

Genomic implications of severe population bottlenecks in two Mustelid species



Rebecca Lianne Shaw

A thesis submitted to the School of Biological Sciences of the University of East Anglia in partial fulfilment of the requirements for the degree of Doctor of Philosophy

August 2025

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that use of any information derived therefrom must be in accordance with current UK Copyright Law. In addition, any quotation or extract must include full attribution.

Thesis Abstract

The accelerating decline in global biodiversity is now recognised as a sixth mass extinction event, representing an unprecedented ecological crisis. Habitat destruction, climate change, direct persecution and overexploitation of natural resources have rendered animal populations increasingly susceptible to decline, with 73% of the 5,495 vertebrate populations monitored by the IUCN Red List exhibiting reductions over the past five decades. Population bottlenecks, a common consequence of such pressures, reduce genetic diversity, increase inbreeding and can compromise long-term viability, yet their genomic consequences vary in magnitude and persistence. This thesis examines these impacts in two mustelids that experienced markedly different bottlenecks: the European polecat (*Mustela putorius*) and the black-footed ferret (*Mustela nigripes*). In Chapter Two, I used whole-genome data from British and mainland European polecats to investigate population structure, diversity and demographic history. The British population showed genetic isolation from the continent, with Welsh polecats distinct from English ones. A nineteenth-century bottleneck in Wales was detected, less severe than that of the black-footed ferret, accompanied by evidence of recent inbreeding. In Chapter Three, I examined coding-region variation in the black-footed ferret using historic, captive and reintroduced genomes. The extreme twentieth-century bottleneck caused substantial diversity loss, fixation of some deleterious alleles and elevated mutational load. In Chapter Four, I extended this to conserved non-coding elements, identifying regulatory variants near genes involved in reproduction, neurodevelopment and immunity, some predicted to alter transcription factor binding and potentially contributing to reduced fertility. I synthesise these findings in a broader conservation genomic context, showing how differences in bottleneck severity and duration influence the retention of diversity and accumulation of deleterious mutations, with implications for the management of small or recovering carnivore populations.



Credit: Rachel Brookes Art

Access Condition and Agreement

Each deposit in UEA Digital Repository is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the Data Collections is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form. You must obtain permission from the copyright holder, usually the author, for any other use. Exceptions only apply where a deposit may be explicitly provided under a stated licence, such as a Creative Commons licence or Open Government licence.

Electronic or print copies may not be offered, whether for sale or otherwise to anyone, unless explicitly stated under a Creative Commons or Open Government license. Unauthorised reproduction, editing or reformatting for resale purposes is explicitly prohibited (except where approved by the copyright holder themselves) and UEA reserves the right to take immediate 'take down' action on behalf of the copyright and/or rights holder if this Access condition of the UEA Digital Repository is breached. Any material in this database has been supplied on the understanding that it is copyright material and that no quotation from the material may be published without proper acknowledgement.

Table of Contents

Thesis Abstract.....	1
Table of Contents.....	2
Acknowledgements.....	3
List of figures and tables.....	5
Publications and Contributions.....	18
1.0 An introduction to conservation genomics, tools, practices and its application in the protection and recovery of endangered mammals.....	19
1.1 Abstract.....	19
1.2 The Global Biodiversity Crisis.....	19
1.3 Underlying population genetics and evolutionary theory in the field of conservation genetics and genomics.....	22
1.3 Conservation genetics and genomics.....	24
1.4 Conservation Genomics in Application.....	27
1.5 The consequences of population bottlenecks.....	31
1.6 Conservation in practice: methods to prevent the loss of species.....	36
1.7 Thesis aims.....	39
2.0 Population genomics of the European polecat (<i>Mustela putorius</i>) in Britain.....	45
2.1 Abstract.....	45
2.2 Introduction.....	46
2.3 Methods.....	49
2.4 Results.....	55
2.5 Discussion.....	67
3.0 Characterisation of the accumulation of deleterious mutations within protein-coding genes associated with severe population bottleneck in the black-footed ferret (<i>Mustela nigripes</i>).....	75
3.1 Abstract.....	75
3.2 Introduction.....	76
3.3 Methods.....	80
3.4 Results.....	86
3.5 Discussion.....	101
4.0 Identification of constrained regions in the endangered Black-footed ferret and implications of noncoding genetic variants.....	108
4.1 Abstract.....	108
4.2 Introduction.....	109
4.3 Methods.....	112
4.4 Results.....	118
4.5 Discussion.....	128
5.0 General Discussion: Genomics in Conservation: Utility, Practicality, and Tool Development.....	134
5.1 Abstract.....	134
5.2 Chapter Two: Demographic Insights and Hybridisation in British Polecats.....	135
5.3 Chapter Three: Genetic Load and Recovery in the Black-Footed Ferret.....	138
5.4 Chapter Four: Tool Development and the Regulatory Genome.....	140
5.5 Conclusion: Bringing Genomics into Broader Conservation Practice.....	143
References.....	144
Appendix.....	207

A.1 Chapter Two - Population Genomics of the European polecat (<i>Mustela putorius</i>) in Britain.....	207
A.2 Chapter Three - Characterisation of deleterious mutations within protein-coding genes in the Black-footed ferret (<i>Mustela nigripes</i>).....	214
A.3 Chapter four - Identification of constrained regions in the endangered Black-footed ferret and implications of noncoding genetic variants.....	222

Acknowledgements

The saying goes that a PhD is more of a marathon than a sprint, but in many ways these years have raced by. Starting a PhD in the midst of a global pandemic was no easy feat, but the people I have met and the community I have developed during my almost five years in Norwich, across the Earlham Institute and UEA have been immensely supportive, and beyond what I imagined I would find when I embarked on this journey.

Firstly, to my supervisors Graham and Wilfried. I am grateful to you both; Graham for taking a chance on me, and bringing me into the fold of mustelid madness and to Wilfried, who adopted me when I was a stray, despite saying you would never work on *Mustela* ever again. Thank you both for your guidance, mentorship and for persisting with me all these years.

A massive thank you to all of the Haerty group, both past and present. Indebted thanks go to: Pete, you lived up to your name's roots and were a brilliant older PhD student to gain guidance from. Will, I am immensely grateful for our discussions, your wisdom and for keeping me *slightly* sane these past years. Dave, you gave me a home away from home with four-legged friends, tea and introduced me to an obsession in nature-based boardgames that will forever relieve me of my money. Ange, thank you for the constant effort you put in, with helping me in my lab work and for getting me back to running! You gave me strength when I was insane enough to embark on two marathon finishes this year.

My time as a PhD student at the Earlham Institute and UEA would not have been the same without the wonderful student community there. To the founding members of the ESB - Matt, Anita, Sam and Peter - thank you for being so welcoming, for our late night chats and for introducing me to the wilderness that is Norwich at 5am. I'm incredibly grateful for being given the opportunity to become a small part of the legacy of ESB alongside Lorcan, Dasha and Mia. Planning the ESB symposium(s) were some of the best and worst times of the past five years, but you all made it a latte more enjoyable! 12:30 was always my favourite time of day thanks to Jess, Sofia, Mia, Kate, Lu and George. I know I wouldn't have gotten through this PhD without our long lunches, laughs and family outings. A massive thank you also to the ARIES Fairies (Lili, George, Karolina and Sarah), for Tuesday quiz nonsense, wildlife trips and for accepting me into the UEA gang. Special thanks as well to my lovely housemates who became like family - Lili, Beth, Jack, Eliza (and by proxy Dan and Meg). I am indebted to you all for making life outside of the PhD much more enjoyable!

I am thankful that science communication and science policy played a massive role during my time as a PhD student. I am grateful to my supervisors for supporting me in my pursuit of these opportunities, as well as to the ARIES DTP. I am also incredibly grateful to my Pint of Science family - Sam, Cansu, Rose, Carolina, Becky and Lizzie, the EnClude team - Aparna and Ev, Dr Carl Harrington for offering me any and every chance to work with him and to my incredible Defra colleagues. It was inspiring to work with you all on various projects and I'll never take for granted what a privilege it was to have had these experiences.

I am forever grateful to my actual family for putting up with me while I pursue "*yet another thing to add to my CV*". Your support has never wavered, and I hope that my insistence at getting you all to call me Dr won't get too annoying. I was lucky to have had and to unfortunately have lost several family members during my PhD. My grandparents, who were always proud that I wanted to study 'an ology', my uncle who instilled in me my love for wildlife and my beautiful boy Hobbes. I dedicate the passion I put into this PhD to you.

Finally, to my partner Fede. My appreciation for you has only strengthened throughout these five years, and I am eternally thankful to have gone through this journey with you.

List of figures and tables

Figure 1.1: (a) The range of the European polecat in Great Britain in 1915. Dark green areas indicate the population stronghold, light green indicates known occurrences, and “?” refers to uncertain occurrences of polecats in those areas (Langley & Yalden, 1977). (b) Records received during the Vincent Wildlife Trust 2014-15 survey of verifiable sightings, based on their phenotype, of true polecats (dark-green), polecat-ferret hybrids (yellow), both true polecats and hybrids (lime green), and unverifiable (grey) (Croose, 2016).....pp42

Figure 1.2: Map of North America and southern parts of Canada. Dotted line shows the historic range of the Black-footed ferret. Green circles depict the 34 reintroduction sites that have been utilised during the reintroduction programme from 1991-present day. The purple star depicts the location in Meeteetse, Wyoming where the last wild population was found in 1981, before going extinct in 1986. Redrawn from (Marinari et al., 2024).....pp44

Figure 2.1: (a) The range of the European polecat in Britain in 1915. Dark green areas indicate the population stronghold, light green indicates known occurrences, and “?” refers to uncertain occurrences of polecats in those areas (Langley & Yalden, 1977). (b) Records received during the Vincent Wildlife Trust 2014-15 survey of verifiable sightings, based on their phenotype, of true polecats (dark-green), polecat-ferret hybrids (yellow), both true polecats and hybrids (lime green), and unverifiable (grey) (Croose, 2016). (c) Samples selected to be sequenced in this study (dark yellow) and (red) samples that were previously sequenced by (Etherington et al., 2022). Figures 2.1a and 2.1b were adapted from Croose, 2016.....pp51

Figure 2.2: Principal component analysis (PCA) of whole genome SNPs for British and European samples. **a)** Without being batch effect aware using bcftools (1,730,122 SNPs) **b)** Variants called using GATK, batch effect aware and accounting for downsampling of higher coverage samples (595,820 SNPs).....pp58

Figure 2.3: Principal component analysis (PCA) of whole genome SNPs for British, Domestic ferret and European samples. **a)** PCA run with SNPs called using batch effect aware GATK, but no downsampling to account for higher coverage samples (397,604 SNPs) **b)** PCA run with the SNP dataset called using GATK, batch effect aware and downsampling of higher coverage samples (595,820 SNPs).....pp58

Figure 2.4: Principal component analysis (PCA) of whole genome SNPs for British samples. **a)** PCA run with SNPs called without batch effect awareness using bcftools (1,663,826 SNPs) **b)** PCA run with the SNP dataset called using GATK, batch effect aware and downsampling of high coverage samples (595,820 SNPs).....**pp59**

Figure 2.5. Population structure of the European polecat in Britain and across mainland Europe. (a) PCA based on whole genome SNPs for all European polecats and domestic ferrets. (b) Map of Britain demonstrating longitude coordinates of samples and PCA based on whole genome SNPs for all British polecats, which is supported by a significant correlation between PC1 and longitude coordinates ($r_s = -0.754$, $p < 0.01$). (c) Admixture results for whole-genome SNPs between the four groups (domestic ferret, English, Welsh and mainland European polecats), with the best fitting value of $K = 3$, supported by a cross validation error of 0.68 (Supplementary Table 2.02).....**pp60**

Figure 2.6: Maximum likelihood phylogeny with 100 bootstrap replicates using genome-wide SNPs (2,941,545 SNPs) from 72 samples, consisting of domestic ferrets and European polecats from Europe and Britain. The number at the nodes refer to bootstrap values of 95 and above. Branch lengths are in expected substitutions per site.....**pp61**

Figure 2.7: Isolation by distance of British samples plotted using a pairwise Hamming genetic distance and geographic distance matrix. Results revealed a positive correlation between genetic distance and geographic distance ($r = 0.333$, $p = 0.001$).....**pp62**

Figure 2.8. Genetic diversity estimates and comparisons across British and European populations (a) Shared and private SNPs between populations. (b) Population nucleotide diversity values (π), calculated in 20 Mb windows across the genome. (c) Proportion of heterozygous sites calculated using VCFtools and (d) Population-level inbreeding coefficient values.....**pp63**

Figure 2.9. The realised inbreeding coefficient (F_{ROH}) in correlation with mean read depth per individual (a) All individuals across British and European populations (64 individuals) (b) All individuals across British and European populations with a mean read depth of 6x or higher (47 individuals: Wales (11), England (23), France (3), Spain (2), Austria (3), Italy (5)).....**pp64**

Figure 2.10. Population estimates of higher coverage individuals (mean depth >6x) and (a) the calculated proportion of the genome in ROH (F_{ROH}) and (b) the distribution of ROH

across different length categories. Data points represent the percentage of an individual's total ROH that falls within each length category.....pp65

Figure 2.11: Mutation load landscape across populations within different functional categories. Loss of function (LoF) (a) homozygous mutational load per population and (b) heterozygous mutational load and missense (c) homozygous mutational load and (d) heterozygous mutational load.....pp67

Figure 2.12: Estimated mean effective population size (solid line) with 95% confidence intervals (shaded area) for English (green) and Welsh (orange) populations across 200 generations (generation time = 4 years) based on whole genome SNP data computed using GONE.....pp68

Figure 3.1: Alignment of the *DVL2* coding sequence from across the eight mammalian species. The red box highlights the site that CODEML reported as a BEB site. Note the upstream and downstream gaps either side of the highlighted box, which resulted in this alignment being categorised as 'bad'.....pp86

Figure 3.2: Alignment of the *ADAMTS3* coding sequence across the eight mammalian species. The red box highlights the site that CODEML reported as a BEB site. Note the lack of gaps downstream and understream of the highlighted box, which resulted in this alignment being categorised as 'good'.....pp86

Figure 3.3: Distribution of relatedness (PHI) values across population groups. The histogram shows the frequency of pairwise relatedness estimates for individuals. Each bar represents the count of relatedness values within a specific range, and the fill color indicates the population group of the individual pairs. The relatedness values range from unrelated individuals (<0.044) to monozygotic twins or duplicates (>0.354) (Manichaikul et al., 2010).....pp88

Figure 3.4. Population structure of black-footed ferret individuals inferred from whole-genome SNP data (944,698 variants). Principal component analysis (PCA) plots show clustering patterns among individuals from Wild (n = 6), Captive (n = 26), and Reintroduced (n = 22) populations: (a) PC1 versus PC2 and (b) PC2 versus PC3. Two Wild individuals (SB10 and SB2) cluster separately from the main population groups,

highlighting elevated genetic divergence relative to the other sampled individuals.....pp89

Figure 3.5: ADMIXTURE analysis of population structure based on 944,698 whole-genome SNPs from 54 black-footed ferret individuals across three population groups: Wild (n = 6), Captive (n = 26), and Reintroduced (n = 22). Results are shown for K = 1–5 ancestral clusters. The analysis supports a single, genetically homogeneous population, with the best-supported model at K = 1.....pp90

Figure 3.6 Comparison of genome-wide genetic diversity among sampled black-footed ferret populations: Wild (n = 6), Captive (n = 26), and Reintroduced (n = 22). Population nucleotide diversity (π) is shown for each group, calculated from a dataset of 1,927,827 SNPs using VCFtools. Estimates were generated in 20 Mb sliding windows across the genome, allowing for assessment of variation in diversity between populations.....pp91

Figure 3.7: Genetic diversity estimates across black-footed ferret population groups: Wild (n = 6), Captive (n = 26), and Reintroduced (n = 22), measured as the proportion of heterozygous sites per individual. Heterozygosity was calculated using VCFtools. Boxplots represent the median and interquartile range of heterozygosity values within each population group. Individual data points are overlaid as jittered dots to show within-group variation.....pp92

Figure 3.8: Inbreeding coefficient values per individual within population groups: Wild (n = 6), Captive (n = 26), and Reintroduced (n = 22), as calculated by VCFtools. Boxplots represent the median and interquartile range of inbreeding coefficient values within each population group. Individual data points are overlaid as jittered dots to show within-group variation.....pp93

Figure 3.9: Mean inbreeding coefficient (F_{ROH}) for each black-footed ferret population group: Wild (n = 6), Captive (n = 26), and Reintroduced (n = 22), as calculated by PLINK. Bars represent the mean Froh value per population, with error bars indicating the standard error of the mean. Froh was calculated as the proportion of the genome in runs of homozygosity (ROH).....pp94

Figure 3.10: Number of runs of homozygosity (ROH) segments as calculated by PLINK per individual across black-footed ferret population groups: Wild (n = 6), Captive (n = 26),

and Reintroduced (n = 22), grouped by ROH length. Each facet corresponds to a different ROH length bin a) 300 Kb–500 Kb, b) 500 Kb–1 Mb and c) >1 Mb. Boxplots show the distribution of segment counts per individual within each population group, representing the median and interquartile range. Jittered points indicate individual-level values.....pp95

Figure 3.11: Inbreeding coefficient (F_{ROH}) distribution across black-footed ferret population groups: Wild (n = 6), Captive (n = 26), and Reintroduced (n = 22), faceted by ROH length. Each facet represents a different ROH length bin (e.g., 300 Kb–500 Kb, 500 Kb–1 Mb, > 1 Mb). F_{ROH} is the percentage of the genome contained within ROH segments of the specified length. Boxplots display the median and interquartile range of F_{ROH} within each population group; individual points are overlaid as jittered dots to show within-group variation.....pp96

Figure 3.12: Analysis of gene family expansions and contractions across 20,363 gene clusters as calculated by Cafe v5.0 with an overall lambda (0.0052).12,097 gene families were shared amongst all eight genomes, 13,248 were shared amongst Mustelidae and 13,553 were shared amongst species belonging to Mustelinae.....pp97

Figure 3.13: Patterns of mutational load across the black-footed ferret populations. Aggregated counts of heterozygous and homozygous variant calls by mutation type (missense and loss-of-function (LoF)) across black-footed ferret population groups: Wild (n = 6), Captive (n = 26), and Reintroduced (n = 22). Variants were annotated using SnpEff and further classified by allele state (heterozygous or homozygous) and evolutionary origin, with derived alleles (unique to black-footed ferrets) shown in red and non-derived (ancestral) alleles shown in blue.....pp99

Figure 3.14: Predicted functional impacts of non-synonymous mutations at sites under positive selection in the black-footed ferret. A total of 364 Bayes Empirical Bayes (BEB) sites with elevated dN/dS ratios were identified using CODEML and analysed with SIFT4G to predict effects on protein function. SIFT generated predictions for 305 sites, classifying 66 variants as deleterious (p < 0.05; red) and 239 as tolerated (green), based on sequence conservation. The remaining 59 sites could not be assessed by SIFT, and their impact is therefore unknown (yellow).....pp100

Figure 3.15: Bar plots show the proportion of individual genotypes at variants annotated by SIFT, separated by prediction (deleterious vs. tolerated) and stratified by population groups (Captive, Reintroduced and Wild). Genotype states are classified as homozygous

reference (dark blue), heterozygous (orange), and homozygous alternate (light green). Each facet represents a different population group, allowing comparison of genotype composition across populations. The y-axis is split to provide a greater resolution for each proportion at every genotype state.....pp101

Figure 4.1: Phylogenetic tree of 11 mammalian genomes used to guide multi-genome alignments with Progressive Cactus, including eight mustelid species and three outgroup mammals. Mustelid genomes included: European polecat (*Mustela putorius*, GCA_902207235.1), domestic ferret (*Mustela putorius furo*, GCA_920103865.1), Steppe polecat (*Mustela eversmanii*, GCA_963422785.1), black-footed ferret (*Mustela nigripes*, GCF_022355385.1), stoat (*Mustela erminea*, GCA_009829155.1), weasel (*Mustela nivalis*), stone marten (*Martes foina*, GCA_040938555.1), and Eurasian otter (*Lutra lutra*, GCA_902655055). Outgroup genomes included human (*Homo sapiens*, GCA_000001405.15), domestic cat (*Felis catus*, GCF_000181335.3), and domestic dog (*Canis lupus familiaris*, GCF_011100685.1). No branch lengths were specified and instead was drawn from the literature (Koepfli et al. 2008) and the tree was visualised using iTOL.....pp114

Figure 4.2: Phylogenetic tree of 11 mammalian genomes used to guide multi-genome alignments with Progressive Cactus, including eight mustelid species and three outgroup mammals. Mustelid genomes included: European polecat (*Mustela putorius*, GCA_902207235.1), domestic ferret (*Mustela putorius furo*, GCA_920103865.1), Steppe polecat (*Mustela eversmanii*, GCA_963422785.1), black-footed ferret (*Mustela nigripes*, GCF_022355385.1), stoat (*Mustela erminea*, GCA_009829155.1), weasel (*Mustela nivalis*), stone marten (*Martes foina*, GCA_040938555.1), and Eurasian otter (*Lutra lutra*, GCA_902655055). Outgroup genomes included human (*Homo sapiens*, GCA_000001405.15), domestic cat (*Felis catus*, GCF_000181335.3), and domestic dog (*Canis lupus familiaris*, GCF_011100685.1). Branch lengths were estimated from a neutral model using PhyloFit (PHAST suite) based on aligned genome sequences, and the tree was visualised using iTOL.....pp119

Figure 4.3: Comparison of multi-genome alignment runs using Progressive Cactus, showing the percentage of each mammalian genome aligned to the black-footed ferret (*Mustela nigripes*) reference genome. Two alignment runs are presented: one without specified branch lengths in the input phylogeny (green) and one with specified branch lengths (orange). Alignment coverage was calculated using HAL alignment depth, revealing that specifying branch lengths improves alignment resolution for more distantly related species. This improvement is particularly evident for *Martes foina*, *Lutra lutra*, *Felis*

catus, and *Homo sapiens*, highlighting the benefit of incorporating phylogenetic branch lengths in multi-genome alignments.....pp120

Figure 4.4: Circos plot depicting a circular representation of the black-footed ferret chromosomes. All data are plotted using 200 Kb windows (a) alignment coverage of the ten other mammalian genomes to the black-footed ferret in the the Cactus HAL alignment, (b) gene density, (c) PhyloP conserved site density, (d) PhyloP accelerated site density, (e) genomic locations of population variants from 54 black-footed ferret individuals.....pp121

Figure 4.5: Frequency distribution of conserved (A) and accelerated (B) PhyloP scores across the black-footed ferret genome. Panel (A) represents the frequency distribution of conserved PhyloP scores >2 (2.48 - 8.46), which indicate genomic regions under evolutionary constraint, while panel (B) shows the distribution of accelerated PhyloP scores (20 - 0), reflecting regions potentially under positive selection or rapid evolution.....pp123

Figure 4.6: Comparison of PhyloP-identified conserved and accelerated bases between the black-footed ferret and domestic ferret genomes. PhyloP conservation and acceleration scores were calculated independently in each species. To allow for direct comparison, the domestic ferret gene model and the corresponding conserved/accelerated regions were lifted over to the black-footed ferret genome coordinates. (A) Fraction of genomic features (e.g., coding sequences, introns, intergenic regions) covered by conserved or accelerated bases. (B) Percentage of conserved or accelerated PhyloP bases overlapping each genomic feature. This approach enables assessment of the overlap between species-specific and shared conserved or accelerated elements.....pp124

Figure 4.7: Distribution of Derived Allele Frequencies (DAF) across black-footed ferret population variants found within A) conserved non-coding elements (intronic and intergenic regions) ($n = 1,844$) and B) non-conserved intergenic and intronic regions ($n = 141,789$). Derived allele frequencies range from 0 to 1 and are binned in intervals of 0.05. Variants are classified as rare ($DAF < 0.1$) or common ($DAF \geq 0.1$).....pp127

Figure 4.8: Visualisation of the variant falling within a non-coding region, classified as highly constrained by PhyloP and within an SP2 motif as classified by FIMO. A) PAX2 gene model depicting the location of the conserved variant, falling within an intergenic region upstream of the transcription start site. B) The plot demonstrates the PhyloP scores

across the constrained region within 25-bp windows; the location of the called motif is highlighted in orange. C) Motif sequence logo for the SP2 motif, demonstrating the most common nucleotide at the variant location is 'C', where the population variant is a 'T'. Multi-species sequence alignment taken from the whole-genome Progressive Cactus alignment demonstrates the conservation of that 'C' allele across all 11 mammalian genomes.....pp128

Figure 4.9: Visualisation of the variant falling within a non-coding region, classified as highly constrained by PhyloP and within an KLF4 motif as classified by FIMO. A) GGNBP2 gene model depicting the location of the conserved variant, falling within the 5' UTR region (which ranges from 35560209-35560512). B) The plot demonstrates the PhyloP scores across the constrained region within 25-bp windows; the location of the called motif is highlighted in orange. C) Motif sequence logo for the KLF4 motif, demonstrating the most common nucleotide at the variant location is 'A', where the population variant is a 'C'. Multi-species sequence alignment taken from the whole-genome Progressive Cactus alignment demonstrates the conservation of that 'A' allele across all 11 mammalian genomes.....pp129

Figure S2.1: Kinship coefficients between each of the 65 polecat and 8 domestic ferret individuals. KING estimates a relatedness-PHI score which can range from >0.354, [0.177, 0.354], [0.088, 0.177] and [0.044, 0.088] and corresponds to duplicate/MZ twin, 1st-degree, 2nd-degree, and 3rd-degree relationships respectively. Starred individuals (S17 & S0012_VWT665) were identified as having a 3rd-degree relationship (0.076) and S17 was removed from further analyses.....pp208

Figure S2.2: Pairwise F_{ST} values for whole genome SNPs between English, Welsh and European populations as calculated in VCFtools version 0.1.16 (Danecek et al., 2011).....pp209

Figure S2.3: The realised inbreeding coefficient (F_{ROH}) as calculated by PLINK version 1.9 for each individual, grouped by sampling location for the British (England and Wales) and European (Austria, France, Germany, Italy and Spain) polecat populations as well as Domestic ferret individuals. Individual F_{ROH} are overlaid on top of boxplots which show the upper, median and lower limits of the calculated F_{ROH} for each group.....pp209

Figure S2.4: Estimated mean effective population size (solid line) with 95% confidence intervals (shaded area) for the Italian polecat population (n = 5) across 200 generations (generations = 4 years) based on whole genome SNP data computed in GONE.....pp210

Supplementary Table 2.01: Details of the downsampled variant dataset and number of variants used for all analyses carried out in chapter two. Variants were called using GATK, with biallelic SNPs and quality filters (Q20, AC3) being applied for all downstream analyses. Further filters were then applied for population structure (LD, MAF).....pp210

Supplementary Table 2.02: Cross-validation errors from ADMIXTURE runs for all 72 samples using a population size (K) of 1-12. Ten replicates for each ancestry run were carried out using the custom python wrapper (admixturewrapper.py), with the standard deviation for each mean CV error presented below.....pp211

Supplementary Table 2.03: Output statistic tables for 1) kruskal wallis chi-squared test plus dunn posthoc test for genome-wide diversity estimates 2) ANOVA for the proportion of heterozygous sites across the genome per population and posthoc HSD tukey test 3) ANOVA for the inbreeding coefficient across populations and posthoc HSD tukey test 4) ANOVA for derived mutation load across populations for heterozygous and homozygous state missense and lof variants.....pp211

Supplementary material S2.1: Geographic location of the British european polecat samples utilised within this study. Grid reference, longitude and latitude references pertain to the location where carcasses were recovered during the Vincent Wildlife Trust 2014-2015 survey. Phenotype refers to the physical identification of the polecat (whether it looks more polecat-like or more ferret-like). Dataset refers to whether the sample was sequenced in this study, or by (Etherington et al., 2022).....pp214

Supplementary material S2.2: Metadata associated with the samples that were utilised in this study. Identifies the library preparation method, sequencing chemistry, depth of read coverage and proportion that the sample was downsampled by if applicable.....pp214

Supplementary material S2.3: Number of variants per sample (73 individuals), mean depth of sample as calculated by VCFtools and the proportion of missingness across whole-genome variants.....pp214

Supplementary material S2.4: Ancestry proportions for European, Welsh and English polecats, as well as domestic ferrets as calculated by ADMIXTURE for $K = 1 - 12$ pp214

Supplementary material S2.5: Population nucleotide diversity values (π), calculated in 20 Mb windows across the genome by VCFtools for English, Welsh and European polecats.....pp214

Supplementary material S2.6: Runs of homozygosity across individuals with a mean depth >6x coverage as calculated by PLINK. Number of segments, Froh and proportion of missingness across sampled sites are reported for each individual (n = 47).....pp214

Supplementary material S2.7: Aggregated snpEff results for European polecat population variants for Welsh, English and European individuals. These are reported as non-derived and derived alleles for two deleterious mutation categories; missense and LoF mutations.....pp214

Supplementary material S2.8: GONE output results for the Welsh population. Effective population size was modelled over 150 generations and replicated ten times. Mean effective population size is reported with the upper and lower confidence limits of the mean.....pp214

Supplementary material S2.9: GONE output results for the English population. Effective population size was modelled over 150 generations and replicated ten times. Mean effective population size is reported with the upper and lower confidence limits of the mean.....pp214

Supplementary material S2.10: Output statistic tables for 1) kruskal wallis chi-squared test plus dunn posthoc test for genome-wide diversity estimates 2) ANOVA for the proportion of heterozygous sites across the genome per population and posthoc HSD tukey test 3) ANOVA for the inbreeding coefficient across populations and posthoc HSD tukey test 4) ANOVA for derived mutation load across populations for heterozygous and homozygous state missense and lof variants.....pp214

Figure S3.1: Pairwise kinship coefficients between black-footed ferret individuals, where estimates of 0.17-0.35 indicate first-degree relations, second-degree relations correspond to 0.08-0.17 and third-degree relations as between 0.04-0.08.....pp215

Supplementary Table 3.01: Details of the variant dataset and number of variants used for all analyses carried out in chapter three. Variants were called using GATK, with biallelic SNPs and quality filters (Q20, AC3) being applied for all downstream analyses. Further filters were then applied for population structure (LD).....	pp215
Supplementary Table 3.02: Cross-validation errors from ADMIXTURE runs for all 54 samples using a population size (K) of 1-12. Ten replicates for each ancestry run were carried out using the custom python wrapper (admixturewrapper.py), with the standard deviation for each mean CV error presented in the table.....	pp216
Supplementary Table 3.03: Output statistic tables for 1) ANOVA plus HSD tukey posthoc test for genome-wide diversity estimates 2) ANOVA for the proportion of heterozygous sites across the genome per population and posthoc HSD tukey test 3) ANOVA for the inbreeding coefficient across populations and posthoc HSD tukey test 4) kruskal wallis chi-squared test plus dunns post hoc test for comparison of Froh across populations 5) ANOVA for derived and non-derived mutation load across populations for heterozygous and homozygous state missense and lof variants.....	pp216
Supplementary material S3.1: Sample ID, sequence ID, population group, geographic location, read depth per individual and sampling year for all 54 black-footed ferrets used within this study.....	pp222
Supplementary material S3.2: Ancestry proportions for Captive, Wild and Reintroduced black-footed ferret individuals as calculated by ADMIXTURE for K = 1 - 12.....	pp222
Supplementary material S3.3: Cross validation error mean estimates plus standard error for each replicate run (K1 -12) of ADMIXTURE to calculate ancestry estimates for all 54 black-footed ferret individuals.....	pp222
Supplementary material S3.4: Gprofiler results detailing enriched gene ontology, kegg pathway and processes for the expanded gene families within <i>Mustela nigripes</i> as identified by CAFE.....	pp222
Supplementary material S3.5: Gprofiler results detailing enriched gene ontology, kegg pathway and processes for the contracted gene families within <i>Mustela nigripes</i> as identified by CAFE.....	pp222

Supplementary material S3.6: Runs of homozygosity across all black-footed ferret individuals as calculated by PLINK. Number of segments, whole genome length of ROH and Froh are reported for each individual (n = 54).....pp222

Supplementary material S3.7: Aggregated snpEff results for population variants (reported as non-derived and derived for missense and lof mutation categories).....pp222

Supplementary material S3.8: Annotated SIFT output for all 364 non-synonymous variants identified with CODEML. The domestic ferret genome reference is reported, the genome location within the black-footed ferret, as well as the REF and ALT variant change and REF and ALT amino acid change. Gene ID and predicted deleterious of the variant are also reported.....pp222

Supplementary material S3.9: Output statistic tables for 1) ANOVA plus HSD tukey posthoc test for genome-wide diversity estimates 2) ANOVA for the proportion of heterozygous sites across the genome per population and posthoc HSD tukey test 3) ANOVA for the inbreeding coefficient across populations and posthoc HSD tukey test 4) kruskal wallis chi-squared test plus dunns post hoc test for comparison of Froh across populations 5) ANOVA for derived and non-derived mutation load across populations for heterozygous and homozygous state missense and lof variants.....pp222

Figure S4.1: Frequency distribution of conserved PhyloP scores across the black-footed ferret genome. PhyloP scores were originally calculated across 25-bp windows, but adjacent windows were merged to increase window size to 50-bp and scores were averaged across this region. The results of this is visualised within this plot. The resulting observed pattern of scores across the black-footed ferret genome is less undulating than the one observed in Figure 4.5, and determines that this pattern may be down to a technical artefact due to calling conservation across short genome regions, as opposed to a biological artefact.....pp223

Supplementary material S4.1: Alignment statistics for a whole genome-alignment of eleven mammalian species. Calculated by halAlignment stats, the coverage of species to the other ten species in the alignment is presented in proportion of the genome it aligned to.....pp223

Supplementary material S4.2: Gprofiler results detailing enriched gene ontology, kegg pathway and processes of the top 5% accelerated genes in *Mustela*

nigripes.....pp223

Supplementary material S4.3: Gprofiler results detailing enriched gene ontology, kegg pathway and processes of the top 5% conserved genes in *Mustela nigripes*.....pp223

Supplementary material S4.4: Describes the 38 predicted disruptive (+/- 0.05 PWM score) CNE variants, identified motif, PhyloP score, reference and alternate motif sequences as well as identified nearby genes. Allele frequencies for each variant within the black-footed ferret population (54 individuals) is also reported.....pp223

Supplementary material S4.5: Reported abnormal phenotypes, associated diseases and associated processes from the MGI database for genes found near CNE variants that were predicted as disruptive (+/-0.05 PWM score).....pp223

Publications and Contributions

Chapter 2 - Scripts for running GATK were adapted from scripts provided by Dr Graham Etherington.

Chapter 3 - Scripts for preparing CODON aware alignments, and the running and parsing of CODEML output were adapted from scripts provided by Dr Wilfried Haerty and Dr Dave Wright. Assistance was provided by Dr Anil Thanki to run GeneSeqtoFamily.

Chapter 4 - Scripts for running Progressive Cactus were adapted from scripts provided by Dr Will Nash and Dr Tarang Mehta.

The work described and presented within Chapter two has been associated with two publications, one which is already published in the Journal of Heredity, and the other published in Molecular Ecology.

Etherington, G. J., Ciezarek, A., Shaw, R., Michaux, J., Croose, E., Haerty, W., & Di Palma, F. (2022), Extensive genome introgression between domestic ferret and European polecat during population recovery in Great Britain. *Journal of Heredity*, 113(5), 500-515. <https://doi.org/10.1093/jhered/esac038>

Shaw, R., MacPherson, J., Kitchener, A.C., Etherington, G.J. and Haerty, W. (2025), Characterisation of the Historic Demographic Decline of the British European Polecat Population. *Mol Ecol*, 34: e70091. <https://doi.org/10.1111/mec.70091>

1.0 An introduction to conservation genomics, tools, practices and its application in the protection and recovery of endangered mammals

1.1 Abstract

This chapter explores the applications of conservation genomics in understanding and mitigating threats to mammalian biodiversity. It begins by outlining the primary challenges facing mammal populations globally, including habitat loss, human-wildlife conflict, and genetic erosion resulting from small population sizes and population bottlenecks. The theoretical foundations of conservation genomics are discussed, with particular emphasis on how genomic tools can be used to assess genetic diversity, detect inbreeding, and inform species recovery strategies. This chapter also introduces the comparative case study underpinning this thesis of the European polecat (*Mustela putorius*) and the black-footed ferret (*Mustela nigripes*), two closely related mustelids that have undergone severe population declines. Using genomic data, this thesis aims to evaluate the long-term genetic consequences of bottlenecks and hybridisation, demonstrating how conservation genomics can contribute to evidence-based management and the effective recovery of threatened species.

1.2 The Global Biodiversity Crisis

The severity of the Earth's ongoing biodiversity crisis has prompted significant global concern and a concerted effort to identify effective strategies for preserving ecological diversity. Anthropogenic pressures (stemming from human activity) are now widely recognized as the primary drivers of what scientists have termed the planet's sixth mass extinction, unfolding at an alarming and unprecedented pace (Barnosky et al., 2011; G. Ceballos et al., 2020). Current projections suggest that approximately 28% (47,187) of all listed species (172,600) on the Red List of the International Union for Conservation of Nature (IUCN) are at risk of extinction, with many expected to disappear within the next few decades if current trends continue (Díaz et al., 2019). A combination of interrelated factors, including climate change (Paniw et al., 2021), the spread of emerging infectious diseases (White & Razgour, 2020), the introduction of non-native species (Tedeschi et al., 2022) and pathogens (Fagre et al., 2022), habitat fragmentation (G. Li et al., 2022), and overexploitation of natural resources (Maxwell et al., 2016), is exacerbating this crisis. Collectively, these forces disrupt ecosystems, endangering species, which leads to impaired gene flow between and within populations, as well as leading to diminishing genetic diversity. Both of these are mechanisms that enable species to adapt to changing environmental conditions, and conserving them is critical in shaping both the survival and continued evolution of these species.

In response to these escalating threats, the scientific and conservation communities have developed a variety of frameworks aimed at enhancing the effectiveness of existing conservation tools. These approaches seek to identify priority areas for intervention, to improve our understanding of anthropogenic impacts on biodiversity, and to inform policy development. For example, the concept of biodiversity hotspots has been employed to focus conservation resources on regions with high species richness and endemism that are also under significant threat (Myers et al., 2000). Other strategies involve assessing the influence of land-use change and other anthropogenic pressures on terrestrial ecosystems (Locke et al., 2019), as well as implementing targeted actions to manage and reduce the impact of invasive species (Baquero et al., 2023; Dueñas et al., 2021; García-Díaz et al., 2022). These frameworks not only guide scientific enquiry and resource allocation, but also underpin the formulation, enactment, and enforcement of legislation designed to protect vulnerable species and habitats (Harris et al., 1995). Key legislative examples include the Wildlife and Countryside Act 1981 in the UK, which offers legal protections for native species, and the European Union's Habitats Directive (1992), which mandates the conservation of rare and endangered species and the habitats they depend on.

Effective conservation management relies on clearly defined conservation units, which inform legislative policies, guide resource allocation, and support enforcement efforts. One of the most widely recognised classifications is the evolutionarily significant unit (ESU) (Moritz, 1994; O. Ryder, 1986), which refers to populations that have undergone adaptive divergence and are thus considered evolutionarily important within a species (Allendorf et al., 2010; Frankham, 2010). Under the U.S. Endangered Species Act (ESA), these are referred to as distinct population segments (DPSs), allowing for the protection of reproductively isolated, genetically distinct populations essential to species survival (Coggins, 1974; Crouse et al., 2002).

The ability to identify and delineate conservation units increasingly relies on the use of genetic and genomic tools. While the significance of these tools has long been recognised in scientific research, there has been much discussion regarding a genomics research and integration gap, with the adoption of genetic tools only recently being implemented into conservation frameworks and policy (Neaves et al., 2024). The management and preservation of genetic diversity within populations has become a main focus of conservation initiatives as the protection of genetic diversity is critical for maintaining a species' adaptive potential and long-term viability (McNeely et al., 1990). Reflecting this shift, the Kunming-Montreal Global Biodiversity Framework - adopted at the 15th Conference of the Parties (COP15) to the UN Convention on Biological Diversity, introduced new genetic metrics as part of its global targets to halt biodiversity loss by 2030

(CBDa, b). By embedding genetic metrics into biodiversity monitoring, the framework aims to bridge the longstanding gap between genetic research and on-the-ground conservation management. This alignment has the potential to transform how species and populations are assessed, managed, and protected, ensuring that genetic considerations are no longer an afterthought but a foundational element of conservation policy.

This growing emphasis on genetics within conservation has been supported by rapid technological advances, particularly in the field of genomics. The increasing accessibility of open-source genomic tools (Allendorf et al., 2010; Hohenlohe et al., 2021; Supple & Shapiro, 2018) has transformed the scale and scope at which biological questions can be addressed. Genomics, which involves the study of large-scale DNA sequence data, has advanced dramatically through milestones such as the Human Genome Project (Lander et al., 2001) and the ENCODE project's investigation of gene regulation (ENCODE Project Consortium et al., 2007). These breakthroughs have catalysed further research into diverse fields, from understanding cancer genetics (Pan-Cancer Whole Genomes Consortium, 2020; Weinstein et al., 2013) and improving crop disease resistance (Greenwood et al., 2023; Zanini et al., 2022) to tracking the evolution of novel viral strains (Damas et al., 2020; Hu et al., 2024; Saravanan et al., 2022).

Similarly, predictions over the past decade were made concerning the revolution of genomics and its utility towards conservation efforts, with conservation biology being able to take advantage of current technological progress (Allendorf et al., 2010; Ouborg et al., 2010). Genomic tools now inform the identification of biodiversity hotspots, support comparative studies, and increase the resolution at which population-level processes can be studied. Advances such as the reduced cost of sequencing and the development of user-friendly de novo assembly software have enabled genome sequencing for non-model species, opening new frontiers in population and conservation genomics. Moreover, the release of high-quality reference genomes (Etherington et al. 2020; Totikov et al. 2021; Kliver et al. 2023; Johnson et al. 2018) has greatly enhanced comparative genomics (Mc Cartney et al., 2024; Rhie et al., 2021; Zoonomia Consortium, 2020) and provided essential reference panels for population-level analyses. The proliferation of genomic software and pipelines has further improved the statistical power and practical utility of these data, solidifying genomics as an indispensable tool in conservation planning and decision-making. For example, an important factor to consider when integrating genomics into conservation management is the implications and practicalities of sampling from threatened or elusive, wild populations. Advances in methods for extracting good quality DNA from samples taken in the field - either from roadkill or faecal samples - and software that can support analyses from low-coverage samples have made it possible to make

inferences on population structure, genetic diversity and population viability (Allio et al., 2021; de Oliveira et al., 2025; Kardos & Waples, 2024).

1.3 Underlying population genetics and evolutionary theory in the field of conservation genetics and genomics

Conservation genetics and genomics are grounded in the principles of evolutionary and population genetics (Kardos et al., 2021; Yoder et al., 2018). A central goal of conservation frameworks is to preserve genetic diversity, both among species and within populations, as it underpins a species' adaptive potential. Genetic variation is the raw material upon which natural selection acts and is essential to evolutionary processes (Frankham et al., 2004). Therefore, understanding how genetic variation arises and changes over time is fundamental to effective conservation management (Willi et al., 2022).

Throughout an organism's genome, nucleotide substitutions occur randomly, arising through errors during DNA replication. Rates of mutation differ greatly across species and regions of the genome (Loewe & Hill, 2010). According to the neutral theory of molecular evolution, most mutations are selectively neutral or nearly neutral, becoming fixed in populations largely through genetic drift and selection (Kimura, 1968). Mutations are often classified into three broad categories: deleterious, advantageous, or neutral. In reality, their effects can vary in intensity across a spectrum within each category. Only a small fraction of mutations confer beneficial effects and are fixed through adaptive processes, while others may be deleterious, reducing individual fitness and potentially contributing to mutation load (Eyre-Walker & Keightley, 2007). In conservation genomics, understanding mutation dynamics is essential for assessing the health and long-term viability of populations. The effectiveness of selection acting upon mutations depends on both the fitness effect of the mutation (the selection coefficient, s) and the effective population size (N_e). Therefore, in smaller populations, with reduced N_e , mutations are more likely to be acted upon by random genetic drift than selection, resulting in the potential for weakly deleterious mutations to accumulate in the genome (Crow, 2017; Kimura et al., 1963; S. Wright, 1931). These signatures have been studied to understand inbreeding depression and genetic erosion (Hedrick & Garcia-Dorado, 2016; Pinto et al., 2024). Mutation rates also play a critical role in determining how quickly genetic diversity can recover following population bottlenecks or founder events, which are key concerns in reintroduction and restoration programs (Kardos et al., 2021; Mueller et al., 2022; Ogden et al., 2020). In addition to studying mutation rates, conservation genomics also investigates the functional consequences of mutations across the genome, distinguishing between neutral variation and those that impact protein-coding or regulatory regions (Castellano et al., 2020; Eyre-Walker & Keightley, 2009; Wilder et al., 2023).

Recombination is the process by which genetic material is exchanged during meiosis, leading to new combinations of alleles and increasing genetic diversity. This process also acts to influence patterns of linkage disequilibrium across the genome by facilitating the breakdown of linkage blocks within the genome (Lynch, 2006). Similarly to mutation, recombination rates vary across the genome and species (Payseur, 2024). Within conservation genomics, studying and understanding recombination rates provides insights into the historical demographic structure of populations and the extent of inbreeding, as inherited blocks of linked DNA reveal patterns of ancestry and recent shared genetic history (Hohenlohe et al., 2021; Leitwein et al., 2020).

Selection acts as the other major driving force on genetic variability within organisms and populations. It acts upon alleles within a population, increasing the frequency of those that confer a fitness advantage (positive selection) and decreasing those with harmful effects (purifying or negative selection) (Gregory, 2009; Hill, 1982; J. Wang & Hill, 1999; Willi et al., 2022). The intensity and effectiveness of selective pressure are influenced by several factors, including environmental conditions, N_e , and the strength of the fitness effect conferred by the allele. In large populations, natural selection is more efficient at favoring beneficial mutations and removing deleterious ones, whereas in small populations, the effects of selection can be overpowered by genetic drift, leading to an accumulation of potentially harmful mutations (Kimura et al., 1963). In conservation contexts, identifying signatures of selection is important for detecting locally adapted traits and maintaining evolutionary potential in changing environments (Meek et al., 2023).

Genetic drift is a process which refers to random changes in allele frequencies, regardless of their fitness consequence, between generations (Frankham, 2005). This process has a more pronounced effect in small populations due to relaxed selective pressures, and as a result, drift can lead to the fixation of alleles and the loss of genetic diversity, increasing the likelihood that harmful mutations will persist and spread within a population (B. Charlesworth, 2022). This loss of genetic diversity harms the potential for a population to adapt to future environmental change and reduces its long-term survival (DeWoody et al., 2021).

Gene flow is the movement of alleles between populations through migration or admixture, and it can act conversely to genetic drift by introducing new genetic material to a population. This influx of variation can enhance a population's genetic diversity and adaptive potential (Frankham, 2015). Monitoring gene flow between populations is fundamental in conservation, as human activities can result in both the restriction and disruption of natural gene flow (Frankham et al., 2019). Habitat fragmentation, for example, can isolate populations and reduce connectivity, while translocations or

environmental changes may facilitate interbreeding between previously separated subpopulations or closely related species. Such human-mediated gene flow can lead to genetic swamping, where locally adapted genotypes are diluted, potentially resulting in outbreeding depression and the loss of fitness in hybrid offspring (Howard-McCombe et al. 2023; Wu et al. 2023).

1.3 Conservation genetics and genomics

Conservation genetics utilises the aforementioned theory and techniques of evolution and population genetics with the long-term goal of preserving species as dynamic entities so that they are capable of adapting to environmental change (Frankham et al., 2004; Primmer, 2009). The field of conservation genetics gained popularity when it became evident that genetic variation was vital for species survival (Frankel et al., 1981; van Oosterhout, 2020). Since the 1960s, population genetics and then conservation biology have relied on using markers that evolve nearly neutrally to infer genetic diversity, gene flow, inbreeding depression, phylogenetics and N_e in wild populations. The first approach in conservation genetics was the use of allozymes for population genetics (Hubby & Lewontin, 1966), which are allelic variants at protein loci that can be determined through electrophoresis. This method allowed researchers to estimate genetic variation within and between populations and played an influential role in early conservation efforts. For instance, high levels of homozygosity at 47 allozyme loci in two cheetah populations suggested a historical population bottleneck (O'Brien et al., 1983). Although allozymes were foundational in the early development of conservation genetics, their usefulness was later questioned, as they detect genetic variation indirectly and with low sensitivity (Schlötterer, 2004). As a result, they were soon surpassed by more accurate and informative DNA-based markers (Botstein et al. 1980; Williams et al. 1990).

Differently from allozymes, DNA markers made it possible to quantify mutations between different alleles, as well as increasing the number of sampled loci, thus increasing the statistical power of the analysis. Mitochondrial DNA (mtDNA) has been widely used in conservation genetics and has significantly contributed to the field (Avice, 2012). The primary advantage of using mitochondrial haplotypes in conservation is the ability to track genetic diversity and reconstruct phylogenies in vertebrates, as mtDNA is maternally inherited and not affected by recombination. Additionally, mtDNA has a higher mutation rate than nuclear DNA, which increases its analytical power (Brown et al., 1979). This method has also been widely used in taxonomic studies (Parson et al., 2000; Senfeld et al., 2020) and in wildlife forensics due to its cost-effectiveness and practical utility (Cronin et al. 1991; Hsieh et al. 2001; Johnson et al. 2014).

However, limitations of mitochondrial markers have been noted. Due to its lack of recombination, the mitochondrial genome is considered a single locus rather than multiple loci, which can limit the depth of genetic information it provides (Allendorf, 2017). This is problematic when inferring gene flow within and between populations, as estimating gene flow based on a single locus is inappropriate (Allendorf, 2017). Therefore, researchers determined that more markers at neutral loci would increase accuracy and power to understand evolutionary forces affecting wild populations (Allendorf, 2017; Ballard & Whitlock, 2004; Galtier et al., 2009). In addition, further research suggests that using mtDNA markers alone does not generate enough power and can lead to confounding results compared to using nuclear markers (Bagda et al., 2012). Despite these limitations, mitochondrial markers are still utilised to aid in conservation efforts due to their efficiency for taxonomic justification, phylogeny and demographic inferences.

In conjunction with the use of mtDNA for conservation genetics purposes, towards the end of the 20th century conservation genetic studies also introduced the use of PCR-based neutral markers known as microsatellites, or single tandem repeats (STRs). These are short DNA motifs (1-6bp) repeated in tandem which are ubiquitous across animal genomes. The advantages that microsatellites offered were two-fold. Firstly, they allowed for multiple polymorphic loci to be sampled across the genome, increasing the statistical power of prior studies. Secondly, was their ability to be amplified using PCR which exemplified the ease and reproducibility of their use within the laboratory (DeFaveri et al., 2013). Inferring genetic information from these small subsets of markers was easy to implement for numerous analyses; from deducing genetic diversity (Barrientos et al., 2024), estimating gene flow (Szatmári et al., 2021), parentage inference (García et al., 2002) and as markers for the resolution of population structure (Costa et al. 2013; Zimmerman et al. 2020). These markers are still used in conservation genetics today (Barrientos et al., 2024; Naude et al., 2020), however advances in technology have somewhat eclipsed their utility for ascertaining many aspects of population and conservation genomics (Zimmerman et al., 2020). Their limitations lie within the fact that many of these loci are neutral and are therefore not subject to selection. Therefore, genetic diversity assessments using these markers are not necessarily indicative of whole-genome level diversity. In addition, the number of these markers that are used to determine these metrics - tens of markers in comparison to the thousands to millions of SNPs that conservation genomics utilise - reduces the level of precision of genetic diversity estimates (Pérez-González et al., 2023; Zimmerman et al., 2020). Although microsatellites are highly polymorphic due to their elevated mutation rates, this same property can lead to homoplasy, where identical alleles arise independently rather than from a shared ancestor, thereby weakening inferences about genetic diversity and its relationship to fitness (Hansson & Westerberg, 2002; Pérez-González et al., 2023;

Zimmerman et al., 2020). Additionally, generating comparable microsatellite data across laboratories remains a challenge due to subjective allele size calling and variation in size determination methods, even when automated software is employed (LaHood et al., 2002; Zimmerman et al., 2020) .

As previously described, prior studies within the field have relied on a small subset of putatively neutral markers such as microsatellites or mitochondrial haplotypes to monitor genetic diversity, and infer phylogenetic relationships, population structure, and N_e . Genomic tools can provide these data with greater precision by using complete genomes or thousands of genome-wide markers that can be independently inherited, providing a sufficient level of variability to be able to detect changes in genetic diversity (Ouborg et al., 2010; Zimmerman et al., 2020) . The ability to distinguish between haplotypes in linkage disequilibrium (LD) can also enhance our understanding of key aspects of conservation genetics, including population structure, demographic history, and patterns of gene flow (Leitwein et al., 2020). Conservation biology has greatly benefited from the advancement and democratisation of sequencing technologies and their applications to the field (Allendorf et al., 2010).

Conservation genomics can be described as the use of genome-wide data to inform conservation biology (Supple & Shapiro, 2018). The application of genome-wide high-density markers, such as those provided by single nucleotide polymorphisms (SNPs). A SNP is a location in the DNA sequence at which individuals can vary by a single nucleotide. These markers come with several advantages, for example, technological advancements have generated SNP datasets which have greatly increased the number of loci that can be sampled across the genome and the fact that they are highly prevalent throughout the genome allows this marker type to provide a more accurate representation of genome-wide variation (Brumfield et al., 2003). SNPs are also found in coding and non-coding parts of the genome which highlights their uses to target both functional and demographic processes (Zimmerman et al., 2020).

The higher density of markers, alongside the genome-wide abundance of SNPs has led to a higher resolution and greater precision of estimating breeding programmes, ascertaining fine-scale population structure, and identifying functional correlation to genomic architecture (Hohenlohe et al., 2021). The transition to genomic data expands the tools which can be used as well as the range of biological questions that can be addressed. However, applying these methods on a large scale for non-model organisms requires extensive data from a sufficient proportion of the targeted population (Schweizer et al., 2021). Many genomic techniques can offer sequence data that represents a reduced

portion of the genome, such as RAD-Seq and SNP arrays (López et al., 2025; von Thaden et al., 2020). Whole genome sequencing allows access to non-coding regions that regulate gene expression without ascertainment bias, as well as whole transcriptome sequencing which allows the ability to quantify variances in gene expression (Christiansen et al., 2021; Theissinger et al., 2023).

Restriction site-associated DNA sequencing (RADSeq) is a widely used next-generation sequencing (NGS) method for capturing sequencing data from markers that are widespread across the genome (López et al., 2025). This method uses restriction enzymes that cut specific sites across the genome and sequences the regions adjacent to the cut sites (Baird et al., 2008). This is a relatively inexpensive method and has been used for sequencing partial genomes of non-model organisms, with no prior genomic information being needed (Andrews et al. 2016). This makes it an incredibly useful tool for conservation genomics (Cano et al., 2022; Grossen et al., 2018; Jackson et al., 2022). However, these methods do present several caveats in their inability to target specific regions of the genome, and this can result in missing data whereby regions across the genome that could be under selection are missed from the pool of sampled loci (Holliday et al., 2018; Shafer et al., 2017).

1.4 Conservation Genomics in Application

The advances in genomic technologies have led to an abundance of tools that can address a wide range of biological questions. These tools are essential for tackling conservation challenges, such as preserving genetic diversity and understanding the effects of hybridisation and adaptation. However, there remains a significant gap between the availability of genomic data and its application in conservation management (Neaves et al., 2024). One major challenge is translating academic research into practical conservation practices and policies (Britt et al., 2018; Vernesi et al., 2008). Analysing genetic and genomic data requires specialised knowledge of statistical models, software, and scientific terminology, which can be a barrier. Additionally, the lack of adequate computational infrastructure, particularly among NGOs, conservation groups, or less-funded research organisations, hinders their ability to fully utilise and analyse the vast amounts of data generated by next-generation sequencing (NGS) technologies. This issue is further compounded by paywalls in scientific journals, which create a divide between research and its practical implementation in conservation management and policy (Shafer et al., 2015).

Another challenge lies in whether conservation practitioners and stakeholders recognise the importance of genetic and genomic data for the survival and recovery of populations. Community inertia and lack of adoption of these novel technologies, the "just because

wecan, doesn't mean weshould" sentiment, was once a significant concern. However, the clear benefits of generating genomic data for non-model populations have since become evident (McMahon et al., 2014).

It is also critical to choose the appropriate data and analytical methods for specific conservation questions. For example, microsatellites and mitochondrial DNA (mtDNA) markers remain widely used in conservation genetics due to their relatively low cost and ease of analysis compared to NGS methods (Barrientos et al., 2024; Klimova et al., 2022; Miller-Butterworth et al., 2021). Although the cost of genome sequencing is decreasing, challenges related to downstream analysis and interpretation persist, meaning that more traditional markers may continue to be preferred (Allendorf et al., 2010).

1.4.1 Comparative genomics

As sequencing technologies become more advanced and accessible, and as researcher expertise grows, the generation of whole-genome data across diverse taxa is accelerating (Challis et al., 2023; Ellegren, 2014; Grueber, 2015; Schell et al., 2025). A key driver of this trend is the significant reduction in the cost of sequencing associated with increased computing capability making large-scale genomic studies more tractable (May et al., 2025; Wetterstrand, 2013). Additionally, improvements in speed and efficiency have further reduced the time required to produce and analyse these datasets (van der Valk & Dalèn, 2024). The increasing accessibility of high-quality genomic data is expanding the potential for comparative genomic analyses across both populations and taxa. This shift has driven the creation of *de novo* genome assemblies through large-scale international initiatives such as the Zoonomia Project (Zoonomia Consortium, 2020), the Vertebrate Genome Project (Koepfli et al., 2015; Rhie et al., 2021), the Earth BioGenome Project (Lewin et al., 2018), Bird 10K (G. Zhang, 2015), and Bat1K (Teeling et al., 2018). Despite differences in scope and focus, these initiatives share a common goal: to produce high-quality reference genomes across a wide diversity of species. The availability of these resources has tremendous potential to support a wide range of multidisciplinary research efforts. This potential was exemplified during the SARS-CoV-2 pandemic, where comparative genomic analyses helped identify potential animal hosts of the virus (Damas et al., 2020), highlighting the critical role of genomics in global health and disease surveillance.

A highly contiguous, chromosome-level reference is the much sought-after gold standard for genomic analysis today (Brandies et al., 2019; Totikov et al., 2021). Once a reference genome has been generated, there are multiple ways in which that data can be integrated with other forms of conservation data, such as population-level whole genome resequencing, reduced representation sequencing, or the production of genome enabled marker panels (Hogg et al. 2025). This provides a valuable resource to researchers to

enable the extraction of millions of variable markers present for genome-wide analyses (S. Zhao et al., 2012). Producing a reference genome for a species in decline has proven to engage practitioners, as well as being a starting resource to inform on genetic diversity, demography, disease susceptibility, hybridisation and adaptation in populations in the wild (Brandies et al., 2019; Hogg, 2024).

Despite the generation of reference genomes being the back-bone to understanding genetic variation across populations, errors can still occur in the alignment to these references (Matthews et al., 2024; Secomandi et al., 2025). This may be due to poorly sequenced regions, or divergent areas within the genome. This can lead to a misrepresentation in the understanding of genetic variability within a population, and lead to reference bias (Huang et al., 2013; Matthews et al., 2024). Developing technology is enabling systematic capturing of genomic variability through the utility of pangenomes (Liao et al. 2023; Secomandi et al. 2023; Secomandi et al. 2025; Affini et al. 2025; Li et al. 2023). Pangenomics involves the curation of genomic information from multiple individuals which enables the inclusion of multiple alternate alleles, encompassing a more comprehensive view of genomic variation within a population (Sherman and Salzberg 2020; Liao et al. 2023). The accessibility of these types of tools for non-model species is still limited, however their potential as a tool within conservation genomic projects has been observed. For example, they have been utilised in detecting large structural variants or genomic rearrangements which underpin fitness traits (Li et al. 2023; Fang and Edwards 2024), crop breeding and improvement (Jayakodi et al., 2020; Zanini et al., 2022), and in uncovering mechanisms of adaptation in populations (Li et al. 2023; Hou et al. 2024). Further utilities of pangenomics in the field of conservation genomics would be to detect functional genomic variation in adaptation to environmental pressures (Secomandi et al., 2023) and discerning gene flow in hybrid zones (Hübner et al., 2019).

Another advantage of generating high-quality, annotated genome data for non-model organisms is that it enables researchers to conduct comparative genomic analyses across species. A key benefit of comparative genomics is its ability to detect signatures of selection acting on the genome and to identify regions of genetic variation associated with unique traits (Alföldi and Lindblad-Toh 2013; Sullivan et al. 2023; Xue et al. 2023; Wright et al. 2024). These approaches also enhance our understanding of evolutionary constraints by detecting regions of the genome that have remained unchanged over millions of years due to purifying selection. Identifying such regions helps us understand the functional impacts of genetic variation within these regions, enabling predictions about which variants may lead to phenotypic changes such as increased disease risk (Sullivan et al., 2023) or adaptive responses to environmental factors (Feng et al., 2020; Sebastianelli et al., 2024).

1.4.2 Functional genomics, ultra conserved elements and the annotation of variants

Beyond evolutionary insights, comparative genomics is increasingly recognised as a valuable tool in conservation biology (Grueber, 2015; van der Valk et al., 2024; Wilder et al., 2023). It enables the assessment of selective pressures acting on small or endangered populations, the identification of genomic regions impacted by hybridisation and introgression (Howard-McCombe et al. 2023; Zhang et al. 2023; Chen et al. 2025), and the discovery of species-defining gene losses or gains (Johnson et al. 2018; Long et al. 2024). Moreover, recent studies have shown that constrained common variants are more enriched for heritable diseases, which may pose a greater risk in genetically depauperate populations (Wright et al. 2017; Fraik et al. 2020; Sullivan et al. 2023).

Ultraconserved elements (UCEs) are DNA sequences that are highly conserved across diverse species and evolutionary lineages. They are typically defined as regions of approximately 100–200 base pairs in length with a high degree of sequence identity, often ranging from 50% to 97% identity between species (Bejerano et al., 2004; Cummins et al., 2024; Faircloth et al., 2012). Based on the hypothesis that such high conservation is driven by purifying selection, the conservation of these elements across taxa suggests that they serve important biological functions (Boffelli et al. 2004; Margulies et al. 2007; Christmas et al. 2023; Kuderna et al. 2024; Thomas et al. 2003). While exonic constraint has advanced our understanding of gene models in various organisms, comparative genomics has also highlighted the significance of conserved non-coding regions. In particular, highly conserved non-coding elements have been associated with regulatory functions, including enhancer activity in mammals (Hecker & Hiller, 2020; Kaplow et al., 2023; Lindblad-Toh et al., 2005). UCEs have been found to cluster around developmental and neurological genes in vertebrates, suggesting that variation found within these regions could potentially disrupt the regulation of complex phenotypes (McLean et al., 2011) or reflect lineage specific adaptations (Feng et al., 2020). Since UCEs and other conserved elements can be identified across genomes, variants occurring within these regions can be scored for their potential functional impact. Comparative genomics provides powerful tools for this purpose: phylogenetic methods such as Genome Evolution Rate Profiling (GERP) (Davydov et al., 2010; Huber et al., 2020), PhastCons, PhyloP (Pollard et al., 2010) and CADD (Kircher et al., 2014) frameworks quantify evolutionary constraint by evaluating sequence conservation across multiple species. These scores reflect the strength of purifying selection acting to maintain sequence integrity and, as such, variants in highly constrained regions are often interpreted as putatively deleterious. This ability to prioritise variants based on evolutionary signatures is especially important for interpreting non-coding regions (Groß, Bortoluzzi, et al., 2020a), where functional annotation is limited

and especially in non-model organisms where experimental validation is often challenging (Speak et al., 2024).

However, nucleotide-level conservation alone does not necessarily imply conservation of biological function; recent studies have shown that even in the absence of strict sequence conservation, these regions can retain conserved regulatory patterns, such as transcription factor binding motifs that contribute to gene regulation (Kaplow et al., 2022; Snetkova et al., 2021). Leveraging these conserved patterns of regulatory activity, recent work has demonstrated the advancement of technologies in the utility of machine learning approaches that can detect shared features across taxa and use them to predict regulatory elements in less-characterised or divergent systems (Zhao et al. 2022; Kaplow et al. 2023).

1.5 The consequences of population bottlenecks

This thesis focuses on the genetic consequences of population bottlenecks - events in which a species undergoes a sharp decline in population size - and applies genomic tools to detect the signatures of these demographic shifts. As previously discussed, foundational theories in conservation genomics highlight how such bottlenecks can leave detectable marks on the genome through reduced genetic diversity, increased homozygosity, and elevated inbreeding. These genetic consequences can hinder conservation efforts and threaten long-term species viability.

1.5.1 Small populations

Population size is an important factor in managing conservation priorities for species. One main objective of conservation biology is to understand the mechanisms behind small or declining populations for their protection and long-term survival (Caughley, 1994). Threatened or fragmented populations, as well as populations of endemic island species, can have a relatively low population size. This often leaves them vulnerable to extinction risk through several mechanisms. These mechanisms are otherwise referred to as Allee effects and they can be both ecological and genetic. These effects highlight a situation where population size is related to the overall fitness of the individuals within that population. Weak Allee effects will not detrimentally affect the survival of a population, whereas strong Allee effects can (Wittmann et al., 2018).

Small populations with high relatedness among individuals are particularly vulnerable to negative fitness effects. Population genetics theory predicts that larger populations will have both higher neutral and deleterious genetic mutations (Kimura 1968; Charlesworth and Charlesworth 1987; Kardos et al. 2021). However, genetic drift and low N_e can have severe consequences, as evolutionary pressures like drift can quickly lead to the fixation

of harmful mutations within the gene pool (discussed in section 1.2) (Lande 1994; Frankham 2005). This may result in individuals with reduced fitness in subsequent generations. The loss of genetic variability can also lead to inbreeding depression, where individuals exhibit lower fitness compared to outbred conspecifics (Newman and Pilson 1997; Bijlsma et al. 2000; Keller & Waller 2002; Shorter et al. 2017).

The Florida panther (*Puma concolor c cougar*) is one example of a small population that was heavily impacted by the loss of genetic variation and accumulation of deleterious mutations. The population had shrunk to between 20-25 individuals in the 1990s before conservation management frameworks intervened. Phenotypic abnormalities, including kinked tails and thoracic cowlicks, were monitored in individuals, alongside indicators of reduced health such as poor sperm quality and increased susceptibility to infectious pathogens (Johnson et al. 2010).

1.5.2 Inbreeding low effective population size (N_e)

Inbreeding is common in small, isolated, and fragmented populations, where it can lead to inbreeding depression. This occurs when an increase in homozygosity unmasks and fixes deleterious mutations, contributing to a downward spiral known as the extinction vortex (Gilpin and Soule 1986; Hedrick and Garcia-Dorado 2016), which threatens a population's viability. Breeding among closely related individuals can result in genotypes more susceptible to genetic drift and less phenotypic variation due to reduced genetic diversity (Frankham et al. 2017). In a changing environment, this lack of adaptability poses a significant risk to wild populations (Harrisson et al. 2019; Robinson et al. 2023).

Conservation genetics can reveal reductions in genetic diversity, but it often lacks the ability to reconstruct the demographic history leading to inbreeding events (Costa et al. 2013). Genomic tools, however, provide a more detailed analysis without needing historical data. For instance, (Ekblom et al., 2018) used whole-genome sequencing of 10 Scandinavian wolverines to estimate N_e and structure. Their findings showed a long-term decline in genetic diversity, spanning thousands of years, providing valuable insights for conservation management in a non-invasive way.

Genomic tools thus offer powerful insights into population demography, N_e , and inbreeding levels, even from a small sample of individuals. These are imperative parameters to measure as the relationship between the consequences of low effective population size, inbreeding and mutational load is also often variable. Evidence demonstrates that some populations which have undergone bottlenecks and sustained at a low effective population size for years, have lower than expected mutational loads - most likely due to purifying selection acting to purge deleterious mutations (van der Valk et al. 2019; von Seth et al.

2021; Grossen et al. 2020; Wootton et al. 2023; Lan et al. 2024; Yang et al. 2025; Robinson et al. 2022). However, other examples such as the Scandinavian Wolf (*Canis lupus*), Whooping Crane (*Grus americana*) and Bull Kelp (*Nereocystis luetkeana*) populations, demonstrate a history of low N_e and high inbreeding levels leading to fixation rather than purging of deleterious mutations (Smeds and Ellegren 2023; Fontseré et al. 2025; Bemmels et al. 2025). This demonstrates that these measurements (inbreeding, mutational load and N_e) should be wholly considered when estimating the genomic health and long-term viability of a population (van der Valk et al., 2019; Wootton et al., 2023).

1.5.3 Runs of Homozygosity (ROH) as a consequence of low N_e and inbreeding

Genome-wide scans using microsatellites in the 1990s led to the discovery of uninterrupted runs of homozygous genotypes in individuals that were closely related (Broman & Weber, 1999). These long ROH were then discovered to be prevalent throughout the human genome and across different populations (Gibson et al., 2006).

With the advancement in genomics, more studies have been using these tools to more accurately infer inbreeding coefficients. The inbreeding coefficient (F) measures and identifies the relative proportion of an individual's genome that is identical by descent (IBD) and was made possible through the use of SNP markers (Wright 1969; Kardos et al. 2016). Furthermore, by being able to generate short genomic reads and mapping them back to a reference genome, I can genotype whole portions of the genome that are homozygous, no matter how small these may be. Estimating the amount of ROH in the genome, or the proportion that is IBD, has been associated with the concept of inbreeding as it is the result of relatedness between an individual and their parents (Ceballos et al. 2018; Ceballos et al. 2018). When the inherited haplotypes from both parents are homozygous at each locus, it is suggestive of them being related to a common ancestor (Saremi et al., 2019).

This analysis is extremely useful in population genomics today, as I can recreate the demographic history of a population by looking at the presence and length of ROH within the genome (Saremi et al. 2019; Grossen et al. 2020; Robinson et al. 2018). Long runs of these homozygous sites can infer little to no recombination within the population, which suggests a fairly recent inbreeding event (Xue et al. 2015; Saremi et al. 2019; Stuart et al. 2025), although some haplotypes may be common in the population due to other factors such as selection and drift. Shorter runs of homozygous segments are indicative of more historical inbreeding events due to recombination breaking up segments of ROH (Saremi et al., 2019). ROHs are more enriched areas for homozygous alleles, potentially increasing the number of deleterious mutations occurring within the genome (Szpiech et al., 2013).

1.5.4 Accumulation of deleterious mutations in small populations

In small populations, genetic drift plays a stronger role relative to selection, reducing the efficacy of purifying selection in removing harmful mutations (Ohta 1973). As a result, weakly deleterious mutations are more likely to persist and even reach fixation. This accumulation of deleterious alleles contributes to the genetic load, which is the reduction in the mean fitness of a population due to the presence of harmful alleles, reducing population viability over time. The component of genetic load caused specifically by recent and recurrent deleterious mutations is referred to as the mutation load (Grossen and Ramakrishnan 2024). The elevated homozygosity, typical of small, often recently bottlenecked populations further exposes recessive deleterious mutations, compounding the effects of inbreeding depression (Bortoluzzi et al., 2020; Kimura et al., 1963). This process of converting the masked load (recessive mutations with no fitness bearing on current population) into the realised load (impactful, homozygous mutations) can accelerate genetic decline, increasing extinction risk in already vulnerable species (Dusseux et al., 2023; Mathur & DeWoody, 2021; Pinto et al., 2024).

However, this same exposure of deleterious alleles through inbreeding also creates an opportunity for purifying selection to act. In this way, population bottlenecks can lead to the purging of highly deleterious recessive mutations, even as mildly deleterious ones continue to drift and accumulate. This dynamic, where inbreeding both reveals and removes harmful variants, has been observed across multiple species that have experienced varying degrees of demographic contraction (Grossen et al. 2020; López-Cortegano et al. 2021; Stuart et al. 2025; Robinson et al. 2018; Robinson et al. 2022). These findings underscore the importance of preventing populations from reaching critically low effective sizes, where the balance between purging and accumulation can tip toward long-term genetic decline (Bortoluzzi et al., 2020).

1.5.5 Hybridisation

Hybridisation leads to genetic material being transferred after interbreeding of genetically distinguishable taxa (Mallet, 2005). The study of hybridisation has been an important tool in understanding evolutionary biology and is thought to be a relatively common phenomenon within nature (Taylor & Larson, 2019). Biodiversity loss is a major consequence of hybridisation and can be driven by several factors, such as human activities (Bohling & Waits, 2011; Smith et al., 2018), species invasions (Delgado-Acevedo et al., 2021), and habitat fragmentation (Heppenheimer et al., 2018).

Several categories of hybridisation have been described (Allendorf et al., 2001; Quilodrán et al., 2020) to elucidate the consequences of hybridisation events in conservation policy. They highlight two possible scenarios whereby hybridisation events can result in 1)

infertile offspring or 2) genome exclusion where introgression is prevented (Quilodrán et al., 2018). A third scenario describes introgressive hybridisation (Abbott et al., 2013), which describes the interspecific gene exchange between species that can result in adaptive radiation (Seehausen, 2004) and initiate new evolutionary trajectories (Grant and Grant 1994).

Understanding the impacts that introgression and hybridisation have on wild populations is important for conservation implications, especially in light of changing environmental conditions. Hybridisation has been argued as the basis of evolutionary change, which can bring about events that allow for adaptive radiation in species (Jones et al., 2018; Kays et al., 2010). Prior to the advances in high-throughput sequencing, reliably tracking hybridisation events in the genome was difficult (Harrison & Larson, 2016). Genomics can offer the opportunity to understand the frequency of hybridisation within evolutionary lineages (Figueiró et al., 2017), identify potential 'hybrid zones' where these events occur, and determine the likelihood of genomic introgression after hybridisation.

Hybridisation between snowshoe hares (*Lepus americanus*) and black-tailed jack rabbits (*Lepus californicus*) has been demonstrated to enable rapid adaptation of a brown phenotype in snow-shoe hares in areas with a warmer winter period (Jones et al., 2018). Additionally, rapid adaptation in Killifish (Cyprinodontiformes) to an increasingly polluted environment was possible through introgressed loci (Oziolor et al., 2019). These events have also been demonstrated to boost genetic diversity in populations (Kolbe et al., 2004). Therefore, in some cases, hybridisation can be beneficial to populations that are suffering from the effects of inbreeding as a way to increase fitness and counteract inbreeding depression (Grant et al. 2003). Interspecies hybridisation has been demonstrated as a potential route for evolutionary rescue, with botanists first noting the important role that natural hybridisation can play in the generation of new biodiversity (Goulet et al. 2017). By using genomic tools, these events have been well documented across different taxa, including mammals (Figueiró et al., 2017), inclusive of hominids (de Manuel et al., 2016), fishes (Meier et al., 2017), reptiles (Barley and Cole 2025), insects (Hinojosa et al. 2022), amphibians (Chen and Pfennig 2020) and birds (Blom et al. 2024). This leads to an argument that hybridisation can benefit the preservation of biodiversity by producing new species, despite driving the extinction of native species (Quilodrán et al., 2020).

While the advantages of hybridisation have been extensively demonstrated, hybridisation can also have negative impacts on the fitness of the wild counterpart (Allendorf et al., 2001). It has been demonstrated that hybridisation can impact the genetic integrity of a species and therefore potentially the legal protection offered to the wild or hybrid populations (Daniels et al., 2001; Howard-McCombe et al., 2021, 2023; Senn et al., 2019;

Senn & Pemberton, 2009; Smith et al., 2018). In instances where extensive admixture occurs - hybrid swarms, it can be difficult to distinguish between lineages (Howard-McCombe et al., 2023; Senn et al., 2019). For example, the critically endangered Scottish wildcat (*Felis silvestris silvestris*), which is a subpopulation of the European wildcat and Great Britain's only surviving native felid (Howard-McCombe et al., 2021; Kitchener et al., 2005). The main threat to this species is introgression of domestic cat (*Felis catus*) genes (Beaumont et al., 2001; Breitenmoser et al., 2019). This has led to difficulty in discerning a genetic and phenotypic difference between true wildcats and wildcat hybrids. This also causes increasing difficulty in being able to manage populations, implement protective legislation and monitor the current population (Driscoll et al., 2011; Kilshaw, 2015; Senn et al., 2019).

The increased genomic resolution at which hybridisation can now be studied, especially in those of elusive populations, can provide a beneficial insight into the role hybridisation has played in the evolution of species. These tools also provide conservationists with a better understanding of how hybridisation can impact species. For example, if a population can benefit from such events, to align the management of species and protective policies to reflect this (Quilodrán et al., 2020).

1.6 Conservation in practice: methods to prevent the loss of species

1.6.1 Genetic rescue

Genetic rescue is a powerful tool in the restoration of species and was fundamentally driven by conservation genetics. This conservation strategy is aimed at restoring declining populations by increasing genetic diversity and fitness through gene flow (Tallmon et al., 2004; Whiteley et al., 2015). This typically involves translocating individuals between populations, either through managed interventions or natural dispersal. While genetic rescue has led to notable successes (Frankham, 2015), such as the recoveries of the Californian condor (*Gymnogyps californianus*) (Ralls & Ballou, 2004) and Florida panther (*Puma concolor coudar*) (Johnson et al. 2010), its effectiveness depends heavily on the careful selection of source populations.

A key consideration in genetic rescue is the assessment of the most suitable source population. One of the risks associated with genetic rescue is outbreeding depression, where mixing genetically distinct populations can reduce overall fitness (Frankham et al., 2011). Additionally, genetic swamping where introduced genes overwhelm locally adapted variants can result in the loss of local adaptation and even the extinction of unique conservation units (Edmands, 2007; Meek et al., 2023). Another risk that may arise in the implementation of genetic rescue is the introduction of rare, deleterious alleles that may be additive to the population's mutational load. These risks underscore the importance of

using population genomics to assess functional differences and adaptive variation across populations, informing translocation decisions and improving the likelihood of long-term success (Fitzpatrick et al., 2020; Whiteley et al., 2015).

The case of the Isle Royale wolf (*Canis lupus*) exemplifies the complexity of genetic rescue. The introduction of a single male migrant initially improved fitness, but later generations suffered from increased inbreeding and a declining population due to the fixation of recessive deleterious alleles (Hedrick et al., 2014, 2017, 2019). Genomic analyses revealed high homozygosity, helping identify the underlying causes and highlighting the role of genomic tools in refining future management (O'Brien et al., 2017; Supple & Shapiro, 2018). This further highlights the need to wholly assess the aforementioned parameters (ROH, mutational load, N_e , inbreeding and genetic diversity) to determine genomic health for selecting individuals for genetic rescue (van der Valk & Dalèn, 2024).

1.6.2 Captive breeding programmes and reproductive technologies

Captive breeding programmes play a crucial role in modern conservation, serving as a safeguard against extinction for species facing severe population declines. These efforts aim to maintain genetic diversity and ensure that captive populations remain representative of their wild counterparts, with the long-term goal of supporting reintroductions and facilitating genetic rescue. Best practices for successful conservation breeding and the IUCN's Species Survival Commission have developed guidelines to help determine when *ex-situ* management is appropriate, emphasizing the importance of sourcing genetically diverse founders and sustaining releases over extended periods to maximize genetic contributions (Frankham et al., 2002; McGowan et al., 2017).

However, in the most recent past, captive populations have been established when populations are already declining, and as a result, founders for insurance populations are often drawn from already declining or genetically compromised populations (Klimova et al., 2022; Wisely et al., 2003), potentially undermining recovery efforts (Klimova et al., 2022; Robertson, 2006; Swinnerton et al., 2004; Willoughby et al., 2017). Moreover, establishing new populations with a limited number of individuals can lead to additional genetic bottlenecks and increased inbreeding, compounding the risk of inbreeding depression (Charlesworth and Charlesworth 1987; Hedrick and Kalinowski 2000; Boakes et al. 2007).

Traditionally, the genetic management of captive populations has relied on pedigrees and studbook records. While these tools aim to minimise inbreeding, they often assume unrelated founders, an assumption increasingly challenged by genomic evidence (Hogg et

al., 2019) Recent advances in genetic and genomic technologies now enable more accurate estimates of relatedness and genetic diversity, enhancing the practical implementation of conservation breeding and improving the genetic quality of reintroduced individuals (Humble et al., 2023; Nelson et al., 2024; Speak et al., 2024).

The increasing integration of genomic tools into conservation has expanded the range of strategies available to manage genetic diversity in endangered species. Advanced reproductive and emerging biotechnologies - including in vitro fertilisation (IVF), gene editing, and cloning via somatic cell nuclear transfer (SCNT), are now being explored as means to mitigate the impacts of genetic drift, low effective population size, and the loss of genetic diversity in endangered populations. These approaches hold significant potential for restoring previous allelic diversity to populations, particularly when combined with cryopreserved tissue or cell lines stored in biobanks (Swegen et al., 2023; Wisely et al., 2015).

Notable examples include the Przewalski's horse (*Equus przewalskii*) and the Black-footed ferret (*Mustela nigripes*), both of which were rescued from severe genetic bottlenecks, founded by just 12 and 7 individuals respectively and have since been sustained through captive breeding and reintroduction programmes (Der Sarkissian et al., 2015; Wisely et al., 2015). In both cases, SCNT has been successfully used to produce genetically viable individuals from cryopreserved material (Novak et al., 2024, 2025). SCNT, the technique that famously led to the cloning of Dolly the sheep (*Ovis aries*) in 1996 (Campbell et al., 1996), involves transferring the nuclear DNA from a somatic (non-reproductive) cell into an enucleated oocyte, producing an offspring that is genetically identical to the donor.

The value of this technology in supporting captive breeding goals has spurred a broader interest in advanced reproductive techniques and long-term genetic preservation. Drawing from developments in medicine and agriculture, zoos and conservation programmes are increasingly integrating cryopreservation, popularly referred to as 'frozen zoos', into their management strategies through the use of biobanks (Breithoff & Harrison, 2020; Pérez-Espona, 2021). These resources not only safeguard genetic material from endangered species but also provide the foundation for future applications of cutting-edge techniques.

One promising area involves the generation of induced pluripotent stem cells (iPSCs) from somatic cells, which can be differentiated into gametes (sperm or eggs) for use in assisted reproduction or cloning. While still largely experimental, this technique holds transformative potential for conservation, especially for functionally extinct species where few or no reproductive individuals remain (Stanton et al. 2019; Wu et al. 2024). For

example, these technologies are hoped to support the critical case of the northern white rhinoceros (*Ceratotherium simum cottoni*), a subspecies now functionally extinct due to poaching, with only two non-reproductive females alive today. Conservation efforts are focused on leveraging cryopreserved somatic cell lines from 12 deceased individuals, with the goal of generating viable embryos through iPSC-derived gametes or SCNT (Wilder et al., 2024). A benefit to the utilisation of this technology is that embryos produced via cloning or in vitro gametogenesis from iPSCs can be repeatedly reintroduced to the population over multiple generations. This would ultimately reduce the impacts of a single founder event, establishing a genetically diverse founder population and ultimately restore the species (Wilder et al., 2024; Wisely et al., 2015).

1.7 Thesis aims

The overarching aim of this thesis is to use genomic tools to detect signatures of population bottlenecks. To do this, I will generate genome-wide data and develop bioinformatic tools and pipelines to compare two species that have experienced bottlenecks of differing severity. Both species are actively monitored and protected under species recovery and biodiversity frameworks, and the findings are intended to support ongoing conservation efforts. Comparative analysis of two closely related species with divergent demographic histories provides a powerful natural framework for assessing both the impact of demographic history, but also the effectiveness of conservation measures applied in each case. Here, I will introduce the study species, the European polecat (*Mustela putorius*) and Black-footed ferret (*Mustela nigripes*). Below, I briefly outline the biology, demographic history and conservation management of these species which offer the essential context for analyses, interpretations and broader implications of findings discussed in subsequent chapters.

1.7.1 The European polecat

The European polecat is a medium-sized carnivore from the Mustelidae family, found throughout much of mainland Europe. The polecat's range extends from the Atlantic coastline in the southwest, across central and eastern Europe, all the way to the eastern edges of the Ural Mountains (Croose et al., 2018; Skumatov et al., 2016). Like other members of the Mustelidae family, the polecat has evolved to thrive in diverse habitats, including grasslands, woodland margins, agricultural landscapes, coniferous forests, and riparian environments (Birks and Kitchener 1999; Virgós 2003; Baghli et al. 2005). The polecat's adaptation to its environment also extends to its hunting ability, where it has adopted a diverse generalist diet ranging from rodents, lagomorphs, amphibians, and carrion (Sainsbury et al., 2020). However, its dietary requirements are limited by and specific to particular regions. For example, in Great Britain and Mediterranean Europe, the diet of the polecat predominantly consists of European rabbits (*Oryctolagus cuniculus*),

which also serve as a key source of shelter, as polecats have been known to occupy and repurpose their burrows (Birks and Kitchener 1999; Barrientos and de Dios Miranda 2012).

The status of the European polecat population across its range in Europe is fragmented, mostly due to data on polecat being limited due to their low population density, nocturnal behaviour, and lack of distinctive signs in the field (Birks and Kitchener 1999; Baghli et al. 2005; Croose et al. 2018). This makes the species systematically difficult to monitor, and makes studying these individuals through non-invasive, genetic methods, such as sampling from roadkill necessary (Barg et al., 2022; Barrientos et al., 2024). Due to lack of data, population trends for the polecat are difficult to discern. It is currently listed as Least Concern on the IUCN Red List, with several countries listing either that the polecat population was in decline (Baghli & Verhagen, 2003; Croose et al., 2018; Skumatov et al., 2016; Virgós, 2003) or reported insufficient data (Baghli & Verhagen, 2003; Hofmeester et al., 2019).

However, the polecat is currently listed under legal protections in some countries - including Great Britain, due to a perceived or observed reduction in population numbers. Across the continent, European polecat populations are threatened due to secondary exposure to toxicants (Sainsbury et al., 2020; Szapu et al., 2024), persecution through hunting (Sainsbury et al. 2019), hybridisation (Davison et al. 1999; Costa et al. 2013; Szatmári et al. 2021), road casualties (Barg et al., 2022; Barrientos & de Dios Miranda, 2012), fluctuations in prey species populations (Barrientos & Plaza, 2016) and habitat fragmentation (Barrientos et al., 2024).

In Great Britain, the European polecat experienced a severe population decline in the late 1900s, primarily due to intensive hunting by gamekeepers, which led to a significant contraction of its range (Birks and Kitchener 1999; Costa et al. 2013). A small population persisted in a few western English counties, with a major stronghold remaining in central mid Wales (Langley and Yalden 1977; Birks and Kitchener 1999). Following the introduction of protective legislation - specifically the Wildlife and Countryside Act 1981, and a reduction in persecution during the Second World War, polecat numbers began to recover (Costa et al. 2013). This recovery has since extended across much of their former range in Wales, England, and parts of Scotland. However, monitoring efforts have revealed that hybridisation with feral domestic ferrets in England has contributed to this resurgence, indicating that interbreeding has potentially aided its recovery (Davison et al. 1999; Costa et al. 2013; Etherington et al. 2022).

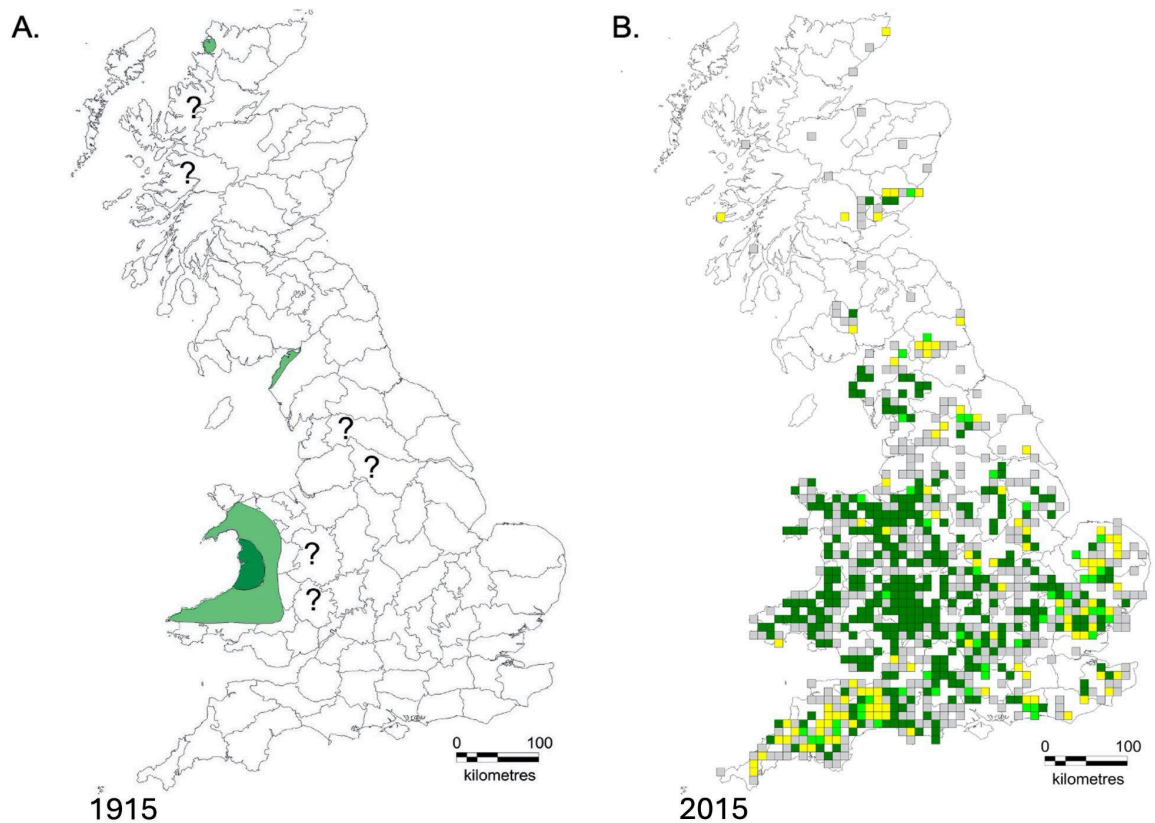


Figure 1.1: (a) The range of the European polecat in Great Britain in 1915. Dark green areas indicate the population stronghold, light green indicates known occurrences, and “?” refers to uncertain occurrences of polecats in those areas (Langley & Yalden, 1977). (b) Records received during the Vincent Wildlife Trust 2014-15 survey of verifiable sightings, based on their phenotype, of true polecats (dark-green), polecat-ferret hybrids (yellow), both true polecats and hybrids (lime green), and unverifiable (grey) (Croose, 2016).

1.7.2 The Black-footed ferret

In contrast to the European polecat, the decline and subsequent recovery efforts of the black-footed ferret (*Mustela nigripes*) have been extensively documented. Over the past four decades, this species has been the focus of intensive conservation and captive management programs aimed at restoring population numbers (Powell et al. 1985; Forrest et al. 1988; Miller et al. 1996; Dobson and Lyles 2000; Wisely et al. 2003; Wisely et al. 2015; Kliver et al. 2023; Santymire et al. 2014).

Native to the Great Plains of North America, the black-footed ferret has become a flagship species for this ecosystem. It is an obligate predator of the prairie dog (*Cynomys* spp.), relying on it not only as its primary food source but also for shelter (Biggins et al., 2024). Prairie dogs are recognised as ecosystem engineers due to their extensive burrowing and grazing behaviour, and black-footed ferrets utilise these burrow systems for both

protection and hunting (Davidson et al., 2012). However, during the 20th century, much of the ecosystem within the Great Plains habitat were subject to land changes due to agriculture and urbanisation (Biggins & Godbey, 2003; Dodds et al., 2004; Wisely et al., 2002). This led to habitat fragmentation and habitat loss of suitable environments for prairie dogs, ultimately leading to a decline in the population within these affected areas (Biggins & Godbey, 2003). The Sylvatic Plague, caused by the bacterium *Yersinia pestis*, also played a major additive role in the destabilisation of the prairie dog population at this time. Introduced to North America in the early 1900s, the sylvatic plague became widespread across much of the Great Plains ecosystems by the 1950s (Davidson et al., 2012, 2022, 2023). This invasive disease decimated populations of both prairie dog and the black-footed ferret due to not having evolved immunity to the disease (Davidson et al., 2022). It is estimated that due to habitat fragmentation and disease, prairie dog populations have declined to 95% of their previous range, which has had immense repercussions on the survival of the black-footed ferret (Forrest et al., 1988).

Subsequently, the black-footed ferret was listed as endangered by the US Fish and Wildlife Service in 1967, with a captive breeding programme being established in 1971. After reproductive complications led to the failure of the breeding programme and outbreaks of canine distemper affecting the last observed wild population in South Dakota, ferrets were pronounced as extinct in the wild by 1976 (Cain et al., 2011; Forrest et al., 1988). A dramatic discovery led to an abrupt change in their extinction status when a previously unknown black-footed ferret population was found near a private ranch in Meeteetse, Wyoming in 1981. This population was monitored from 1981-1985, with the population rising to an estimated 129 individuals in 1984 (Forrest et al., 1988). In 1985, sylvatic plague was detected in nearby prairie dog towns. This quickly led to a second attempt at an assisted breeding programme, with efforts ultimately resulting in trapping the last 24 wild individuals on-site during 1985-1987 due to the decimation of the population during that time (Marinari et al., 2024). Six of the 24 captured ferrets died whilst in captivity due to canine distemper virus (CDV), a common yet highly infectious virus that affects domestic and wild living Canidae and Mustelidae (Deem et al., 2000). Infected ferrets were found to have no circulating antibodies for the virus, highlighting another confounding factor to the survival of the remaining individuals (Williams et al. 1988). Following this period, no remaining free-living individuals were found in the area, leaving only the 11 females and seven males in captivity. Of the 18 remaining individuals, only seven produced offspring, which make up the founder population of the captive breeding programme (Forrest et al. 1988; Miller et al. 1996; Wisely et al. 2002; Cain et al. 2011; Santymire et al. 2019).

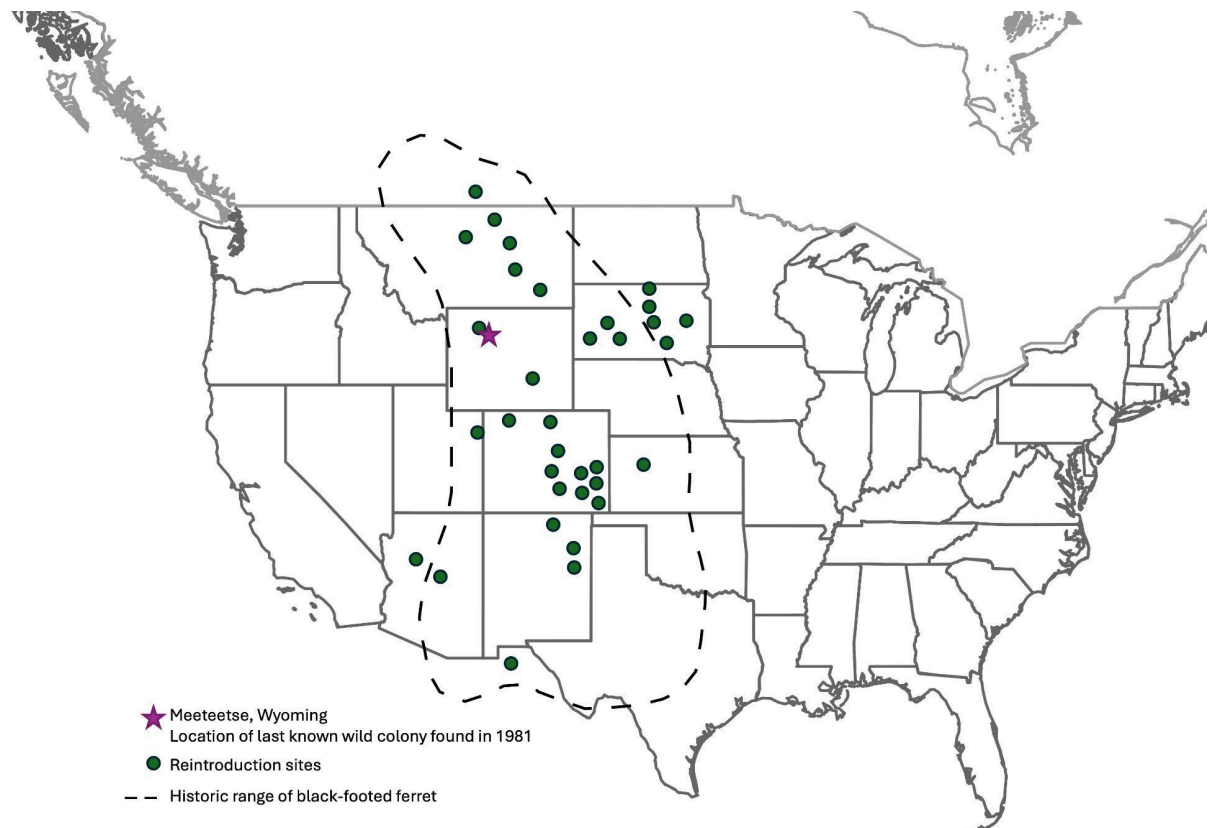


Figure 1.2: Map of North America and southern parts of Canada. Dotted line shows the historic range of the Black-footed ferret. Green circles depict the 34 reintroduction sites that have been utilised during the reintroduction programme from 1991-present day. The purple star depicts the location in Meeteetse, Wyoming where the last wild population was found in 1981, before going extinct in 1986. Redrawn from (Marinari et al., 2024).

From those seven founders, over 11,000 black-footed ferrets have been born which have boosted numbers throughout several national breeding programmes, with over 5,500 of those individuals being released back into the wild at 34 reintroduction sites across 9 states - inclusive of the black-footed ferrets historic range (Marinari et al., 2024; Novak et al., 2024) (Figure 1.2). However, the removal of prairie dogs from this ecosystem has led to a decline in connectivity between black-footed ferret reintroduction sites, leaving populations vulnerable to environmental changes due to limited gene flow (Marinari et al., 2024). Plague is also still determined as a limiting factor in the recovery of both black-footed ferret and prairie dog populations (Biggins et al., 2024). Furthermore, the implications of long term inbreeding on the population were quickly realised at the onset of the captive breeding programme with sperm traits, litter sizes and genetic diversity estimates showing a demonstrable decline over successive years (Wisely et al. 2002; Santymire et al. 2019). Not only that, but environmental factors associated with captivity -

such as reduced nutritional value of diets - have been implicated in a decline in sperm quality demonstrating the effects of environmental-induced inbreeding depression (Santymire et al. 2015; Ali et al. 2024).

Interventions by the way of genetic rescue have attempted to reintroduce genetic variation back into the population through assisted reproductive technologies using cryopreserved sperm (Santymire et al. 2014; Howard et al. 2016), and cloning using biobanked tissue of an unrepresented lineage in the current population (Novak et al., 2024). Further advances in technology have led to intervention methods exploring uses of gene editing - through gene drives and the application of CRISPR - to infer disease resistance to sylvatic plague and canine distemper in the black-footed ferret population (Novak et al., 2018; Tibbetts, 2022).

2.0 Population genomics of the European polecat (*Mustela putorius*) in Britain

2.1 Abstract

The European polecat (*Mustela putorius*) has a widespread distribution spanning mainland Europe, but is declining within this range. In Britain, direct persecution led to a severe decline of the polecat population during the 19th century. Unlike European mainland populations, it is now recovering across much of its former British range. The genomic and conservation implications of such a severe demographic decline, followed by the current recovery, have still to be characterised in the European polecat in Britain. Here, I carry out population-level whole-genome analyses of 65 polecats from Britain (Wales and England) and the European mainland. Our analyses show that Welsh polecats show genetic variability from both English and European polecats, while British polecats as a whole exhibit signs of genetic isolation from mainland European populations. I also reconstructed the demographic history of the Welsh polecat to quantify the magnitude of the population bottleneck experienced by this population. Our analyses confirm the drastic decline of the Welsh polecat's effective population size, with a severe genetic bottleneck around 40 generations ago (1854-1872). I investigated whether whole genome diversity reflected this demographic event and found that Welsh polecats had significantly less genetic diversity than English polecats, but not less than European mainland polecats. Runs of homozygosity and genetic load present in Welsh and English polecat genomes also indicate recent historic inbreeding. Our findings suggest that the increase in the British polecat population size may be attributed to admixture events. Additionally, I demonstrate that the Welsh polecat constitutes a genetically distinct population, which could be crucial for the overall conservation of European polecats by preserving unique genetic diversity.

2.2 Introduction

Anthropogenic impacts are considered some of the most serious threats to the survival of small carnivore populations (Marneweck et al., 2021; Willcox, 2020). This includes active persecution and habitat fragmentation, which reduces connectivity between individuals, limits gene flow and reduces effective population size (N_e) (Croose et al., 2018; Marneweck et al., 2021). The ensuing decline in genetic diversity decreases future resilience to challenges including environmental changes (Frankham et al., 2017). Smaller populations are highly susceptible to inbreeding and genetic drift, limiting the species' ability to purge deleterious mutations due to reduced efficacy of purifying selection at low N_e , coupled with increasing efficacy of drift which increases their probability of fixation. This phenomenon allows for weakly deleterious mutations to segregate at higher frequencies which can negatively affect the population and lead to an increased risk of extinction (Khan et al., 2021; Lande, 1994; Spielman et al., 2004).

Populations that have recently undergone rapid expansion often display similar genetic signatures of low N_e , partly due to allele surfing, where certain gene variants increase in frequency among individuals at the forefront of the expansion (Cisternas-Fuentes & Koski, 2023; Excoffier & Ray, 2008; Graciá et al., 2013; Welles & Dlugosch, 2018). Likewise, the 'founder effect' can leave a distinct genetic imprint by reducing genetic diversity, often resulting in greater genetic differentiation between populations (Jarausch et al., 2021; Szewczyk et al., 2019). These signals are often associated with historical (post-glacial) contemporary (often human-mediated) introductions of species (Sendell-Price et al., 2020) or by range expansions. The latter has been well documented in the recent, rapid range expansion of the grey wolf (*Canis lupus*) across Europe, where recolonisation by expanding populations can lead to altered genetic structure across the landscape (Gustafson et al., 2018; Jarausch et al., 2021; Szewczyk et al., 2019).

Habitat fragmentation or population displacement can also threaten the genetic integrity of species through introgression and hybridisation. As species population boundaries overlap, or through introductions of invasive species, populations may become more at risk of interbreeding with sister species. Hybridisation with closely related species has been shown to have negative consequences, particularly in endangered species (Allendorf et al., 2001). These arise through outbreeding depression which causes a reduction in the fitness of offspring (Hedrick et al., 2019) or genomic swamping whereby the genetics of the original population are overwhelmed by the hybrid, potentially leading to their extinction (Howard-McCombe et al., 2023). Hybridisation can also impact how a species is managed, with uncertainty over the classification and protection of hybrids in the wild (Moroni et al., 2022; Senn et al., 2019).

Despite these issues, adaptive introgression through successful hybridisation between two species, can be particularly beneficial for populations recovering from a bottleneck (Hedrick, 2013). In such cases, where genetic diversity is low, introgression can introduce new variants into the offspring's genome. This phenomenon has been observed in the Alpine ibex (*Capra ibex*) (Brambilla et al., 2024) and Scottish wildcat (*Felis silvestris*) (Howard-McCombe et al., 2023), where adaptive introgression has replenished genetic diversity in previously depleted genomic regions, such as immune-related genes, thereby increasing the populations' adaptive potential. This can be beneficial for recovering populations, yet adds to the complexity of managing hybrid individuals as distinguishing signals of introgression are complex and require well-defined local ancestry estimates (Brambilla et al., 2024; Howard-McCombe et al., 2023).

It is evident that multiple factors influence genomics on a population level and that these factors interact in complex ways. Therefore, it is crucial to monitor and investigate genomic signals in populations of conservation concern to understand how they adapt to and recover from environmental change. This knowledge can inform conservation strategies, such as prioritizing populations for protection (von Takach et al., 2024), monitoring reintroductions (Gajdárová et al., 2023) or translocations (Farquharson et al., 2022), identifying hybrid individuals to inform action plans (Senn et al., 2019) and contributing to ex-situ breeding programmes (Speak et al., 2024). Additionally, capturing the temporal fluctuations of genomic changes in a recovering population undergoing range expansion is essential to inform conservation management strategies to prevent long-term genetic fragmentation and monitor loss (or gain) of adaptive potential (N. E. Thomas et al., 2022).

The European polecat (*Mustela putorius*) is a medium-sized carnivore of the Mustelidae family, which is the largest in the order Carnivora, with ~60 extant species. This family is a great example of adaptive radiation, as species belonging to this family occupy a wide range of niches within different climates and habitats (Koepfli et al., 2008; Schluter, 2000). Despite this, and despite the polecat inhabiting a wide geographical range across mainland Europe (Brzeziński et al., 2010; Mestre et al., 2007; Virgós, 2003), in several of these countries their populations are listed as, or suspected to be, declining (Croose et al., 2018; Maran et al., 2016). Several of the aforementioned threats, such as habitat fragmentation, prey availability, poisoning, and persecution, are known to severely impact polecat populations in this region (Barrientos et al., 2024; Eeraerts et al., 2022; Szapu et al., 2024). Hybridisation has also been known to occur between the European polecat and other closely related mustelid species, such as the Steppe polecat (*Mustela eversmanni*) and the severely endangered European mink (*Mustela lutreola*) (Cabria et al., 2011; Cserkés et al., 2021).

2.2.1 Polecat population in Britain

In Britain, the European polecat has undergone an extensive population reduction and range contraction, followed by a subsequent rapid range and population expansion over the last 70 years (Langley and Yalden 1977; Birks and Kitchener 1999; Birks 2008; Costa et al. 2013; Croose 2016). The population experienced a severe decline in numbers during the late 19th century as the species was persecuted heavily by gamekeepers and regarded as pests which impacted gamebird populations (Birks and Kitchener 1999; Costa et al. 2013). This led to a contraction in the species range, with small, isolated populations in Cumbria and Scotland persisting until the early 20th century and a stronghold in central Wales and the neighbouring English border counties of Herefordshire and Shropshire (Harris et al., 1995; Langley & Yalden, 1977).

The existing stronghold population in Wales started to recover in size after the end of the First World War. This resulted in a period of reduced persecution and coincided with the recovery of a key prey species in the polecats diet (Langley & Yalden, 1977), the rabbit (*Oryctolagus cuniculus*) in its post-myxomatosis period. To aid its protection, the polecat was listed on schedule 6 of the UK Wildlife and Countryside Act in 1981, protecting it from being persecuted and since 2007 has been listed as a priority species on the UK Biodiversity Action Plan (BRIG, 2007). As the population began to expand eastwards across its former range, it came into contact with populations of feral domestic ferrets (*Mustela putorius furo*) (Davison et al. 1999; Costa et al. 2013; Etherington et al. 2022). Ferrets are thought to be derived from the European polecat through domestication that occurred over 2000 years ago (Blandford, 1987; Davison et al., 1999) and have long been used for hunting purposes. In Britain, they are known to have been used predominantly for hunting rabbits, and are thought to have been released into the wild or accidentally introduced. Owing to the short evolutionary distance between the two species, the European polecat and the domestic ferret readily hybridise and this has been well-documented in Britain (Costa et al. 2013; Etherington et al. 2022).

Previous work on the British population of polecats has focussed on assessing genome introgression between European polecats and feral domestic ferrets (Davison et al. 1999; Costa et al. 2013; Etherington et al. 2022) or has aimed to model their demographic history (Costa et al. 2013). These studies identified high degrees of genome introgression in British polecats outside of their stronghold, emphasising the importance of the Welsh polecat population for the conservation of the British European polecat. However, genetic assessments for the Welsh population have either been conducted using microsatellites and mitochondrial genes, which did not provide enough power to confidently assess the size of the genetic bottleneck (Costa et al. 2013), or were limited in the number of Welsh individuals used in their assessment (Etherington et al., 2022).

In this chapter, I aim to characterise the demographic history and contemporary genomic structure of the British European polecat using whole-genome resequencing of individuals from England, Wales, and mainland Europe. I will evaluate the population structure, genetic diversity, and mutational load of the British polecat population following the historical population bottleneck. These results will aim to resolve whether British polecats form a genetically homogeneous group distinct from mainland European populations, and to uncover the degree of introgression with domestic ferrets within the British polecat population. Finally, this chapter also aims to explore how past demographic events and recent population recovery have shaped the genome of British polecats, providing a comparative framework for assessing the conservation implications of population bottlenecks in wild carnivores.

2.3 Methods

2.3.1 Sample Collection

31 British polecat samples were resourced from the Vincent Wildlife Trust 2014-2015 survey (Croose, 2016) from roadkill (Figure 2.1b). Carcasses were collected by members of the public who were requested to freeze the carcass where possible. These were then sent to the Centre for Ecology and Hydrology (CEH) via a pre-paid postage box and the carcasses were frozen upon receipt.

Samples for this study were selected based on geographical location from areas around Wales and from a trajectory across England from west to east to capture any diversity across the predicted range expansion of the polecat in Britain (Figure 2.1c), (Supplementary material S2.1). The home ranges of polecats in Britain are known to vary seasonally, and male polecats tend to disperse further than females (Birks and Kitchener 1999), with males travelling as far as 8 km between dens (Blandford, 1987). To avoid incorporating familial bias into our selection by sampling related individuals, I selected individual carcasses that were collected greater than 10 km apart.

I further increased our sample size through the inclusion of publicly available data consisting of eight domestic ferret genomes, 19 British polecats, and 15 European mainland polecat samples previously published in (Etherington et al., 2022), and accessible under ENA project PRJEB48359.

2.3.2 DNA Extraction

DNA extraction from tissue samples was carried out using a Qiagen™ DNEasy Blood & Tissue kit according to the manufacturer's protocol. Samples were quality checked using a Nanodrop Spectrophotometer and DNA quantity was checked using a Qubit Fluorometer. Samples with a low yield, and with 260/280 and 260/230, absorbance values outside the desired range (260/280 <1.80, 260/230 <2.0) were further cleaned up using Beckman Coulter™ AMPureXP Beads. A total of 31 samples were selected for library preparation and sent for sequencing. Figure 2.1c shows the distribution of the samples sent for sequencing and the location of British polecats already sequenced.

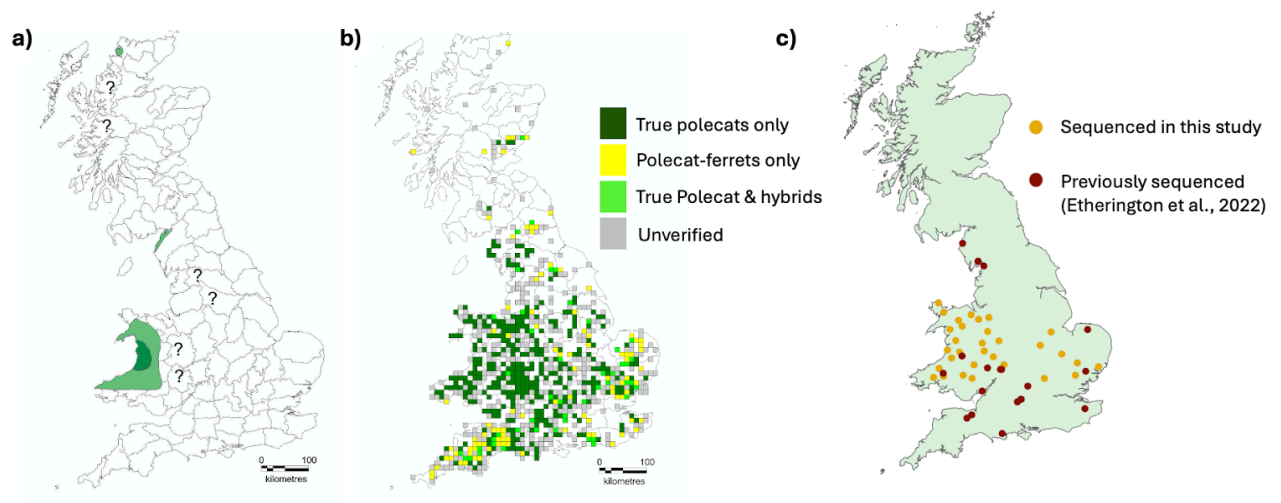


Figure 2.1: (a) The range of the European polecat in Britain in 1915. Dark green areas indicate the population stronghold, light green indicates known occurrences, and “?” refers to uncertain occurrences of polecats in those areas (Langley & Yalden, 1977). (b) Records received during the Vincent Wildlife Trust 2014-15 survey of verifiable sightings, based on their phenotype, of true polecats (dark-green), polecat-ferret hybrids (yellow), both true polecats and hybrids (lime green), and unverifiable (grey) (Croose, 2016). (c) Samples selected to be sequenced in this study (dark yellow) and (red) samples that were previously sequenced by (Etherington et al., 2022). Figures 2.1a and 2.1b were adapted from Croose, 2016.

2.3.2 Library preparation & sequencing

Samples were sequenced by The Earlham Institute Technical Genomics group, with LITE library preparation (Perez-Sepulveda et al., 2021) and sequenced using an Illumina NovaSeq 6000 generating paired-end (2 x 150bp) reads. Samples were run on one lane with a predicted coverage of 10x.

2.3.3 Sequencing QC, Read mapping & SNP calling

Raw read and sequence quality for all British and European samples were first assessed using fastQC version 0.11.9 (Andrews 2010), and any reads below the threshold (Q20) were trimmed using TRIMMOMATIC version 0.39 (Bolger et al., 2014). Raw reads were mapped to a new version of the domestic ferret genome MusPutFur1.0 scaffolded with Bionano data (GCA_920103865.1) (Etherington et al., 2022; Peng et al., 2014) using BWA MEM version 0.7.17 (H. Li & Durbin, 2009). Duplicate reads were then marked using PicardTools version 2.1.1 ("Picard Toolkit," 2019). Single nucleotide polymorphisms (SNPs) were then called using GATK version 4.5.0 (McKenna et al., 2010) following the best practice workflow for SNP discovery (van der Auwera & O'Connor, 2020). Briefly, BaseRecalibrator was run on each sample bam file to detect and correct patterns of systematic errors in the base quality scores. The recalibration metric was then applied to each sample bam file, and HaplotypeCaller was run for each sample before consolidating all of them together using GATK GenomicsDBImport. For downstream analysis, BCFtools version 1.15.1 (Danecek et al., 2021) was used to select only biallelic SNPs with a quality score (QUAL) of at least 20 and a minor allele count of at least three (AC3). A further step for population structure analyses of pruning variants over a linkage disequilibrium (LD) threshold of r^2 0.6 within 50 Kb windows was carried out using the bcftools prune command. Table 2.01 provides a breakdown of the variant dataset for each analyses carried out and filters applied.

2.3.4 Relatedness

assessed relatedness between samples using the KING algorithm in PLINK version 2 (Chang et al., 2015; Purcell et al., 2007). KING kinship coefficients determine first-degree relations (parent-child, full siblings) as values between 0.17-0.35, second-degree relations correspond to 0.08-0.17 and third-degree relations as between 0.04-0.08 (Manichaikul et al., 2010). One pairing between samples S17 and S0012_VWT665 had a kinship coefficient of 0.07, suggesting a third-degree level of relatedness (Supplementary Figure S2.1). Both of these individuals were English polecats but were found in very different geographical locations (~291km apart). I removed S17 from all of the downstream analysis as this sample was found in Waterrow, England, and was therefore outside of the west-to-east trajectory that I was hypothesising was the route of the Welsh polecat expansion.

2.3.5 Population Structure

I first generated a principal component analysis (PCA) using the whole-genome SNP dataset of 65 polecats and 8 domestic ferrets in PLINK version 1.9 (Chang et al., 2015; Purcell et al., 2007) to determine the British population structure. A Spearman correlation test was implemented in R version 4.1.1 (R Core Team, 2024), using values from PC1 and

sample longitude coordinates, to ascertain whether population structure in Britain could be correlated with geographical location. I used ADMIXTURE version 1.3 (D. H. Alexander et al., 2009) to perform ancestry runs on the SNP dataset for K values 1-12, and a cross-validation (CV) error was used to compare all K values. Replicates for each ancestry run were carried out using the custom python wrapper (admixturerwrapper.py). Pairwise F_{st} values between populations were computed using VCFtools version 0.1.16 (Danecek et al., 2011).

To further detect structure within the British population, I tested for isolation by distance between individuals. I did this by first calculating the genetic distance (Hamming distance) between British samples using PLINK version 1.9 (Chang et al., 2015; Purcell et al., 2007). I then used a geographic distance matrix between British samples using the geosphere package in R version 4.1.1 (Hijmans, 2024; R Core Team, 2024), inputting longitudinal and latitudinal coordinates to output distance in km. A mantel test was used to test for isolation by distance between the two matrices.

I used RAXML-ng version 0.9.0 (Stamatakis, 2014) to infer a phylogenetic tree from our whole-genome SNP dataset. The phylogeny was generated using the GTRGAMMA model of evolution with the Lewis bias ascertainment correction and 100 rapid bootstraps.

2.3.6 Genomic Diversity

To assess the genetic diversity between populations, I used VCFtools version 0.1.16 (Danecek et al., 2011) to detect nucleotide diversity in 20 Mb windows across all samples. I tested for significant differences in the genetic diversity estimates between populations by using a Kruskal Wallis chi-squared test followed by a Dunn post-hoc test in R version 4.1.1 (R Core Team, 2024). SNPs that were private to one population, or that were shared across populations, were identified using VCFtools. The observed proportion of heterozygous sites and the inbreeding coefficient (f) for each individual per population were also calculated using VCFtools. I tested for significant differences in the heterozygosity and inbreeding coefficient estimates between populations by using an analysis of variance test (ANOVA) and post-hoc Tukey tests implemented in R version 4.1.1 (R Core Team, 2024) (Supplementary material S2.10).

2.3.7 Detection of historic bottleneck: Runs of homozygosity and identification of deleterious mutations

To detect signatures associated with a genetic bottleneck, I calculated the distribution of runs of homozygosity (ROH) using PLINK version 1.9 (Chang et al., 2015; Purcell et al., 2007). Since calling ROH can be sensitive to low sequencing depth in areas of the genome, which can introduce genotype errors (Ceballos et al. 2018), I only calculated

ROH for individuals per population that had an average read depth of 6.0 or higher. The following recommended parameters were used to call ROH for low coverage samples

```
-plink --geno 0.5 --maf 0.1 --homozyg --homozy-density 50 --homozyg-gap 1000 --homozyg-kb 300 --homozyg-snp 50 --homozyg-window-het 4 --homozyg-window-missing 5 --homozyg-window-snp 50 --homozyg-window-snp 50 --homozyg-window-snp 50 --homozyg-window-threshold 0.05 --aec --double-id
```

(F. C. Ceballos, Joshi, et al., 2018). I tested for a correlation between F_{ROH} and mean read depth of samples using a spearman rank correlation test in R version 4.1.1 (R Core Team, 2024). A Kruskal Wallis chi-squared test and Dunn post-hoc test were performed in R to assess significant differences between F_{ROH} estimates across European, English and Welsh populations (Supplementary material S2.10). ROH were then categorised into three depth classes; 300-500 Kb, 500 Kb-1Mb, and >1Mb.

SnEff version 5.1 (Cingolani et al., 2012) was used to detect the functional impacts of the biallelic SNPs on the genome of the European polecat. SnEff compares a set of variants to a known annotation database and compares the derived alleles within genic regions to predict the impact of that variant. In this case, I used the domestic ferret annotation as the input database. I assessed whether mutational load differed between populations and this was calculated based on homozygous and heterozygous mutational load within missense and loss-of-function (LoF) categories. I tested for significant differences in the mutation categories between populations by using an analysis of variance test (ANOVA) followed by post-hoc Tukey tests in R version 4.1.1 (R Core Team, 2024) (Supplementary material S2.10).

Due to the potential bias that could arise by aligning reads to the domestic ferret reference genome to infer deleterious mutations within the polecat genome landscape, I polarised our variant calls using the Steppe polecat (*Mustela eversmanii*) and weasel (*Mustela nivalis*) as outgroups to infer the ancestral allele at sites called by SnEff version 5.1 (Cingolani et al., 2012) as deleterious.

I selected species from the Mustelidae family with available genomes: European polecat (*Mustela putorius*) (GCA_902207235.1) (Etherington et al. 2020), domestic ferret (*Mustela putorius furo*) (GCA_920103865.1) (Peng et al. 2014; Etherington et al. 2022), Steppe polecat (*Mustela eversmanii*) (GCA_963422785.1), black-footed ferret (*Mustela nigripes*) (GCF_022355385.1) (Kliver et al. 2023), stoat (*Mustela erminea*) (GCA_009829155.1), stone marten (*Martes foina*) (GCA_040938555.1) (Tomarovsky et al. 2025), and Eurasian otter (*Lutra lutra*) (GCA_902655055) (Mead et al. 2020). Additionally, I included three mammalian outgroups: human (*Homo sapiens*) (GCA_000001405.15) (Schneider et al.

2017), domestic cat (*Felis catus*) (GCF_000181335.3) (Buckley et al. 2020), and domestic dog (*Canis lupus familiaris*) (GCF_011100685.1) (Wang et al. 2021) to construct a whole-genome alignment.

I first generated a whole-genome alignment using Progressive Cactus v2.2.0 (Armstrong et al. 2020) without specifying branch lengths. Next, I extracted ancestral repeats from the alignment and inferred a phylogenetic tree under a neutral model using phyloFit v1.6 (part of the PHAST toolkit) (Siepel & Haussler 2004). This tree was then used as a guide for rerunning Progressive Cactus to refine the alignment.

I subsequently extracted variants from four sister species (domestic ferret, European polecat, Steppe polecat, and weasel), using the domestic ferret as the reference, via *halSNPs* v2.2 (Hickey et al. 2013). Finally, I intersected the extracted variants with our called variants to identify alleles derived in the European polecat.

2.3.8 Historic effective population size

Recent historic N_e was estimated using GONE (Santiago et al., 2020). GONE estimates recent N_e by calculating LD between pairs of SNPs and finding the N_e which best fits the observed LD spectrum. GONE assumes that the population being tested is closed, with no recent admixture to allow for unbiased estimates in the calculation of effective population size. Therefore, I ran GONE for Welsh, English, and European populations separately. I restricted GONE runs for the English population to those individuals in the south, central and eastern regions of England. This was to account for any potential gene flow between individuals located within the border counties of Wales and Welsh individuals, as well as lack of gene flow due to the geographic isolation. I further removed an individual (S0024_VWT1661) from the Welsh population run due to apparent recent hybridisation identified by PCA and ADMIXTURE analyses. Low sequencing depth was further accounted for by further removing individuals from the analyses that had a mean whole-genome sequencing depth of less than 5x coverage. For each population, I simulated 2000 generations and calculated 400 bins. GONE takes up to 200 chromosomes as input data, so I removed scaffolds that overlapped with the annotated X chromosome of *Mustela nigripes* and used the largest 100 autosomal scaffolds as input, which constitutes 78% of the whole genome assembly. GONE calculated LD for up to 10,000 SNPs per scaffold (Supplementary material S2.8 and S2.9). Each simulation was run 100 times, with a new subsample of 10,000 SNPs for each run. For each population, I plotted the mean N_e for the 100 runs.

2.4 Results

2.4.1 Whole Genome Resequencing

I generated whole-genome sequence data for 31 European polecats from Wales and England (Supplementary material S2.1, S2.2). Samples had been chosen from the location of the known refugium (Figure 2.1a), across a trajectory of the historic range expansion following the population's recovery from a genetic bottleneck. Data curated for this study were produced from different sequencing experiments and had been generated using different library preparation methods, run on different sequencing technology platforms and varied in sequence length and coverage (Supplementary material S2.1). These were combined with an already published dataset of 34 European polecats from Italy (5), Germany (2), Austria (3), France (3), Spain (2), Wales (3) and England (16), along with 8 domestic ferrets. Sequence reads were aligned to the domestic ferret genome (*Mustela putorius furo*) (GCA_920103865.1) (Peng et al. 2014; Etherington et al. 2022).

2.4.2 Method for correcting batch effects within samples

The first attempt at calling variants was done so using `bcftools -mpileup` version 1.15.1 (Danecek et al., 2021). The total number of called SNPs using `bcftools` before filtering steps was 18,711,787, whereas the number of variants after quality filtering (Q20 and AC3) were 8,057,962, after pruning for LD ($r^2 > 0.6$ within 50 Kb windows) were 2,450,228 SNPs for downstream analysis. For population structure, I further pruned for a minor allele frequency (MAF) of < 0.05 with 1,730,122 SNPs remaining for all samples, 1,663,826 SNPs for British samples. Figure 2.2a and 2.4a demonstrate the principal component analysis (PCA) run using these variant datasets, demonstrating a clear batch effect arising within the data associated with the different sequencing technologies and sequencing chemistries used.

To tackle the impact of batch effect on the real biological signal (Leek et al., 2010) and consistent with previous studies (Leigh et al., 2018), sequence read lengths were trimmed so that all reads were of similar sizes (100bp) using TRIMMOMATIC version 0.39 (Bolger et al., 2014). Reads were then re-aligned to the reference genome using BWA MEM version 0.7.17 (Li and Durbin 2009), and then processed using GATK version 4.5.0 (McKenna et al., 2010) for variant calling. The total number of called SNPs using GATK before filtering steps was 10,294,109. The number of variants after quality filtering (Q20 and AC3) was 6,327,326, after pruning for LD ($r^2 > 0.6$ within 50 Kb windows) 1,600,115 SNPs remain for downstream analysis. For population structure, the SNP dataset was further pruned for a minor allele frequency (MAF) of < 0.05 with 397,604 SNPs remaining for all samples and 392,299 for British samples. Figure 2.2b shows the PCA performed on the corrected dataset, demonstrating reduced batch effects associated with technology

and improved biological signal, as samples sequenced using different technologies but originating from a similar location now cluster together.

Previous work by (Lou & Therkildsen, 2022) demonstrated that downsampling and base recalibration of quality scores on datasets from different sequencing runs could successfully alleviate batch effects. Therefore, I carried out a further alleviation step of downsampling samples to a similar mean (~5x) sequencing depth using samtools. Briefly, I calculated proportions (Supplementary material S2.2) by which to downsample reads within bam files, to reduce the average read coverage to be consistent with other samples. I then used the BaseRecalibrator approach implemented by the GATK pipeline based on their best practice data pre-processing and best practice workflow for SNP discovery (van der Auwera & O'Connor, 2020). Variants were then called per sample, before being consolidated into one VCF for all samples. These were then filtered to include only SNPs passing the quality score threshold, resulting in a total of 10,009,461 SNPs. The number of variants after quality filtering (Q20 and AC3) was 6,115,457, after pruning for LD ($r^2 > 0.6$ within 50 Kb windows) 2,941,545 SNPs remained for downstream analysis (Supplementary material S2.3). For population structure, SNPs further pruned for a minor allele frequency (MAF) of < 0.05 with 595,820 SNPs remaining for all samples.

Figure 2.2 shows the geographic clustering of samples from mainland Europe and Britain, with PC2 showing greater resolution of population structure within samples from Europe. However, for the variants called without batch correction (Figure 2.2a) a clear effect of the sequencing run (technology) is noticeable as British and Welsh data produced on HiSeq and NovaSeq (run1) cluster together and away from the data produced through the second NovaSeq run. In contrast, Figure 2.2b demonstrates a greater resolution in both Welsh and English samples from different runs clustering together. The addition of the domestic ferret samples (Figure 2.3) further resolves the population structure within British samples.

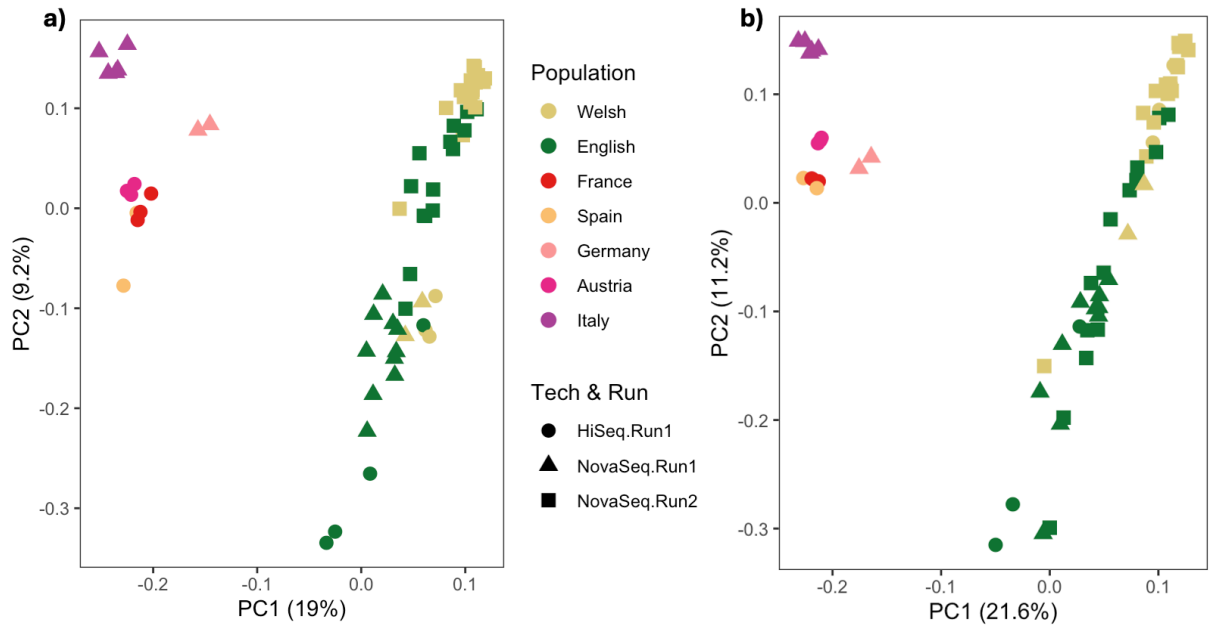


Figure 2.2: Principal component analysis (PCA) of whole genome SNPs for British and European samples. **a)** Without being batch effect aware using bcftools (1,730,122 SNPs) **b)** Variants called using GATK, batch effect aware and accounting for downsampling of higher coverage samples (595,820 SNPs).

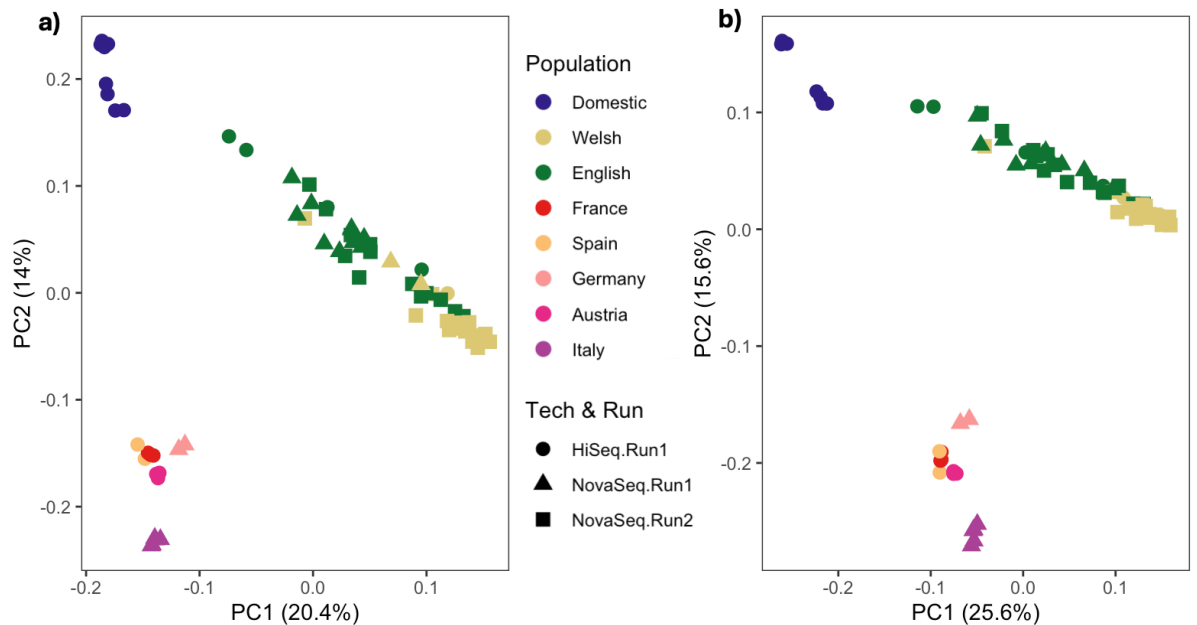


Figure 2.3: Principal component analysis (PCA) of whole genome SNPs for British, Domestic ferret and European samples. **a)** PCA run with SNPs called using batch effect aware GATK, but no downsampling to account for higher coverage samples (397,604 SNPs) **b)** PCA run with the SNP dataset called using GATK, batch effect aware and downsampling of higher coverage samples (595,820 SNPs).

Removing the European samples and rerunning the PCA with only British SNP data clearly shows clustering based on sequencing runs (Figure 2.4a). This has then been resolved using the methods above in Figure 2.4b.

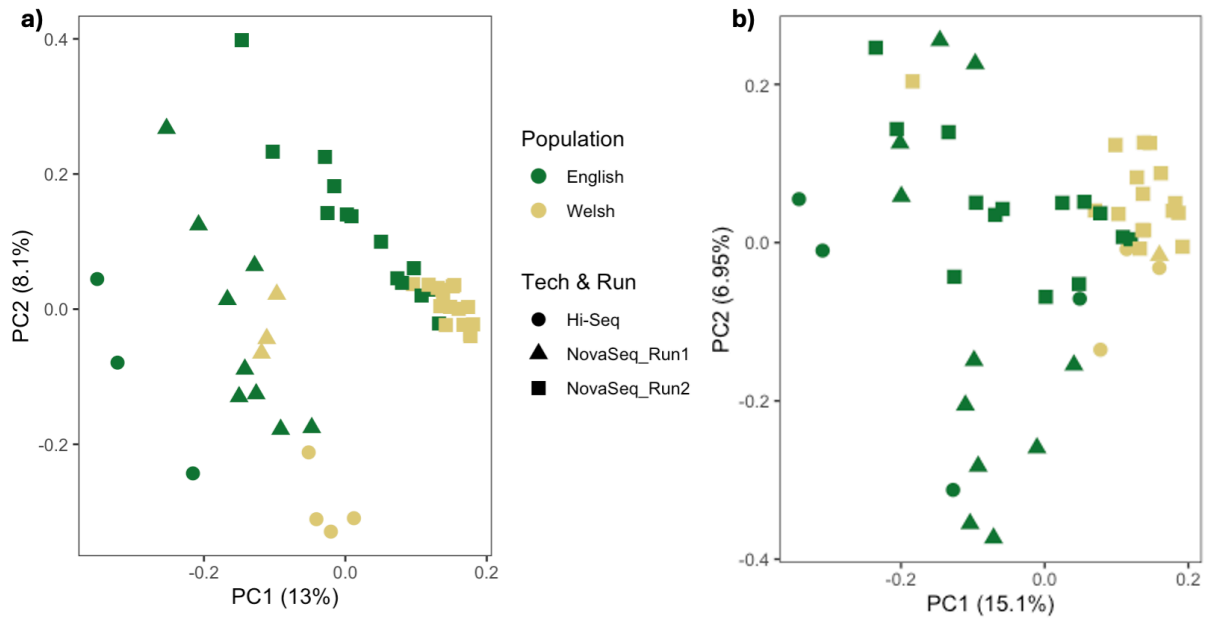


Figure 2.4: Principal component analysis (PCA) of whole genome SNPs for British samples. **a)** PCA run with SNPs called without batch effect awareness using bcftools (1,663,826 SNPs) **b)** PCA run with the SNP dataset called using GATK, batch effect aware and downsampling of high coverage samples (595,820 SNPs).

2.4.3 Population Structure Using Genomic Data

Principal component analysis (PCA) of whole-genome SNPs revealed 3 distinct clusters of samples of polecats from (1) mainland Europe, (2) Britain, and (3) domestic ferrets (Figure 2.5a). Within this, PCs 1 (25.6%) & 2 (15.6%) revealed the English and Welsh samples clustered together and away from mainland European samples. To assess structure within the British population, I repeated the PCA only on the 49 British samples, which revealed a geographical separation of samples from West (Welsh) and East (English) (Figure 2.5b). Longitudinal coordinates strongly correlated with PC1 ($r_s = -0.754$, $p < 0.01$). Outliers to this relationship that clustered with the Eastern individuals are from Northwestern and Southern England, suggesting that Welsh samples were distinct from English samples.

ADMIXTURE analysis identified three genetic clusters ($K=3$), supported by the lowest cross-validation error (Table 2.02). This indicates the existence of three genetically distinct populations across all of the sampled populations; domestic ferret, British (encompassing English and Welsh populations) and European polecats. These results resolve the British population as a homogenous group distinct from the European mainland population,

however this could also be due to the small sample size representing different European populations (Figure 2.5c). Under these parameters, Welsh individuals demonstrated a lower degree of admixture with the domestic ferret than English individuals (Figure 2.5c) (Supplementary material S2.4), with the exception of one individual, which is likely a polecat-ferret hybrid (S0024_VWT1661). These results suggest repeated hybridisation and continuing introgression between English polecats and domestic ferrets.

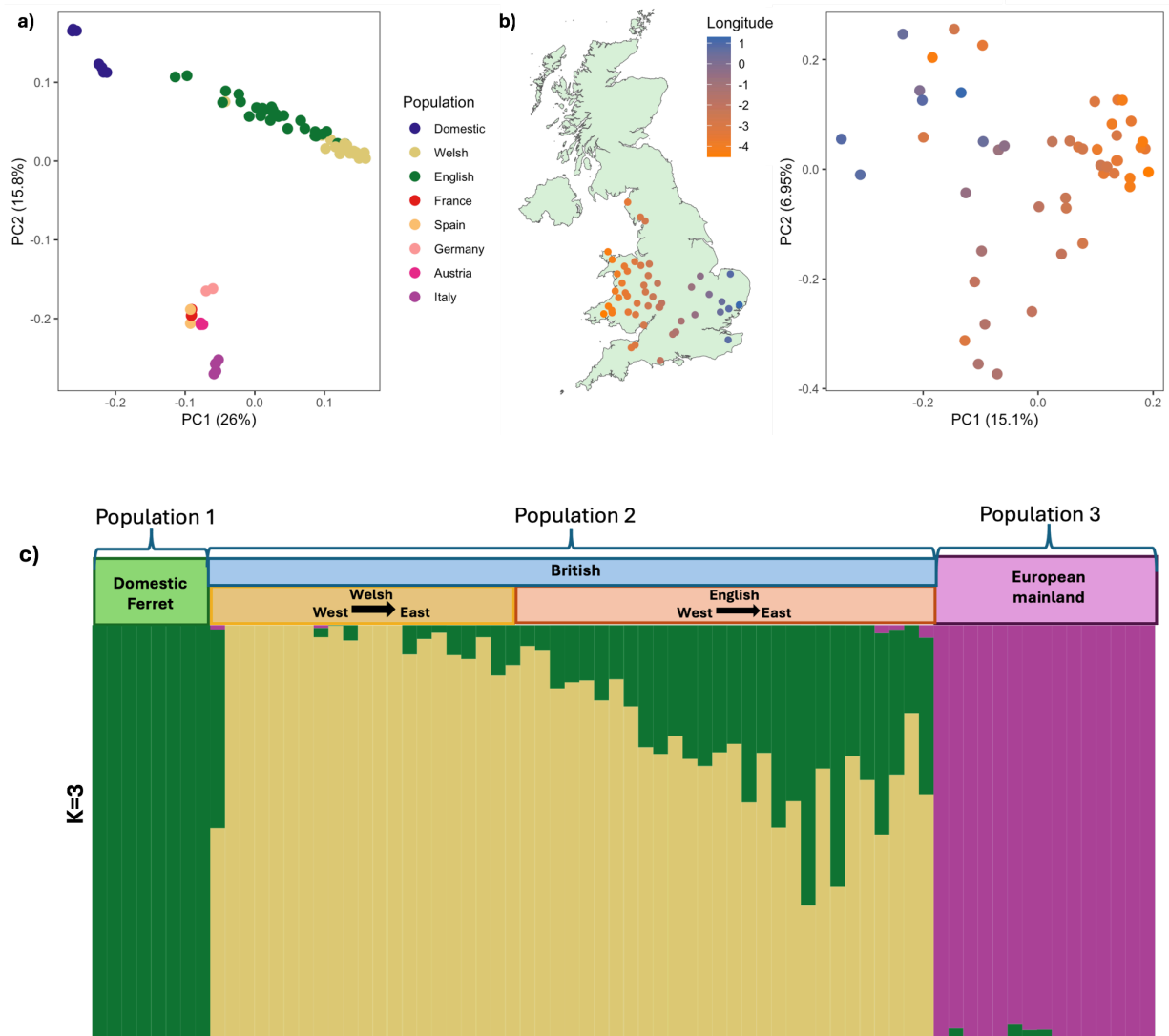


Figure 2.5. Population structure of the European polecat in Britain and across mainland Europe. (a) PCA based on whole genome SNPs for all European polecats and domestic ferrets. (b) Map of Britain demonstrating longitude coordinates of samples and PCA based on whole genome SNPs for all British polecats, which is supported by a significant correlation between PC1 and longitude coordinates ($r_s = -0.75$, $p < 0.01$). (c) Admixture results for whole-genome SNPs between the four groups (domestic ferret, English, Welsh and mainland European polecats), with the best fitting value of $K = 3$, supported by a cross validation error of 0.68 (Supplementary table 2.02).

Population structure was further confirmed across European polecat samples from Europe and Britain, as well as domestic ferret samples through a maximum-likelihood phylogeny (Figure 2.6).

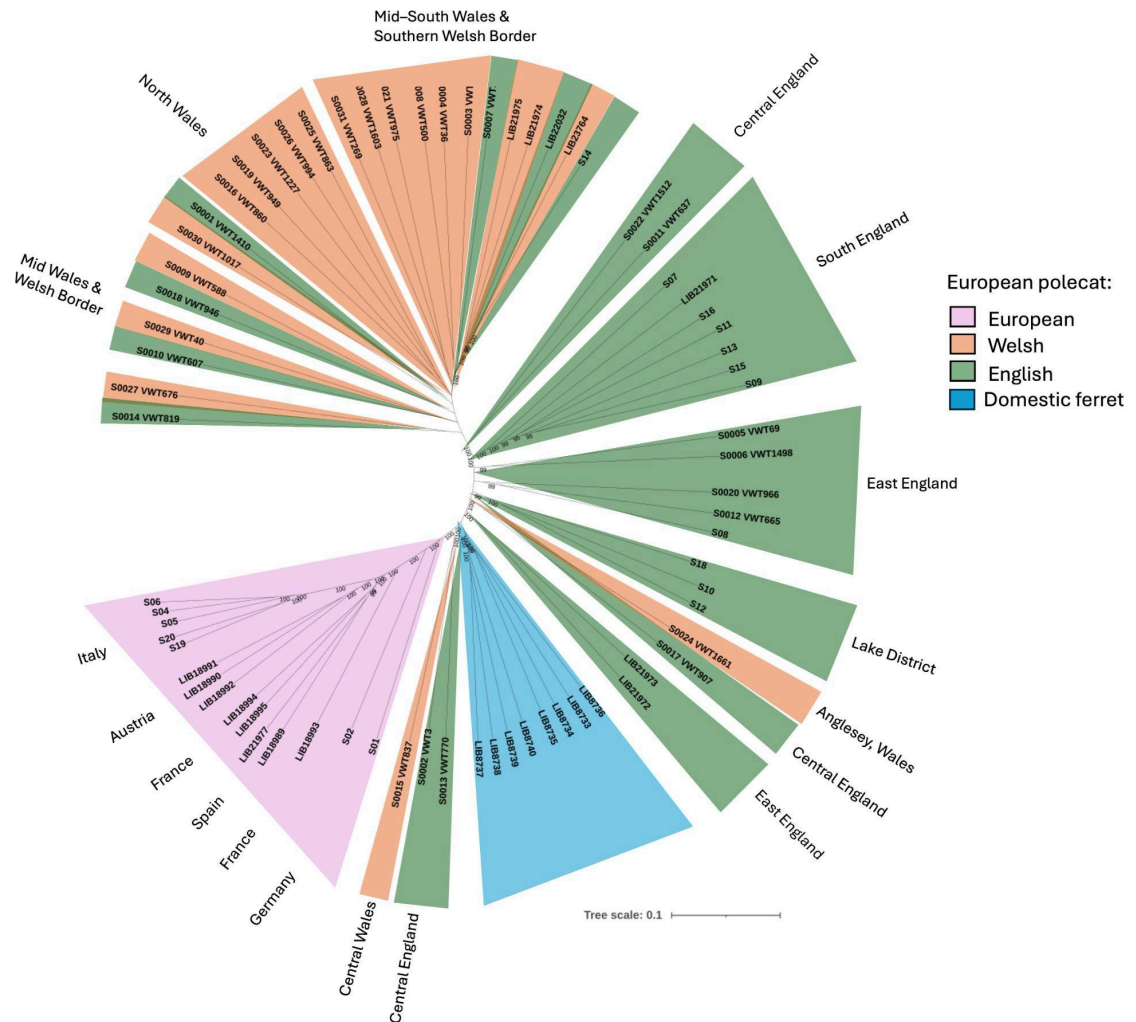


Figure 2.6: Maximum likelihood phylogeny with 100 bootstrap replicates using genome-wide SNPs (2,941,545 SNPs) from 72 samples, consisting of domestic ferrets and European polecats from Europe and Britain. The number at the nodes refer to bootstrap values of 95 and above. Branch lengths are in expected substitutions per site.

From the phylogeny, there is a distinct separation of European and British samples, as well as between Welsh individuals (besides S0015 and S0024) and Domestic ferret samples. Individuals from the neighboring border counties in England (S0014, S0010, S0018, S001, S007, S14 and LIB22032) cluster with the majority of Welsh individuals, further demonstrating the separation of samples from Wales and the West of England, with those from the east and south of England. The phylogeny appears to resolve some population structure based on geographic location, due to clustering of individuals sampled from similar geographic regions. However, to further understand this relationship,

I performed a mantel test to examine the correlation between genetic distance and geographic distance across individuals from Wales and England (Figure 2.7). The analysis revealed a significant positive correlation (Mantel's $r = 0.33$, $p = <0.01$), suggesting that genetic differentiation increases with geographic distance, demonstrating that individuals in Britain are isolated by distance.

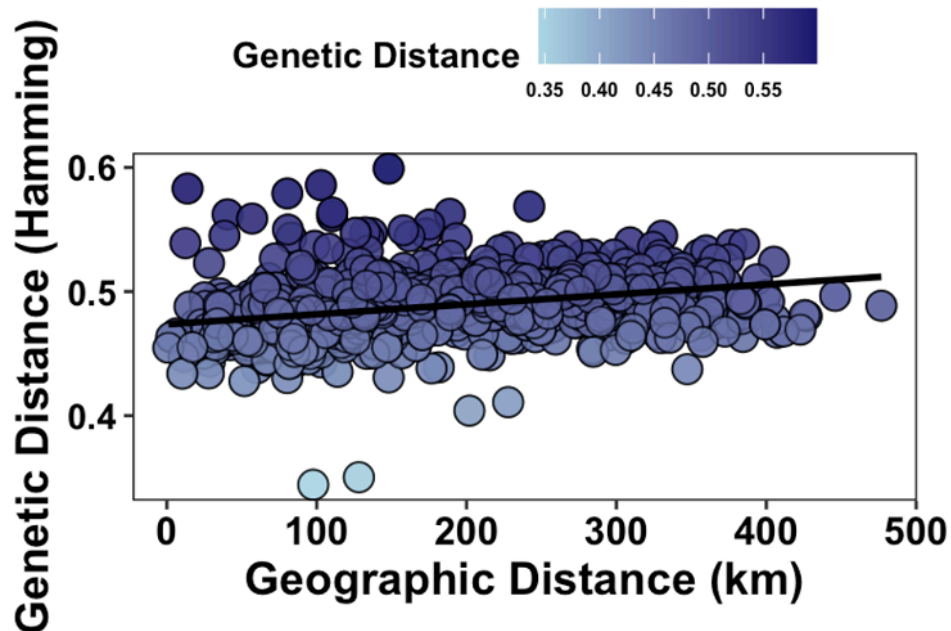


Figure 2.7: Isolation by distance of British samples plotted using a pairwise Hamming genetic distance and geographic distance matrix. Results revealed a positive correlation between genetic distance and geographic distance ($r = 0.33$, $p = <0.01$).

Pairwise F_{st} values were computed to assess genetic divergence between populations. Results showed a comparison between English and Welsh individuals having an F_{st} value of 0.036, compared to 0.106 between English and European and 0.145 between Welsh and European (Supplementary figure S2.2).

2.4.6 Genomic Diversity in Britain in Comparison to Mainland Europe

As the British populations have undergone a severe population contraction prior to their current recovery, I investigated the impact of these demographic events on the genomic diversity of the different populations. European populations demonstrated a higher number of private SNPs (2,290,307) in comparison to English (838,471) and Welsh (578,869) (Figure 2.8a). This could be attributed to population structure and also demonstrates the genetic divergence between British and European mainland polecats. Welsh individuals showed a lower number of private SNPs in comparison to English individuals, although still demonstrating that unique genomic diversity resides within the refugia population.

Nucleotide diversity estimates (Tajima's π) (Supplementary material S5) revealed a significant difference amongst populations ($H = 7.14$, $df = 2$, $p = 0.02$), with the English population having higher diversity when compared to the Welsh population ($p < 0.01$) and European population ($p = 0.01$). However, there was no significant difference between the Welsh and European populations (Figure 2.8b) (Supplementary table 2.03).

Inbreeding coefficients ($F_{2,62} = 18.06$, $p < 0.01$) and proportion of heterozygous sites ($F_{2,62} = 18.65$, $p < 0.01$) significantly differed across the three populations (Figure 2.8b & 2.8d). Post-hoc comparisons found that the European population had a significantly higher inbreeding coefficient ($p < 0.01$) and significantly lower proportion of heterozygous sites ($p < 0.01$) in comparison to individuals in the English sub-population. However, I did not account for population structure in the grouping of individuals due to the total sample size. French and Austrian populations had higher proportions of heterozygous sites and a lower inbreeding coefficient than German, Italian, and Spanish populations. There was no significant difference between the Welsh and European populations for either inbreeding coefficient or the proportion of heterozygous site estimates. English individuals did show a significantly higher proportion of heterozygous sites ($p < 0.01$) and significantly lower inbreeding coefficient ($p < 0.01$) than those in the Welsh sub-population (Supplementary table 2.03).

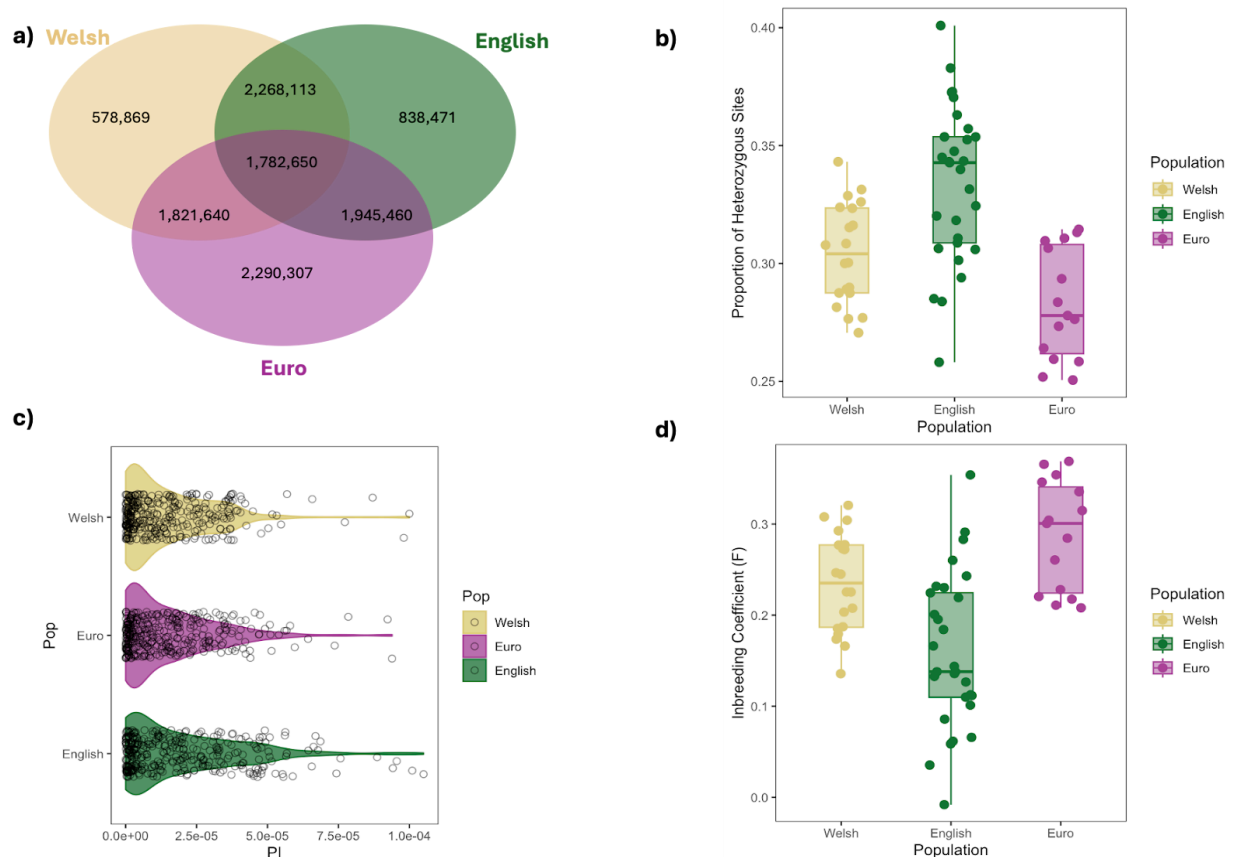


Figure 2.8. Genetic diversity estimates and comparisons across British and European populations (a) Shared and private SNPs between populations. (b) Proportion of heterozygous sites calculated using VCFtools. (c) Population nucleotide diversity values (π), calculated in 20 Mb windows across the genome. and (d) Population-level inbreeding coefficient values.

2.4.7 Runs of Homozygosity & Deleterious Mutations

To assess whether the British polecat population had undergone prolonged periods of inbreeding due to the population bottleneck, I investigated the presence of ROH within polecat genomes. Low coverage whole-genome sequencing (4-6x), as utilised in this study, can result in higher genotyping errors and therefore inaccuracy in calling runs of homozygosity across the genome (Ceballos et al. 2018). I therefore filtered for ROH segments that would help to tolerate genotyping errors within windows - `--homozyg-window-het 4`, which is recommended for low coverage sequencing data (Ceballos et al. 2018). However, I still assessed our data to ascertain whether there was any correlation between sequencing depth and called ROH. I calculated the realised inbreeding coefficient (F_{ROH}), the proportion of the genome that fell within ROH. When including all samples within our analyses, I found a significant negative correlation between the mean depth and F_{ROH} ($\rho = -0.4$, $S = 58152$, $p < 0.01$) (Figure 2.9a, Supplementary material S2.6). To alleviate this correlation, I removed all individuals from our analyses that had a mean depth < 6 . A spearman's rank-order correlation test was used to assess the filtered individuals and their calculated F_{ROH} and individual mean depth and found no significant correlation ($\rho = -0.07$, $S = 17472$, $p = 0.6$) (Figure 2.9b).

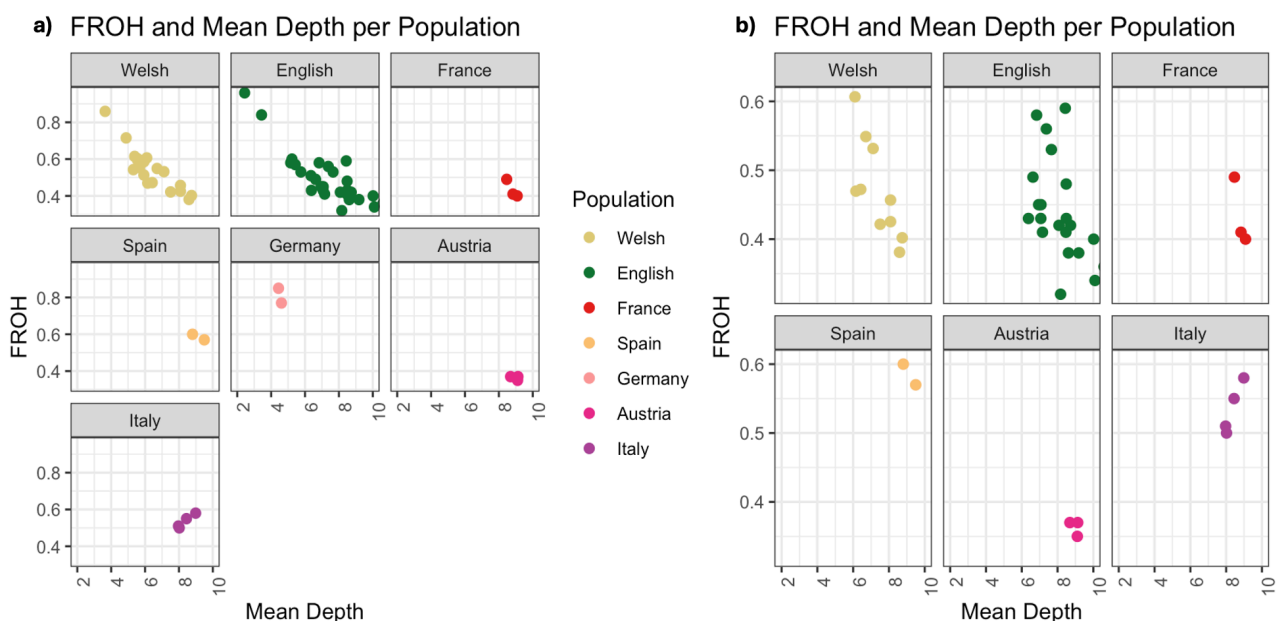


Figure 2.9. The realised inbreeding coefficient (F_{ROH}) in correlation with mean read depth per individual (a) All individuals across British and European populations (64 individuals) (b) All individuals across British and European populations with a mean read depth of 6x or higher (47 individuals: Wales (11), England (23), France (3), Spain (2), Austria (3), Italy (5)).

The realised inbreeding coefficient (F_{ROH}) for higher coverage individuals ranged from 0.33 - 0.76, with results showing that European polecats had a higher mean F_{ROH} than British polecats. Post-hoc comparisons showed that F_{ROH} values were significantly different between European and English individuals ($Z = -5.28$, $p = <0.01$) European and Welsh ($Z = 2.96$, $p = <0.01$) individuals, but not between English and Welsh ($Z = -1.54$, $p = 0.06$) (Figure 2.10a) (Supplementary table 2.03).

ROH lengths were separated into different classes to visualise short (300-500 Kb), medium (500-1 Mb), and long (>1 Mb) ROH across the genomes of individuals belonging to the British (England and Wales) and European populations (Figure 2.10b). Longer ROH are indicative of recent inbreeding events, with shorter ROH being indicative of historic events. Individuals from Wales were seen to have the lowest number of ROH segments in all length classes. Individuals from English sub-populations were seen to have a greater proportion of their genome within longer ROH segments (>1mb) ($58.3\% \pm 1.50$ SE) in comparison to the European ($53.8\% \pm 2.14$ SE) and Welsh ($55.0\% \pm 1.90$ SE) individuals (Figure 2.10b).

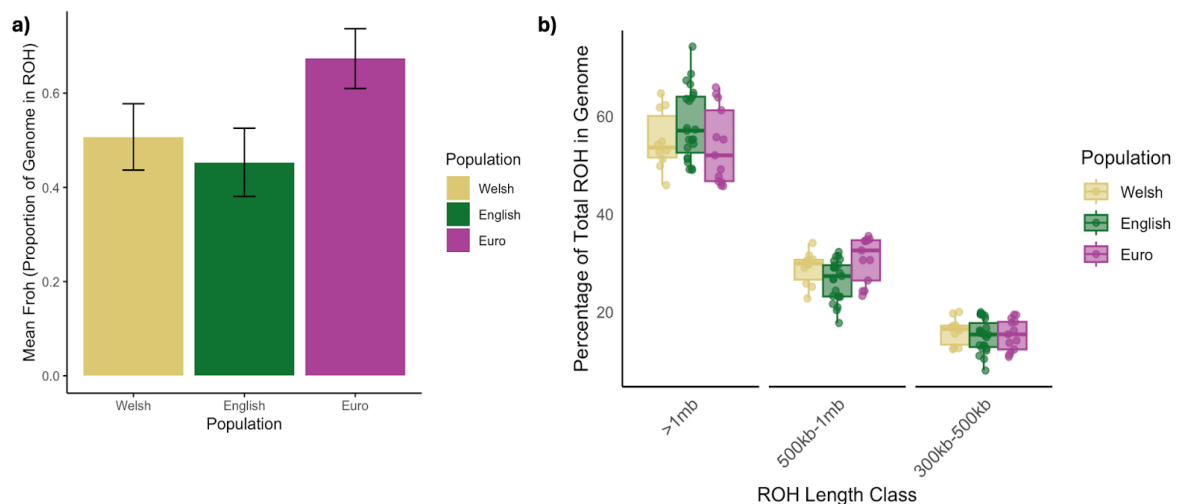


Figure 2.10. Population estimates of higher coverage individuals (mean depth >6x) and (a) the calculated proportion of the genome in ROH (F_{ROH}) and (b) the distribution of ROH across different length categories. Data points represent the percentage of an individual's total ROH that falls within each length category.

I investigated whether the mutation landscape differed among individuals belonging to European, Welsh, and English populations (Supplementary material S2.7). Mutation load was assessed using the polarised variant dataset, across two functional categories as reported by SnpEff; missense and loss of function (LoF). The homozygous and heterozygous load per individual, per category was calculated, with homozygous mutations in both categories expected to be the most deleterious.

Homozygous LoF mutation load differed significantly between populations ($F_{2,60} = 16.81$, $p < 0.01$), with both Welsh and English individuals showing a significantly higher homozygous load within LoF mutations in comparison to the European populations ($p < 0.01$) (Figure 2.11a). There was also a significant difference amongst populations in LoF masked (heterozygous) load ($F_{2,60} = 5.58$, $p < 0.01$), with individuals from English sub-populations having a significantly higher masked load than Welsh individuals ($p < 0.01$) (Figure 2.11b) (Supplementary table 2.03).

Missense homozygous mutations significantly differed between populations ($F_{2,62} = 31.39$, $p < 0.01$). Post-hoc comparisons revealed that Welsh individuals had significantly more missense homozygous mutations than English individuals ($p = < 0.01$) (Figure 2.11c). Masked, heterozygous missense load also significantly differed among populations, ($F_{2,62} = 14.77$, $p < 0.01$), with English individuals showing a higher masked load than both Welsh ($p < 0.01$) and European individuals ($p < 0.01$) (Figure 2.11d) (Supplementary table 2.03).

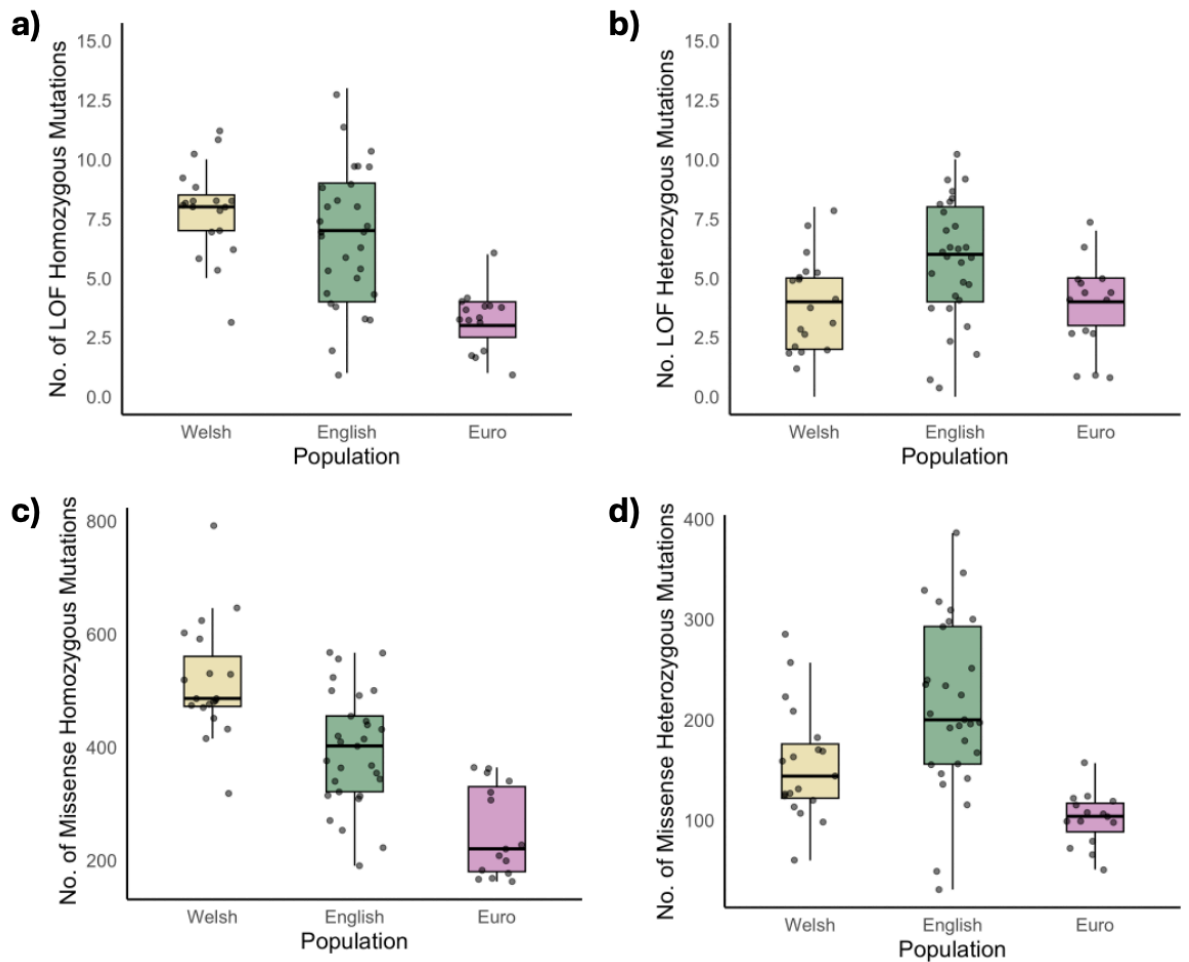


Figure 2.11: Mutation load landscape across populations within different functional categories. Loss of function (LoF) (a) homozygous mutational load per population and (b) heterozygous mutational load and missense (c) homozygous mutational load and (d) heterozygous mutational load.

2.4.8 Historic effective population size (N_e)

GONE was used to estimate the recent historic effective population size for the British population of European polecats. Estimates using GONE correlate with a demographic decline in population size occurring around 30-40 generations ago - which using an estimate for generation time of the polecat of 4 years (Di Marco et al., 2013), corresponds to the mid-late 19th century (1854-1872) (Figure 2.12). Recovery of the polecat population in Wales based on GONE estimates fits with previous estimates of 10-12 generations ago (1966-1974) (Supplementary material S2.8).

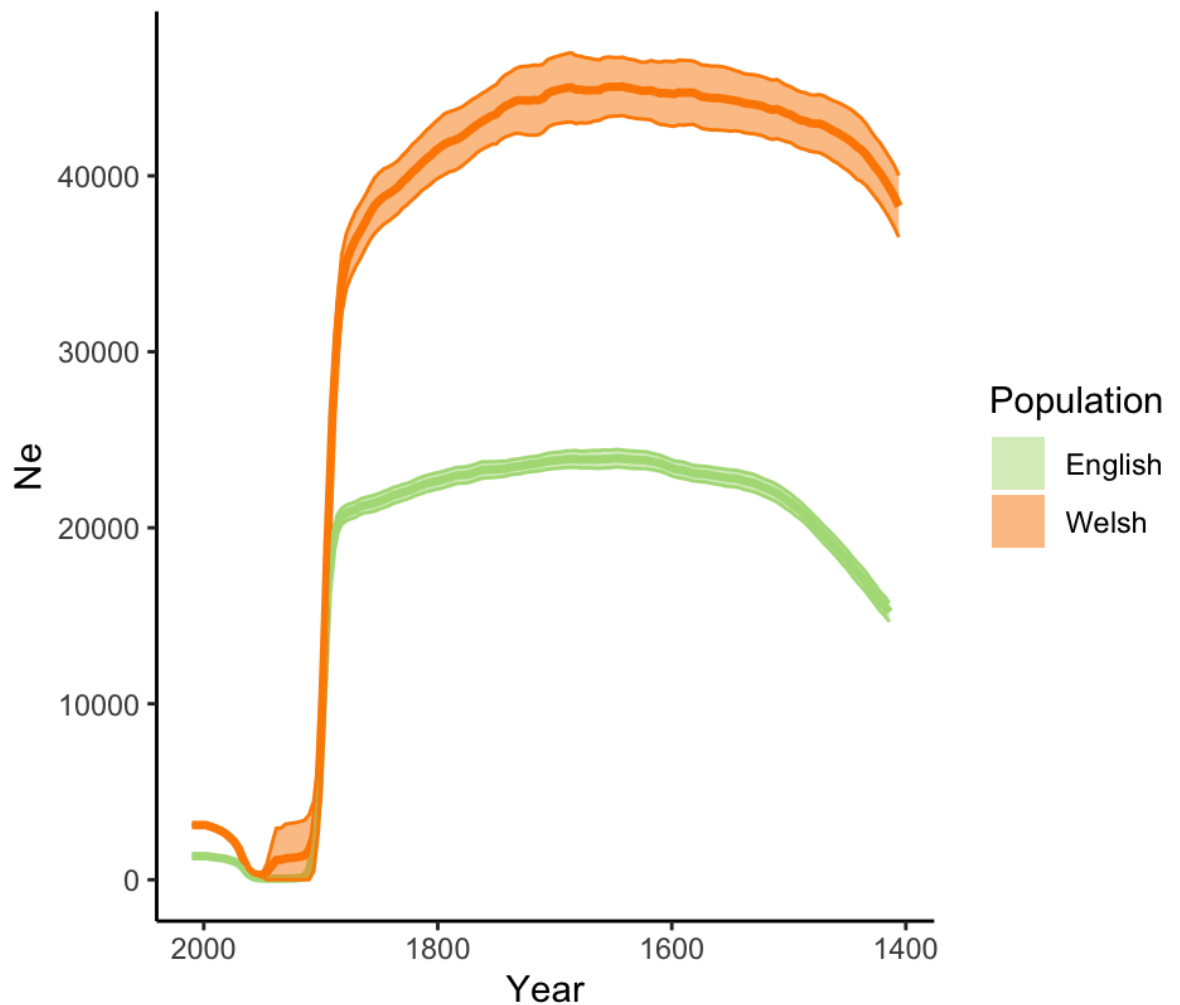


Figure 2.12: Estimated mean effective population size (solid line) with 95% confidence intervals (shaded area) for English (green) and Welsh (orange) populations across 200 generations (generation time = 4 years) based on whole genome SNP data computed using GONE.

2.5 Discussion

Here, I provide the first genomic evidence of a bottleneck having occurred within the British population of polecats. Our analyses indicate that the population drastically decreased in effective population size from the late 1800s and this population crash persisted for ca. 100 years. Our analyses also indicate that the population began recovering from 1960-70 onwards. This coincides with a prolonged period of relaxed persecution from gamekeepers, the recovery of lagomorph populations following myxomatosis, the banning of gin traps, and the introduction of protective legislation. Our results demonstrate that the British polecat population is distinct from European mainland populations, and is recovering from the genetic bottleneck. Conservation efforts should focus on increased protection measures and monitoring of the Welsh sub-population, as it represents unique genetic integrity in comparison to admixed, English sub-populations.

Our study also highlights the importance of genomic data in being able to detect signatures within populations to fully detect and monitor the success of their recovery.

2.5.1 Comparison of British to Mainland European populations

Our analyses indicate that the British polecat population is genetically divergent from European mainland populations. This is highlighted in pairwise F_{st} estimates per population where a higher F_{st} value (closer to 1.0) suggests a population has low levels of breeding and therefore is more genetically distinct from populations compared to lower F_{st} values. Our results indicate that English and Welsh populations have a low fixation index (0.03) and are therefore likely to share genetic variation. F_{st} values between English and European (0.106) and European and Welsh population groups (0.145) are slightly higher, suggesting a greater degree of genetic distinction between mainland European and British polecats.

Divergence between British and European populations is also evident from our ADMIXTURE analyses which also corroborates the genetic divergence seen between Welsh and European polecats (Etherington et al., 2022). British and European polecats are thought to have been separated for ca. 8ky (Montgomery et al., 2014), since the flooding of the Doggerland bridge after the Last Glacial Maximum. Several widespread mammal populations that are native to Britain demonstrate genetic distinction due to isolation over time from mainland European populations. For example, Irish and British populations of red fox (*Vulpes vulpes*) demonstrate unique genetic diversity that distinguishes them from mainland European populations (McDevitt et al., 2022). Likewise, native British populations of roe deer (*Capreolus capreolus*) demonstrate genetic differentiation from mainland European populations, most likely due to colonisation and isolation events (de Jong et al., 2020). Populations of British red deer (*Cervus elaphus*) also demonstrate genetic distinction from mainland European populations, but show a slightly lower genetic distance to Iberian populations, potentially suggesting retention of genetic diversity within populations from similar migration waves during postglacial colonisation (Carranza et al., 2024). Therefore, the genetic distinction I observe between Welsh and European mainland polecats likely results from their recent isolation, mirroring the pattern of genetic divergence seen across other British populations of wide-ranging mammals (Carranza et al., 2024; McDevitt et al., 2022; Montgomery et al., 2014).

2.5.2 Population Structure

Our analyses indicate population structuring based on geographic location within polecats in Britain (Figure 2.2b). Ancestry estimates also demonstrate a higher incidence of introgression in polecats that are found at the furthest edges of the polecat range from the Welsh, refugium population (Figure 2.2c). I also observe a similar relationship in our

maximum likelihood phylogeny (Figure 2.6), whereby English individuals that are found at the more eastern and southern edge of the polecat's range are clustered more closely to domestic ferrets than to the Welsh individuals. In agreement with (M. Costa et al., 2013), I find the least introgression of domestic ferret genes within individuals found in Wales and English border counties. This observed pattern fits the narrative that as the polecat population expanded eastwards, individuals hybridised with feral domestic ferrets within those regions (Birks and Kitchener 1999; Davison et al. 1999; Birks 2008; Etherington et al. 2022). However, the presence of introgressed polecats within these regions may also have been exacerbated by covert reintroductions of individuals from hybrid stocks (Birks and Kitchener 1999).

2.5.3 Demographic Analyses

The British polecat bottleneck was thought to have occurred between the late 1800s to the early 1900s, with populations showing recovery since the 1970s (Langley & Yalden, 1977). The most recent historic genetic bottleneck was evident in our results using GONE, with a sharp decrease in N_e occurring in the Welsh population around 30-40 generations (120-160 years ago), from the sampling time of 2014-2015. This coincides with the known history of persecution of polecats from gamekeepers, as with other populations of terrestrial carnivorans in Britain, which subsided for a period during and following World War I, and again after legal protections came into place (Langley & Yalden, 1977; Sainsbury et al., 2019). Alongside legal protections, it is thought that increasing rabbit populations, which had declined due to myxomatosis and began recovering from the mid-20th century, are also closely linked with the recovery of the polecat (Langley & Yalden, 1977; Sainsbury et al., 2019; Sumption & Flowerdew, 1985).

Historic N_e estimates were lower for the English population than the Welsh population (Figure 2.12, Supplementary material S2.9). This could be explained both by the rapid expansion of individuals from Wales to England and by the hybridisation and subsequent introgression of those individuals with populations of feral domestic ferrets. Our outputs also highlight that modelling demographic histories can be impacted by hybridisation and population structure (Novo et al., 2023). Previous work (Costa et al. 2013), using microsatellites in British polecats could not detect any evidence of a genetic bottleneck. This could be due to the lower resolution from the smaller representation of genetic markers used, but could also be due to the admixture from domestic ferret populations as suggested by the authors. This pattern was observed from the results as English individuals had greater levels of introgression in comparison to the Welsh individuals and therefore demonstrates not only a lower historic effective population size but also a lower present day effective population size when just modelled using the English sub-population. These results also show that population structure can obscure the recent

historic effective population size when trying to model the European population's demographic histories. Modelling these as a whole population is difficult due to the different environmental pressures that each population faces. GONE authors advise not to model demographic histories with less than 5 individuals (Novo et al., 2023), of which only the Italian population was viable from our European samples. However, our model reconstruction demonstrated the unreliability of this with low power and possible population structure (Supplementary figure S2.3).

Census data in 2018 estimated that the polecat population in Britain was currently 88,300 (Mathews et al., 2018). Data estimates from this study corroborate survey data from the Vincent Wildlife Trust that the Welsh sub-population is expanding, with more 'true polecat' sightings in Welsh counties and border counties such as Herefordshire and Shropshire (Croose, 2016), which are areas that were thought to contribute to the historic stronghold of the polecat (Langley & Yalden, 1977), and these areas also having the individuals with the least admixture (Figure 2.2c).

2.5.4 Genomic signatures associated with the population bottleneck

I hypothesised that these analyses would demonstrate genomic signatures within the population associated with the prolonged genetic bottleneck period. The results demonstrate the presence of potentially harmful deleterious mutations within both Welsh and English individuals. It also shows the presence of heterozygous, potentially masked, deleterious variants being more prevalent in admixed English individuals, and a higher unmasked load of missense variants in the Welsh sub-population. This is as expected, as periods of inbreeding in Welsh populations would lead to a higher presence of homozygous deleterious variants within the population. English individuals appear to have higher genetic diversity, due to the introgression of domestic ferret genes. This is also confirmed in the higher nucleotide diversity across the genomes of English individuals. However, no significant difference amongst populations was found in mutations that were categorised as homozygous LoF, suggesting these mutations may have been purged from Welsh individuals due to their expansion and purifying selection acting upon them. Further investigation into historic genomic signatures through the addition of historical samples would provide an insight into temporal changes of mutational load over time, which would assist in resolving potential signatures of purifying selection within the Welsh population, which would be consistent with other studies that have observed this (Dussex et al., 2021; Grossen et al., 2020; van der Valk et al., 2019).

Individuals from Wales and bordering counties had an overall higher estimation of F_{ROH} in comparison to English individuals (Figure 2.10a). This could be due to the long-term, low N_e within the Welsh refugium. Welsh individuals also demonstrated a higher mean of

shorter ROH within the genome in comparison to the English individuals. This is indicative of more ancestral inbreeding events, and is a strong indicator that the population underwent a bottleneck further back in time (Saremi et al., 2019). Longer ROH lengths can indicate recent inbreeding events, as recombination has not yet acted on breaking apart these segments (Ceballos et al. 2018). Individuals from eastern and southern English counties were shown to have a higher mean number of long ROH in their genome than individuals from Wales and the bordering English counties. This could be associated with the introgression with feral domestic ferrets that have been subjected to years of selective, human-mediated breeding. However, it may also be indicative of the historic bottleneck, followed by the expansion of some individuals to low areas of genetic connectivity and therefore increased occurrence of inbreeding events (Colpitts et al., 2022; Manunza et al., 2016).

European individuals overall showed a higher mean F_{ROH} than both English and Welsh individuals, as well as a higher mean proportion of medium length ROH. However, across the length classes and F_{ROH} values there was a varied ROH landscape per population (Figure S2.3) Figure S2.3 demonstrates this, where it can be observed that individuals from Austria, Spain and France have higher mean F_{ROH} than those sampled from Italy and Germany. This could be due to population-level pressures persisting in their environment at the time (Croose et al., 2018) and may indicate the necessity for similar genomic studies in polecat populations in regions of Europe where the populations are declining. However, this could also be due to the small sample sizes used to represent populations from these regions. Increased sampling within these regions would allow for a better understanding of their population history, and a fairer comparison to the British polecat population.

2.5.5 Use of low coverage sequencing to inform conservation strategies

The rapid advancement and decreasing cost of sequencing technologies have enabled researchers to study non-model organisms (Hendricks et al., 2018). However, low-coverage sequencing can introduce challenges, particularly in estimating genetic diversity and effective population size which are key metrics for conservation. In this study, I observed a correlation between sequencing depth and the proportion of homozygous calls across the genome. Kardos and Waples (Kardos & Waples, 2024) stated that at sequencing depths of 5, 4 and 3x lead to the probability of missing one of the two alleles within sequencing reads at heterozygous locus is 0.06, 0.12 and 0.25 respectively. This demonstrates the possibility that reduced site coverage has to increase genotyping errors, leading to an overestimation of homozygosity due to insufficient power to accurately call both genotypes (Duntsch et al., 2021; O'Leary et al., 2018). To mitigate this within this study, I excluded individuals with an average sequencing depth below 6x for calling runs of

homozygosity. While this improved data reliability, it also reduced our sample size. A larger sample size is critical for accurately estimating effective population size, as it better captures the population's genetic structure when using high-quality loci.

In order to fully resolve these issues, future work would be designed to ensure individuals were sequenced at a higher depth. (Kardos & Waples, 2024) recommend that loci should be confidently called at 15-20x in order to accurately estimate heterozygosity and therefore inbreeding estimates for a population. However, the cost of sequencing at this depth (~\$270 per individual) may constrain the sample size feasible within the budget for non-model studies (Watowich et al., 2023). Therefore, taking into account the metrics important to report on for conservation management, future studies may benefit from using low coverage sequencing (0.5 - 5x) and imputation of genotypes for larger sample sizes. (Zhang et al. 2021; Watowich et al. 2023) demonstrated the power that imputation can have at alleviating genotyping error from low coverage sequencing, however highlight that a high quality, diverse reference panel is needed in order to accurately infer genotypes (Wragg et al., 2024). Future efforts to monitor recovering populations should carefully account for the challenges of low-coverage sequencing highlighted in this study and the above cited literature. While trade-offs between sequencing depth and sample size may still be necessary in experimental design, it is crucial to consider how these choices impact the information available for informing conservation strategies.

2.5.6 Conservation Implications

Here, I provide the first evidence of the severity and extent of a genetic bottleneck experienced by the Welsh polecat population from the late 1800s to the mid-1950s. The results demonstrate that the population started to expand when persecution pressures were lessened and after protective legislation came into action, thus supporting legal protection as an effective conservation management tool.

I present a unique scenario of population recovery in a once heavily persecuted mammal in Britain. The recovery can largely be attributed to protective legislation and the resurgence of rabbit populations following the myxomatosis outbreak, as well as to the introgression with domestic ferrets, which has facilitated the eastern range expansion of the population. Genomic introgression with a domesticated species is not a typical conservation management tool but has been witnessed to assist in the recovery of other species, such as the American bison (*Bison bison*) (Stroupe et al., 2022), and Iberian ibex (*Capra pyrenaica*) (Münger et al., 2024). However, the decline of the wildcat (*Felis silvestris*) population in Scotland has been severely hampered by hybridisation and subsequent introgression with the domestic cat (Howard-McCombe et al., 2023). It could be that due to the relatively recent nature of the hybridisation of the European polecat and

domestic ferret in Britain, the harmful consequences are yet to surface. However, (Costa et al. 2013) demonstrated that the lack of F1 hybrids observed within the British population suggests that these hybridisation events are rare. Feral ferret populations are hard to distinguish in mainland Britain but are known to thrive on surrounding islands where there is a lack of, or reduction in, competition from polecats and other predatory mammals (Kitchener & Birks, 2008). This suggests that continued hybridisation events in wild settings could be extremely rare, but would still pose a threat to the genetic integrity of the recovering population.

Our analyses also demonstrate that individuals from the Welsh refugium population hold unique genetic diversity that could be important for the resilience of the population as a whole. This raises important conservation considerations for reintroducing and augmenting polecats from Wales to other parts of the country. Additionally, conservation concerns are underscored by the unmasked missense mutational load persisting in the genome as a result of the prolonged low effective population size (Robinson et al. 2023). The monitoring of this elusive species is hard, due to its low population density and nocturnal habits. Monitoring through non-invasive methods, such as the Vincent Wildlife Trust survey (Birks 2008; Croose 2016), where sightings and suspected roadkill are reported, has proven successful and has helped to document the polecat's recovery within Wales and England. However, these efforts are not matched in some parts of mainland Europe and therefore its status in some countries needs to be clarified. In one study that aimed to document the status of the polecat across its known range, of the 34 countries included, 20 reported that the polecat population was declining (Croose et al., 2018).

Only two countries reported the polecat population as increasing; Estonia and Britain. Therefore, conservation priorities should be to continue monitoring the British population within its known ranges to protect its genetic legacy and ensure the population's continued expansion. Given the reported declines in other known polecat ranges, conserving the British population is crucial for supporting the global population. Unfortunately, polecats appear to not only be threatened by hybridisation with feral domestic ferrets but road traffic collisions have been recorded as negatively impacting the dispersal of polecats during peak mobility periods (Barg et al., 2022). The focus should be to avoid habitat fragmentation and the frequency of polecats having to disperse across primary roads. One implementation that could be considered are wildlife crossings, where studies have demonstrated the benefit and utility of wildlife crossing structures in urbanised areas (Mata et al., 2008; Mysłajek et al., 2020). These structures have provided the means for species to overcome habitat fragmentation and allowed gene flow between populations that have been isolated due to urban development (Warnock-Juteau et al., 2022). Many studies monitoring these structures have also highlighted the frequency of use by polecats and

other mustelid species, providing further evidence for their use to prevent fragmentation within the Welsh polecat population (Mata et al., 2008; van Vuurde & van der Grift, 2005).

It should be noted that the genetic recovery of a population can often lag behind its demographic recovery. In the case of the British polecat population, genetic structuring is observed due to admixture across parts of England. Another British mustelid, the Eurasian Otter (*Lutra lutra*), has subsequently undergone a recent historic genetic bottleneck and is recovering across its previous ranges (du Plessis et al., 2023). However, barriers, such as landscape changes, have potentially contributed to the fragmentation and low genetic connectivity between populations. The authors recommend that continued monitoring be complemented with continued analysis of genetic data, to ensure that populations in Wales do not become genetically isolated due to landscape change and continue to recover both demographically and genetically (Thomas et al. 2022).

This study examined the recovery of polecats across much of Britain, however it did not include data on the population in Scotland. According to survey records, it was thought that by 1915, polecat populations were eradicated within Scotland. Covert re-introductions of polecat were thought to have occurred across some areas in southern England, Cumbria and Argyll (Birks and Kitchener 1999; Davison et al. 1999), with re-discovery of polecats in Scotland in 2004 potentially being attributed to further reintroductions, or demonstrating the low population density and elusiveness of the species in this area (Solow et al., 2006). Previous work has highlighted the more 'ferret-like' ancestry of polecats found within regions of Scotland (Costa et al. 2013), with feral ferret populations being found on offshore Scottish islands. The 2014-15 survey conducted by the Vincent Wildlife Trust recorded only 29 sightings of polecats in Scotland, representing just 2% of the total records from Britain (Croose, 2016). However, some of these sightings occurred in a newly identified distribution area in Dumfries, which may indicate a range expansion of the Cumbrian populations. Additional sampling and genomic analyses would provide further insights into the population structure and conservation status of the remaining population within Scotland.

3.0 Characterisation of the accumulation of deleterious mutations within protein-coding genes associated with severe population bottleneck in the black-footed ferret (*Mustela nigripes*).

3.1 Abstract

The black-footed ferret (*Mustela nigripes*) underwent a severe bottleneck during the last century as a result of habitat fragmentation, loss of prey species, and disease. This led to an estimated 30-50% loss of genetic diversity within the *M. nigripes* population. Subsequent generations suffered from inbreeding, with effects such as low fecundity, poor sperm motility and higher disease susceptibility. To limit the impact of inbreeding effects in future offspring, it is important to understand the genome-wide impact of low effective population size and reduced genetic diversity. To investigate this, I use population-level whole-genome SNPs to identify the landscape of deleterious mutations segregating within the genomes of captive, reintroduced and historic (1970-1980s) individuals, investigating population structure, genetic diversity and functional impact of mutation. Results showed that the reintroduced population had a greater level of genetic diversity, reflected also in the increased degree of heterozygous sites within mutational categories in comparison to both captive and wild population groups. I also investigate genetic load using comparative methods, and identified sites within coding genes, that are fixed within the past (wild) and current (captive and reintroduced) population. These fixed, deleterious mutations are found within genes associated with fertility, immune response and development, which can be attributed to the severe population bottleneck experienced by this conservation priority species. This work highlights the genomic consequences of population bottlenecks and demonstrates how genomic approaches can provide valuable insights into the genomic health of species recovering from demographic decline.

3.2 Introduction

The accelerating decline in global biodiversity is now recognised as a sixth mass extinction event (Ceballos et al. 2020), representing an unprecedented ecological crisis. Key drivers of biodiversity loss, including habitat destruction, climate change, direct persecution, and overexploitation of natural resources, have rendered animal populations across all biomes increasingly susceptible to decline (Keck et al., 2025). According to the most recent *Living Planet Report*, 73% of the 5,495 vertebrate populations currently monitored by the IUCN Red List have exhibited declines over the past five decades (*Living Planet Report 2024: LPR*, 2024; McRae et al., 2024). Declining populations are often fragmented and geographically isolated. Such isolation increases their susceptibility to genomic erosion, which is characterised by the loss of genetic diversity and the accumulation of deleterious genetic variants. This reduction in genetic diversity limits the adaptive potential and contributes to the reduced mean fitness of the population. At the individual level, genetic load refers to the fitness impact and overall burden of all mutations present in the genome in either a homozygous or heterozygous state. Mutation load represents the component of genetic load that arises specifically from newly acquired deleterious mutations that directly reduce fitness (Henn et al. 2015). Whereas at the population level, genetic load is reflected in the frequency of these deleterious mutations, both of which contribute to a higher risk of extinction (Dussex et al., 2023; Pinto et al., 2024; van Oosterhout et al., 2022). As populations become isolated in this way, gene flow between subpopulations is restricted, exacerbating genetic drift and increasing the risk of inbreeding, further aggravating the genetic burden.

Small populations are particularly vulnerable to the accumulation of weakly deleterious mutations. In large, genetically diverse populations, such mutations are typically maintained at low frequencies due to their recessive nature and the effective action of purifying selection, meaning they pose minimal immediate threat to individual fitness. However, during population decline, as effective population size is reduced the influence of genetic drift becomes more pronounced. Under these conditions, weakly deleterious mutations may segregate at elevated frequencies and are more likely to become fixed within the genome (Dwivedi et al., 2023; Peers et al., 2025; Smeds & Ellegren, 2023; Yoshida et al., 2020). Furthermore, the increased risk of inbreeding can reduce genome-wide heterozygosity, resulting in the potential 'un-masking' of these mutations. The reduced efficiency of purifying selection at removing these harmful mutations facilitates their persistence within the population and enables the eventual fixation of such alleles, accelerating genomic deterioration (García-Dorado 2012; Charlesworth 2018; Robinson et al. 2023). Several studies have shown that endangered mammal populations experiencing prolonged low effective population sizes accumulate deleterious mutations through genetic drift, which can lead to increased levels of phenotypic abnormality

(Grossen et al., 2020; Peers et al., 2025; Smeds & Ellegren, 2023; Yoshida et al., 2020). Combined, increased genetic load and reduced effective population size can promote gene pseudogenisation, the loss of coding potential within previously functional genes. This process occurs when deleterious mutations accumulate and purifying selection becomes less effective in small populations (Kumar et al. 2023; Peers et al. 2025). While individual weakly deleterious mutations may have subtle effects on phenotype and fitness, their accumulation in small populations can collectively pose significant challenges to population viability. Consequently, examining the genomic landscape of endangered species can provide important insights into how deleterious variation and gene loss may influence long-term population persistence and should be considered in conservation strategies (Humble et al. 2023; Kumar et al. 2023).

Conservation interventions, both in situ and ex situ, can halt and reverse population declines. Such measures have been instrumental in the recovery of several species, including the Californian condor (*Gymnogyps californianus*), which fell to just 22 individuals in 1982 and has since rebounded to over 500 individuals across captivity and the wild (Cohn 1999; Snyder and Snyder 2000; Robinson et al. 2021); the kākāpō (*Strigops habroptilus*), which declined to 51 individuals in 1995 and now numbers over 200 (Clout & Merton, 1998; Foster et al., 2021; Robertson, 2006); and the Przewalski's horse (*Equus ferus przewalskii*), which went extinct in the wild in 1969 but has recovered from just 12 founders to nearly 2,000 wild-roaming individuals (Ryder and Wedemeyer 1982; Der Sarkissian et al. 2015; Turghan et al. 2022). Captive management programmes have strived to act as a safeguard against extinction, with efforts focussed on ensuring managed populations are representative of those in the wild, with the view that they become important sources for reintroductions or genetic rescue. However, utilisation of these programmes should be carefully assessed, as there is high variability across species in the impacts arising from captive breeding (Farquharson et al., 2021). The IUCN's Species Survival Commission established guidelines that identify when ex-situ management strategies may be beneficial and highlight the necessity for sourcing genetically diverse individuals for insurance populations and to release large numbers of animals over a long period of time (McGowan et al., 2017). These methods would likely ensure constant gene flow amongst populations and keep any potentially deleterious mutations segregating within the population masked (Frankham, 2015; Humble et al., 2023; Ralls et al., 2020). While there are increasing numbers of genetically informed captive-breeding management and reintroduction strategies (Humble et al., 2023; Ogden et al., 2020), limited conservation resources and the challenges of managing chronically small populations often hinder their effectiveness. Therefore, founders for insurance populations are frequently drawn from already declining populations, which can undermine recovery efforts (Klimova et al., 2022; Robertson, 2006; Swinnerton et al., 2004;

Willoughby et al., 2017). Furthermore, the process of population rescue with a limited number of available individuals will lead to a further bottleneck event, and inbreeding depression (Charlesworth and Charlesworth 1987; Hedrick and Kalinowski 2000; Boakes et al. 2007).

The black-footed ferret (*Mustela nigripes*) has become a flagship species for conservation biology and genetic rescue (Williams et al. 1988; Santymire et al. 2014; Livieri et al. 2022). Once widespread across much of North America, populations of black-footed ferret have suffered greatly due to habitat fragmentation, loss of prey species, and disease, resulting in a severe genetic bottleneck, leading to an estimated 30-50% loss of genetic diversity within the species (Wisely et al., 2002). From the mid-1970s through to the 1980s, black-footed ferrets were listed as extinct in the wild, with no known sightings being recorded within this time. However, a previously undetected population was discovered in 1981, and due to the threat of sylvatic plague, conservationists decided to instigate a human-mediated management plan through captive breeding (Forrest et al., 1988). Of the 18 individuals alive at the time, only seven proved to be viable breeders, significantly limiting the gene pool for the newly established captive population. In response, forward-thinking conservationists chose to biobank two of the 18 non-breeding individuals, cryopreserving their somatic cell lines to safeguard their genetic ancestry. SB2 was a male who died from Canine Distemper, while SB10, a female, would not breed in captivity but was believed to have had a son, SB14. SB14 was also captured but, like his mother, was not a successful breeder.

After the removal of the 18 remaining individuals from the wild, the black-footed ferret was declared extinct in the wild in 1987 (Santymire et al. 2014). However, collaborative captive breeding programmes across the US (Jachowski and Lockhart 2009) led to numerous captive-born kits, providing a significant boost to the black-footed ferret insurance population, as of 2024 11,000 black-footed ferrets have been born in captivity with over 4,000 individuals being released to 34 rewilding sites (Novak et al. 2024; Santymire and Graves).

Despite extensive efforts to preserve genetic diversity in the black-footed ferret captive breeding population, recent studbook pedigree analyses estimate the mean kinship of individuals to range from 0.1293 to 0.1496, slightly higher than the relatedness expected between first cousins (0.125) (Novak et al., 2024). The breeding programme has observed elevated inbreeding coefficients, estimated from both microsatellite and genome-wide data, and revealed a decline in reproductive performance in both sexes, including reduced sperm concentration and motility in males, as well as decreased fecundity in females. These effects have been accompanied by a measurable loss of heterozygosity, likely

indicative of inbreeding depression (Santymire et al. 2019). Comparable outcomes have been observed in other species, with inbreeding-associated reductions in testicular volume and sperm motility in the Florida panther (*Puma concolor coryi*) (Barone et al., 1994) , and impaired spermiogenesis in African lions (*Panthera leo*) (Munson et al., 1996). While Santymire et al. (2019) noted a potential link between reduced heterozygosity and diminished fecundity in black-footed ferrets, they emphasised that the underlying biological mechanisms remain poorly understood.

Previous genetic monitoring of the black-footed ferret population has assessed declines in genetic diversity and associated functional consequences using genome-wide markers such as microsatellites and mitochondrial DNA sequences (Cain et al., 2011; Wisely et al., 2002, 2003, 2008). Although beneficial to monitoring the population, these methods lack the genome-wide resolution necessary to reliably identify deleterious mutations or assess patterns of purifying selection (Hohenlohe et al., 2021; Supple & Shapiro, 2018). These techniques capture only a small fraction of the genome, often biased toward neutral or non-coding regions, and therefore provide limited insight into the functional consequences of genetic variation. In contrast, whole-genome sequencing enables the identification of putatively deleterious variants across coding and regulatory regions, offering a more comprehensive view of genomic erosion and its impact on population viability (Bosse & van Loon, 2022). These analyses are highly dependent on the availability and accessibility of genomic resources, which can become essential tools in species recovery, long-term genomic monitoring, and genetic rescue efforts (Brandies et al., 2019; Kliver et al., 2023; Theissinger et al., 2023; Totikov et al., 2021). For black-footed ferrets, the availability of a reference genome has been instrumental in advancing the application of biotechnology in the conservation genomics of this species, enabling innovations such as somatic cell nuclear transfer and gene editing to be applied (Novak et al., 2024; Sandler et al., 2021; Wisely et al., 2015).

This chapter aims to use population-scale whole-genome sequencing to characterise how generations of severe inbreeding and captive management have shaped the genomic architecture of the black-footed ferret. Here, I will conduct a genome-wide investigation to characterise the genetic consequences of an extreme demographic bottleneck and subsequent captive management in the black-footed ferret. Using whole-genome sequencing data from historical, captive, and reintroduced individuals, I aim to quantify patterns of genetic diversity, relatedness, runs of homozygosity, and assess population structure across contemporary populations. I aim to evaluate further the distribution and frequency of deleterious variants using functional annotation and comparative evolutionary analyses to test whether the severe demographic bottleneck reduced the efficacy of purifying selection and allowed deleterious mutations to accumulate or drift to fixation

within the population. I hypothesise that these findings will demonstrate whether captive breeding has prevented further loss of diversity and helped maintain genetic variation across reintroduced populations, and how this may impact the long-term viability of the species.

3.3 Methods

3.3.1 Data curation

I utilised publicly available genomes for eleven mammalian species. From there I selected species from within the family Mustelidae that had annotated genomes: European polecat (*Mustela putorius*) (GCA_902207235.1) (Etherington et al., 2020), domestic ferret (*Mustela putorius furo*) (GCA_920103865.1) (Etherington et al., 2022; Peng et al., 2014), Steppe polecat (*Mustela eversmanii*) (GCA_963422785.1), black-footed ferret (*Mustela nigripes*) (GCF_022355385.1) (Kliver et al., 2023), stoat (*Mustela erminea*) (GCA_009829155.1), weasel (*Mustela nivalis*) (Miranda et al., 2021), stone marten (*Martes foina*) (GCA_040938555.1) (Tomarovsky et al., 2025), and Eurasian otter (*Lutra lutra*) (GCA_902655055) (Mead et al., 2020). Additionally, three mammalian outgroup species were included: human (*Homo sapiens*) (GCA_000001405.15) (Schneider et al., 2017), domestic cat (*Felis catus*) (GCF_000181335.3) (Buckley et al., 2020), and domestic dog (*Canis lupus familiaris*) (GCF_011100685.1) (Wang et al. 2021).

Additionally, I received genome-wide SNP data from 54 black-footed ferrets generated from whole-genome resequencing of six individuals originating from the last wild populations (four from Wyoming and two from South Dakota), 26 captive individuals, and 22 reintroduced individuals (Supplementary material S3.1). Variant data were generated by mapping high-coverage (Supplementary table 3.01) raw reads against the black-footed ferret reference genome (GCA_022355385.1) (Kliver et al. 2023) using the BWA software version 0.7.1 and the BWA-mem algorithm (Li and Durbin 2009). The variant calling process was performed using the Genomic Analysis Toolkit version 4.2.5 (GATK) pipeline (McKenna et al. 2010). Individual samples were merged into a single file using GATK CombineGVCF and joint genotype calling was performed using GATK GenotypeGVCF. The following hard filters were applied to single nucleotide polymorphisms (SNPs), to ensure the accuracy of the variant calling results: "QD < 2.0", "QUAL < 30.0", "QUAL30", "SOR > 3.0", "FS > 60.0", "MQ < 40.0", "MQRankSum < -12.5", "ReadPosRankSum < -8.0" (Figueiro et al., *in preparation*). For downstream analysis, BCFtools filter v1.15.1 (Danecek et al., 2021) was used to extract biallelic SNPs with a quality score (QUAL) of ≥ 20 and a minor allele count of $\geq$ three (AC3). The resulting SNP set was then pruned to a linkage disequilibrium (LD) threshold of r^2 0.6 in 50 Kb windows using bcftools prune for some population genomics analysis. The variant

datasets used for analyses throughout this chapter and filters applied are detailed in Supplementary table 3.01.

3.3.2 Relatedness

The relatedness between samples was assessed using the KING algorithm in PLINK v 2.0 (Chang et al., 2015; Purcell et al., 2007). KING coefficients assign related as a value between 0.17-0.35 would suggest a first-degree relationship (parent-child, full siblings), 0.08-0.17 suggests second-degree and third-degree relationships are suggested between 0.04-0.08 (Manichaikul et al. 2010).

3.3.3 Population structure

To determine the population structure within the historic modern-day black footed ferret population, I generated a principal component analysis (PCA) using the filtered and LD-pruned whole-genome SNP dataset of 54 black-footed ferrets in PLINK v1.9 (Chang et al., 2015; Purcell et al., 2007). ADMIXTURE v1.3 (Alexander et al. 2009) was used to perform ancestry runs on the SNP dataset for K values 1-12, and used CV error to compare all K values. Each ancestry run was replicated ten times using the custom python wrapper (admixturewrapper.py).

3.3.4 Genetic Diversity

To assess genetic diversity in modern-day and historic samples, VCFtools v0.1.16 (Danecek et al., 2011) was used to detect nucleotide diversity in 20 Mb windows across all 54 individuals. The observed proportion of heterozygous sites and the inbreeding coefficient (f) for each individual per population were also calculated using VCFtools. I tested for significant differences in the genetic diversity, heterozygous sites and inbreeding coefficients between populations (wild, captive, and reintroduced) by using an analysis of variance test (ANOVA) followed by post-hoc Tukey tests implemented in R v 4.1.1 (R Core Team, 2024) (Table 3.03).

3.3.5 Detection of signatures of genetic bottleneck: Runs of homozygosity and identification of deleterious mutations

To detect signatures associated with a genetic bottleneck, I calculated the distribution of runs of homozygosity (ROH) using PLINK v1.9 (Chang et al., 2015; Purcell et al., 2007) using these parameters `--homozyg-window-snp 50 --homozyg-snp 50 --homozyg-kb 500 --homozyg-gap 1000 --homozyg-density 50 --homozyg-window-missing 5 and --homozyg-window-het 3` (Supplementary material S3.6) (Humble et al. 2023). The ROH were then categorised into three length classes; 300-500 Kb, 500 Kb-1 Mb, and >1 Mb. I tested for significant differences in the F_{roh} between populations by using a Kruskal Wallis

chi-squared test followed by a Dunn post-hoc test in R v 4.1.1 (R Core Team, 2024) (Table 3.03)).

SnEff v5.1 (Cingolani et al., 2012) was used to detect the functional impact of biallelic SNPs present in the black-footed ferret population. SnEff uses a known annotation database to predict the impact of derived alleles within genic regions. I used the black-footed ferret annotation (Kliver et al. 2023) as the input database. I then assessed whether mutational load differed between populations and this was calculated based on homozygous and heterozygous mutational load within missense and loss-of-function (LoF) categories. Significant differences in mutational load across populations were tested using an analysis of variance test (ANOVA) followed by post-hoc Tukey tests implemented in R v 4.1.1 (R Core Team, 2024) (Supplementary material S3.7 and Supplementary Table 3.03).

Due to the issue that could arise from calling deleterious mutations based on the reference genome of the black-footed ferret, which is derived from a captive-bred, and therefore potentially highly inbred, male (Kliver et al., 2023), a comparative approach was used to polarise the variant calls.

To achieve this, I first generated a whole-genome alignment with Progressive Cactus v2.2.0 (Armstrong et al., 2020), using the eleven publicly available mammalian genomes (eight mustelids and three outgroups) described above. Next, I extracted the ancestral genome from the Cactus alignment. This was run with RepeatMasker v 4.0.8 (<http://www.repeatmasker.org>) (Tarailo-Graovac & Chen, 2009) to identify repetitive regions within the ancestral genome. The ancestral repeats were used to infer a phylogenetic tree under a neutral model using phyloFit (part of the PHAST toolkit) (Pollard et al., 2010; Siepel & Haussler, 2004). This tree was then used as a guide for rerunning Progressive Cactus to refine the alignment.

Using this alignment, I identified the nucleotide identities of positions called as LoF or missense in the black footed ferret from its four sister species (domestic ferret, European polecat, Steppe polecat, and weasel), using halLiftover v 2.2 (Hickey et al., 2013). Finally, I intersected the nucleotide identities with our called variants to identify derived alleles in the black-footed ferret.

3.3.6 Gathering Gene Family clusters

To investigate potential changes in gene structure associated with mutation accumulation in small populations, I examined patterns of gene family expansion and contraction across mustelid genomes. In populations with reduced effective population size, the accumulation of deleterious mutations and reduced efficacy of purifying selection can lead to gene loss, pseudogenisation, or alterations in gene copy number (Kumar et al. 2023; Peers et al. 2025). Examining gene family evolution can therefore provide insight into how genomic architecture may be affected by prolonged demographic decline.

To identify patterns of gene family expansion and contraction within mustelids, I used the GeneSeqToFamily (GSTF) pipeline (Thanki et al., 2018). GSTF incorporates the Ensembl GeneTrees pipeline in a workflow comprising six main steps. For input to GSTF, I used eight of the eleven curated genomes and their publicly available genome annotations. At the time of this analysis, there was no publicly available genome annotation alongside the genome for the weasel (*Mustela nivalis*). Additionally, I also decided to remove the stone marten (*Martes foina*) and steppe polecat (*Mustela eversmanii*) due to errors found within the genome annotations. The following steps were taken by running GSTF:

1. Firstly, the longest coding sequence (CDS) transcript per gene was extracted from the input sequences. This required an additional step of extracting coding sequences from each species.
2. Once the longest CDS for each gene was extracted, the CDS for each species was collated for input to the next step.
3. CDS were then translated to protein sequences.
4. GSTF generated pairwise alignments of protein sequences and sorted them into clusters based on similarity. As a result, multiple sequence alignments were generated for each cluster.
5. Gene trees for each gene family were then generated using the cluster alignment and user-specified species tree as input. The species tree was generated from extracted ancestral repeats from the whole-genome alignment and inferred using phyloFit (part of the PHAST toolkit) (Siepel & Haussler 2004).
6. Finally, these trees were aggregated to form a data table detailing gene copies per species. From this, orthologous gene copies and families can be identified, and gene family expansions or contractions analysed.

3.3.7 Gene Family Expansions and Contractions

CAFE v5.0 (Mendes et al., 2021) was used to consolidate gene family expansion and contraction predictions across eight mammalian species. For input to CAFE, a species tree was inferred using 10,212 single-copy orthologs identified through GSTF and used as input to ASTRAL v5.14.2 (Mirarab et al. 2014; Zhang et al. 2018). An ultrametric tree was

then produced from the species tree input into CAFE. The overall lambda option was used to estimate gene gains and losses across branches. For each gene family, CAFE modeled the likelihood of observing each gene count under lambda and tested whether the gene gain or loss on each node is expected. This approach was used to calculate the overall gene family gains and losses over each node, as well as significant gene family gains and losses ($p < 0.01$). Gene ontology (GO) enrichment was then carried out using DAVID v6.8 (Dennis et al. 2003; Sherman et al. 2022) on all significant gene family gains and losses in the black-footed ferret branch.

3.3.8 Detecting signatures of selection

Population bottlenecks can lead to the accumulation of weakly deleterious mutations in the genomes of surviving individuals, as reduced population size limits the effectiveness of purifying selection in removing these mutations. These mutations can then quickly become fixed within the population due to drift. Therefore, elevated dN/dS can reflect both adaptive evolution or the accumulation of weakly deleterious mutations due to relaxed purifying selection in small or bottlenecked populations (Bortoluzzi et al. 2020; Beichman et al. 2023; Dobrynin et al. 2015). The CODEML program, implemented by PAML (Yang 2007; Álvarez-Carretero et al. 2023) allows for the detection of genes under positive selection, by comparing synonymous (d_S) and nonsynonymous (d_N) substitution rates across coding sequences from different species. Any gene with a d_N/d_S (ω) > 1 , is said to have an accelerated nonsynonymous substitution rate and is under positive selection. It should be noted that CODEML assumes constant population sizes among the species included in the comparison (Yang 2007; Álvarez-Carretero et al. 2023). However, given the well-documented demographic history of the black-footed ferret, I hypothesised that elevated dN/dS values may instead reflect the accumulation of mutations associated with a reduced effective population size rather than signatures of positive selection.

Using the previously identified 10,212 GSTF single-copy orthologs present in all species, I extracted and aligned nucleotide sequences for input into CODEML. First, stringent filtering of alignments was carried out to alleviate false positives that could occur due to errors or gaps occurring within the sequence alignments as CODEML is highly sensitive to errors that can arise in the alignment process so several steps were taken to avoid this. Orthologous sequences were aligned in a codon-aware manner using PRANK v 150803 (Löytynoja, 2014), and any excess residues were masked using GUIDANCE2 v 2.0.2 (Sela et al., 2015). Any alignments that were less than 300 nucleotides in length were discarded from the analysis.

Additionally, I used the branch-site test implemented in CODEML (Yang 2007) using the models M0 and M2, where M0 assumes the same d_N/d_S ratio across all branches and M2

assumes differing d_N/d_S ratios across the background (human, cat, dog, Eurasian otter, stoat, European polecat and domestic ferret) and the foreground (black-footed ferret) branches.

A likelihood ratio test (LRT) was used to compare models M0 (one-ratio) and M2 (positive selection) across the dataset to determine the best-fitting model for each gene. Genes showing a statistically significant elevation in d_N/d_S , following Storey-Tibshirani false discovery rate correction (Storey & Tibshirani, 2003), and which also had at least one codon site identified as being under positive selection by the Bayes Empirical Bayes (BEB) method were subjected to further manual inspection.

The BEB method estimates the posterior probability that a codon belongs to a specific ω (d_N/d_S) class under the best-fitting model (e.g., M2), and highlights sites with high probability (typically >0.95) of being under positive selection ($\omega > 1$). To reduce the likelihood of false positives due to misalignment, stringent filtering was applied: any BEB-identified site was flagged as unreliable if a gap was present within five codons upstream or downstream of the site in the alignment. Additionally, any sequences showing evidence of misannotation or misalignment in the vicinity of a BEB-flagged site were excluded from further analysis.

Figures 3.1 and 3.2 illustrate examples of alignments considered poor and high quality, respectively, based on these criteria.

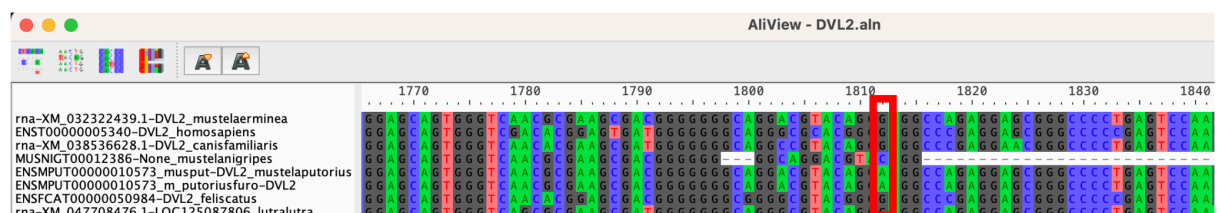


Figure 3.1: Alignment of the *DVL2* coding sequence from across the eight mammalian species. The red box highlights the site that CODEML reported as a BEB site. Note the upstream and downstream gaps either side of the highlighted box, which resulted in this alignment being categorised as ‘bad’.

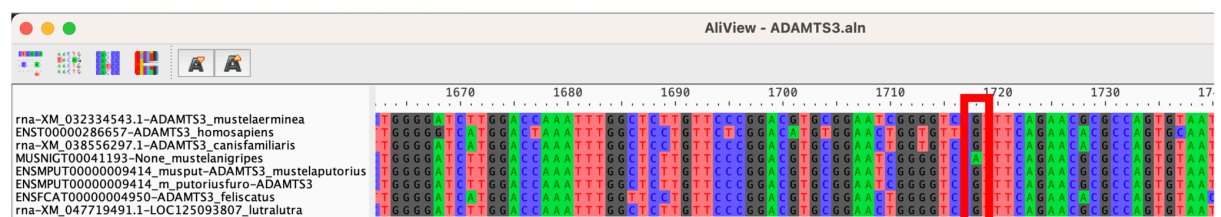


Figure 3.2: Alignment of the *ADAMTS3* coding sequence across the eight mammalian species. The red box highlights the site that CODEML reported as a BEB site. Note the

lack of gaps downstream and understream of the highlighted box, which resulted in this alignment being categorised as 'good'.

3.3.9 Assessing the impact of BEB sites

To investigate the impact of the nucleotide changes that were flagged by CODEML, I first translated the black-footed ferret genomic position to the genomic position in the domestic ferret. The black-footed ferret individual used to produce the reference genome was a captive-bred male. Therefore, this presented an issue whereby this individual could have deleterious mutations already fixed within its genome. These mutations could be masked from population genomic analyses when aligning other individuals from the population to this reference if they are also fixed within those individuals. To alleviate this, the genomic positions of the sites that were highlighted as being under selection were manually aligned to the domestic ferret genome. Briefly, each black-footed ferret coding sequence that had a reliable BEB site was aligned against the Ensembl domestic ferret genome using BLAST. The genomic position of the BEB sites in domestic ferret were then manually confirmed by checking the sequence identity in the black-footed ferret 10 nucleotides $\pm$ the BEB site.

In order to predict the functional impact of the variants identified in this way, these genomic positions were used as input to SIFTv4G (Ng & Henikoff, 2003). SIFT can predict whether an amino acid change impacts protein function based on conserved sequence homology and does this by searching for similar orthologous sequences from a known database (in this case the Uniref90 database (Suzek et al., 2007)). It then obtains the alignment of these chosen sequences and calculates normalised probabilities for all possible substitutions within those alignments. Candidate sites were functionally annotated using the domestic ferret MusPutFurv0.83 database (Ng & Henikoff, 2003; Vaser et al., 2016). Any positions that have a normalised probability of <0.05 are predicted to be deleterious, those ≥ 0.05 are predicted to be tolerated. Genes containing deleterious substitutions were then examined in relation to their known functions using literature and database searches, including the Mouse Genome Informatics (MGI) database (Eppig, 2017).

SNPs identified as deleterious were assessed for their presence across the whole-genome population dataset using BEDTools v 2.31 (Quinlan & Hall, 2010). The *bedtools -intersect* function was used to compare the positions of the deleterious variants with the population data, allowing determination of allele counts at each site and identification of whether the variants were fixed among the sampled individuals. To investigate their potential impact, literature and database searches - including the Mouse Genome Informatics (MGI) database (Eppig 2017), were conducted.

3.4 Results

Whole-genome variants were called for 54 individuals belonging to three population groups (Wild, Captive, Reintroduced). Before filtering, this consisted of 2,881,885 SNPs. After quality filtering and selecting for bi-allelic SNPs, the dataset consisted of 1,927,827 SNPs and 944,698 SNPs remained after pruning for LD. Supplementary Table 3.01 breaks down which variant datasets were used for each analyses throughout this chapter and the filtering steps applied.

3.4.1 Relatedness

I first tested for relatedness between individuals using the KING relatedness coefficient. The analysis demonstrated the high degree of relatedness between individuals, with coefficients ranging from 0.144-0.355 (Figure 3.3).

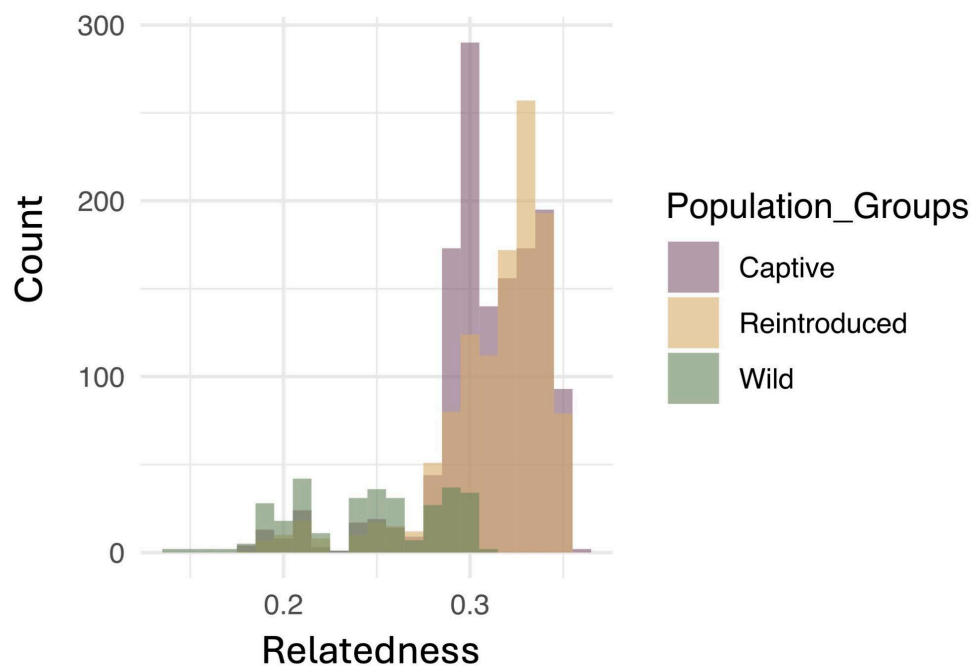


Figure 3.3: Distribution of relatedness (PHI) values across population groups. The histogram shows the frequency of pairwise relatedness estimates for individuals. Each bar represents the count of relatedness values within a specific range, and the fill color indicates the population group of the individual pairs. The relatedness values range from unrelated individuals (<0.044) to monozygotic twins or duplicates (>0.354) (Manichaikul et al., 2010).

The lowest degree of relatedness was observed between individuals from the wild South Dakota population and those from the Wyoming captive and reintroduced populations (Supplementary figure S3.1). This is particularly evident as the two individuals from South Dakota are not genetically represented in the present-day population and originated from the distinct South Dakota population, in contrast to individuals from Wyoming, the source of the founding population (Supplementary material S3.1). Based on these estimates, all sampled black-footed ferrets exhibit first-degree relatedness, with individuals from South Dakota displaying second-degree relatedness.

3.4.2 Population structure

After filtering the dataset consisting of whole-genome SNPs for 54 individuals, 944,698 SNPs remained. Principal component analysis (PCA) of whole-genome SNPs revealed distinct separation between the main captive/reintroduced population and four individuals from the wild populations (South Dakota and Wyoming) described in PC1 (16.9%) and PC2 (8.3%) (Figure 3.4). The two individuals from Wyoming - SB2 and SB10 - were from the original 18 founding individuals that were brought into the captive breeding programme in 1981. The clustering observed demonstrates the distinct genetic diversity represented in these two lineages in comparison to the captive and reintroduced populations. SB10's genetic lineage had been underrepresented in the current population until the use of somatic cell nuclear transfer, which produced three cloned individuals.

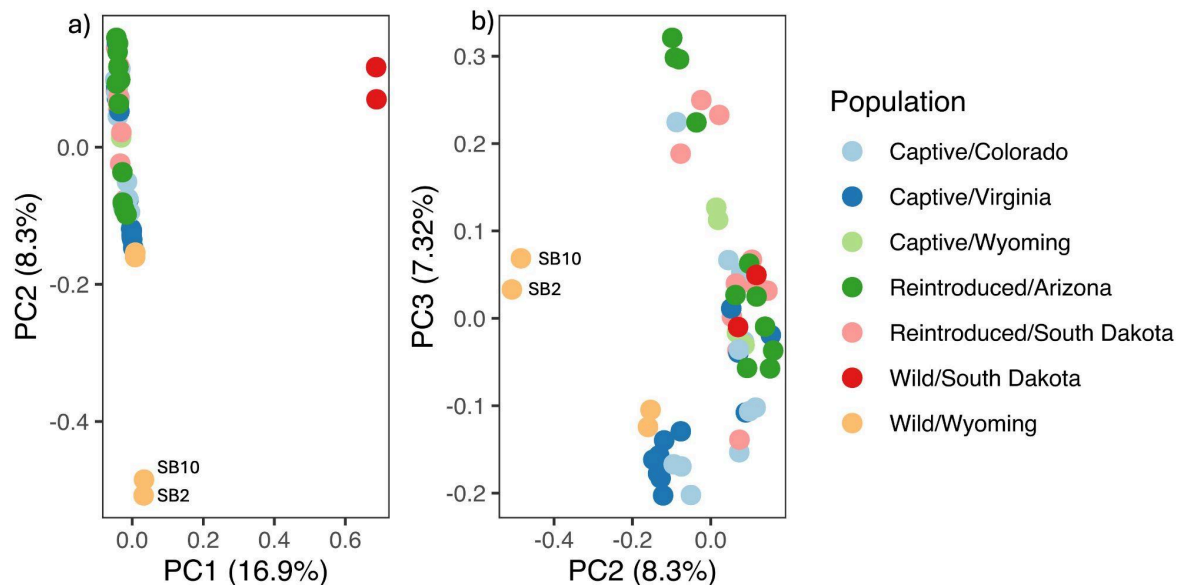


Figure 3.4. Population structure of black-footed ferret individuals inferred from whole-genome SNP data (944,698 variants). Principal component analysis (PCA) plots show clustering patterns among individuals from Wild ($n = 6$), Captive ($n = 26$), and Reintroduced ($n = 22$) populations: (a) PC1 versus PC2 and (b) PC2 versus PC3. Two Wild individuals (SB10 and SB2) cluster separately from the main population groups, highlighting elevated genetic divergence relative to the other sampled individuals.

I then investigated the population structure of the current and historic population of the black-footed ferret by running ADMIXTURE on whole-genome SNPs (Figure 3.5). For this, I ran values of $K = 1-12$, and recovered admixture proportions pertaining to the different population groups (Wild, Captive, Reintroduced) under higher estimations of K . However the output of this analysis was best supported by a cross-validation error of $K=1$ (Supplementary material S3.2, Supplementary table 3.02). This suggests that the sampled individuals are a homogenous population with no distinct genetic clusters.

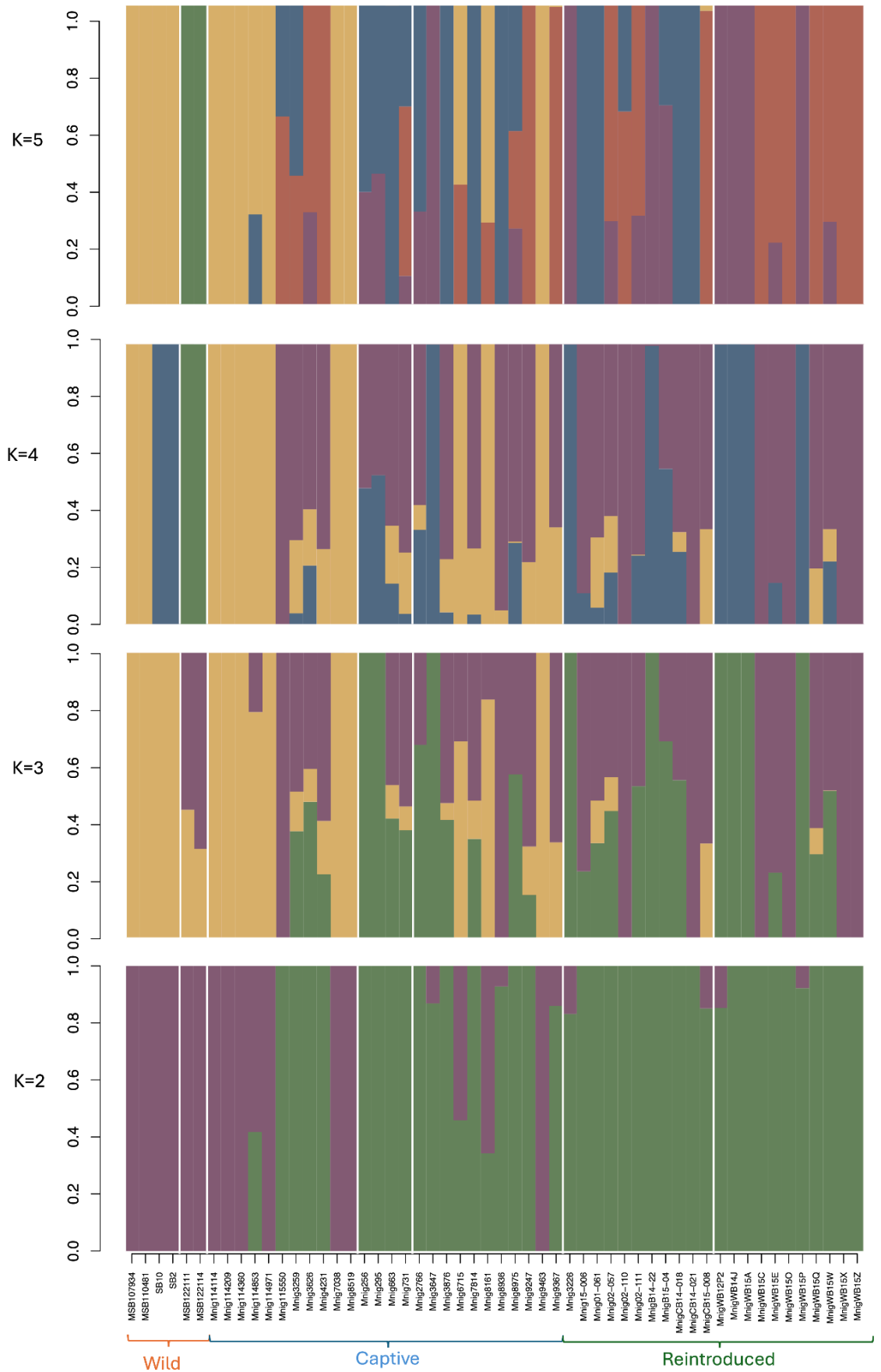


Figure 3.5: ADMIXTURE analysis of population structure based on 944,698 whole-genome SNPs from 54 black-footed ferret individuals across three population groups: Wild (n = 6), Captive (n = 26), and Reintroduced (n = 22). Results are shown for K

= 1–5 ancestral clusters. The analysis supports a single, genetically homogeneous population, with the best-supported model at $K = 1$.

3.4.3 Genetic Diversity

To investigate the genomic impact of the severe population contraction prior to a founder bottleneck, and further sustained inbreeding black-footed ferret populations have experienced, I examined the genome-wide genetic diversity of the individuals.

Nucleotide diversity (Tajima's π) estimates showed a significant difference among populations ($F_{2, 38} = 18.98$, $p < 0.01$). Post-hoc comparisons using Tukey's HSD test revealed that the reintroduced population had significantly higher nucleotide diversity than the wild population ($p < 0.01$), and the captive and reintroduced populations also differed significantly ($p < 0.01$). However, there was no significant difference between the captive and wild populations ($p = 0.33$) (Figure 3.6) (Supplementary table 3.03).

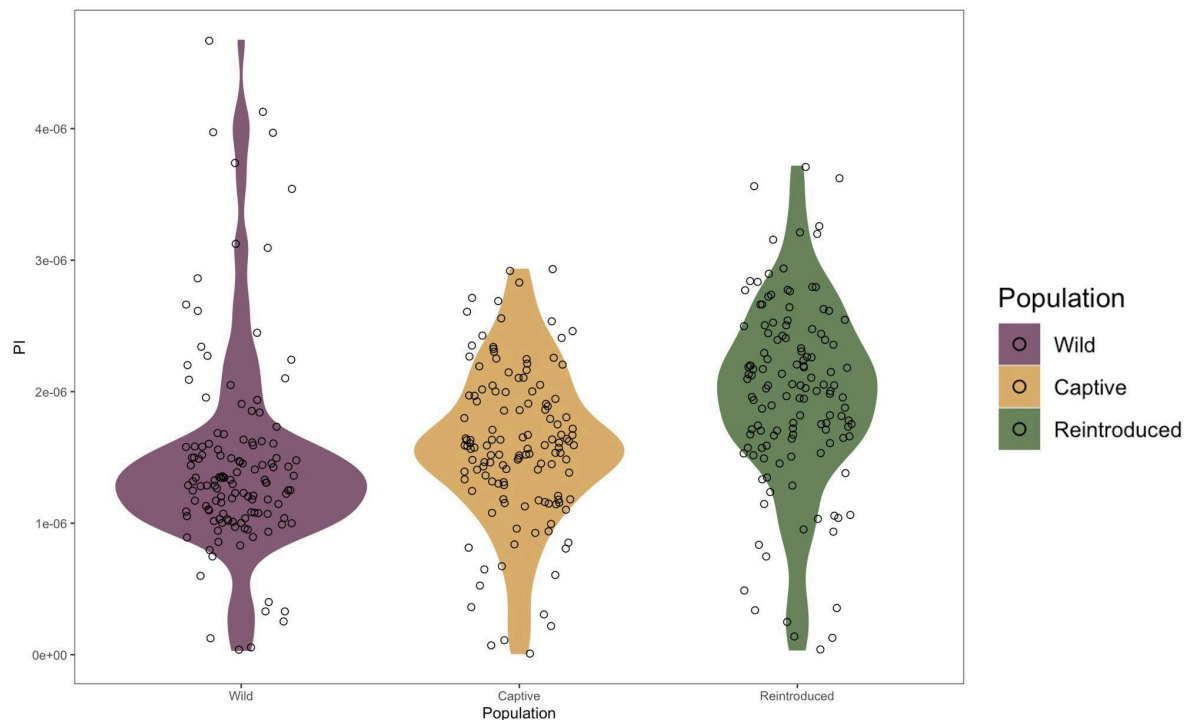


Figure 3.6 Comparison of genome-wide genetic diversity among sampled black-footed ferret populations: Wild ($n = 6$), Captive ($n = 26$), and Reintroduced ($n = 22$). Population nucleotide diversity (π) is shown for each group, calculated from a dataset of 1,927,827 SNPs using VCFtools. Estimates were generated in 20 Mb sliding windows across the genome, allowing for assessment of variation in diversity between populations.

The proportion of heterozygous sites revealed a significant difference among populations ($F_{6, 47} = 9.70$, $p < 0.01$). Following posthoc Tukey tests, I found that the wild (Wyoming and South Dakota) populations had significantly lower heterozygosity than the captive

Wyoming population ($p = <0.01$ for Wyoming, $p = <0.01$ for South Dakota), captive Colorado population ($p = <0.01$ for Wyoming, $p = 0.04$ for South Dakota), reintroduced Arizona population ($p = <0.01$ for Wyoming, $p = <0.01$ for South Dakota) and reintroduced South Dakota population ($p = <0.01$ for Wyoming, $p = <0.01$ for South Dakota). However, no significant differences were observed between the wild South Dakota and captive Virginia populations ($p = 0.66$) or between the wild Wyoming and captive Virginia populations ($p = 0.49$). The reintroduced populations (Arizona and South Dakota) exhibited significantly higher heterozygosity than the captive Virginia population ($p = <0.01$ for Arizona, $p = <0.01$ for South Dakota). However, no other significant differences were observed between captive and reintroduced populations or between wild populations (Figure 3.7) (Supplementary table 3.03).

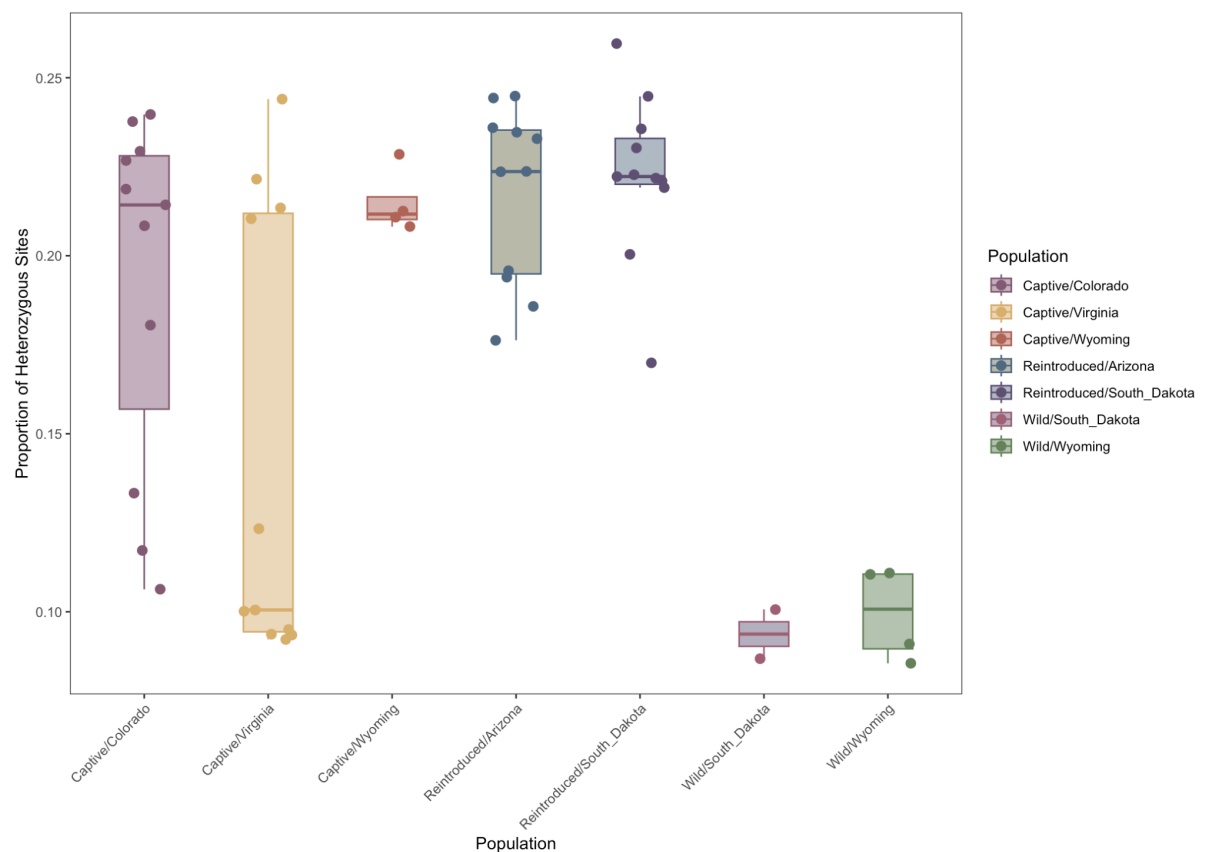


Figure 3.7: Genetic diversity estimates across black-footed ferret population groups: Wild ($n = 6$), Captive ($n = 26$), and Reintroduced ($n = 22$), measured as the proportion of heterozygous sites per individual. Heterozygosity was calculated using VCFtools. Boxplots represent the median and interquartile range of heterozygosity values within each population group. Individual data points are overlaid as jittered dots to show within-group variation.

Inbreeding coefficients (f), calculated using VCFtools, were found to differ significantly across the population groups ($F_{6,47} = 9.49$, $p < 0.01$) (Figure 3.8). Estimates ranged from

-0.39 to 0.48. Both wild populations (Wyoming and South Dakota) were found to possess significantly higher inbreeding coefficients in comparison to both reintroduced populations ($p < 0.01$) and captive populations in Colorado and Wyoming ($p < 0.01$). The captive Virginia population also had a significantly higher measure of inbreeding in comparison to reintroduced groups ($p < 0.01$) and the captive Wyoming population ($p = 0.05$) but not compared to the captive population in Colorado ($p = 0.08$) (Supplementary table 3.03).

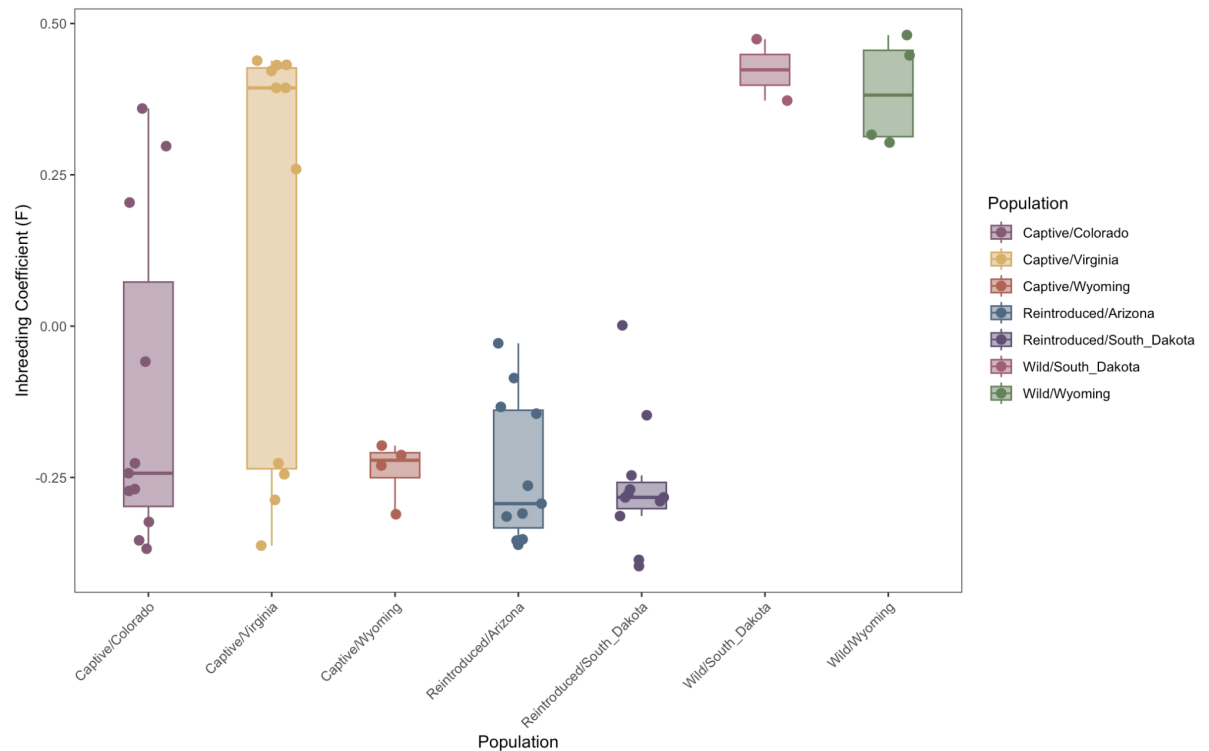


Figure 3.8: Inbreeding coefficient values per individual within population groups: Wild ($n = 6$), Captive ($n = 26$), and Reintroduced ($n = 22$), as calculated by VCFtools. Boxplots represent the median and interquartile range of inbreeding coefficient values within each population group. Individual data points are overlaid as jittered dots to show within-group variation.

3.4.4 Runs of Homozygosity

I used PLINK v 1.9 to calculate the realised inbreeding coefficient (F_{ROH}), which ranged from 0.24 - 0.71, and was significantly different between population groups (captive, reintroduced and wild) ($H = 8.87$, $df = 2$, $p = 0.01$). Dunn's post-hoc comparisons showed that F_{ROH} values were not significantly different between captive individuals compared to reintroduced ($Z = 0.91$, $p = 0.17$), but significantly higher between wild and captive ($Z = -2.44$, $p < 0.01$) as well as between wild and reintroduced individuals ($Z = -2.97$, $p < 0.01$) (Figure 3.9) (Supplementary material S3.6 and Supplementary table 3.03.).

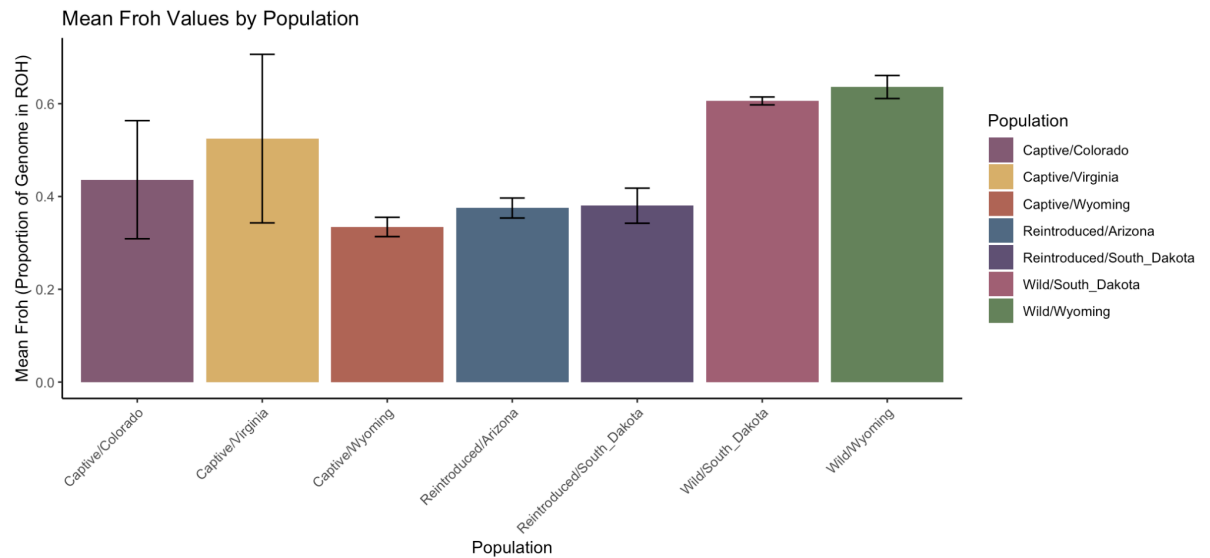


Figure 3.9: Mean inbreeding coefficient (F_{ROH}) for each black-footed ferret population group: Wild ($n = 6$), Captive ($n = 26$), and Reintroduced ($n = 22$), as calculated by PLINK. Bars represent the mean Froh value per population, with error bars indicating the standard error of the mean. Froh was calculated as the proportion of the genome in runs of homozygosity (ROH).

I separated ROH by length to visualise variation in counts of short (300-500 Kb), medium (500-1 Mb), and long (>1 Mb) ROH between captive, reintroduced, and wild populations (Figure 3.10, Figure 3.11) (Supplementary material S3.6). Longer ROH are indicative of recent inbreeding events, with shorter ROH being indicative of historic events. Individuals from both of the reintroduced populations had the lowest mean number of short and long ROH but had a higher average medium ROH length than both captive and wild populations.

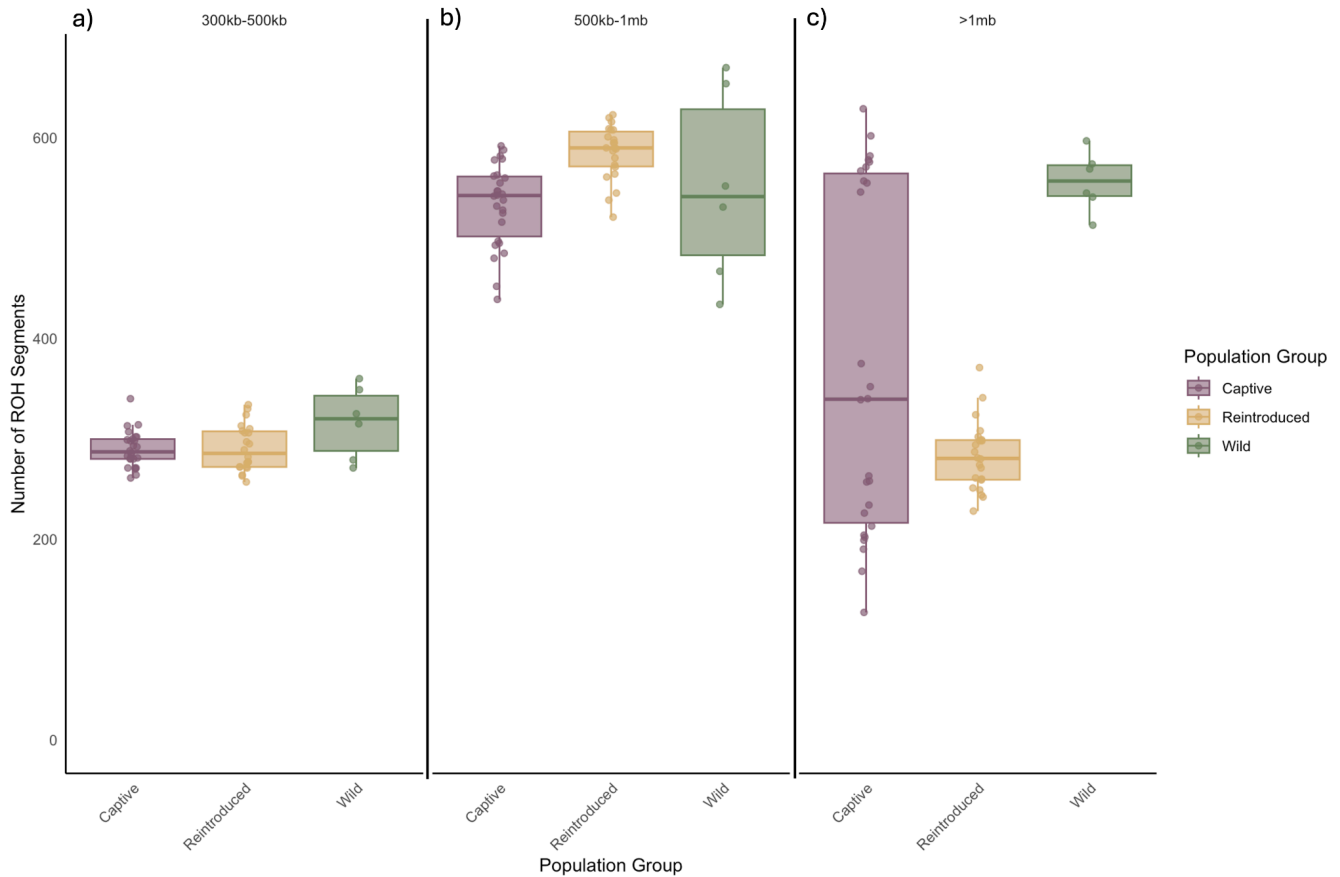


Figure 3.10: Number of runs of homozygosity (ROH) segments, as calculated by PLINK per individual across black-footed ferret population groups: Wild ($n = 6$), Captive ($n = 26$), and Reintroduced ($n = 22$), grouped by ROH length. Each facet corresponds to a different ROH length bin a) 300 Kb–500 Kb, b) 500 Kb–1 Mb and c) >1 Mb. Boxplots show the distribution of segment counts per individual within each population group, representing the median and interquartile range. Jittered points indicate individual-level values.

Individuals belonging to the wild population were seen to have a greater proportion of their genome within longer ROH segments (>1 Mb) ($39.3\% \pm 1.86$ SE) in comparison to the reintroduced ($24.33\% \pm 0.361$ SE) and captive ($29.0\% \pm 2.0$ SE) individuals (Figure 3.12). However, individuals from Mellette County (South Dakota) were observed to have lower counts of ROH in all length classes than individuals from both the wild Wyoming population and Meteetsee (South Dakota).

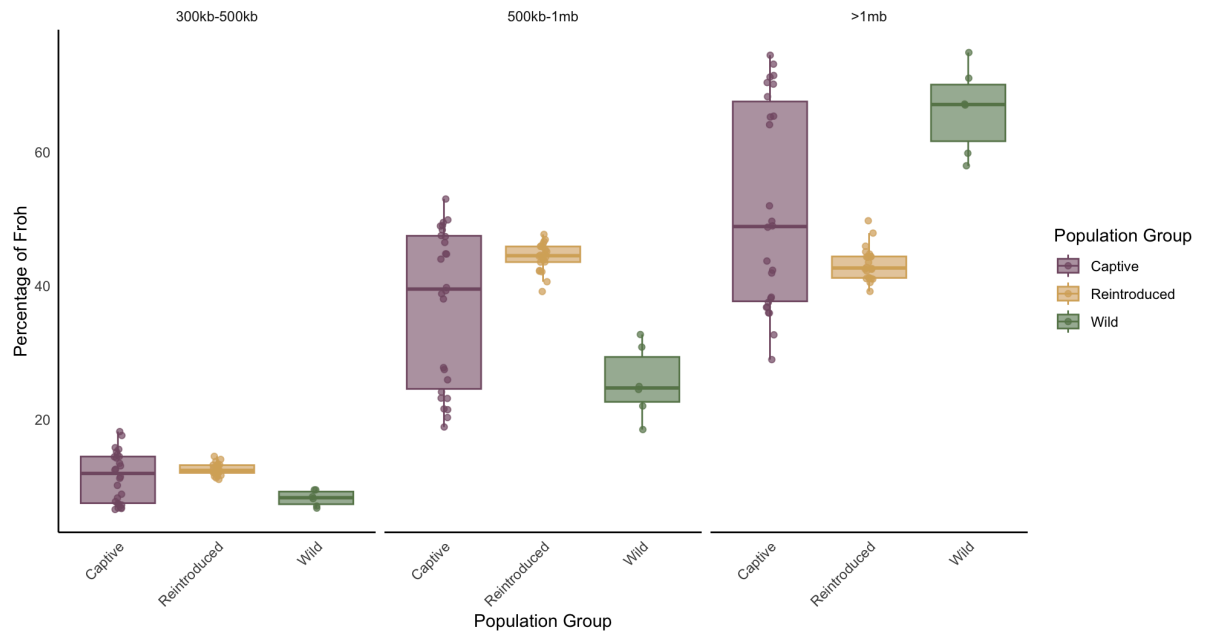


Figure 3.11: Inbreeding coefficient (F_{ROH}) distribution across black-footed ferret population groups: Wild ($n = 6$), Captive ($n = 26$), and Reintroduced ($n = 22$), faceted by ROH length. Each facet represents a different ROH length bin (e.g., 300 Kb–500 Kb, 500 Kb–1 Mb, > 1Mb). F_{ROH} is the percentage of the genome contained within ROH segments of the specified length. Boxplots display the median and interquartile range of F_{ROH} within each population group; individual points are overlaid as jittered dots to show within-group variation.

3.4.5 Gene Family Expansions and Contractions

Gene clusters were constructed using eight mammalian genomes to infer gene family expansions and contractions along each branch. This analysis revealed that the black-footed ferret contained 16,649 orthologous gene families. Among these, 12,097 were shared across all eight genomes analysed. A total of 13,248 gene families were shared amongst Mustelidae and 13,553 were shared amongst the Mustelinae.

I then applied the ‘overall lambda’ option within CAFE v5.0 to determine gene family expansion (births) and contraction (deaths) rates along all eight branches of the 20,363 GSTF gene family trees. CAFE estimates the best-fitting lambda (λ) - the rate of gene gain and loss - by maximising the log-likelihood across a range of candidate values. CAFE identified 0.0052 as the optimal rate of gene gain and loss (λ).

Under these assumptions, the black-footed ferret genome has 469 expanded and 1,162 contracted gene families. Of these gene families, CAFE reported 172 expanded and 318 contracted families as significant ($p < 0.05$) (Figure 3.12).

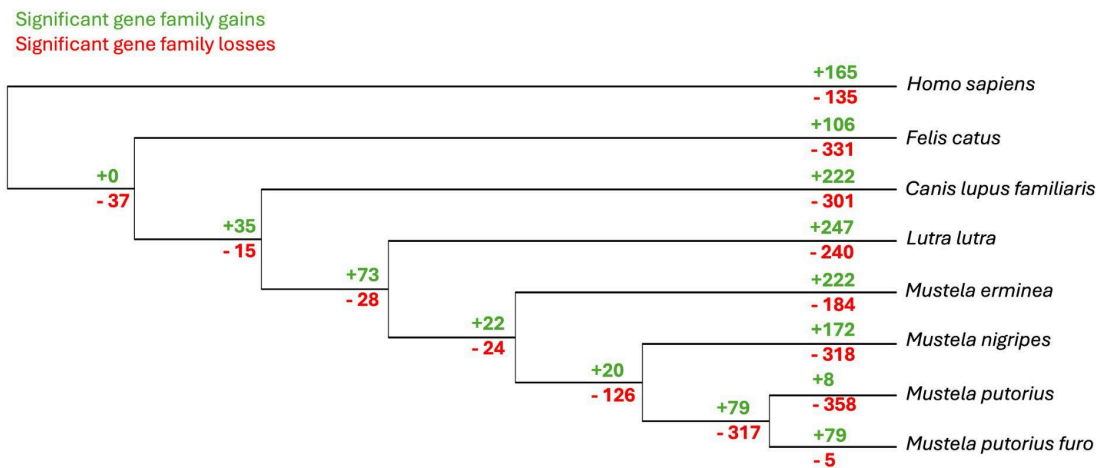


Figure 3.12: Analysis of gene family expansions and contractions across 20,363 gene clusters as calculated by Cafe v5.0 with an overall lambda (0.0052). 12,097 gene families were shared amongst all eight genomes, 13,248 were shared amongst Mustelidae and 13,553 were shared amongst species belonging to Mustelinae.

I used GProfiler to detect the gene ontology enrichment of genes within statistically significant gene family expansions and contractions across the black-footed ferret branch. Significant gene family expansions along the black-footed ferret branch were associated with olfactory receptor activity (23 genes), detection of chemical stimulus (26 genes) and the regulation of biological processes (158 genes). Significant gene family contractions along the black-footed ferret branch are associated with sensory perception of smell (26 genes), G protein receptor activity (30 genes) and olfactory receptor activity (26 genes) (Supplementary material S3.4, S3.5).

3.4.6 Positive Selection analysis

Using the branch site test within the CODEML program, I identified 813 genes that had statistically significant elevated d_N/d_S ($q \leq 0.01$). Of these 813 genes, 809 reported, with 708 genes reporting an $\omega = 999$ which is indicative of one nonsynonymous mutation being present within the coding sequence. Eight of these gene sequences did not report a BEB site, which determines the position under positive selection within the gene. After visual inspection of the remaining 1091 reported BEB sites, 364 BEB sites in 231 genes were determined as high-confidence candidates for positive selection.

3.4.7 Assessing the impact of deleterious variants across the black-footed ferret mutation landscape - using SnpEff and SIFT

SnpEff

Using SnpEff, I characterised deleterious populations mutations segregating across the genome of the black-footed ferret. SnpEff characterised 6,415 mutations as having low impact, 7,162 as moderate impact and 401 high impact mutations (Supplementary material S3.7).

I then investigated whether the mutation landscape differed across the three population groups (Captive, Reintroduced, and Wild) (Figure 3.13). Missense and loss of function (LoF) mutations were assessed, as reported by SnpEff, to quantify mutational load in the black footed ferret. I identified whether alleles at these loci were homozygous or heterozygous per individual, per category. Additionally, I also assessed whether the allele changes were derived (found only in the black-footed ferret) versus ancestral (conserved across other mustelids).

For LoF mutations, heterozygous, non-derived counts differed significantly among populations ($F_{2, 51} = 7.88$, $p = 0.001$), with Reintroduced individuals exhibiting higher counts than both Captive and Wild groups. Homozygous, non-derived, LoF mutations also varied significantly ($F_{2, 51} = 13.39$, $p < 0.001$), with Wild individuals carrying a higher number than both Reintroduced and Captive. In contrast, neither heterozygous nor homozygous derived LoF alleles showed significant differences between groups.

Missense mutations followed a similar pattern. Heterozygous, non-derived mutations were significantly different across populations ($F_{2, 51} = 7.94$, $p < 0.01$), with Wild individuals exhibiting lower counts. A significant effect was also detected for heterozygous derived missense mutations ($F_{2, 51} = 6.26$, $p = <0.01$), with Reintroduced individuals showing higher counts than Captive individuals. Homozygous non-derived missense counts showed a strong population effect ($F_{2, 51} = 11.59$, $p = <0.01$), with Captive individuals carrying significantly fewer mutations compared to both Reintroduced and Wild groups. Significant homozygous derived missense mutations differences were observed between the Wild population group and Captive (<0.01) and Reintroduced (<0.01) individuals. It was observed that two individuals from the 1970s wild South Dakota population (*MSB122111*, *MSB122114*) had the highest overall realised (homozygous) load for both mutation categories (Figure 3.13).

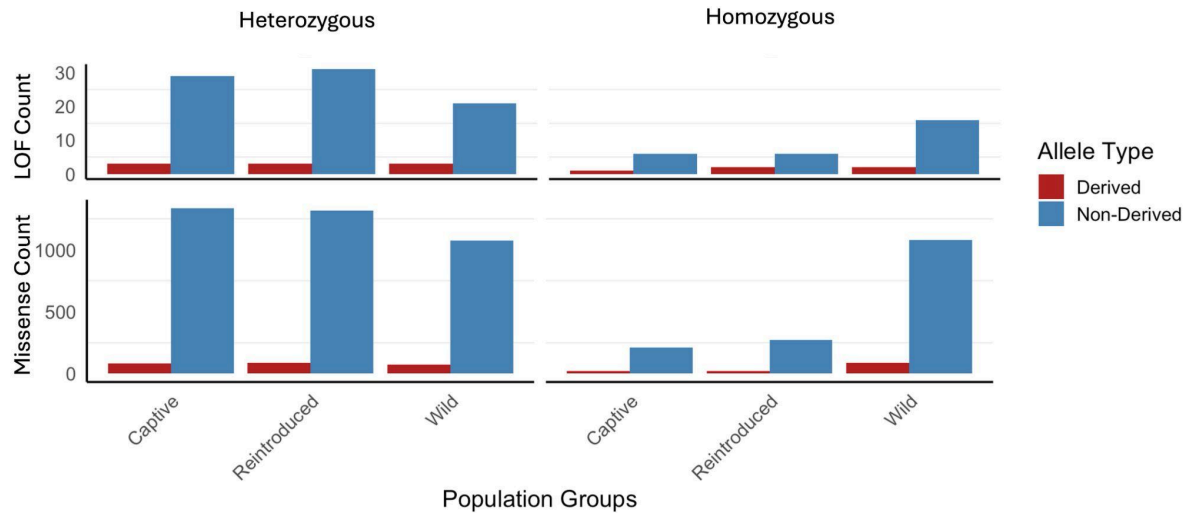


Figure 3.13: Patterns of mutational load across the black-footed ferret populations. Aggregated counts of heterozygous and homozygous variant calls by mutation type (missense and loss-of-function (LoF)) across black-footed ferret population groups: Wild (n = 6), Captive (n = 26), and Reintroduced (n = 22). Variants were annotated using SnpEff and further classified by allele state (heterozygous or homozygous) and evolutionary origin, with derived alleles (unique to black-footed ferrets) shown in red and non-derived (ancestral) alleles shown in blue.

SIFT

Using SIFTv4G, I predicted the impact of mutations on protein function in the black-footed ferret. For input to SIFT I used the 364 BEB sites (sites within genes that showed elevated d_N/d_S ratios) identified using CODEML. Of these, SIFT generated predictions for 305 sites, classifying 66 ($p < 0.05$) as deleterious and 239 as tolerated (Figure 3.14) (Supplementary material S3.8).

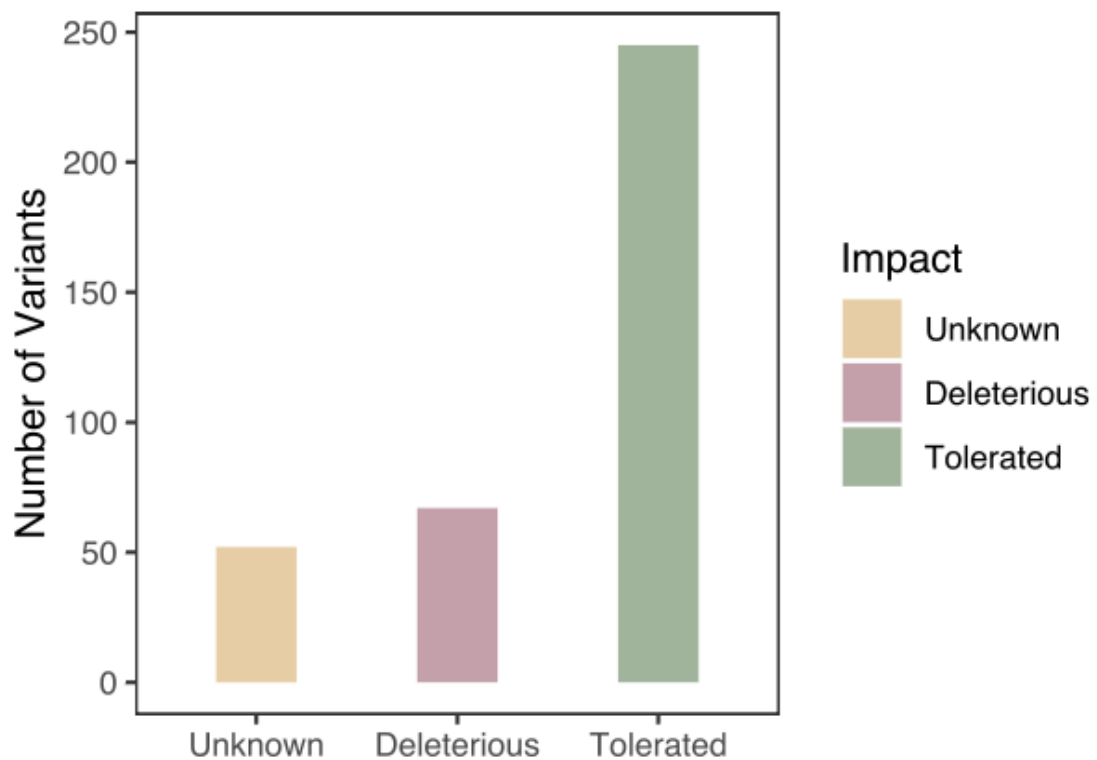


Figure 3.14: Predicted functional impacts of non-synonymous mutations at sites under positive selection in the black-footed ferret. A total of 364 Bayes Empirical Bayes (BEB) sites with elevated dN/dS ratios were identified using CODEML and analysed with SIFT4G to predict effects on protein function. SIFT generated predictions for 305 sites, classifying 66 variants as deleterious ($p < 0.05$; red) and 239 as tolerated (green), based on sequence conservation. The remaining 59 sites could not be assessed by SIFT, and their impact is therefore unknown (yellow).

Using the Mouse Genome Informatics (MGI) database to explore functions and identify possible consequences of mutations in those genes I found that, of the 66 genes that were predicted to be deleterious, 8 are associated with abnormal skeletal morphology, decrease in bone density or a decrease in bone mineral content (*GHRHR*, *EXOC3L1*, *NBEAL2*, *ADAMTS3*, *H6PD*, *COL12A1*, *BAZ2B*, *DNAJC28*), and five genes were reported to have an abnormal phenotype of preweaning lethality (*GBF1*, *TBCD*, *ZWILCH*, *SLIT2*, *FSD1*). There were also two genes associated with abnormal kidney function or morphology (*SLIT2*, *PLXNB2*). 5 genes were involved in behaviour and learning (*EXOC3L1*, *DNAJC28*, *TECTA*, *SRBD1*, *PRRC2A*). Two genes had reported abnormal phenotypes involved in the reproductive system such as abnormal epididymis (*GRIN3A*, *TST*). There were 7 genes involved in the hematopoietic system with abnormal phenotypes reported in a decrease in normal cell type levels (mostly white blood cells) (*MAMLD1*, *ASB16*, *ITGAE*, *DNAJC28*, *NBEAL2*, *TYK2*, *BAZ2B*). There were 6 genes

associated with immune response (*MAMLD1*, *ASB16*, *DNAJC28*, *TYK2*, *BAZ2B*, *FCRL5*) (Eppig, 2017) (Supplementary material S3.8).

3.4.8 Presence of putative deleterious nonsynonymous variants in the black-footed ferret population

To understand the impact that the 66 deleterious mutations identified using SIFT could have in the population of black-footed ferrets, I investigated whether these mutations were fixed in the population or if it was an artefact present in the reference genome. Using the whole-genome SNP data from 54 individuals, I assessed the allele frequency at each of the reported BEB sites (Figure 3.15).

Using bedtools *intersect* I found that 280 of the 305 annotated SIFT variants were fixed within the 54 black-footed ferret individuals. Of the 66 potentially deleterious mutations, 58 sites showed no variation within the population (Figure 3.6). The 9 sites that showed variation were within the genes *LUC7L3*, *DNAJC28*, *TYK2*, *TECTA*, *FCRL5*, *SRBD1*, *H6PD*, *SYNE1* & *MAMLD1*. One of the sites that was found within *MAMLD1* appeared to only be fixed in the reference genome (AF=1.0), whereas all other variations at these sites occurred at very low frequencies in the population (AF=0.009-0.0370). Only two individuals from the wild, 1970s, South Dakota population (*MSB122111* and *MSB122114*) showed any diversity at these sites.

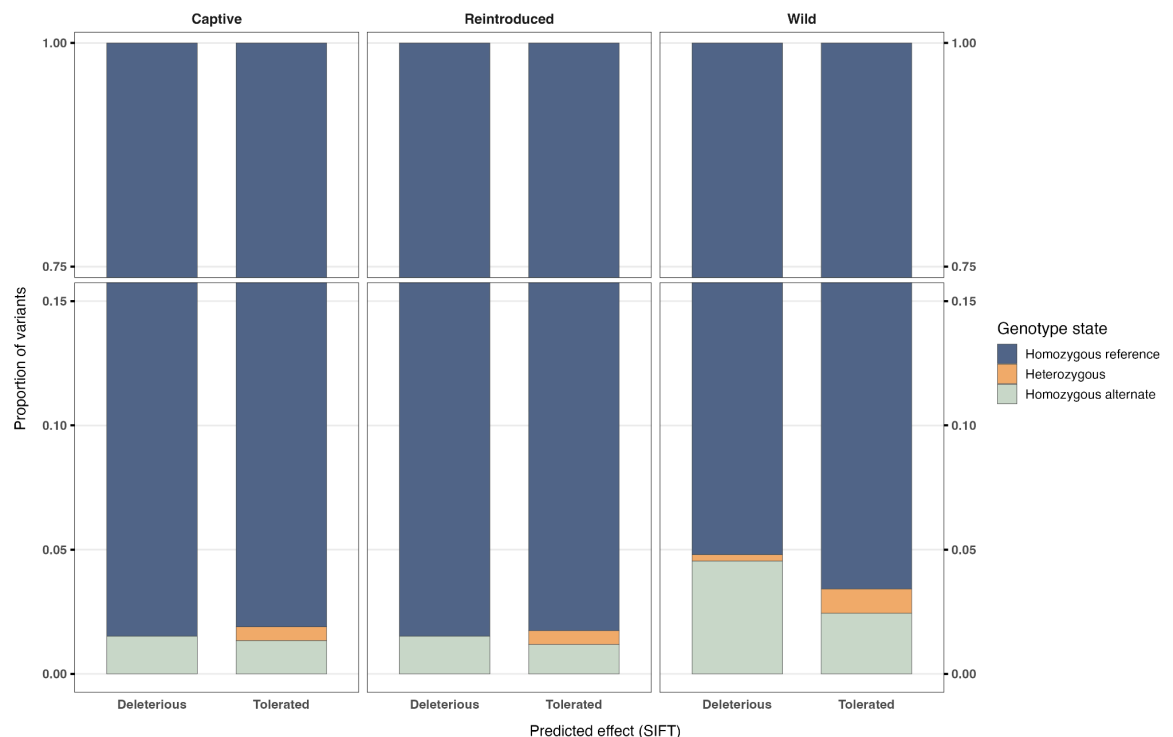


Figure 3.15: Bar plots show the proportion of individual genotypes at variants annotated by SIFT, separated by prediction (deleterious vs. tolerated) and stratified by population

groups (Captive, Reintroduced and Wild). Genotype states are classified as homozygous reference (dark blue), heterozygous (orange), and homozygous alternate (light green). Each facet represents a different population group, allowing comparison of genotype composition across populations. The y-axis is split to provide a greater resolution for each proportion at every genotype state.

3.5 Discussion

Populations that have undergone severe bottlenecks are often characterised by reduced effective population sizes and accumulation of weakly deleterious mutations leading to a decline in overall fitness. The black-footed ferret represents a striking example: after a drastic population crash, the species passed through a tiny founder group and endured generations of inbreeding in captivity. This history provides a rare opportunity to examine how reduced effective population size, inbreeding, and genetic drift interact to shape genomic variation. In this chapter, I characterise the genetic consequences of this bottleneck by comparing whole-genome SNP data from historic populations (1970s–1980s), with data from both captive-managed and reintroduced contemporary populations.

3.5.1 Genetic homogeneity found across black-footed ferret populations

I tested whether any population structure could be detected across black-footed ferret individuals from wild, captive or reintroduced populations using principal component analyses (PCA) and ADMIXTURE. Analyses revealed no significant genetic structure among the three populations, suggesting a largely homogeneous population..

This is consistent with the PCs, where no distinct clustering was noticed within groups from either captive or reintroduced individuals. A distinct cluster was indeed noticed within wild populations, with individuals from Mellette County (*MSB122111* and *MSB122114*), South Dakota (Figure 5) and two individuals from the Wyoming population (*SB2* and *SB10*) in PC1. This is resolved somewhat in PC2 for the Mellette county individuals, again clustering alongside the other individuals. However, the clustering of *SB2* and *SB10* remains in PC2, suggestive of some distinct genetic structure. *SB2* and *SB10* are individuals that were brought into the captive breeding programme in 1981, but were not successful breeders and so did not contribute to the founder population. This would explain the indication of distinct genetic structure between these two individuals and the rest of the sampled individuals.

Similarly, the Mellette County individuals were from another genetically isolated population that did not contribute to the founder population. This is observed within the output from our relatedness analyses, whereas lower coefficient of relatedness was seen between

MSB122111 and MSB122114 and the rest of individuals, as well as, to a lower degree, relatedness coefficients between SB2 and SB10 and the rest of individuals (Figure 3.4). Due to both SB10 and SB2 not contributing to the founding population in 1981, conservationists biobanked somatic tissue from these individuals in the hope that this would become a useful tool in the future.

Additionally I observed a higher degree of relatedness across the rest of the sampled black-footed ferret individuals from both reintroduced and captive population groups. Pedigree estimates from the current studbook suggest that the average relatedness coefficient is over 0.125, indicative of a relatedness level between first cousins (Novak et al., 2024). Estimates for sampled individuals were higher at 0.144-0.355, thus demonstrating the high inbreeding levels within the contemporary black-footed ferret populations despite careful captive management practices. Similar relatedness levels have been observed in other captive management programmes for vulnerable or endangered populations such as the Cuvier's Gazelle (*Gazella cuvieri*) (Moreno et al., 2024), Tasmanian Devils (*Sarcophilus harrisii*) (Hogg et al., 2019) and African painted dog (*Lycaon pictus*) (Miller-Butterworth et al., 2021). These studies highlighted the importance of resolving kinship between individuals prior to mating. It also demonstrates that pedigree measures are still an important tool when combined with genetic or genomic data in order to preserve the genetic diversity of captive populations (Galla et al., 2022), and ensuring the selection of optimal mates (Speak et al., 2024). This highlights the importance of utilising genomic data to inform captive matings for the preservation of genetic diversity.

Genetic diversity differs significantly between the three population groups. Our analyses indicate a higher level of inbreeding and lower heterozygosity within individuals from wild populations. Similarly, our results demonstrate that individuals from wild populations had a higher proportion of their genome in runs of homozygosity, with the greatest proportion of these being made up of longer (>1Mb) runs. Longer stretches of ROH are indicative of more recent shared ancestry (Kardos et al. 2015; Ceballos et al. 2018; Stuart et al. 2025). Both the wild populations from Mellette County and Meeteetse were subject to prolonged isolation, and therefore faced high-levels of inbreeding. Individuals captured from Mellette county populations had high susceptibility to canine distemper virus, as well as other health issues such as neoplasia (abnormal cell growth) (Carpenter et al., 1981). Individuals from the Meeteetse population also suffered from canine distemper virus, as well as sylvatic plague which drastically reduced population numbers from 1983 leading to their extinction in the wild in 1987 (Cain et al., 2011).

Individuals from the reintroduced populations were observed to have higher genetic diversity than wild populations, as well as lower inbreeding coefficients. This is the likely consequence of managed breeding and is consistent with previous studies (Cain et al., 2011; Wisely et al., 2003) that observed that individuals within reintroduced populations could sustain genetic diversity. Genetic diversity differed across captive bred populations, with individuals from Wyoming and Colorado having lower levels of inbreeding and greater genome-wide heterozygosity than those in Virginia. I also observed captive individuals having high degrees of both long (>1Mb) and short (300-500 Kb) ROH, indicative of historic inbreeding. This highlights the critical role of the captive breeding programme in both the rescue and ongoing maintenance of the black-footed ferret population. Earlier estimates suggest that the programme successfully preserved up to 86% of the genetic diversity present in the founding population (Santymire et al. 2014; Wisely et al. 2015; Garelle et al. 2013).

3.5.2 Genetic load comparison and functional consequences within the black-footed ferret

I show that weakly deleterious mutations are segregating within both captive and reintroduced populations, with many also present in historic wild individuals. Using SnpEff, I quantified both missense and loss-of-function (LoF) variants across the population and found that missense mutations were more abundant. Weakly deleterious missense mutations are expected to persist at higher frequencies than LoF mutations in bottlenecked populations due to reduced efficacy of purifying selection, making them more likely to drift to fixation (Charlesworth 2018; Kyriazis et al. 2021; Humble et al. 2023).

Despite a higher abundance, our comparison of realised (homozygous) versus masked (heterozygous) mutation load revealed a greater proportion of homozygous missense mutations than homozygous LoF variants across all populations. Notably, captive and reintroduced individuals exhibited lower numbers of homozygous deleterious mutations in both mutation categories when compared to wild individuals. These managed populations showed a higher proportion of masked deleterious load, which may reflect effective purging of more deleterious variants through inbreeding and selection, or the influence of managed breeding strategies designed to maintain genetic health (Robinson et al. 2018; Grossen et al. 2020; Stuart et al. 2025).

Analysis of gene family evolution across mustelids revealed relatively limited lineage-specific gene family turnover in the black-footed ferret genome. Although several expanded and contracted gene families were detected, only a subset were statistically significant, and the associated functional categories were primarily related to sensory perception and olfactory receptor activity. Gene family expansions and contractions within olfactory receptor pathways are commonly observed across mammalian genomes and

often reflect the highly dynamic nature of these gene families rather than lineage-specific adaptive processes (Beichman et al. 2023). Consequently, these results suggest that large-scale gene family expansion or contraction has not been a dominant driver of genomic change in the black-footed ferret relative to other mustelids.

Importantly, this analysis provided a set of high-confidence one-to-one orthologous genes across the species examined. These orthologs form the basis for downstream evolutionary analyses, enabling more precise estimation of lineage-specific selective pressures using codon-based models of sequence evolution. Thus, although gene family expansion and contraction analyses revealed limited functional divergence in the black-footed ferret, they provided a critical framework for subsequent analyses of molecular evolution.

By incorporating CODEML and SIFT into our pipeline using the aforementioned one-to-one orthologous protein sequences, I identified weakly deleterious mutations segregating within the black-footed ferret population that were not captured by conventional annotation methods such as SnpEff. These mutations, flagged by elevated d_N/d_S ratios in protein-coding genes, suggest signatures of relaxed purifying selection. I hypothesise that these variants likely emerged or increased in frequency during the severe population bottleneck experienced by the founding Wyoming population. However, their presence in a masked (heterozygous) state within the Mellette County population suggests they may have been present at low frequencies historically but remained effectively recessive prior to the bottleneck.

The identification of 58 fixed, deleterious mutations within the sampled black-footed ferret population suggest that these mutations have accumulated due to the species history of low effective population size and inbreeding. Further analysis would be needed to date the timing of these mutations, such as being able to obtain the genome sequence of individuals before the historic population bottleneck. However, the detrimental phenotypes which are being reported in at least the captive population - reduced litter size, sperm abnormalities, reduced immunity (Wisely et al. 2002; Santymire et al. 2014; Santymire et al. 2019) - aligns with findings from other species that have experienced severe demographic bottlenecks (Dobrynin et al., 2015; Dussex et al., 2021; Feng et al., 2019; Fraik et al., 2020; Hasselgren et al., 2024).

For instance, the African cheetah (*Acinonyx jubatus*) has undergone at least two major bottlenecks, resulting in a profound reduction in genetic diversity among extant populations (Dobrynin et al., 2015; O'Brien et al., 2017). A comparative study in the cheetah reported elevated d_N/d_S ratios in three of the same genes, *SPATA5*, *SLIT2*, and *NBEAL2*, that were also identified as harbouring potentially deleterious mutations in the

black-footed ferret. *SPATA5* is a spermatogenesis-associated protein and is thought to be involved with maintaining the structural integrity of mitochondria during spermatogenesis. Mutations within this gene have been associated with microcephaly, hearing loss and intellectual disability in humans (Puusepp et al., 2018; Tanaka et al., 2015). *SLIT2* encodes for a ligand receptor and is known to have a major role in axon migration and the development of the central nervous system. Mutations within this gene in humans have been linked to abnormalities in the kidney and urinary tract (Hwang et al., 2015). *NBEAL2* and recessive mutations within this gene have been linked to a rare bleeding disorder in humans and mice (Delage et al., 2023). Deficiencies of *NBEAL2* have also been linked to low CTLA-4 expression, a protein which helps to maintain immune homeostasis (Lo & Abdel-Motal, 2017). *FAM111A* and *CEP350* also exhibited increased d_N/d_S ratios in the black-footed ferret lineage, and the same signal was seen within the southern sea otter (*Enhydra lutris*), another species that has suffered substantial losses in genetic diversity (Beichman et al., 2019). *FAM111A* is a nuclear trypsin-like protease which plays a role in defense against viral infections, as well as mutations within this gene being associated with known skeletal disorders (Nie et al., 2021). *CEP350* is a tumor-suppressor gene within human melanoma, and is also found to play a major role in centriole sustainability, and ciliogenesis - the process of flagella and cilia formation (Karasu et al., 2022; Mojarad et al., 2017). Additionally, a mutation in the *ADAMTS3* gene has been associated with respiratory disease and upper airway syndrome in the Norwich terrier, likely resulting from prolonged inbreeding (Marchant et al., 2019). Therefore, these mutations may be linked to known phenotypes associated with the black-footed ferret population (immunity, sperm motility). However, further experimental validation would be necessary to verify whether the protein-coding mutations I observe in the black-footed ferret impact the gene function in the same way.

From the results where I detected elevated (d_N/d_S) across protein-coding genes of the black-footed ferret, there were 52 nonsynonymous mutations that a SIFT prediction could not be made for. This was due to SIFT not reporting an impact for 12 variants, and making unreliable predictions for 40 variants. Unreliable SIFT predictions occurred when there were multiple reported nonsynonymous mutations within a gene, which resulted in SIFT reporting a different amino acid change than was observed within the gene sequence. Multinucleotide changes within a coding sequence have been shown to inflate the signal of positive selection (Venkat et al., 2018), leading to model violations and the identification of false sites under selection. However, here I was using the model assumptions to determine whether putatively deleterious, nonsynonymous mutations could be accumulating in gene regions and not to find signatures of true positive selection that could infer adaptive changes. Therefore, I still classified these genes to have elevated

signals of d_N/d_S , potentially as a result of the population bottleneck, but cannot attribute the impact of these regions on gene function.

3.5.3 Future considerations

These findings mirror genetic patterns observed in other endangered mammal populations and further underscore the urgent need to safeguard genetic diversity in the black-footed ferret. These results highlight that genomic data, particularly information on deleterious variation, should be integrated into the design and implementation of conservation strategies. Doing so will be critical for maintaining genetic health and promoting the long-term persistence of small, at-risk populations. Previous studies have demonstrated how the accumulation of weakly deleterious mutations, due to long-term low effective population size can have detrimental impacts on phenotype. For example, increased genetic load within fertility genes leading to reduced reproductive success, linked to impacts from population bottlenecks, have been reported in the cheetah (*Acinonyx jubatus*) (Peers et al., 2025), Arctic fox (*Vulpes lagopus*) (Hasselgren et al., 2024) and Kakapo (*Strigops habroptilus*) (Dussex et al., 2021). Understanding the genomic impact of bottlenecks within the population can help conservation managers make better informed decisions to alleviate further fixation of deleterious mutations within a population.

However, this study also highlighted a significant limitation in the application of genomic tools to non-model organisms. Many widely used methods rely on the availability of a high-quality reference genome and well-annotated gene models, which are often lacking for rare or understudied species (Brandies et al., 2019; Totikov et al., 2021). In this case, the absence of these resources for two key mustelid species, *Mustela eversmannii*, the closest known sister taxon to the black-footed ferret, and *Mustela nivalis*, prevented their inclusion in the positive selection analyses. Additionally, I observed inconsistencies between the gene annotations of *Mustela putorius furo* (the domestic ferret) and *Mustela nigripes*, where certain genes were missing from the black-footed ferret annotation but could be identified when lifted over to the domestic ferret genome. Therefore demonstrating that some resources may not be at a high enough resolution to fully understand the genomic impact of inbreeding and low effective population size. These challenges highlight the urgent need for greater investment in generating and refining genomic resources for non-model and conservation-priority species.

The recovery of the black-footed ferret has become a cornerstone of modern conservation efforts, with significant resources dedicated to ensuring its survival (Sandler et al., 2021; Wisely et al., 2015). These investments have led to groundbreaking advancements including the successful cloning of ferrets in 2020 and 2023 using biobanked genetic

material. This has paved the way for the reintroduction of lost genetic diversity into both captive and reintroduced populations (Novak et al., 2024). These achievements exemplify the transformative potential of genomics in conservation (Gedman et al., 2025; McMahon et al., 2014; Theissinger et al., 2023). However, the recovery of the black-footed ferret serves as a critical reminder that genomic tools should complement, not replace, traditional conservation strategies. Conservation efforts must prioritise a holistic, ecosystem-based approach, as black-footed ferrets are entirely dependent on prairie dogs for both habitat and prey (Jachowski et al. 2011). Both species are integral to the prairie ecosystem, and for the recovery of the ferret to be truly sustainable, genomic innovations must be integrated within a broader, ecosystem-focused conservation framework, rather than solely serving as a tool for novelty (McMahon et al., 2014; Titus & Jachowski, 2021).

4.0 Identification of constrained regions in the endangered Black-footed ferret and implications of noncoding genetic variants

4.1 Abstract

The black-footed ferret (*Mustela nigripes*) represents a compelling case study of extreme genetic bottlenecks and their consequences on genome function and fitness. While prior research has focused primarily on protein-coding variation, the consequences of mutations within conserved non-coding elements (CNEs) remain underexplored in endangered species. In this chapter, a comparative genomics approach was applied, using a reference-free whole-genome alignment of 11 mammalian species to identify evolutionarily constrained regions across the mustelid genome. By integrating conservation metrics (PhyloP, GERP++) with gene annotations, functional motifs, and population-level variant data, I discovered 174,493 conserved and 67,183 accelerated elements in the black-footed ferret genome. A subset of these variants, enriched in regulatory elements proximal to genes involved in reproduction, neurodevelopment, and immunity, were predicted to alter transcription factor binding affinity. Notably, several candidate mutations occur near genes implicated in fertility-related phenotypes in model organisms, suggesting a potential molecular basis for reduced reproductive performance observed in captive populations. Our findings underscore the relevance of conserved non-coding variation in understanding inbreeding depression and fitness-related traits in conservation genomics, offering a framework for integrating evolutionary constraint and functional annotation to prioritize deleterious non-coding mutations in endangered species.

4.2 Introduction

The advancement and increasing accessibility of sequencing technologies have led to a rapid expansion in the availability of genomic data. This has greatly enhanced our understanding of the non-protein-coding fraction of the genome and its functional relevance. Since the sequencing of the human genome, research has increasingly focused on inferring functional importance from DNA sequences (R. P. Alexander et al., 2010; ENCODE Project Consortium et al., 2007; Haudry et al., 2013; Lander et al., 2001; Pollard et al., 2010). It is now well established that protein-coding genes comprise only a small portion of the genome (~1.8% of the human genome). Additionally, computational tools like SnpEff (Cingolani et al., 2012), SIFT (Ng & Henikoff, 2003), Variant Effect Predictor (VEP) (McLaren et al., 2016) and PolyPhen (Adzhubei et al., 2013) have improved our understanding of the impact of mutations within coding regions. In contrast, deciphering the functional implications of variants within non-coding regulatory elements has proven more challenging, initiatives such as the ENCODE project (ENCODE Project Consortium, 2004), Roadmap Epigenomics (Bernstein et al., 2010), Functional Annotation of Animal Genomes (FAANG) (Giuffra et al., 2019), and the Genotype-Tissue Expression (GTEx) programme (GTEx Consortium, 2013, 2020) have contributed significantly to the annotation of genomic elements using mainly high throughput experimental approaches applying chromatin immunoprecipitation sequencing, ATAC-Seq, bisulfite sequencing, RNA-Seq in model systems.

In contrast to experimental identification of putative functional noncoding elements, comparative genomics, which involves comparing genomes across related or phylogenetically distant species, has become a powerful approach for identifying regions under evolutionary constraint (Alföldi & Lindblad-Toh, 2013; Lindblad-Toh et al., 2011). This approach relies on the hypothesis that purifying selection is the main evolutionary driving force of functional sequences, and therefore conserved sequences across species are suggestive of their functional relevance. This has led to the identification of conserved non-coding elements (CNEs), many of which play essential roles in regulating gene expression and developmental pathways (Cooper et al., 2005; Lindblad-Toh et al., 2011; Siepel et al., 2005). These regions have been implicated in adaptive evolution, trait variation, and disease susceptibility. In fact, up to 95% of high-confidence GWAS hits in humans fall within non-coding regions, underscoring their importance in phenotype determination (Lunter et al., 2006). Despite this, the role of conserved regulatory regions in shaping fitness-related traits, particularly in non-model and endangered species, remains poorly understood, although available resources are steadily increasing (Christmas et al., 2023; Groß, Bortoluzzi, et al., 2020a; Kaplow et al., 2023; Zoonomia Consortium, 2020).

Evolutionary constraint is often assessed using phylogenetic approaches that compare observed sequence conservation to what would be expected under neutral evolution. Tools such as PhastCons, PhyloP, and GERP++ are commonly used for this purpose (Davydov et al., 2010; Pollard et al., 2010; Siepel et al., 2005). These methods examine the genome in fixed-size windows, typically ranging from around 12 base pairs to several hundred, in order to identify regions that have evolved more slowly than expected, an indication of active purifying selection (Bejerano et al., 2004; Lindblad-Toh et al., 2011). Among the most striking examples of conserved sequences are ultraconserved elements (UCEs), which represent an extreme subset of conserved noncoding elements and are usually defined as regions longer than 100 base pairs that show near perfect sequence identity across distantly related vertebrate species (Cummins et al., 2024). These elements are thought to play essential roles in gene regulation (Snetkova et al., 2021) or genome structure (McCole et al., 2018). The functional relevance of these constrained regions can often be inferred from their size, level of conservation, and genomic context. In contrast, accelerated regions, which evolve more rapidly than expected, offer valuable insights into positive selection and species-specific adaptations to environmental pressures. However, focusing on evolutionary constraint enables the identification of potentially harmful mutations accumulating in important noncoding elements, which is crucial for understanding genomic mutation load in populations with historic low effective population size (Bortoluzzi et al. 2020; Groß et al. 2020).

Although sequence conservation provides a robust and cell-type-independent signal of purifying selection, it offers only partial insight into biological function. Conservation highlights evolutionary constraints, but does not always reveal the precise regulatory or biochemical roles of sequences. To increase the accuracy in identifying functional non-coding sequences, comparative genomics is increasingly integrated with functional genomic data, such as RNA sequencing (RNA-seq) for detecting transcriptional activity and chromatin immunoprecipitation followed by sequencing (ChIP-seq) to identify transcription factor binding regions (Kaplou et al., 2022, 2023). Additional data on histone modifications, DNA methylation, and DNase hypersensitivity further support the identification of biologically active regions (Alemu et al., 2025; Kuderna et al., 2024). While these tools are widely used in human genetics to predict the regulatory consequences of non-coding variation, they remain severely underutilised in nonmodel species, where a lack of detailed functional datasets often limits their application (Speak et al., 2024).

Our understanding of the fitness effects of non-coding mutations remains limited compared to that of coding mutations that alter amino acids. Protein-coding changes have been studied more extensively, in part because their functional consequences are often easier to predict and interpret (Margulies et al., 2007). In contrast, non-coding mutations

can influence gene expression by disrupting promoters (Martin-Geary et al., 2023), enhancers (Claringbould & Zaugg, 2021), transcription factor binding sites (Liu et al., 2021), or miRNA-binding regions (Ghanbari et al., 2016). While their effects are harder to quantify, they can play an equally important role in regulating biological processes. Mutations within conserved non-coding elements have been strongly linked to human diseases, particularly where they disrupt gene regulation (Alföldi & Lindblad-Toh, 2013; Hawkes et al., 2025; Vitsios et al., 2021). The integration of constraint-based predictions with functional annotations, therefore, offers a powerful strategy for identifying non-coding variants that may be biologically or clinically significant. One context in which this approach may be especially useful is the study of fertility-related traits (Forutan et al., 2024). Spermatogenesis, for example, is known to be tightly regulated by non-coding RNAs and other regulatory elements (Mukherjee et al., 2014), suggesting that mutations in conserved non-coding regions could plausibly impair reproductive function. This is particularly relevant for endangered species or inbred populations, where reproductive performance is often compromised, yet the molecular causes remain poorly understood (Santymire et al. 2006; Costa et al. 2008; Terrell et al. 2016).

These methodological advances are particularly valuable for understanding the genetic consequences of severe population bottlenecks, which are common in many threatened or reintroduced species. Bottlenecks result in a substantial reduction in genetic diversity and an increase in homozygosity, thereby elevating the risk of deleterious allele fixation and reducing overall fitness (Charlesworth and Charlesworth 1987; Frankham 2005; Hedrick and Garcia-Dorado 2016). While conservation genetics has traditionally focused on protein-coding regions due to the relative ease of functional interpretation, there is increasing recognition that mutations within non-coding regulatory regions can also play a significant role in shaping phenotypic traits and fitness (Ellingford et al., 2022; Wray, 2007; Zhen & Andolfatto, 2012). Such variants may influence gene expression, developmental regulation or reproductive function, all of which are critical for the persistence of small or recovering populations. However, these regulatory variants have often gone undetected in non-model organisms due to limited genomic resources and a historical emphasis on coding sequence variation. The growing availability of high-throughput sequencing and comparative genomic tools now enables the detection of both coding and non-coding mutations, allowing for a more comprehensive assessment of mutational burden and its potential impact on population recovery (Christmas et al., 2023; Kaplow et al., 2023; Kronenberg et al., 2018). It also facilitates the identification of structural variants, which are increasingly recognised as important contributors to gene regulation and phenotypic variation (Kronenberg et al., 2018; Wold et al., 2021).

The black-footed ferret (*Mustela nigripes*) has experienced a severe genetic bottleneck that has profoundly impacted its genetic health. Captive breeding efforts have documented clear signs of inbreeding depression, including reduced heterozygosity and lowered reproductive success in both males and females (Santymire et al. 2014; Santymire et al. 2019). This bottleneck has had long-lasting effects on the species' genetic health. Captive breeding has revealed signs of inbreeding depression, including reduced heterozygosity and lowered reproductive success in both sexes. Santymire et al. (2019) reported that increased homozygosity correlated with reduced sperm concentration and motility, potentially contributing to decreased pregnancy rates. Similar associations between inbreeding and reduced fertility have been documented in other large carnivores, such as the Florida panther (Barone et al., 1994) and African lions (Munson et al., 1996). While these patterns are well described, the molecular mechanisms underlying reduced fecundity remain poorly understood.

In this chapter, I aim to investigate whether the demographic history of the black-footed ferret has resulted in the accumulation of potentially deleterious mutations within evolutionarily constrained non-coding regions of the genome. By using a reference-free whole-genome alignment across mustelids and phylogenetic methods to infer evolutionary constraint, I aim to identify conserved and accelerated elements across the black-footed ferret genome and characterise their distribution relative to genes and regulatory regions. I predict that these findings will extend on my previous analyses of coding variation by demonstrating that the genomic consequences of the black-footed ferret bottleneck also affect the regulatory genome. More broadly, I aim to provide evidence that this work highlights the importance of incorporating non-coding functional regions into assessments of genetic load and can demonstrate how comparative phylogenomic approaches can be used to identify candidate non-coding variants of conservation concern in endangered species.

4.3 Methods

4.3.1 Curation of genomes

I utilised publicly available genomes for eleven mammalian species. I selected species from within the family Mustelidae that had annotated genomes: European polecat (*Mustela putorius*) (GCA_902207235.1) (Etherington et al., 2020), domestic ferret (*Mustela putorius furo*) (GCA_920103865.1) (Etherington et al., 2022; Peng et al., 2014), Steppe polecat (*Mustela eversmanii*) (GCA_963422785.1), black-footed ferret (*Mustela nigripes*) (GCF_022355385.1) (Kliver et al., 2023), stoat (*Mustela erminea*) (GCA_009829155.1), weasel (*Mustela nivalis*) (Miranda et al., 2021), stone marten (*Martes foina*) (GCA_040938555.1) (Tomarovsky et al., 2025), and Eurasian otter (*Lutra lutra*) (GCA_902655055) (Mead et al., 2020). Additionally, I included three mammalian outgroup

species: human (*Homo sapiens*) (GCA_000001405.15) (Schneider et al., 2017), domestic cat (*Felis catus*) (GCF_000181335.3) (Buckley et al., 2020), and domestic dog (*Canis lupus familiaris*) (GCF_011100685.1) (Wang et al. 2021).

4.3.2 Progressive Cactus

Progressive cactus v2.20 (Armstrong et al., 2020) was used to generate a reference-free alignment consisting of 11 mammalian genomes. To achieve this, the genomes were first soft-masked using RepeatMasker version 4.0.8 (<http://www.repeatmasker.org>) (Tarailo-Graovac & Chen, 2009) using the ‘mammals’ repeat library to mask interspersed repeats specific to mammalian species. Softmasking genomes is a requirement of Cactus for input genomes to avoid false alignments caused by repetitive regions, which can greatly increase the computational burden, alignment quality, and run-time of the whole-genome alignment.

Cactus requires a guide tree to inform the phylogenetic relationships among the genomes being aligned. Initially, I ran Progressive Cactus using a topology inferred from the literature (Koepfli et al., 2008), without specifying branch lengths (Figure 4.1).

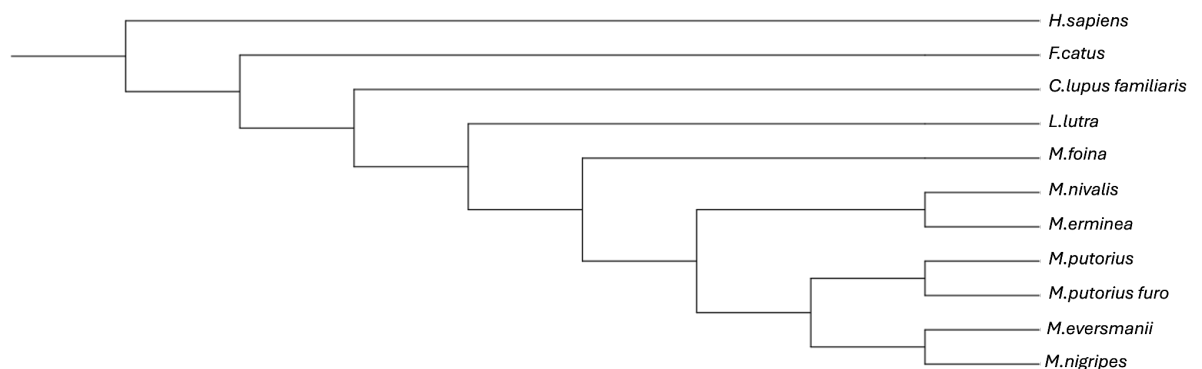


Figure 4.1: Phylogenetic tree of 11 mammalian genomes used to guide multi-genome alignments with Progressive Cactus, including eight mustelid species and three outgroup mammals. Mustelid genomes included: European polecat (*Mustela putorius*, GCA_902207235.1), domestic ferret (*Mustela putorius furo*, GCA_920103865.1), Steppe polecat (*Mustela eversmanii*, GCA_963422785.1), black-footed ferret (*Mustela nigripes*, GCF_022355385.1), stoat (*Mustela erminea*, GCA_009829155.1), weasel (*Mustela nivalis*), stone marten (*Martes foina*, GCA_040938555.1), and Eurasian otter (*Lutra lutra*, GCA_902655055). Outgroup genomes included human (*Homo sapiens*, GCA_000001405.15), domestic cat (*Felis catus*, GCF_000181335.3), and domestic dog (*Canis lupus familiaris*, GCF_011100685.1). No branch lengths were specified and instead was drawn from the literature (Koepfli et al. 2008) and the tree was visualised using iTOL.

Cactus uses this guide tree to divide the alignment task into smaller, phylogenetically-informed subproblems. When branch lengths are not provided, Cactus

assumes all branches have a length of 1, which limits its ability to optimise the partitioning of the alignment problem. Although Cactus can infer branch lengths during alignment, this process is computationally intensive. To improve efficiency and accuracy, I used the output of the initial alignment to estimate a branch-length-informed guide tree based on ancestral repeats (described below), using the Phylogenetic Analysis with Space/Time (PHAST) v1.6 (Pollard et al., 2010; Siepel & Haussler, 2004) under a neutral model. This refined tree was then used as input for a second, final iteration of the alignment. Alignment coverage for each species was calculated and assessed using `halAlignmentDepth` v2.2 (Hickey et al., 2013), with the domestic ferret and the black-footed ferret as target species (Supplementary material S4.1).

4.3.3 Generation of a neutral model

PHAST v1.6 (Pollard et al., 2010; Siepel & Haussler, 2004) was used to determine selection rates across the alignment. Firstly, an alignment in the MAF format was extracted specifying the domestic ferret as the reference, using `hal2maf` v2.2 with the `--noAncestors` to remove ancestral inferred sequences, `--noDups` to avoid duplicate sequences being reported, and `--onlyOrthologs`. The MAF was then post-processed using `maf_stream merge_dups consensus` (https://github.com/joelarmstrong/maf_stream) to replace any duplicated entries with a single entry.

To detect regions of the genome that were evolving slower or faster than expected, I first had to generate a neutral model of evolution across the genome. This model was trained using ancestral repeats that were previously demonstrated to evolve nearly neutrally (Ellegren et al., 2003; Margulies et al., 2007). To generate the neutral model, I first extracted the most basal ancestral genome from the Cactus generated whole-genome alignment. RepeatMasker version 4.0.8 (<http://www.repeatmasker.org>) (Tarailo-Graovac & Chen, 2009) was run on the ancestral genome to extract repetitive regions using the ‘mammal’ database. Low complexity repeats and tRNAs were filtered out of these results to ensure annotated regions were truly neutral. I then carried out a further step to ensure that these repeat regions did not overlap with genic regions. Using `halLiftover` v2.2, the ancestral repeat regions were lifted over to the domestic ferret annotation. Non-coding repeat regions were then inferred using *bedtools intersect* to remove any regions that intersect with genes based on the domestic ferret gene model. The resulting repeat regions were used as input to `phyloFit` v1.6 (Siepel & Haussler, 2004) (`--subst-mod REV --EM`) to generate the neutral model.

4.3.4 Calculation of conserved and accelerated elements

PhyloP from PHAST (version 1.6) was used to identify highly conserved loci. First, the generated MAF file was split into individual chromosomes for the black-footed ferret (HiC

Scaffolds 1–19) and into 297 super-scaffolds for the domestic ferret, which together represent 97.95% of its genome. Each genomic region was then divided into non-overlapping 25-bp windows to improve the statistical power of the constraint analysis (Lindblad-Toh et al., 2011; Secomandi et al., 2023).

PhyloP was run separately on each 25-bp window, using the neutral model and the Likelihood Ratio Test (LRT) in CONACC mode. This analysis produced PhyloP scores, where positive values indicate conservation and negative values indicate acceleration (faster-than-neutral evolution). Windows shorter than 25-bp were excluded from further analysis. To control for multiple hypothesis testing, p-values were adjusted using the Benjamini-Hochberg (BH) procedure and retained only windows with an adjusted p-value < 0.05. Continuous regions were merged using BEDTools version 2.31 (Quinlan and Hall, 2010), with the bedtools *merge* parameter and with a maximum gap of 10 bp to define putative conserved or accelerated elements (CEs and AEs, respectively). To identify conserved elements (CEs), an additional filter was applied, keeping only regions with PhyloP scores > 2 and a minimum length of 100 nucleotides.

To test whether I could identify constraint at a per-base or smaller window size than 25-bp, I further calculated constrained elements using Genomic Evolutionary Rate Profiling (GERP++) version 2.1 (Davydov et al., 2010; Pollard et al., 2010). As input, GERP++ requires a phylogenetic tree, which was generated using the neutral substitution model output from phyloFit. The gerpcol function was then applied to each split MAF file to compute rejected substitution (RS) scores for each alignment column. These RS scores quantify the degree of evolutionary constraint by comparing the estimated substitution rate at each position to the neutral rate. Positive RS scores indicate evolutionary constraint, with scores above 1 signifying highly constrained regions, while scores below 1 are considered putatively neutral. To eliminate bias to the focal species, the *-j* parameter in gerpcol was used to project out the reference genome. Finally, to characterise conserved elements across the black-footed ferret genome, the gerpelem function in GERP++ was used. This step identifies contiguous constrained elements based on the RS scores output by the gerpcol function. This step was also run in order to check the concordance of conserved elements identified by both GERP and PhyloP.

4.3.5 Mapping of constraint scores to annotated regions

Due to the limited availability of a high-quality gene annotation for the black-footed ferret, annotations from the domestic ferret genome were transferred to the black-footed ferret assembly using Liftoff version 1.5 (Shumate & Salzberg, 2021). Liftoff maps genes between assemblies or species by aligning sequences at the transcript and exon level, accounting for gene structure and preserving synteny.

Liftoff was applied using the domestic ferret genome and its annotation file, along with the target genome assembly (the black-footed ferret). It was executed with default parameters, retaining only successfully mapped genes. From the resulting transferred gene annotation, core genomic features were extracted, including genes, coding sequences (CDS), exons, untranslated regions (UTRs), introns, as well as annotated non-coding RNAs and pseudogenes. Introns were identified by subtracting exons from gene regions, while intergenic regions were defined as non-overlapping sequences located outside all known gene annotations.

Intergenic regions were further classified by distance from the nearest gene. Regions within 5 kilobases (Kb) of a gene boundary were defined as proximal intergenic, while those beyond 5 Kb were designated deep intergenic. This distinction was used to reflect varying likelihoods of regulatory influence. Promoter regions were defined based on the predicted transcription start site (TSS) location, up to 5 Kb upstream. Each conserved or accelerated region was intersected with annotated genomic features from both the domestic ferret and the Liftoff-projected gene model in the black-footed ferret using bedtools *intersect*. For each feature, the number of base pairs overlapping the conserved or accelerated region was reported, allowing elements to be associated with multiple annotations if applicable.

4.3.6 Intersection of population variants within constrained regions

To estimate mutation load within conserved non-coding regions across the black-footed ferret population, genome-wide SNP data from 54 individuals were generated. Of these, six were from the last wild populations (four from Wyoming and two from South Dakota), 26 were captive individuals, and 22 were from the reintroduced population (Supplementary material S3.1). Variant data were generated by mapping high-coverage (Supplementary table 3.01) raw reads against the black-footed ferret reference genome (GCA_022355385.1) (Kliver et al. 2023) using the BWA software version 0.7.1 and the BWA-mem algorithm (Li and Durbin 2009). The variant calling process was performed using the Genomic Analysis Toolkit version 4.2.5 (GATK) pipeline (McKenna et al. 2010). Individual samples were merged into a single file using GATK CombineGVCF and joint genotype calling was performed using GATK GenotypeGVCF. The following hard filters were applied to single nucleotide polymorphisms (SNPs), to ensure the accuracy of the variant calling results: "QD < 2.0", "QUAL < 30.0", "QUAL30", "SOR > 3.0", "FS > 60.0", "MQ < 40.0", "MQRankSum < -12.5", "ReadPosRankSum < -8.0" (Figueiro et al., *in preparation*). For downstream analysis, BCFtools filter v1.15.1 (Danecek et al. 2021) was used to extract biallelic SNPs with a quality score (QUAL) of ≥ 20 and a minor allele count of $\geq$ three (AC3). Bedtools *-intersect* was employed to identify variants located within annotated, constrained non-coding regions, as defined by PhyloP scores. Polarized

variants within conserved non-coding regions were determined using a comparative approach. The whole-genome alignment was used to identify nucleotide identities at positions classified within non-coding regions of the black-footed ferret, based on comparisons with four sister species (domestic ferret, European polecat, steppe polecat, and weasel) using *halLiftover* v2.2 (Hickey et al., 2013). These nucleotide identities were then intersected with the called variants to identify derived alleles in the black-footed ferret population. Only SNPs for which either the black-footed ferret reference or alternative allele matched the ancestral allele and was confirmed across three of the sister species were considered. A derived allele frequency cutoff of 10% was used to distinguish rare from common SNPs (Groß, Bortoluzzi, et al., 2020b).

4.3.7 Motif scanning and impact of variants on position weight matrices

To identify potential regulatory regions across the black-footed ferret genome, I obtained transcription factor binding sites (TFSBs) of known vertebrate transcription factors from the JASPAR vertebrate core motif (2022 release). Using the previously defined promoter regions, I used FIMO, part of the MEME suite, version 5.1.0 (Grant et al. 2011) to scan for motifs 5 Kb upstream of a gene's TSS using the default p-value ($1e-4$) for JASPAR. I then filtered for statistically significant motifs using a q-value cut-off value (FDR) < 0.05 . I then intersected population-level variants within conserved non-coding regions with our statistically significant motifs using *bedtools intersect*. If there were duplicate motifs called at each region, I used the q-value to identify the most statistically significant motif.

I then used *bedtools intersect* to identify black-footed ferret population variants within constrained regions that fell within the annotated motifs. Transcription factors recognise binding sites by short (~10-20 nucleotides long) sequence motifs which can be described by position-weight matrices (PWMs) (Ambrosini et al., 2020; Castro-Mondragon et al., 2022; Stormo, 2013). The PWM represents the probability of each base occurring at each given position of the motif, which are then summed to compute the total binding score for that motif. Variants within these regions, which are essential for gene expression and regulation, can alter the binding affinity of transcription factors and therefore could alter the regulation and expression of a gene (Lee & Young, 2013). I therefore assessed the impact of population variants found within motifs in CNEs by calculating the change in the position weight matrix of the total motif score for the given transcription factor binding site by imputing the variant to the motif and comparing the alternative score. Any score that was found to alter the overall motif score over a threshold of 0.5 was determined as disruptive in comparison to the reference sequence to the binding affinity of the transcription factor, where a score of -1 to -0.5 would reduce the binding affinity and +0.5 to +1 would enhance the binding of the transcription factor (Steinhaus et al., 2022). Similar thresholds have previously been used to indicate biologically meaningful effects on

transcription factor binding (Steinhaus et al., 2022; Weirauch et al., 2013). This threshold reflects a moderate-to-strong change in binding likelihood based on position weight matrix models, but it does not account for broader chromatin context or cooperative binding effects. I further investigated the potential impact of variants within non-coding regions through literature and database searches - including the Mouse Genome Informatics (MGI) database (Eppig 2017).

4.4 Results

4.4.1 Progressive Cactus

To identify regions under evolutionary constraint in the black-footed ferret genome, I constructed a mustelid-specific, reference-free whole-genome multiple alignment using Cactus. This alignment included eight mustelid species along with three mammalian outgroups. The extent and quality of alignment coverage against the black-footed ferret reference genome varied significantly depending on the phylogenetic distance of each species. Specifically, species more closely related to the black-footed ferret (*M. eversmanii*, *M. putorius*, *M. putorius furo*) exhibited higher alignment coverage and greater sequence similarity, resulting in more precise and comprehensive alignments. In contrast, more distantly related species showed reduced alignment resolution and lower coverage (*H. sapiens*, *F. catus*, *C. lupus familiaris*) (Figure 4.2, Figure 4.3).

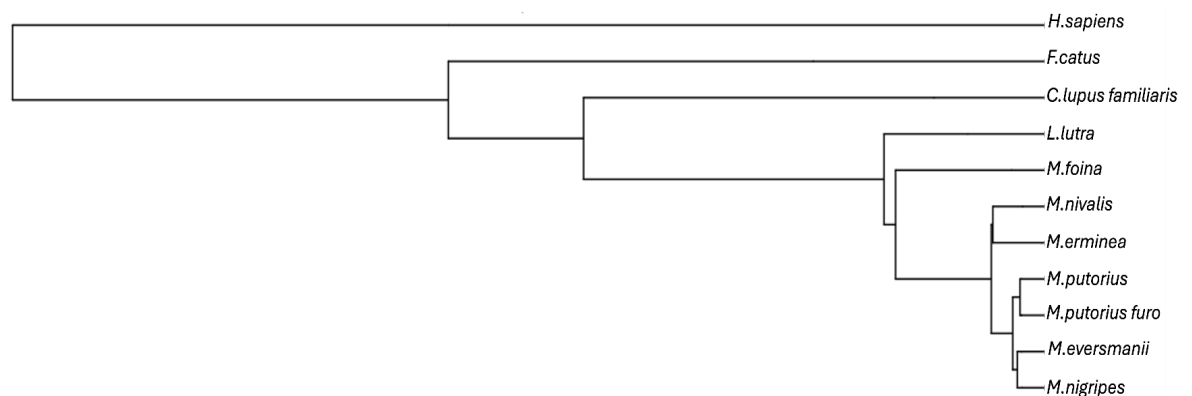


Figure 4.2: Phylogenetic tree of 11 mammalian genomes used to guide multi-genome alignments with Progressive Cactus, including eight mustelid species and three outgroup mammals. Mustelid genomes included: European polecat (*Mustela putorius*, GCA_902207235.1), domestic ferret (*Mustela putorius furo*, GCA_920103865.1), Steppe polecat (*Mustela eversmanii*, GCA_963422785.1), black-footed ferret (*Mustela nigripes*, GCF_022355385.1), stoat (*Mustela erminea*, GCA_009829155.1), weasel (*Mustela nivalis*), stone marten (*Martes foina*, GCA_040938555.1), and Eurasian otter (*Lutra lutra*, GCA_902655055). Outgroup genomes included human (*Homo sapiens*, GCA_000001405.15), domestic cat (*Felis catus*, GCF_000181335.3), and domestic dog

(*Canis lupus familiaris*, GCF_011100685.1). Branch lengths were estimated from a neutral model using PhyloFit (PHAST suite) based on aligned genome sequences, and the tree was visualised using iTOL.

Alignment statistics were calculated using HAL alignment depth for the two Progressive Cactus runs (with and without specified branch lengths). The coverage depth showed a higher resolution in the run with specified branch lengths for species more distantly related to the black-footed ferret genome (Figure 4.3). This improvement was particularly noticeable in the alignment coverage of *Martes foina*, *Lutra lutra*, *Felis catus* and *Homo sapiens*, demonstrating that specifying branch lengths enhances the quality of alignments for more evolutionarily distant species.

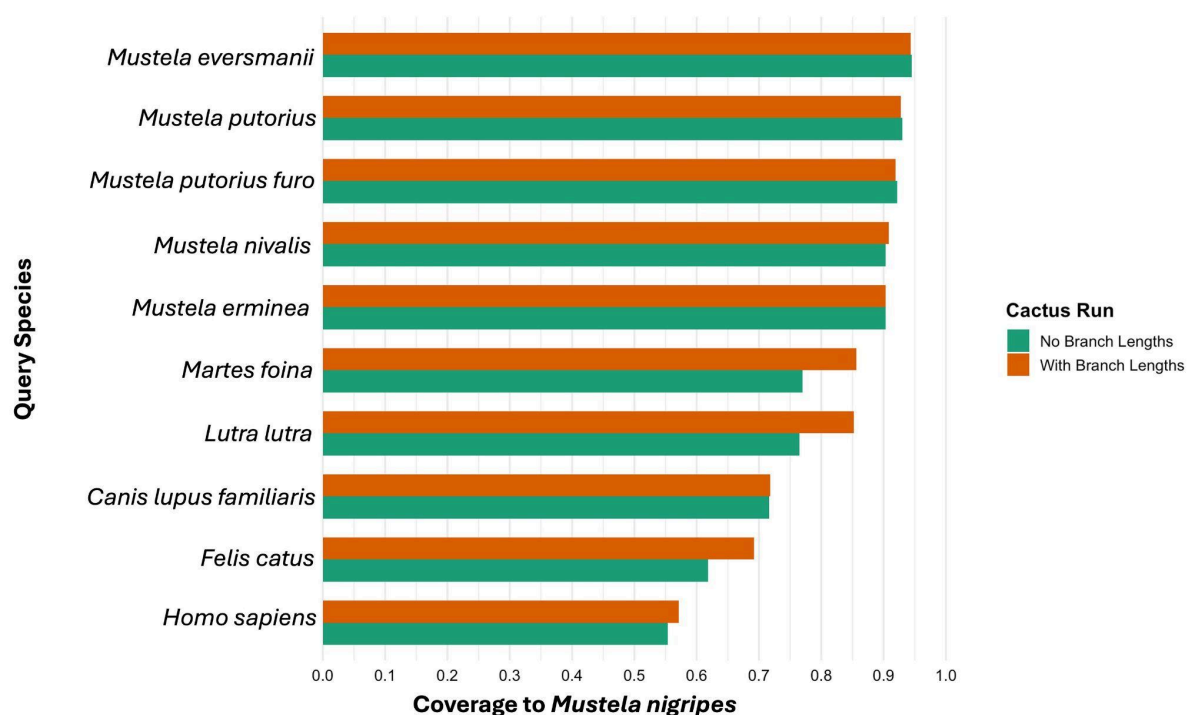


Figure 4.3: Comparison of multi-genome alignment runs using Progressive Cactus, showing the percentage of each mammalian genome aligned to the black-footed ferret (*Mustela nigripes*) reference genome. Two alignment runs are presented: one without specified branch lengths in the input phylogeny (green) and one with specified branch lengths (orange). Alignment coverage was calculated using HAL alignment depth, revealing that specifying branch lengths improves alignment resolution for more distantly related species. This improvement is particularly evident for *Martes foina*, *Lutra lutra*, *Felis catus*, and *Homo sapiens*, highlighting the benefit of incorporating phylogenetic branch lengths in multi-genome alignments.

4.4.2 Calculation of conserved elements

Using an ancestral repeat neutral model and the Cactus alignment in PhyloP, I found that 0.40% of genome-wide black-footed ferret bases are accelerated and 1.23% are conserved after false discovery rate (FDR) correction (Figure 4.4).

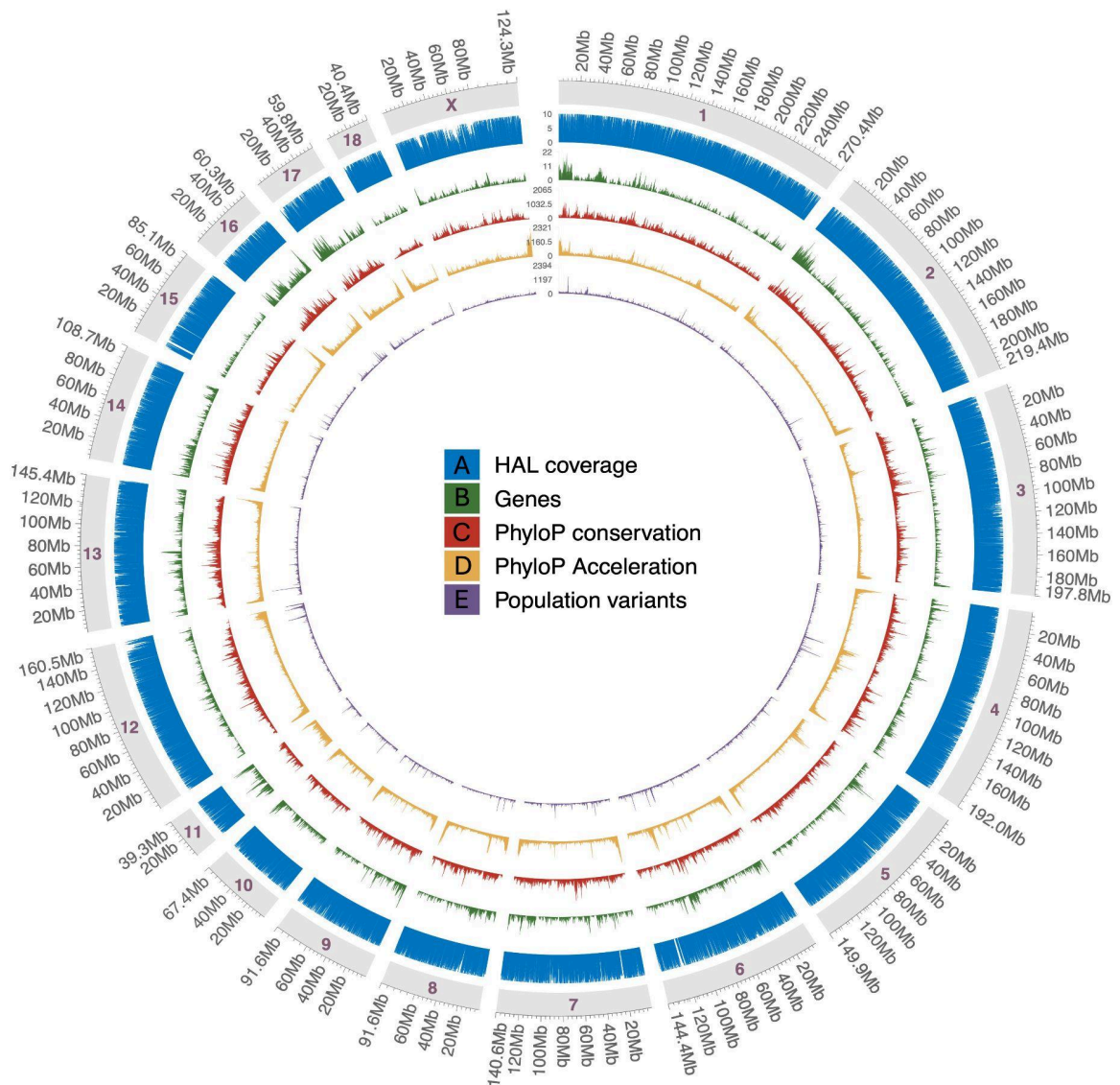


Figure 4.4: Circos plot depicting a circular representation of the black-footed ferret chromosomes. All data are plotted using 200 Kb windows (a) alignment coverage of the ten other mammalian genomes to the black-footed ferret in the the Cactus HAL alignment, (b) gene density, (c) PhyloP conserved site density, (d) PhyloP accelerated site density, (e) genomic locations of population variants from 54 black-footed ferret individuals.

Locations of conserved and accelerated elements were annotated based on the domestic ferret gene model. From the data 174,493 (consecutive 25-bp windows totalling more than 100 bp) conserved elements and 67,183 accelerated elements were identified across the black-footed ferret genome. Of these 174,493 elements, 45% fell into intergenic regions

(79,122) and 21% (37,692) were found within intronic regions. The remaining 33% of these elements were found within genic regions, where 50,038 elements were found in CDS, gene or exonic regions, 3,013 in five prime 'UTR regions and 4,625 in three prime 'UTR regions. Lengths of conserved elements found across the black-footed ferret genome varied from 100bp to 3900bp. Constraint scores in 25-bp windows across the genome ranged between 2.48 - 8.46, with accelerated scores in 25-bp windows ranging from -20 - 0 (Figure 4.5). I observe an undulating pattern within constraint scores across the black-footed ferret genome, which I hypothesise is primarily driven by the window-based approach used to summarise conservation. Because conservation scores were aggregated across relatively small 25-bp windows rather than calculated on a per-base basis, the number and magnitude of conserved sites within each window can produce discretised score distributions. When neighbouring windows were merged to create larger 50 bp windows (Figure S4.1), the resulting histogram displayed a smoother distribution, supporting the interpretation that this pattern reflects a methodological artefact of window size rather than a biological signal.

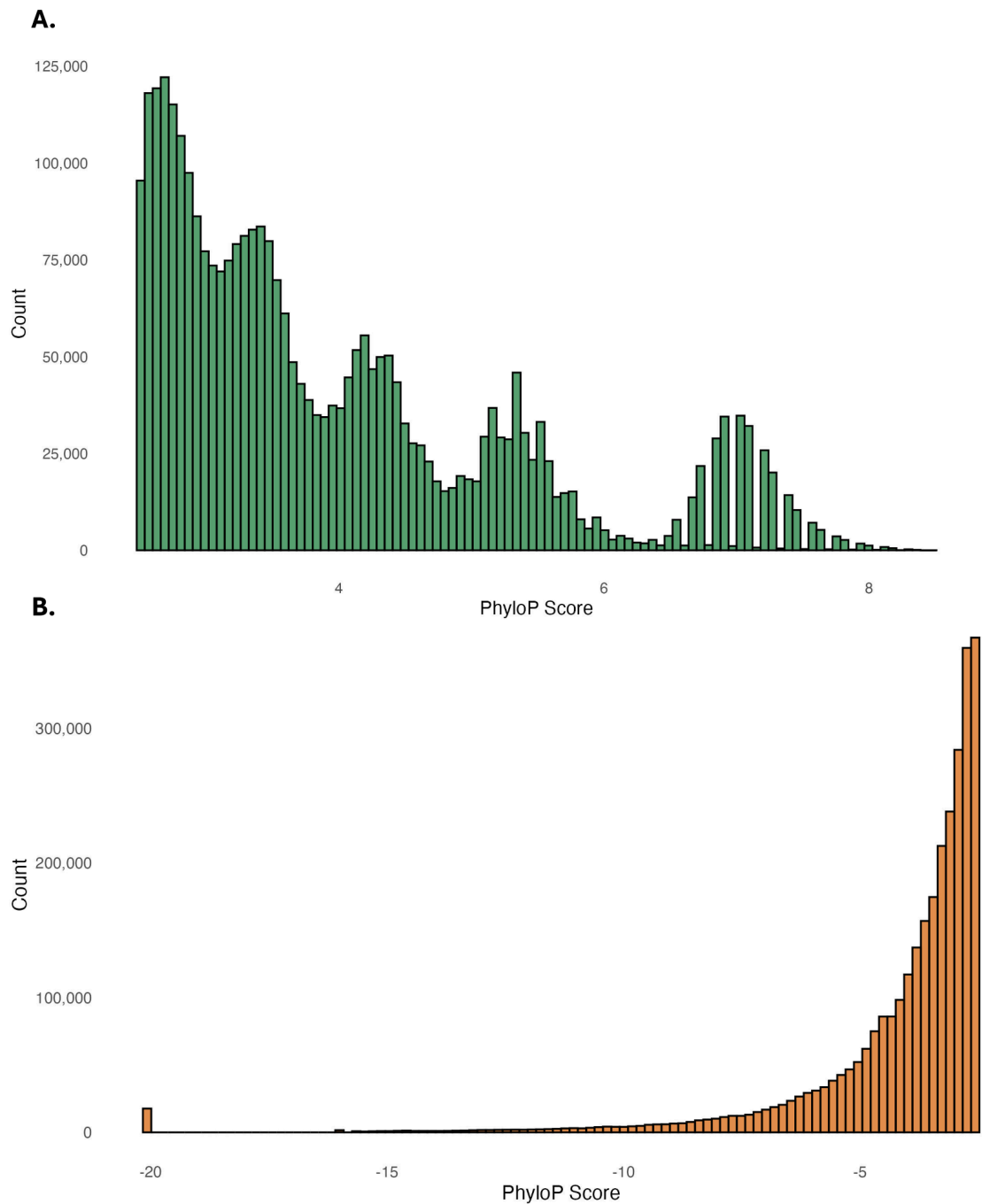


Figure 4.5: Frequency distribution of calculated conserved (A) and accelerated (B) PhyloP scores in 25-bp windows across the black-footed ferret genome. Panel (A) represents the frequency distribution of conserved PhyloP scores >2 (2.48 - 8.46), which indicate genomic regions under evolutionary constraint, while panel (B) shows the distribution of accelerated PhyloP scores (20 - 0), reflecting regions potentially under positive selection or rapid evolution.

Across the accelerated and conserved bases detected by PhyloP, about 29% and 49%, respectively, fall within genes, while only ~0.9% and ~19% overlapped with CDSs, similar to previous studies (Figure 4.6). Intergenic and intronic regions made up a large proportion of the accelerated and conserved bases, with 71% of accelerated bases falling

within intergenic regions and 27% within intronic regions. 47% of conserved bases fell within intergenic regions and 25% fell within intronic regions.

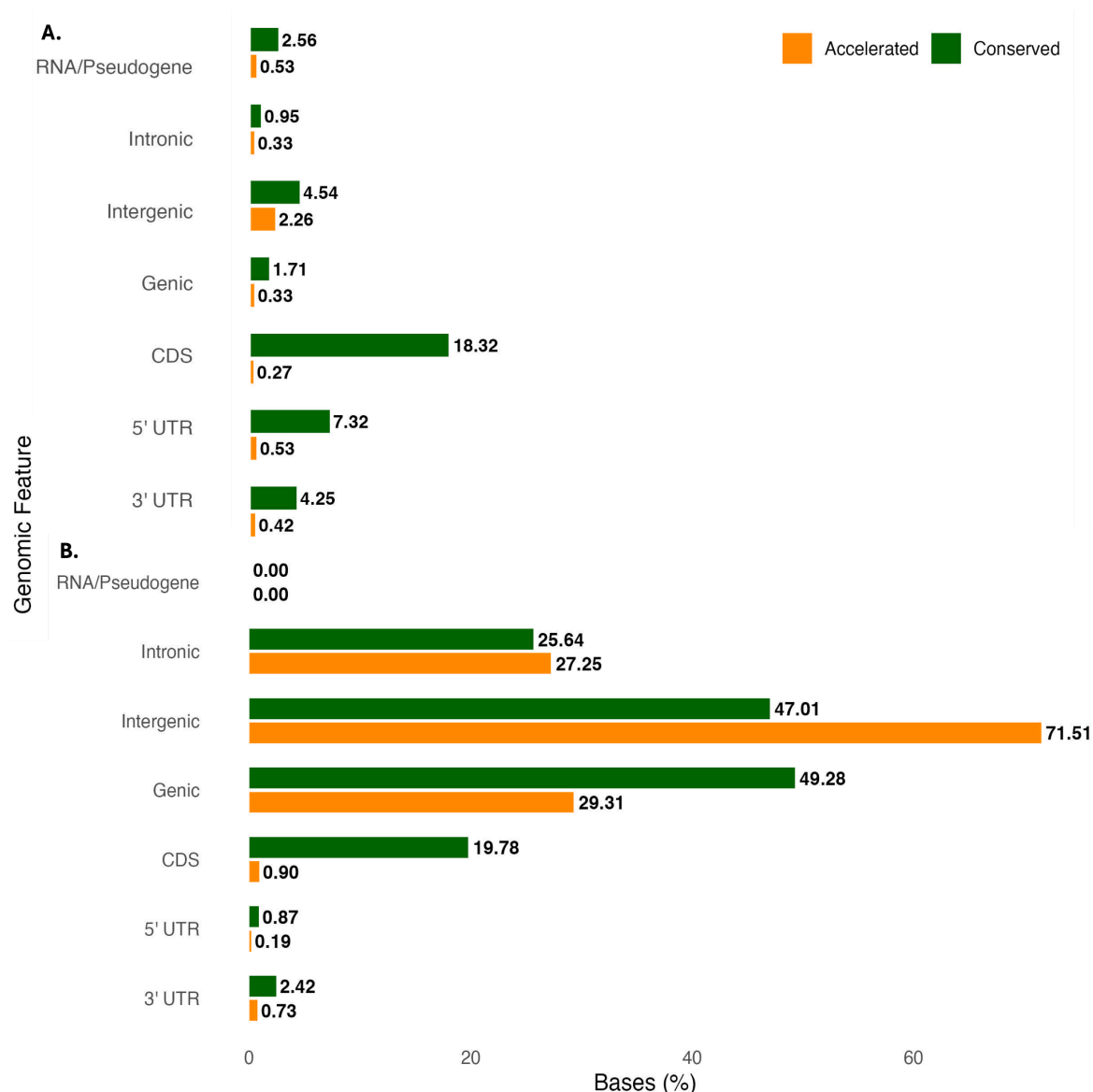


Figure 4.6: Comparison of PhyloP-identified conserved and accelerated bases between the black-footed ferret and domestic ferret genomes. PhyloP conservation and acceleration scores were calculated independently in each species. To allow for direct comparison, the domestic ferret gene model and the corresponding conserved/accelerated regions were lifted over to the black-footed ferret genome coordinates. (A) Fraction of genomic features (e.g., coding sequences, introns, intergenic regions) covered by conserved or accelerated bases. (B) Percentage of conserved or accelerated PhyloP bases overlapping each genomic feature. This approach enables assessment of the overlap between species-specific and shared conserved or accelerated elements.

To assess the distribution of conserved and accelerated elements in the black-footed ferret in comparison to other mustelids, PhyloP scores were also calculated using the domestic ferret genome as a reference. I found that 0.50% of bases in the domestic ferret genome are accelerated and 1.04% are conserved, following false discovery rate (FDR) correction. In total, I identified 148,367 conserved elements and 86,229 accelerated elements across the domestic ferret genome.

The domestic ferret gene model and its associated conserved and accelerated regions were lifted over to the black-footed ferret genome to allow direct comparison between species. This approach enabled the assessment of both shared and species-specific evolutionary patterns within a consistent genomic framework. 65% of conserved elements identified in the domestic ferret overlapped with CEs also identified in the black-footed ferret. In the reciprocal comparison, 59% of CEs called in the black-footed ferret were also conserved in the domestic ferret. Among the conserved elements present in the domestic ferret but absent from the black-footed ferret, 52% were located within genic regions, while 48% were found in intergenic regions. Similarly, of the conserved elements unique to the black-footed ferret, 54% resided within genic regions and 46% in intergenic regions. This shows that conserved elements unique to each species are found both within genes and in the non-coding regions between genes, suggesting that species-specific functional constraints affect multiple types of genomic regions.

A gene ontology (GO) analysis was carried out on the top 5% of genes with the highest number of PhyloP-conserved sites within gene regions (Supplementary material S4.3), which revealed enrichment for genes involved in the regulation of RNA and nucleobase-containing compound metabolic processes, transcriptional regulation, protein binding, and developmental processes. In contrast, the GO analysis of the top 5% of genes with the highest number of PhyloP-accelerated sites (Supplementary material S4.2) showed enrichment for genes associated with ion and small molecule binding, protein interactions, and neural development, including the regulation of neuron migration and cell projection organisation.

4.4.3 GERP scores

GERP ++ was used to generate rejected substitution (RS) scores on each column of the multi-genome alignment to find bases under constraint. A positive RS score is indicative of constraint, with a score >2 indicative of a highly constrained region. However, I did not find any single base region with a score >2. The average RS score across chromosome-aligned scaffolds ranged from 0.56 - 0.058.

Gerpelem was then utilised to detect constrained elements across both the black-footed ferret and domestic ferret genomes. In the black-footed ferret genome, I identified a total of

1,661,089 constrained elements, with an average length of 69 nucleotides and average RS scores ranging from 24.5 to 34.8. Similarly, in the domestic ferret genome, I found 1,634,664 constrained elements, which had an average length of 63.5 nucleotides and average RS scores ranging from 9.5 to 91.8.

I assessed the overlap of called constrained regions from PhyloP and gerpelem, and found that 172,710 PhyloP conserved regions overlapped with 340,126 called gerp elements. Of those 1783 PhyloP regions that did not overlap with gerp elements, 54% (975) were found within intergenic regions, and 18% (321) were found within intronic regions. Based on my results, I found that the regions called by PhyloP as under constraint were more conservative than those identified by Gerpelem and could be considered as a higher confidence subset. Therefore, subsequent analysis was performed using the PhyloP dataset.

4.4.4 Identification of variants within constrained elements

After applying quality filtering, the population variant dataset comprised 1,927,827 bi-allelic SNPs. These variants were intersected with PhyloP-defined 25-bp window, constrained regions, revealing that 5,370 SNPs were located within these regions. In addition, 2,611 variants overlapped with UCEs. Among those overlapping with UCEs, 1,475 were found within intergenic regions, 694 within introns, and 442 within genic regions, including 3' UTRs, 5' UTRs, and coding sequences (CDS).

To assess whether conservation within conserved non-coding elements (CNEs) in the black-footed ferret genome is driven by selective constraint or local variation in mutation rate, I compared the derived allele frequency (DAF) spectra of CNEs to those of putatively neutral regions (intergenic and intronic non-CNEs). Derived alleles were inferred using a multispecies whole-genome alignment of closely related mustelids, including the domestic ferret, European polecat, steppe polecat, and weasel. Sites where the black-footed ferret reference allele matched at least three of the sister species were assigned as ancestral.

Across the black-footed ferret genome, I observed a significantly lower SNP density in CNEs compared to non-CNEs, suggesting reduced levels of new variants arising within constrained regions (Figure 4.7). Moreover, CNEs were enriched for rare variants: 78% of derived alleles in CNEs had a frequency of less than 10% in the population ($n = 1455$), in contrast to non-CNE regions, where 59% of derived alleles were rare and a greater proportion were common variants (40% vs 21% in CNEs). These patterns are consistent with the action of purifying selection, which is expected to limit the rise in frequency of deleterious mutations in functionally important regions.

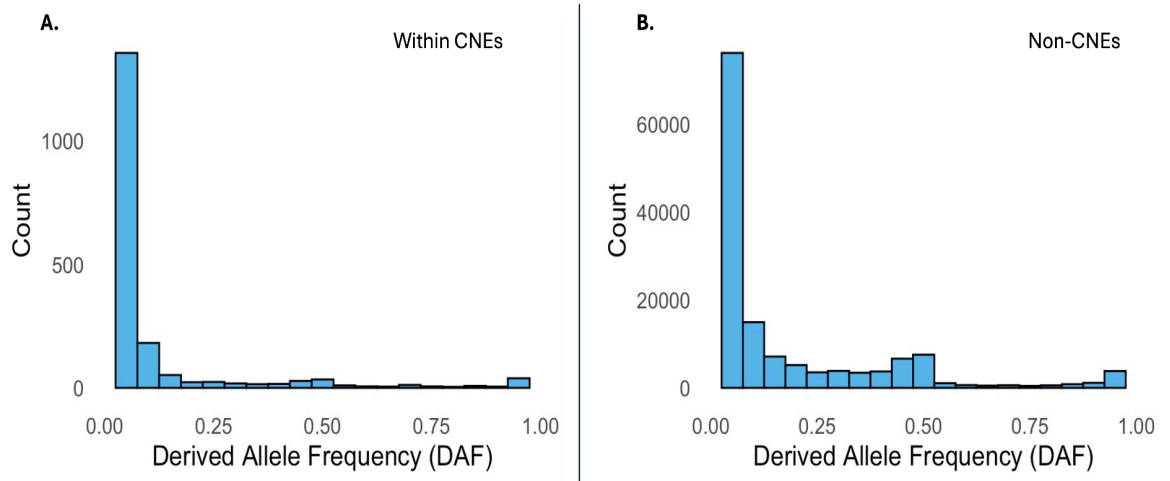


Figure 4.7: Distribution of Derived Allele Frequencies (DAF) across black-footed ferret population variants found within A) conserved non-coding elements (intronic and intergenic regions) (n = 1,844) and B) non-conserved intergenic and intronic regions (n = 141,789). Derived allele frequencies range from 0 to 1 and are binned in intervals of 0.05. Variants are classified as rare (DAF < 0.1) or common (DAF ≥ 0.1).

2.3.6 Motif annotation and intersection with population variants.

As input to FIMO for transcription factor binding site motif scanning, input 19,656 sequences that correspond to 5 Kb upstream of annotated gene TSS. After filtering for significance at a q-value of <0.05, there were 7,369,476 motif sequences found.

I intersected our conserved variants (intergenic, intronic, 3'UTR and 5'UTR) with the annotated motif regions and found 158 variants that overlapped within these regions. After calculating the change in PWM of the motif by the inferred allele change at the variant position, I found 38 (Figure 4.8 & 4.9, Supplementary material S4.4) variants that were predicted to be disruptive to the binding affinity of the transcription factor based on a PWM score change >0.5. These variants fell within or closest to 30 different protein-coding genes. 17 of these variants fell within the intronic regions of genes, 10 fell upstream (<5 Kb) of TSS sites of closest genes, 10 within 5' UTR regions of a gene and 1 within a 3' UTR region of a gene. I assessed whether the population variant was the derived allele by aligning the variant position to the position within the multi-genome Cactus alignment. I found that 24 positions were conserved across all mammals, with six differing in one or more mammals, but not across those belonging to mustelidae and eight conserved across seven or more mammals, but sites were missing in the genome of one or more mustelids.

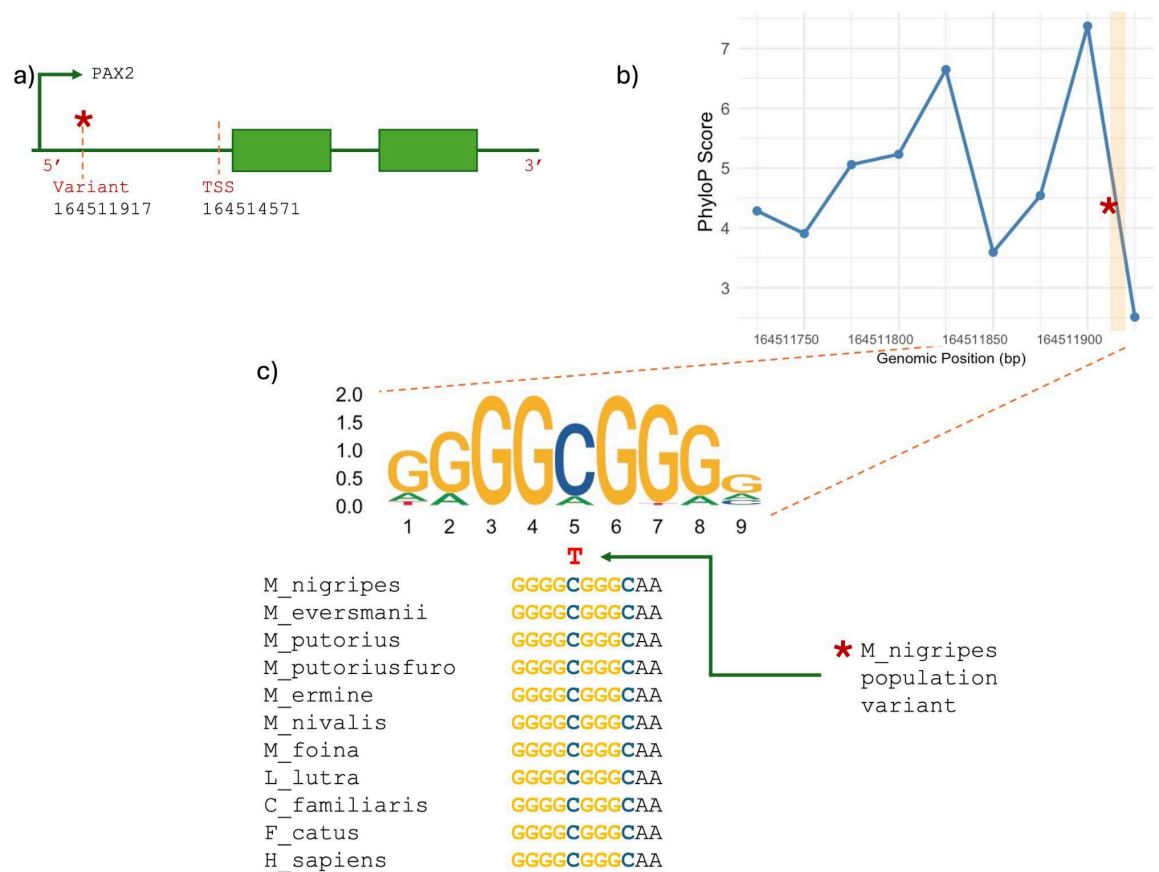


Figure 4.8: Visualisation of the variant falling within a non-coding region, classified as highly constrained by PhyloP and within an SP2 motif as classified by FIMO. A) PAX2 gene model depicting the location of the conserved variant, falling within an intergenic region upstream of the transcription start site. B) The plot demonstrates the PhyloP scores across the constrained region within 25-bp windows; the location of the called motif is highlighted in orange. C) Motif sequence logo for the SP2 motif, demonstrating the most common nucleotide at the variant location is 'C', where the population variant is a 'T'. Multi-species sequence alignment taken from the whole-genome Progressive Cactus alignment demonstrates the conservation of that 'C' allele across all 11 mammalian genomes.

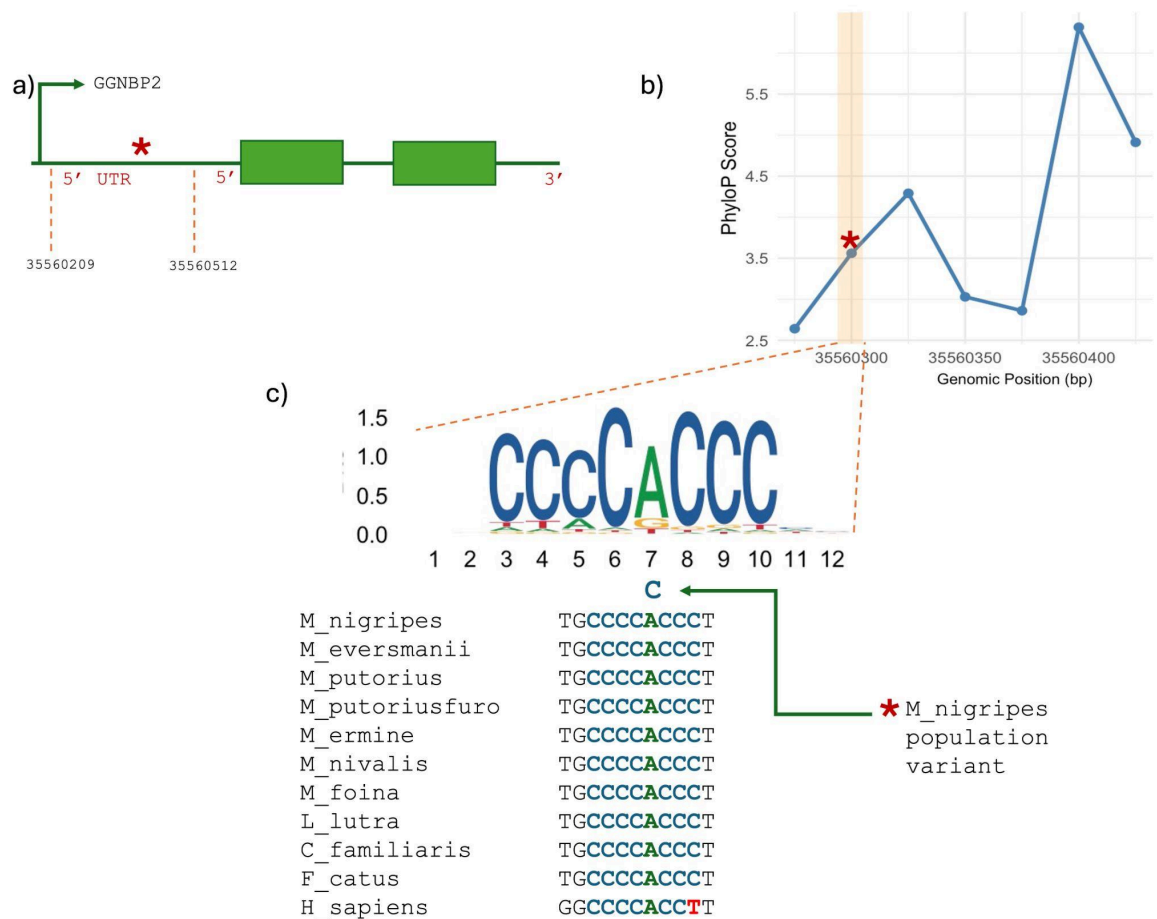


Figure 4.9: Visualisation of the variant falling within a non-coding region, classified as highly constrained by PhyloP and within an KLF4 motif as classified by FIMO. A) GGNBP2 gene model depicting the location of the conserved variant, falling within the 5' UTR region (which ranges from 35560209-35560512). B) The plot demonstrates the PhyloP scores across the constrained region within 25-bp windows; the location of the called motif is highlighted in orange. C) Motif sequence logo for the KLF4 motif, demonstrating the most common nucleotide at the variant location is 'A', where the population variant is a 'C'. Multi-species sequence alignment taken from the whole-genome Progressive Cactus alignment demonstrates the conservation of that 'A' allele across all 11 mammalian genomes.

4.5 Discussion

In this chapter, the aim was to investigate the extent to which deleterious mutations may reside within non-coding elements of the black-footed ferret genome. Such regions, including promoters, enhancers, and other regulatory sequences, play critical roles in controlling gene expression and are increasingly recognised as contributors to complex traits and disease phenotypes. This work aimed to investigate the presence of regulatory mutations within the non-coding genome that may adversely impact important biological functions in the black-footed ferret. It subsequently identified potentially deleterious regulatory mutations linked to these processes, providing insight into the genetic factors influencing the species' conservation.

4.5.1 Detection of constrained regions using a reference-free, multi-species whole-genome alignment

A comparative genomics tool was created by creating a whole-genome alignment of 11 mammalian species using Progressive Cactus. This allowed for the detection of constrained and accelerated regions across the black-footed ferret genome. Using this tool, I was able to identify that 1.24% of black-footed ferret genome bases were significantly conserved, with 0.4% of bases significantly accelerated.

Similar studies have found ranging levels of constraint and acceleration across genomes, based on the number of genomes used within the comparative analysis (Cooper et al., 2005; Feng et al., 2020; Kuderna et al., 2024; Lindblad-Toh et al., 2011; Sullivan et al., 2023). Higher taxonomic coverage within a multi-species alignment increases the power to call evolutionary conserved regions (Eddy, 2005; Feng et al., 2020; Hecker & Hiller, 2020; Sullivan et al., 2023). The Zoonomia project utilised a 240-way mammalian genome alignment to detect genomic regions under evolutionary constraint at a base-pair level. Their analyses found that 3.3% of the human genome was under constraint (Sullivan et al., 2023), consistent with previous estimates that 3-15% of the human genome could be under constraint due to purifying selection (Hecker & Hiller, 2020; Ponting & Hardison, 2011).

Our estimates for the level of constraint across the black-footed ferret genome are therefore slightly lower than those for humans, and potentially could be due to the low number of species aligned. This is consistent with our lack of resolution to define conserved regions at a single base-pair, and I had to increase our window size to 25-bp to increase statistical power (Lindblad-Toh et al., 2011; Secomandi et al., 2023). Our limited power to detect constraint at a single base-pair was also similarly reflected in results obtained when running GERP++. This potentially poses a limitation to understanding a

more realistic level of constraint across the black-footed ferret genome due to the constrained elements detected within larger window sizes.

However, using these methods, I was able to detect accelerated regions and regions under constraint in both coding and non-coding regions of the black-footed ferret. Our analyses demonstrated a high percentage of regions annotated by PhyloP scores as constrained were also called by gerpelem. Similarly, I found a high number of overlap between regions called as constrained in domestic ferret and black-footed ferret. This infers the potential functional importance of these regions within gene regulation and expression. The GO analysis consisting of 518 genes with the highest number of conserved elements revealed an enrichment for processes involved in transcription, regulation, development and protein binding. These results are similar to those found in previous comparative studies in birds (Secomandi et al., 2023), and mammals (Kuderna et al., 2024; Zemke et al., 2023) suggesting the functional importance of these processes.

The comparison of conserved elements between the black-footed ferret and the domestic ferret revealed a substantial overlap, reflecting their close evolutionary relationship. Nevertheless, a significant proportion of conserved regions were unique to each species. These species-specific conserved elements are distributed almost equally between genic and intergenic regions, indicating that both gene-associated sequences and distal regulatory regions contribute to lineage-specific conservation patterns. The presence of unique conserved elements may represent species-specific regulatory adaptations or differences in selective pressures that have acted since their divergence. Additionally, some of these unique elements may arise from methodological factors, such as variation in genome assembly quality or annotation completeness. The comparable proportions of genic and intergenic elements among species-specific conserved regions suggest that functional innovation or divergence occurs both within gene bodies and in regulatory landscapes distant from coding sequences. Further work may consider to elucidate the functional significance of these unique conserved elements which may provide valuable insights into species-specific traits and evolutionary histories, especially considering the distinct ecological niches and demographic trajectories (domestication and population bottleneck) of these mustelid species.

4.5.2 Population variants within constrained elements

Given the historically low effective population size of the black-footed ferret, I sought to investigate whether purifying selection remains efficient at removing potentially deleterious mutations within functionally important regions of the genome. Specifically, I assessed the DAF spectrum in CNEs compared to non-conserved intergenic and intronic regions. This approach allowed us to test whether patterns of variation in CNEs were consistent with

ongoing selective constraint, or whether reduced population size had compromised the efficacy of selection. Consistent with previous studies that have examined the role of purifying selection in CNEs, I observed a clear skew in the DAF spectrum within CNEs towards rarer variants, alongside a lower density of single nucleotide polymorphisms compared to neutral regions (Christmas et al., 2023; Groß, Bortoluzzi, et al., 2020b; Lindblad-Toh et al., 2011). These patterns are characteristic of purifying selection acting to suppress the accumulation and increase in frequency of harmful mutations within regulatory elements. While demographic bottlenecks typically reduce the strength of selection by amplifying the role of genetic drift, our findings suggest that purifying selection is still acting strongly on at least a subset of functionally constrained non-coding regions (Bortoluzzi et al., 2020; Lynch, 2010). This may reflect the essential regulatory roles of these elements, where even mild disruption could incur significant fitness costs. Alternatively, the observed patterns could be shaped in part by historical purging of deleterious variants during the black-footed ferrets population contraction. Together, these results indicate that, despite the genetic consequences of a severe bottleneck, selective mechanisms continue to play a role in maintaining the integrity of the non-coding genome.

I further investigated the impact that variants were having within functional regions of the genome. Our analyses identified 158 population variants that fell within highly conserved, regulatory sequences. Using a change in the PWM scores, I identified that 38 of these variants may be having an impact on the binding affinity of the transcription factor, leading to a potential disruption in the regulation or expression of the gene. Of the 30 genes that these variants fell within or close to, six genes have been described on the Mouse Genome Informatics (MGI) database (Eppig 2017) as having roles in the reproduction systems in mice (*TAOK1*, *FOXA3*, *HECTD4*, *MORC2*, *GGNBP2*, *PRMT1*) (Supplementary material S4.5). Where mutations in these genes have led to male sterility in mice (*GGNBP2*, *MORC2*, *PRMT1*), azoospermia (*GGNBP2*) or abnormal testis morphology (*MORC2*). Likewise, mutations within these genes have also been found to have an impact on reduced female fertility in mice (*MORC2*).

Several of these genes have also been reported as having important roles in neuronal development (*HECTD4*, *MORC2*, *TAOK1*, *ZBTB7A*, *TRIM71*, *NCDN*), where mutations within these genes can lead to diseases, such as Charcot-Marie-Tooth (*MORC2*), which affects the peripheral nervous system. Others have been reported as having important roles within immune response (*RBM10*, *TAOK1*, *ZBTB7A*, *SRF*), behaviour (*IGF2BP3*, *RILPL1*, *PRMT1*, *TAOK1*) and skeleton development or morphology (*COL1A2*, *TGIF1*, *PRMT1*, *TRIM71*, *SRF*).

Previous studies have found that utilising similar comparative methods have identified causal variants for common and rare diseases (Sullivan et al., 2023). Similarly to findings from these studies, I found that variants within the identified conserved regulatory regions were present in the black-footed ferret population at low frequencies (0.02 - 0.12). This would be consistent with findings that variants that potentially have a large effect size - i.e a profound effect on the trait it is associated with, are kept at low frequency by selection (Gazal et al., 2018; Sullivan et al., 2023). This suggests the importance of further investigation into the frequency of these variants throughout the black-footed ferret population. Here, I present variants from only 54 individuals. Understanding the heritability of these variants, and combining phenotype data alongside this information, could provide an understanding of the impacts that these variants could be having on individuals.

4.5.3 Considerations for further work

The work presented in this chapter introduces a mustelid-specific tool for the identification of functionally conserved genomic regions. Using this resource, I have identified variants within conserved, non-coding regions of the black-footed ferret genome that may influence tightly regulated biological processes. These variants are located within regulatory elements or are proximal to genes involved in critical functions such as immune responses, spermatogenesis, neurodevelopment, and skeletal development.

Recent studies have reported signs of inbreeding depression in captive black-footed ferret populations, including compromised sperm motility and morphology, reduced litter sizes, and diminished immune responses (Santymire et al. 2006; Santymire et al. 2014; Santymire et al. 2019; Wisely et al. 2002). Morphological changes have also been observed, with captive individuals exhibiting reduced cranial dimensions relative to those in the founding population, and reintroduced individuals displaying smaller body sizes compared to their historical counterparts. These changes are believed to be a consequence of long-term captivity (Wisely et al., 2002).

Our findings provide a set of candidate variants that may be useful in evaluating the downstream effects on protein-coding genes, particularly those associated with the phenotypes observed in captive populations. Future research could focus on examining the presence and frequency of these variants in individuals exhibiting traits such as poor sperm motility, lower reproductive success, or reduced immune function.

Furthermore, our study highlights the value of generating additional genome-wide data. In particular, incorporating chromatin conformation data, such as Hi-C, could enhance our understanding of the structural and functional consequences of non-coding mutations. While our PWM-based approach offers a practical means of assessing the potential regulatory impact of non-coding variants, it relies on simplified models of transcription

factor binding. PWM scores assume that each nucleotide contributes independently and additively to binding affinity, and they do not take into account chromatin accessibility, nucleosome positioning, or the influence of co-binding transcription factors (Steinhaus et al., 2022; Weirauch et al., 2014). Consequently, some predicted motif disruptions may not exert regulatory effects *in vivo*, whereas others with relatively modest score changes could have substantial impacts if they occur in accessible regulatory regions (Spivakov, 2014). Incorporating functional genomic data, such as ATAC-seq or ChIP-seq from related species or relevant tissues, would improve the interpretability of these predictions and support more accurate prioritisation of candidate variants for experimental follow-up. This would also facilitate the identification of topologically associated domains (TADs), thereby improving resolution of gene regulatory landscapes and the interpretation of non-coding variant effects (Long et al. 2022; Rajderkar et al. 2023; Groß et al. 2020).

Overall, this work underscores the growing importance of genomic tools in elucidating the genetic consequences of small population size and inbreeding in endangered species. These insights offer meaningful opportunities to inform and enhance conservation management strategies. Moreover, this research highlights the potential of comparative and functional genomics to inform conservation beyond traditional genetic diversity metrics, offering a pathway to assess the functional health of genomes in populations under threat. As high-quality reference genomes and increasing understanding of the regulatory genome become increasingly accessible across taxa, the approach presented here can be applied to other endangered species facing similar challenges and further evaluated. In doing so, it provides a valuable model for linking genotype to phenotype in a conservation context.

5.0 General Discussion: Genomics in Conservation: Utility, Practicality, and Tool Development

5.1 Abstract

The advent of high-throughput sequencing has transformed conservation biology by providing powerful insights into how environmental pressures shape the genomes of wild populations. This enhanced genomic resolution allows researchers to investigate population structure, demographic history, and functional genetic variation in greater detail. Such information enables conservationists and wildlife managers to make more informed, genomics-led decisions for the management and preservation of threatened species. Throughout this thesis, whole-genome sequencing was employed to investigate the genomic consequences of population bottlenecks of varying intensities in two mustelid species: the European polecat (*Mustela putorius*) and the black-footed ferret (*Mustela nigripes*). The case studies presented here illustrate both the potential and current limitations of genomics for informing conservation decisions, while also highlighting methodological innovations and areas for further development.

5.2 Chapter Two: Demographic Insights and Hybridisation in British Polecats

Chapter Two aimed to investigate the extent of the decline and subsequent recovery of the British polecat population. To do this, I used genome-wide data to reconstruct the demographic history of the British polecat, revealing the legacy of a historical population bottleneck and the complexity of the polecats' subsequent recovery. Genomic analysis showed a clear signature of decline, followed by recovery in British polecat populations. I investigated whether whole genome diversity reflected this demographic event and found that Welsh polecats had significantly less genetic diversity than English polecats, but not European polecats. Runs of homozygosity and genetic load present in Welsh and English polecat genomes also indicated recent historic inbreeding. Analyses also suggested that the Welsh population retained distinctive genetic variation and showed less evidence of recent admixture with domestic ferrets. In contrast, English populations exhibited varying levels of introgression, complicating efforts to infer pre-bottleneck genetic health or to estimate the long-term impact of hybridisation.

Genomic data enable conservationists to move beyond traditional markers, such as microsatellites or mitochondrial DNA, offering a much finer resolution of inbreeding, genetic load, and adaptive potential. Within Chapter Two, I have demonstrated the higher resolution that whole-genome data provides in comparison to reduced marker panels, resolving fine-scaled population structure within British populations. This is further demonstrated by the results obtained with GONE (Santiago et al., 2020), where it was possible to use whole-genome data to model effective population sizes across multiple generations, capturing signatures of the historic population bottleneck. This approach was able to infer the timing of the population crash to within several generations (30-40), which is consistent with expected findings due to the increasing pressure of the agricultural industry and intensity of persecution through land management practices during this time. This signal was not previously elucidated by using a reduced representation of the genome (Costa et al. 2013), thus showing the advantage and benefit of using genome-wide SNPs for such use-cases. However, certain limitations remain. It was difficult to establish the pre-bottleneck condition of the polecat population, and uncertainty persists regarding the long-term effects of introgression on fitness and adaptability. Future work would aim to sample from museum specimens, which would provide potential insight into the genetic health of the pre-bottleneck population (Dussex et al., 2021; van der Valk et al., 2019; von Seth et al., 2021). This would allow for a more comprehensive understanding of the consequences that the population bottleneck has had on the current population. Similar aligned work on otter populations in Britain highlights the viability of these methods through the use of preserved hunting trophies pre-dating the population crash of otters in Britain (du Plessis et al., 2025). Preliminary findings elucidate that current-day population structuring within British otters arose post-bottleneck, through

forces such as genetic isolation and drift driven by reduced population sizes and reduced gene flow. These findings have direct implications for the conservation management of present-day populations to bridge gaps in gene flow between populations. Understanding the mechanisms impacting the genomic health of populations can assist in the management of anthropogenic impacts on wild populations.

This shift in genomic data availability for conservation problems brings new challenges. Questions remain regarding how to interpret the functional consequences of genetic erosion, how to account for admixture or hybridisation events, and how to contextualise genomic data within the broader ecological and management landscape. This is particularly highlighted within this chapter in the context of the evidence through the extent to which hybridisation has occurred between polecats in England and feral domestic ferrets. Although previous reports have estimated that events leading to F1 hybrids have been rare in Britain (Costa et al. 2013), these results indicate that introgression with domestic ferrets is markedly present in English polecats, as well as a suspected F1 hybrid present within Anglesey, Wales. The understanding of this finding is limited by the samples available for analysis; however, it could potentially indicate that hybridisation events are still occurring, and may now be taking place in Wales.

As previously described, these events can create issues in conservation protections for threatened species due to the genetic integrity around hybrid populations (Barnes et al., 2022; Brambilla et al., 2024; Senn et al., 2019). These findings do align with the argument within the literature in recent years, where hybridisation can be seen as an adaptive management tool (Barnes et al., 2022). This is evident in my findings through the increased genetic diversity and reduced realised genetic load within English polecats in comparison to Welsh and mainland European individuals. Similar hybridised populations of threatened or near-threatened species have resulted in adaptive introgression, leading to an increase in genetic fitness in the resulting offspring. This has been seen in the Scottish wildcat (*Felis catus silvestris*) and Alpine Ibex (*Capra ibex*), where selective pressures have led to introgression within MHC regions of the genome (Howard-McCombe et al., 2023; Münger et al., 2024). Increasing genetic diversity within these regions of the genome has led to protection of hybrid individuals from some diseases, increasing their adaptive potential (Gaczorek et al., 2024; Hedrick, 2013). Whether adaptive introgression can be attributed to the recovery of the polecat in Britain requires further investigation. A previous study found regions of introgression within the protein-coding genes of English polecats associated with cognitive function and sight (Etherington et al., 2022). Future work could look to expand this by producing local ancestry estimates to delineate ancestry across regions of the genome (Howard-McCombe et al., 2023; Lawson et al., 2024).

This scenario in Britain will be of interest to the investigation of polecat populations across mainland Europe as well, where populations are thought to be declining or are understudied (Barrientos et al., 2024; Croose et al., 2018). Previous studies have investigated links of introgression between the European polecat and American Mink (*Neovison vison*) (Martínez-Cruz et al., 2022), Steppe polecat (*Mustela eversmanii*) (Cserkész et al., 2021), and the threatened European Mink (*Mustela lutreola*) (Cabria et al., 2011). These studies have utilised microsatellite or mitochondrial markers to study the population genetics of polecats in Europe, finding limited evidence of introgression or structure. Whole-genome data for polecat individuals in mainland Europe are sparse and are limited to those found in this study and an additional study that utilised restriction site-associated DNA sequencing (RADseq) (Szatmári et al., 2021). Szatmári and colleagues found a marked difference in the investigation of population structure and introgression of *M. eversmanii* and *M. putorius* between traditional markers and whole-genome SNPs. In this study, hybrids that were identified by analysis using microsatellite data could not be confirmed using whole-genome SNPs, demonstrating the limited resolution that a small panel of markers can provide. Additional genomic resources could provide better resolution for fine-scale structure and could provide greater estimates for evidence of introgression between populations.

Further to this, bridging the gap towards the accessibility of genomic data tools, data storage, and the cost of sequencing are still fundamental issues to overcome. This thesis demonstrates that while genomic tools are increasing in their accessibility, their full integration into conservation practice remains limited. To achieve meaningful impact, genomics must be embedded not only within academic studies but also within policy frameworks and biodiversity conservation legislation. Doing so could unlock greater funding opportunities, increase equity in the understanding of genomic data, which would aim to ensure the long-term protection of genetic diversity, and support the development of genomic baselines for threatened species.

The chapter also touched upon the growing importance of public engagement and citizen science initiatives in conservation monitoring. Increased awareness of polecats among the public and wildlife professionals has led to improved non-invasive sampling and more comprehensive spatial coverage (Birks 2008; Croose 2016; Croose et al. 2018). The efforts of the Vincent Wildlife Trust in their extensive efforts of monitoring the British polecat recovery are incomparable to those across mainland Europe, where many populations are seemingly understudied or are persisting at such low density that sightings are negligible (Croose et al., 2018). Without such concerted efforts from the Vincent Wildlife Trust and members of the public who participated in the 2014-2015 survey, I would have been unable to carry out the genomic analyses presented in this study.

Future research should prioritise the integration of historical DNA from museum specimens to provide a full temporal picture of genetic diversity across Great Britain, as well as into the further investigation of introgression within English polecats. Broader sampling across both Britain and continental Europe would also support efforts to clarify the origin of the European polecat in Great Britain. These results indicate a degree of genetic differentiation between British and mainland European polecats, a pattern observed in several other British mammals, where post-glacial isolation from continental populations has led to distinct genetic signatures (Carranza et al., 2024; McDevitt et al., 2022). Additionally, these findings contribute to understanding the evolutionary history of the species and may have important conservation implications, particularly in identifying populations at risk and informing future reintroduction or management strategies (Croose et al., 2018; Martínez-Cruz et al., 2022).

5.3 Chapter Three: Genetic Load and Recovery in the Black-Footed Ferret

Chapter Three focused on the critically endangered black-footed ferret, a species that underwent an extreme genetic bottleneck during the 20th century, largely due to habitat loss, loss of its prey species, and disease. The analysis of genomic data from 54 individuals across present-day captive, reintroduced, and members of the founding population revealed the persistence of deleterious variants within the population, as well as the presence of fixed, weakly deleterious mutations likely introduced through inbreeding. Nonetheless, evidence suggested some degree of purging of unmasked deleterious variants within the captive-bred populations.

Importantly, the chapter highlighted the role of captive breeding management in mitigating further genetic erosion. Genetic diversity metrics varied between populations (founding - 0.00000147, captive - 0.00000159, reintroduced - 0.00000198), reflecting both the legacy of the founding bottleneck and the effectiveness of subsequent breeding strategies. Although all current individuals trace their ancestry to a limited founder group, careful pedigree management and assisted reproduction have helped to retain a degree of genetic variation.

Despite these efforts, the continued presence of genomic erosion from inbreeding and functionally relevant mutations underscores the need for long-term genomic monitoring. In comparison to the European polecat populations that were analysed in Chapter Two, the black-footed ferret population shows an extreme loss of genetic diversity and genome-wide heterozygosity. The European polecat population in Wales, England, and across mainland Europe shows a ten-fold difference in nucleotide diversity in comparison to black-footed ferret populations. Inbreeding coefficients were also demonstrably higher across some black-footed ferret populations (-0.39 - 0.48), particularly in the founding population and some captive populations, in comparison to those seen in European

polecat populations (0.16 - 0.35). This demonstrates that the strength of the genetic bottleneck does shape the population viability of future generations (Bortoluzzi et al., 2020) and highlights that the intervention of a captive breeding programme and reintroduction of individuals has helped to reduce incidences of inbreeding in the black-footed ferret. Future directions include sequencing the offspring of cloned individuals to assess the recovery of lost diversity and integrating historical genomes to compare modern ferrets with their pre-bottleneck ancestors to determine the extent of the genetic bottleneck. Genomic monitoring of the reintroduced populations would also help to inform on the genomic health of the population and whether deleterious mutations were naturally being purged from the genome. Such work would further inform strategies for managing mutational load and identifying candidates for future genetic rescue (Speak et al., 2024; van der Valk et al., 2019; van Oosterhout et al., 2022).

This chapter also highlights the potential uses in genomics to assist in the restoration of genetic diversity to a population to increase population viability. Restoration genomic approaches (van der Valk & Dalèn, 2024) are increasingly being utilised to boost threatened populations, informed by the insight that increased resolution of genomic data can bring. Approaches include selecting suitable individuals for captive breeding programmes (Humble et al., 2023; Speak et al., 2024) or reintroductions (Farquharson et al., 2022; Quinn et al., 2024), and artificial insemination to aid assisted breeding (von Seth et al., 2021).

Scientists and conservationists are looking to more novel methods, alongside these traditional approaches to preserving genetic diversity within populations. One such method is the utility of cryopreservation of biological materials, which has revolutionised conservation thinking by providing opportunities to reintroduce genetic diversity from historic individuals. This is especially relevant in the context of the black-footed ferret, where the cloning of two individuals from the historic wild population using biobanked somatic cells demonstrates the validity of this technique and the opportunities available to other endangered species to reintroduce genetic diversity that was lost from the extant population. Zoos and animal management institutions are currently preserving genetic material for these reasons and to prevent any further loss of biodiversity and genomic diversity from endangered populations. Techniques for the preservation of biological materials have been further developed, as embryos and germ cells can often be difficult to collect from endangered species (Hutchinson et al., 2024). Focus has turned to stem cell technology, and the use of induced pluripotent stem cells (iPSCs), which have the potential to develop into multiple different cell types and can be directly derived from somatic (tissue) cells and programmed into germ cells (Hutchinson et al., 2024; Novak et al., 2024, 2025). The European Association for Zoos and Aquariums (EAZA) biobank is one of the largest concerted efforts for storing animal tissue and blood from zoos, wildlife

parks, and conservation centres across Europe. It currently holds over 28,000 samples from over 1,100 subspecies and is poised to act as a 'doomsday' resource to prevent the impact of genetic bottlenecks by enhancing gene flow within the population (*EAZA Biobank*, n.d.). An example of this approach is seen in the revival effort for the critically endangered Northern white rhino (*Ceratotherium simum cottoni*), where skin cells from 12 individuals have been banked for future use in population restoration. Simulations using these frozen genomes suggest that the viability and genetic diversity of a revived Northern white rhino population could exceed that of the extant Southern white rhino, highlighting the potential impact of this technology for the species recovery (Wilder et al., 2024). However, this technique is still seen as a last resort due to the current costs of this measure sitting at around \$40,000 (Fritts, 2022; Wetzel, 2024). Furthermore, these technological advances are still subject to familiar conservation debates regarding which individual would be most beneficial to reintroduce into the population. Genomic assessments of biobanked material would be required to determine the mutation load of the individual and therefore the potential for their genetics to reintroduce deleterious mutations that may have already been purged from the population (Dussex et al., 2023).

5.4 Chapter Four: Tool Development and the Regulatory Genome

Chapter Four addressed a broader methodological gap in conservation genomics: the need to better understand and annotate the non-coding genome. This chapter introduced a novel tool to identify constrained elements across the mustelid lineage, offering a valuable resource for interpreting regulatory mutations in both model and non-model species. By identifying genomic regions under purifying selection, the tool provides insights into the functional landscape of mammalian genomes, which is particularly relevant for understanding the impact of bottlenecks and the accumulation of mutations on regulatory variation.

The analyses within this chapter identified novel mutations within the black-footed ferret population that have potentially accumulated due to the population bottleneck and subsequent inbreeding, with purging of these mutations not yet occurring. Our analyses also highlight mutations that may be impacting the regulation of genes involved in key processes, such as spermatogenesis and immune response. These are phenotypes which have been associated with changes having occurred since the timing of the bottleneck and founding of the captive breeding population in the 1980s (R. M. Santymire et al., 2019; Wisely et al., 2002). Further analysis would aim to generate a genome-wide framework to assess the deleteriousness of these mutations, clarify their functional context, and evaluate their potential impact. Mutation scores, such as Combined Annotation-Dependent Depletion scores (CADD), assess the impact that a mutation is having based on surrounding sequence context, gene annotations, evolutionary constraint, and functional data such as RNA-seq and ATAC-seq experimental data (Groß,

Bortoluzzi, et al., 2020b; Groß, Derks, et al., 2020; Kircher et al., 2014). This framework enables the prioritisation of mutations based on their predicted deleteriousness, allowing for the ranking of variants found within both coding and non-coding regions according to their impact. It has been used to prioritise variants found within functionally important regions in humans (Rentzsch et al., 2019), domesticated species (Groß, Bortoluzzi, et al., 2020b; Groß, Derks, et al., 2020), and species that have undergone genetic bottlenecks (Fontseré et al., 2024; Speak et al., 2024). The development of such a model for the black-footed ferret, integrated with studbook data and phenotypic information from the captive population, could support efforts to mitigate inbreeding and alleviate the associated genomic load.

The chapter has highlighted key challenges in applying such tools to conservation contexts. High-quality reference genomes remain a limiting factor, especially for non-model taxa (Brandies et al., 2019; Formenti et al., 2022). Conservation genomic studies are often limited in their capacity to interpret regulatory variants due to data availability or quality (Almeida-Silva & Van de Peer, 2023; Feron & Waterhouse, 2022). Short-read sequencing and poor genome annotation limit accurate interpretation of regulatory variants because they cannot capture long-range interactions between enhancer regions and their target genes. Hi-C data provides a 3D map of the genome, showing how different regions physically interact and help to identify which regulatory elements control which genes (Chen et al., 2023; Rao et al., 2014). This work therefore advocates for greater investment in long-read sequencing, improved genome annotation, and the generation of contact maps. Hi-C can resolve genome-wide interactions between regulatory regions and genes, improving our understanding of how mutations in these areas affect function. Developing these resources is crucial to move beyond just describing inbreeding and diversity. A better understanding of gene regulation and constraint will help link genomic erosion to specific phenotypic effects, enhancing the predictive power of genomic tools in conservation biology (Groß, Bortoluzzi, et al., 2020b).

Due to the sparsity of this data, especially for some taxa (Cowie et al., 2022), conservation and evolutionary genomic approaches are starting to utilise machine-learning methods to infer functionality and regulation, as well as estimating long-term viability and extinction risk of species (van Oosterhout, 2024). AI models can be trained on large datasets from well-studied model organisms to infer functional constraints, identify regulatory regions, and predict the deleteriousness of genetic variants (Labes et al. 2022; Kaplow et al. 2023; Zhao et al. 2022). These approaches have enabled researchers to bypass some of the limitations of traditional experimental methods for identifying functional genomic regions, such as reporter assays, allowing for rapid, large-scale predictions (Kumari et al., 2025). However, several studies have highlighted the challenges of transferring these predictions across species, as patterns of sequence conservation and variant pathogenicity may not

be consistent across taxonomic boundaries (Labes et al., 2022; Malhis et al., 2019). This inconsistency can arise from differences in evolutionary rates and selective pressures acting on genes across taxa, leading to varying nucleotide changes (Wong et al., 2020). Furthermore, this emphasises that transferring the pathogenicity of variants from one species to another is complicated because even within seemingly conserved regions, variants may be adaptive as opposed to deleterious (Kaplow et al., 2023; Sundaram et al., 2018; Wong et al., 2020). This highlights a major limitation of using model species as references for inferring deleterious mutations in non-coding regions of the black-footed ferret genome.

To address these challenges, models have looked to incorporate low-level sequence conservation (Kaplow et al., 2023) or evolutionary distance (Karollus et al., 2024; Labes et al., 2022) to improve the prediction of regulatory regions. These approaches have shown increased power in identifying conserved regulatory elements across taxa and in predicting base-level constraint within those regions. For example, the TACIT framework trains machine learning models on open chromatin data from a small number of species, capturing tissue and cell-type-specific enhancer activity (Kaplow et al., 2023). By focusing on functional conservation rather than strict sequence identity, TACIT has linked candidate enhancers to phenotypes such as brain size and social behaviour. However, authors have highlighted that the accuracy of prediction is still currently constrained by the scarcity of high-quality functional genomic data across a broad range of species.

Pangenome graphs are also increasingly being recognised as powerful tools in conservation genomics, as they allow for a more inclusive representation of genetic diversity within and across populations (Affini et al., 2025; Secomandi et al., 2023, 2025) and could be used to address some of these limitations. Incorporating pangenome graphs into AI-based prediction frameworks could improve cross-species functional inference for non-model species by capturing a more complete spectrum of genetic variation, reducing reference bias, and enabling models to account for lineage-specific patterns of conservation and constraint (Secomandi et al., 2025). The generation and incorporation of such data would enhance our understanding of the genomic consequences of population bottlenecks and subsequent inbreeding. This knowledge would facilitate the development of more precise conservation management strategies and genomically informed breeding programs tailored to halting genomic erosion and mitigating the accumulation of deleterious mutations.

5.5 Conclusion: Bringing Genomics into Broader Conservation Practice

The case studies in this thesis demonstrate that genomics can substantially inform conservation strategies, particularly in understanding historical demography, quantifying genetic load, and monitoring hybridisation. The findings highlight how genomic data can improve our understanding of how populations adapt to environmental pressures and can allow for monitoring and assessment of future population viability. However, despite clear potential, widespread uptake of genomic tools within conservation practice remains uneven. In part, this reflects a disconnect between researchers and decision-makers, as well as uncertainty over how to interpret and act on complex genomic data.

To ensure that genomics plays a central role in biodiversity conservation, it must be more fully integrated into legal and policy frameworks. Genomic data should complement existing conservation metrics and be considered when assessing species status, planning translocations, or allocating funding. Building cross-disciplinary partnerships, training conservation practitioners in genomics, and creating accessible analytical tools will also be key. Greater legislative recognition of the importance of genetic resources may also incentivise investment in genomic monitoring and long-term data stewardship. This would not only aid in the recovery of individual species but also contribute to broader goals in preserving evolutionary potential across ecosystems.

In conclusion, this thesis demonstrates that genomics can reveal nuanced and often unexpected insights into the consequences of population decline and recovery. Across two mustelid species with contrasting conservation histories, it was possible to alleviate the impact of genetic bottlenecks, detect introgression, assess mutation load, and develop tools for broader application. Yet the full potential of genomics in conservation will only be realised through its integration into practical decision-making and policy development.

Future conservation efforts must continue to invest in the genomic characterisation of threatened species, particularly those with complex management histories. By doing so, I can move towards a more predictive, functional, and effective model of conservation, one that recognises the value of multiple genomic indicators in the assessment of the population viability and long-term adaptive potential of species under threat (van Oosterhout et al., 2022).

References

- Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., Boughman, J., Brelsford, A., Buerkle, C. A., Buggs, R., Butlin, R. K., Dieckmann, U., Eroukhmanoff, F., Grill, A., Cahan, S. H., Hermansen, J. S., Hewitt, G., Hudson, A. G., Jiggins, C., ... Zinner, D. (2013). Hybridization and speciation. *Journal of Evolutionary Biology*, 26(2), 229–246.
- Adzhubei, I., Jordan, D. M., & Sunyaev, S. R. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. *Et Al [Current Protocols in Human Genetics]*, Chapter 7(1), Unit7.20.
- Affini, A., Baranowski, H., Forbes, S., Foust, E. N., Hatley, K., Ni, E. L., McGuire, A., Paist, K., Rutter, M., Smith, R. N., Vargas, N., Akella, H., Castellano, K., Pauloski, N., Reid, N., O'Neill, R., King, T., Jockusch, E. L., Wegrzyn, J. L., & Hapeman, P. (2025). Mink by mink: stitching together signatures of subspecies adaptation through a pangenome of threatened mustelids. In *bioRxiv* (p. 2025.07.25.666817). <https://doi.org/10.1101/2025.07.25.666817>
- Alemu, S. W., Lopdell, T. J., Trevarton, A. J., Snell, R. G., Littlejohn, M. D., & Garrick, D. J. (2025). Comparison of genomic prediction accuracies in dairy cattle lactation traits using five classes of functional variants versus generic SNP. *Genetics, Selection, Evolution: GSE*, 57(1), 20.
- Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19(9), 1655–1664.
- Alexander, R. P., Fang, G., Rozowsky, J., Snyder, M., & Gerstein, M. B. (2010). Annotating non-coding regions of the genome. *Nature Reviews. Genetics*, 11(8), 559–571.
- Alföldi, J., & Lindblad-Toh, K. (2013). Comparative genomics as a tool to understand evolution and disease. *Genome Research*, 23(7), 1063–1068.
- Ali, N., Amelkina, O., Santymire, R. M., Koepfli, K.-P., Comizzoli, P., & Vazquez, J. M. (2024). Semen proteome and transcriptome of the endangered black-footed ferret (*Mustela nigripes*) show association with the environment and fertility outcome.

Scientific Reports, 14(1), 7063.

Allendorf, F. W. (2017). Genetics and the conservation of natural populations: allozymes to genomes. *Molecular Ecology*, 26(2), 420–430.

Allendorf, F. W., Hohenlohe, P. A., & Luikart, G. (2010). Genomics and the future of conservation genetics. *Nature Reviews. Genetics*, 11(10), 697–709.

Allendorf, F. W., Leary, R. F., Spruell, P., & Wenburg, J. K. (2001). The problems with hybrids: setting conservation guidelines. *Trends in Ecology & Evolution*, 16(11), 613–622.

Allio, R., Tilak, M.-K., Scornavacca, C., Avenant, N. L., Kitchener, A. C., Corre, E., Nabholz, B., & Delsuc, F. (2021). High-quality carnivoran genomes from roadkill samples enable comparative species delineation in aardwolf and bat-eared fox. *eLife*, 10. <https://doi.org/10.7554/eLife.63167>

Almeida-Silva, F., & Van de Peer, Y. (2023). Assessing the quality of comparative genomics data and results with the *cogeqc* R/Bioconductor package. *Methods in Ecology and Evolution*, 14(12), 2942–2952.

Álvarez-Carretero, S., Kapli, P., & Yang, Z. (2023). Beginner's Guide on the Use of PAML to Detect Positive Selection. *Molecular Biology and Evolution*, 40(4). <https://doi.org/10.1093/molbev/msad041>

Ambrosini, G., Vorontsov, I., Penzar, D., Groux, R., Fornes, O., Nikolaeva, D. D., Ballester, B., Grau, J., Grosse, I., Makeev, V., Kulakovskiy, I., & Bucher, P. (2020). Insights gained from a comprehensive all-against-all transcription factor binding motif benchmarking study. *Genome Biology*, 21(1), 114.

Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews. Genetics*, 17(2), 81–92.

Andrews, S. (2010). *FastQC: Quality Control Tool High Throughput Sequence Data*.

Armstrong, J., Hickey, G., Diekhans, M., Fiddes, I. T., Novak, A. M., Deran, A., Fang, Q., Xie, D., Feng, S., Stiller, J., Genereux, D., Johnson, J., Marinescu, V. D., Alföldi, J., Harris, R. S., Lindblad-Toh, K., Haussler, D., Karlsson, E., Jarvis, E. D., ... Paten, B.

- (2020). Progressive Cactus is a multiple-genome aligner for the thousand-genome era. *Nature*, 587(7833), 246–251.
- Avise, J. C. (2012). *Molecular markers, natural history and evolution* (1994th ed.) [PDF]. Chapman and Hall.
- Bagda, E., Bardakci, F., & Turkozcan, O. (2012). Lower genetic structuring in mitochondrial DNA than nuclear DNA among the nesting colonies of green turtles (*Chelonia mydas*) in the Mediterranean. *Biochemical Systematics and Ecology*, 43, 192–199.
- Baghli, A., & Verhagen, R. (2003). The distribution and status of the polecat *Mustela putorius* in Luxembourg: Polecats in Luxembourg. *Mammal Review*, 33(1), 57–68.
- Baghli, A., Walzberg, C., & Verhagen, R. (2005). Habitat use by the European polecat *Mustela putorius* at low density in a fragmented landscape. *Wildlife Biology*, 11(4), 331–339.
- Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., Selker, E. U., Cresko, W. A., & Johnson, E. A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PloS One*, 3(10), e3376.
- Ballard, J. W. O., & Whitlock, M. C. (2004). The incomplete natural history of mitochondria. *Molecular Ecology*, 13(4), 729–744.
- Baquero, R. A., Oficialdegui, F. J., Ayllón, D., & Nicola, G. G. (2023). The challenge of managing threatened invasive species at a continental scale. *Conservation Biology: The Journal of the Society for Conservation Biology*, 37(5), e14165.
- Barg, A., MacPherson, J., & Caravaggi, A. (2022). Spatial and temporal trends in western polecat road mortality in Wales. *PeerJ*, 10(e14291), e14291.
- Barnes, T. M., Karlin, M., vonHoldt, B. M., Adams, J. R., Waits, L. P., Hinton, J. W., Henderson, J., & Brzeski, K. E. (2022). Genetic diversity and family groups detected in a coyote population with red wolf ancestry on Galveston Island, Texas. *BMC Ecology and Evolution*, 22(1), 134.
- Barnosky, A. D., Matzke, N., Tomiya, S., Wogan, G. O. U., Swartz, B., Quental, T. B., Marshall, C., McGuire, J. L., Lindsey, E. L., Maguire, K. C., Mersey, B., & Ferrer, E. A. (2011). Has the Earth's sixth mass extinction already arrived? *Nature*, 471(7336),

- Barone, M. A., Roelke, M. E., Howard, J., Brown, J. L., Anderson, A. E., & Wildt, D. E. (1994). Reproductive Characteristics of Male Florida Panthers: Comparative Studies from Florida, Texas, Colorado, Latin America, and North American Zoos. *Journal of Mammalogy*, 75(1), 150–162.
- Barrientos, R., Carmona, G., Burgos, T., Martín-García, S., Vargas-Ramírez, M., Hernández-Hernández, J., Bandeira, V. J., Quiles, P., Palacín, C., Martín, C. A., Virgós, E., & Horreo, J. L. (2024). The genetics of the European polecat in the Iberian Peninsula. *The Journal of Wildlife Management*, 88(7), e22628.
- Barrientos, R., & de Dios Miranda, J. (2012). Can we explain regional abundance and road-kill patterns with variables derived from local-scale road-kill models? Evaluating transferability with the European polecat: From local- to regional-scale road-kill patterns. *Diversity & Distributions*, 18(7), 635–647.
- Barrientos, R., & Plaza, M. (2016). Road-kill hot spots can change over the time, variables explaining them do not. *IENE, Infra Eco Network Europe*.
- Beaumont, M., Barratt, E. M., Gottelli, D., Kitchener, A. C., Daniels, M. J., Pritchard, J. K., & Bruford, M. W. (2001). Genetic diversity and introgression in the Scottish wildcat. *Molecular Ecology*, 10(2), 319–336.
- Beichman, A. C., Koepfli, K.-P., Li, G., Murphy, W., Dobrynin, P., Kliver, S., Tinker, M. T., Murray, M. J., Johnson, J., Lindblad-Toh, K., Karlsson, E. K., Lohmueller, K. E., & Wayne, R. K. (2019). Aquatic Adaptation and Depleted Diversity: A Deep Dive into the Genomes of the Sea Otter and Giant Otter. *Molecular Biology and Evolution*, 36(12), 2631–2655.
- Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W. J., Mattick, J. S., & Haussler, D. (2004). Ultraconserved elements in the human genome. *Science (New York, N.Y.)*, 304(5675), 1321–1325.
- Bernstein, B. E., Stamatoyannopoulos, J. A., Costello, J. F., Ren, B., Milosavljevic, A., Meissner, A., Kellis, M., Marra, M. A., Beaudet, A. L., Ecker, J. R., Farnham, P. J., Hirst, M., Lander, E. S., Mikkelsen, T. S., & Thomson, J. A. (2010). The NIH

- Roadmap Epigenomics Mapping Consortium. *Nature Biotechnology*, 28(10), 1045–1048.
- Biggins, D. E., Eads, D. A., Ramakrishnan, S., Goldberg, A. R., Eads, S. L., Hardin, J., & Konkel, D. (2024). Prey selection by black-footed ferrets (*Mustela nigripes*): implications for intersexual resource partitioning and conservation. *Journal of Mammalogy*, 105(2), 221–229.
- Biggins, D. E., & Godbey, J. L. (2003). Challenges to reestablishment of free-ranging populations of black-footed ferrets. *Comptes Rendus Biologies*, 326 Suppl 1, S104–S111.
- Birks, J. D. (2008). *The polecat survey of Britain 2004-2006: A report on the polecats distribution, status and conservation*.
- Birks, J. D. S., & Kitchener, A. C. (1999). *The Distribution and Status of the Polecat Mustela Putorius in Britain in the 1990s*. Vincent Wildlife Trust.
- Blandford, P. R. S. (1987). Biology of the Polecat *Mustela putorius*: a literature review. *Mammal Review*, 17(4), 155–198.
- Boakes, E. H., Wang, J., & Amos, W. (2007). An investigation of inbreeding depression and purging in captive pedigreed populations. *Heredity*, 98(3), 172–182.
- Boffelli, D., Nobrega, M. A., & Rubin, E. M. (2004). Comparative genomics at the vertebrate extremes. *Nature Reviews. Genetics*, 5(6), 456–465.
- Bohling, J. H., & Waits, L. P. (2011). Assessing the prevalence of hybridization between sympatric *Canis* species surrounding the red wolf (*Canis rufus*) recovery area in North Carolina: CANIS HYBRIDIZATION IN EASTERN NORTH CAROLINA. *Molecular Ecology*, 20(10), 2142–2156.
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)*, 30(15), 2114–2120.
- Bortoluzzi, C., Bosse, M., Derks, M. F. L., Crooijmans, R. P. M. A., Groenen, M. A. M., & Megens, H.-J. (2020). The type of bottleneck matters: Insights into the deleterious variation landscape of small managed populations. *Evolutionary Applications*, 13(2), 330–341.

- Bosse, M., & van Loon, S. (2022). Challenges in quantifying genome erosion for conservation. *Frontiers in Genetics*, 13, 960958.
- Botstein, D., White, R. L., Skolnick, M., & Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *The American Journal of Human Genetics*, 32(3), 314–331.
- Brambilla, A., Zehnder, N., Bassano, B., Rossi, L., & Grossen, C. (2024). Genetic evidence of a hybrid swarm between Alpine ibex (*Capra ibex*) and domestic goat (*C. hircus*). *Evolutionary Applications*, 17(8), e13761.
- Brandies, P., Peel, E., Hogg, C. J., & Belov, K. (2019). The value of reference genomes in the conservation of threatened species. *Genes*, 10(11), 846.
- Breitenmoser, U., Lanz, T., & Breitenmoser-Würsten, C. (2019). *Conservation of the wildcat (Felis silvestris) in Scotland: Review of the conservation status and assessment of conservation activities*.
https://www.nature.scot/sites/default/files/2019-02/Wildcat%20in%20Scotland%20-%20Review%20of%20conservation%20status%20and%20activities_1.pdf
- Breithoff, E., & Harrison, R. (2020). From ark to bank: extinction, proxies and biocapitals in ex-situ biodiversity conservation practices. *International Journal of Heritage Studies*, 26(1), 37–55.
- BRIG. (2007). *Report on the Species and Habitat Review*.
<https://data.jncc.gov.uk/data/bdd8ad64-c247-4b69-ab33-19c2e0d63736/UKBAP-Species-HabitatsReview-2007.pdf>
- Britt, M., Haworth, S. E., Johnson, J. B., Martchenko, D., & Shafer, A. B. A. (2018). The importance of non-academic coauthors in bridging the conservation genetics gap. *Biological Conservation*, 218, 118–123.
- Broman, K. W., & Weber, J. L. (1999). Long homozygous chromosomal segments in reference families from the centre d'Étude du polymorphisme humain. *The American Journal of Human Genetics*, 65(6), 1493–1500.
- Brown, W. M., George, M., Jr, & Wilson, A. C. (1979). Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences of the United*

States of America, 76(4), 1967–1971.

- Brumfield, R. T., Beerli, P., Nickerson, D. A., & Edwards, S. V. (2003). The utility of single nucleotide polymorphisms in inferences of population history. *Trends in Ecology & Evolution*, 18(5), 249–256.
- Brzeziński, M., Marzec, M., & Żmihorski, M. (2010). Spatial distribution, activity, habitat selection of American mink (*Neovison vison*) and polecats (*Mustela putorius*) inhabiting the vicinity of eutrophic lakes in NE Poland. *Folia Zoologica*, 59(3), 183–191.
- Buckley, R. M., Davis, B. W., Brashear, W. A., Farias, F. H. G., Kuroki, K., Graves, T., Hillier, L. W., Kremitzki, M., Li, G., Middleton, R. P., Minx, P., Tomlinson, C., Lyons, L. A., Murphy, W. J., & Warren, W. C. (2020). A new domestic cat genome assembly based on long sequence reads empowers feline genomic medicine and identifies a novel gene for dwarfism. *PLoS Genetics*, 16(10), e1008926.
- Cabria, M. T., Michaux, J. R., Gómez-Moliner, B. J., Skumatov, D., Maran, T., Fournier, P., de Luzuriaga, J. L., & Zardoya, R. (2011). Bayesian analysis of hybridization and introgression between the endangered european mink (*Mustela lutreola*) and the polecat (*Mustela putorius*): HYBRID IDENTIFICATION BETWEEN TWO MUSTELID SPECIES. *Molecular Ecology*, 20(6), 1176–1190.
- Cahill, J. A., Heintzman, P. D., Harris, K., Teasdale, M. D., Kapp, J., Soares, A. E. R., Stirling, I., Bradley, D., Edwards, C. J., Graim, K., Kisleika, A. A., Malev, A. V., Monaghan, N., Green, R. E., & Shapiro, B. (2018). Genomic evidence of widespread admixture from polar bears into brown bears during the last ice age. *Molecular Biology and Evolution*, 35(5), 1120–1129.
- Cain, C. M., Livieri, T. M., & Swanson, B. J. (2011). Genetic evaluation of a reintroduced population of black-footed ferrets (*Mustela nigripes*). *Journal of Mammalogy*, 92(4), 751–759.
- Campbell, K. H., McWhir, J., Ritchie, W. A., & Wilmut, I. (1996). Sheep cloned by nuclear transfer from a cultured cell line. *Nature*, 380(6569), 64–66.
- Cano, M. J., Twyford, A. D., & Hollingsworth, P. M. (2022). Population and conservation

- genetics using RAD sequencing in four endemic conifers from South America. *Biodiversity and Conservation*, 31(13-14), 3093–3112.
- Carpenter, J. W., Novilla, M., & Kaiser, H. E. (1981). *Neoplasia and other disease problems in black-footed ferrets: Implications for an endangered species*. 739–746.
- Carranza, J., Pérez-González, J., Anaya, G., de Jong, M., Broggini, C., Zachos, F. E., McDevitt, A. D., Niedziałkowska, M., Sykut, M., Csányi, S., Bleier, N., Csirke, L., Røed, K., Saint-Andrieux, C., Barboiron, A., Gort-Esteve, A., Ruiz-Olmo, J., Seoane, J. M., Godoy, J. A., ... Membrillo, A. (2024). Genome-wide SNP assessment of contemporary European red deer genetic structure highlights the distinction of peripheral populations and the main admixture zones in Europe. *Molecular Ecology*, e17508.
- Castellano, D., Eyre-Walker, A., & Munch, K. (2020). Impact of mutation rate and selection at linked sites on DNA variation across the genomes of humans and other homininae. *Genome Biology and Evolution*, 12(1), 3550–3561.
- Castro-Mondragon, J. A., Riudavets-Puig, R., Rauluseviciute, I., Lemma, R. B., Turchi, L., Blanc-Mathieu, R., Lucas, J., Boddie, P., Khan, A., Manosalva Pérez, N., Fornes, O., Leung, T. Y., Aguirre, A., Hammal, F., Schmelter, D., Baranasic, D., Ballester, B., Sandelin, A., Lenhard, B., ... Mathelier, A. (2022). JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Research*, 50(D1), D165–D173.
- Caughley, G. (1994). Directions in conservation biology. *Journal of Animal Ecology*, 63(2), 215–244.
- Ceballos, F. C., Hazelhurst, S., & Ramsay, M. (2018). Assessing runs of Homozygosity: a comparison of SNP Array and whole genome sequence low coverage data. *BMC Genomics*, 19(1), 106.
- Ceballos, F. C., Joshi, P. K., Clark, D. W., Ramsay, M., & Wilson, J. F. (2018). Runs of homozygosity: windows into population history and trait architecture. *Nature Reviews. Genetics*, 19(4), 220–234.
- Ceballos, G., Ehrlich, P. R., & Raven, P. H. (2020). Vertebrates on the brink as indicators

- of biological annihilation and the sixth mass extinction. *Proceedings of the National Academy of Sciences of the United States of America*, 117(24), 13596–13602.
- Challis, R., Kumar, S., Sotero-Caio, C., Brown, M., & Blaxter, M. (2023). Genomes on a Tree (GoaT): A versatile, scalable search engine for genomic and sequencing project metadata across the eukaryotic tree of life. *Wellcome Open Research*, 8, 24.
- Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, 4(1), 7.
- Charlesworth, B. (2018). Mutational load, inbreeding depression and heterosis in subdivided populations. *Molecular Ecology*, 27(24), 4991–5003.
- Charlesworth, B. (2022). The effects of weak selection on neutral diversity at linked sites. *Genetics*, 221(1), iyac027.
- Charlesworth, D., & Charlesworth, B. (1987). Inbreeding depression and its evolutionary consequences. *Annual Review of Ecology and Systematics*, 18(1), 237–268.
- Chen, M., Liu, X., Liu, Q., Shi, D., & Li, H. (2023). 3D genomics and its applications in precision medicine. *Cellular & Molecular Biology Letters*, 28(1), 19.
- Christiansen, H., Heindler, F. M., Hellemans, B., Jossart, Q., Pasotti, F., Robert, H., Verheye, M., Danis, B., Kochzius, M., Leliaert, F., Moreau, C., Patel, T., Van de Putte, A. P., Vanreusel, A., Volckaert, F. A. M., & Schön, I. (2021). Facilitating population genomics of non-model organisms through optimized experimental design for reduced representation sequencing. *BMC Genomics*, 22(1), 625.
- Christmas, M. J., Kaplow, I. M., Genereux, D. P., Dong, M. X., Hughes, G. M., Li, X., Sullivan, P. F., Hindle, A. G., Andrews, G., Armstrong, J. C., Bianchi, M., Breit, A. M., Diekhans, M., Fanter, C., Foley, N. M., Goodman, D. B., Goodman, L., Keough, K. C., Kirilenko, B., ... Karlsson, E. K. (2023). Evolutionary constraint and innovation across hundreds of placental mammals. *Science (New York, N.Y.)*, 380(6643), eabn3943.
- Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., & Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster*

- strain w1118; iso-2; iso-3: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*, 6(2), 80–92.
- Cisternas-Fuentes, A., & Koski, M. H. (2023). Drivers of strong isolation and small effective population size at a leading range edge of a widespread plant. *Heredity*, 130(6), 347–357.
- Claringbould, A., & Zaugg, J. B. (2021). Enhancers in disease: molecular basis and emerging treatment strategies. *Trends in Molecular Medicine*, 27(11), 1060–1073.
- Clout, M. N., & Merton, D. V. (1998). Saving the Kakapo: the conservation of the world's most peculiar parrot. *Bird Conservation International*, 8(3), 281–296.
- Coggins, G. C. (1974). Conserving wildlife resources: An overview of the endangered species act of 1973. *NDL Rev.*
- <https://commons.und.edu/cgi/viewcontent.cgi?article=2921&context=ndlr>
- Cohn, J. P. (1999). Saving the California condor. *Bioscience*, 49(11), 864–868.
- Colpitts, J., McLoughlin, P. D., & Poissant, J. (2022). Runs of homozygosity in Sable Island feral horses reveal the genomic consequences of inbreeding and divergence from domestic breeds. *BMC Genomics*, 23(1), 501.
- Cooper, G. M., Stone, E. A., Asimenos, G., NISC Comparative Sequencing Program, Green, E. D., Batzoglou, S., & Sidow, A. (2005). Distribution and intensity of constraint in mammalian genomic sequence. *Genome Research*, 15(7), 901–913.
- Costa, G. M. J., Chiarini-Garcia, H., Morato, R. G., Alvarenga, R. L. L. S., & França, L. R. (2008). Duration of spermatogenesis and daily sperm production in the jaguar (*Panthera onca*). *Theriogenology*, 70(7), 1136–1146.
- Costa, M., Fernandes, C., Birks, J. D. S., Kitchener, A. C., Santos-Reis, M., & Bruford, M. W. (2013). The genetic legacy of the 19th-century decline of the British polecat: evidence for extensive introgression from feral ferrets. *Molecular Ecology*, 22(20), 5130–5147.
- Cowie, R. H., Bouchet, P., & Fontaine, B. (2022). The Sixth Mass Extinction: fact, fiction or speculation? *Biological Reviews of the Cambridge Philosophical Society*, 97(2), 640–663.

- Cronin, M., Palmisciano, D. A., Vyse, E., & Cameron, D. (1991). Mitochondrial DNA in wildlife forensic science: Species identification of tissues. *Wildlife Society Bulletin*, 19(1), 94–105.
- Croose, E. (2016). *The Distribution and Status of the Polecat (Mustela putorius) in Britain 2014-2015*.
<https://www.vwt.org.uk/wp-content/uploads/2016/04/Polecat-Report-2016.pdf>
- Croose, E., Duckworth, J. W., Ruetten, S., Skumatov, D. V., Kolesnikov, V. V., & Saveljev, A. P. (2018). A review of the status of the Western polecat *Mustela putorius*: a neglected and declining species? *Mammalia*, 82(6), 550–564.
- Crouse, D. T., Mehrhoff, L. A., Parkin, M. J., Elam, D. R., & Chen, L. Y. (2002). ENDANGERED SPECIES RECOVERY AND THE SCB STUDY: A U.S. FISH AND WILDLIFE SERVICE PERSPECTIVE. *Ecological Applications*, 12(3), 719–723.
- Crow, J. F. (2017). *An introduction to population genetics theory*.
https://books.google.com/books?hl=en&lr=&id=mpVEEAAAQBAJ&oi=fnd&pg=PP1&ots=KnfuRYk_r6&sig=6yxSDgx1FlggVfUyIWoaIoj7p9s
- Cserkés, T., Kiss, C., Barkaszi, Z., Görföl, T., Zagorodniuk, I., Sramkó, G., & Csorba, G. (2021). Intra- and interspecific morphological variation in sympatric and allopatric populations of *Mustela putorius* and *M. eversmanii* (Carnivora: Mustelidae) and detection of potential hybrids. *Mammal Research*, 66(1), 103–114.
- Cummins, M., Watson, C., Edwards, R. J., & Mattick, J. S. (2024). The evolution of ultraconserved elements in vertebrates. *Molecular Biology and Evolution*, 41(7), msae146.
- Damas, J., Hughes, G. M., Keough, K. C., Painter, C. A., Persky, N. S., Corbo, M., Hiller, M., Koepfli, K.-P., Pfenning, A. R., Zhao, H., Genereux, D. P., Swofford, R., Pollard, K. S., Ryder, O. A., Nweeia, M. T., Lindblad-Toh, K., Teeling, E. C., Karlsson, E. K., & Lewin, H. A. (2020). Broad host range of SARS-CoV-2 predicted by comparative and structural analysis of ACE2 in vertebrates. *Proceedings of the National Academy of Sciences of the United States of America*, 117(36), 22311–22322.
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A.,

- Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics (Oxford, England)*, 27(15), 2156–2158.
- Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, 10(2), giab008.
- Daniels, M. J., Beaumont, M. A., Johnson, P. J., Balharry, D., Macdonald, D. W., & Barratt, E. (2001). Ecology and genetics of wild-living cats in the north-east of Scotland and the implications for the conservation of the wildcat: *Wildcat ecology and genetics*. *The Journal of Applied Ecology*, 38(1), 146–161.
- Davidson, A. D., Augustine, D. J., Jacobsen, H., Pellatz, D., Porensky, L. M., McKee, G., & Duchardt, C. (2022). Boom and bust cycles of black-tailed prairie dog populations in the Thunder Basin grassland ecosystem. *Journal of Mammalogy*, 103(5), 1112–1126.
- Davidson, A. D., Detling, J. K., & Brown, J. H. (2012). Ecological roles and conservation challenges of social, burrowing, herbivorous mammals in the world's grasslands. *Frontiers in Ecology and the Environment*, 10(9), 477–486.
- Davidson, A. D., Fink, M., Menefee, M., Sterling-Krank, L., Van Pelt, W., & Augustine, D. J. (2023). Present and future suitable habitat for the black-tailed prairie dog ecosystem. *Biological Conservation*, 286(110241), 110241.
- Davison, A., Birks, J. D. S., Griffiths, H. I., Kitchener, A. C., Biggins, D., & Butlin, R. K. (1999). Hybridization and the phylogenetic relationship between polecats and domestic ferrets in Britain. *Biological Conservation*, 87(2), 155–161.
- Davydov, E. V., Goode, D. L., Sirota, M., Cooper, G. M., Sidow, A., & Batzoglou, S. (2010). Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Computational Biology*, 6(12), e1001025.
- Deem, S. L., Spelman, L. H., Yates, R. A., & Montali, R. J. (2000). Canine distemper in terrestrial carnivores: a review. *Journal of Zoo and Wildlife Medicine: Official Publication of the American Association of Zoo Veterinarians*, 31(4), 441–451.

- DeFaveri, J., Viitaniemi, H., Leder, E., & Merilä, J. (2013). Characterizing genic and nongenic molecular markers: comparison of microsatellites and SNPs. *Molecular Ecology Resources*, 13(3), 377–392.
- de Jong, M. J., Li, Z., Qin, Y., Quéméré, E., Baker, K., Wang, W., & Hoelzel, A. R. (2020). Demography and adaptation promoting evolutionary transitions in a mammalian genus that diversified during the Pleistocene. *Molecular Ecology*, 29(15), 2777–2792.
- Delage, L., Carbone, F., Riller, Q., Zacharyus, J.-L., Kerbellec, E., Buzy, A., Stolzenberg, M.-C., Luka, M., de Cevins, C., Kalouche, G., Favier, R., Michel, A., Meynier, S., Corneau, A., Evrard, C., Neveux, N., Roudières, S., Pérot, B. P., Fusaro, M., ... Rieux-Laucat, F. (2023). NBEAL2 deficiency in humans leads to low CTLA-4 expression in activated conventional T cells. *Nature Communications*, 14(1), 3728.
- Delgado-Acevedo, J., Zamorano, A., DeYoung, R. W., & Campbell, T. A. (2021). Genetic population structure of wild pigs in southern Texas. *Animals: An Open Access Journal from MDPI*, 11(1), 168.
- de Manuel, M., Kuhlwilm, M., Frandsen, P., Sousa, V. C., Desai, T., Prado-Martinez, J., Hernandez-Rodriguez, J., Dupanloup, I., Lao, O., Hallast, P., Schmidt, J. M., Heredia-Genestar, J. M., Benazzo, A., Barbujani, G., Peter, B. M., Kuderna, L. F. K., Casals, F., Angedakin, S., Arandjelovic, M., ... Marques-Bonet, T. (2016). Chimpanzee genomic diversity reveals ancient admixture with bonobos. *Science (New York, N.Y.)*, 354(6311), 477–481.
- Dennis, G., Jr, Sherman, B. T., Hosack, D. A., Yang, J., Gao, W., Lane, H. C., & Lempicki, R. A. (2003). DAVID: Database for annotation, visualization, and integrated discovery. *Genome Biology*, 4(9), R60.
- de Oliveira, M. E., Gonçalves, F., & Galetti, P. M., Jr. (2025). Scat DNA applications for low-density carnivore survey: Techniques, efficiency and future directions. *Mammal Review*, e70007.
- Der Sarkissian, C., Ermini, L., Schubert, M., Yang, M. A., Librado, P., Fumagalli, M., Jónsson, H., Bar-Gal, G. K., Albrechtsen, A., Vieira, F. G., Petersen, B., Ginolhac, A., Seguin-Orlando, A., Magnussen, K., Fages, A., Gamba, C., Lorente-Galdos, B.,

- Polani, S., Steiner, C., ... Orlando, L. (2015). Evolutionary genomics and conservation of the endangered przewalski's horse. *Current Biology: CB*, 25(19), 2577–2583.
- DeWoody, J. A., Harder, A. M., Mathur, S., & Willoughby, J. R. (2021). The long-standing significance of genetic diversity in conservation. *Molecular Ecology*, 30(17), 4147–4154.
- Díaz, S., Settele, J., Brondízio, E. S., Ngo, H. T., Agard, J., Arneth, A., Balvanera, P., Brauman, K. A., Butchart, S. H. M., Chan, K. M. A., Garibaldi, L. A., Ichii, K., Liu, J., Subramanian, S. M., Midgley, G. F., Miloslavich, P., Molnár, Z., Obura, D., Pfaff, A., ... Zayas, C. N. (2019). Pervasive human-driven decline of life on Earth points to the need for transformative change. *Science (New York, N.Y.)*, 366(6471), eaax3100.
- Di Marco, M., Pacifici, M., Santini, L., Baisero, D., Francucci, L., Grottolo Marasini, G., Visconti, P., & Rondinini, C. (2013). Generation length for mammals. *Nature Conservation*, 5, 89–94.
- Dobrynin, P., Liu, S., Tamazian, G., Xiong, Z., Yurchenko, A. A., Krasheninnikova, K., Kliver, S., Schmidt-Küntzel, A., Koepfli, K.-P., Johnson, W., Kuderna, L. F. K., García-Pérez, R., Manuel, M. de, Godinez, R., Komissarov, A., Makunin, A., Brukhin, V., Qiu, W., Zhou, L., ... O'Brien, S. J. (2015). Genomic legacy of the African cheetah, *Acinonyx jubatus*. *Genome Biology*, 16, 277.
- Dobson, A., & Lyles, A. (2000). Black-footed ferret recovery. *Science (New York, N.Y.)*, 288(5468), 985–988.
- Dodds, W. K., Gido, K., Whiles, M. R., Fritz, K. M., & Matthews, W. J. (2004). Life on the Edge: The Ecology of Great Plains Prairie Streams. *BioScience*, 54(3), 205–216.
- Driscoll, C., Yamaguchi, N., O'Brien, S. J., & Macdonald, D. W. (2011). A suite of genetic markers useful in assessing wildcat (*Felis silvestris* ssp.)-domestic cat (*Felis silvestris catus*) admixture. *The Journal of Heredity*, 102 Suppl 1(Suppl 1), S87–S90.
- Dueñas, M.-A., Hemming, D. J., Roberts, A., & Diaz-Soltero, H. (2021). The threat of invasive species to IUCN-listed critically endangered species: A systematic review. *Global Ecology and Conservation*, 26(e01476), e01476.

- Duntsch, L., Whibley, A., Brekke, P., Ewen, J. G., & Santure, A. W. (2021). Genomic data of different resolutions reveal consistent inbreeding estimates but contrasting homozygosity landscapes for the threatened Aotearoa New Zealand hihi. *Molecular Ecology*, 30(23), 6006–6020.
- du Plessis, S. J., Blaxter, M., Koepfli, K.-P., Chadwick, E. A., & Hailer, F. (2023). Genomics Reveals Complex Population History and Unexpected Diversity of Eurasian Otters (*Lutra lutra*) in Britain Relative to Genetic Methods. *Molecular Biology and Evolution*, 40(11). <https://doi.org/10.1093/molbev/msad207>
- du Plessis, S. J., Martins, N. F., Sun, X., Mak, S. S. T., Stanton, D. W. G., Koepfli, K.-P., Chadwick, E. A., Gilbert, M. P. T., & Hailer, F. (2025). Historical genomic variation of Eurasian otters (*Lutra lutra*) in Britain, from hunting trophies. *Conservation Genetics Resources*, 17(2), 101–110.
- Dussex, N., Morales, H. E., Grossen, C., Dalén, L., & van Oosterhout, C. (2023). Purging and accumulation of genetic load in conservation. *Trends in Ecology & Evolution*, 38(10), 961–969.
- Dussex, N., van der Valk, T., Morales, H. E., Wheat, C. W., Díez-Del-Molino, D., von Seth, J., Foster, Y., Kutschera, V. E., Guschanski, K., Rhie, A., Phillippy, A. M., Korlach, J., Howe, K., Chow, W., Pelan, S., Mendes Damas, J. D., Lewin, H. A., Hastie, A. R., Formenti, G., ... Dalén, L. (2021). Population genomics of the critically endangered kākāpō. *Cell Genomics*, 1(1), 100002.
- Dwivedi, S. L., Heslop-Harrison, P., Spillane, C., McKeown, P. C., Edwards, D., Goldman, I., & Ortiz, R. (2023). Evolutionary dynamics and adaptive benefits of deleterious mutations in crop gene pools. *Trends in Plant Science*, 28(6), 685–697.
- EAZA Biobank. (n.d.). Retrieved August 5, 2025, from <https://www.eaza.net/biobank/>
- Eddy, S. R. (2005). A model of the statistical power of comparative genome sequence analysis. *PLoS Biology*, 3(1), e10.
- Edmands, S. (2007). Between a rock and a hard place: evaluating the relative risks of inbreeding and outbreeding for conservation and management: RELATIVE RISKS OF INBREEDING AND OUTBREEDING. *Molecular Ecology*, 16(3), 463–475.

- Eeraerts, M., Verschelde, P., Gouwy, J., Van Den Berge, S., & Van Den Berge, K. (2022). Shifting dietary habits linked with contrasting population trends of two sympatric mustelids (*Martes foina* and *Mustela putorius*) over 16 years. *Global Ecology and Conservation*, 35(e02108), e02108.
- Ekblom, R., Brechlin, B., Persson, J., Smeds, L., Johansson, M., Magnusson, J., Flagstad, Ø., & Ellegren, H. (2018). Genome sequencing and conservation genomics in the Scandinavian wolverine population. *Conservation Biology: The Journal of the Society for Conservation Biology*, 32(6), 1301–1312.
- Ellegren, H. (2014). Genome sequencing and population genomics in non-model organisms. *Trends in Ecology & Evolution*, 29(1), 51–63.
- Ellegren, H., Smith, N. G. C., & Webster, M. T. (2003). Mutation rate variation in the mammalian genome. *Current Opinion in Genetics & Development*, 13(6), 562–568.
- Ellingford, J. M., Ahn, J. W., Bagnall, R. D., Baralle, D., Barton, S., Campbell, C., Downes, K., Ellard, S., Duff-Farrier, C., FitzPatrick, D. R., Greally, J. M., Ingles, J., Krishnan, N., Lord, J., Martin, H. C., Newman, W. G., O'Donnell-Luria, A., Ramsden, S. C., Rehm, H. L., ... Whiffin, N. (2022). Recommendations for clinical interpretation of variants found in non-coding regions of the genome. *Genome Medicine*, 14(1), 73.
- ENCODE Project Consortium. (2004). The ENCODE (ENCyclopedia Of DNA elements) Project. *Science (New York, N.Y.)*, 306(5696), 636–640.
- ENCODE Project Consortium, Birney, E., Stamatoyannopoulos, J. A., Dutta, A., Guigó, R., Gingeras, T. R., Margulies, E. H., Weng, Z., Snyder, M., Dermitzakis, E. T., Thurman, R. E., Kuehn, M. S., Taylor, C. M., Neph, S., Koch, C. M., Asthana, S., Malhotra, A., Adzhubei, I., Greenbaum, J. A., ... de Jong, P. J. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, 447(7146), 799–816.
- Eppig, J. (2017). Mouse genome informatics (MGI) resource: Genetic, genomic, and biological knowledgebase for the laboratory mouse. *ILAR Journal*, 58, 17–41.
- Etherington, G. J., Ciezarek, A., Shaw, R., Michaux, J., Croose, E., Haerty, W., & Di Palma, F. (2022). Extensive genome introgression between domestic ferret and

- European polecat during population recovery in Great Britain. *The Journal of Heredity*, 113(5), 500–515.
- Etherington, G. J., Heavens, D., Baker, D., Lister, A., McNelly, R., Garcia, G., Clavijo, B., Macaulay, I., Haerty, W., & Di Palma, F. (2020). Sequencing smart: De novo sequencing and assembly approaches for a non-model mammal. *GigaScience*, 9(5), giaa045.
- Excoffier, L., & Ray, N. (2008). Surfing during population expansions promotes genetic revolutions and structuration. *Trends in Ecology & Evolution*, 23(7), 347–351.
- Eyre-Walker, A., & Keightley, P. D. (2007). The distribution of fitness effects of new mutations. *Nature Reviews. Genetics*, 8(8), 610–618.
- Eyre-Walker, A., & Keightley, P. D. (2009). Estimating the rate of adaptive molecular evolution in the presence of slightly deleterious mutations and population size change. *Molecular Biology and Evolution*, 26(9), 2097–2108.
- Fagre, A. C., Cohen, L. E., Eskew, E. A., Farrell, M., Glennon, E., Joseph, M. B., Frank, H. K., Ryan, S. J., Carlson, C. J., & Albery, G. F. (2022). Assessing the risk of human-to-wildlife pathogen transmission for conservation and public health. *Ecology Letters*, 25(6), 1534–1549.
- Faircloth, B. C., McCormack, J. E., Crawford, N. G., Harvey, M. G., Brumfield, R. T., & Glenn, T. C. (2012). Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Systematic Biology*, 61(5), 717–726.
- Fang, B., & Edwards, S. V. (2024). Fitness consequences of structural variation inferred from a House Finch pangenome. *Proceedings of the National Academy of Sciences of the United States of America*, 121(47), e2409943121.
- Farquharson, K. A., Hogg, C. J., & Grueber, C. E. (2021). Offspring survival changes over generations of captive breeding. *Nature Communications*, 12(1), 3045.
- Farquharson, K. A., McLennan, E. A., Cheng, Y., Alexander, L., Fox, S., Lee, A. V., Belov, K., & Hogg, C. J. (2022). Restoring faith in conservation action: Maintaining wild genetic diversity through the Tasmanian devil insurance program. *iScience*, 25(7), 104474.

- Feng, S., Fang, Q., Barnett, R., Li, C., Han, S., Kuhlilm, M., Zhou, L., Pan, H., Deng, Y., Chen, G., Gamauf, A., Woog, F., Prys-Jones, R., Marques-Bonet, T., Gilbert, M. T. P., & Zhang, G. (2019). The genomic footprints of the fall and recovery of the crested ibis. *Current Biology: CB*, 29(2), 340–349.e7.
- Feng, S., Stiller, J., Deng, Y., Armstrong, J., Fang, Q., Reeve, A. H., Xie, D., Chen, G., Guo, C., Faircloth, B. C., Petersen, B., Wang, Z., Zhou, Q., Diekhans, M., Chen, W., Andreu-Sánchez, S., Margaryan, A., Howard, J. T., Parent, C., ... Zhang, G. (2020). Dense sampling of bird diversity increases power of comparative genomics. *Nature*, 587(7833), 252–257.
- Feron, R., & Waterhouse, R. M. (2022). Assessing species coverage and assembly quality of rapidly accumulating sequenced genomes. *GigaScience*, 11.
<https://doi.org/10.1093/gigascience/giac006>
- Figueiró, H. V., Li, G., Trindade, F. J., Assis, J., Pais, F., Fernandes, G., Santos, S. H. D., Hughes, G. M., Komissarov, A., Antunes, A., Trinca, C. S., Rodrigues, M. R., Linderth, T., Bi, K., Silveira, L., Azevedo, F. C. C., Kantek, D., Ramalho, E., Brassaloti, R. A., ... Eizirik, E. (2017). Genome-wide signatures of complex introgression and adaptive evolution in the big cats. *Science Advances*, 3(7), e1700299.
- Fitzpatrick, S. W., Bradburd, G. S., Kremer, C. T., Salerno, P. E., Angeloni, L. M., & Funk, W. C. (2020). Genomic and fitness consequences of genetic rescue in wild populations. *Current Biology: CB*, 30(3), 517–522.e5.
- Fontsere, C., Speak, S. A., Caven, A. J., Rodríguez, J. A., Wang, X., Pacheco, C., Cassatt-Johnstone, M., Femerling, G., Maloney, B., Balacco, J., Collins, J., Sims, Y., Abueg, L., Fedrigo, O., Jarvis, E. D., Hartup, B. K., Shapiro, B., Gilbert, M. T. P., van Oosterhout, C., & Morales, H. E. (2024). Persistent genomic erosion in whooping cranes despite demographic recovery. In *bioRxiv* (p. 2024.12.12.628160).
<https://doi.org/10.1101/2024.12.12.628160>
- Formenti, G., Theissinger, K., Fernandes, C., Bista, I., Bombarely, A., Bleidorn, C., Ciofi, C., Crottini, A., Godoy, J. A., Höglund, J., Malukiewicz, J., Mouton, A., Oomen, R. A.,

- Paez, S., Palsbøll, P. J., Pampoulie, C., Ruiz-López, M. J., Svardal, H., Theofanopoulou, C., ... European Reference Genome Atlas (ERGA) Consortium. (2022). The era of reference genomes in conservation genomics. *Trends in Ecology & Evolution*, 37(3), 197–202.
- Forrest, S. C., Biggins, D. E., Richardson, L., Clark, T. W., Campbell, T. M., Fagerstone, K. A., & Thorne, E. T. (1988). Population attributes for the black-footed ferret (*Mustela nigripes*) at Meeteetse, Wyoming, 1981-1985. *Journal of Mammalogy*, 69(2), 261–273.
- Forutan, M., Engle, B. N., Chamberlain, A. J., Ross, E. M., Nguyen, L. T., D'Occhio, M. J., Snr, A. C., Kho, E. A., Fordyce, G., Speight, S., Goddard, M. E., & Hayes, B. J. (2024). Genome-wide association and expression quantitative trait loci in cattle reveals common genes regulating mammalian fertility. *Communications Biology*, 7(1), 724.
- Foster, Y., Dutoit, L., Grosser, S., Dussex, N., Foster, B. J., Dodds, K. G., Brauning, R., Van Stijn, T., Robertson, F., McEwan, J. C., Jacobs, J. M. E., & Robertson, B. C. (2021). Genomic signatures of inbreeding in a critically endangered parrot, the kākāpō. *G3 (Bethesda, Md.)*, 11(11), jkab307.
- Fraik, A. K., Margres, M. J., Epstein, B., Barbosa, S., Jones, M., Hendricks, S., Schönfeld, B., Stahlke, A. R., Veillet, A., Hamede, R., McCallum, H., Lopez-Contreras, E., Kallinen, S. J., Hohenlohe, P. A., Kelley, J. L., & Storfer, A. (2020). Disease swamps molecular signatures of genetic-environmental associations to abiotic factors in Tasmanian devil (*Sarcophilus harrisii*) populations. *Evolution; International Journal of Organic Evolution*, 74(7), 1392–1408.
- Frankel, O. H., Soulé, M. E., & Others. (1981). *Conservation and evolution*.
<https://www.sidalc.net/search/Record/KOHA-OAI-ECOSUR:60234/Description>
- Frankham, R. (2005). Genetics and extinction. *Biological Conservation*, 126(2), 131–140.
- Frankham, R. (2010). Challenges and opportunities of genetic approaches to biological conservation. *Biological Conservation*, 143(9), 1919–1927.
- Frankham, R. (2015). Genetic rescue of small inbred populations: meta-analysis reveals

- large and consistent benefits of gene flow. *Molecular Ecology*, 24(11), 2610–2618.
- Frankham, R., Ballou, J. D., & Briscoe, D. A. (2004). *A primer of conservation genetics*.
<https://books.google.com/books?hl=en&lr=&id=vvAfAwAAQBAJ&oi=fnd&pg=PR9&ots=LGVOHReEpb&sig=xmTCouv92zX-Xm1u7iKBf7wybV8>
- Frankham, R., Ballou, J. D., Eldridge, M. D. B., Lacy, R. C., Ralls, K., Dudash, M. R., & Fenster, C. B. (2011). Predicting the probability of outbreeding depression: Predicting outbreeding depression. *Conservation Biology: The Journal of the Society for Conservation Biology*, 25(3), 465–475.
- Frankham, R., Ballou, J. D., Ralls, K., Eldridge, M. D. B., Dudash, M. R., Fenster, C. B., Lacy, R. C., & Sunnucks, P. (2017). *Loss of genetic diversity reduces ability to adapt*. Oxford University Press.
- Frankham, R., Ballou, J. D., Ralls, K., Eldridge, M. D. B., Dudash, M. R., Fenster, C. B., Lacy, R. C., & Sunnucks, P. (2019). Managing gene flow among isolated population fragments. In *A Practical Guide for Genetic Management of Fragmented Animal and Plant Populations* (pp. 115–134). Oxford University Press.
- Frankham, R., Briscoe, D. A., & Ballou, J. D. (2002). *Introduction to conservation genetics*.
https://books.google.com/books?hl=en&lr=&id=F-XB8hqZ4s8C&oi=fnd&pg=PR13&ots=CNEfw_aK2M&sig=785W2AjquurWmSFbFwtpuvuOFJc
- Fritts, R. (2022, January 13). *In a conservation first, a cloned ferret could help save her species*. American Association for the Advancement of Science.
<https://www.science.org/content/article/conservation-first-cloned-ferret-could-help-save-her-species>
- Gaczorek, T., Dudek, K., Fritz, U., Bahri-Sfar, L., Baird, S. J. E., Bonhomme, F., Dufresnes, C., Gvoždík, V., Irwin, D., Kotlík, P., Marková, S., McGinnity, P., Migalska, M., Moravec, J., Natola, L., Pabijan, M., Phillips, K. P., Schöneberg, Y., Souissi, A., ... Babik, W. (2024). Widespread adaptive introgression of major histocompatibility complex genes across vertebrate hybrid zones. *Molecular Biology and Evolution*, 41(10), msae201.
- Gajdárová, B., Belotti, E., Bufka, L., Volfová, J., Wölfl, S., Mináriková, T., Hollerbach, L.,

- Duľa, M., Kleven, O., Kotal, M., Nowak, C., Ozoliņš, J., Tām, B., Bryja, J., Koubek, P., & Krojerová-Prokešová, J. (2023). Long-term genetic monitoring of a reintroduced Eurasian lynx population does not indicate an ongoing loss of genetic diversity. *Global Ecology and Conservation*, 42(e02399), e02399.
- Galla, S. J., Brown, L., Couch-Lewis Ngāi Tahu Te Hapū O Ngāti Wheke Ngāti Waewae, Y., Cubrinovska, I., Eason, D., Gooley, R. M., Hamilton, J. A., Heath, J. A., Hauser, S. S., Latch, E. K., Matocq, M. D., Richardson, A., Wold, J. R., Hogg, C. J., Santure, A. W., & Steeves, T. E. (2022). The relevance of pedigrees in the conservation genomics era. *Molecular Ecology*, 31(1), 41–54.
- Galtier, N., Nabholz, B., Glémin, S., & Hurst, G. D. D. (2009). Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Molecular Ecology*, 18(22), 4541–4550.
- García, D., Carleos, C., Parra, D., & Cañón, J. (2002). Sib-parentage testing using molecular markers when parents are unknown: Sib-parentage testing. *Animal Genetics*, 33(5), 364–371.
- García-Díaz, P., Montti, L., Powell, P. A., Phimister, E., Pizarro, J. C., Fasola, L., Langdon, B., Pauchard, A., Raffo, E., Bastías, J., Damasceno, G., Fidelis, A., Huerta, M. F., Linardaki, E., Moyano, J., Núñez, M. A., Ortiz, M. I., Rodríguez-Jorquera, I., Roesler, I., ... Lambin, X. (2022). Identifying priorities, targets, and actions for the long-term social and ecological management of invasive non-native species. *Environmental Management*, 69(1), 140–153.
- García-Dorado, A. (2012). Understanding and predicting the fitness decline of shrunk populations: inbreeding, purging, mutation, and standard selection. *Genetics*, 190(4), 1461–1476.
- Garelle, D., Marinari, P., & Lynch, C. (2013). Population analysis and breeding and transfer plan ferret-black footed (*Mustela nigripes*). *Chicago: Association of Zoos and Aquariums*.
- Gazal, S., Loh, P.-R., Finucane, H. K., Ganna, A., Schoech, A., Sunyaev, S., & Price, A. L. (2018). Functional architecture of low-frequency variants highlights strength of negative selection across coding and non-coding annotations. *Nature Genetics*,

50(11), 1600–1607.

- Gedman, G., Morrill Pirovich, K., Oppenheimer, J., Hyseni, C., Cassatt-Johnstone, M., Alexandre, N., Troy, W., Chao, C., Fedrigo, O., Hoyt, S. J., Grady, P. G. S., Sacco, S., Seligmann, W., Dash, A., Chokshi, M., Knecht, L., Papizan, J. B., Miyawaki, T., Bocklandt, S., ... Shapiro, B. (2025). On the ancestry and evolution of the extinct dire wolf. In *bioRxiv* (p. 2025.04.09.647074). <https://doi.org/10.1101/2025.04.09.647074>
- Ghanbari, M., Ikram, M. A., de Looper, H. W. J., Hofman, A., Erkeland, S. J., Franco, O. H., & Dehghan, A. (2016). Genome-wide identification of microRNA-related variants associated with risk of Alzheimer's disease. *Scientific Reports*, 6(1), 28387.
- Gibson, J., Morton, N. E., & Collins, A. (2006). Extended tracts of homozygosity in outbred human populations. *Human Molecular Genetics*, 15(5), 789–795.
- Gilpin, M. E., & Soule, M. E. (1986). Conservation Biology: The Science of Scarcity and Diversity. Sinsuer Associates. Inc, Sunderland, Massachusetts, USA.
- Giuffra, E., Tuggle, C. K., & FAANG Consortium. (2019). Functional Annotation of Animal Genomes (FAANG): Current achievements and roadmap. *Annual Review of Animal Biosciences*, 7(1), 65–88.
- Graciá, E., Botella, F., Anadón, J. D., Edelaar, P., Harris, D. J., & Giménez, A. (2013). Surfing in tortoises? Empirical signs of genetic structuring owing to range expansion. *Biology Letters*, 9(3), 20121091.
- Grant, C. E., Bailey, T. L., & Noble, W. S. (2011). FIMO: scanning for occurrences of a given motif. *Bioinformatics (Oxford, England)*, 27(7), 1017–1018.
- Grant, P. R., & Grant, B. R. (1994). Phenotypic and genetic effects of hybridization in Darwin's finches. *Evolution; International Journal of Organic Evolution*, 48(2), 297–316.
- Grant, P. R., Grant, B. R., Keller, L. F., Markert, J. A., & Petren, K. (2003). Inbreeding and interbreeding in Darwin's finches. *Evolution; International Journal of Organic Evolution*, 57(12), 2911–2916.
- Greenwood, J. R., Zhang, X., & Rathjen, J. P. (2023). Precision genome editing of crops for improved disease resistance. *Current Biology: CB*, 33(11), R650–R657.

- Gregory, T. R. (2009). Understanding natural selection: Essential concepts and common misconceptions. *Evolution*, 2(2), 156–175.
- Groß, C., Bortoluzzi, C., de Ridder, D., Megens, H.-J., Groenen, M. A. M., Reinders, M., & Bosse, M. (2020a). Evolutionarily conserved non-protein-coding regions in the chicken genome harbor functionally important variation. In *bioRxiv*. bioRxiv. <https://doi.org/10.1101/2020.03.27.012005>
- Groß, C., Bortoluzzi, C., de Ridder, D., Megens, H.-J., Groenen, M. A. M., Reinders, M., & Bosse, M. (2020b). Prioritizing sequence variants in conserved non-coding elements in the chicken genome using chCADD. *PLoS Genetics*, 16(9), e1009027.
- Groß, C., Derks, M., Megens, H.-J., Bosse, M., Groenen, M. A. M., Reinders, M., & de Ridder, D. (2020). pCADD: SNV prioritisation in *Sus scrofa*. *Genetics, Selection, Evolution: GSE*, 52(1), 4.
- Grossen, C., Biebach, I., Angelone-Alasaad, S., Keller, L. F., & Croll, D. (2018). Population genomics analyses of European ibex species show lower diversity and higher inbreeding in reintroduced populations. *Evolutionary Applications*, 11(2), 123–139.
- Grossen, C., Guillaume, F., Keller, L. F., & Croll, D. (2020). Purging of highly deleterious mutations through severe bottlenecks in Alpine ibex. *Nature Communications*, 11(1), 1001.
- Grueber, C. E. (2015). Comparative genomics for biodiversity conservation. *Computational and Structural Biotechnology Journal*, 13, 370–375.
- GTEx Consortium. (2013). The Genotype-Tissue Expression (GTEx) project. *Nature Genetics*, 45(6), 580–585.
- GTEx Consortium. (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science (New York, N.Y.)*, 369(6509), 1318–1330.
- Gustafson, K. D., Hawkins, M. G., Drazenovich, T. L., Church, R., Brown, S. A., & Ernest, H. B. (2018). Founder events, isolation, and inbreeding: Intercontinental genetic structure of the domestic ferret. *Evolutionary Applications*, 11(5), 694–704.
- Hansson, B., & Westerberg, L. (2002). On the correlation between heterozygosity and fitness in natural populations: HETEROZYGOSITY-FITNESS CORRELATIONS.

Molecular Ecology, 11(12), 2467–2474.

Harrison, R. G., & Larson, E. L. (2016). Heterogeneous genome divergence, differential introgression, and the origin and structure of hybrid zones. *Molecular Ecology*, 25(11), 2454–2466.

Harris, S., Morris, P., Wray, S., & Yalden, D. (1995). *A Review of British Mammals: Population Estimates and Conservation Status of British Mammals Other Than Cetaceans*.

https://www.researchgate.net/publication/242583185_A_Review_of_British_Mammals_Population_Estimates_and_Conservation_Status_of_British_Mammals_Other_Than_Cetaceans

Hasselgren, M., Dussex, N., von Seth, J., Angerbjörn, A., Dalén, L., & Norén, K. (2024). Strongly deleterious mutations influence reproductive output and longevity in an endangered population. *Nature Communications*, 15(1), 8378.

Haudry, A., Platts, A. E., Vello, E., Hoen, D. R., Leclercq, M., Williamson, R. J., Forczek, E., Joly-Lopez, Z., Steffen, J. G., Hazzouri, K. M., Dewar, K., Stinchcombe, J. R., Schoen, D. J., Wang, X., Schmutz, J., Town, C. D., Edger, P. P., Pires, J. C., Schumaker, K. S., ... Blanchette, M. (2013). An atlas of over 90,000 conserved noncoding sequences provides insight into crucifer regulatory regions. *Nature Genetics*, 45(8), 891–898.

Hawkes, G., Chundru, K., Jackson, L., Patel, K. A., Murray, A., Wood, A. R., Wright, C. F., Weedon, M. N., Frayling, T. M., & Beaumont, R. N. (2025). Whole-genome sequencing analysis identifies rare, large-effect noncoding variants and regulatory regions associated with circulating protein levels. *Nature Genetics*, 57(3), 626–634.

Hecker, N., & Hiller, M. (2020). A genome alignment of 120 mammals highlights ultraconserved element variability and placenta-associated enhancers. *GigaScience*, 9(1). <https://doi.org/10.1093/gigascience/giz159>

Hedrick, P. W. (2013). Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. *Molecular Ecology*, 22(18), 4606–4618.

- Hedrick, P. W., & Garcia-Dorado, A. (2016). Understanding inbreeding depression, purging, and genetic rescue. *Trends in Ecology & Evolution*, 31(12), 940–952.
- Hedrick, P. W., & Kalinowski, S. T. (2000). Inbreeding Depression in Conservation Biology. *Annual Review of Ecology and Systematics*, 31(1), 139–162.
- Hedrick, P. W., Kardos, M., Peterson, R. O., & Vucetich, J. A. (2017). Genomic variation of inbreeding and ancestry in the remaining two Isle Royale wolves. *The Journal of Heredity*, 108(2), 120–126.
- Hedrick, P. W., Peterson, R. O., Vucetich, L. M., Adams, J. R., & Vucetich, J. A. (2014). Genetic rescue in Isle Royale wolves: genetic analysis and the collapse of the population. *Conservation Genetics*, 15(5), 1111–1121.
- Hedrick, P. W., Robinson, J. A., Peterson, R. O., & Vucetich, J. A. (2019). Genetics and extinction and the example of Isle Royale wolves. *Animal Conservation*, 22(3), 302–309.
- Hendricks, S., Anderson, E. C., Antao, T., Bernatchez, L., Forester, B. R., Garner, B., Hand, B. K., Hohenlohe, P. A., Kardos, M., Koop, B., Sethuraman, A., Waples, R. S., & Luikart, G. (2018). Recent advances in conservation and population genomics data analysis. *Evolutionary Applications*, 11(8), 1197–1211.
- Heppenheimer, E., Harrigan, R., Rutledge, L., Koepfli, K., DeCandia, A., Brzeski, K. E., Benson, J. F., Wheeldon, T., Patterson, B., Kays, R., Hohenlohe, P. A., & von Holdt, B. M. (2018). Population genomic analysis of north American eastern wolves (*Canis lycaon*) supports their conservation priority status. *Genes*, 9.
<https://doi.org/10.3390/genes9120606>
- Hickey, G., Paten, B., Earl, D., Zerbino, D., & Haussler, D. (2013). HAL: a hierarchical format for storing and analyzing multiple genome alignments. *Bioinformatics (Oxford, England)*, 29(10), 1341–1342.
- Hijmans, R. (2024). *geosphere: Spherical Trigonometry. R package version 1.5-20*,
<https://github.com/rspatial/geosphere>.
- Hill, W. G. (1982). Predictions of response to artificial selection from new mutations. *Genetical Research*, 40(3), 255–278.

- Hofmeester, T. R., Dijkstra, V., Dekker, J. J. A., van der Meij, T., & Broekhuizen, S. (2019). The status of the Dutch polecat population: correction of a recently published error. *Mammalia*, 83(5), 453–454.
- Hogg, C. J. (2024). Translating genomic advances into biodiversity conservation. *Nature Reviews. Genetics*, 25(5), 362–373.
- Hogg, C. J., Wright, B., Morris, K. M., Lee, A. V., Ivy, J. A., Grueber, C. E., & Belov, K. (2019). Founder relationships and conservation management: empirical kinships reveal the effect on breeding programmes when founders are assumed to be unrelated. *Animal Conservation*, 22(4), 348–361.
- Hohenlohe, P. A., Funk, W. C., & Rajora, O. P. (2021). Population genomics for wildlife conservation and management. *Molecular Ecology*, 30(1), 62–82.
- Holliday, J. A., Hallerman, E. M., & Haak, D. C. (2018). Genotyping and Sequencing Technologies in Population Genetics and Genomics. In *Population Genomics* (pp. 83–125). Springer International Publishing.
- Howard, J. G., Lynch, C., Santymire, R. M., Marinari, P. E., & Wildt, D. E. (2016). Recovery of gene diversity using long-term cryopreserved spermatozoa and artificial insemination in the endangered black-footed ferret: Black-footed ferret gene restoration. *Animal Conservation*, 19(2), 102–111.
- Howard-McCombe, J., Jamieson, A., Carmagnini, A., Russo, I.-R. M., Ghazali, M., Campbell, R., Driscoll, C., Murphy, W. J., Nowak, C., O'Connor, T., Tomsett, L., Lyons, L. A., Muñoz-Fuentes, V., Bruford, M. W., Kitchener, A. C., Larson, G., Frantz, L., Senn, H., Lawson, D. J., & Beaumont, M. A. (2023). Genetic swamping of the critically endangered Scottish wildcat was recent and accelerated by disease. *Current Biology: CB*, 33(21), 4761–4769.e5.
- Howard-McCombe, J., Ward, D., Kitchener, A. C., Lawson, D., Senn, H. V., & Beaumont, M. (2021). On the use of genome-wide data to model and date the time of anthropogenic hybridisation: An example from the Scottish wildcat. *Molecular Ecology*, 30(15), 3688–3702.
- Hsieh, H. M., Chiang, H. L., Tsai, L. C., Lai, S. Y., Huang, N. E., Linacre, A., & Lee, J. C.

- (2001). Cytochrome b gene for species identification of the conservation animals. *Forensic Science International*, 122(1), 7–18.
- Huang, L., Popic, V., & Batzoglou, S. (2013). Short read alignment with populations of genomes. *Bioinformatics (Oxford, England)*, 29(13), i361–i370.
- Hubby, J. L., & Lewontin, R. C. (1966). A molecular approach to the study of genic heterozygosity in natural populations. I. The number of alleles at different loci in *Drosophila pseudoobscura*. *Genetics*, 54(2), 577–594.
- Huber, C. D., Kim, B. Y., & Lohmueller, K. E. (2020). Population genetic models of GERP scores suggest pervasive turnover of constrained sites across mammalian evolution. *PLoS Genetics*, 16(5), e1008827.
- Hübner, S., Bercovich, N., Todesco, M., Mandel, J. R., Odenheimer, J., Ziegler, E., Lee, J. S., Baute, G. J., Owens, G. L., Grassa, C. J., Ebert, D. P., Ostevik, K. L., Moyers, B. T., Yakimowski, S., Masalia, R. R., Gao, L., Čalić, I., Bowers, J. E., Kane, N. C., ... Rieseberg, L. H. (2019). Sunflower pan-genome analysis shows that hybridization altered gene content and disease resistance. *Nature Plants*, 5(1), 54–62.
- Humble, E., Stoffel, M. A., Dicks, K., Ball, A. D., Gooley, R. M., Chuven, J., Pusey, R., Remeithi, M. A., Koepfli, K.-P., Pukazhenthi, B., Senn, H., & Ogden, R. (2023). Conservation management strategy impacts inbreeding and mutation load in scimitar-horned oryx. *Proceedings of the National Academy of Sciences of the United States of America*, 120(18), e2210756120.
- Hutchinson, A. M., Appeltant, R., Burdon, T., Bao, Q., Bargaje, R., Bodnar, A., Chambers, S., Comizzoli, P., Cook, L., Endo, Y., Harman, B., Hayashi, K., Hildebrandt, T., Korody, M. L., Lakshmipathy, U., Loring, J. F., Munger, C., Ng, A. H. M., Novak, B., ... Kohler, T. N. (2024). Advancing stem cell technologies for conservation of wildlife biodiversity. *Development (Cambridge, England)*, 151(20), dev203116.
- Hu, X., Saxena, A., Magstadt, D. R., Gauger, P. C., Burrough, E. R., Zhang, J., Siepker, C., Mainenti, M., Gorden, P. J., Plummer, P. J., & Li, G. (2024). Genomic characterization of highly pathogenic avian influenza A H5N1 virus newly emerged in dairy cattle. *Emerging Microbes & Infections*, 13(1), 2380421.

- Hwang, D.-Y., Kohl, S., Fan, X., Vivante, A., Chan, S., Dworschak, G. C., Schulz, J., van Eerde, A. M., Hilger, A. C., Gee, H. Y., Pennimpede, T., Herrmann, B. G., van de Hoek, G., Renkema, K. Y., Schell, C., Huber, T. B., Reutter, H. M., Soliman, N. A., Stajic, N., ... Hildebrandt, F. (2015). Mutations of the SLIT2-ROBO2 pathway genes SLIT2 and SRGAP1 confer risk for congenital anomalies of the kidney and urinary tract. *Human Genetics*, 134(8), 905–916.
- Jachowski, D., & Lockhart, J. M. (2009). Reintroducing the black-footed ferret *Mustela nigripes* to the great plains of north America. *Small Carnivore Conservation*.
https://www.researchgate.net/profile/David-Jachowski-2/publication/228484317_Reintroducing_the_black-footed_ferret_Mustela_nigripes_to_the_Great_Plains_of_North_America/links/53fb2cf10cf20a4549704da6/Reintroducing-the-black-footed-ferret-Mustela-nigripes-to-the-Great-Plains-of-North-America.pdf
- Jachowski, D. S., Gitzen, R. A., Grenier, M. B., Holmes, B., & Millsaugh, J. J. (2011). The importance of thinking big: Large-scale prey conservation drives black-footed ferret reintroduction success. *Biological Conservation*, 144(5), 1560–1566.
- Jackson, H. A., Percival-Alwyn, L., Ryan, C., Albeshr, M. F., Venturi, L., Morales, H. E., Mathers, T. C., Cocker, J., Speak, S. A., Accinelli, G. G., Barker, T., Heavens, D., Willman, F., Dawson, D., Ward, L., Tatayah, V., Zuël, N., Young, R., Concannon, L., ... Van Oosterhout, C. (2022). Genomic erosion in a demographically recovered bird species during conservation rescue. *Conservation Biology: The Journal of the Society for Conservation Biology*, 36(4), e13918.
- Jarausch, A., Harms, V., Kluth, G., Reinhardt, I., & Nowak, C. (2021). How the west was won: genetic reconstruction of rapid wolf recolonization into Germany's anthropogenic landscapes. *Heredity*, 127(1), 92–106.
- Jayakodi, M., Padmarasu, S., Haberer, G., Bonthala, V. S., Gundlach, H., Monat, C., Lux, T., Kamal, N., Lang, D., Himmelbach, A., Ens, J., Zhang, X.-Q., Angessa, T. T., Zhou, G., Tan, C., Hill, C., Wang, P., Schreiber, M., Boston, L. B., ... Stein, N. (2020). The barley pan-genome reveals the hidden legacy of mutation breeding. *Nature*, 588(7837), 284–289.

- Johnson, R. N., O'Meally, D., Chen, Z., Etherington, G. J., Ho, S. Y. W., Nash, W. J., Grueber, C. E., Cheng, Y., Whittington, C. M., Dennison, S., Peel, E., Haerty, W., O'Neill, R. J., Colgan, D., Russell, T. L., Alquezar-Planas, D. E., Attenbrow, V., Bragg, J. G., Brandies, P. A., ... Belov, K. (2018). Adaptation and conservation insights from the koala genome. *Nature Genetics*, 50(8), 1102–1111.
- Johnson, R. N., Wilson-Wilde, L., & Linacre, A. (2014). Current and future directions of DNA in wildlife forensic science. *Forensic Science International. Genetics*, 10, 1–11.
- Johnson, W. E., Onorato, D. P., Roelke, M. E., Land, E. D., Cunningham, M., Belden, R. C., McBride, R., Jansen, D., Lotz, M., Shindle, D., Howard, J., Wildt, D. E., Penfold, L. M., Hostetler, J. A., Oli, M. K., & O'Brien, S. J. (2010). Genetic restoration of the Florida panther. *Science (New York, N.Y.)*, 329(5999), 1641–1645.
- Jones, M. R., Mills, L. S., Alves, P. C., Callahan, C. M., Alves, J. M., Lafferty, D. J. R., Jiggins, F. M., Jensen, J. D., Melo-Ferreira, J., & Good, J. M. (2018). Adaptive introgression underlies polymorphic seasonal camouflage in snowshoe hares. *Science (New York, N.Y.)*, 360(6395), 1355–1358.
- Kaplow, I. M., Lawler, A. J., Schäffer, D. E., Srinivasan, C., Sestili, H. H., Wirthlin, M. E., Phan, B. N., Prasad, K., Brown, A. R., Zhang, X., Foley, K., Genereux, D. P., Zoonomia Consortium**, Karlsson, E. K., Lindblad-Toh, K., Meyer, W. K., & Pfenning, A. R. (2023). Relating enhancer genetic variation across mammals to complex phenotypes using machine learning. *Science (New York, N.Y.)*, 380(6643), eabm7993.
- Kaplow, I. M., Schäffer, D. E., Wirthlin, M. E., Lawler, A. J., Brown, A. R., Kleyman, M., & Pfenning, A. R. (2022). Inferring mammalian tissue-specific regulatory conservation by predicting tissue-specific differences in open chromatin. *BMC Genomics*, 23(1), 291.
- Karasu, O. R., Neuner, A., Atorino, E. S., Pereira, G., & Schiebel, E. (2022). The central scaffold protein CEP350 coordinates centriole length, stability, and maturation. *The Journal of Cell Biology*, 221(12). <https://doi.org/10.1083/jcb.202203081>
- Kardos, M., Armstrong, E. E., Fitzpatrick, S. W., Hauser, S., Hedrick, P. W., Miller, J. M.,

- Tallmon, D. A., & Funk, W. C. (2021). The crucial role of genome-wide genetic variation in conservation. *Proceedings of the National Academy of Sciences of the United States of America*, 118(48). <https://doi.org/10.1073/pnas.2104642118>
- Kardos, M., Luikart, G., & Allendorf, F. W. (2015). Measuring individual inbreeding in the age of genomics: marker-based measures are better than pedigrees. *Heredity*, 115(1), 63–72.
- Kardos, M., Taylor, H. R., Ellegren, H., Luikart, G., & Allendorf, F. W. (2016). Genomics advances the study of inbreeding depression in the wild. *Evolutionary Applications*, 9(10), 1205–1218.
- Kardos, M., & Waples, R. S. (2024). Low-coverage sequencing and Wahlund effect severely bias estimates of inbreeding, heterozygosity and effective population size in North American wolves. *Molecular Ecology*, e17415.
- Karollus, A., Hingerl, J., Gankin, D., Grosshauser, M., Klemon, K., & Gagneur, J. (2024). Species-aware DNA language models capture regulatory elements and their evolution. *Genome Biology*, 25(1), 83.
- Kays, R., Curtis, A., & Kirchman, J. J. (2010). Rapid adaptive evolution of northeastern coyotes via hybridization with wolves. *Biology Letters*, 6(1), 89–93.
- Keck, F., Peller, T., Alther, R., Barouillet, C., Blackman, R., Capo, E., Chonova, T., Couton, M., Fehlinger, L., Kirschner, D., Knüsel, M., Muneret, L., Oester, R., Tapolczai, K., Zhang, H., & Altermatt, F. (2025). The global human impact on biodiversity. *Nature*, 1–6.
- Keller & Waller. (2002). Inbreeding effects in wild populations. *Trends in Ecology & Evolution*, 17(5), 230–241.
- Khan, A., Patel, K., Shukla, H., Viswanathan, A., van der Valk, T., Borthakur, U., Nigam, P., Zachariah, A., Jhala, Y. V., Kardos, M., & Ramakrishnan, U. (2021). Genomic evidence for inbreeding depression and purging of deleterious genetic variation in Indian tigers. *Proceedings of the National Academy of Sciences of the United States of America*, 118(49). <https://doi.org/10.1073/pnas.2023018118>
- Kilshaw, K. (2015). *Introgression and the current status of the Scottish wildcat (Felis*

silvestris silvestris).

<https://ora.ox.ac.uk/objects/uuid:11c27f05-c051-4cbb-b498-538561005d87>

- Kimura, M. (1968). Evolutionary rate at the molecular level. *Nature*, 217(5129), 624–626.
- Kimura, M., Maruyama, T., & Crow, J. F. (1963). THE MUTATION LOAD IN SMALL POPULATIONS. *Genetics*, 48(10), 1303–1312.
- Kircher, M., Witten, D. M., Jain, P., O’Roak, B. J., Cooper, G. M., & Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics*, 46(3), 310–315.
- Kitchener, A. C., & Birks, J. D. S. (2008). *Feral ferret Mustela furo*.
<https://scholar.google.com/citations?user=b6MOxGsAAAAJ&hl=en&oi=sra>
- Kitchener, A. C., Yamaguchi, N., Ward, J. M., & Macdonald, D. W. (2005). A diagnosis for the Scottish wildcat (*Felis silvestris*): a tool for conservation action for a critically-endangered felid. *Animal Conservation*, 8(3), 223–237.
- Klimova, A., Gutiérrez-Rivera, J. N., Sánchez-Sotomayor, V., & Hoffman, J. I. (2022). The genetic consequences of captive breeding, environmental change and human exploitation in the endangered peninsular pronghorn. *Scientific Reports*, 12(1), 11253.
- Kliver, S., Houck, M. L., Perelman, P. L., Totikov, A., Tomarovsky, A., Dudchenko, O., Omer, A. D., Colaric, Z., Weisz, D., Aiden, E. L., Chan, S., Hastie, A., Komissarov, A., Ryder, O. A., Graphodatsky, A., Johnson, W. E., Maldonado, J. E., Pukazhenth, B. S., Marinari, P. E., ... Koepfli, K.-P. (2023). Chromosome-length genome assembly and karyotype of the endangered black-footed ferret (*Mustela nigripes*). *The Journal of Heredity*, 114(5), 539–548.
- Koepfli, K.-P., Deere, K. A., Slater, G. J., Begg, C., Begg, K., Grassman, L., Lucherini, M., Veron, G., & Wayne, R. K. (2008). Multigene phylogeny of the Mustelidae: resolving relationships, tempo and biogeographic history of a mammalian adaptive radiation. *BMC Biology*, 6(1), 10.
- Koepfli, K.-P., Paten, B., Genome 10K Community of Scientists, & O’Brien, S. J. (2015). The Genome 10K Project: a way forward. *Annual Review of Animal Biosciences*, 3,

- Kolbe, J. J., Glor, R. E., Rodríguez Schettino, L., Lara, A. C., Larson, A., & Losos, J. B. (2004). Genetic variation increases during biological invasion by a Cuban lizard. *Nature*, 431(7005), 177–181.
- Kronenberg, Z. N., Fiddes, I. T., Gordon, D., Murali, S., Cantsilieris, S., Meyerson, O. S., Underwood, J. G., Nelson, B. J., Chaisson, M. J. P., Dougherty, M. L., Munson, K. M., Hastie, A. R., Diekhans, M., Hormozdiari, F., Lorusso, N., Hoekzema, K., Qiu, R., Clark, K., Raja, A., ... Eichler, E. E. (2018). High-resolution comparative analysis of great ape genomes. *Science (New York, N.Y.)*, 360(6393).
<https://doi.org/10.1126/science.aar6343>
- Kuderna, L. F. K., Ulirsch, J. C., Rashid, S., Ameen, M., Sundaram, L., Hickey, G., Cox, A. J., Gao, H., Kumar, A., Aguet, F., Christmas, M. J., Clawson, H., Haeussler, M., Janiak, M. C., Kuhlwilm, M., Orkin, J. D., Bataillon, T., Manu, S., Valenzuela, A., ... Farh, K. K.-H. (2024). Identification of constrained sequence elements across 239 primate genomes. *Nature*, 625(7996), 735–742.
- Kumari, P., Friedman, R. Z., Curtis, S. W., Pi, L., Paraiso, K., Visel, A., Rhea, L., Dunnwald, M., Patni, A. P., Mar, D., Bomsztyk, K., Mathieu, J., Ruohola-Baker, H., Leslie-Clarkson, E. J., White, M. A., Cohen, B. A., & Cornell, R. A. (2025). Identification of functional non-coding variants associated with orofacial cleft. *Nature Communications*, 16(1), 6545.
- Kyriazis, C. C., Wayne, R. K., & Lohmueller, K. E. (2021). Strongly deleterious mutations are a primary determinant of extinction risk due to inbreeding depression. *Evolution Letters*, 5(1), 33–47.
- Labes, S., Stupp, D., Wagner, N., Bloch, I., Lotem, M., L Lahad, E., Polak, P., Pupko, T., & Tabach, Y. (2022). Machine-learning of complex evolutionary signals improves classification of SNVs. *NAR Genomics and Bioinformatics*, 4(2), lqac025.
- LaHood, E. S., Moran, P., Olsen, J., Grant, W. S., & Park, L. K. (2002). Microsatellite allele ladders in two species of Pacific salmon: preparation and field-test results. *Molecular Ecology Notes*, 2(2), 187–190.

- Lande, R. (1994). Risk of population extinction from fixation of new deleterious mutations. *Evolution; International Journal of Organic Evolution*, 48(5), 1460–1469.
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., ... International Human Genome Sequencing Consortium. (2001). Initial sequencing and analysis of the human genome. *Nature*, 409(6822), 860–921.
- Langley, P. J. W., & Yalden, D. W. (1977). The decline of the rarer carnivores in Great Britain during the nineteenth century. *Mammal Review*, 7(3-4), 95–116.
- Lan, T., Li, H., Liu, B., Shi, M., Tian, Y., Sahu, S. K., Cui, L., Dussex, N., Liu, D., Ma, Y., Kong, W., Liu, S., Fan, J., Zhao, Y., Fu, Y., Li, Q., Lin, C., Dalén, L., Liu, H., ... Xu, Y. (2024). Revealing extensive inbreeding and less-efficient purging of deleterious mutations in wild Amur tigers in China. *Yi Chuan Xue Bao [Journal of Genetics and Genomics]*. <https://doi.org/10.1016/j.jgg.2024.12.004>
- Lawson, D. J., Howard-McCombe, J., Beaumont, M., & Senn, H. (2024). How admixed captive breeding populations could be rescued using local ancestry information. *Molecular Ecology*, e17349.
- Leek, J. T., Scharpf, R. B., Bravo, H. C., Simcha, D., Langmead, B., Johnson, W. E., Geman, D., Baggerly, K., & Irizarry, R. A. (2010). Tackling the widespread and critical impact of batch effects in high-throughput data. *Nature Reviews. Genetics*, 11(10), 733–739.
- Lee, T. I., & Young, R. A. (2013). Transcriptional regulation and its misregulation in disease. *Cell*, 152(6), 1237–1251.
- Leigh, D. M., Lischer, H. E. L., Grossen, C., & Keller, L. F. (2018). Batch effects in a multiyear sequencing study: False biological trends due to changes in read lengths. *Molecular Ecology Resources*, 18(4), 778–788.
- Leitwein, M., Duranton, M., Rougemont, Q., Gagnaire, P.-A., & Bernatchez, L. (2020). Using haplotype information for conservation genomics. *Trends in Ecology & Evolution*, 35(3), 245–258.

- Lewin, H. A., Robinson, G. E., Kress, W. J., Baker, W. J., Coddington, J., Crandall, K. A., Durbin, R., Edwards, S. V., Forest, F., Gilbert, M. T. P., Goldstein, M. M., Grigoriev, I. V., Hackett, K. J., Haussler, D., Jarvis, E. D., Johnson, W. E., Patrinos, A., Richards, S., Castilla-Rubio, J. C., ... Zhang, G. (2018). Earth BioGenome Project: Sequencing life for the future of life. *Proceedings of the National Academy of Sciences of the United States of America*, 115(17), 4325–4333.
- Liao, W.-W., Asri, M., Ebler, J., Doerr, D., Haukness, M., Hickey, G., Lu, S., Lucas, J. K., Monlong, J., Abel, H. J., Buonaiuto, S., Chang, X. H., Cheng, H., Chu, J., Colonna, V., Eizenga, J. M., Feng, X., Fischer, C., Fulton, R. S., ... Paten, B. (2023). A draft human pangenome reference. *Nature*, 617(7960), 312–324.
- Li, G., Fang, C., Li, Y., Wang, Z., Sun, S., He, S., Qi, W., Bao, C., Ma, H., Fan, Y., Feng, Y., & Liu, X. (2022). Global impacts of future urban expansion on terrestrial vertebrate diversity. *Nature Communications*, 13(1), 1628.
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, 25(14), 1754–1760.
- Lindblad-Toh, K., Garber, M., Zuk, O., Lin, M. F., Parker, B. J., Washietl, S., Kheradpour, P., Ernst, J., Jordan, G., Mauceli, E., Ward, L. D., Lowe, C. B., Holloway, A. K., Clamp, M., Gnerre, S., Alföldi, J., Beal, K., Chang, J., Clawson, H., ... Kellis, M. (2011). A high-resolution map of human evolutionary constraint using 29 mammals. *Nature*, 478(7370), 476–482.
- Lindblad-Toh, K., Wade, C. M., Mikkelsen, T. S., Karlsson, E. K., Jaffe, D. B., Kamal, M., Clamp, M., Chang, J. L., Kulbokas, E. J., 3rd, Zody, M. C., Mauceli, E., Xie, X., Breen, M., Wayne, R. K., Ostrander, E. A., Ponting, C. P., Galibert, F., Smith, D. R., DeJong, P. J., ... Lander, E. S. (2005). Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature*, 438(7069), 803–819.
- Li, R., Gong, M., Zhang, X., Wang, F., Liu, Z., Zhang, L., Yang, Q., Xu, Y., Xu, M., Zhang, H., Zhang, Y., Dai, X., Gao, Y., Zhang, Z., Fang, W., Yang, Y., Fu, W., Cao, C., Yang, P., ... Jiang, Y. (2023). A sheep pangenome reveals the spectrum of structural variations and their effects on tail phenotypes. *Genome Research*, 33(3), 463–477.

- Liu, M., Boot, A., Ng, A. W. T., Gordân, R., & Rozen, S. G. (2021). Mutational processes in cancer preferentially affect binding of particular transcription factors. *Scientific Reports*, 11(1), 3339.
- Livieri, T. M., Forrest, S. C., Matchett, M. R., & Breck, S. (2022). *Conserving Endangered Black-Footed Ferrets: Biological Threats, Political Challenges, and Lessons Learned*. https://digitalcommons.unl.edu/icwdm_usdanwrc/2638/
- Living Planet Report 2024: LPR*. (2024). WWF International.
- Lo, B., & Abdel-Motal, U. M. (2017). Lessons from CTLA-4 deficiency and checkpoint inhibition. *Current Opinion in Immunology*, 49, 14–19.
- Locke, H., Ellis, E. C., Venter, O., Schuster, R., Ma, K., Shen, X., Woodley, S., Kingston, N., Bhola, N., Strassburg, B. B. N., Paulsch, A., Williams, B., & Watson, J. E. M. (2019). Three global conditions for biodiversity conservation and sustainable use: an implementation framework. *National Science Review*, 6(6), 1080–1082.
- Loewe, L., & Hill, W. G. (2010). The population genetics of mutations: good, bad and indifferent. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 365(1544), 1153–1167.
- López, A., Carreras, C., Pascual, M., & Pegueroles, C. (2025). Evaluating restriction enzyme selection for reduced representation sequencing in conservation genomics. *Molecular Ecology Resources*, 25(5), e13865.
- López-Cortegano, E., Moreno, E., & García-Dorado, A. (2021). Genetic purging in captive endangered ungulates with extremely low effective population sizes. *Heredity*, 127(5), 433–442.
- Lou, R. N., & Therikildsen, N. O. (2022). Batch effects in population genomic studies with low-coverage whole genome sequencing data: Causes, detection and mitigation. *Molecular Ecology Resources*, 22(5), 1678–1692.
- Löytynoja, A. (2014). Phylogeny-aware alignment with PRANK. *Methods in Molecular Biology (Clifton, N.J.)*, 1079, 155–170.
- Lunter, G., Ponting, C. P., & Hein, J. (2006). Genome-wide identification of human functional DNA using a neutral indel model. *PLoS Computational Biology*, 2(1), e5.

- Lynch, M. (2006). The origins of eukaryotic gene structure. *Molecular Biology and Evolution*, 23(2), 450–468.
- Lynch, M. (2010). Evolution of the mutation rate. *Trends in Genetics: TIG*, 26(8), 345–352.
- Malhis, N., Jones, S. J. M., & Gsponer, J. (2019). Improved measures for evolutionary conservation that exploit taxonomy distances. *Nature Communications*, 10(1), 1556.
- Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends in Ecology & Evolution*, 20(5), 229–237.
- Manichaikul, A., Mychaleckyj, J. C., Rich, S. S., Daly, K., Sale, M., & Chen, W.-M. (2010). Robust relationship inference in genome-wide association studies. *Bioinformatics (Oxford, England)*, 26(22), 2867–2873.
- Manunza, A., Amills, M., Noce, A., Cabrera, B., Zidi, A., Eghbalsaied, S., de Albornoz, E. C., Portell, M., Mercadé, A., Sánchez, A., & Balteanu, V. (2016). Romanian wild boars and Mangalitza pigs have a European ancestry and harbour genetic signatures compatible with past population bottlenecks. *Scientific Reports*, 6(1), 29913.
- Maran, T., Skumatov, D., Abramov, A. V., & Kranz, A. (2016).
https://www.researchgate.net/profile/Alexei-Abramov/publication/308778184_Mustela_eversmanii/links/57ef8f5208ae886b897417c4/Mustela-eversmanii.pdf
- Marchant, T. W., Dietschi, E., Rytz, U., Schawalder, P., Jagannathan, V., Hadji Rasouliha, S., Gurtner, C., Waldvogel, A. S., Harrington, R. S., Drögemüller, M., Kidd, J., Ostrander, E. A., Warr, A., Watson, M., Argyle, D., Ter Haar, G., Clements, D. N., Leeb, T., & Schoenebeck, J. J. (2019). An ADAMTS3 missense variant is associated with Norwich Terrier upper airway syndrome. *PLoS Genetics*, 15(5), e1008102.
- Margulies, E. H., Cooper, G. M., Asimenos, G., Thomas, D. J., Dewey, C. N., Siepel, A., Birney, E., Keefe, D., Schwartz, A. S., Hou, M., Taylor, J., Nikolaev, S., Montoya-Burgos, J. I., Löytynoja, A., Whelan, S., Pardi, F., Massingham, T., Brown, J. B., Bickel, P., ... Sidow, A. (2007). Analyses of deep mammalian sequence alignments and constraint predictions for 1% of the human genome. *Genome Research*, 17(6), 760–774.
- Marinari, P., Baughman, J., Heuring, W., & Mantz, N. (2024). SAFE black-footed ferret

- program action plan 2024-2027. *Association of Zoos and Aquariums, Silver Spring, Maryland, USA*. <https://assets.speakcdn.com/assets/2332/sf2024-2027bfffpp.pdf>
- Marneweck, C., Butler, A. R., Gigliotti, L. C., Harris, S. N., Jensen, A. J., Muthersbaugh, M., Newman, B. A., Saldo, E. A., Shute, K., Titus, K. L., Yu, S. W., & Jachowski, D. S. (2021). Shining the spotlight on small mammalian carnivores: Global status and threats. *Biological Conservation*, 255(109005), 109005.
- Martínez-Cruz, B., Zalewska, H., & Zalewski, A. (2022). The genetic diversity and structure in the European polecat were not affected by the introduction of the American mink in Poland. *PloS One*, 17(9), e0266161.
- Martin-Geary, A. C., Blakes, A. J. M., Dawes, R., Findlay, S. D., Lord, J., Walker, S., Talbot-Martin, J., Wieder, N., D'Souza, E. N., Fernandes, M., Hilton, S., Lahiri, N., Campbell, C., Jenkinson, S., DeGoede, C. G. E. L., Anderson, E. R., Burge, C. B., Sanders, S. J., Ellingford, J., ... Whiffin, N. (2023). Systematic identification of disease-causing promoter and untranslated region variants in 8,040 undiagnosed individuals with rare disease. *medRxiv: The Preprint Server for Health Sciences*, 2023.09.12.23295416.
- Mata, C., Hervás, I., Herranz, J., Suárez, F., & Malo, J. E. (2008). Are motorway wildlife passages worth building? Vertebrate use of road-crossing structures on a Spanish motorway. *Journal of Environmental Management*, 88(3), 407–415.
- Mathews, F., Kubasiewicz, L., Gurnell, J., Harrower, C., McDonald, R., & Shore, R. (2018). *A review of the population and conservation status of British mammals*. Unpublished. <https://doi.org/10.13140/RG.2.2.28059.31521>
- Mathur, S., & DeWoody, J. A. (2021). Genetic load has potential in large populations but is realized in small inbred populations. *Evolutionary Applications*, 14(6), 1540–1557.
- Matthews, C. A., Watson-Haigh, N. S., Burton, R. A., & Sheppard, A. E. (2024). A gentle introduction to pangenomics. *Briefings in Bioinformatics*, 25(6), bbae588.
- Maxwell, S. L., Fuller, R. A., Brooks, T. M., & Watson, J. E. M. (2016). Biodiversity: The ravages of guns, nets and bulldozers. *Nature*, 536(7615), 143–145.
- May, S. A., Rosenbaum, S. W., Pearse, D. E., Kardos, M., Primmer, C. R., Baetscher, D.

- S., & Waples, R. S. (2025). The genomics revolution in nonmodel species: Predictions vs. Reality for salmonids. *Molecular Ecology*, e17758.
- Mc Cartney, A. M., Formenti, G., Mouton, A., De Panis, D., Marins, L. S., Leitão, H. G., Diedericks, G., Kirangwa, J., Morselli, M., Salces-Ortiz, J., Escudero, N., Iannucci, A., Natali, C., Svardal, H., Fernández, R., De Pooter, T., Joris, G., Strazisar, M., Wood, J. M. D., ... Mazzoni, C. J. (2024). The European Reference Genome Atlas: piloting a decentralised approach to equitable biodiversity genomics. *Npj Biodiversity*, 3(1), 28.
- McCole, R. B., Erceg, J., Saylor, W., & Wu, C.-T. (2018). Ultraconserved elements occupy specific arenas of three-dimensional mammalian genome organization. *Cell Reports*, 24(2), 479–488.
- 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., Ågren, E. O., Learmount, J., Basto, M., Fernandes, C., Stuart, P., Tosh, D. G., Sindicic, M., Andreanszky, T., Isomursu, M., ... Wójcik, J. M. (2022). Next-generation phylogeography resolves post-glacial colonization patterns in a widespread carnivore, the red fox (*Vulpes vulpes*), in Europe. *Molecular Ecology*, 31(3), 993–1006.
- McGowan, P. J. K., Traylor-Holzer, K., & Leus, K. (2017). IUCN Guidelines for Determining When and How Ex Situ Management Should Be Used in Species Conservation: IUCN ex situ management guidelines. *Conservation Letters*, 10(3), 361–366.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9), 1297–1303.
- McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R. S., Thormann, A., Flicek, P., & Cunningham, F. (2016). The Ensembl Variant Effect Predictor. *Genome Biology*, 17(1), 122.
- McLean, C. Y., Reno, P. L., Pollen, A. A., Bassan, A. I., Capellini, T. D., Guenther, C., Indjeian, V. B., Lim, X., Menke, D. B., Schaar, B. T., Wenger, A. M., Bejerano, G., & Kingsley, D. M. (2011). Human-specific loss of regulatory DNA and the evolution of

- human-specific traits. *Nature*, 471(7337), 216–219.
- McMahon, B. J., Teeling, E. C., & Höglund, J. (2014). How and why should we implement genomics into conservation? *Evolutionary Applications*, 7(9), 999–1007.
- McNeely, J. A., Miller, K. R., Reid, W. V., Mittermeier, R. A., & Werner, T. B. (1990). Strategies for conserving biodiversity. *Environment*, 32(3), 16–40.
- McRae, L., Marconi, V., Deinet, S., Puleston, H., Living Planet Index Team, & Freeman, R. (2024). Living Planet Index technical supplement 2024. In *stats.livingplanetindex.org*. OSF. <https://doi.org/10.17605/OSF.IO/6TE9H>
- Mead, D., Hailer, F., Chadwick, E., Portela Miguez, R., Smith, M., Corton, C., Oliver, K., Skelton, J., Betteridge, E., Doulcan, J. D., Dudchenko, O., Omer, A., Weisz, D., Lieberman Aiden, E., McCarthy, S., Howe, K., Sims, Y., Torrance, J., Tracey, A., ... Blaxter, M. (2020). The genome sequence of the Eurasian river otter, *Lutra lutra* Linnaeus 1758. *Wellcome Open Research*, 5, 33.
- Meek, M. H., Beever, E. A., Barbosa, S., Fitzpatrick, S. W., Fletcher, N. K., Mittan-Moreau, C. S., Reid, B. N., Campbell-Staton, S. C., Green, N. F., & Hellmann, J. J. (2023). Understanding local adaptation to prepare populations for climate change. *Bioscience*, 73(1), 36–47.
- Meier, J. I., Marques, D. A., Mwaiko, S., Wagner, C. E., Excoffier, L., & Seehausen, O. (2017). Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nature Communications*, 8(1), 14363.
- Mendes, F. K., Vanderpool, D., Fulton, B., & Hahn, M. W. (2021). CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics*, 36(22-23), 5516–5518.
- Mestre, F. M., Ferreira, J. P., & Mira, A. (2007). Modelling the distribution of the European polecat *Mustela putorius* in a Mediterranean agricultural landscape. *Revue D'écologie*, 62(1), 35–47.
- Miller, B., Reading, R. P., & Forrest, S. (1996). *Prairie night: black-footed ferrets and the recovery of endangered species*.
- Miller-Butterworth, C. M., Vacco, K., Russell, A. L., & Gaspard, J. C., 3rd. (2021). Genetic diversity and relatedness among captive African painted dogs in North America.

Genes, 12(10), 1463.

- Miranda, I., Giska, I., Farelo, L., Pimenta, J., Zimova, M., Bryk, J., Dalén, L., Mills, L. S., Zub, K., & Melo-Ferreira, J. (2021). Museomics dissects the genetic basis for adaptive seasonal coloration in the least weasel. *Molecular Biology and Evolution*, 38(10), 4388–4402.
- Mirarab, S., Reaz, R., Bayzid, M. S., Zimmermann, T., Swenson, M. S., & Warnow, T. (2014). ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics (Oxford, England)*, 30(17), i541–i548.
- Mojarad, B. A., Gupta, G. D., Hasegan, M., Goudiam, O., Basto, R., Gingras, A.-C., & Pelletier, L. (2017). CEP19 cooperates with FOP and CEP350 to drive early steps in the ciliogenesis programme. *Open Biology*, 7(6), 170114.
- Montgomery, W. I., Provan, J., McCabe, A. M., & Yalden, D. W. (2014). Origin of British and Irish mammals: disparate post-glacial colonisation and species introductions. *Quaternary Science Reviews*, 98, 144–165.
- Moreno, E., Cervantes, I., Gutiérrez, J. P., Fernández, I., & Goyache, F. (2024). Analysing the pedigree to identify undesirable losses of genetic diversity and to prioritize management decisions in captive breeding: a case study. *Heredity*, 133(6), 400–409.
- Moritz, C. (1994). Defining “Evolutionarily Significant Units” for conservation. *Trends in Ecology & Evolution*, 9(10), 373–375.
- Moroni, B., Brambilla, A., Rossi, L., Meneguz, P. G., Bassano, B., & Tizzani, P. (2022). Hybridization between Alpine ibex and domestic goat in the Alps: A sporadic and localized phenomenon? *Animals: An Open Access Journal from MDPI*, 12(6), 751.
- Mueller, S. A., Prost, S., Anders, O., Breitenmoser-Würsten, C., Kleven, O., Klinga, P., Konec, M., Kopatz, A., Krojerová-Prokešová, J., Middelhoff, T. L., Obexer-Ruff, G., Reiners, T. E., Schmidt, K., Sindičič, M., Skrbinšek, T., Tám, B., Saveljev, A. P., Naranbaatar, G., & Nowak, C. (2022). Genome-wide diversity loss in reintroduced Eurasian lynx populations urges immediate conservation management. *Biological Conservation*, 266(109442), 109442.
- Münger, X., Robin, M., Dalén, L., & Grossen, C. (2024). Facilitated introgression from

- domestic goat into Alpine ibex at immune loci. *Molecular Ecology*, 33(14), e17429.
- Munson, L., Brown, J. L., Bush, M., Packer, C., Janssen, D., Reiziss, S. M., & Wildt, D. E. (1996). Genetic diversity affects testicular morphology in free-ranging lions (*Panthera leo*) of the Serengeti Plains and Ngorongoro Crater. *Journal of Reproduction and Fertility*, 108(1), 11–15.
- Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A., & Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, 403(6772), 853–858.
- Mysłajek, R. W., Olkowska, E., Wronka-Tomulewicz, M., & Nowak, S. (2020). Mammal use of wildlife crossing structures along a new motorway in an area recently recolonized by wolves. *European Journal of Wildlife Research*, 66(5), 1–14.
- Naude, V. N., Balme, G. A., O’Riain, J., Hunter, L. T. B., Fattebert, J., Dickerson, T., & Bishop, J. M. (2020). Unsustainable anthropogenic mortality disrupts natal dispersal and promotes inbreeding in leopards. *Ecology and Evolution*, 10(8), 3605–3619.
- Neaves, L. E., Brockett, B. M., Evans, M. J., Pierson, J. C., & Sarre, S. D. (2024). Technological advances have enhanced and expanded conservation genomics research but are yet to be integrated fully into biodiversity management. *Biological Conservation*, 299(110788), 110788.
- Nelson, H. V., Farquharson, K. A., Georges, A., McLennan, E. A., DeGabriel, J. L., Giese, M., Ormond, C., McFadden, M., Skidmore, A., Prangell, J., Belov, K., & Hogg, C. J. (2024). A genomic framework to assist conservation breeding and translocation success: A case study of a critically endangered turtle. *Conservation Science and Practice*, 6(10), e13204.
- Ng, P. C., & Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Research*, 31(13), 3812–3814.
- Nie, M., Oravcová, M., Jami-Alahmadi, Y., Wohlschlegel, J. A., Lazzerini-Denchi, E., & Boddy, M. N. (2021). FAM111A induces nuclear dysfunction in disease and viral restriction. *EMBO Reports*, 22(2), e50803.
- Novak, B. J., Gober, P., Bortner, R., Garelle, D., Wright, M., Chuven, J., Houck, M. L., Ryder, O. A., Milutinovich, D., Benavidez, J., Ryan, K., Walker, S., Arenivas, S. S.,

- Aston, L., Russell, B., Marinari, P., Crosier, A., Helmick, K., Gibson, M. R., ... Phelan, R. (2024). First endangered black-footed ferrets, *Mustela nigripes*, cloned for genetic rescue. In *bioRxiv* (p. 2024.04.17.589896).
<https://doi.org/10.1101/2024.04.17.589896>
- Novak, B. J., Maloney, T., & Phelan, R. (2018). Advancing a new toolkit for conservation: From science to policy. *The CRISPR Journal*, 1, 11–15.
- Novak, B. J., Ryder, O. A., Houck, M. L., Walker, K., Russell, L., Russell, B., Walker, S., Arenivas, S. S., Aston, L., Veneklasen, G., Ivy, J. A., Koepfli, K.-P., Rusnak, A., Simek, J., Zhuk, A., Putnam, A. S., & Phelan, R. (2025). Endangered Przewalski's horse, *Equus przewalskii*, cloned from historically cryopreserved cells. *Animals: An Open Access Journal from MDPI*, 15(5), 613.
- Novo, I., Ordás, P., Moraga, N., Santiago, E., Quesada, H., & Caballero, A. (2023). Impact of population structure in the estimation of recent historical effective population size by the software GONE. *Genetics, Selection, Evolution: GSE*, 55(1), 86.
- O'Brien, S. J., Johnson, W. E., Driscoll, C. A., Dobrynin, P., & Marker, L. (2017). Conservation genetics of the cheetah: Lessons learned and new opportunities. *The Journal of Heredity*, 108(6), 671–677.
- O'brien, S. J., Wildt, D. E., Goldman, D., Merrill, C. R., & Bush, M. (1983). The cheetah is depauperate in genetic variation. *Science (New York, N.Y.)*, 221(4609), 459–462.
- Ogden, R., Chuven, J., Gilbert, T., Hosking, C., Gharbi, K., Craig, M., Al Dhaheri, S. S., & Senn, H. (2020). Benefits and pitfalls of captive conservation genetic management: Evaluating diversity in scimitar-horned oryx to support reintroduction planning. *Biological Conservation*, 241(108244), 108244.
- Ohta, T. (1973). Slightly deleterious mutant substitutions in evolution. *Nature*, 246(5428), 96–98.
- O'Leary, S. J., Puritz, J. B., Willis, S. C., Hollenbeck, C. M., & Portnoy, D. S. (2018). These aren't the loci you're looking for: Principles of effective SNP filtering for molecular ecologists. *Molecular Ecology*, 27(16), 3193–3206.
- Ouborg, N. J., Pertoldi, C., Loeschcke, V., Bijlsma, R. K., & Hedrick, P. W. (2010).

- Conservation genetics in transition to conservation genomics. *Trends in Genetics: TIG*, 26(4), 177–187.
- Oziolor, E. M., Reid, N. M., Yair, S., Lee, K. M., Guberman VerPloeg, S., Bruns, P. C., Shaw, J. R., Whitehead, A., & Matson, C. W. (2019). Adaptive introgression enables evolutionary rescue from extreme environmental pollution. *Science (New York, N.Y.)*, 364(6439), 455–457.
- Pan-Cancer Whole Genomes Consortium. (2020). Pan-cancer analysis of whole genomes. *Nature*, 578(7793), 82–93.
- Paniw, M., James, T. D., Ruth Archer, C., Römer, G., Levin, S., Compagnoni, A., Che-Castaldo, J., Bennett, J. M., Mooney, A., Childs, D. Z., Ozgul, A., Jones, O. R., Burns, J. H., Beckerman, A. P., Patwary, A., Sanchez-Gassen, N., Knight, T. M., & Salguero-Gómez, R. (2021). The myriad of complex demographic responses of terrestrial mammals to climate change and gaps of knowledge: A global analysis. *The Journal of Animal Ecology*, 90(6), 1398–1407.
- Parson, W., Pegoraro, K., Niederstätter, H., Föger, M., & Steinlechner, M. (2000). Species identification by means of the cytochrome b gene. *International Journal of Legal Medicine*, 114(1-2), 23–28.
- Payseur, B. A. (2024). Genetics of recombination rate variation within and between species. *Journal of Evolutionary Biology*, voae158.
- Peers, J. A., Nash, W. J., & Haerty, W. (2025). Gene pseudogenization in fertility-associated genes in cheetah (*Acinonyx jubatus*), a species with long-term low effective population size. *Evolution; International Journal of Organic Evolution*, 79(4), 574–585.
- Peng, X., Alföldi, J., Gori, K., Eisfeld, A. J., Tyler, S. R., Tisoncik-Go, J., Brawand, D., Law, G. L., Skunca, N., Hatta, M., Gasper, D. J., Kelly, S. M., Chang, J., Thomas, M. J., Johnson, J., Berlin, A. M., Lara, M., Russell, P., Swofford, R., ... Katze, M. G. (2014). The draft genome sequence of the ferret (*Mustela putorius furo*) facilitates study of human respiratory disease. *Nature Biotechnology*, 32(12), 1250–1255.
- Pérez-Espona, S. (2021). Conservation-focused biobanks: A valuable resource for wildlife

- DNA forensics. *Forensic Science International. Animals and Environments*, 1(100017), 100017.
- Pérez-González, J., Carranza, J., Anaya, G., Broggin, C., Vedel, G., de la Peña, E., & Membrillo, A. (2023). Comparative analysis of microsatellite and SNP markers for genetic management of Red Deer. *Animals: An Open Access Journal from MDPI*, 13(21). <https://doi.org/10.3390/ani13213374>
- Perez-Sepulveda, B. M., Heavens, D., Pulford, C. V., Predeus, A. V., Low, R., Webster, H., Dykes, G. F., Schudoma, C., Rowe, W., Lipscombe, J., Watkins, C., Kumwenda, B., Shearer, N., Costigan, K., Baker, K. S., Feasey, N. A., Hinton, J. C. D., Hall, N., & 10KSG consortium. (2021). An accessible, efficient and global approach for the large-scale sequencing of bacterial genomes. *Genome Biology*, 22(1), 349.
- Picard toolkit. (2019). In *Broad Institute, GitHub repository*. Broad Institute. <https://broadinstitute.github.io/picard/>
- Pinto, A. V., Hansson, B., Patramanis, I., Morales, H. E., & van Oosterhout, C. (2024). The impact of habitat loss and population fragmentation on genomic erosion. *Conservation Genetics*, 25(1), 49–57.
- Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R., & Siepel, A. (2010). Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Research*, 20(1), 110–121.
- Ponting, C. P., & Hardison, R. C. (2011). What fraction of the human genome is functional? *Genome Research*, 21(11), 1769–1776.
- Powell, R. A., Clark, T. W., Richardson, L., & Forrest, S. C. (1985). Black-footed ferret *Mustela nigripes* energy expenditure and prey requirements. *Biological Conservation*, 34(1), 1–15.
- Primmer, C. R. (2009). From conservation genetics to conservation genomics. *Annals of the New York Academy of Sciences*, 1162(1), 357–368.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American*

Journal of Human Genetics, 81(3), 559–575.

- Puusepp, S., Kovacs-Nagy, R., Alhaddad, B., Braunisch, M., Hoffmann, G. F., Kotzaeridou, U., Lichvarova, L., Liiv, M., Makowski, C., Mandel, M., Meitinger, T., Pajusalu, S., Rodenburg, R. J., Safiulina, D., Strom, T. M., Talvik, I., Vaarmann, A., Wilson, C., Kaasik, A., ... Õunap, K. (2018). Compound heterozygous SPATA5 variants in four families and functional studies of SPATA5 deficiency. *European Journal of Human Genetics: EJHG*, 26(3), 407–419.
- Quilodrán, C. S., Currat, M., & Montoya-Burgos, J. I. (2018). Effect of hybridization with genome exclusion on extinction risk. *Conservation Biology: The Journal of the Society for Conservation Biology*, 32(5), 1139–1149.
- Quilodrán, C. S., Montoya-Burgos, J. I., & Currat, M. (2020). Harmonizing hybridization dissonance in conservation. *Communications Biology*, 3(1), 391.
- Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)*, 26(6), 841–842.
- Quinn, C. B., Preckler-Quisquater, S., Buchalski, M. R., & Sacks, B. N. (2024). Whole genomes inform genetic rescue strategy for Montane red foxes in north America. *Molecular Biology and Evolution*, 41(9), msae193.
- Ralls, K., & Ballou, J. D. (2004). Genetic status and management of California condors. *The Condor*, 106(2), 215–228.
- Ralls, K., Sunnucks, P., Lacy, R. C., & Frankham, R. (2020). Genetic rescue: A critique of the evidence supports maximizing genetic diversity rather than minimizing the introduction of putatively harmful genetic variation. *Biological Conservation*, 251(108784), 108784.
- Rao, S. S. P., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., Sanborn, A. L., Machol, I., Omer, A. D., Lander, E. S., & Aiden, E. L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, 159(7), 1665–1680.
- R Core Team. (2024). *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>.

- Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J., & Kircher, M. (2019). CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Research*, 47(D1), D886–D894.
- Rhie, A., McCarthy, S. A., Fedrigo, O., Damas, J., Formenti, G., Koren, S., Uliano-Silva, M., Chow, W., Fungtammasan, A., Kim, J., Lee, C., Ko, B. J., Chaisson, M., Gedman, G. L., Cantin, L. J., Thibaud-Nissen, F., Haggerty, L., Bista, I., Smith, M., ... Jarvis, E. D. (2021). Towards complete and error-free genome assemblies of all vertebrate species. *Nature*, 592(7856), 737–746.
- Robertson, B. (2006). The role of genetics in kakapo recovery. *Notornis*.
<https://scholar.google.com/citations?user=BI8wSxMAAAAJ&hl=en&oi=sra>
- Robinson, J. A., Bowie, R. C. K., Dudchenko, O., Aiden, E. L., Hendrickson, S. L., Steiner, C. C., Ryder, O. A., Mindell, D. P., & Wall, J. D. (2021). Genome-wide diversity in the California condor tracks its prehistoric abundance and decline. *Current Biology: CB*, 31(13), 2939–2946.e5.
- Robinson, J. A., Brown, C., Kim, B. Y., Lohmueller, K. E., & Wayne, R. K. (2018). Purging of strongly deleterious mutations explains long-term persistence and absence of inbreeding depression in island foxes. *Current Biology: CB*, 28(21), 3487–3494.e4.
- Robinson, J. A., Kyriazis, C. C., Nigenda-Morales, S. F., Beichman, A. C., Rojas-Bracho, L., Robertson, K. M., Fontaine, M. C., Wayne, R. K., Lohmueller, K. E., Taylor, B. L., & Morin, P. A. (2022). The critically endangered vaquita is not doomed to extinction by inbreeding depression. *Science (New York, N.Y.)*, 376(6593), 635–639.
- Robinson, J., Kyriazis, C. C., Yuan, S. C., & Lohmueller, K. E. (2023). Deleterious Variation in Natural Populations and Implications for Conservation Genetics. *Annual Review of Animal Biosciences*, 11, 93–114.
- Ryder, O. (1986). Species conservation and systematics: the dilemma of subspecies. *Trends in Ecology and Evolution*, 1, 9–10.
- Ryder, O. A., & Wedemeyer, E. A. (1982). A cooperative breeding programme for the Mongolian wild horse *Equus przewalskii* in the United States. *Biological Conservation*, 22(4), 259–271.

- Sainsbury, K. A., Shore, R. F., Schofield, H., Croose, E., Campbell, R. D., & McDonald, R. A. (2019). Recent history, current status, conservation and management of native mammalian carnivore species in Great Britain. *Mammal Review*, 49(2), 171–188.
- Sainsbury, K. A., Shore, R. F., Schofield, H., Croose, E., Hantke, G., Kitchener, A. C., & McDonald, R. A. (2020). Diets of European polecat *Mustela putorius* in Great Britain during fifty years of population recovery. *Mammal Research*, 65(2), 181–190.
- Sandler, R. L., Moses, L., & Wisely, S. M. (2021). An ethical analysis of cloning for genetic rescue: Case study of the black-footed ferret. *Biological Conservation*, 257(109118), 109118.
- Santiago, E., Novo, I., Pardiñas, A. F., Saura, M., Wang, J., & Caballero, A. (2020). Recent Demographic History Inferred by High-Resolution Analysis of Linkage Disequilibrium. *Molecular Biology and Evolution*, 37(12), 3642–3653.
- Santymire, R., Branvold-Faber, H., & Marinari, P. E. (2014). Recovery of the black-footed ferret. In *Biology and Diseases of the Ferret* (pp. 219–231). John Wiley & Sons, Inc.
- Santymire, R., & Graves, G. (n.d.). *Black-footed Ferret SAFE Program Action Plan 2019-2021*.
https://assets.speakcdn.com/assets/2332/2019-2021_black-footed_ferret_safe_program_plan-updated_032020.pdf
- Santymire, R. M., Lavin, S. R., Branvold-Faber, H., Kreeger, J., & Marinari, P. (2015). Effect of dietary vitamin E and prey supplementation on semen quality in male black-footed ferrets (*Mustela nigripes*). *Theriogenology*, 84(2), 217–225.
- Santymire, R. M., Livieri, T. M., Branvold-Faber, H., & Marinari, P. E. (2014). The black-footed ferret: on the brink of recovery? *Advances in Experimental Medicine and Biology*, 753, 119–134.
- Santymire, R. M., Lonsdorf, E. V., Lynch, C. M., Wildt, D. E., Marinari, P. E., Kreeger, J. S., & Howard, J. G. (2019). Inbreeding causes decreased seminal quality affecting pregnancy and litter size in the endangered black-footed ferret. *Animal Conservation*, 22(4), 331–340.
- Santymire, R. M., Marinari, P. E., Kreeger, J. S., Wildt, D. E., & Howard, J. (2006). Sperm

- viability in the black-footed ferret (*Mustela nigripes*) is influenced by seminal and medium osmolality. *Cryobiology*, 53(1), 37–50.
- Saravanan, K. A., Panigrahi, M., Kumar, H., Rajawat, D., Nayak, S. S., Bhushan, B., & Dutt, T. (2022). Role of genomics in combating COVID-19 pandemic. *Gene*, 823(146387), 146387.
- Saremi, N. F., Supple, M. A., Byrne, A., Cahill, J. A., Coutinho, L. L., Dalén, L., Figueiró, H. V., Johnson, W. E., Milne, H. J., O'Brien, S. J., O'Connell, B., Onorato, D. P., Riley, S. P. D., Sikich, J. A., Stahler, D. R., Villela, P. M. S., Vollmers, C., Wayne, R. K., Eizirik, E., ... Shapiro, B. (2019). Puma genomes from North and South America provide insights into the genomic consequences of inbreeding. *Nature Communications*, 10(1), 4769.
- Schell, T., Greve, C., & Podsiadlowski, L. (2025). Establishing genome sequencing and assembly for non-model and emerging model organisms: a brief guide. *Frontiers in Zoology*, 22(1), 7.
- Schlötterer, C. (2004). The evolution of molecular markers--just a matter of fashion? *Nature Reviews. Genetics*, 5(1), 63–69.
- Schluter, D. (2000). *The ecology of adaptive radiation*. Oxford University Press.
- Schneider, V. A., Graves-Lindsay, T., Howe, K., Bouk, N., Chen, H.-C., Kitts, P. A., Murphy, T. D., Pruitt, K. D., Thibaud-Nissen, F., Albracht, D., Fulton, R. S., Kremitzki, M., Magrini, V., Markovic, C., McGrath, S., Steinberg, K. M., Auger, K., Chow, W., Collins, J., ... Church, D. M. (2017). Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. *Genome Research*, 27(5), 849–864.
- Schweizer, R. M., Saarman, N., Ramstad, K. M., Forester, B. R., Kelley, J. L., Hand, B. K., Malison, R. L., Ackiss, A. S., Watsa, M., Nelson, T. C., Beja-Pereira, A., Waples, R. S., Funk, W. C., & Luikart, G. (2021). Big Data in Conservation Genomics: Boosting Skills, Hedging Bets, and Staying Current in the Field. *The Journal of Heredity*, 112(4), 313–327.
- Sebastianelli, M., Lukhele, S. M., Secomandi, S., de Souza, S. G., Haase, B., Moysi, M.,

- Nikiforou, C., Hutfluss, A., Mountcastle, J., Balacco, J., Pelan, S., Chow, W., Fedrigo, O., Downs, C. T., Monadjem, A., Dingemanse, N. J., Jarvis, E. D., Brelsford, A., vonHoldt, B. M., & Kirschel, A. N. G. (2024). A genomic basis of vocal rhythm in birds. *Nature Communications*, 15(1), 3095.
- Secomandi, S., Gallo, G. R., Rossi, R., Rodríguez Fernandes, C., Jarvis, E. D., Bonisoli-Alquati, A., Gianfranceschi, L., & Formenti, G. (2025). Pangenome graphs and their applications in biodiversity genomics. *Nature Genetics*, 57(1), 13–26.
- Secomandi, S., Gallo, G. R., Sozzoni, M., Iannucci, A., Galati, E., Abueg, L., Balacco, J., Caprioli, M., Chow, W., Ciofi, C., Collins, J., Fedrigo, O., Ferretti, L., Fungtammasan, A., Haase, B., Howe, K., Kwak, W., Lombardo, G., Masterson, P., ... Formenti, G. (2023). A chromosome-level reference genome and pangenome for barn swallow population genomics. *Cell Reports*, 42(1), 111992.
- Seehausen, O. (2004). Hybridization and adaptive radiation. *Trends in Ecology & Evolution*, 19(4), 198–207.
- Sela, I., Ashkenazy, H., Katoh, K., & Pupko, T. (2015). GUIDANCE2: accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. *Nucleic Acids Research*, 43(W1), W7–W14.
- Sendell-Price, A. T., Ruegg, K. C., & Clegg, S. M. (2020). Rapid morphological divergence following a human-mediated introduction: the role of drift and directional selection. *Heredity*, 124(4), 535–549.
- Senfeld, T., Shannon, T. J., Grouw, H., Pajmans, D. M., Tavares, E. S., Baker, A. J., Lees, A. C., & Collinson, J. M. (2020). Taxonomic status of the extinct Canary Islands Oystercatcher *Haematopus meadewaldoi*. *The Ibis*, 162(3), 1068–1074.
- Senn, H. V., Ghazali, M., Kaden, J., Barclay, D., Harrower, B., Campbell, R. D., Macdonald, D. W., & Kitchener, A. C. (2019). Distinguishing the victim from the threat: SNP-based methods reveal the extent of introgressive hybridization between wildcats and domestic cats in Scotland and inform future in situ and ex situ management options for species restoration. *Evolutionary Applications*, 12(3), 399–414.

- Senn, H. V., & Pemberton, J. M. (2009). Variable extent of hybridization between invasive sika (*Cervus nippon*) and native red deer (*C. elaphus*) in a small geographical area. *Molecular Ecology*, 18(5), 862–876.
- Shafer, A. B. A., Peart, C. R., Tusso, S., Maayan, I., Brelsford, A., Wheat, C. W., & Wolf, J. B. W. (2017). Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. *Methods in Ecology and Evolution*, 8(8), 907–917.
- Shafer, A. B. A., Wolf, J. B. W., Alves, P. C., Bergström, L., Bruford, M. W., Brännström, I., Colling, G., Dalén, L., De Meester, L., Ekblom, R., Fawcett, K. D., Fior, S., Hajibabaei, M., Hill, J. A., Hoezel, A. R., Höglund, J., Jensen, E. L., Krause, J., Kristensen, T. N., ... Zieliński, P. (2015). Genomics and the challenging translation into conservation practice. *Trends in Ecology & Evolution*, 30(2), 78–87.
- 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(W1), W216–W221.
- Sherman, R. M., & Salzberg, S. L. (2020). Pan-genomics in the human genome era. *Nature Reviews. Genetics*, 21(4), 243–254.
- Shumate, A., & Salzberg, S. L. (2021). Liftoff: accurate mapping of gene annotations. *Bioinformatics (Oxford, England)*, 37(12), 1639–1643.
- Siepel, A., Bejerano, G., Pedersen, J. S., Hinrichs, A. S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L. W., Richards, S., Weinstock, G. M., Wilson, R. K., Gibbs, R. A., Kent, W. J., Miller, W., & Haussler, D. (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Research*, 15(8), 1034–1050.
- Siepel, A., & Haussler, D. (2004). Phylogenetic estimation of context-dependent substitution rates by maximum likelihood. *Molecular Biology and Evolution*, 21(3), 468–488.
- Skumatov, D., Abramov, A. V., Herrero, J., Kitchener, A., Maran, T., Kranz, A., Sándor, A.,

- Saveljev, A., Saviour-Soubelet, A., Guinot-Ghestem, M., & Others. (2016). *Mustela putorius*. *The IUCN Red List of Threatened Species*, 2016.
- <https://scholar.google.com/citations?user=J6DSGDMAAAAJ&hl=en&oi=sra>
- Smeds, L., & Ellegren, H. (2023). From high masked to high realized genetic load in inbred Scandinavian wolves. *Molecular Ecology*, 32(7), 1567–1580.
- Smith, S. L., Senn, H. V., Pérez-Espona, S., Wyman, M. T., Heap, E., & Pemberton, J. M. (2018). Introgression of exotic *Cervus* (*nippon* and *canadensis*) into red deer (*Cervus elaphus*) populations in Scotland and the English Lake District. *Ecology and Evolution*, 8(4), 2122–2134.
- Snetkova, V., Ypsilanti, A. R., Akiyama, J. A., Mannion, B. J., Plajzer-Frick, I., Novak, C. S., Harrington, A. N., Pham, Q. T., Kato, M., Zhu, Y., Godoy, J., Meky, E., Hunter, R. D., Shi, M., Kvon, E. Z., Afzal, V., Tran, S., Rubenstein, J. L. R., Visel, A., ... Dickel, D. E. (2021). Ultraconserved enhancer function does not require perfect sequence conservation. *Nature Genetics*, 53(4), 521–528.
- Snyder, N. F. R., & Snyder, H. (2000). The California condor: a saga of natural history and conservation. (*No Title*). <https://cir.nii.ac.jp/crid/1130000794361666048>
- Solow, A. R., Kitchener, A. C., Roberts, D. L., & Birks, J. D. S. (2006). Rediscovery of the Scottish polecat, *Mustela putorius*: Survival or reintroduction? *Biological Conservation*, 128(4), 574–575.
- Speak, S. A., Birley, T., Bortoluzzi, C., Clark, M. D., Percival-Alwyn, L., Morales, H. E., & van Oosterhout, C. (2024). Genomics-informed captive breeding can reduce inbreeding depression and the genetic load in zoo populations. *Molecular Ecology Resources*, 24(7), e13967.
- Spielman, D., Brook, B. W., & Frankham, R. (2004). Most species are not driven to extinction before genetic factors impact them. *Proceedings of the National Academy of Sciences of the United States of America*, 101(42), 15261–15264.
- Spivakov, M. (2014). Spurious transcription factor binding: non-functional or genetically redundant?: Prospects & Overviews. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, 36(8), 798–806.

- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics (Oxford, England)*, 30(9), 1312–1313.
- Stanton, M. M., Tzatzalos, E., Donne, M., Kolundzic, N., Helgason, I., & Ilic, D. (2019). Prospects for the use of induced pluripotent stem cells in animal conservation and environmental protection: Wildlife conservation from a different perspective. *Stem Cells Translational Medicine*, 8(1), 7–13.
- Steinhaus, R., Robinson, P. N., & Seelow, D. (2022). FABIAN-variant: predicting the effects of DNA variants on transcription factor binding. *Nucleic Acids Research*, 50(W1), W322–W329.
- Storey, J. D., & Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences of the United States of America*, 100(16), 9440–9445.
- Stormo, G. D. (2013). Modeling the specificity of protein-DNA interactions. *Quantitative Biology (Beijing, China)*, 1(2), 115–130.
- Stroupe, S., Forgacs, D., Harris, A., Derr, J. N., & Davis, B. W. (2022). Genomic evaluation of hybridization in historic and modern North American Bison (*Bison bison*). *Scientific Reports*, 12(1), 6397.
- Stuart, O. P., Cleave, R., Pearce, K., Magrath, M. J. L., & Mikheyev, A. S. (2025). Purging of highly deleterious mutations through an extreme bottleneck. *Molecular Biology and Evolution*, msaf079.
- Sullivan, P. F., Meadows, J. R. S., Gazal, S., Phan, B. N., Li, X., Genereux, D. P., Dong, M. X., Bianchi, M., Andrews, G., Sakthikumar, S., Nordin, J., Roy, A., Christmas, M. J., Marinescu, V. D., Wang, C., Wallerman, O., Xue, J., Yao, S., Sun, Q., ... Lindblad-Toh, K. (2023). Leveraging base-pair mammalian constraint to understand genetic variation and human disease. *Science (New York, N.Y.)*, 380(6643), eabn2937.
- Sumption, K. J., & Flowerdew, J. R. (1985). The ecological effects of the decline in Rabbits (*Oryctolagus cuniculus* L.) due to myxomatosis. *Mammal Review*, 15(4),

151–186.

- Sundaram, L., Gao, H., Padigepati, S. R., McRae, J. F., Li, Y., Kosmicki, J. A., Fritzilas, N., Hakenberg, J., Dutta, A., Shon, J., Xu, J., Batzoglou, S., Li, X., & Farh, K. K.-H. (2018). Predicting the clinical impact of human mutation with deep neural networks. *Nature Genetics*, 50(8), 1161–1170.
- Supple, M. A., & Shapiro, B. (2018). Conservation of biodiversity in the genomics era. *Genome Biology*, 19(1), 131.
- Suzek, B. E., Huang, H., McGarvey, P., Mazumder, R., & Wu, C. H. (2007). UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics (Oxford, England)*, 23(10), 1282–1288.
- Swegen, A., Appeltant, R., & Williams, S. A. (2023). Cloning in action: can embryo splitting, induced pluripotency and somatic cell nuclear transfer contribute to endangered species conservation? *Biological Reviews of the Cambridge Philosophical Society*, 98(4), 1225–1249.
- Swinnerton, K. J., Groombridge, J. J., Jones, C. G., Burn, R. W., & Mungroo, Y. (2004). Inbreeding depression and founder diversity among captive and free-living populations of the endangered pink pigeon *Columba mayeri*. *Animal Conservation*, 7(4), 353–364.
- Szapu, J. S., Cserkés, T., Pirger, Z., Kiss, C., & Lanszki, J. (2024). Exposure to anticoagulant rodenticides in steppe polecat (*Mustela eversmanii*) and European polecat (*Mustela putorius*) in central Europe. *The Science of the Total Environment*, 948(174282), 174282.
- Szatmári, L., Cserkés, T., Laczkó, L., Lanszki, J., Pertoldi, C., Abramov, A. V., Elmeros, M., Ottlecz, B., Hegyeli, Z., & Sramkó, G. (2021). A comparison of microsatellites and genome-wide SNPs for the detection of admixture brings the first molecular evidence for hybridization between *Mustela eversmanii* and *M. putorius* (Mustelidae, Carnivora). *Evolutionary Applications*, 14(9), 2286–2304.
- Szewczyk, M., Nowak, S., Niedźwiecka, N., Hulva, P., Špinkytė-Bačkaitienė, R., Demjanovičová, K., Bolfíková, B. Č., Antal, V., Fenchuk, V., Figura, M., Tomczak, P.,

- Stachyra, P., Stępnia, K. M., Zwijacz-Kozica, T., & Mysłajek, R. W. (2019). Dynamic range expansion leads to establishment of a new, genetically distinct wolf population in Central Europe. *Scientific Reports*, 9(1), 19003.
- Szpiech, Z. A., Xu, J., Pemberton, T. J., Peng, W., Zöllner, S., Rosenberg, N. A., & Li, J. Z. (2013). Long runs of homozygosity are enriched for deleterious variation. *The American Journal of Human Genetics*, 93(1), 90–102.
- Tallmon, D. A., Luikart, G., & Waples, R. S. (2004). The alluring simplicity and complex reality of genetic rescue. *Trends in Ecology & Evolution*, 19(9), 489–496.
- Tanaka, A. J., Cho, M. T., Millan, F., Juusola, J., Retterer, K., Joshi, C., Niyazov, D., Garnica, A., Gratz, E., Deardorff, M., Wilkins, A., Ortiz-Gonzalez, X., Mathews, K., Panzer, K., Brilstra, E., van Gassen, K. L. I., Volker-Touw, C. M. L., van Binsbergen, E., Sobreira, N., ... Chung, W. K. (2015). Mutations in SPATA5 are associated with microcephaly, intellectual disability, seizures, and hearing loss. *The American Journal of Human Genetics*, 97(3), 457–464.
- Tarailo-Graovac, M., & Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. *Et Al [Current Protocols in Bioinformatics]*, Chapter 4(1), 4.10.1–4.10.14.
- Taylor, S. A., & Larson, E. L. (2019). Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. *Nature Ecology & Evolution*, 3(2), 170–177.
- Tedeschi, L., Biancolini, D., Capinha, C., Rondinini, C., & Essl, F. (2022). Introduction, spread, and impacts of invasive alien mammal species in Europe. *Mammal Review*, 52(2), 252–266.
- Teeling, E. C., Vernes, S. C., Dávalos, L. M., Ray, D. A., Gilbert, M. T. P., Myers, E., & Bat1K Consortium. (2018). Bat biology, genomes, and the Bat1K project: To generate chromosome-level genomes for all living bat species. *Annual Review of Animal Biosciences*, 6(1), 23–46.
- Terrell, K. A., Crosier, A. E., Wildt, D. E., O'Brien, S. J., Anthony, N. M., Marker, L., & Johnson, W. E. (2016). Continued decline in genetic diversity among wild cheetahs

- (*Acinonyx jubatus*) without further loss of semen quality. *Biological Conservation*, 200, 192–199.
- Thanki, A. S., Soranzo, N., Haerty, W., & Davey, R. P. (2018). GeneSeqToFamily: a Galaxy workflow to find gene families based on the Ensembl Compara GeneTrees pipeline. *GigaScience*, 7(3), 1–10.
- Theissinger, K., Fernandes, C., Formenti, G., Bista, I., Berg, P. R., Bleidorn, C., Bombarely, A., Crottini, A., Gallo, G. R., Godoy, J. A., Jentoft, S., Malukiewicz, J., Mouton, A., Oomen, R. A., Paez, S., Palsbøll, P. J., Pampoulie, C., Ruiz-López, M. J., Secomandi, S., ... European Reference Genome Atlas Consortium. (2023). How genomics can help biodiversity conservation. *Trends in Genetics: TIG*, 39(7), 545–559.
- Thomas, J. W., Touchman, J. W., Blakesley, R. W., Bouffard, G. G., Beckstrom-Sternberg, S. M., Margulies, E. H., Blanchette, M., Siepel, A. C., Thomas, P. J., McDowell, J. C., Maskeri, B., Hansen, N. F., Schwartz, M. S., Weber, R. J., Kent, W. J., Karolchik, D., Bruen, T. C., Bevan, R., Cutler, D. J., ... Green, E. D. (2003). Comparative analyses of multi-species sequences from targeted genomic regions. *Nature*, 424(6950), 788–793.
- Thomas, N. E., Hailer, F., Bruford, M. W., & Chadwick, E. A. (2022). Country-wide genetic monitoring over 21 years reveals lag in genetic recovery despite spatial connectivity in an expanding carnivore (Eurasian otter, *Lutra lutra*) population. *Evolutionary Applications*, 15(12), 2125–2141.
- Thorne, E. T. (1989). *Conservation biology and the black-footed ferret*.
- Tibbetts, J. H. (2022). Synthetic biology and endangered species. *Bioscience*, 72(7), 610–617.
- Titus, K. L., & Jachowski, D. S. (2021). Persistent negative stakeholder perspectives limit recovery of a critically endangered carnivore. *Conservation Science and Practice*, 3(11), e526.
- Tomarovsky, A., Khan, R., Dudchenko, O., Totikov, A., Serdyukova, N. A., Weisz, D., Vorobieva, N. V., Bulyonkova, T., Abramov, A. V., Nie, W., Wang, J., Romanenko, S.

- A., Proskuryakova, A. A., Cherkasov, N., Ferguson-Smith, M. A., Yang, F., Balanovskaya, E., Gilbert, M. T. P., Graphodatsky, A. S., ... Kliver, S. (2025). Chromosome length genome assembly of the stone marten (*Martes foina*, Mustelidae): a new view on one of the cornerstones in carnivore cytogenetics. *The Journal of Heredity*, esaf001.
- Totikov, A., Tomarovsky, A., Prokopov, D., Yakupova, A., Bulyonkova, T., Derezanin, L., Rasskazov, D., Wolfsberger, W. W., Koepfli, K.-P., Oleksyk, T. K., & Kliver, S. (2021). Chromosome-level genome assemblies expand capabilities of genomics for conservation biology. *Genes*, 12(9), 1336.
- Turghan, M. A., Jiang, Z., & Niu, Z. (2022). An Update on Status and Conservation of the Przewalski's Horse (*Equus ferus przewalskii*): Captive Breeding and Reintroduction Projects. *Animals: An Open Access Journal from MDPI*, 12(22), 3158.
- van der Auwera, G., & O'Connor, B. D. (2020). *Genomics in the cloud: Using docker, GATK, and WDL in Terra*. O'Reilly Media.
- van der Valk, T., & Dalén, L. (2024). From genomic threat assessment to conservation action. *Cell*, 187(5), 1038–1041.
- van der Valk, T., Díez-Del-Molino, D., Marques-Bonet, T., Guschanski, K., & Dalén, L. (2019). Historical genomes reveal the genomic consequences of recent population decline in eastern gorillas. *Current Biology: CB*, 29(1), 165–170.e6.
- van der Valk, T., Jensen, A., Caillaud, D., & Guschanski, K. (2024). Comparative genomic analyses provide new insights into evolutionary history and conservation genomics of gorillas. *BMC Ecology and Evolution*, 24(1), 14.
- van Oosterhout, C. (2020). Conservation genetics: 50 Years and counting. *Conservation Letters*. <https://doi.org/10.1111/conl.12789>
- van Oosterhout, C. (2024). AI-informed conservation genomics. *Heredity*, 132(1), 1–4.
- van Oosterhout, C., Speak, S. A., Birley, T., Hitchings, L. W. G., Bortoluzzi, C., Percival-Alwyn, L., Urban, L., Groombridge, J. J., Segelbacher, G., & Morales, H. E. (2022). Genomic erosion in the assessment of species' extinction risk and recovery potential. In *bioRxiv* (p. 2022.09.13.507768).

- van Vuurde, M. R., & van der Grift, E. A. (2005). The effects of landscape attributes on the use of small wildlife underpasses by weasel (*Mustela nivalis*) and stoat (*Mustela erminea*). *Lutra*, 48(2), 91–108.
- Vaser, R., Adusumalli, S., Leng, S. N., Sikic, M., & Ng, P. C. (2016). SIFT missense predictions for genomes. *Nature Protocols*, 11(1), 1–9.
- Venkat, A., Hahn, M. W., & Thornton, J. W. (2018). Multinucleotide mutations cause false inferences of lineage-specific positive selection. *Nature Ecology & Evolution*, 2(8), 1280–1288.
- Vernesi, C., Bruford, M. W., Bertorelle, G., Pecchioli, E., Rizzoli, A., & Hauffe, H. C. (2008). Where's the conservation in conservation genetics? *Conservation Biology: The Journal of the Society for Conservation Biology*, 22(3), 802–804.
- Virgós, E. (2003). Association of the polecat *Mustela putorius* in eastern Spain with montane pine forests. *Oryx*, 37(4), 484–487.
- Vitsios, D., Dhindsa, R. S., Middleton, L., Gussow, A. B., & Petrovski, S. (2021). Prioritizing non-coding regions based on human genomic constraint and sequence context with deep learning. *Nature Communications*, 12(1), 1504.
- vonHoldt, B. M., Cahill, J. A., Fan, Z., Gronau, I., Robinson, J., Pollinger, J. P., Shapiro, B., Wall, J., & Wayne, R. K. (2016). Whole-genome sequence analysis shows that two endemic species of North American wolf are admixtures of the coyote and gray wolf. *Science Advances*, 2(7), e1501714.
- von Seth, J., Dussex, N., Díez-Del-Molino, D., van der Valk, T., Kutschera, V. E., Kierczak, M., Steiner, C. C., Liu, S., Gilbert, M. T. P., Sinding, M.-H. S., Prost, S., Guschanski, K., Nathan, S. K. S. S., Brace, S., Chan, Y. L., Wheat, C. W., Skoglund, P., Ryder, O. A., Goossens, B., ... Dalén, L. (2021). Genomic insights into the conservation status of the world's last remaining Sumatran rhinoceros populations. *Nature Communications*, 12(1), 2393.
- von Takach, B., Cameron, S. F., Cremona, T., Eldridge, M. D. B., Fisher, D. O., Hohnen, R., Jolly, C. J., Kelly, E., Phillips, B. L., Radford, I. J., Rick, K., Spencer, P. B. S.,

- Trewella, G. J., Umbrello, L. S., & Banks, S. C. (2024). Conservation prioritisation of genomic diversity to inform management of a declining mammal species. *Biological Conservation*, 291(110467), 110467.
- von Thaden, A., Nowak, C., Tiesmeyer, A., Reiners, T. E., Alves, P. C., Lyons, L. A., Mattucci, F., Randi, E., Cragolini, M., Galián, J., Hegyeli, Z., Kitchener, A. C., Lambinet, C., Lucas, J. M., Mölich, T., Ramos, L., Schockert, V., & Cocchiararo, B. (2020). Applying genomic data in wildlife monitoring: Development guidelines for genotyping degraded samples with reduced single nucleotide polymorphism panels. *Molecular Ecology Resources*, 20(3), 662–680.
- Wang, C., Wallerman, O., Arendt, M.-L., Sundström, E., Karlsson, Å., Nordin, J., Mäkeläinen, S., Pielberg, G. R., Hanson, J., Ohlsson, Å., Saellström, S., Rönnerberg, H., Ljungvall, I., Häggström, J., Bergström, T. F., Hedhammar, Å., Meadows, J. R. S., & Lindblad-Toh, K. (2021). A novel canine reference genome resolves genomic architecture and uncovers transcript complexity. *Communications Biology*, 4(1), 185.
- Wang, J., & Hill, W. G. (1999). Effect of selection against deleterious mutations on the decline in heterozygosity at neutral loci in closely inbreeding populations. *Genetics*, 153(3), 1475–1489.
- Warnock-Juteau, K., Bolduc, V., LoScerbo, D., Anderson, M., Daguet, C., & Jaeger, J. A. G. (2022). Co-use of existing crossing structures along roads by wildlife and humans: Wishful thinking? *Nature Conservation*, 47, 235–270.
- Watowich, M. M., Chiou, K. L., Graves, B., Montague, M. J., Brent, L. J. N., Higham, J. P., Horvath, J. E., Lu, A., Martinez, M. I., Platt, M. L., Schneider-Crease, I. A., Lea, A. J., & Snyder-Mackler, N. (2023). Best practices for genotype imputation from low-coverage sequencing data in natural populations. *Molecular Ecology Resources*. <https://doi.org/10.1111/1755-0998.13854>
- Weinstein, J. N., Collisson, E. A., Mills, G. B., Shaw, K. R., Ozenberger, B. A., Ellrott, K., & Stuart J. (2013). The cancer genome atlas pan-cancer analysis project. *Nature Genetics*, 45(10), 1113–1120.
- Weirauch, M. T., Cote, A., Norel, R., Annala, M., Zhao, Y., Riley, T. R., Saez-Rodriguez, J.,

- Cokelaer, T., Vedenko, A., Talukder, S., DREAM5 Consortium, Bussemaker, H. J., Morris, Q. D., Bulyk, M. L., Stolovitzky, G., & Hughes, T. R. (2013). Evaluation of methods for modeling transcription factor sequence specificity. *Nature Biotechnology*, 31(2), 126–134.
- Weirauch, M. T., Yang, A., Albu, M., Cote, A. G., Montenegro-Montero, A., Drewe, P., Najafabadi, H. S., Lambert, S. A., Mann, I., Cook, K., Zheng, H., Goity, A., van Bakel, H., Lozano, J.-C., Galli, M., Lewsey, M. G., Huang, E., Mukherjee, T., Chen, X., ... Hughes, T. R. (2014). Determination and inference of eukaryotic transcription factor sequence specificity. *Cell*, 158(6), 1431–1443.
- Welles, S. R., & Dlugosch, K. M. (2018). Population Genomics of Colonization and Invasion. In *Population Genomics* (pp. 655–683). Springer International Publishing.
- Wetterstrand, K. A. (2013). DNA sequencing costs: data from the NHGRI Genome Sequencing Program (GSP). *National Human Genome Research Institute*.
<https://cir.nii.ac.jp/crid/1370002219373721113>
- Wetzel, C. (2024, April 9). *Northern white rhino could be saved from extinction using frozen skin*. New Scientist.
<https://www.newscientist.com/article/2425679-northern-white-rhino-could-be-saved-from-extinction-using-frozen-skin/>
- Whiteley, A. R., Fitzpatrick, S. W., Funk, W. C., & Tallmon, D. A. (2015). Genetic rescue to the rescue. *Trends in Ecology & Evolution*, 30(1), 42–49.
- White, R. J., & Razgour, O. (2020). Emerging zoonotic diseases originating in mammals: a systematic review of effects of anthropogenic land-use change. *Mammal Review*, 50(4), 336–352.
- Wilder, A. P., Steiner, C. C., Hendricks, S., Haller, B. C., Kim, C., Korody, M. L., & Ryder, O. A. (2024). Genetic load and viability of a future restored northern white rhino population. *Evolutionary Applications*, 17(4), e13683.
- Wilder, A. P., Supple, M. A., Subramanian, A., Mudide, A., Swofford, R., Serres-Armero, A., Steiner, C., Koepfli, K.-P., Genereux, D. P., Karlsson, E. K., Lindblad-Toh, K., Marques-Bonet, T., Munoz Fuentes, V., Foley, K., Meyer, W. K., Zoonomia

- Consortium†, Ryder, O. A., & Shapiro, B. (2023). The contribution of historical processes to contemporary extinction risk in placental mammals. *Science (New York, N.Y.)*, 380(6643), eabn5856.
- Willcox, D. (2020). Conservation status, extirpation priorities and emerging threats to small carnivores. *The International Zoo Yearbook*, 54(1), 19–34.
- Williams, E. S., Thorne, E. T., Appel, M. J., & Belitsky, D. W. (1988). Canine distemper in black-footed ferrets (*Mustela nigripes*) from Wyoming. *Journal of Wildlife Diseases*, 24(3), 385–398.
- Williams, J. G., Kubelik, A. R., Livak, K. J., Rafalski, J. A., & Tingey, S. V. (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research*, 18(22), 6531–6535.
- Willi, Y., Kristensen, T. N., Sgrò, C. M., Weeks, A. R., Ørsted, M., & Hoffmann, A. A. (2022). Conservation genetics as a management tool: The five best-supported paradigms to assist the management of threatened species. *Proceedings of the National Academy of Sciences of the United States of America*, 119(1), e2105076119.
- Willoughby, J. R., Ivy, J. A., Lacy, R. C., Doyle, J. M., & DeWoody, J. A. (2017). Inbreeding and selection shape genomic diversity in captive populations: Implications for the conservation of endangered species. *PloS One*, 12(4), e0175996.
- Wisely, S. M., Buskirk, S. W., Fleming, M. A., McDonald, D. B., & Ostrander, E. A. (2002). Genetic diversity and fitness in black-footed ferrets before and during a bottleneck. *The Journal of Heredity*, 93(4), 231–237.
- Wisely, S. M., McDonald, D. B., & Buskirk, S. W. (2003). Evaluation of the genetic management of the endangered black-footed ferret (*Mustela nigripes*). *Zoo Biology*, 22(3), 287–298.
- Wisely, S. M., Ryder, O. A., Santymire, R. M., Engelhardt, J. F., & Novak, B. J. (2015). A road map for 21st century genetic restoration: Gene pool enrichment of the black-footed ferret. *The Journal of Heredity*, 106(5), 581–592.
- Wisely, S. M., Statham, M. J., & Fleischer, R. C. (2008). Pleistocene Refugia and

- Holocene expansion of a grassland-dependent species, the black-footed ferret (*Mustela nigripes*). *Journal of Mammalogy*, 89(1), 87–96.
- Wittmann, M. J., Stuis, H., & Metzler, D. (2018). Genetic Allee effects and their interaction with ecological Allee effects. *The Journal of Animal Ecology*, 87(1), 11–23.
- Wold, J., Koepfli, K.-P., Galla, S. J., Eccles, D., Hogg, C. J., Le Lec, M. F., Guhlin, J., Santure, A. W., & Steeves, T. E. (2021). Expanding the conservation genomics toolbox: Incorporating structural variants to enhance genomic studies for species of conservation concern. *Molecular Ecology*, 30(23), 5949–5965.
- Wong, E. S., Zheng, D., Tan, S. Z., Bower, N. L., Garside, V., Vanwalleghem, G., Gaiti, F., Scott, E., Hogan, B. M., Kikuchi, K., McGlinn, E., Francois, M., & Degnan, B. M. (2020). Deep conservation of the enhancer regulatory code in animals. *Science (New York, N.Y.)*, 370(6517), eaax8137.
- Wootton, E., Robert, C., Taillon, J., Côté, S. D., & Shafer, A. B. A. (2023). Genomic health is dependent on long-term population demographic history. *Molecular Ecology*, 32(8), 1943–1954.
- Wragg, D., Zhang, W., Peterson, S., Yerramilli, M., Mellanby, R., Schoenebeck, J. J., & Clements, D. N. (2024). A cautionary tale of low-pass sequencing and imputation with respect to haplotype accuracy. *Genetics, Selection, Evolution: GSE*, 56(1), 6.
- Wray, G. (2007). The evolutionary significance of cis-regulatory mutations. *Nature Reviews. Genetics*, 8, 206–216.
- Wright, B., Willet, C. E., Hamede, R., Jones, M., Belov, K., & Wade, C. M. (2017). Variants in the host genome may inhibit tumour growth in devil facial tumours: evidence from genome-wide association. *Scientific Reports*, 7(1), 423.
- Wright, S. (1931). Evolution in Mendelian populations. *Genetics*, 16(2), 97–159.
- Wright, S. (1969). *Evolution and the genetics of populations. Vol. 2. The theory of gene frequencies*. <https://doi.org/10.5555/19710101332>
- Wu, M. Y., Forcina, G., Low, G. W., Sadanandan, K. R., Gwee, C. Y., van Grouw, H., Wu, S., Edwards, S. V., Baldwin, M. W., & Rheindt, F. E. (2023). Historic samples reveal loss of wild genotype through domestic chicken introgression during the

- Anthropocene. *PLoS Genetics*, 19(1), e1010551.
- Wu, Y., Wang, C., Fan, X., Ma, Y., Liu, Z., Ye, X., Shen, C., & Wu, C. (2024). The impact of induced pluripotent stem cells in animal conservation. *Veterinary Research Communications*, 48(2), 649–663.
- Xue, J. R., Mackay-Smith, A., Mouri, K., Garcia, M. F., Dong, M. X., Akers, J. F., Noble, M., Li, X., Zoonomia Consortium†, Lindblad-Toh, K., Karlsson, E. K., Noonan, J. P., Capellini, T. D., Brennand, K. J., Tewhey, R., Sabeti, P. C., & Reilly, S. K. (2023). The functional and evolutionary impacts of human-specific deletions in conserved elements. *Science (New York, N.Y.)*, 380(6643), eabn2253.
- Xue, Y., Prado-Martinez, J., Sudmant, P. H., Narasimhan, V., Ayub, Q., Szpak, M., Frandsen, P., Chen, Y., Yngvadottir, B., Cooper, D. N., de Manuel, M., Hernandez-Rodriguez, J., Lobon, I., Siegismund, H. R., Pagani, L., Quail, M. A., Hvilsom, C., Mudakikwa, A., Eichler, E. E., ... Scally, A. (2015). Mountain gorilla genomes reveal the impact of long-term population decline and inbreeding. *Science (New York, N.Y.)*, 348(6231), 242–245.
- Yang, L., Jin, H., Yang, Q., Poyarkov, A., Korablev, M., Rozhnov, V., Shao, J., Fu, Q., Hernandez-Blanco, J. A., Zhan, X., Yu, L., Alexandrov, D., Dai, Q., Munkhtsog, B., Du, X., Munkhtsog, B., Ma, L., Chen, W., Malykh, S., ... Hu, Y. (2025). Genomic evidence for low genetic diversity but purging of strong deleterious variants in snow leopards. *Genome Biology*, 26(1), 94.
- Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, 24(8), 1586–1591.
- Yoder, A. D., Poelstra, J. W., Tiley, G. P., & Williams, R. C. (2018). Neutral theory is the foundation of conservation genetics. *Molecular Biology and Evolution*, 35(6), 1322–1326.
- Yoshida, K., Ravinet, M., Makino, T., Toyoda, A., Kokita, T., Mori, S., & Kitano, J. (2020). Accumulation of Deleterious Mutations in Landlocked Threespine Stickleback Populations. *Genome Biology and Evolution*, 12(4), 479–492.
- Zanini, S. F., Bayer, P. E., Wells, R., Snowdon, R. J., Batley, J., Varshney, R. K., Nguyen,

- H. T., Edwards, D., & Golicz, A. A. (2022). Pangenomics in crop improvement-from coding structural variations to finding regulatory variants with pangenome graphs. *The Plant Genome*, 15(1), e20177.
- Zemke, N. R., Armand, E. J., Wang, W., Lee, S., Zhou, J., Li, Y. E., Liu, H., Tian, W., Nery, J. R., Castanon, R. G., Bartlett, A., Osteen, J. K., Li, D., Zhuo, X., Xu, V., Chang, L., Dong, K., Indralingam, H. S., Rink, J. A., ... Ren, B. (2023). Conserved and divergent gene regulatory programs of the mammalian neocortex. *Nature*, 624(7991), 390–402.
- Zhang, C., Rabiee, M., Sayyari, E., & Mirarab, S. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*, 19(Suppl 6), 153.
- Zhang, G. (2015). Bird sequencing project takes off. *Nature*, 522(7554), 34–34.
- Zhang, W., Li, W., Liu, G., Gu, L., Ye, K., Zhang, Y., Li, W., Jiang, D., Wang, Z., & Fang, M. (2021). Evaluation for the effect of low-coverage sequencing on genomic selection in large yellow croaker. *Aquaculture (Amsterdam, Netherlands)*, 534(736323), 736323.
- Zhao, R., Talenti, A., Fang, L., Liu, S., Liu, G., Chue Hong, N. P., Tenesa, A., Hassan, M., & Prendergast, J. G. D. (2022). The conservation of human functional variants and their effects across livestock species. *Communications Biology*, 5(1), 1003.
- Zhao, S., Zheng, P., Dong, S., Zhan, X., Wu, Q., Guo, X., Hu, Y., He, W., Zhang, S., Fan, W., Zhu, L., Li, D., Zhang, X., Chen, Q., Zhang, H.-M., Zhang, Z., Jin, X.-L., Zhang, J., Yang, H., ... Wei, F. (2012). Whole-genome sequencing of giant pandas provides insights into demographic history and local adaptation. *Nature Genetics*, 45, 67–71.
- Zhen, Y., & Andolfatto, P. (2012). Methods to detect selection on noncoding DNA. *Methods in Molecular Biology (Clifton, N.J.)*, 856, 141–159.
- Zimmerman, S. J., Aldridge, C. L., & Oyler-McCance, S. J. (2020). An empirical comparison of population genetic analyses using microsatellite and SNP data for a species of conservation concern. *BMC Genomics*, 21(1), 382.
- Zoonomia Consortium. (2020). A comparative genomics multitool for scientific discovery and conservation. *Nature*, 587(7833), 240–245.

Appendix

A.1 Chapter Two - Population Genomics of the European polecat (*Mustela putorius*) in Britain.

Supplementary Figures and Tables:

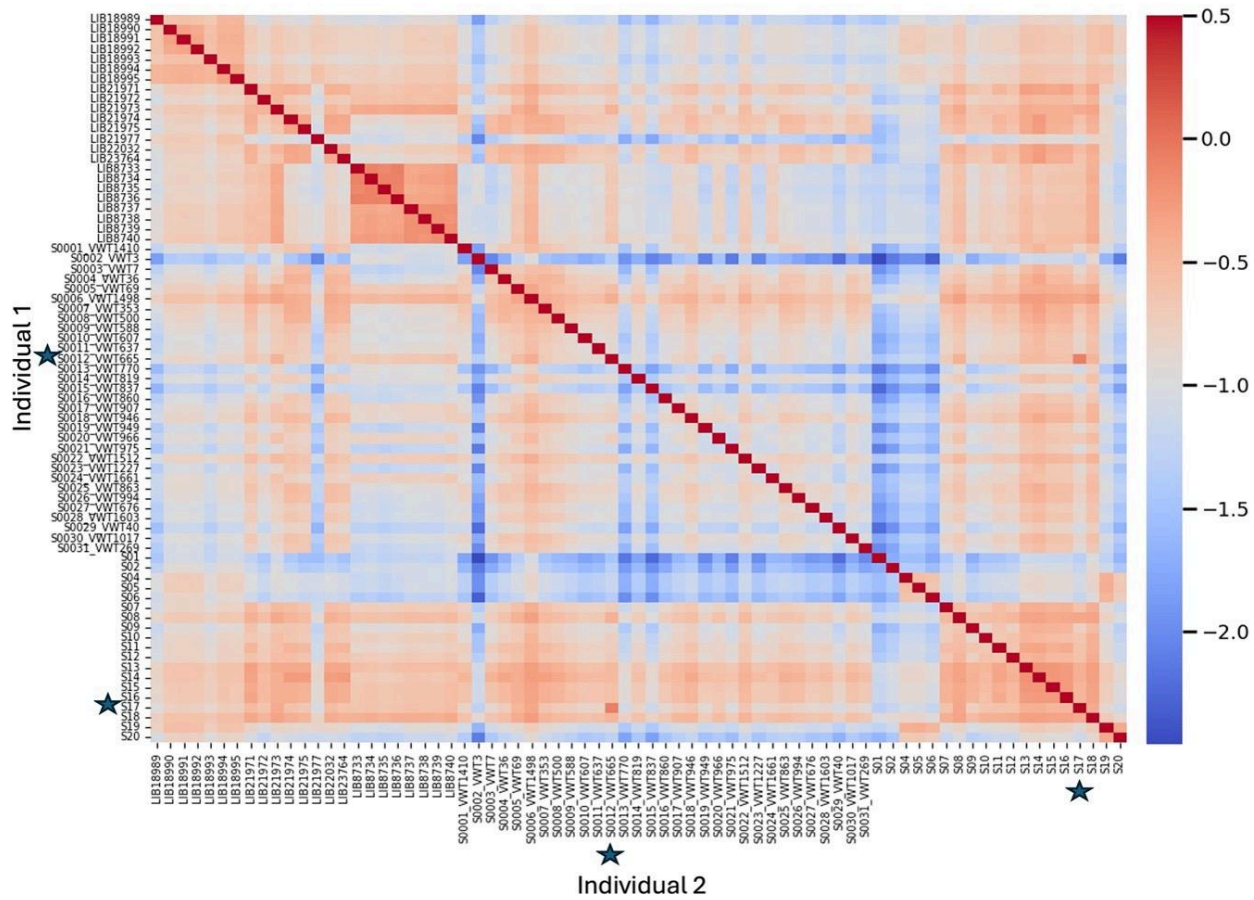


Figure S2.1: Kinship coefficients between each of the 65 polecat and 8 domestic ferret individuals. KING estimates a relatedness-PHI score which can range from >0.354 , $[0.177, 0.354]$, $[0.0884, 0.177]$ and $[0.0442, 0.0884]$ and corresponds to duplicate/MZ twin, 1st-degree, 2nd-degree, and 3rd-degree relationships respectively. Starred individuals (S17 & S0012_VWT665) were identified as having a 3rd-degree relationship (0.076) and S17 was removed from further analyses.

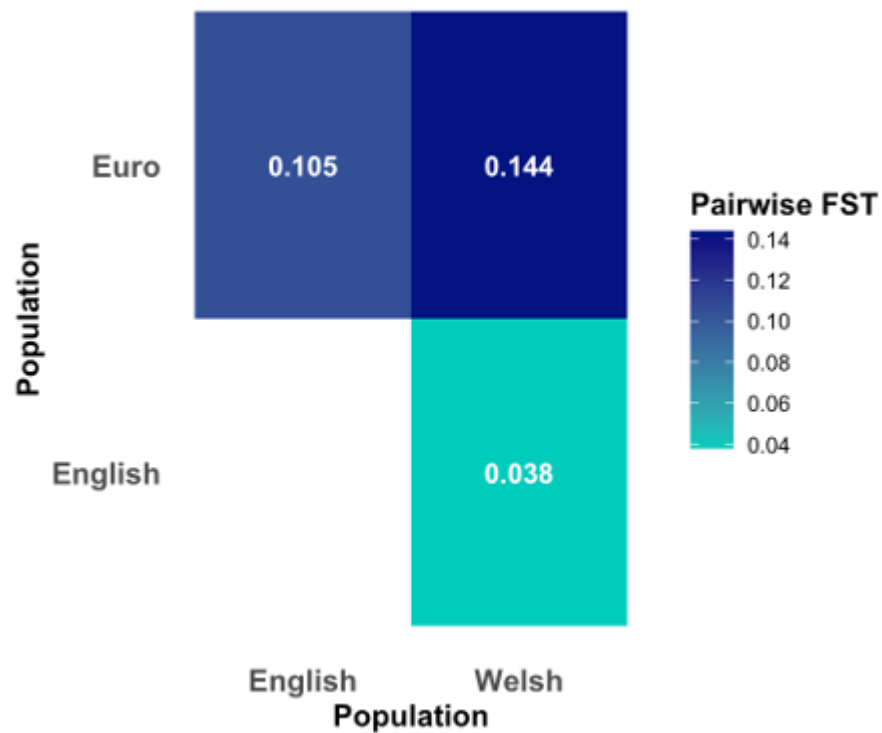


Figure S2.2: Pairwise F_{ST} values for whole genome SNPs between English, Welsh and European populations as calculated in VCFtools version 0.1.16 (Danecek et al., 2011).

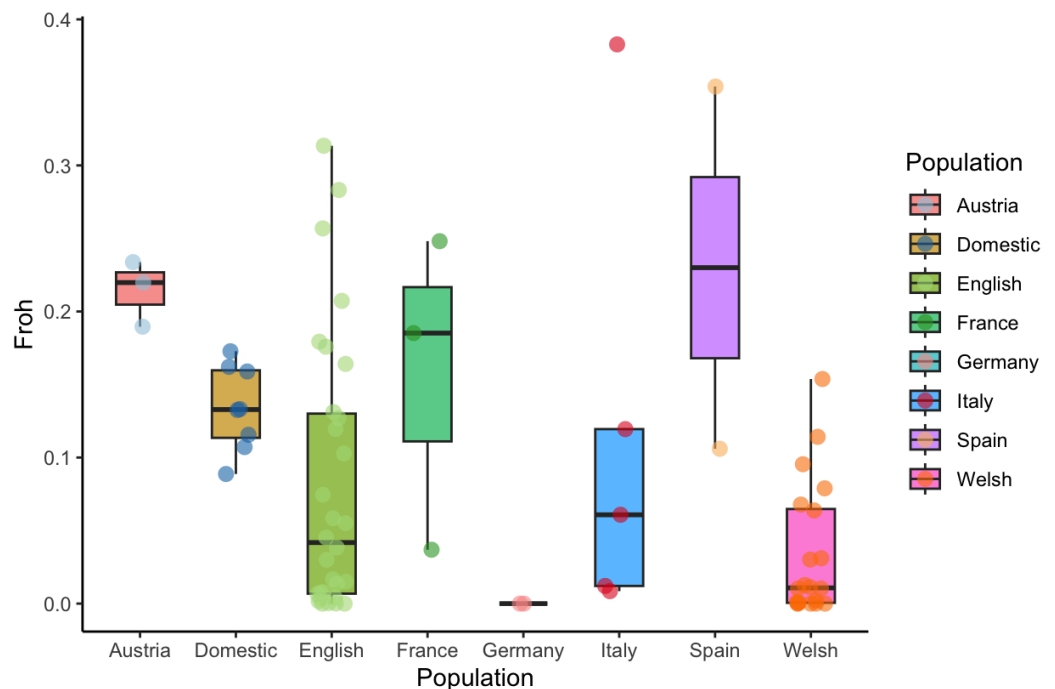


Figure S2.3: The realised inbreeding coefficient (F_{ROH}) as calculated by PLINK version 1.9 for each individual, grouped by sampling location for the British (England and Wales) and European (Austria, France, Germany, Italy and Spain) polecat populations as well as Domestic ferret individuals. Individual F_{ROH} are overlaid on top of boxplots which show the upper, median and lower limits of the calculated F_{ROH} for each group.

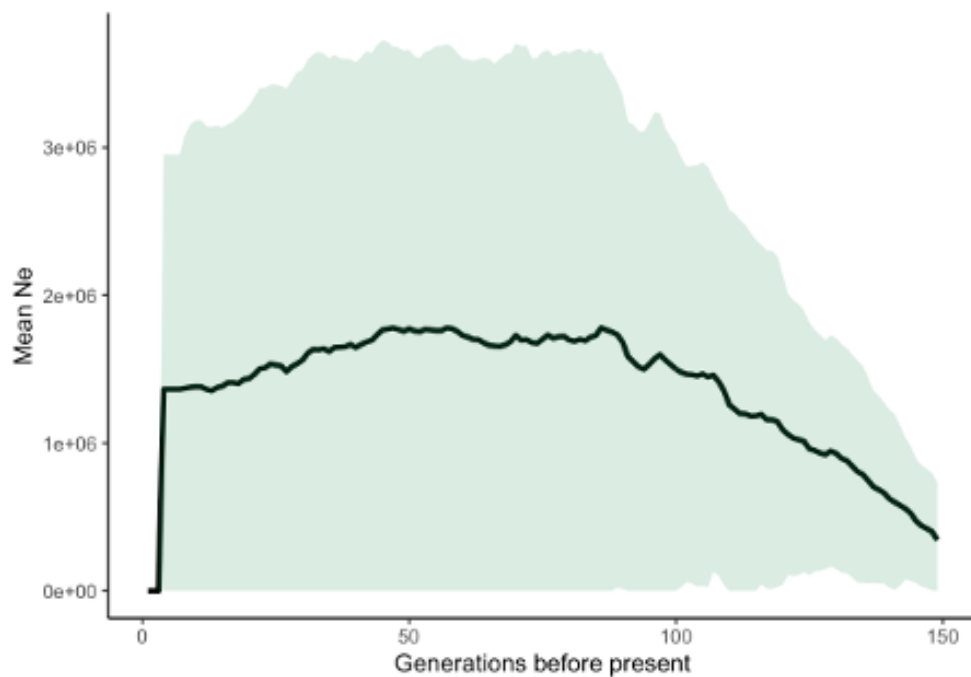


Figure S2.4: Estimated mean effective population size (solid line) with 95% confidence intervals (shaded area) for the Italian polecat population ($n = 5$) across 200 generations (generations = 4 years) based on whole genome SNP data computed in GONE.

Table 2.01: Details of the downsampled variant dataset and number of variants used for all analyses carried out in chapter two. Variants were called using GATK, with biallelic SNPs and quality filters (Q20, AC3) being applied for all downstream analyses. Further filters were then applied for population structure (LD, MAF).

Analyses	Number of Variants	Filters Applied
Relatedness, Whole-genome SNP phylogeny	2,941,546	Biallelic SNPs, Q20, AC3, LD pruned ($r^2 > 0.6$)
PCA, Admixture	595,820	Biallelic SNPs, Q20, AC3, LD pruned ($r^2 > 0.6$), MAF < 0.05
Genomic Diversity, ROH, Mutation load, Demographic Reconstruction	3,836,092 (welsh), 5,126,466 (european), 4,126,068 (english)	Biallelic SNPS, Q20, AC3

Table 2.02: Cross-validation errors from ADMIXTURE runs for all 72 samples using a population size (K) of 1-12. Ten replicates for each ancestry run were carried out using the custom python wrapper (admixturerwrapper.py), with the standard deviation for each mean CV error presented below.

K	CV Average	CV Standard Deviation
1	0.7504	0.0001
2	0.6969	0.0001
3	0.6833	0.0004
4	0.7170	0.0068
5	0.7696	0.0049
6	0.8253	0.006
7	0.9138	0.0275
8	0.9853	0.0258
9	1.0646	0.0238
10	1.1774	0.0363
11	1.2884	0.0438
12	1.3962	0.044

Table 2.03: Output statistic tables for 1) kruskal wallis chi-squared test plus dunn posthoc test for genome-wide diversity estimates 2) ANOVA for the proportion of heterozygous sites across the genome per population and posthoc HSD tukey test 3) ANOVA for the inbreeding coefficient across populations and posthoc HSD tukey test 4) ANOVA for derived mutation load across populations for heterozygous and homozygous state missense and lof variants.

1. Genome-wide nucleotide diversity Df		chi-squared(H)	P-value
	2	7.1443	0.03
Dunn's Test			
Population	Z stat	P-value	
Euro x English	2.189	0.0143*	
Welsh x English	2.423	0.0077*	
Welsh x Euro	0.234	0.41	

2. Proportion Het Sites	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Population	2	0.02859	0.014296	18.65	4.55E-07 ***
Residuals	62	0.04751	0.000766		
HSD Tukey					
Population	diff	lwr	upr	p adj	
English-Welsh	0.03027198	0.01108232	0.04946164	0.0009953	
Euro-Welsh	-0.0208865	-0.043592	0.001819	0.0775475	
Euro-English	-0.0511585	-0.0721797	-0.0301373	0.0000006	
2. Inbreeding coefficient	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Population	2	0.1798	0.08991	18.06	6.61E07 ***
Residuals	62	0.3087	0.00498		
HSD Tukey					
Population	diff	lwr	upr	p adj	
English-Welsh	-0.0764358	-0.1253476	-0.0275241	0.0011157	
Euro-Welsh	0.05165117	-0.006222	0.10952436	0.089492	
Euro-English	0.128087	0.07450683	0.18166717	0.0000009	
4. Runs of Homozygosity	Df	chi-squared(H) P-value			
	2	28.0230	8.22E-07		
Dunn's Test					
Population	Z stat	P-value			
Euro x English	-5.289	0.000			
Welsh x English	-1.548	0.061			
Welsh x Euro	2.969	0.002			
5. Mutation Load					
5.1 Heterozygous LOF derived	Df	Sum_Sq	Mean_Sq	F value	Pr(>F)
Population	2	55.96	27.98	5.584	0.00597 **
Residuals	60	300.64	5.011		

HSD Tukey					
Population	diff	lwr	upr	p adj	
Euro-English	-1.92E+00	-3.63E+00	-2.11E-01	0.0241113	
Welsh-English	-1.87E+00	-3.45E+00	-2.78E-01	0.0174249	
Welsh-Euro	0.05614035	-1.80192	1.9142006	0.9970975	
5.2 Homozygous LOF derived					
	Df	Sum_Sq	Mean_Sq	F value	Pr(>F)
Population	2	183.1	91.54	16.81	1.59E-06
Residuals	60	326.6	5.44		
HSD Tukey					
Population	diff	lwr	upr	p adj	
Euro-English	-3.388506	-5.1718517	-1.60516	0.000074	
Welsh-English	1.134301	-0.5207015	2.789304	0.2341289	
Welsh-Euro	4.522807	2.5860688	6.459545	0.0000016	
5.3 Heterozygous missense derived					
	Df	Sum_Sq	Mean_Sq	F value	Pr(>F)
Population	2	132333	66166	14.77	6.1E-06 ***
Residuals	60	268878	4481		
HSD Tukey					
Population	diff	lwr	upr	p adj	
Euro-English	-113.38851	-164.55393	-62.22308	0.0000047	
Welsh-English	-58.49728	-105.98045	-11.01411	0.0120458	
Welsh-Euro	54.89123	-0.6751293	110.45759	0.0535361	
5.4 Homozygous missense derived					
	Df	Sum_Sq	Mean_Sq	F value	Pr(>F)
Population	2	591323	295661	31.39	4.68E-10 ***
Residuals	60	565123	9419		
HSD Tukey					
Population	diff	lwr	upr	p adj	
Euro-English	-144.4966	-218.67377	-70.31933	0.0000492	
Welsh-English	121.0508	52.21196	189.88967	0.0002399	
Welsh-Euro	265.5474	184.98989	346.10485	0.000	

All supplementary tables listed within this chapter are available at [https://github.com/EarlhamInst/RShaw-PhD-Thesis/blob/main/RSThesis_ChapterTwo_Supplementary_Materials%20\(5\).xlsx](https://github.com/EarlhamInst/RShaw-PhD-Thesis/blob/main/RSThesis_ChapterTwo_Supplementary_Materials%20(5).xlsx)

Supplementary Material:

S1 - Geographic Origin of British Samples

S2 - Sample sequencing metadata

S3 - Site data per individual

S4 - Ancestry group output identified by Admixture

S5 - Nucleotide diversity across the genome per individual

S6 - ROH data per individual

S7 - SnpEff summary output

S8 - GONE effective population size predictions for Welsh population

S9 - GONE effective population size predictions for English population

S10 - Chapter two statistic summary table

A.2 Chapter Three - Characterisation of deleterious mutations within protein-coding genes in the Black-footed ferret (*Mustela nigripes*).

Supplementary Figures and Tables:

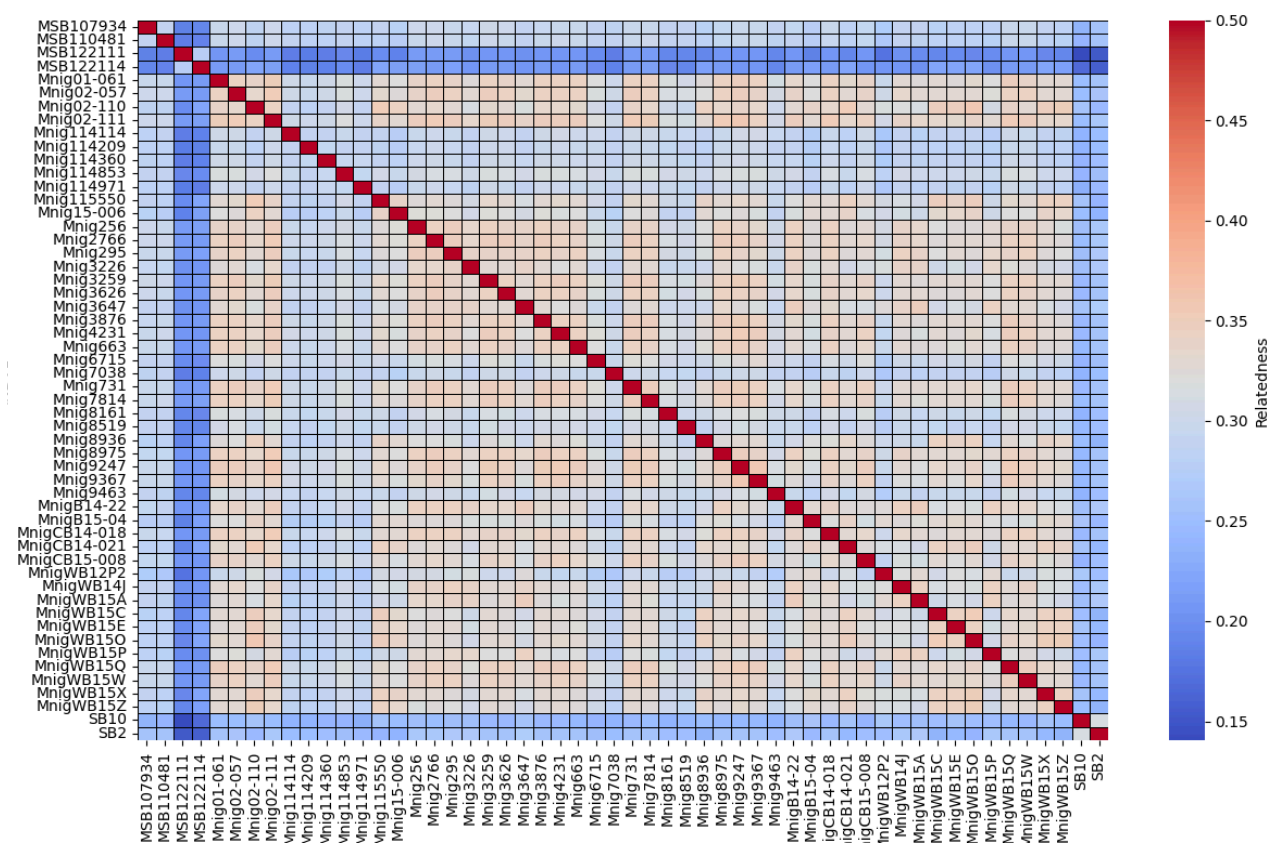


Figure S3.1: Pairwise kinship coefficients between black-footed ferret individuals, where estimates of 0.17-0.35 indicate first-degree relations, second-degree relations correspond to 0.08-0.17 and third-degree relations as between 0.04-0.08.

Table 3.01: Details of the variant dataset and number of variants used for all analyses carried out in chapter three. Variants were called using GATK, with biallelic SNPs and quality filters (Q20, AC3) being applied for all downstream analyses. Further filters were then applied for population structure (LD).

Analyses	Number of Variants	Filters Applied
Relatedness, PCA, Admixture	944,698	Biallelic SNPs, Q20, AC3, LD pruned (r2 >0.6)
Genomic Diversity, ROH, Mutation load	1,927,827	Biallelic SNPS, Q20, AC3

Table 3.02: Cross-validation errors from ADMIXTURE runs for all 54 samples using a population size (K) of 1-12. Ten replicates for each ancestry run were carried out using the custom python wrapper (admixturewrapper.py), with the standard deviation for each mean CV error presented in the table.

K	CV Average	CV Standard Deviation
1	0.4073	0.0005
2	0.4696	0.0017
3	0.5538	0.002
4	0.6428	0.0102
5	0.7289	0.011
6	0.8004	0.0059
7	0.877	0.0089
8	0.9423	0.0091
9	1.0088	0.0116
10	1.061	0.0191
11	1.1183	0.0174
12	1.1599	0.0147

Table 3.03: Output statistic tables for 1) ANOVA plus HSD tukey posthoc test for genome-wide diversity estimates 2) ANOVA for the proportion of heterozygous sites across the genome per population and posthoc HSD tukey test 3) ANOVA for the inbreeding coefficient across populations and posthoc HSD tukey test 4) kruskal wallis chi-squared test plus dunns post hoc test for comparison of Froh across populations 5) ANOVA for derived and non-derived mutation load across populations for heterozygous and homozygous state missense and lof variants.

1. Genome-wide nucleotide diversity					
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Population	2	1.81E-11	9.04E-12	18.98	1.37e-08 ***
Residuals	388	1.85E-10	4.76E-13		
HSD Tukey					
Population	diff	lwr	upr	p adj	

Captive-Wild	1.22E-07	-7.95E-08	3.23E-07	0.3301864	
Reintroduced-Wild	5.05E-07	3.04E-07	7.07E-07	0	
Reintroduced-Captive	3.84E-07	1.83E-07	5.85E-07	0.000028	
2. Proportion Het Sites					
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Population	6	0.09529	0.015881	9.698	6.12e-07***
Residuals	47	0.07696	0.001638		
HSD Tukey					
Population	diff	lwr	upr	p adj	
	0.0230012				
#Captive/Wyoming-Captive/Colorado	6	-0.0497591	0.0957616	0.9572756	
			0.0054682		
#Captive/Virginia-Captive/Colorado	-0.0476684	-0.100805	8	0.1053334	
#Reintroduced/Arizona-Captive/Colorado	0.0254049		0.0785415		
	4	-0.0277317	7	0.7593175	
#Reintroduced/South_Dakota-Captive/Colorado	0.0304938		0.0836305		
	9	-0.0226427	2	0.5759084	
#Wild/South_Dakota-Captive/Colorado	-0.0982748	-0.1940682	-0.0024814	0.0409426	
#Wild/Wyoming-Captive/Colorado	-0.0925552	-0.1653155	-0.0197949	0.0050179	
	0.0706696		0.1434299		
#Captive/Wyoming-Captive/Virginia	2	-0.0020907	5	0.0620639	
#Reintroduced/Arizona-Captive/Virginia	0.0730732		0.1262099		
	9	0.01993666	3	0.0019108	
#Reintroduced/South_Dakota-Captive/Virginia	0.0781622		0.1312988		
	4	0.02502561	7	0.0007537	
			0.0451869		
#Wild/South_Dakota-Captive/Virginia	-0.0506064	-0.1463999	8	0.6660848	
			0.0278734		
#Wild/Wyoming-Captive/Virginia	-0.0448868	-0.1176472	9	0.4905921	
#Reintroduced/Arizona-Captive/Wyoming	0.0024036		0.0751640		
	8	-0.0703567	1	0.9999999	
#Reintroduced/South_Dakota-Captive/Wyoming	0.0074926		0.0802529		
	2	-0.0652677	5	0.9999087	
#Wild/South_Dakota-Captive/Wyoming	-0.1212761	-0.2291971	-0.0133551	0.0185039	
#Wild/Wyoming-Captive/Wyoming	-0.1155565	-0.2036736	-0.0274393	0.0034893	
#Reintroduced/South_Dakota-Reintroduced/Arizona	0.0050889		0.0582255		
	5	-0.0480477	8	0.9999403	
#Wild/South_Dakota-Reintroduced/Arizona	-0.1236797	-0.2194732	-0.0278863	0.0042116	

#Wild/Wyoming-Reintroduced/Arizona	-0.1179601	-0.1907205	-0.0451998	0.0001664
#Wild/South_Dakota-Reintroduced/South_Dakota	-0.1287687	-0.2245621	-0.0329753	0.0025642
#Wild/Wyoming-Reintroduced/South_Dakota	-0.1230491	-0.1958094	-0.0502887	0.000081
			0.1136406	
#Wild/Wyoming-Wild/South_Dakota	0.0057196	-0.1022014		1 0.9999982

3. Inbreeding coefficient	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Population	6	2.715	0.4526	9.49	7.96e-07 ***
Residuals	47	2.241	0.0477		
HSD Tukey					
Population	diff	lwr	upr	p adj	
			0.5506879		
Captive/Virginia-Captive/Colorado	0.26394	-0.022808		9 0.0895844	
			0.2687076		
Captive/Wyoming-Captive/Colorado	-0.1239382	-0.516584		8 0.9575801	
Reintroduced/Arizona-Captive/Colorado			0.1606470		
	-0.1261009	-0.4128489		9 0.822517	
Reintroduced/South_Dakota-Captive/Colorado			0.1379098		
	-0.1488382	-0.4355862		1 0.6838917	
	0.5374168		1.0543591		
Wild/South_Dakota-Captive/Colorado		2 0.02047452		2 0.0367669	
	0.5009168		0.8935626		
Wild/Wyoming-Captive/Colorado		2 0.10827096		8 0.004851	
			0.0047676		
Captive/Wyoming-Captive/Virginia	-0.3878782	-0.780524		8 0.0548217	
Reintroduced/Arizona-Captive/Virginia					
	-0.3900409	-0.6767889	-0.1032929	0.0022038	
Reintroduced/South_Dakota-Captive/Virginia					
	-0.4127782	-0.6995262	-0.1260302	0.0010262	
	0.2734768		0.7904191		
Wild/South_Dakota-Captive/Virginia		2 -0.2434655		2 0.6646469	
	0.2369768		0.6296226		
Wild/Wyoming-Captive/Virginia		2 -0.155669		8 0.5168814	
Reintroduced/Arizona-Captive/Wyoming			0.3904831		
	-0.0021627	-0.3948086		3 1	
Reintroduced/South_Dakota-Captive/Wyoming			0.3677458		
	-0.0249	-0.4175459		6 0.9999948	
			1.2437429		
Wild/South_Dakota-Captive/Wyoming	0.661355	0.07896707		3 0.0167426	

Wild/Wyoming-Captive/Wyoming	0.624855	0.14933725	1.1003727	5	0.0034039
Reintroduced/South_Dakota-Reintroduced/Arizona	-0.0227373	-0.3094853	0.2640107	2	0.9999804
Wild/South_Dakota-Reintroduced/Arizona	0.6635177	3 0.14657543	1.1804600	3	0.0045158
Wild/Wyoming-Reintroduced/Arizona	0.6270177	3 0.23437187	1.0196635	9	0.0002132
Wild/South_Dakota-Reintroduced/South_Dakota	0.686255	0.1693127	1.2031973	0.0030008	
Wild/Wyoming-Reintroduced/South_Dakota	0.649755	0.25710914	1.0424008	6	0.0001179
Wild/Wyoming-Wild/South_Dakota	-0.0365	-0.6188879	0.5458879	3	0.9999951
4. Runs of Homozygosity	Df	chi-square d(H)	P-value		
	2	8.8787	0.01		
Dunn's Test					
Population	Z stat	P-value			
Reintroduced x Captive	0.916888	0.1796			
Wild x Reintroduced	-2.977907	0.0015			
Wild x Captive	-2.441798	0.0073			
5. Mutation Load					
5.1 Heterozygous LOF non-dervied	Df	Sum_Sq	Mean_Sq	F value	Pr(>F)
Population_Groups	2	179.2	89.59	7.882	0.00104**
Residuals	51	579.7	11.37		
HSD Tukey					
Population_Groups	diff	lwr	upr	p adj	
Reintroduced-Captive	2.765734	0.4082044	5.1232641	0.0178426	
Wild-Captive	-2.79E+00	-6.48E+00	8.91E-01	0.1699908	
Wild-Reintroduced	-5.56E+00	-9.31E+00	-1.81E+00	0.002166	
5.2 Homozygous LOF non-derived	Df	Sum_Sq	Mean_Sq	F value	Pr(>F)
Population_Groups	2	130.6	65.32	13.39	2.11e-05***
Residuals	51	248.7	4.88		

HSD Tukey					
Population_Groups	diff	lwr	upr	p adj	
Reintroduced-Captive	7.73E-01	-7.71E-01	2.32E+00	0.4539908	
Wild-Captive	5.17E+00	2.75E+00	7.58E+00	0.000012	
Wild-Reintroduced	4.39E+00	1.94E+00	6.85E+00	0.0002101	
5.3 Heterozygous LOF derived					
	Df	Sum_Sq	Mean_Sq	F value	Pr(>F)
Population_Groups	2	3.07	1.5348	3.156	0.051
Residuals	51	24.8	0.4863		
HSD Tukey					
Population_Groups	diff	lwr	upr	p adj	
Reintroduced-Captive	4.69E-01	-1.91E-02	9.56E-01	0.0620768	
Wild-Captive	-7.69E-02	-8.39E-01	6.85E-01	0.967849	
Wild-Reintroduced	-5.45E-01	-1.32E+00	2.30E-01	0.2156755	
5.4 Homozygous LOF derived					
	Df	Sum_Sq	Mean_Sq	F value	Pr(>F)
Population_Groups	2	0.28	0.1401	0.477	0.623
Residuals	51	14.98	0.2937		
HSD Tukey					
Population_Groups	diff	lwr	upr	p adj	
Reintroduced-Captive	3.50E-03	-3.75E-01	3.82E-01	0.9997266	
Wild-Captive	2.31E-01	-3.62E-01	8.23E-01	0.617648	
Wild-Reintroduced	2.27E-01	-3.75E-01	8.30E-01	0.6361928	
5.5 Heterozygous missense non-derived					
	Df	Sum_Sq	Mean_Sq	F value	Pr(>F)
Population_Groups	2	132119	66059	7.94	0.000996 ***
Residuals	51	424329	8320		
HSD Tukey					
Population_Groups	diff	lwr	upr	p adj	
Reintroduced-Captive	4.89E+01	-1.49E+01	1.13E+02	0.1637864	
Wild-Captive	-1.17E+02	-2.17E+02	-1.73E+01	0.017851	
Wild-Reintroduced	-1.66E+02	-2.67E+02	-6.45E+01	0.0006974	

5.6 Homozygous missense non-derived					
	Df	Sum_Sq	Mean_Sq	F value	Pr(>F)
Population_Groups	2	510031	255015	11.59	7.09e-05 ***
Residuals	51	1122157	22003		
HSD Tukey					
Population_Groups	diff	lwr	upr	p adj	
Reintroduced-Captive	3.19E+01	-7.18E+01	1.36E+02	0.7394334	
Wild-Captive	3.20E+02	1.58E+02	4.82E+02	0.000047	
Wild-Reintroduced	2.88E+02	1.23E+02	4.53E+02	0.0002924	
5.7 Heterozygous missense derived					
	Df	Sum_Sq	Mean_Sq	F value	Pr(>F)
Population_Groups	2	638.4	319.2	6.267	0.00369 **
Residuals	51	2597.6	50.9		
HSD Tukey					
Population_Groups	diff	lwr	upr	p adj	
Reintroduced-Captive	6.83E+00	1.84E+00	1.18E+01	0.0049064	
Wild-Captive	-8.08E-01	-8.61E+00	7.00E+00	0.966188	
Wild-Reintroduced	-7.64E+00	-1.56E+01	2.98E-01	0.0615396	
5.8 Homozygous missense derived					
	Df	Sum_Sq	Mean_Sq	F value	Pr(>F)
Population_Groups	2	3114	1557	9.854	0.000241 ***
Residuals	51	8058	158		
HSD Tukey					
Population_Groups	diff	lwr	upr	p adj	
Reintroduced-Captive	1.87E+00	-6.92E+00	1.07E+01	0.8645402	
Wild-Captive	2.49E+01	1.11E+01	3.86E+01	0.00018	
Wild-Reintroduced	2.30E+01	9.01E+00	3.70E+01	0.0006508	

All supplementary tables listed within this chapter are available at [https://github.com/EarlhamInst/RShaw-PhD-Thesis/blob/main/RSThesis_ChapterThree_Supplementary_Materials%20\(3\).xlsx](https://github.com/EarlhamInst/RShaw-PhD-Thesis/blob/main/RSThesis_ChapterThree_Supplementary_Materials%20(3).xlsx)

Supplementary Material:

S3.1 - Black-footed ferret sample location and population grouping

S3.2 - Ancestry group output identified by Admixture

S3.3 - Admixture cross validation error

S3.4 - GProfiler Results on Expanded Gene Families identified by CAFE

S3.5 - GProfiler Results on Contracted Gene Families identified by CAFE

S3.6 - ROH data per individual

S3.7 - SnpEff summary output

S3.8 - SIFT summary output and annotation

S3.9 - Chapter 3 statistic summary table

A.3 Chapter four - Identification of constrained regions in the endangered Black-footed ferret and implications of noncoding genetic variants

All supplementary tables listed within this chapter are available at [https://github.com/TGAC/RShaw-PhD-Thesis/blob/main/RSThesis_ChapterFour_Supplementary_Materials%20\(1\).xlsx](https://github.com/TGAC/RShaw-PhD-Thesis/blob/main/RSThesis_ChapterFour_Supplementary_Materials%20(1).xlsx)

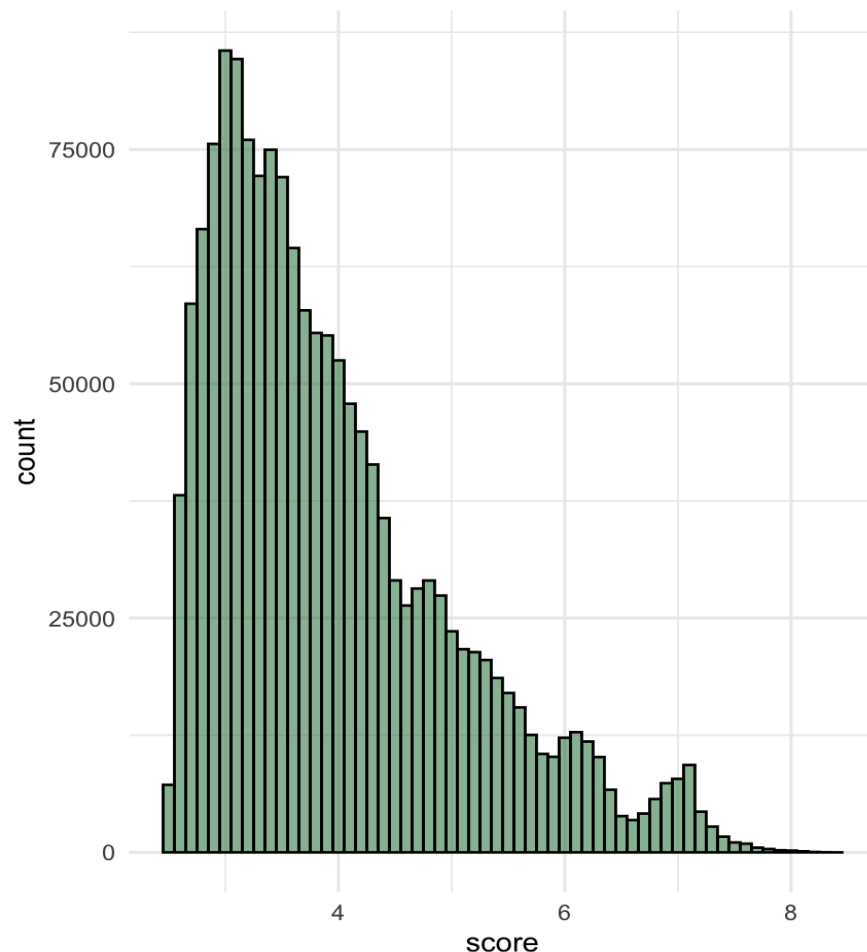


Figure S4.1: Frequency distribution of conserved PhyloP scores across the black-footed ferret genome. PhyloP scores were originally calculated across 25-bp windows, but adjacent windows were merged to increase window size to 50-bp and scores were averaged across this region. The results of this is visualised within this plot. The resulting observed pattern of scores across the black-footed ferret genome is less undulating than the one observed in Figure 4.5, and determines that this pattern may be down to a technical artefact due to calling conservation across short genome regions, as opposed to a biological artefact.

Supplementary Table:

S4.1 - Cactus Alignment Statistics

S4.2 - Top 5% Accelerated Gene GO Terms

S4.3 - Top 5% Conserved Gene GO Terms

S4.4 - Annotated CNE motif variants

S4.5 - Phenotype associated with gene CNE variants fall near