Island demographics and trait associations in white-tailed deer

Brooklyn S. Cars[1,2], Camille Kessler[1], Eric A. Hoffman[3], Steeve D. Côté[4], Daniel Koelsch[5,6], Aaron B.A Shafer[1,2]

[1]*Environmental and Life Sciences Graduate Program, Trent University, 2140 East Bank Drive, Peterborough, ON K9J 7B8, Canada*

[2]*Department of Forensics, Trent University, 2140 East Bank Drive, Peterborough, ON K9J 7B8, Canada*

[3]*Department of Biology, University of Central Florida, 4000 Central Florida Blvd, Orlando, FL, USA*

[4]*Département de Biologie and Centre d'Études Nordiques, Université Laval, G1V 0A6, Québec, QC, Canada*

[5]*Fédération des chasseurs de Saint-Pierre et Miquelon. Saint-Pierre et Miquelon, France*

[6]*Direction des Territoires de l'Alimentation et de la Mer, service Biodiversité. Saint-Pierre et Miquelon, France.*

Corresponding author: Aaron Shafer (aaronshafer@trentu.ca)

Word count: 4,577

Abstract

When a population is isolated and composed of few individuals, genetic drift is the paramount evolutionary force. Inbreeding might also occur, resulting in genomic regions that are identical by descent, manifesting as runs of homozygosity (ROHs) and the expression of recessive traits; accordingly, ROH can be used to map phenotypic traits. Likewise, the genes underlying traits of interest can be revealed by comparing fixed SNPs and divergent haplotypes between affected and unaffected individuals. Populations of white-tailed deer (*Odocoileus virginianus*) on islands of Saint Pierre and Miquelon (SPM, France) have high incidences of leucism and malocclusions; on the Florida Keys islands (USA), deer exhibit smaller body sizes. Here we aimed to reconstruct island demography and identify the genes associated with these traits using ROHs, fixed SNPs, and divergent haplotypes as markers in a pseudo case-control design. The two island populations of interest showed reduced levels of genomic diversity, a low *Ne*, and a build-up of genetic load with significant divergence in Key deer. Key deer showed higher inbreeding levels, but not longer ROHs, consistent with long-term isolation and a historical bottleneck. We identified multiple trait-related genes in ROHs including *LAMTOR2* which has links to hypopigmentation and HES1 which is involved in the shortening of the maxilla and mandible. Our mixed approach of linking ROH, fixed SNPs and haplotypes matched a high number (~50) of a-priori body size candidate genes in Key deer. All three phenotypes showed patterns consistent with a complex trait and non-mendelian inheritance.

Introduction

Genomic architecture is the organized arrangement of genes and other functional elements (Koonin, 2009). Targeted population sampling combined with high-throughput sequencing can characterize the architecture and effect size of genomic variation underlying traits (Ceballos *et al.*, 2018). Specifically, the genomes of individuals with and without traits of interest can be compared for the identification of divergent single nucleotide polymorphisms (SNPs) (Yang *et al.*, 2008; Russell *et al.*, 2012; Getmantseva *et al.*, 2020). Reconstructing haplotypes in also common for developing diagnostic tools and identifying candidate genes, with clear regions underlying body size variation in dogs detected for example (Parker *et al.*, 2009).

The distribution of SNPs present in the genome is also useful for the identification of regions with reduced heterozygosity and segments of autozygosity (Kim *et al.*, 2013); long rows of sequential homozygous sites known as runs of homozygosity (ROHs) can be used to map the genetic regions of recessive traits (Hildebrandt *et al.*, 2009). Runs of homozygosity result from the passing of identical haplotypes from one generation to the next with the length of ROH corresponding to recency of the common ancestor (McQuillan *et al.*, 2008; Kirin *et al.*, 2010). Expression of recessive deleterious traits require ROHs to be present (Robinson *et al.*, 2023), though purging might remove them (Wootton *et al.*, 2023). A populations' demographic history, notably recent trends in effective size (N_e), also can correspond to ROH dynamics (Kardos *et al.*, 2023) and estimates of ROH are now commonly included in assessments of endangered species, with for example unmanaged scimitar-horned oryx (*Oryx dammah*) populations manifesting significantly higher inbreeding (F_{ROH}) than those in captivity (Humble *et al.*, 2023). Therefore, detecting and mapping ROHs in a population can identify the genomic impact of demographic change and map genes that are associated with recessive and deleterious traits.

White-tailed deer (*Odocoileus virginianus*) are herbivorous, hooved ungulates that are native to the western hemisphere, primarily North America, but their range has expanded through human assistance to parts of Europe and various remote islands (Hewitt, 2011; Villanova *et al.*, 2017; Fuller *et al.*, 2019; Entfellner *et al.*, 2021). The ensuing island dynamics have resulted in various phenotypic variants or anomalies in some deer populations. For example, the native population of white-tailed deer that inhabit the Florida Keys - referred to as Key deer - are significantly smaller compared to the white-tailed deer populations on the mainland (Maffei *et al.*, 1988; Folk and Klimstra, 1991; Karns, 2010). These patterns are consistent with the Island rule (Foster, 1964; Benítez-López *et al.*, 2021) but also follow a general clinal decline in deer body sizes at southern latitudes (Wolverton *et al.*, 2009). The average adult mass of white-tailed deer ranges from 60 to 100 kg, while the Key deer typically weigh an average of 30 kg for females and 40 kg for males (Hewitt, 2011). Stature is generally a complex trait driven by many large and small effect genes (Kemper *et al.*, 2012; Sidorenko *et al.*, 2022) and Anderson et al. (2022) identified a suite of candidate genes impacting body size in deer (See supplemental file S1). Key deer are genetically differentiated (Villanova *et al.*, 2017), making disentangling drift from selection, and the Island rule, particularly challenging without first reconstructing their demographic history.

In contrast, the introduced population of white-tailed deer on the islands of Saint Pierre and Miquelon (SPM) have high rates of leucism and malocclusions; here, it is likely that the genes underlying these traits exhibit an increased frequency due to their presence in a small founding population. While white-tailed deer are normally a uniform reddish-brown-grey with white fur scattered on appendages, face, and rear, deer with leucism present a partial to entire lack of pigmentation. Coat colour composition in ungulates is mainly determined by loci that

control melanocyte activity and distribution (Fontanesi *et al.*, 2010). Mutations in the melanocortin receptor locus (MC1R) in particular have been found to be major contributors to leucism (Haase *et al.*, 2013; Reiner *et al.*, 2020). Brachygnathism and prognathism are both malocclusions involving the incorrect alignment of the maxillary bone and the mandible (Faragalla, 2000; Guan *et al.*, 2015). Brachygnathism has been seen in both mule deer (*O. hemionus*, Robinette and Aldous, 1955) and white-tailed deer (Smits and Bubenik, 1990). Candidate genes have been identified for both leucism (Supplementary Table S1) and jaw malocclusions (Supplementary Table S2); leucism is considered a recessive trait (Pereira *et al.*, 2023) but malocclusion inheritance patterns are less clear (Cakan *et al.*, 2012). Expression of both traits in deer likely comes with a fitness cost (Wolff *et al.*, 1993; Faragalla, 2000; Reissmann and Ludwig, 2013).

The emergence of divergent (Key deer) and costly (SPM) traits on islands present a pseudo case-control design allowing for characterizing the genetic architecture of phenotypes. By leveraging these phenotypically distinct populations and comparing to a large mainland white-tailed deer genomic data set (Kessler *et al.*, 2023), we aim to reconstruct each population's recent demographic history and genetic load, and test the prediction that ROH and candidate genes are associated with expression of recessive traits in SPM, while unique haplotypes with candidate body size genes underly the divergence of Key deer.

Materials and Methods

Sampling and Sequencing

Phenotype information for leucism or malocclusions from SPM was obtained from local managers and photos were examined to select individuals with one or more of the visible traits of interest. Twenty-six white-tailed deer tissue samples were then selected from the two island

populations of interest: SPM (France, but coastal Canada) and the Florida Keys (United States) in addition to four samples from Anticosti Island, Quebec (Canada). We leveraged available genome sequence data from four additional Key deer, one SPM sample (normal phenotype), and 21 deer from the North American mainland (normal phenotypes) that are part of a larger project (Fig. 1; PRJNA830519). DNA extraction from the tissue samples was carried out using the QIAGEN DNeasy Blood & Tissue Kit following the manufacturer's instructions (*DNeasy® Blood & Tissue Handbook*, 2020). We quantified the DNA concentrations using the Invitrogen Qubit assay. Samples were sent to The Centre for Applied Genomics in Toronto, Canada for low coverage whole genome sequencing (WGS) on an Illumina NovaSeq.

Data Preparation

The quality of the raw data libraries was examined using FastQC (v0.11.9; Andrews, 2010) and regions of poor quality were removed using Trimmomatic (v0.39; Bolger, Lohse and Usadel, 2014). The trimmed sequence data was then re-examined using FastQC, mapped to the annotated *O. virginianus* reference genome (GCA_023699985.1; Anderson *et al.*, 2022) using BWA-MEM (v0.7.17; Li, 2013), and sorted using SAMtools sort (v1.15.1; Li *et al.*, 2009). We detected duplicated reads using Picard MarkDuplicates (v2.23.2; Broad Institute, GitHub Repository, 2019) and removed the duplicated reads using Sambamba (v0.8.2; Tarasov *et al.*, 2015). Some samples had multiple sequence files and were merged before removing the duplicated reads using SAMtools merge. To ensure that no errors had been made throughout the process and that we had the desired coverage and depth to continue analysis we quantified metrics with SAMtools coverage and depth, and sambamba flagstat.

SNPs were called using ANGSD (v0.939; Korneliussen, Albrechtsen and Nielsen, 2014) following the GATK model (-gl 2) to estimate genotype likelihoods (GLs) in a beagle genotype

likelihood format (-doGlf 2). The genotypes were printed (-doGeno 4) and required SNPs with a minimum p-value of 1e−6 (-SNP_pval 1e-6), a minimum mapQ quality of 20 (-minMapQ 20), and a minimum base quality of 20 (-minQ 20). Further, we assumed that the major and minor alleles were known (-domaf 1) and that they can be inferred directly from likelihoods (-doMajorMinor 1).

Estimating Subpopulation Diversity and Demography

We estimated the nucleotide diversity (π) and Tajima's D, among each subpopulation of our dataset using VCFtools (v0.1.16; Danecek *et al.*, 2011). We calculated π using a 10 kb step size in between windows (-window-pi 10000), and Tajima's D in windows of 10 kb (-TajimaD 10000). We estimated fixation indices (F_{ST}) between populations using window sizes of 10 kb (-fst-window-size 10000). Using plink (v1.9) we generated a PCA based on the variance-standardized relationship matrix for the visualization of population stratification of the subpopulations. All plots were done in R version 4.2.2 (R Core Team, 2022). To estimate the historical N_e within the last 200 generations, we used the program GONE that uses the linkage disequilibrium information for markers across the genome (Santiago *et al.*, 2020). We selected a maximum of 30,000 SNPs per chromosome, 0.5 cM/Mb (Kessler *et al.*, 2023), an *hc* of 0.04, and applied the Haldane correction for genetic distance; we repeated the analysis 500 times.

Calling Runs of Homozygosity (ROHs)

Scaffolds making up the N90 were extracted using tabix (v.0.2.6.; Li, 2011). We created tped and tfam files using plink (v2.0) and identified ROH genotypes using the plink window-based approach (v1.9). We used a sliding window size of 100 kb (-homozyg-kb 100), a minimum density of 1 SNP per 50 kb (-homozyg-density 50) and required a minimum of 50 SNPs to be considered a ROH (-homozyg-snp 50). We used the default scanning window size of 50 SNPs (-

homozyg-window-snp 50), the default SNP hit rate of 0.05 (-homozyg-window-threshold 0.05) and allowed up to 5 heterozygote sites (-homozyg-window-het 5) and 5 missing calls (-homozyg-window-missing 5) for each 100 kb window within called ROHs. A 1,000 bp gap between variant sites (-homozyg-gap 1000) was required for two ROHs to be considered separate and we grouped the samples in phenotypic pools to identify overlapping, and possibly matching ROH segments (-homozyg-group). The percentage of the genome covered by ROHs (F_{ROH}) per individual was estimated by summing the total distance spanned by the ROH segments and dividing this by the total length of the N90 scaffolds.

We identified ROHs unique to each phenotypic group and required that the majority of individuals have the ROH while none of the SPM normal, mainland or Anticosti Island comparators presented this ROH. The genomic coordinates of the ROHs were identified and matched to the *O. virginianus* gene annotation file (Anderson *et al.*, 2022) and the biological function information of the discovered genes within ROH were identified through Uniprot (Pundir *et al.*, 2016) and our candidate gene literature review (Supplementary Tables S1, S2, file S1). Any gene without an entry ID was screened on NCBI blast to identify a putative function. We inferred the approximate age of the isolation by descent (IBD) event (no. of generations or g) based on ROH length, where $g = 100/(2rL)$, with r being the recombination rate of 0.50 cM/Mbp (Kessler *et al.*, 2023) length (L) of ROH in Mbp (Browning and Browning, 2015; Kardos *et al.*, 2017).

Fixed SNPs and haplotypes

We identified SNPs that were fixed with opposite alleles between the phenotypic group and wild-type samples (i.e. south mainland vs Key deer, SPM normal and mainland vs. leucism and malocclusions; Fig. 1). Using VCFtools (v0.1.16) we extracted lists of SNPs fixed in the

focal phenotypic group, requiring the alternate allele to be fixed in the mainland comparator. For SPM and the Key deer, we extracted and assayed all fixed SNPs. We complemented this with a formal haplotype analysis; here samples were first phased using SHAPEIT2 (Delaneau *et al.*, 2013) then analyzed through a likelihood ratio test HapLRT (Garrison *et al.*, 2022) to model the haplotype length distributions of the target (island) and background (comparator) populations; note the entire island sample of SPM was used to increase power. Here a positive sign from HapLRT reflect longer haplotypes in the target population; haplotype blocks were then defined as any region having a positive sign with successive p values <0.01 spanning 100 kb or longer and separated by at least 5 kb from other blocks. All fixed SNPs and haplotype blocks were matched to the *O. virginianus* gene annotation file (Anderson *et al.*, 2022) and we identified genes within haplotype blocks or within 50 kb of the fixed SNP grouping. Lastly, for all phenotypic and mainland groupings we used SNPeff (Cingolani *et al.*, 2012) to estimate and summarize the putative impact of all present SNPs; we focused on genetic load by measuring homozygote missense and loss of function (LOF) mutation, the latter being SNPs that resulted in a premature stop codon.

Results

Sampling, Sequencing and Data Preparation

All deer samples were sequenced to ~4x coverage; this included 16 individuals from SPM of which four presented with leucism and seven with malocclusions, as well as 10 Key deer samples (smaller stature size). As comparators, we used 21 samples from the North American mainland split into North and South groupings and four from Anticosti Island, all of which were normal presenting (Fig. 1). We identified a total of 100,544,312 SNPs in all individuals and the analyses were performed exclusively on the N90 scaffolds that harboured 90,993,235 SNPs.

Subpopulation Diversity

Tajima's D for SPM was the highest followed by Key deer that had the lowest value of π (Table 1). Mainland and Anticosti deer showed negative Tajima's D values. The fixation index (F_{ST}) and PCA showed moderate levels of genetic differentiation between the Key deer and every other subpopulation; SPM formed its own cluster, while Anticosti Island was embedded within the North Mainland samples (Supplementary Table S3; Supplementary Fig. S1). Recent demographic trends revealed a stable long-term low N_e of Key deer, with a large crash in SPM that predates the translocation (Fig. 2), which is also seen in mainland deer (Kessler *et al.*, 2023).

Distribution of ROH among phenotypes

The populations with the highest F_{ROH} were the Key deer and SPM, while the populations from the North American Mainland and Anticosti Island all had a relatively low F_{ROH} (Fig. 3A). The range of average ROH lengths within the Key deer individuals was shorter than the other populations with ROH lengths of 138.47 kb – 191.24 kb (Table 1); this range of ROH likely corresponds to 524 – 725 generations to the most recent common ancestor.

In the pools for all four SPM white-tailed deer with leucism, we found 46 shared ROHs. Of those, one was unique to the individuals with leucism, and it contained 18 genes; no a-priori candidate genes for leucism were detected (Supplementary Table S1). One ROH was detected in 3/4 animals with leucism and absent from the comparators; this region contained six genes and notably harboured the gene *LAMTOR2* (Endosomal adaptor protein p14) (Table 2). A total of 23 ROHs were found to be shared between the SPM white-tailed deer with malocclusions; none of these ROHs were unique to this phenotype. However, five out of seven of the white-tailed deer with malocclusions shared an ROH that contained eight genes, including *HES1* (Table 2). Lastly, 16 ROHs were shared among the smaller-stature Key deer, but none of these ROHs were unique

to this population. Within a pool of all 10 Key deer, with one mainland and one SPM deer, there were two genes that matched the Anderson et al. (2022) candidate gene table: *ART1* and *ART5* (Table 2). In addition, nine out of 10 of the Key deer shared four ROHs that were unique to them, they collectively harboured six genes (Table 2).

Fixed SNPs and haplotypes

For each fixed SNP we screened 50 kb flanking regions on either end to identify genes that were in close proximity; here we identified 72,291 fixed SNPs for Key deer population that collectively encompassed 9,081 genes. Of these, 12 overlapped with the ROHs and 427 matched the Anderson et al. (2022) candidate gene table. There were 159 fixed SNPs within the SPM leucism group that contained 65 genes, five fixed SNPs within the SPM malocclusion group that contained two genes and four fixed SNPs within the Anticosti population that contained one gene. There was no overlap for fixed SNPs and ROHs or candidate genes in either SPM phenotypic group.

There were 695 haplotype blocks detected in Key deer, which encompassed 455 genes. One haplotype overlapped with the ROHs for Key deer, 384 overlapped with the fixed SNPs for Key deer, and 50 genes matched the Anderson et al. (2022) candidate gene table, and notably included the gene IGF1R (Supplementary File S2). For the SPM population, we identified four haplotype blocks; these blocks harboured two genes, neither of which overlapped with the ROHs in SPM deer. The absolute distribution variants, including deleterious, were similar amongst populations (Table S4). Homozygote LOF and missense mutations were notably elevated in Key deer (Fig. 3 B/C); there was a modest increase in SPM, specifically relative to Anticosti and mainland deer (Fig. 3 B/C).

Discussion

Island populations are prone to drift; further, isolated populations composed of a small number of individuals are more likely to suffer from inbreeding, accumulate deleterious alleles, and manifest undesirable traits (Soule, 1973; Villanova *et al.*, 2017). SPM deer have a relatively high frequency of malocclusions and leucism, and Key deer are significantly smaller than their mainland counterparts (Folk and Klimstra, 1991). SPM samples clustered adjacent to northern mainland deer due to their recent translocation, while the Key deer are significantly diverged (see Supplementary Fig. S1). The Key deer became isolated approximately 8,000 years ago (Lazell, 1989) while SPM were introduced in 1953; consequently, the confounding effects of drift are much less of a factor in the gene-trait associations detected in SPM compared to the Key deer. Still, clear patterns emerged that not only impact our understanding of the genetic architecture of traits in wild populations, but also the build-up of ROH and genetic load in light of recent demographic change.

A tale of two island populations

Key deer are an endangered species under the US Federal Endangered Species Act. Genetic mark recapture studies suggests a population of ~1,000 individuals, with an N_e of only 11 (Villanova *et al.*, 2017). Our genome wide estimates are higher, with a contemporary point estimate of ~350. Here we suspect the cross-species amplification and small number of microsatellites resulted in a downward bias estimate in Villanova et al. (2017) compared to genome wide estimates based on linkage disequilibrium. Values of π, which are reflective of long-term or coalescent N_e (Peart et al. 2020) are lowest in Key deer compared to all comparators; the coalescent N_e / N_C ratio does appear to be linked to conservation status and extinction risk (Peart *et al.*, 2020; Dedato *et al.*, 2022; Wilder *et al.*, 2023); solving for $\pi = 4N_e\mu$ and with N_C estimates of Villanova et al. (2017) and μ of Kessler et al. (2023), would generate a

ratio coalescent N_e / N_C ratio >1, thereby supporting the endangered species designation.

Key deer were almost hunted to extinction in the 1950s (U.S. Fish and Wildlife Service, 1999); despite this, the N_e has been stable for >200 generations (Fig. 2) and no signal of overharvest is detected, as seen in mainland deer (Kessler *et al.*, 2023). Collectively, this suggests the moderate buildup of ROH (Fig. 3A), and genetic load (Fig. 3B/C) is due to the historical demographic bottleneck, and not to contemporary inbreeding or overharvest. Indeed, while Key deer have the highest mean F_{ROH} of white-tailed deer (Table 1), the average length of ROH is not longer, with the number of generations to the common ancestor of the ROH being >500. While there is an elevated genetic load, with absolute values higher than relevant comparators (Humble *et al.*, 2023), there does not appear to be a link to population viability given the temporal stability of N_e. This relatively high N_e might also have limited purging, as seen in ungulates with much lower N_e (e.g. mountain goats; Wootton et al. 2023; Martchenko & Shafer 2023), but also reiterates the importance of genome-wide metrics (Kardos et al. 2022). Collectively, these data support Villanova et al. (2017) who predicted long-term persistence of the Key deer population.

The SPM Island population originated from 12 white-tailed deer and the two inhabited islands are small (242 km²); in contrast 220 white-tailed deer were introduced to Anticosti Island (7,953 km²) off the coast of Quebec in 1896 (Fuller *et al.*, 2019). Accordingly, diversity values (Table 1) reflect this difference in founding number and island size, with Anticosti Island largely indistinguishable from the mainland. Tajima's D values were also highest on SPM, suggesting a recent loss of rare alleles due to this founding effect. In addition, the crash in temporal N_e (Fig. 2) predates the translocation event; as genomes carry a historical record of the species, this signal is a carry-over from the source population and colonial overharvest on the mainland (Kessler *et al.*,

2023). There does, however, appear to be a slight build-up of genetic load from the translocation event (Fig. 3B/C). As with Key deer, purging has yet to take place and temporal monitoring of SPM will shed light on this important evolutionary process given the recency of the founding event.

Inbreeding and isolation lead to detecting candidate genes

The Islands of SPM have a relatively high number of individuals with leucism and malocclusions. None of the candidate genes for leucism (Hauswirth *et al.*, 2012; Haase *et al.*, 2013; Reiner *et al.*, 2020) were found within our ROH segments. There were 18 new candidate genes harboured by the ROHs unique to all leucism individuals, though none had a clear biological link. Screening ROHs shared by 3/4 animals in the leucism pool identified the gene *LAMTOR2* (Endosomal adaptor protein p14) which has links to hypopigmentation (Bohn *et al.*, 2007). Hypopigmentation is caused by the hindrance of melanocytes which are responsible for the pigmentation of hair and skin (Dessinioti *et al.*, 2009). *LAMTOR2* is a target gene of the *MITF* transcription factor which regulates melanocyte activity and development; as leucism is known to be caused by mutations in genes that are associated with melanocytes, including *MITF*, *LAMTOR2* could be the leucism pathway in white-tailed deer. Collectively, these hint at non-mendelian patterns of inheritance as suggested in other vertebrates (Edelaar *et al.*, 2011; Jablonski *et al.*, 2017). No ROHs were found to be unique to the white-tailed deer with malocclusions; however, ROH segments shared by 5/7 white-tailed deer with malocclusions contained the gene *HES1*. This gene helps control the maintenance of undifferentiated cells and has been found to cause various craniofacial malformations in mice including a shortened maxilla and mandible (Akimoto *et al.*, 2010). While this gene certainly appears functionally relevant, there are other white-tailed deer who do not have this abnormality that have this ROH,

suggesting again a polygenic trait or incomplete penetrance (Stiles and Luke, 1953).

Stature size is a complex trait influenced by many potential genes that have both large and small effects. Anderson et al. (2022) sequenced the genomes of extreme phenotypes in deer, specifically the tail ends of the distribution for body size controlling for age and created an extensive list of candidate genes (supplemental file S1). Our study found close to 50 genes that overlapped previous genome-association work (Anderson et al. 2022), suggesting these are indeed likely linked to body size. Body size is clearly polygenic and it would be tempting to create a story of how these genes influence body size (Pavlidis *et al.*, 2012), so we simply note that one haplotype block contained the gene IGF1R which has been seen to cause a reduced size in dogs (*Canis familiaris*, Hoopes *et al.*, 2012).

Conclusion

Island white-tailed deer populations arose from a mix of human-assisted and natural colonization or isolation. Demographic assessment shows a stable, low *Ne* in both SPM and Key deer, with an accumulation of genetic load but generally low levels of ROH. Our results could be useful for wildlife management and conservationists who are concerned with small and isolated populations that may be experiencing a loss of genetic diversity due to inbreeding. Notably, combining Tajima's D with demographic models, and ROH with genetic load metrics, helps elucidate the role of demographic change in shaping population genomic health. Interestingly, the island populations do not have longer ROH; thus expression of putative recessive traits on SPM is not due to recent inbreeding, rather chance autozygosity due to a high frequency of the gene variant(s) in the founding population. Collectively, this also supports more detailed demographic models of Key suggesting they have been isolated for a considerable length of time (Kessler *et al.*, 2023), with this study showing no evidence of recent inbreeding, despite an

increased genetic load.

Our mixed approach of ROH, fixed SNPs and haplotypes did pull out a-priori candidate genes and discovered others. Storytelling of gene ontologies is common and leads to false positives (Pavlidis *et al.*, 2012), but our findings support all three traits not having a simple architecture, and identified promising genes involved with leucism (*LAMTOR2*), malocclusions (*HES1*), and stature size (*IGF1R*) in white-tailed deer. Despite the divergence in Key deer limiting some approaches (i.e. fixed SNPs were associated with almost half the total genes), both haplotype and ROH inferences overlapped with a large number of candidate genes allowing us to weed out background noise. Still, the assessment of autozygosity in independent samples and mouse models will be required to validate these genes.

Acknowledgements

Participation from hunters, managers and community members is essential to this research. We want to thank the community of St-Pierre and Miquelon that provided samples. Additional Florida Key samples were provided by George Zaragoza; the remaining providers are found listed in (Kessler *et al.*, 2023). BC was supported by a Natural Sciences and Engineering Research Council (NSERC) Undergraduate Student Research Award and the Trent Summer Work Experience Program. This work was supported by NSERC Discovery Grant (ABAS; SDC); ComputeCanada Resources for Research Groups (ABAS); Canadian Foundation for Innovation (ABAS): John R. Evans Leaders Fund (ABAS) and the Ontario Early Researcher Award (ABAS). Trent University is located on the traditional territory of the Michi Saagiig Anishnaabeg.

Author contribution statement

BC and ABAS conceived the study, BC, CK, EH, SDC, DK, ABAS collected the samples, BC and CK performed the molecular laboratory work and the bioinformatic analyses with contribution from ABAS. BC and ABAS wrote the initial draft of the manuscript, all authors reviewed and commented on the manuscript.

Conflict of Interest

Authors declare no competing interests.

Data archiving

Raw reads for the 25 deer additional individuals were deposited on the NCBI Sequence Read Archive (Accession number PRJNA830519). All scripts are available on GitLab (https://gitlab.com/WiDGeT_TrentU/undergrad-theses/-/tree/master/%20Cars_2023)

References

Akimoto M, Kameda Y, Arai Y, Miura M, Nishimaki T, Takeda A, *et al.* (2010). Hes1 is required for the development of craniofacial structures derived from ectomesenchymal neural crest cells. *J Craniofac Surg* **21**: 1443–1449.

Anderson SJ, Côté SD, Richard JH, Shafer ABA (2022). Genomic architecture of phenotypic extremes in a wild cervid. *BMC Genomics* **23**: 126.

Andrews S (2010). Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data.

Benítez-López A, Santini L, Gallego-Zamorano J, Milá B, Walkden P, Huijbregts MAJ, *et al.* (2021). The island rule explains consistent patterns of body size evolution in terrestrial vertebrates. *Nat Ecol Evol* **5**: 768–786.

Bohn G, Allroth A, Brandes G, Thiel J, Glocker E, Schaffer AA, *et al.* (2007). A novel human primary immunodeficiency syndrome caused by deficiency of the endosomal adaptor protein p14. *Nat Med* **13**: 38–46.

Bolger AM, Lohse M, Usadel B (2014). USADELLAB.org - Trimmomatic: A flexible read trimming tool for Illumina NGS data.

Broad Institute, GitHub Repository (2019). Picard Tools.

Browning SR, Browning BL (2015). Accurate Non-parametric Estimation of Recent Effective Population Size from Segments of Identity by Descent. *Am J Hum Genet* **97**: 404–418.

Cakan DG, Ulkur F, Taner T (Uğur) (2012). The genetic basis of facial skeletal characteristics and its relation with orthodontics. *Eur J Dent* **6**: 340–345.

Ceballos FC, Joshi PK, Clark DW, Ramsay M, Wilson JF (2018). Runs of homozygosity: windows into population history and trait architecture. *Nat Rev Genet* **19**: 220–235.

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, *et al.* (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly (Austin)* **6**: 80–92.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, *et al.* (2011). The variant call format and VCFtools. *Bioinformatics* **27**: 2156–2158.

Dedato MN, Robert C, Taillon J, Shafer ABA, Côté SD (2022). Demographic history and conservation genomics of caribou (Rangifer tarandus) in Québec. *Evol Appl* **15**: 2043–2053.

Delaneau O, Howie B, Cox AJ, Zagury J-F, Marchini J (2013). Haplotype Estimation Using Sequencing Reads. *Am J Hum Genet* **93**: 687–696.

Dessinioti C, Stratigos AJ, Rigopoulos D, Katsambas AD (2009). A review of genetic disorders of hypopigmentation: lessons learned from the biology of melanocytes. *Exp Dermatol* **18**: 741–749.

DNeasy® Blood & Tissue Handbook (2020). Qiagen.

Edelaar P, Donazar J, Soriano M, Santillan MA, Gonzalez Zevallos DR, Garcia Borboroglu JP, *et al.* (2011). Apparent selective advantage of leucism in a coastal population of Southern caracaras (Falconidae).

Entfellner J-BD, Muchadeyi FC, Mpolya EA, Chebii VJ (2021). Genomics of Adaptations in Ungulates. *Animals* **11**.

Faragalla (2000). Maxillary Brachygnathism in an Arab Mare. *J Vet Dent* **17**: 148–149.

Folk MJ, Klimstra WD (1991). Antlers of White-tailed Deer (Odocoieus virginianus) from Insular and Mainland Florida. **19**: 97–105.

Fontanesi L, Martelli PL, Beretti F, Riggio V, Dall'Olio S, Colombo M, *et al.* (2010). An initial comparative map of copy number variations in the goat (Capra hircus) genome. *BMC Genomics* **11**: 639.

Foster JB (1964). Evolution of Mammals on Islands. *Nature* **202**: 234–235.

Fuller J, Ferchaud A, Laporte M, Le Luyer J, Davis TB, Côté SD, *et al.* (2019). Absence of founder effect and evidence for adaptive divergence in a recently introduced insular population of white-tailed deer (Odocoileus virginianus). *Mol Ecol* **29**: 86–104.

Garrison E, Kronenberg ZN, Dawson ET, Pedersen BS, Prins P (2022). A spectrum of free software tools for processing the VCF variant call format: vcflib, bio-vcf, cyvcf2, hts-nim and slivar. *PLOS Comput Biol* **18**: e1009123.

Getmantseva LV, Bakoev SY, Shevtsova VS, Kolosov AY, Bakoev NF, Kolosova MA (2020). Assessing the Effect of SNPs on Litter Traits in Pigs. *Scientifica* **2020**: e5243689.

Guan X, Song Y, Ott J, Zhang Y, Li C, Xin T, *et al.* (2015). The ADAMTS1 Gene Is Associated with Familial Mandibular Prognathism. *J Dent Res* **94**: 1196–1201.

Haase B, Signer-Hasler H, Binns MM, Obexer-Ruff G, Hauswirth R, Bellone RR, *et al.* (2013). Accumulating Mutations in Series of Haplotypes at the KIT and MITF Loci Are Major Determinants of White Markings in Franches-Montagnes Horses. *Plos One* **8**: e75071-e75071.

Hauswirth R, Haase B, Blatter M, Brooks SA, Burger D, Drögemüller C, *et al.* (2012). Mutations in MITF and PAX3 Cause “Splashed White” and Other White Spotting Phenotypes in Horses. *PLOS Genet* **8**: e1002653.

Hewitt D (2011). *Biology and management of white-tailed deer*. CRC Press: Boca Raton, FL.

Hildebrandt F, Heeringa SF, Ruschendorf F, Attanasio M, Nurnberg G, Becker C, *et al.* (2009). A systematic approach to mapping recessive disease genes in individuals from outbred populations. *PLoS Genet* **5**.

Hoopes BC, Rimbault M, Liebers D, Ostrander EA, Sutter NB (2012). The insulin-like growth factor 1 receptor (IGF1R) contributes to reduced size in dogs. *Mamm Genome Off J Int Mamm Genome Soc* **23**: 780–790.

Humble E, Stoffel MA, Dicks K, Ball AD, Gooley RM, Chuven J, *et al.* (2023). Conservation management strategy impacts inbreeding and mutation load in scimitar-horned oryx. *Proc Natl Acad Sci* **120**: e2210756120.

Jablonski D, Grula D, Christophoryová J (2017). Unusual colour and pattern variation of Lacerta agilis (Squamata: Lacertidae) recorded from Central Europe. *Biharean Biol* **11**: 126–128.

Kardos M, Qvarnström A, Ellegren H (2017). Inferring Individual Inbreeding and Demographic History from Segments of Identity by Descent in Ficedula Flycatcher Genome Sequences. *Genetics* **205**: 1319–1334.

Kardos M, Zhang Y, Parsons KM, A Y, Kang H, Xu X, *et al.* (2023). Inbreeding depression explains killer whale population dynamics. *Nat Ecol Evol* **7**: 675–686.

Karns G (2010). Understanding White-Tailed Deer: Florida and the Southeast by William M. Giuliano; Elina Garrison; Brandon J. Schad. *J Wildl Manag* **74**: 1935–1936.

Kemper KE, Visscher PM, Goddard ME (2012). Genetic architecture of body size in mammals. *Genome Biol* **13**: 244.

Kessler C, Wootton E, Shafer ABA (2023). Speciation without gene-flow in hybridizing deer. *Mol Ecol* **32**: 1117–1132.

Kim E-S, Cole JB, Huson H, Wiggans GR, Van Tassell CP, Crooker BA, *et al.* (2013). Effect of Artificial Selection on Runs of Homozygosity in U.S. Holstein Cattle. *PLoS ONE* **8**: e80813–e80813.

Kirin M, McQuillan R, Franklin CS, Campbell H, McKeigue PM, Wilson JF (2010). Genomic runs of homozygosity record population history and consanguinity. *PloS One* **5**: e13996.

Koonin EV (2009). Evolution of genome architecture. *Int J Biochem Cell Biol* **41**: 298–306.

Korneliussen TS, Albrechtsen A, Nielsen R (2014). ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* **15**: 356.

Lazell JD (1989). Wildlife of the Florida Keys: a natural history.

Li H (2011). Tabix: fast retrieval of sequence features from generic TAB-delimited files. *Bioinformatics* **27**: 718–719.

Li H (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, *et al.* (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.

Maffei MD, Klimstra WD, Thomas J. W (1988). Cranial and Mandibular Characteristics of the Key Deer (Odocoileus virginianus clavium). *J Mammal* **69**: 403–407.

McQuillan R, Leutenegger A-L, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, *et al.* (2008). Runs of homozygosity in European populations. *Am J Hum Genet* **83**: 359–372.

Parker HG, VonHoldt BM, Quignon P, Margulies EH, Shao S, Mosher DS, *et al.* (2009). An Expressed Fgf4 Retrogene Is Associated with Breed-Defining Chondrodysplasia in Domestic Dogs. *Science* **325**: 995–998.

Pavlidis P, Jensen JD, Stephan W, Stamatakis A (2012). A Critical Assessment of Storytelling: Gene Ontology Categories and the Importance of Validating Genomic Scans. *Mol Biol Evol* **29**: 3237–3248.

Peart CR, Tusso S, Pophaly SD, Botero-Castro F, Wu C-C, Aurioles-Gamboa D, *et al.* (2020). Determinants of genetic variation across eco-evolutionary scales in pinnipeds. *Nat Ecol Evol* **4**: 1095–1104.

Pereira JA, Wolfenson LI, Artero DF, Argerich EC, Varela D, Fracassi NG (2023). First records of pigmentation anomalies in the marsh deer at its southernmost stronghold: a warning message? *Eur J Wildl Res* **69**: 59.

Pundir S, Martin MJ, O'Donovan C, Consortium TU (2016). UniProt Tools. *Curr Protoc Bioinforma* **53**: 1.29.1-1.29.15.

R Core Team (2022). R: The R Project for Statistical Computing.

Reiner G, Weber T, Nietfeld F, Fischer D, Wurmser C, Fries R, *et al.* (2020). A genome-wide scan study identifies a single nucleotide substitution in MC1R gene associated with white coat colour in fallow deer (Dama dama). *BMC Genet* **21**.

Reissmann M, Ludwig A (2013). Pleiotropic effects of coat colour-associated mutations in humans, mice and other mammals. *Semin Cell Dev Biol* **24**: 576–586.

Robinette W, Aldous C (1955). 'Parrot Mouth' in Rocky Mountain Mule Deer. *J Mammal* **36**: 308–310.

Robinson J, Kyriazis CC, Yuan SC, Lohmueller KE (2023). Deleterious Variation in Natural Populations and Implications for Conservation Genetics. *Annu Rev Anim Biosci* **11**: 93–114.

Russell CD, Widdison S, Leigh JA, Coffey TJ (2012). Identification of single nucleotide polymorphisms in the bovine Toll-like receptor 1 gene and association with health traits in cattle. *Vet Res* **43**: 17.

Santiago E, Novo I, Pardiñas AF, Saura M, Wang J, Caballero A (2020). Recent Demographic History Inferred by High-Resolution Analysis of Linkage Disequilibrium. *Mol Biol Evol* **37**: 3642–3653.

Sidorenko J, Arias JD, Graham SE, Highland HH, Lüll K, Bielak LF, *et al.* (2022). A saturated map of common genetic variants associated with human height. *Nat Lond* **610**: 704–712.

Smits B, Bubenik GA (1990). Congenital osteopetrosis in white-tailed deer (Odocoileus virginianus). *J Wildl Dis* **26**: 567–571.

Soule M (1973). The Epistasis Cycle: A Theory of Marginal Populations.

Stiles KA, Luke JE (1953). The inheritance of malocclusion due to mandibular prognathism. *J Hered* **44**: 241–245.

Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P (2015). Sambamba: fast processing of NGS alignment formats. *Bioinformatics* **31**: 2032–2034.

U.S. Fish and Wildlife Service (1999). South Florida multi-species recovery plan.

Villanova VL, Hughes PT, Hoffman EA (2017). Combining genetic structure and demographic analyses to estimate persistence in endangered Key deer (Odocoileus virginianus clavium). **18**: 1061–1076.

Wilder AP, Supple MA, Subramanian A, Mudide A, Swofford R, Serres-Armero A, *et al.* (2023). The contribution of historical processes to contemporary extinction risk in placental mammals. *Science* **380**: eabn5856.

Wolff G, Wienker TF, Sander H (1993). On the genetics of mandibular prognathism: analysis of large European noble families. *J Med Genet* **30**: 112–116.

Wolverton S, Huston MA, Kennedy JH, Cagle K, Cornelius JD (2009). Conformation to Bergmann's rule in white-tailed deer can be explained by food availability. *Am Midl Nat* **162**: 403–418.

Wootton E, Robert C, Taillon J, Côté S, Shafer ABA (2023). Genomic health is dependent on long-term population demographic history. : 2022.08.16.503900.

Yang Y, Li SS, Chien JW, Andriesen J, Zhao LP (2008). A systematic search for SNPs/haplotypes associated with disease phenotypes using a haplotype-based stepwise procedure. *BMC Genet* **9**: 90.

Figure legends

Fig. 1: Sampling locations with phenotype represented by colours: Normal (black), Leucism (yellow), Malocclusion (blue), and Small Stature (red).

Fig. 2: Effective population size changes for the last 400 years, x-axis as generations before present with a generation time of two years (SPM in blue and Key deer in red)

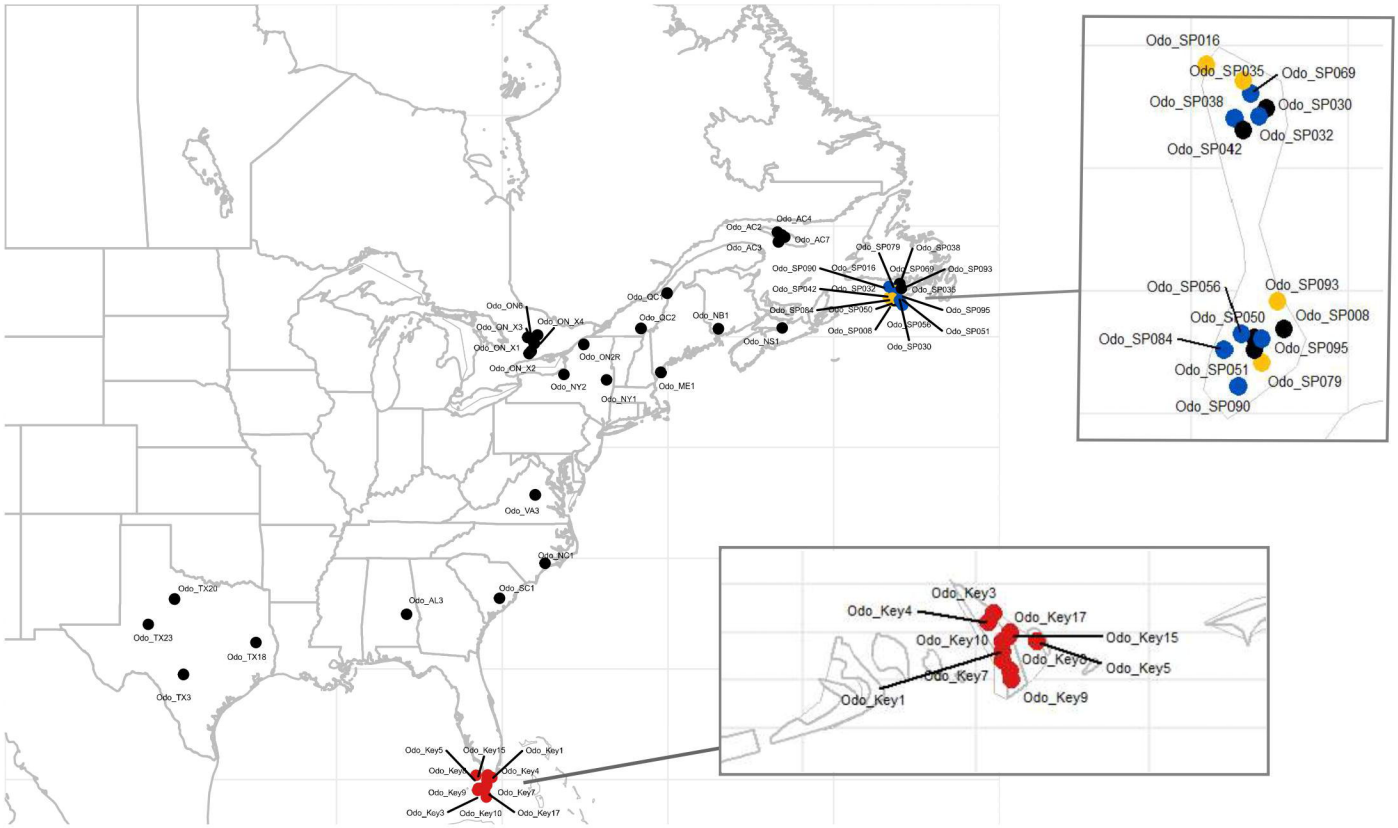
Fig. 3: A) Percentage of the genome covered by ROHs (F_{ROH}) for each population. B) Number of homozygous missense mutations based on SNPeff annotations. C) Number of homozygous loss of function mutations based on SNPeff annotations. Populations are colour coded by phenotype: Normal (black), Leucism (yellow), Malocclusion (blue), and Small Stature (red).

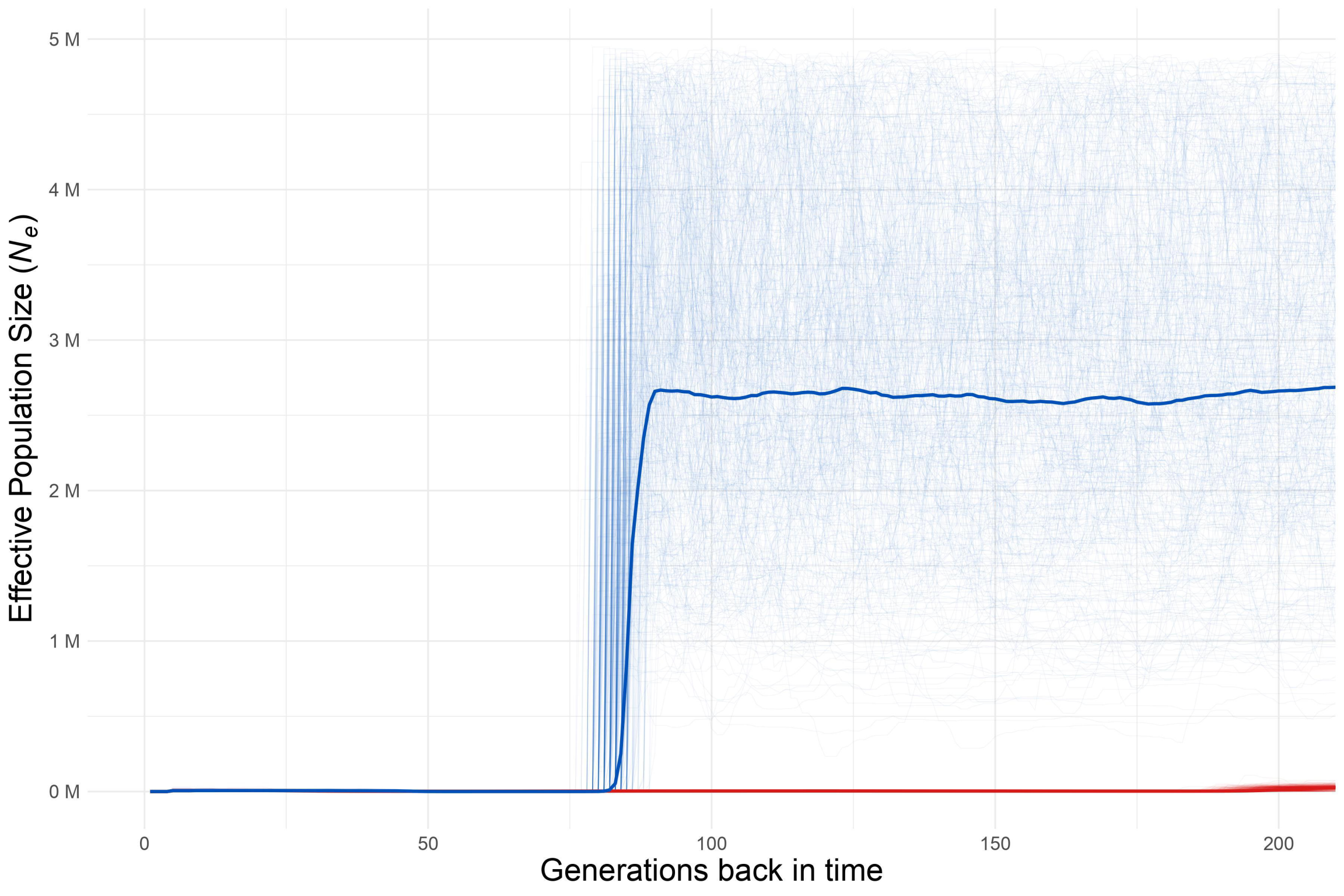
Table 1: Summary statistics of subpopulations analyzed.

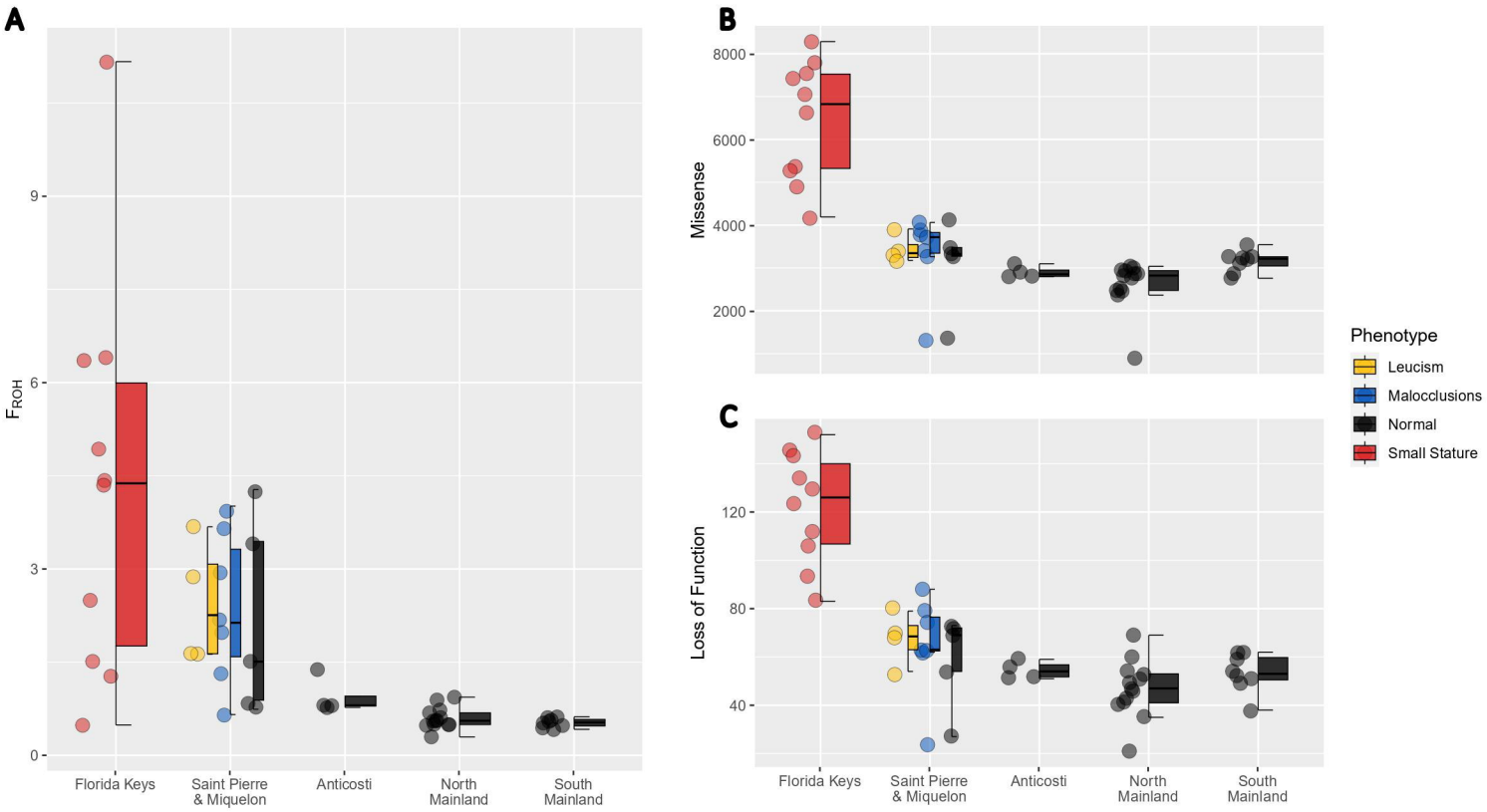
Subpopulation	Tajima's D	Nucleotide Diversity (π)	Mean F_{ROH}	Range of average ROH length (kb)
Anticosti	-0.09	0.007	0.01	156.73 - 170.84
SPM	0.54	0.005	0.02	154.26 - 199.05
North Mainland	-0.39	0.006	<0.01	155.81 - 215.74
South Mainland	-0.51	0.007	<0.01	164.45 - 186.57
Florida Keys	0.22	0.003	0.04	138.47 - 191.24

Table 2: List of ROHs shared by subsets of each group and the genes found within them (Matching candidate gene table = red, fixed SNPs = blue, haplotypes = orange, and haplotypes and fixed SNPs = green)

Group	Scaffold	Position	Genes Found
Leucism (4/4)	ref0002511	916,264 – 1,348,461	GPATCH4, NAXE, TTC24, MEF2D, IQGAP3, RBMS2, RHBG, TSACC, GLMP, CCT3, PAQR6, SMG5, SLC25A44, PMF1, TMEM79, BGLAP, MEX3A, SEMA4A
Leucism (3/4)	ref0002511	1,368,505 – 1,525,167	SSR2, UBQLN4, **LAMTOR2**, RAB25, ARHGEF2, RXFP4
Malocclusions (5/7)	ref0002179	1,508,045 – 1,749,819	TMEM44, LSG1, ATP13A3, **HES1**, OPA1, ATP13A4, ATP13A5, PLAAT1
Small Stature (10/10)	ref0000849	25096754 - 25528504	**LAMTOR1**, **TOMT**, **FOLR1**, **NUMA1**, **IL18BP**, **RNF121**, LOC110131318, LOC110131306, **ART1**, **ART5**, CHRNA10
Small Stature (9/10)	ref0001825	20,573,557 – 20,840,670	None
Small Stature (9/10)	ref0001994	104,322 – 495,467	None
Small Stature (9/10)	ref0002507	3,971,081 – 4,232,589	**TBC1D14**
Small Stature (9/10)	ref0002754	8,036,746 – 8,316,086	**MARCHF5**, **IDE**, **HHEX**, **KIF11**, **LOC110122451**


Odo_AC2
Odo_AC4
Odo_AC7
Odo_AC3
Odo_SP079
Odo_SP038
Odo_SP090
Odo_SP019
Odo_SP069
Odo_SP093
Odo_SP042
Odo_SP037
Odo_SP035
Odo_SP084
Odo_SP050
Odo_SP095
Odo_SP008
Odo_SP056
Odo_SP051
Odo_SP030
Odo_NS1
Odo_NB1
Odo_QC1
Odo_QC2
Odo_ON6
Odo_ON_X3
Odo_ON_X4
Odo_ON_X1
Odo_ON_X2
Odo_ON2R
Odo_NY2
Odo_NY1
Odo_ME1
Odo_VA3
Odo_NC1
Odo_SC1
Odo_AL3
Odo_TX20
Odo_TX23
Odo_TX18
Odo_TX3
Odo_Key5
Odo_Key15
Odo_Key1
Odo_Key6
Odo_Key4
Odo_Key9
Odo_Key7
Odo_Key3
Odo_Key10
Odo_Key17
Odo_SP016
Odo_SP035
Odo_SP069
Odo_SP038
Odo_SP030
Odo_SP032
Odo_SP042
Odo_SP056
Odo_SP093
Odo_SP050
Odo_SP008
Odo_SP084
Odo_SP095
Odo_SP051
Odo_SP079
Odo_SP090
Odo_Key3
Odo_Key4
Odo_Key17
Odo_Key10
Odo_Key15
Odo_Key6
Odo_Key5
Odo_Key7
Odo_Key1
Odo_Key9
Leucism
Malocclusions
Small Stature
Normal


5 M
4 M
3 M
2 M
1 M
0 M
Effective Population Size (N_e)
0
50
100
150
200
Generations back in time


A
B
C
F_{ROH}
Missense
Loss of Function
Florida Keys
Saint Pierre & Miquelon
Anticosti
North Mainland
South Mainland
Phenotype
Leucism
Malocclusions
Normal
Small Stature