Eide, N.E., Erlandsson, R., Flagstad, Ø., Landa, A., Wallén, J., Norén, K., 2018. Genetic rescue in an inbred Arctic fox (*Vulpes lagopus*) population. *Proceedings of the Royal Society B: Biological Sciences* 285, 20172814.
- He, B., Zhu, R., Yang, H., Lu, Q., Wang, W., Song, L., Sun, X., Zhang, G., Li, S., Yang, J., 2020. Assessing the impact of data preprocessing on analyzing next generation sequencing data. *Frontiers in Bioengineering and Biotechnology* 8, 817.
- Hemstrom, W., Grummer, J.A., Luikart, G., Christie, M.R., 2024. Next-generation data filtering in the genomics era. *Nature Reviews Genetics* 25, 750-767.
- Hersteinsson, P., Macdonald, D.W., 1992. Interspecific competition and the geographical distribution of red and arctic foxes (*Vulpes vulpes*) and (*Alopex lagopus*). *Oikos*, 505-515.
- Hewitt, G.M., 1999. Post-glacial re-colonization of European biota. *Biological journal of the Linnean Society* 68, 87-112.
- Hoffmann, M., Sillero-Zubiri, C., 2016. *Vulpes vulpes*. The IUCN red list of threatened species 2016: e. T23062A46190249. IUCN, <http://dx.doi.org/10.2305/IUCN.UK.2016>.
- Holm, S., Larsen, R.M., Holst, C.M., Heide-Jørgensen, M.P., Steffensen, J.F., Stevnsner, T., Larsen, K., 2023. Bowhead NEIL1: Molecular cloning, characterization, and enzymatic properties. *Biochimie* 206, 136-149.
- Horn, S., Prost, S., Stiller, M., Makowiecki, D., Kuznetsova, T., Benecke, N., Pucher, E., Hufthammer, A.K., Schouwenburg, C., Shapiro, B., 2014. Ancient mitochondrial DNA and the genetic history of Eurasian beaver (*Castor fiber*) in Europe. *Molecular Ecology* 23, 1717-1729.

- Howard, C., Marjakangas, E.-L., Morán-Ordóñez, A., Milanese, P., Abuladze, A., Aghababayan, K., Ajder, V., Arkumarev, V., Balmer, D.E., Bauer, H.-G., 2023. Local colonisations and extinctions of European birds are poorly explained by changes in climate suitability. *Nature Communications* 14, 4304.
- Hu, T., Chitnis, N., Monos, D., Dinh, A., 2021. Next-generation sequencing technologies: An overview. *Human Immunology* 82, 801-811.
- Huang, D.W., Sherman, B.T., Lempicki, R.A., 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* 4, 44-57.
- Huson, D.H., Bryant, D., 2005. Application of Phylogenetic Networks in Evolutionary Studies. *Molecular Biology and Evolution* 23, 254-267.
- Huson, D.H., Bryant, D., 2024. The SplitsTree App: interactive analysis and visualization using phylogenetic trees and networks. *Nature methods*, 1-2.
- InkscapeProject, 2020. Inkscape.
- Inoue, T., Nonaka, N., Mizuno, A., Morishima, Y., Sato, H., Katakura, K., Oku, Y., 2007. Mitochondrial DNA phylogeography of the red fox (*Vulpes vulpes*) in northern Japan. *Zoological Science* 24, 1178-1186.
- Jombart, T., Collins, C., 2015. Analysing genome-wide SNP data using adegenet 2.0. 0.
- Kanaka, K., Sukhija, N., Goli, R.C., Singh, S., Ganguly, I., Dixit, S., Dash, A., Malik, A.A., 2023. On the concepts and measures of diversity in the genomics era. *Current Plant Biology* 33, 100278.
- Kern, A.D., Hahn, M.W., 2018. The neutral theory in light of natural selection. *Molecular biology and evolution* 35, 1366-1371.
- Kimura, M., 1977. Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. *Nature* 267, 275-276.
- Kopelman, N.M., Mayzel, J., Jakobsson, M., Rosenberg, N.A., Mayrose, I., 2015. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular ecology resources* 15, 1179-1191.
- Korneliussen, T.S., Albrechtsen, A., Nielsen, R., 2014. ANGSD: analysis of next generation sequencing data. *BMC bioinformatics* 15, 1-13.
- Koropoulis, A., Alachiotis, N., Pavlidis, P., 2020. Detecting Positive Selection in Populations Using Genetic Data. *Methods Mol Biol* 2090, 87-123.
- Kozlov, A.M., Darriba, D., Flouri, T., Morel, B., Stamatakis, A., 2019. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* 35, 4453-4455.
- Kukekova, A.V., Johnson, J.L., Xiang, X., Feng, S., Liu, S., Rando, H.M., Kharlamova, A.V., Herbeck, Y., Serdyukova, N.A., Xiong, Z., 2018. Red fox genome assembly identifies

genomic regions associated with tame and aggressive behaviours. *Nature ecology & evolution* 2, 1479-1491.

Kumar, V., Kutschera, V.E., Nilsson, M.A., Janke, A., 2015. Genetic signatures of adaptation revealed from transcriptome sequencing of Arctic and red foxes. *BMC genomics* 16, 1-13.

Kutschera, V.E., Lecomte, N., Janke, A., Selva, N., Sokolov, A.A., Haun, T., Steyer, K., Nowak, C., Hailer, F., 2013. A range-wide synthesis and timeline for phylogeographic events in the red fox (*Vulpes vulpes*). *BMC evolutionary biology* 13, 1-15.

Lai, S., Warret Rodrigues, C., Gallant, D., Roth, J.D., Berteaux, D., 2022. Red foxes at their northern edge: competition with the Arctic fox and winter movements. *Journal of Mammalogy* 103, 586-597.

Larivière, S., Pasitschniak-Arts, M., 1996. *Vulpes vulpes*. *Mammalian species*, 1-11.

Leaché, A.D., Harris, R.B., Rannala, B., Yang, Z., 2014. The influence of gene flow on species tree estimation: a simulation study. *Systematic biology* 63, 17-30.

Leipzig, J., 2017. A review of bioinformatic pipeline frameworks. *Brief Bioinform* 18, 530-536.

Lewis, P.O., Holder, M.T., Swofford, D.L., 2015. Phycas: software for Bayesian phylogenetic analysis. *Systematic Biology* 64, 525-531.

Li, F., Zhang, H., 2019. Lysosomal acid lipase in lipid metabolism and beyond.

Arteriosclerosis, thrombosis, and vascular biology 39, 850-856.

Li, H., 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv preprint arXiv:1303.3997*.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009a. The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078-2079.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., Subgroup, G.P.D.P., 2009b. The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078-2079.

Li, M., Li, X., Wu, Z., Zhang, G., Wang, N., Dou, M., Liu, S., Yang, C., Meng, G., Sun, H., 2023. Convergent molecular evolution of thermogenesis and circadian rhythm in Arctic ruminants. *Proceedings of the Royal Society B* 290, 20230538.

Li, X., Wang, H., Wang, X., Bao, M., Sun, R., Dai, W., Sun, K., Feng, J., 2024. Molecular adaptations underlying high-frequency hearing in the brain of CF bats species. *BMC genomics* 25, 279.

Li, Y., Sidore, C., Kang, H.M., Boehnke, M., Abecasis, G.R., 2011. Low-coverage sequencing: implications for design of complex trait association studies. *Genome research* 21, 940-951.

Librado, P., Der Sarkissian, C., Ermini, L., Schubert, M., Jónsson, H., Albrechtsen, A., Fumagalli, M., Yang, M.A., Gamba, C., Seguin-Orlando, A., 2015. Tracking the origins of

- Yakutian horses and the genetic basis for their fast adaptation to subarctic environments. *Proceedings of the National Academy of Sciences* 112, E6889-E6897.
- Liu, S., Lorenzen, E.D., Fumagalli, M., Li, B., Harris, K., Xiong, Z., Zhou, L., Korneliussen, T.S., Somel, M., Babbitt, C., 2014. Population genomics reveal recent speciation and rapid evolutionary adaptation in polar bears. *Cell* 157, 785-794.
- Lonsinger, R.C., Adams, J.R., Waits, L.P., 2018. Evaluating effective population size and genetic diversity of a declining kit fox population using contemporary and historical specimens. *Ecology and Evolution* 8, 12011-12021.
- López, A., Bonasora, M.G., 2017. Phylogeography, genetic diversity and population structure in a Patagonian endemic plant. *AoB Plants* 9, plx017.
- Lou, R.N., Jacobs, A., Wilder, A.P., Therkildsen, N.O., 2021. A beginner's guide to low-coverage whole genome sequencing for population genomics. *Molecular Ecology* 30, 5966-5993.
- Lucenti, S.B., Madurell-Malapeira, J., 2020. Unraveling the fossil record of foxes: An updated review on the Plio-Pleistocene *Vulpes* spp. from Europe. *Quaternary Science Reviews* 236, 106296.
- Lynch, V.J., Bedoya-Reina, O.C., Ratan, A., Sulak, M., Drautz-Moses, D.I., Perry, G.H., Miller, W., Schuster, S.C., 2015. Elephantid genomes reveal the molecular bases of woolly mammoth adaptations to the Arctic. *Cell reports* 12, 217-228.
- Macdonald, D., Courtenay, O., Forbes, S., Mathews, F., 1999. The red fox (*Vulpes vulpes*) in Saudi Arabia: loose-knit groupings in the absence of territoriality. *Journal of Zoology* 249, 383-391.
- MacKenzie, D., 1994. *gnu Coreutils*. Free Software Foundation, Inc 2022.
- Magory Cohen, T., King, R., Dolev, A., Boldo, A., Lichter-Peled, A., Kahila Bar-Gal, G., 2013. Genetic characterization of populations of the golden jackal and the red fox in Israel. *Conservation Genetics* 14, 55-63.
- Manel, S., Perrier, C., Pratlong, M., Abi-Rached, L., Paganini, J., Pontarotti, P., Aurelle, D., 2016. Genomic resources and their influence on the detection of the signal of positive selection in genome scans. *Molecular ecology* 25, 170-184.
- Mangan, M., Kaiserman, D., Bird, P.I., 2008. The role of serpins in vertebrate immunity. *Tissue antigens* 72, 1-10.
- Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.-J., Chen, Z., Dewell, S.B., Du, L., Fierro, J.M., Gomes, X.V., Godwin, B.C., He, W., Helgesen, S., Ho, C.H., Irzyk, G.P., Jando, S.C., Alenquer, M.L.I., Jarvie, T.P., Jirage, K.B., Kim, J.-B., Knight, J.R., Lanza, J.R., Leamon, J.H., Lefkowitz, S.M., Lei, M., Li, J., Lohman, K.L., Lu, H., Makhijani, V.B., McDade, K.E., McKenna, M.P., Myers, E.W., Nickerson, E., Nobile, J.R., Plant, R., Puc, B.P., Ronan, M.T., Roth, G.T., Sarkis, G.J.,

- Simons, J.F., Simpson, J.W., Srinivasan, M., Tartaro, K.R., Tomasz, A., Vogt, K.A., Volkmer, G.A., Wang, S.H., Wang, Y., Weiner, M.P., Yu, P., Begley, R.F., Rothberg, J.M., 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437, 376-380.
- McCarty, J.P., 2001. Ecological consequences of recent climate change. *Conservation biology* 15, 320-331.
- McDevitt, A.D., Coscia, I., Browett, S.S., Ruiz-González, A., Statham, M.J., Ruczyńska, I., Roberts, L., Stojak, J., Frantz, A.C., Norén, K., 2022. Next-generation phylogeography resolves post-glacial colonization patterns in a widespread carnivore, the red fox (*Vulpes vulpes*), in Europe. *Molecular ecology* 31, 993-1006.
- McDonald, J.H., Kreitman, M., 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* 351, 652-654.
- McLean, S., Nichols, D.S., Davies, N.W., 2021. Volatile scent chemicals in the urine of the red fox, *Vulpes vulpes*. *PLoS One* 16, e0248961.
- Meisner, J., Albrechtsen, A., 2018. Inferring Population Structure and Admixture Proportions in Low-Depth NGS Data. *Genetics* 210, 719-731.
- Mierzejewska, E.J., Dwużnik, D., Koczwarska, J., Stańczak, Ł., Opalińska, P., Krokowska-Paluszak, M., Wierzbicka, A., Górecki, G., Bajer, A., 2021. The red fox (*Vulpes vulpes*), a possible reservoir of *Babesia vulpes*, *B. canis* and *Hepatozoon canis* and its association with the tick *Dermacentor reticulatus* occurrence. *Ticks and tick-borne diseases* 12, 101551.
- Mieth, B., Kloft, M., Rodríguez, J.A., Sonnenburg, S., Vobruha, R., Morcillo-Suárez, C., Farré, X., Marigorta, U.M., Fehr, E., Dickhaus, T., Blanchard, G., Schunk, D., Navarro, A., Müller, K.-R., 2016. Combining Multiple Hypothesis Testing with Machine Learning Increases the Statistical Power of Genome-wide Association Studies. *Scientific Reports* 6, 36671.
- Miko, I., LeJeune, L., Chial, H., Drovdic, C., Koopman, M., Nelson, S.C., Spivey, A., Smith, R., 2009. The genetic variation in a population is caused by multiple factors. *Essentials of genetics*, Unit 5, 125.
- Moltke, I., Albrechtsen, A., 2014. *RelateAdmix*: a software tool for estimating relatedness between admixed individuals. *Bioinformatics* 30, 1027-1028.
- Morales-Barbero, J., Martinez, P.A., Ferrer-Castán, D., Olalla-Tárraga, M.Á., 2018. Quaternary refugia are associated with higher speciation rates in mammalian faunas of the Western Palaearctic. *Ecography* 41, 607-621.
- Moritz, C., 1994. Defining 'evolutionarily significant units' for conservation. *Trends in ecology & evolution* 9, 373-375.
- Mullins, J., McDevitt, A.D., Kowalczyk, R., Ruczyńska, I., Górny, M., Wójcik, J.M., 2014. The influence of habitat structure on genetic differentiation in red fox populations in north-eastern Poland. *Acta theriologica* 59, 367-376.

- Nøhr, A.K., Hanghøj, K., Garcia-Erill, G., Li, Z., Moltke, I., Albrechtsen, A., 2021. NGSremix: a software tool for estimating pairwise relatedness between admixed individuals from next-generation sequencing data. *G3 (Bethesda)* 11.
- Norén, K., Dalén, L., Flagstad, Ø., Berteaux, D., Wallén, J., Angerbjörn, A., 2017. Evolution, ecology and conservation—revisiting three decades of Arctic fox population genetic research. *Polar Research* 36, 4.
- Norén, K., Statham, M.J., Ågren, E.O., Isomursu, M., Flagstad, Ø., Eide, N.E., Berg, T.B.G., Bech-Sanderhoff, L., Sacks, B.N., 2015. Genetic footprints reveal geographic patterns of expansion in Fennoscandian red foxes. *Global Change Biology* 21, 3299-3312.
- Okonechnikov, K., Conesa, A., García-Alcalde, F., 2015. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* 32, 292-294.
- Ortiz-Leal, I., Torres, M.V., Vargas-Barroso, V., Fidalgo, L.E., López-Beceiro, A.M., Larriva-Sahd, J.A., Sánchez-Quinteiro, P., 2023. The olfactory limb of the red fox (*Vulpes vulpes*). New insights regarding a noncanonical olfactory bulb pathway. *Frontiers in Neuroanatomy* 16, 1097467.
- Pacifici, M., Rondinini, C., Rhodes, J.R., Burbidge, A.A., Cristiano, A., Watson, J.E., Woinarski, J.C., Di Marco, M., 2020. Global correlates of range contractions and expansions in terrestrial mammals. *Nature Communications* 11, 2840.
- Park, J.Y., An, Y.-R., Kanda, N., An, C.-M., An, H.S., Kang, J.-H., Kim, E.M., An, D.-H., Jung, H., Joung, M., 2015. Cetaceans evolution: insights from the genome sequences of common minke whales. *BMC genomics* 16, 1-8.
- Park, N., Kim, H., Oh, J., Kim, J., Heo, C., Kim, J., 2024. PAPIpe: A Pipeline for Comprehensive Population Genetic Analysis. *Mol Biol Evol* 41.
- Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., Reich, D., 2012. Ancient admixture in human history. *Genetics* 192, 1065-1093.
- Pavlidis, P., Živković, D., Stamatakis, A., Alachiotis, N., 2013. SweeD: likelihood-based detection of selective sweeps in thousands of genomes. *Molecular biology and evolution* 30, 2224-2234.
- Pecl, G.T., Araújo, M.B., Bell, J.D., Blanchard, J., Bonebrake, T.C., Chen, I.-C., Clark, T.D., Colwell, R.K., Danielsen, F., Evengård, B., 2017. Biodiversity redistribution under climate change: Impacts on ecosystems and human well-being. *Science* 355, eaai9214.
- Pedersen, B.S., Quinlan, A.R., 2018. Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics* 34, 867-868.
- PicardToolkit, 2019. Picard Toolkit, Broad Institute, GitHub Repository.
- Pierce, S.B., Walsh, T., Chisholm, K.M., Lee, M.K., Thornton, A.M., Fiumara, A., Opitz, J.M., Levy-Lahad, E., Klevit, R.E., King, M.-C., 2010. Mutations in the DBP-deficiency protein

- HSD17B4 cause ovarian dysgenesis, hearing loss, and ataxia of Perrault Syndrome. *The American Journal of Human Genetics* 87, 282-288.
- Porto, L.M., Etienne, R.S., Maestri, R., 2023. Evolutionary radiation in canids following continental colonizations. *Evolution* 77, 971-979.
- Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. *Genetics* 155, 945-959.
- Prost, S., Machado, A.P., Zumbroich, J., Preier, L., Mahtani-Williams, S., Meissner, R., Guschanski, K., Brealey, J.C., Fernandes, C.R., Vercammen, P., 2022. Genomic analyses show extremely perilous conservation status of African and Asiatic cheetahs (*Acinonyx jubatus*). *Molecular ecology* 31, 4208-4223.
- Quan, J., Bode, A.M., Luo, X., 2021. ACSL family: The regulatory mechanisms and therapeutic implications in cancer. *European journal of pharmacology* 909, 174397.
- Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841-842.
- Rantanen, M., Karpechko, A.Y., Lipponen, A., Nordling, K., Hyvärinen, O., Ruosteenoja, K., Vihma, T., Laaksonen, A., 2022. The Arctic has warmed nearly four times faster than the globe since 1979. *Communications earth & environment* 3, 168.
- Reich, D., Thangaraj, K., Patterson, N., Price, A.L., Singh, L., 2009. Reconstructing Indian population history. *Nature* 461, 489-494.
- Roberts, L., 2019. Signatures of Selection in the Red Fox (*Vulpes vulpes*) in Europe using genotype-by-sequencing, University of Salford (United Kingdom).
- Rocha, J.L., Silva, P., Santos, N., Nakamura, M., Afonso, S., Qninba, A., Boratynski, Z., Sudmant, P.H., Brito, J.C., Nielsen, R., 2023. North African fox genomes show signatures of repeated introgression and adaptation to life in deserts. *Nature ecology & evolution* 7, 1267-1286.
- Rochesequencing, KAPA HyperPrep Kits.
- Rodnikova, A., Ims, R.A., Sokolov, A., Skogstad, G., Sokolov, V., Shtro, V., Fuglei, E., 2011. Red fox takeover of arctic fox breeding den: an observation from Yamal Peninsula, Russia. *Polar Biology* 34, 1609-1614.
- Sacks, B.N., Lounsberry, Z.T., Statham, M.J., 2018. Nuclear genetic analysis of the red fox across its trans-Pacific range. *Journal of Heredity* 109, 573-584.
- Safran, M., Dalah, I., Alexander, J., Rosen, N., Iny Stein, T., Shmoish, M., Nativ, N., Bahir, I., Doniger, T., Krug, H., 2010. GeneCards Version 3: the human gene integrator. *Database* 2010, baq020.
- Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution* 4, 406-425.

- Sánchez-Barreiro, F., De Cahsan, B., Westbury, M.V., Sun, X., Margaryan, A., Fontseré, C., Bruford, M.W., Russo, I.-R.M., Kalthoff, D.C., Sicheritz-Pontén, T., 2023. Historic sampling of a vanishing beast: population structure and diversity in the black rhinoceros. *Molecular biology and evolution* 40, msad180.
- Sanger, F., Nicklen, S., Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the national academy of sciences* 74, 5463-5467.
- Santos-Cortez, R.L.P., Lee, K., Giese, A.P., Ansar, M., Amin-Ud-Din, M., Rehn, K., Wang, X., Aziz, A., Chiu, I., Hussain Ali, R., 2014. Adenylate cyclase 1 (ADCY1) mutations cause recessive hearing impairment in humans and defects in hair cell function and hearing in zebrafish. *Human molecular genetics* 23, 3289-3298.
- Schubert, M., Ermini, L., Sarkissian, C.D., Jónsson, H., Ginolhac, A., Schaefer, R., Martin, M.D., Fernández, R., Kircher, M., McCue, M., 2014. Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nature protocols* 9, 1056-1082.
- Schubert, M., Lindgreen, S., Orlando, L., 2016. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Research Notes* 9, 88.
- Selås, V., Vik, J., 2006. Possible impact of snow depth and ungulate carcasses on red fox (*Vulpes vulpes*) populations in Norway, 1897–1976. *Journal of Zoology* 269, 299-308.
- Sherman, B.T., Hao, M., Qiu, J., Jiao, X., Baseler, M.W., Lane, H.C., Imamichi, T., Chang, W., 2022. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic acids research* 50, W216-W221.
- Shpak, M., Lawrence, K.N., Pool, J.E., 2024. The Precision and Power of Population Branch Statistics in Identifying the Genomic Signatures of Local Adaptation. *BioRxiv*.
- Sillero-Zubiri, C., Hoffmann, M., Macdonald, D.W., 2004. Canids: foxes, wolves, jackals, and dogs: status survey and conservation action plan. IUCN Gland, Switzerland.
- Simonsen, K.L., Churchill, G.A., Aquadro, C.F., 1995. Properties of Statistical Tests of the Neutral Model in Molecular Evolution.
- Skotte, L., Korneliussen, T.S., Albrechtsen, A., 2013. Estimating individual admixture proportions from next generation sequencing data. *Genetics* 195, 693-702.
- Slatkin, M., 1985. Gene flow in natural populations. *Annual review of ecology and systematics*, 393-430.
- Slatkin, M., 2008. Linkage disequilibrium—understanding the evolutionary past and mapping the medical future. *Nat Rev Genet* 9, 477-485.
- Slatkin, M., Maddison, W.P., 1989. A cladistic measure of gene flow inferred from the phylogenies of alleles. *Genetics* 123, 603-613.
- Sokal, R.R., Michener, C.D., 1958. A statistical method for evaluating systematic relationships. *Univ Kans sci bull* 38, 1409-1438.

- Sokolov, A.A., Sokolova, N.A., Ims, R.A., Brucker, L., Ehrich, D., 2016. Emergent rainy winter warm spells may promote boreal predator expansion into the Arctic. *Arctic*, 121-129.
- Sommer, R.S., 2020. Late Pleistocene and Holocene History of Mammals in Europe, in: Hackländer, K., Zachos, F.E. (Eds.), *Mammals of Europe - Past, Present, and Future*, Springer International Publishing, Cham, pp. 83-98.
- Sommer, R.S., Nadachowski, A., 2006. Glacial refugia of mammals in Europe: evidence from fossil records. *Mammal Review* 36, 251-265.
- Soraggi, S., Wiuf, C., Albrechtsen, A., 2018. Powerful Inference with the D-Statistic on Low-Coverage Whole-Genome Data. *G3 Genes|Genomes|Genetics* 8, 551-566.
- Sousa da Mota, B., Rubinacci, S., Cruz Dávalos, D.I., G. Amorim, C.E., Sikora, M., Johannsen, N.N., Szmyt, M.H., Włodarczak, P., Szczepanek, A., Przybyła, M.M., Schroeder, H., Allentoft, M.E., Willerslev, E., Malaspinas, A.-S., Delaneau, O., 2023. Imputation of ancient human genomes. *Nature Communications* 14, 3660.
- Spence, A.R., Tingley, M.W., 2020. The challenge of novel abiotic conditions for species undergoing climate-induced range shifts. *Ecography* 43, 1571-1590.
- Statham, M.J., Edwards, C.J., Norén, K., Soulsbury, C.D., Sacks, B.N., 2018. Genetic analysis of European red foxes reveals multiple distinct peripheral populations and central continental admixture. *Quaternary Science Reviews* 197, 257-266.
- Statham, M.J., Murdoch, J., Janecka, J., Aubry, K.B., Edwards, C.J., Soulsbury, C.D., Berry, O., Wang, Z., Harrison, D., Pearch, M., 2014. Range-wide multilocus phylogeography of the red fox reveals ancient continental divergence, minimal genomic exchange and distinct demographic histories. *Molecular Ecology* 23, 4813-4830.
- Stern, A.J., Nielsen, R., 2019. Detecting natural selection. *Handbook of Statistical Genomics: Two Volume Set*, 397-340.
- Sternberg, M., Gabay, O., Angel, D., Barneah, O., Gafny, S., Gasith, A., Grünzweig, J.M., HersHKovitz, Y., Israel, A., Milstein, D., 2015. Impacts of climate change on biodiversity in Israel: an expert assessment approach. *Regional Environmental Change* 15, 895-906.
- Storm, G.L., Andrews, R.D., Phillips, R.L., Bishop, R.A., Siniff, D.B., Tester, J.R., 1976. Morphology, reproduction, dispersal, and mortality of midwestern red fox populations. *Wildlife monographs*, 3-82.
- Strasser, A., Jost, P.J., Nagata, S., 2009. The many roles of FAS receptor signaling in the immune system. *Immunity* 30, 180-192.
- Taberlet, P., Fumagalli, L., WUST-SAUCY, A.G., COSSON, J.F., 1998. Comparative phylogeography and postglacial colonization routes in Europe. *Molecular ecology* 7, 453-464.

- Tamura, R., Tanaka, T., Akasaki, Y., Murayama, Y., Yoshida, K., Sasaki, H., 2020. The role of vascular endothelial growth factor in the hypoxic and immunosuppressive tumor microenvironment: perspectives for therapeutic implications. *Medical Oncology* 37, 1-14.
- Teacher, A.G., Thomas, J.A., Barnes, I., 2011. Modern and ancient red fox (*Vulpes vulpes*) in Europe show an unusual lack of geographical and temporal structuring, and differing responses within the carnivores to historical climatic change. *BMC evolutionary biology* 11, 1-10.
- Turner, S.D., 2014. qqman: an R package for visualizing GWAS results using QQ and manhattan plots. *Biorxiv*, 005165.
- Valasek, M.A., Repa, J.J., Quan, G., Dietschy, J.M., Turley, S.D., 2008. Inhibiting intestinal NPC1L1 activity prevents diet-induced increase in biliary cholesterol in Golden Syrian hamsters. *American Journal of Physiology-Gastrointestinal and Liver Physiology* 295, G813-G822.
- Vieira, F.G., Lassalle, F., Korneliussen, T.S., Fumagalli, M., 2016. Improving the estimation of genetic distances from Next-Generation Sequencing data. *Biological journal of the Linnean Society* 117, 139-149.
- Vitti, J.J., Grossman, S.R., Sabeti, P.C., 2013. Detecting natural selection in genomic data. *Annual review of genetics* 47, 97-120.
- Vonholdt, B., Driscoll, C., 2016. Origins of the dog: Genetic insights into dog domestication, in: Serpell, J. (Ed.), *The Domestic Dog*
- Its Evolution, Behavior and Interactions with People*, Cambridge University Press, pp. 22-41.
- Walton, Z., Samelius, G., Odden, M., Willebrand, T., 2018. Long-distance dispersal in red foxes (*Vulpes vulpes*) revealed by GPS tracking. *European Journal of Wildlife Research* 64, 1-6.
- Wang, J., 2011. COANCESTRY: a program for simulating, estimating and analysing relatedness and inbreeding coefficients. *Molecular ecology resources* 11, 141-145.
- Wang, S.-Y., Gao, K., Deng, D.-L., Cai, J.-J., Xiao, Z.-Y., He, L.-Q., Jiao, H.-L., Ye, Y.-P., Yang, R.-W., Li, T.-T., 2015. TLE4 promotes colorectal cancer progression through activation of JNK/c-Jun signaling pathway. *Oncotarget* 7, 2878.
- Wang, X., Tedford, R.H., Van Valkenburgh, B., Wayne, R.K., 2004. 39Ancestry: Evolutionary history, molecular systematics, and evolutionary ecology of Canidae, in: Macdonald, D.W., Sillero-Zubiri, C. (Eds.), *The Biology and Conservation of Wild Canids*, Oxford University Press, p. 0.
- Warret Rodrigues, C., 2023. Use of space and resources by red foxes and Arctic foxes in a coastal tundra transitional ecosystem.
- Weisstein, E.W., 2004. Bonferroni correction. <https://mathworld.wolfram.com/>.

- Welch, A.J., Bedoya-Reina, O.C., Carretero-Paulet, L., Miller, W., Rode, K.D., Lindqvist, C., 2014. Polar bears exhibit genome-wide signatures of bioenergetic adaptation to life in the arctic environment. *Genome biology and evolution* 6, 433-450.
- Westbury, M.V., Barnett, R., Sandoval-Velasco, M., Gower, G., Vieira, F.G., de Manuel, M., Hansen, A.J., Yamaguchi, N., Werdelin, L., Marques-Bonet, T., Gilbert, M.T.P., Lorenzen, E.D., 2021. A genomic exploration of the early evolution of extant cats and their sabre-toothed relatives. *Open Res Eur* 1, 25.
- Whiteside, H.M., Dawson, D.A., Soulsbury, C.D., Harris, S., 2011. Mother knows best: dominant females determine offspring dispersal in red foxes (*Vulpes vulpes*). *Plos One* 6, e22145.
- Whitlock, M., Schluter, D., 2015. *The Analysis of Biological Data: textbook and website*.
- Widdig, A., 2017. Coefficient of Relatedness, in: Vonk, J., Shackelford, T. (Eds.), *Encyclopedia of Animal Cognition and Behavior*, Springer International Publishing, Cham, pp. 1-3.
- Wiener, P., Robert, C., Ahbara, A., Salavati, M., Abebe, A., Kebede, A., Wragg, D., Friedrich, J., Vasoya, D., Hume, D.A., 2021. Whole-genome sequence data suggest environmental adaptation of Ethiopian sheep populations. *Genome Biology and Evolution* 13, evab014.
- Wilkinson, C., Vigués, J., Stoessel, M., Vinka, M., Angerbjörn, A., Norén, K., 2024. Phase-dependent red fox expansion into the tundra: implications for management. *The Journal of Wildlife Management* 88, e22569.
- Wright, S., 1922. Coefficients of inbreeding and relationship. *The American Naturalist* 56, 330-338.
- Wright, S., 1949. The genetical structure of populations. *Annals of eugenics* 15, 323-354.
- Xia, Z., Huang, M., Zhu, Q., Li, Y., Ma, Q., Wang, Y., Chen, X., Li, J., Qiu, L., Zhang, J., 2019. Cadherin related family member 2 acts as A tumor suppressor by inactivating akt in human hepatocellular carcinoma. *Journal of Cancer* 10, 864.
- Xiao, J., Dong, L.-W., Liu, S., Meng, F.-H., Xie, C., Lu, X.-Y., Zhang, W.J., Luo, J., Song, B.-L., 2023. Bile acids-mediated intracellular cholesterol transport promotes intestinal cholesterol absorption and NPC1L1 recycling. *Nature Communications* 14, 6469.
- Xin, S., Mueller, C., Pfeiffer, S., Kraft, V.A., Merl-Pham, J., Bao, X., Feederle, R., Jin, X., Hauck, S.M., Schmitt-Kopplin, P., 2022. MS4A15 drives ferroptosis resistance through calcium-restricted lipid remodeling. *Cell Death & Differentiation* 29, 670-686.
- Yelmen, B., 2021. Characterization of ancient Eurasian influences within modern human genomes, Tartu.
- Yelmen, B., Marnetto, D., Molinaro, L., Flores, R., Mondal, M., Pagani, L., 2021. Improving selection detection with population branch statistic on admixed populations. *Genome biology and evolution* 13, evab039.

- Yen, C.-L.E., Stone, S.J., Cases, S., Zhou, P., Farese Jr, R.V., 2002. Identification of a gene encoding MGAT1, a monoacylglycerol acyltransferase. *Proceedings of the National Academy of Sciences* 99, 8512-8517.
- Yi, X., Liang, Y., Huerta-Sanchez, E., Jin, X., Cuo, Z.X.P., Pool, J.E., Xu, X., Jiang, H., Vinckenbosch, N., Korneliussen, T.S., 2010. Sequencing of 50 human exomes reveals adaptation to high altitude. *science* 329, 75-78.
- Yim, H.-S., Cho, Y.S., Guang, X., Kang, S.G., Jeong, J.-Y., Cha, S.-S., Oh, H.-M., Lee, J.-H., Yang, E.C., Kwon, K.K., 2014. Minke whale genome and aquatic adaptation in cetaceans. *Nature genetics* 46, 88-92.
- Young, A.D., Gillung, J.P., 2020. Phylogenomics—principles, opportunities and pitfalls of big-data phylogenetics. *Systematic entomology* 45, 225-247.
- Yudin, N.S., Larkin, D.M., Ignatieva, E.V., 2017. A compendium and functional characterization of mammalian genes involved in adaptation to Arctic or Antarctic environments. *BMC genetics* 18, 33-43.
- Zecchin, B., De Nardi, M., Nouvellet, P., Vernesi, C., Babbucci, M., Crestanello, B., Bagó, Z., Bedeković, T., Hostnik, P., Milani, A., 2019. Genetic and spatial characterization of the red fox (*Vulpes vulpes*) population in the area stretching between the Eastern and Dinaric Alps and its relationship with rabies and canine distemper dynamics. *PloS one* 14, e0213515.
- Zhang, H., Sun, Z., Feng, S., Zhang, J., Zhang, F., Wang, W., Hu, H., Zhang, W., Bao, M., 2022. The C2H2-type zinc finger protein PhZFP1 regulates cold stress tolerance by modulating galactinol synthesis in *Petunia hybrida*. *Journal of Experimental Botany* 73, 6434-6448.
- Zhang, Z., Xia, T., Zhou, S., Yang, X., Lyu, T., Wang, L., Fang, J., Wang, Q., Dou, H., Zhang, H., 2023. High-quality chromosome-level genome assembly of the Corsac fox (*Vulpes corsac*) reveals adaptation to semiarid and harsh environments. *International Journal of Molecular Sciences* 24, 9599.
- Zhu, Y., Yuan, Y., Si, H., Li, S., Zhao, F., Mu, R., Lin, Z., Wang, X., Qiu, Q., Xu, C., 2024. Lipidomic and transcriptomic profiles provide new insights into the triacylglycerol and glucose handling capacities of the Arctic fox. *Frontiers in Veterinary Science* 11, 1388532.

Appendix 1

==Genome sequencing data preprocessing and quality control==

FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>)

MultiQC (<https://docs.seqera.io/multiqc>)

==Mapping==

Paleomix (<https://paleomix.readthedocs.io/en/stable/>)

==Relatedness==

- `angsd -b redfoxbamlist.txt -ref GCF_003160815.1_VulVul2.2_genomic.fasta -GL 1 -domajorminor 1 -snp_pval 1e-6 -domaf 2 -minmaf 0.007 -doGlf 3 -minQ 20 -minMapQ 20 -minInd 73 -doCounts 1 -only_proper_pairs 0 -baq 1 -setMinDepthInd 1 -out Vulpes_Relatedness_baqmindep_newmaftestdp1`
- `zcat Vulpes_Relatedness_baqmindep.mafs.gz | tail -n +2 | cut -f 6 > Vulpes_Relatedness_baqmindep.freq`
- `ngsRelate -g Vulpes_Relatedness_baqmindep.glf.gz -n 91 -f Vulpes_Relatedness_baqmindep.freq -O Vulpes_Relatedness_baqmindep -z redfoxsamplenames.txt`
- `awk '$15 > 0.3 {print $3,$4,$15,$16,$17}' Vulpes_Relatedness_baqmindep > Vulpes_Related_baqmindep.txt`
- `nano Vulpes_Related_baqmindep.txt`

==Genetic diversity==

- `angsd -i A_09_56_S65.rmdup.normal.bam -GL 1 -anc GCF_003160815.1_VulVul2.2_genomic.fasta -out A_09_56_S65_Het -doSaf 1 -ref GCF_003160815.1_VulVul2.2_genomic.fasta -domajorminor 1 -snp_pval 1e-6 -domaf 2 -minQ 20 -minMapQ 20 -doCounts 1 -only_proper_pairs 0 -baq 1`
- `realSFS -fold 1 A_09_56_S65_Het.saf.idx > A_09_56_S65_Het.ml`

In R

```
a<-scan("A_09_56_S65_Het.ml")
a[2]/sum(a)
```

==Phylogenetic analyses==

==Genetic distance tree==

- `angsd -b redfoxbamlist_firstrelaterm_incl_outgroup.txt -ref GCF_003160815.1_VulVul2.2_genomic.fasta -anc GCF_003160815.1_VulVul2.2_genomic.fasta -GL 1 -doGeno 8 -doPost 1 -doSaf 1 -domajorminor 1 -snp_pval 1e-6 -domaf 2 -minmaf 0.007 -minQ 20 -minMapQ 20 -minInd 70 -doCounts 1 -only_proper_pairs 0 -baq 1 -setMinDepthInd 1 -out Vulpes_Relatedness_baqmindep1_ngsdist2_newmaf`

- `ngsDist --n_boot_rep 100 --boot_block_size 1 --geno Vulpes_Relatedness_baqmindep1_ngsdist.geno.gz --n_ind 87 -labels labels.txt --n_sites 28731 --n_threads 10 --out Vulpes_Relatedness_baqmindep_ngsdist.dist`
- `fastme -i Vulpes_Relatedness_baqmindep_ngsdist.dist -s -D 87 -q -o Vulpes_Relatedness_baqmindep_ngsdist_2_q.nwk`
- `head -n 1 Vulpes_Relatedness_baqmindep_ngsdist_2_q.nwk > Vulpes_Relatedness_baqmindep_ngsdist_2_q.main.nwk`
- `tail -n +2 Vulpes_Relatedness_baqmindep_ngsdist_2_q.nwk | awk 'NF' > Vulpes_Relatedness_baqmindep_ngsdist_2_q.boot.nwk`
- `RAXML-ng --support --tree Vulpes_Relatedness_baqmindep_ngsdist_2_q.main.nwk -bs-trees Vulpes_Relatedness_baqmindep_ngsdist_2_q.boot.nwk --prefix Vulpes_Relatedness_baqmindep_ngsdist_2_q`

==Phylogenetic network==

- `angsd -b redfoxbamlist_firstrelaterm_incl_outgroup.txt -ref GCF_003160815.1_VulVul2.2_genomic.fasta -anc GCF_003160815.1_VulVul2.2_genomic.fasta -GL 1 -doMajorMinor 1 -doGeno 2 -doPost 1 -doSaf 1 -Snp_pval 1e-6 -doMaf 2 -minmaf 0.007 -minInd 70 -doCounts 1 -only_proper_pairs 0 -baq 1 -setMinDepthInd 3 -out Vulpes_70_3_splitstree`

(<https://github.com/CGRL-QB3-UCBerkeley/PopGenTools>)

```
NSITES=`zcat Vulpes_70_1_splitstree.mafs.gz | tail -n+2 | wc -l`
echo $NSITES
```

- `perl /projappl/project_2007265/software/PopGenTools0.901_1.pl Adegnet -g <(gunzip -c Vulpes_70_3_splitstree.geno.gz) -n 87 -s 11586 -o Vulpes_70_3_splitstree.adegenet -h 0 -m 0.2`
- Usage `PopGenomics.pl geno_to_adegenet [options]`
Options: `-g FILE` raw genotype probability generated using ANGSD `"-doGeno 2"`
 `-n INT` sample size
 `-s INT` number of loci

`-o FILE` outfile name

`-h INT 1 = all sites; 0 = one snp per contig [1] ###only keep informative`

In R

```
library(adegenet)
```

```
library(Zecchin et al.)
```

```
library(ape)
```

```
gt <- read.delim("~/Phylonetwork/Vulpes_70_3_splitstree.adegenet", header=T)
```

```
gt <- gt[, -1]
```

```
gtgi <- df2genind(gt, ncode=4)
```

```
D <- dist(gtgi)
```

```
write.matrix(D, file = "~/Phylonetwork/Vulpes_70_3_splitstree.D")
```

- Perl
- perl /projappl/project_2007265/software/splittree.pl Vulpes_70_3_splitstree.D 87 > Vulpes_70_1_splitstree.splitstree

==Visualisation==

- SplitsTree

==Gene flow analyses==

([https://www.popgen.dk/angsd/index.php/Abbababa2#Tutorial for the ABBABABA \(Multipopulation test\)](https://www.popgen.dk/angsd/index.php/Abbababa2#Tutorial%20for%20the%20ABBABABA%20(Multipopulation%20test)))

- angsd -b redfoxbamlist_firstrelaterm_incl_outgroup.txt -ref GCF_003160815.1_VulVul2.2_genomic.fasta -anc GCF_003160815.1_VulVul2.2_genomic.fasta -GL 1 -doAbbababa2 1 -useLast 1 -rmTrans 0 -blockSize 5000000 -remove_bads 1 -doPost 1 -doSaf 1 -domajorminor 1 -snppval 1e-6 -domaf 2 -minmaf 0.007 -minQ 20 -minMapQ 20 -minInd 70 -doCounts 1 -only_proper_pairs 0 -baq 1 -setMinDepthInd 1 -sizeFile sizeFile.size -out Vulpes_baqmindep_dstats_maf_007
- -doAbbababa2 1 tells the program to estimate ABBA-BABA using all bases at each position
- -UseLast 1, invokes the use of the last individual in the BAM list, this individual is the outgroup
- -anc option on the command line
- -remove_bads 1, removes bad reads
- -blockSize, is the size of each block in bases
- -doPost 1 calculate posterior probabilities
- -doSaf 1, calculates site allele frequency
- -sizeFile has the list of populations and their sizes

==Population structure==

- angsd -b redfoxbamlist_firstrelaterm.txt -ref GCF_003160815.1_VulVul2.2_genomic.fasta -GL 1 -domajorminor 1 -snppval 1e-6 -domaf 2 -minmaf 0.007 -doGlf 2 -minQ 20 -minMapQ 20 -minInd 70 -doCounts 1 -only_proper_pairs 0 -baq 1 -setMinDepthInd 1 -out Vulpes_PCA_ADMIX_dep1_newmaf
- angsd -bredfoxbamlist_firstrelaterm.txt -ref GCF_003160815.1_VulVul2.2_genomic.fasta -GL 1 -domajorminor 1 -snppval 1e-6 -domaf 2 -minmaf 0.007 -doGlf 2 -minQ 20 -minMapQ 20 -minInd 70 -doCounts 1 -only_proper_pairs 0 -baq 1 -setMinDepthInd 1 -out Vulpes_baqmindep_unrelatedPCA_dep1_newmaf.beagle.gz

==PCA==

- pcangsd -b Vulpes_baqmindep_unrelatedPCA_dep1_newmaf.beagle.gz -o Vulpes_baqmindep_unrelatedPCA_dep1_newmaf -t 5
- PLOT in R

```
genome <-
as.matrix(read.table("Vulpes_baqmindep_unrelatedPCA_dep1_newmaf.cov",header=
F))

e <- eigen(Li et al.)

png("/media/congo/pcangsd/PCA_1_2_FINAL_redfox.png")

plot(e$vectors[,1:2],col =
c(rep("#00aae4",35),rep("#740e86",34),rep("#387c1a",16)),xlab="PC1,26,17%",ylab="
PC2,14.07%",pch=19)

legend("top",legend= c("Arctic","Central_Europe","Israel"), fill = c("#00aae4",
"#740e86", "#387c1a"))

dev.off()
```

==NGSadmix==

- for (j in 1:50) {for (i in 2:6)


```
{system(paste("/media/zambezi/BLESSING/REDFOX/angsd_41/misc/NGSadmix -
likes Vulpes_baqmindep_unrelatedPCA_ADMIX_dep1_newmaf.beagle.gz -K ", i, " -o
", "Admix_runs_DEP1_K", i, "run", j,sep = ""))}}
```
- (for log in `ls \*.log`; do grep -Po 'like=\\Admix_unrelated_K[^]+' \$log; done) >
 logfile.txt

==Visualisation==

- CLUMPAK and INKSCAPE

==Selection tests==

Population Branch Statistics

(<https://github.com/mfumagalli/Copenhagen/blob/master/Files/day2a.md>)

==Arctic (N=37)==

- angsd -b VV_arctic_pop.txt -ref GCF_003160815.1_VulVul2.2_genomic.fasta -anc
 GCF_003160815.1_VulVul2.2_genomic.fasta -doHWE 1 -doSaf 1 -GL 1 -
 domajorminor 1 -snp_pval 1e-6 -domaf 1 -minmaf 0.02 -minQ 20 -minMapQ 20 -
 minInd 28 -doCounts 1 -only_proper_pairs 0 -baq 1 -setMinDepthInd 1 -out
 arctic_pbs_dep1

==Central Europe(N=35)==

- angsd -b VV_centraleurope_pop.txt -ref
 GCF_003160815.1_VulVul2.2_genomic.fasta -anc
 GCF_003160815.1_VulVul2.2_genomic.fasta -doHWE 1 -doSaf 1 -GL 1 -
 domajorminor 1 -snp_pval 1e-6 -domaf 1 -minmaf 0.02 -minQ 20 -minMapQ 20 -
 minInd 28 -doCounts 1 -only_proper_pairs 0 -baq 1 -setMinDepthInd 1 -out
 centraleurope_pbs_dep1

==Israel(N=19)==

- `angsd -b VV_israel_pop.txt -ref GCF_003160815.1_VulVul2.2_genomic.fasta -anc GCF_003160815.1_VulVul2.2_genomic.fasta -doHWE 1 -doSaf 1 -GL 1 -domajorminor 1 -snp_pval 1e-6 -domaf 1 -minmaf 0.04 -minQ 20 -minMapQ 20 -minInd 13 -doCounts 1 -only_proper_pairs 0 -baq 1 -setMinDepthInd 1 -out israel_pbs_dep1`

==Folded pairwise two-dimensional site frequency spectrum for each pair==

- `realSFS arctic_pbs_dep1.saf.idx israel_pbs_dep1.saf.idx -fold 1 > israel.arctic.ml`
- `realSFS israel_pbs_dep1.saf.idx centraleurope_pbs_dep1.saf.idx -fold 1 > israel.centraleurope.ml`
- `realSFS centraleurope_pbs_dep1.saf.idx arctic_pbs_dep1.saf.idx -fold 1 > centraleurope.arctic.ml`

==FST global estimate==

- `realSFS fst index israel_pbs_dep1.saf.idx centraleurope_pbs_dep1.saf.idx arctic_pbs_dep1.saf.idx -sfs israel.centraleurope.ml -sfs israel.arctic.ml -sfs centraleurope.arctic.ml -fstout VV_TEST_IS_CE_AR`

==Sliding window analysis==

- `realSFS fst stats2 VV_TEST_IS_CE_AR.fst.idx -type 2 -win 50000 -step 10000 > VV_TEST_IS_CE_AR_slidingwindow`

==Editing text file to find and replace negative values with 0==

- `awk '{$10=$10<0?0:$10}1' OFS='\t' VV_TEST_IS_CE_AR_slidingwindow | awk '{$8=$8<0?0:$8}1' OFS='\t' | awk '{$9=$9<0?0:$9}1' OFS='\t'`

==Multiple testing==

- False discovery rate (FDR) in R for all the PBS values for PBS0, 1, and 2
- `setwd("C:/Users/bleess/Documents/1_Redfox_Project/PBS")`
- `vv_data <- read.csv("PBS_arctic_sign.csv", header = FALSE)`
- `PBS2_fdr_adjusted <- p.adjust(data$PBS2, method = 'fdr', n = length(data$PBS2))`
- `print(PBS2_fdr_adjusted)`
- `write.table(PBS2_fdr_adjusted, file = "adjusted_PBS2_fdr_adjusted.txt", sep = "\t", col.names = "PBS2_fdr_adjusted_VV", row.names = FALSE)`

==Manhattan Plot==

- `setwd("~/1_Redfox_Project/PBS")`
- `library("qqman")`
- `PBS_data <- read.table("VV_TEST_IS_CE_AR_slidingwindow_FDR_adjusted.txt", header=TRUE)`
- `names(PBS_data)[-c(1:4)] <- c("wFst_ISR_CE", "wFst_ISR_ARC", "wFst_CE_ARC", "PBS_ISR", "PBS_CE", "PBS_ARC")`
- `mmax <- max(c(PBS_data$PBS_ISR, PBS_data$PBS_CE, PBS_data$PBS_ARC), na.rm=T)`
- `PBS_data$chr_number <- as.numeric(as.factor(PBS_data$chr))`

- `manhattan(PBS_data, chr = "chr_number", bp = "midPos", logp= FALSE, p = "PBS_ARC", snp = "region", ylim = c(0,max(PBS_data$PBS_ARC)*1.1), ylab = "PBS_ARCTIC")`
- `abline(h=0.01, col="red")`

==Extracting genes under selection==

- `sort -k1,1 -k2,2n genomic.gff > genomic.sorted.bed`
- `bedtools closest -D b -a PBS_arctic_sign_adj_pbs2.bed -b genomic_closest_genes_sorted_test.bed -io -t first > PBS_arctic_sign_adj_pbs2_genes_tfirst_io.bed`
- `bedtools closest -D b -a PBS_arctic_sign_adj_pbs2.bed -b genomic_closest_genes_sorted_test.bed -iu -t first > PBS_arctic_sign_adj_pbs2_genes_tfirst_iu.bed`
- `bedtools closest -D b -a PBS_arctic_sign_adj_pbs2.bed -b genomic_closest_genes_sorted_test.bed -id -t first > PBS_arctic_sign_adj_pbs2_genes_tfirst_id.bed`
- `bedtools closest -D b -a PBS_arctic_sign_adj_pbs2.bed -b genomic_closest_genes_sorted_test.bed -id -t all > PBS_arctic_sign_adj_pbs2_genes_tall_id.bed`

==Functional Annotation==

==Extract overlapping upstream, downstream genes ==

- `grep -o "Name=[^;]*" | cut -d'=' -f2 | sort | uniq > PBS_arctic_sign_adj_pbs2_genes_withoverlap_protein_coding_noduplicates.txt`
- `grep -o "Name=[^;]*" PBS_arctic_sign_adj_pbs2_genes_tall_k_iu_protein_coding.txt | cut -d'=' -f2 | sort | uniq > PBS_arctic_sign_adj_pbs2_genes_tall_k_iu_protein_coding_noduplicates.txt`
- `grep -o "Name=[^;]*" PBS_arctic_sign_adj_pbs2_genes_tall_k_io_protein_coding.txt | cut -d'=' -f2 | sort | uniq > PBS_arctic_sign_adj_pbs2_genes_tall_k_io_protein_coding_noduplicates.txt`
- The list of genes is used as input in Database for Annotation, Visualisation and Integrated Discovery(DAVID), GENECARDS and NCBI GENE