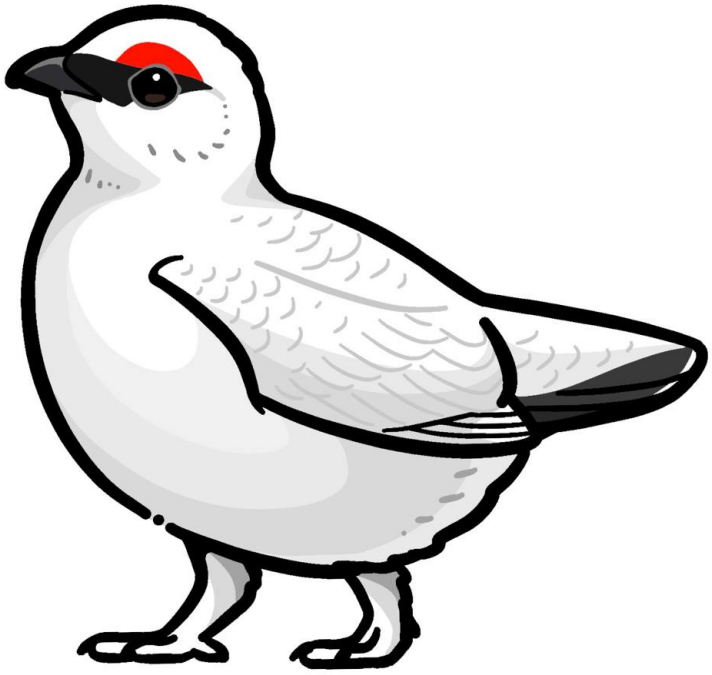
UPPSALA
UNIVERSITET

Genomic Analysis of Icelandic Rock Ptarmigan (*Lagopus muta*) Using Time Series Data



Wanyi Wei

Degree project in biology, Master of science (2 years), 2024
Examensarbete i biologi 60 hp till masterexamen, 2024
Biology Education Centre
Supervisors: Jacob Höglund and Theodore Squires
External opponent: Martyna Zwoinska

Abstract

To prevent the cyclic population of Icelandic rock ptarmigan (Lagopus muta) from collapsing due to a long-term decline, a hunting prohibition was implemented between 2003 and 2005. This not only decreased mortality in the rock ptarmigan but also unexpectedly changed its cyclic period from 10 years to 5 years. To assess the genomic features in response to the environmental changes in Icelandic rock ptarmigan, 99 first year birds from 2007 to 2018 were sequenced in this study. We examined changes in their inbreeding levels, genetic diversity, differentiation, and allele frequencies, finding a significant decrease in runs of homozygosity and observed temporal fluctuations in allele frequency patterns, which could be considered as an evidence of balancing selection. To further improve understanding about how balancing selection affects polymorphisms in Icelandic rock ptarmigan during population cycling, more genomic and ecological studies are required.

Table of Contents

List of Figures and Tables

Chapter 1 Introduction

1.1 Ecological Background of Rock Ptarmigan from NE Iceland

Rock Ptarmigan (*Lagopus muta*) are cold-adapted and sedentary grouse with a circumpolar distribution around the northern hemisphere, primarily inhabiting the arctic and alpine tundra regions of the Palearctic (Robert & Holder 2020). However, the arctic and alpine ecosystems are predicted to experience rapid changes caused by climate changes, marked by advanced timing of spring phenological events, elevated summer temperatures, uphill and poleward range contraction, etc. (Ernakovich *et al.* 2014). Compared to long migratory birds, year-round residents like ptarmigans are more vulnerable to these changes. Two ptarmigan species, willow ptarmigan (*Lagopus lagopus*) and rock ptarmigan have already been observed to experience the most severe declines among 14 studied species in northern Europe (Lehikoinen *et al.* 2014, Henden *et al.* 2017). Among the rock ptarmigan populations, the Icelandic population is also recommended to be managed as one single evolutionarily significant unit (ESU), *Lagopus mutus islandorum* (Faber, 1822), for future conservation (Holder *et al.* 2004).

In Iceland, the rock ptarmigan is the only of native vertebrate herbivore which show population cycling and no other sympatric cyclic herbivores occurr within the same range (Garðarsson 1988). Thus, the Icelandic populations are embedded in a relatively simple food web and play a pivotal role in the ecosystem as a keystone species, being the primary prey of the specialist predator gyrfalcon (*Falco rusticolus*) while also facing predation from generalist predators like the arctic fox (*Alopex lagopus*), raven (*Corvus corax*), gulls (*Larus* spp.), skuas (*Stercorarius* spp.), and the invasive American mink (*Neovison vison*) (Nielsen 1999, Nielsen 2011a). Although rock ptarmigan can be found throughout Iceland, the population on the northeast coast hosts the highest density, possibly attributed to less mammalian predators and higher chick production (Snæþórsson 2012). The gyrfalcon is the primary predator of rock ptarmigan in northeast Iceland, with the latter constituting 72% of the falcon's dietary intake (Nielsen 1999).

The census size of the Icelandic population has been documented for over six decades since this species plays a crucial role not only within the local ecosystem but also in the lives of local people. Before 2003 and 2004, the population cycled in a 10-12 year period with peak numbers dampened lower than data recorded in the early 20th century (Nielsen 2011b), and the density also experienced an average decline of 4% per year from 1981 to 2003 (Brynjarsdóttir *et al.* 2003). Due to the long-term decline, although rock ptarmigan is assessed as least concern (LC) by the International Union

for Conservation of Nature (IUCN), it is listed as near threatened (NT) in Iceland (BirdLife International 2016, Icelandic Institute of Natural History). Hence, commercial hunting for rock ptarmigan in Iceland was prohibited in 2003 and 2004 to prevent the cyclic population from collapsing and to ensure a sustainable harvest, although it has been allowed again since 2005 along with some restrictions (Icelandic Ministry of the Environment 2005, Nielsen 2011b). Following the hunting prohibition, a notable cyclic period shift from 10-12 years to 3-5 years has been documented (Figure 1.1), and the Icelandic population size is now cycling between less than 20,000 and over 100,000 (Icelandic Institute of Natural History). The possible driving forces of the cycles are excess juvenile autumn and winter loss and the negative impact of population size on chick production (Garðarsson 1988, Magnússon *et al.* 2004).

What restrictions?
I would mention
what they

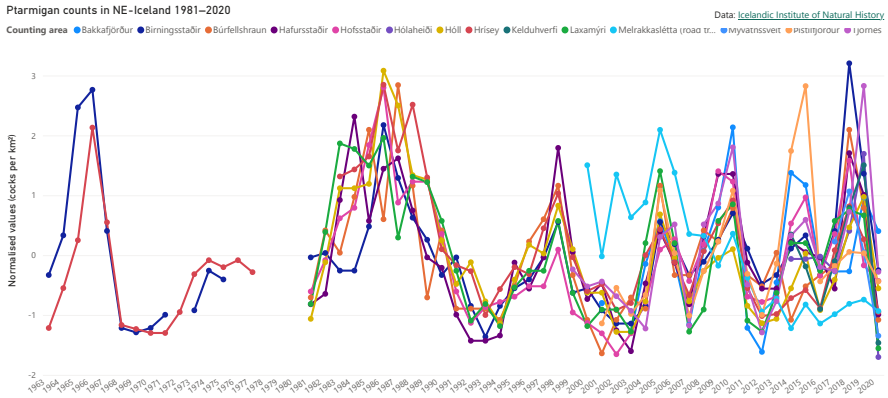


Figure 1.1 Rock Ptarmigan Population Counts in NE Iceland (1981-2020)
Adapted from Icelandic Institute of Natural History.

1.2 Preserving the Blueprint of Nature: The Role of Genetics in Conservation

Notwithstanding some researchers argue that extinction is a natural phenomenon, an inevitable outcome of evolution whereby certain species are selected out (Ceballos & Ehrlich 2018), one widely recognised notion in conservation is that we are currently experiencing the sixth massive extinction (Cowie *et al.* 2022). Indeed, extinction is a continual natural process occurring consistently throughout time. When a species cannot adapt to its surrounding environment, extinction happens to when the last member perishes. However, with the current extinction rate being estimated as 1000 times the possible background rate (Barnosky *et al.* 2011, Pimm *et al.* 2014), it is hard to believe the extinction events occurring now can be regarded as natural. Throughout the history of Earth, there have been five mass extinctions that happened in the past 540 million years where the extinction rate is greater than the speciation rate (Raup 1986, Barnosky *et al.* 2011), and it generally takes 5-10 million years for biodiversity to recover from these events (Jablonski 1994). Despite the potential for recovery, every loss of an evolutionary clade is an irreversible loss. Furthermore, since humans are also

part of the ecosystem, the inability to preserve our integrity in the face of such events is evident as well (Cardinale *et al.* 2012).

Biological diversity is usually considered on three levels: ecosystem, species, and genetic diversity (Primack 2006). Although it was not until the late 1970s to 1980s that scientists began to acknowledge the crucial importance of genetic diversity in both conservation efforts and understanding extinction dynamics (Allendorf *et al.* 2022), it is imperative to note that the role of genetic diversity in conservation is intricate and pivotal. While there are some cases where species experienced extremely severe bottlenecks and currently show no inbreeding depression, e.g. the Mauritius kestrels (*Falco punctatus*) (Groombridge *et al.* 2000), some species are showing low recovery potentials after their bottleneck events (Feng *et al.* 2019, Kardos *et al.* 2023). Debates persist regarding whether genetic diversity impacts a species before its extinction, wherein the species may perish before genetic factors come into play, or if reduced genetic diversity expedites the extinction process (Lande 1988, Spielman *et al.* 2004). Still, small populations are usually considered to be affected by inbreeding depression, loss of genetic diversity, mutation load, and involved in extinction vortex (Frankham 2005, Höglund 2009, Allendorf *et al.* 2022). Genetic variation serves as the raw materials of evolution, wherein genetic diversity manifests the inherent potential for species to adapt effectively to rapidly changing environments, particularly amidst the challenges posed by climate change. Therefore, it is critical to understand the mechanisms behind genetic diversity maintenance.

By now, two plausible driving forces have been identified as responsible for preserving genetic diversity: the balance among mutations, selection, and drift, and the alternative is balancing selection, with the former primarily removing conditionally deleterious alleles and the latter being recognised as the main reason behind preserved diversity (Hartl 2020, Futuyma & Kirkpatrick 2023). Balancing selection is mainly categorised into three types: Overdominance, negative frequency-dependent selection and fluctuating selection with reversal dominance (Hedrick 2007, Wittmann *et al.* 2017). Overdominance, also known as heterozygote advantage or heterosis, is where the heterozygote has higher fitness than both homozygotes, resulting in populations reaching a stable polymorphism equilibrium, a classic example of this would be the relationship between sickle cell anemia and malaria (Futuyma & Kirkpatrick 2023). The second one is negative frequency-dependent selection, in which the fitness of one allele changes opposite to its frequency in the population, often resulting from niche release attributed to some natural processes such as ecological competition, predation, sexual selection, etc. (Allendorf *et al.* 2022, Futuyma & Kirkpatrick 2023). The last one is fluctuating selection, where different alleles are favoured in alternating environments (Bell 2010).

1.3 Aims

Understanding how genetic diversity is sustained in nature and how organisms respond to their environment remains critical for conservation efforts. Therefore the Icelandic rock ptarmigan provides an ideal system to study these questions, particularly when decades of phenotypic data are available. The following questions will be addressed with the whole genome sequence data of the Icelandic rock ptarmigan:

1. Will inbreeding be exacerbated during troughs and broken up during peaks?
2. Whether genetic diversity is affected by population size fluctuations or determined by the long-term geometric mean population size?
3. Identify genomic regions and genetic variants linked to specific phenotypic traits to elucidate the genetic factors influencing population fluctuations.

Chapter 2 Materials and Methods

2.1 Sampling Methods

99 first-year birds were sampled in NE Iceland (Table 2.1, Figure 2.1) from two trough years (2007 and 2013) and two peak years (2010 and 2018) respectively. All specimens adhered to legal shooting permissions granted by Icelandic authorities.

Table 2.1 Year and category of specific sample size of rock ptarmigan sampled in NE Iceland (n = 99)

CollYear	2007	2010	2013	2018
Sample Size	27	23	26	23
Category	Trough	Peak	Trough	Peak

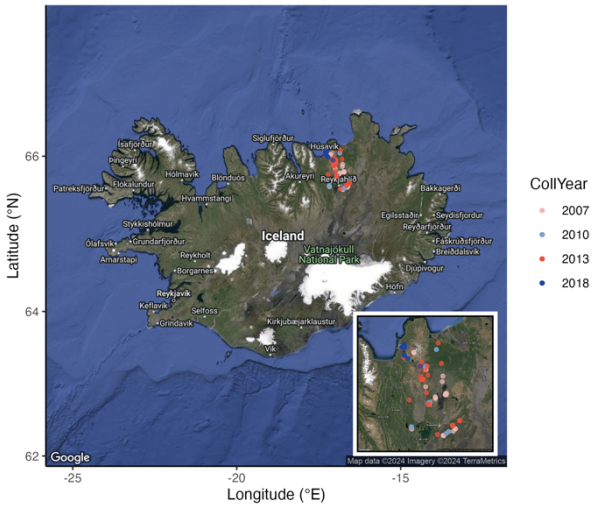


Figure 2.1 Map of the sampling area in NE Iceland of Rock Ptarmigan (2007, n = 27; 2010, n = 23; 2013, n = 26; 2018, n = 23).

The `.vcf` file that is used for the analysis in this thesis contains 7,259,761 SNPs and was generated following the methods desbribed in (Rödin-Mörch *et al.* 2023):

Initially, DNA extraction of muscle tissue was carried out employing the Monach™ Genomic DNA Purification Kit (New England BioLabs, USA). The extracted DNA underwent assessment for concentration and purity utilizing both a NanoDrop® 2000 spectrophotometer and Qubit® 3.0 fluorometer Quantitation Kit (Invitrogen™). Libraries were then generated using the TruSeq Nano DNA library preparation kit, followed by sequencing on an Illumina Novaseq 6000 S4 flowcell through the SNP&SEQ technology platform at the Royal Institute of Technology, Solna, Sweden.

The analysis was intiated by removing adapters from the raw reads using Trimmomatic v.0.36 (Bolger *et al.* 2014) and then conducted an assessment of read quality using FastQC v.0.11.9 (Andrews 2010). Subsequently, the trimmed reads were

later mapped onto the rock ptarmigan reference genome (GenBank accession: GCA_023343835.1) (Squires *et al.* 2023) using BWA-mem v.0.7.17 (Li & Durbin 2009). Evaluation of mapping quality and average read coverage was performed using Qualimap v.2.2.1 (García-Alcalde *et al.* 2012, Okonechnikov *et al.* 2016). We then proceeded to sort bam files and mark read duplicates using Picard v.2.20.4 (http://broadinstitute.github.io/picard/). Variant calling was conducted separately for each species utilizing GATK v.4.1.1.0 (McKenna *et al.* 2010) and following best practices (DePristo *et al.* 2011, Van der Auwera & O'Connor 2020). First, we employed Haplotypecaller with -ERC GVCF enabled on the duplicate marked and down-sampled bam files. The resultant gVCFs were combined for each individual and chromosome per species using CombineGVCFs, and then jointly genotyped using GenotypeGVCFs. Biallelic SNPs and INDELs were separately selected from the subsequent `.vcf` with genotyped variants using SelectVariants. We applied "hard filtering" standards to filter the resulting VCFs for both SNP and INDEL variants. For SNPs, this involved filtering out variants with Phred quality (QUAL) < 30.0, root mean square mapping quality (MQ) < 40.00, strand odds ratio (SOR) > 4.00, variant quality by depth (QD) < 2, fisher strand bias (FS) > 60.00, mapping quality rank sum test (MQRankSum) < -12.5, and read position rank sum test (ReadPosRankSym) < -8.00. For INDELs, filtering criteria included (QUAL) < 30.0, MQ < 40.00, SOR > 10.00, QD < 2.00, FS > 200.00, and ReadPosRankSum < -20.00. We then filtered the final VCF by read depth (10 < DP < 41), allowing a maximum of 10% missing data across samples using vcftools v.0.1.15 (Danecek et al. 2011), resulting in 7,259,761 SNPs.

2.2 Runs of Homozygosity (RoH), Inbreeding Coefficient and Heterozygosity

The RoH analysis is done by first converting the `.vcf` file into `.ped` and `.map` files using VCFtools v.0.1.16 (Danecek *et al.* 2011). RoH segments are later defined by applying the option `-homozyg` in PLINK v.1.90b4.9 (Chang *et al.* 2015) and setting the minimum RoH length at 100Kb, the maximum gap between consecutive SNPs considered part of the same RoH at 1000Kb, and the minimum SNP count for a RoH at 50. The proportion of the autosomal genome in runs of homozygosity (F_{RoH}) is estimated using $F_{RoH} = \frac{\sum L_{RoH}}{L_{auto}}$, in which $\sum L_{RoH}$ is the accumulative length of RoH in each individual, and L_{auto} represents the length of the autosomal genome covered by SNPs (McQuillan *et al.* 2008). The inbreeding coefficient and heterozygosity were computed using VCFtools v.0.1.16 (Danecek *et al.* 2011). Heterozygosity specifically was determined using the formula $\frac{N_{sites}-N_{hom\_obs}}{N_{sites}}$, where N_{sites} represents the total number of sites calculated and N_{hom_obs} denotes the number of observed homozygous

sites. For each comparison, a t-test was conducted using the `pairwise_t_test` function with the `paired = FALSE` option from rstatix (Kassambara 2023a). Results were plotted using ggpubr v.0.6.0 (Kassambara 2023b), ggplot2 v.3.4.2 and tidyverse v.2.0.0 (Wickham 2016, Wickham *et al.* 2019) in R v4.3.0 (R Core Team 2023) and RStudio v.2023.12.1+402 (RStudio Team 2020). The distribution of Runs of Homozygosity (RoH) was visualised using CMplot v.4.5.0 (Yin *et al.* 2021).

2.3 Genetic Diverisity (π) and Differentiation (F_{st} and d_{xy})

The autosomal nucleotide diversity (π), between population nucleotide divergence (d_{xy}), and genetic differentiation (F_{st}) was computated utilizing pixy v.1.2.5.beta1 (Korunes & Samuk 2021) which offers an unbiased estimation of nucleotide diversity by factoring invariant sites into its computation. The sliding window size was set to be 10k. To incorporate these invariant sites in our `.vcf` file, we re-executed GenotypeGVCFs in GATK v.4.1.1.0 (McKenna *et al.* 2010), specifying the option `--include-non-variant-sites`. The π and F_{st} values of each population and comparison were visualised using ggnetwork v.0.5.13 (Briatte 2024).

To standardize F_{st} values, Z-transform was conducted using the `scale()` function, setting `center = TRUE` and `scale = TRUE`, and ensuring negative values were converted to 0 beforehand. gene regions were identified overlapping with sliding windows using bedtools v2.31.0 (Quinlan & Hall 2010, Quinlan 2014) with the `"-intersect"` option. Windows with Top 10 values for each analysis and also have predicted functions available on NCBI were selected and labelled. Results of the first 28 chromosomes were plotted using ggrepel v.0.9.3 (Slowikowski 2024), ggplot2 v.3.4.2 and tidyverse v.2.0.0 (Wickham 2016, Wickham *et al.* 2019) in R v4.3.0 (R Core Team 2023) and RStudio v.2023.12.1+402 (RStudio Team 2020).

2.4 Allele Frequency Shifts Calculation

The allele frequency was computed for each group based on the collection year using VCFtools v.0.1.16 (Danecek *et al.* 2011). To monitor changes in allele frequency, we selected SNPs with minor allele frequency (MAF) in the year 2007, and also verified whether the SNP is located within gene regions that overlap with sliding windows. Results of windows with Top 10 ZF_{st} values were plotted using ggrepel v.0.9.3 (Slowikowski 2024), ggplot2 v.3.4.2 and tidyverse v.2.0.0 (Wickham 2016, Wickham *et al.* 2019) in R v4.3.0 (R Core Team 2023) and RStudio v.2023.12.1+402 (RStudio Team 2020).

Chapter 3 Results

3.1 Changes of Inbreeding Level Over Time

The number and length of ROHs can provide insights into individual demographic history. Long ROHs typically result from recent inbreeding events, whereas shorter ROHs possibly represent more distant inbreeding events, as they may result from ROHs fragmented by recombination (Ceballos FC *et al.* 2018). In our results, RoH were detected in 33 chromosome and 1 scaffold (Supplementary Figure 4.1). All individuals exhibited short RoH ranging from 100 Kb to 1 Mb. Additionally, long RoH (> 1 Mb) were identified in a combined total of 16 individuals, distributed as follows: 6 individuals in 2007, 6 in 2010, 1 in 2013, and 3 in 2018 (Figure 3.1). Regarding the length of ROHs, the majority of shorter ones span approximately 165 Kb, while long RoHs range from 1 Mb to 2.07 Mb (Supplementary Figure 4.2).

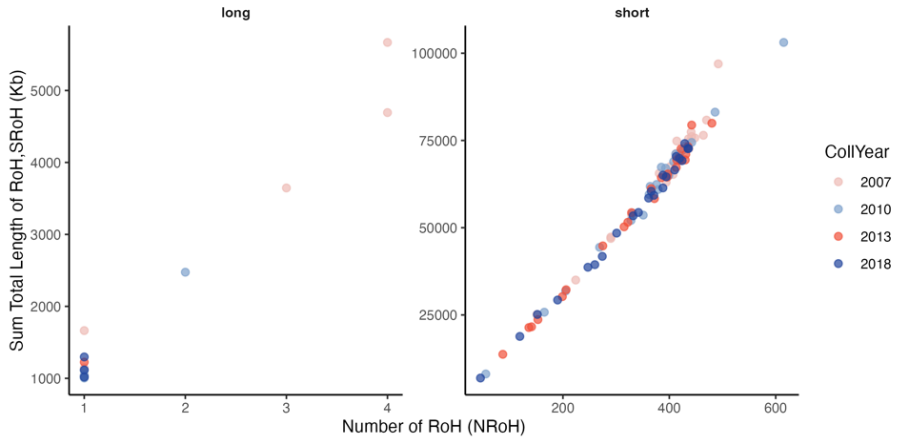


Figure 3.1 Sum total length of RoH (SRoH) and number of RoH (NRoH) of each Individual in different length categories

Categories: short: 100 Kb ~ 1 Mb; long: > 1Mb; Points are coloured according to collection years (2007, n = 27; 2010, n = 23; 2013, n = 26; 2018, n = 23).

The prevalence of ROH burden generally correlates with population size and admixture. Larger and more admixed populations usually exhibit fewer and shorter ROHs compared to smaller and more inbred populations (Ceballos *et al.* 2018). Within our populations, the average values of the number of ROHs (NRoH) and the sum total length of ROHs (SRoH) for both length categories experienced a decrease from 2007 to 2018 (Figure 3.2). As for the length of short ROHs, there is a significant difference between 2007 and 2013 ($p = 0.029$, t-test), and a highly significant difference between 2007 and 2018 ($p = 0.009$, t-test) (Supplementary Figure 4.3). Additionally, the difference in the number of short ROHs between 2007 and 2018 also exhibits significance ($p = 0.022$, t-test).

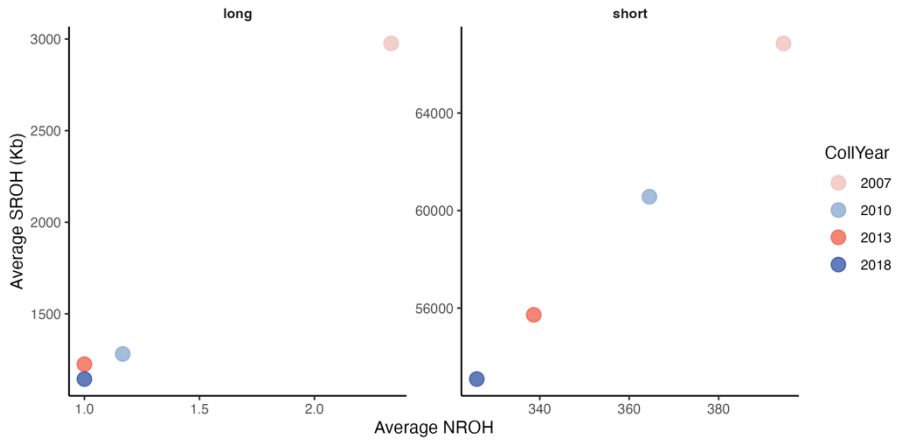


Figure 3.2 The mean of number of RoH (NRoH) and sum total length of RoH (SRoH) of each sampling year

Each point indicate a sampling year coloured by collection years (2007, n = 27; 2010, n = 23; 2013, n = 26; 2018, n = 23).

The genomic regions covered by SNPs was summed up to be 932,577,559 bytes. In the 2007 population, F_{RoH} ranged from 0.027 to 0.109, in 2010 from 0.009 to 0.111, in 2013 from 0.015 to 0.087, and in 2018 from 0.007 to 0.081. The mean value decreased from 0.072 in 2007 to 0.057 in 2018 (Figure 3.3). There was no evidence of any change in inbreeding coefficients and heterozygosity over the years (Supplementary Figure 4.4). These results suggest that from the year 2007 to the year 2018, the inbreeding level of the rock ptarmigan from NE Iceland consistently decreased instead of being exacerbated during trough years and broken up during peak years.

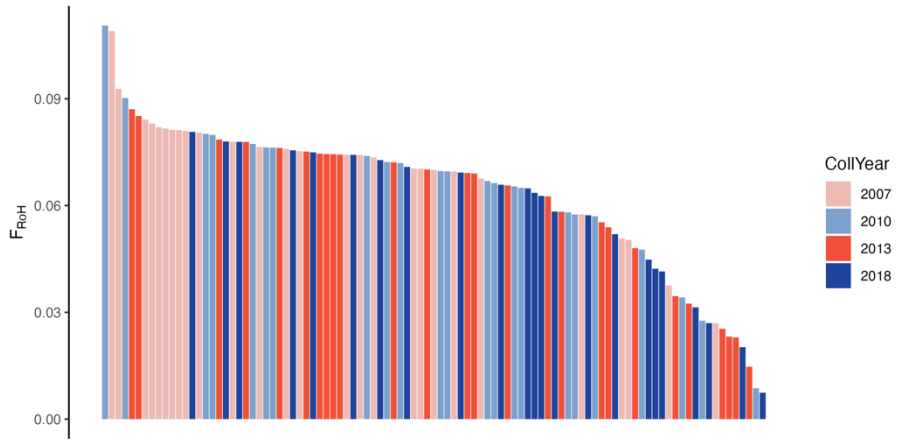


Figure 3.3 The proportion of autosomal genome in runs of homozygosity (F_{RoH})

Each bar represents an individual and are coloured according to collection years (2007, n = 27; 2010, n = 23; 2013, n = 26; 2018, n = 23).

3.2 Genetic Diversity and Differentiation Among Sampling Years

The genetic diversity within each sampling year was around 3.18×10^{-3}, and the genome-wide F_{st} values between each group was notably low as one would expect (Figure 3.4). Genetic diversity appears to be influenced by the long-term effective population size rather than being impacted by fluctuations in population size. These results are reasonable if one consider that the genetic diversity, π, is an estimator of the expected genetic diversity, θ, where $\theta = 4N_e\mu$, with μ representing the mutation rate. And the effective population size, N_e, of a constantly fluctuating population is calculated as the harmonic mean of its population size. Hence, the genetic diversity of a population undergoing constant fluctuations is anticipated to remain stable as we observed.

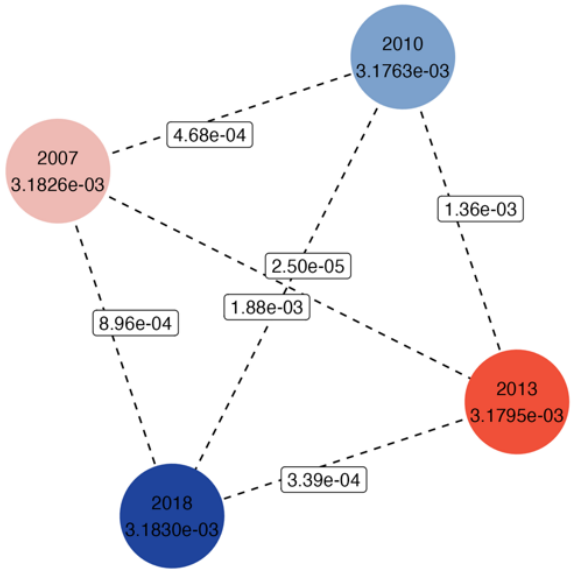


Figure 3.4 The genetic diversity (π) of each sampling year and the genetic differentiation (F_{st}) between different groups

The circles represent collection years and the sizes indicate the π value, coloured by collection years. The values on each line represent the F_{st} value between each two groups (2007, n = 27; 2010, n = 23; 2013, n = 26; 2018, n = 23).

Given the consistent trend in inbreeding levels within our populations from 2007 to 2008 and our aim to investigate genome changes during population fluctuations, for the rest of our analysis, we decided to concentrate on comparing trough years to peak years and also examining the differences between the earliest year, 2007, and the latest year, 2018.

The F_{st} values varied between -0.014 and 0.197 for comparisons between trough and peak years, and between -0.035 and 0.218 for the years 2007 and 2018. Although the range may look similar, among windows with the top 5% highest Fst values, there

were more windows of higher Fst values in the comparison between 2007 and 2018 than in the comparison between trough and peak years. This observation may imply that genomic regions undergo more pronounced changes over time rather than simply exhibiting great differences between trough and peak years. Considered that 2007 represented a trough year and 2018 a peak year, we further compared between trough years (2007 and 2013) and between peak years (2010 and 2018) (Supplementary Figure 4.5). The findings similarly indicated that more windows exhibited differences over time.

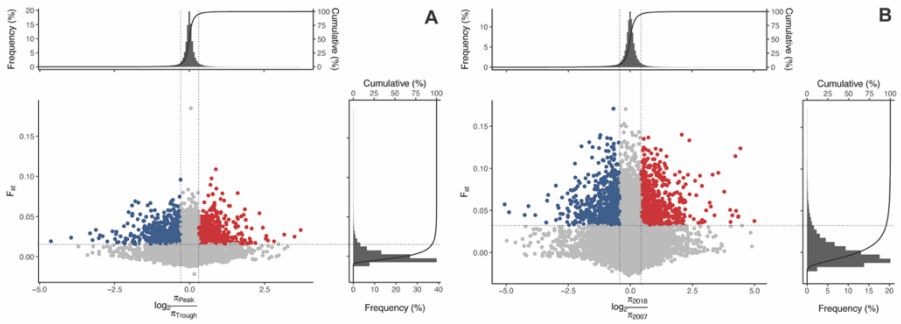


Figure 3.5 Selection analysis based on the π ratio and F_{st} between different groups

(A | The comparisons between trough years, i.e. 2007 and 2013, and peak years, i.e. 2010 and 2018; B | The comparisons between the earliest year, 2007, and the lattest year, 2018; 2007, n = 27; 2010, n = 23; 2013, n = 26; 2018, n = 23). The x axis represents π ratio (log2 transformed), the y axis indicates Fst value, and each point represents a window. The coloured regions are the top 5% regions selected.

3.3 Temporal Allele Frequency Shifts

Despite the current findings indicating greater differences over time rather than between trough and peak years in terms of inbreeding, genetic diversity, and differentiation, it is still worth investigating genomic regions that exhibit significant differences between trough and peak years. This is because the population density of the Icelandic rock ptarmigan undergoes significant cycling, suggesting variations in ecological conditions due to these population density changes. Thus, the genomic regions displaying disparities between trough and peak years could be considered as candidates influenced by balancing selection induced by population cycling (Supplementary Figure 4.6).

Based on the patterns of allele frequency shifts, a total of 7,259,761 SNPs are classified into three groups: directional (showing consistent changes), fluctuating (exhibiting fluctuations), and the remaining (rest). Our analysis indicate that 8.13% of the SNPs exhibited directional allele frequency changes from 2007 to 2018, 41.83% showed fluctuating patterns, while the rest accounted for 50.04% (Table 3.1, Figure 3.6).

Table 3.1 The categories and numbers of SNPs with different allele frequency shift patterns. (n = 7,259,761)

Category	07-10	10-13	13-18	Count
Directional	+	+	+	590,375
	-	-	-	
Fluctuating	+	-	+	3,036,685
	-	+	-	
Rest	+	+	-	3,632,701
	+	-	-	
	-	+	+	
	-	-	+	

Directional indicates the allele frequency changes of the SNPs are consistent; The SNPs with regular fluatuating patterns are in the Fluctuating group; The Rest group contains the ones without any of the patterns mentioned above.

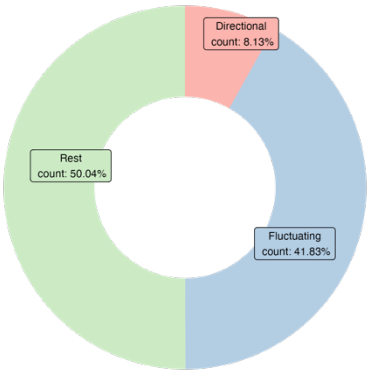


Figure 3.6 The proportion of the SNPs demonstrating different frequency shifts patterns (see Table 3.1).

Each part is coloured by the patterns. The SNPs that are with consistent allele frequency changes are categorised into Directional; those with regular fluctuating patterns are in Fluctuating group; The rest are contained in the Rest group.

Considering that if all of these SNPs were neutral, then one would expect 25% of the SNPs to display directional and fluctuating patterns, with the remaining 50%, our findings therefore suggest evidence of balancing selection occurring in Icelandic rock ptarmigans due to their population cycling. Subsequently, we examined the allele frequency shifts for windows exhibiting the highest ZF_{st} values in the comparisons (Supplementary Figure 4.2). Our analysis revealed that in the comparison between trough and peak years, the windows consistently exhibited fluctuating patterns, suggesting a correlation with our population density cycling (Figure 3.7). Additionally,

in the comparison between 2007 and 2018, the allele frequency of certain windows maintained a directional trend, seemingly unrelated to the population cycling.

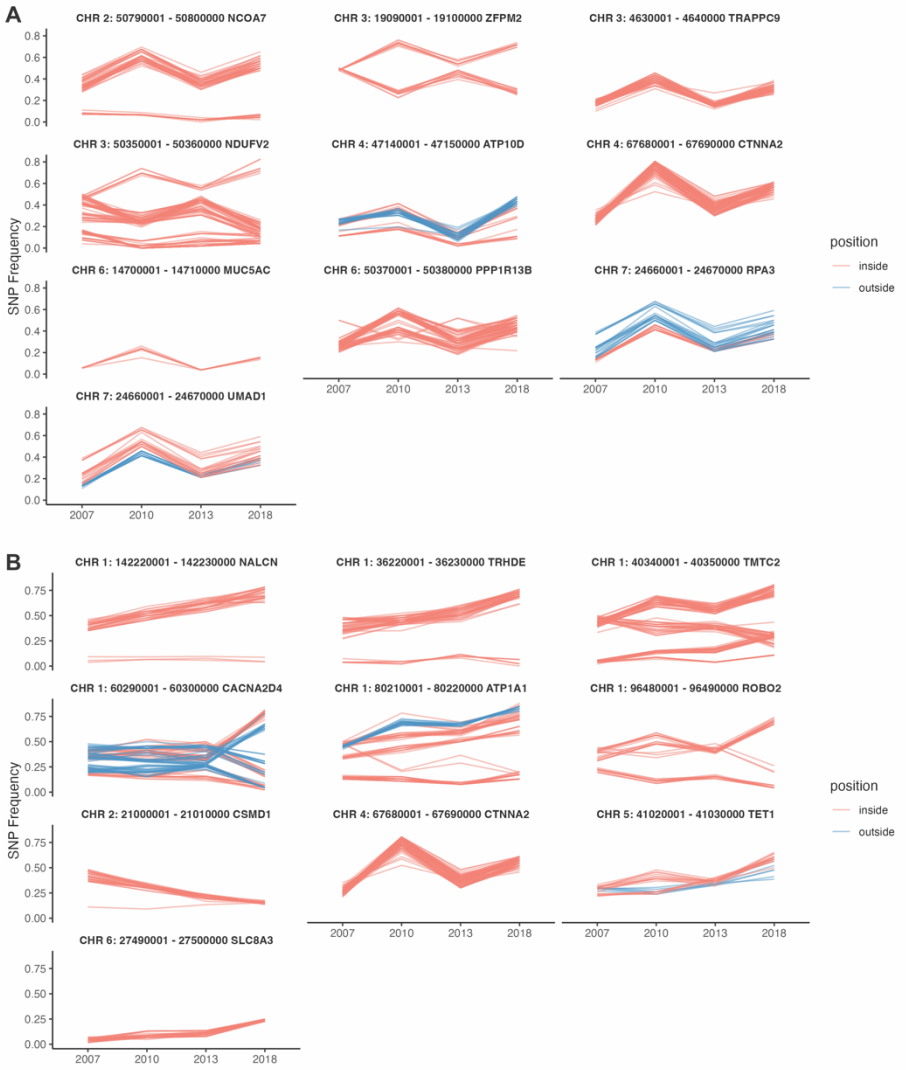


Figure 3.7 Temporal SNP frequency shifts of windows with Top 10 ZF_{st} values in different comparisons

(A | Comparisons between trough years, i.e. 2007 and 2013, and peak years, 2010 and 2018; B | Comparisons between the earliest year, 2007, and the lattest year, 2018). Each panel indicates a window. Each line represents a SNP and coloured by whether it is within a gene region. Protein coding genes found windows are indicated.

Chapter 4 Conclusions and Discussions

4.1 Discussions

Are there any other theories as to why there was a decrease?

The observed decrease in RoH could possibly result from reduced mortality due to the hunting prohibition. In the analysis of inbreeding level in Icelandic rock ptarmigan from 2007 to 2018, we observed a significant decrease in both the numbers and lengths of RoH, which became progressively more pronounced over time. However, this trend was not observed for heterozygosity and inbreeding coefficient. One potential explanation for this could be associated with the hunting prohibition implemented between 2003 and 2005, the harvest mortality of the Icelandic rock ptarmigan reduced (Sturludottir *et al.* 2018). The elevated survival rates could potentially contribute to a rise in recombination events, causing fragmentation of the RoHs. Nevertheless, even though these RoHs were fragmented into segments smaller than 100 Kb, making them undetectable by our scanning method, these segments remained homozygous. This may explain why, despite we observed a decrease in the mean values of heterozygosity and inbreeding coefficient, the differences did not reach significance.

Our findings regarding the genetic diversity and differentiation of Icelandic rock ptarmigan from 2007 to 2018 demonstrate that the majority of genetic windows displayed significant variation over time, rather than between trough and peak years. This pattern of high genetic turnover through time has also been documented in *Drosophila* (Bergland *et al.* 2014), a species that exhibits genomic features of seasonally fluctuating selection.

There are three major types of balancing selection, heterozygote advantage, negative frequency dependent selection, and alternating selection due to time and space. The Icelandic rock ptarmigans are facing varying external environments due to their population cycling, and this environmental heterogeneity is expected to lead to different selective pressures, i.e. alternating selection, on some loci. However, one cannot disregard the other possibilities for the oscillations we observed at polymorphisms, especially when our categorisation is qualitative but not quantitative.

The reasons are addressed below: 1) In the analysis, we did not set a criteria on flitering SNPs with recurrent and significant allele frequency changes, so it is possible that the patterns of some selected windows may be driven by drift. 2) Loci with heterozygote advantage may exhibit fluctuating frequency patterns when influenced by external conditions. The allele frequency of loci with heterozygote advantage theoretically tends to reach a polymorphic equilibrium, represented by $\hat{p} = \frac{1-w_{AA}}{2-w_{AA}-w_{aa}}$, with w is the fitness for the allele (Futuyma & Kirkpatrick 2023). Consequently, it is

plausible that these loci attain this equilibrium while being impacted by external conditions, leading to the observed patterns, particularly for those with cycling patterns of small magnitude. 3) The potential for negative frequency-dependent selection cannot be dismissed as well, as it is conceivable that the allele frequencies at some loci may be cycling independently, unrelated to the cycling of population size, and coincidentally match the sampling years we selected.

However, considering genetic functions and ecological studies, certain identified windows are putatively subject to temporal fluctuating selection, such as the gene *CTNNA2*, the one with the highest ZF_{st} value in the comparison of trough and peak years. In human, this gene encodes α-N-catenin and functions in the control of a startle response (Stelzer *et al.* 2016, Schaffer *et al.* 2018). Interestingly, a study on ptarmigan (*Lagopus* spp.) in southwestern Canada revealed a change in their flushing behavior, particularly noting a reduced flushing response during the year of increased population density (Pelletier & Krebs 1998), and this finding may help to explain the fluctuating allele frequency of *CTNNA2*.

From the long-term monitoring program of Icelandic rock ptarmigan, it has previously been found that, on average, individuals have more fat during trough years compared to peak years (T. Squires *et al.* unpublished data). Considering that rock ptarmigan is a cold-adapted and sedentary brid, fat plays a crucial role for winter survival, especially for the juveniles, and there is a thermoregulation-predation trade-off in wintering birds (Carr & Lima 2012). It is possible that the individuals that are less sensitive to predators, those that flush less, have higher fitness during peak years, but lower fitness during trough years. This can be further confirmed by documenting the flushing distance and the proportion of flushed birds during trough and peak years. The flushing distance is anticipated to be shorter during peak years because individuals avoid flushing to conserve energy for winter survival, despite the predator density being highest during these periods. Some studies have shown that grouse tend to have longer flushing distances when frequently disturbed by humans (Storch 2013), who can be considered potential predators.Since rock ptarmigan respond differently to aerial and ground predators, this assumption remains a possibility. Simarly, genes related to clutch size could also be under fluctuating selection since the clutch size of Icelandic rock ptarmigan varies little but significantly and inversely correlated to population density (Garðarsson 1988), and it has been showed that in great tits, the females with big clutch size would have reduced fitness when the population density increases (Sæther *et al.* 2016).

In future analysis, estimating the recent effective population size would be beneficial for evaluating the extent to which hunting pressure and hunting prohibition impact the Icelandic rock ptarmigan population, particularly in understanding how

genetic diversity responds following the prohibition. Furthermore, within the thesis, the categorisation based on allele frequency shifts was very preliminary since it was qualitative rather than quantitative. Hence, in order to understand the general molecular basis in response to fluctuating external environments for the Icelandic rock ptarmigan, it is necessary to investigate more about the patterns by establishing a threshold at magnitude to identify which SNPs exhibit recurrent and significant changes in frequency. Finally, the selective forces driving the observed patterns remain unclear, and one could test different hypotheses by conducting simulations, such as using SLiM, to provide further confirmation. In our analysis, we have identified temporal allele frequency shifts for some loci, which could be an evidence for alternating selection across time. However, in order to confirm or further investigate the role of balancing selection in the evolution of Icelandic rock ptarmigan genomes, one could seek for different genomic features potentially caused by balancing selection, e.g. excess of polymorphisms and intermediate-frequency alleles, or large number of low-frequency alleles (Bitarello *et al.* 2023).

Nevertheless, it is still challenging to detect the genomic footprints of balancing selection in our data. The reason for this is because balancing selection actually contains various forms of selective forces, and sometimes they are not mutually exclusive from each other. For instance, in the cause of fluctuating selection caused by time, the homozygotes for different alleles may have various fitness across time, but the heterozygotes would have higher geometric mean fitness in a long term. This situation is known as marginal overdominance, and it may also be considered as heterozygote advantage in a broad sense.

4.2 Conclusions

In the thesis, the level of inbreeding, genetic diversity, differentiation, and shifts in allele frequency have been examined in Icelandic rock ptarmigans across four distinct sampling years. First, the findings suggest a decline in both the lengths and numbers of Runs of Homozygosity (RoH), with no notable variance observed in heterozygosity and inbreeding coefficient. Furthermore, the genetic diversity of our fluctuating population remains stable rather than being influenced by the changes. Consequently, the genetic differentiation among our various sampling years was not significant. Finally, it is worth noting that although there were more windows of variations observed over time rather than between trough and peak years, those exhibiting significant differences between these periods displayed consistent fluctuating patterns over time. The patterns are seemingly correlated with our population cycling, but further analysis is required to infer the driving forces behind the patterns.

References

Allendorf FW, Funk WC, Aitken SN, Byrne M, Luikart G, Antunes A. 2022. Conservation and the Genomics of Populations, 3rd ed. doi 10.1093/oso/9780198856566.001.0001.

Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data.

Barnosky AD, Matzke N, Tomiya S, Wogan GOU, Swartz B, Quental TB, Marshall C, McGuire JL, Lindsey EL, Maguire KC, Mersey B, Ferrer EA. 2011. Has the Earth's sixth mass extinction already arrived? Nature 471: 51–57.

Bell G. 2010. Fluctuating selection: the perpetual renewal of adaptation in variable environments. Philosophical Transactions of the Royal Society B: Biological Sciences 365: 87–97.

Bergland AO, Behrman EL, O'Brien KR, Schmidt PS, Petrov DA. 2014. Genomic Evidence of Rapid and Stable Adaptive Oscillations over Seasonal Time Scales in Drosophila. PLOS Genetics 10: e1004775.

BirdLife International. 2016. Lagopus muta. The IUCN Red List of Threatened Species 2016: e.T22679464A113623562.

Bitarello BD, Brandt DY, Meyer D, Andrés AM. 2023. Inferring balancing selection from genome-scale data. Genome biology and evolution 15: evad032.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30: 2114–2120.

Briatte F. 2024. ggnetwork: Geometries to plot networks with 'ggplot2'.

Brynjarsdóttir J, Lund SH, Magnússon KG, Nielsen ÓK. 2003. Analysis of time series for rock ptarmigan and gyrfalcon populations in north-east Iceland. Raunvísindastofnun Háskólans

Cardinale BJ, Duffy JE, Gonzalez A, Hooper DU, Perrings C, Venail P, Narwani A, Mace GM, Tilman D, Wardle DA, Kinzig AP, Daily GC, Loreau M, Grace JB, Larigauderie A, Srivastava DS, Naeem S. 2012. Biodiversity loss and its impact on humanity. Nature 486: 59–67.

Carr JM, Lima SL. 2012. Heat-conserving postures hinder escape: a thermoregulation–predation trade-off in wintering birds. Behavioral Ecology 23: 434–441.

Ceballos FC, Joshi PK, Clark DW, Ramsay M, Wilson JF. 2018. Runs of homozygosity: windows into population history and trait architecture. Nature Reviews Genetics 19: 220–234.

Ceballos G, Ehrlich PR. 2018. The misunderstood sixth mass extinction. Science 360: 1080–1081.

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience 4: s13742-015-0047–8.

Cowie RH, Bouchet P, Fontaine B. 2022. The Sixth Mass Extinction: fact, fiction or speculation? Biological Reviews 97: 640–663.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, 1000 Genomes Project Analysis Group. 2011. The variant call format and VCFtools. Bioinformatics 27: 2156–2158.

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature Genetics 43: 491–498.

Ernakovich JG, Hopping KA, Berdanier AB, Simpson RT, Kachergis EJ, Steltzer H, Wallenstein MD. 2014. Predicted responses of arctic and alpine ecosystems to altered seasonality under climate change. Global Change Biology 20: 3256–3269.

Feng S, Fang Q, Barnett R, Li C, Han S, Kuhlwilm M, Zhou L, Pan H, Deng Y, Chen G, Gamauf A, Woog F, Prys-Jones R, Marques-Bonet T, Gilbert MTP, Zhang G. 2019. The Genomic Footprints of the Fall and Recovery of the Crested Ibis. Current Biology 29: 340-349.e7.

Frankham R. 2005. Genetics and extinction. Biological Conservation 126: 131–140.

Futuyma DJ, Kirkpatrick M. 2023. Evolution, Fifth edition. Oxford University Press, New York, NY.

García-Alcalde F, Okonechnikov K, Carbonell J, Cruz LM, Götz S, Tarazona S, Dopazo J, Meyer TF, Conesa A. 2012. Qualimap: evaluating next-generation sequencing alignment data. Bioinformatics 28: 2678–2679.

Garðarsson A. 1988. Cyclic Population Changes and Some Related Events in Rock Ptarmigan in Iceland. In: Gratson MW (ed.). Adaptive Strategies and Population Ecology of Northern Grouse, pp. 300–329. University of Minnesota Press, Fitzhenry & Whiteside Limited, the United States of America.

Groombridge JJ, Jones CG, Bruford MW, Nichols RA. 2000. ‘Ghost’ alleles of the Mauritius kestrel. Nature 403: 616–616.

Hartl DL. 2020. A primer of population genetics and genomics. A primer of population genetics and genomics

Hedrick PW. 2007. Balancing selection. Current Biology 17: R230–R231.

Henden J-A, Ims RA, Fuglei E, Pedersen ÅØ. 2017. Changed Arctic-alpine food web interactions under rapid climate warming: implication for ptarmigan research. Wildlife Biology 2017: wlb.00240.

Höglund J. 2009. Evolutionary conservation genetics. Oxford University Press, Oxford ; New York.

Holder K, Montgomerie R, Friesen VL. 2004. Genetic diversity and management of Nearctic Rock Ptarmigan (Lagopus mutus). Canadian Journal of Zoology 82: 564–575.

Icelandic Institute of Natural History. Rock ptarmigan | Icelandic Institute of Natural History. WWW document: https://www.ni.is/en/fauna/birds/breeding-birds/rock-ptarmigan. Accessed 17 April 2024.

Icelandic Ministry of the Environment. 2005. Amendment to Legal Regulation No. 456/1994 on hunting of birds and the exploitation of benefits from wild birds. Reglugerdir 800/2005. WWW document 2005: https://www.reglugerd.is/reglugerdir/allar/nr/800-2005. Accessed 16 April 2024.

Jablonski D. 1994. Extinctions in the fossil record. Philosophical Transactions of the Royal Society of London Series B: Biological Sciences 344: 11–17.

Kardos M, Zhang Y, Parsons KM, A Y, Kang H, Xu X, Liu X, Matkin CO, Zhang P, Ward EJ, Hanson MB, Emmons C, Ford MJ, Fan G, Li S. 2023. Inbreeding depression explains killer whale population dynamics. Nature Ecology & Evolution 7: 675–686.

Kassambara A. 2023a. rstatix: Pipe-friendly framework for basic statistical tests.

Kassambara A. 2023b. ggpubr: 'ggplot2' based publication ready plots.

Korunes KL, Samuk K. 2021. pixy: Unbiased estimation of nucleotide diversity and divergence in the presence of missing data. Molecular Ecology Resources 21: 1359–1368.

Lande R. 1988. Genetics and Demography in Biological Conservation. Science 241: 1455–1460.

Lehikoinen A, Green M, Husby M, Kålås JA, Lindström Å. 2014. Common montane birds are declining in northern Europe. Journal of Avian Biology 45: 3–14.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25: 1754–1760.

Magnússon KG, Brynjarsdóttir J, Nielsen ÓK. 2004. Population cycles in rock ptarmigan Lagopus muta: modelling and parameter estimation.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, others. 2010. The Genome Analysis

Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome research 20: 1297–1303.

McQuillan R, Leutenegger A-L, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, Smolej-Narancic N, Janicijevic B, Polasek O, Tenesa A, MacLeod AK, Farrington SM, Rudan P, Hayward C, Vitart V, Rudan I, Wild SH, Dunlop MG, Wright AF, Campbell H, Wilson JF. 2008. Runs of Homozygosity in European Populations. American Journal of Human Genetics 83: 359–372.

Nielsen ÓK. 1999. Gyrfalcon predation on ptarmigan: numerical and functional responses. Journal of Animal Ecology 68: 1034–1050.

Nielsen ÓK. 2011a. Gyrfalcon Population and Reproduction in Relation to Rock Ptarmigan Numbers in Iceland. Gyrfalcons and Ptarmigan in a Changing World, doi 10.4080/gpcw.2011.0210.

Nielsen ÓK. 2011b. Harvest and Population Change of Rock Ptarmigan in Iceland. Gyrfalcons and Ptarmigan in a Changing World, doi 10.4080/gpcw.2011.0212.

Okonechnikov K, Conesa A, García-Alcalde F. 2016. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics 32: 292–294.

Pelletier L, Krebs CJ. 1998. Evaluation of aerial surveys of ptarmigan *Lagopus* species. Journal of Applied Ecology 35: 941–947.

Pimm SL, Jenkins CN, Abell R, Brooks TM, Gittleman JL, Joppa LN, Raven PH, Roberts CM, Sexton JO. 2014. The biodiversity of species and their rates of extinction, distribution, and protection. Science 344: 1246752.

Primack RB. 2006. Essentials of conservation biology. Sinauer Associates Sunderland

Quinlan AR. 2014. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. Current Protocols in Bioinformatics 47: 11.12.1-11.12.34.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26: 841–842.

R Core Team. 2023. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Raup DM. 1986. Biological Extinction in Earth History. Science 231: 1528–1533.

Robert M, Holder K. 2020. Rock Ptarmigan (Lagopus muta), version 1.0. In Birds of the World. WWW document 2020: https://doi.org/10.2173/bow.rocpta1.01. Accessed 16 June 2024.

Rödin-Mörch P, Squires T, Magnússon KP, Höglund J. 2023. Genomic vulnerability to climate change and mutation load are affected by past declines in effective population size in two sedentary arctic bird species. 2023.01.30.526273.

RStudio Team. 2020. RStudio: Integrated development environment for R. RStudio, PBC., Boston, MA.

Sæther B-E, Visser ME, Grøtan V, Engen S. 2016. Evidence for r- and K-selection in a wild bird population: a reciprocal link between ecology and evolution. Proceedings of the Royal Society B: Biological Sciences 283: 20152411.

Schaffer AE, Breuss MW, Caglayan AO, Al-Sanaa N, Al-Abdulwahed HY, Kaymakcalan H, Yılmaz C, Zaki MS, Rosti RO, Copeland B, others. 2018. Biallelic loss of human CTNNA2, encoding αN-catenin, leads to ARP2/3 complex overactivity and disordered cortical neuronal migration. Nature genetics 50: 1093–1101.

Slowikowski K. 2024. ggrepel: Automatically position non-overlapping text labels with 'ggplot2'.

Snæþórsson AÖ. 2012. Reproductive success and survival of hen rock ptarmigan (Lagopus muta) during summer.

Spielman D, Brook BW, Frankham R. 2004. Most species are not driven to extinction before genetic factors impact them. Proceedings of the National Academy of Sciences 101: 15261–15264.

Squires TE, Rödin-Mörch P, Formenti G, Tracey A, Abueg L, Brajuka N, Jarvis E, Halapi EC, Melsted P, Höglund J, Magnússon KP. 2023. A chromosome-level genome assembly for the Rock Ptarmigan (Lagopus muta). G3 Genes|Genomes|Genetics 13: jkad099.

Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, Stein TI, Nudel R, Lieder I, Mazor Y, others. 2016. The GeneCards suite: from gene data mining to disease genome sequence analyses. Current protocols in bioinformatics 54: 1–30.

Storch I. 2013. Human disturbance of grouse - why and when? Wildlife Biology 19: 390–403.

Sturludottir E, Nielsen OK, Stefansson G. 2018. Evaluation of Ptarmigan Management with a Population Reconstruction Model. The Journal of Wildlife Management 82: 958–965.

Van der Auwera GA, O'Connor BD. 2020. Genomics in the cloud: using docker, GATK, and WDL in terra. O'Reilly Media

Wickham H. 2016. ggplot2: Elegant graphics for data analysis. Springer-Verlag New York

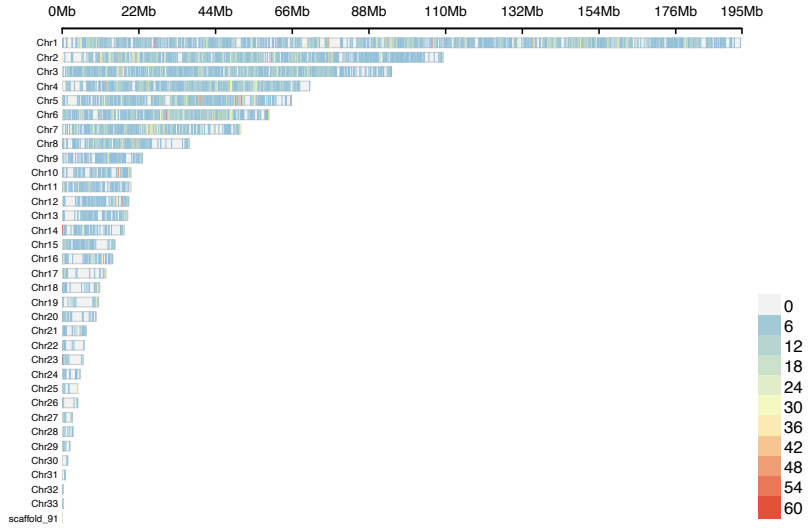
Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache

SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H. 2019. Welcome to the tidyverse. Journal of Open Source Software 4: 1686.

Wittmann MJ, Bergland AO, Feldman MW, Schmidt PS, Petrov DA. 2017. Seasonally fluctuating selection can maintain polymorphism at many loci via segregation lift. Proceedings of the National Academy of Sciences, doi 10.1073/pnas.1702994114.

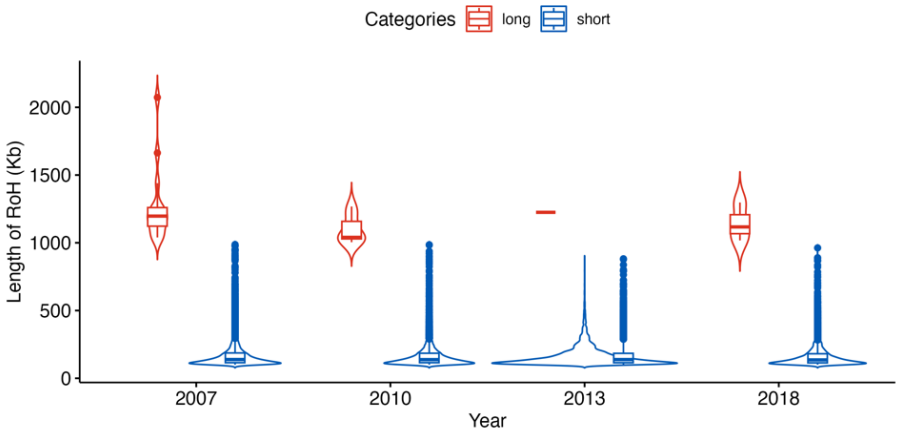
Yin L, Zhang H, Tang Z, Xu J, Yin D, Zhang Z, Yuan X, Zhu M, Zhao S, Li X, Liu X. 2021. rMVP: A Memory-efficient, Visualization-enhanced, and Parallel-accelerated Tool for Genome-wide Association Study. Genomics, Proteomics & Bioinformatics 19: 619–628.

Appendix



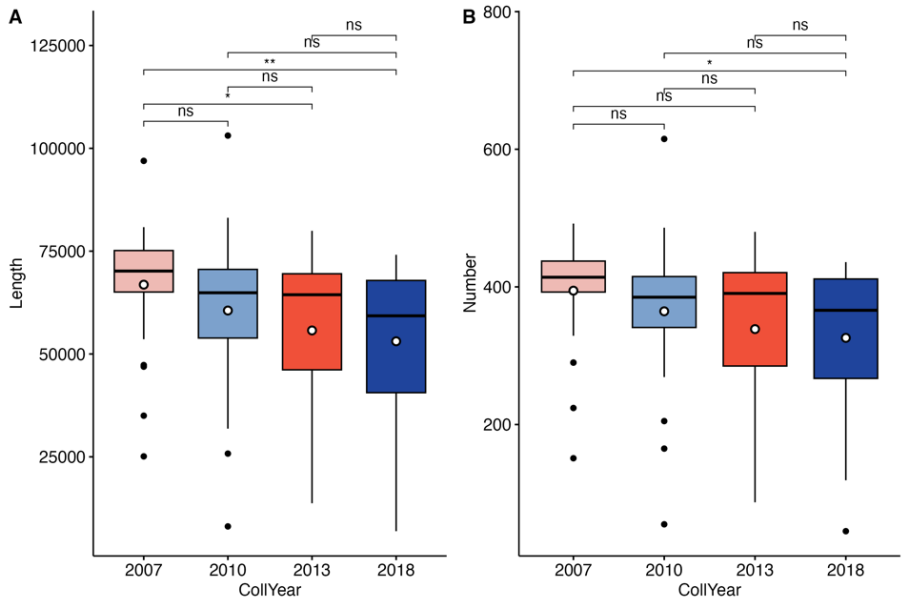
Supplementary Figure 4.1 Distribution of Runs of Homozygosity (RoH) (n = 99)

The length represents the total distance covered by RoH, while the colors indicate the number of individuals with RoH in those areas.

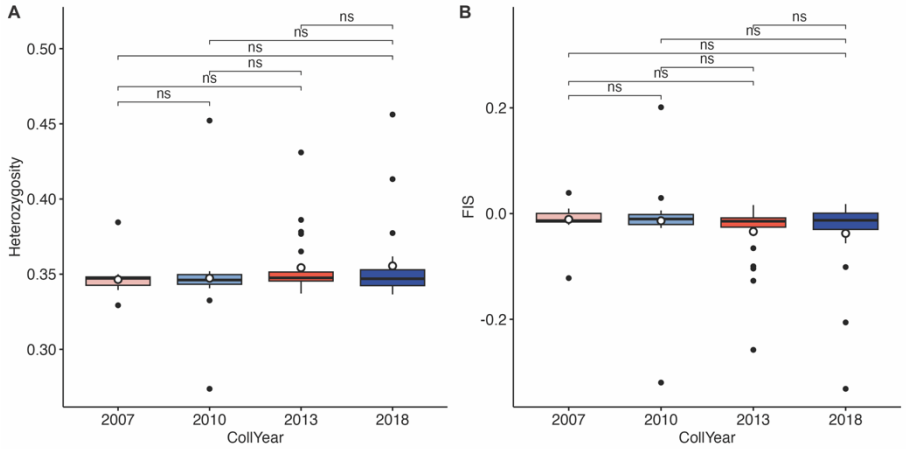


Supplementary Figure 4.2 Distribution of length of RoH in different length categories for each population

Categories: short: 100 Kb ~ 1 Mb; long: > 1Mb; groups are coloured by length categories (2007, n = 27; 2010, n = 23; 2013, n = 26; 2018, n = 23).

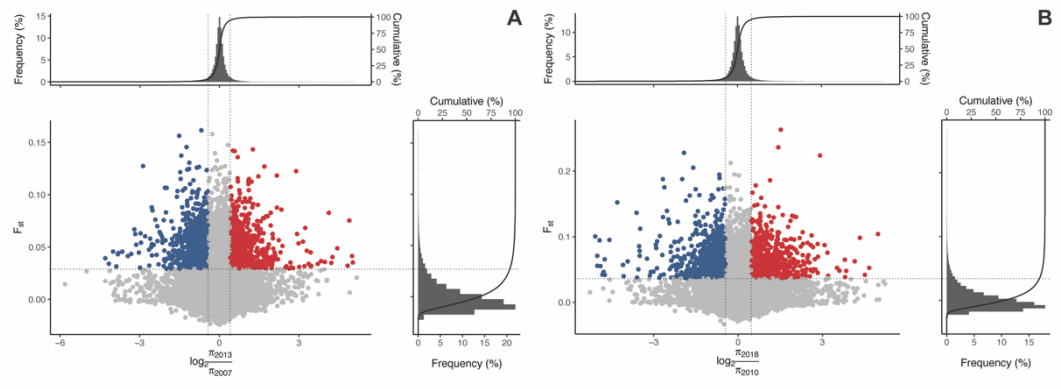


Supplementary Figure 4.3 Comparisons of short RoH (100 Kb ~ 1Mb) for each collecting year (A | Length of short RoH; B | Number of short RoH); The white dots represent the mean values for each group (2007, n = 27; 2010, n = 23; 2013, n = 26; 2018, n = 23).



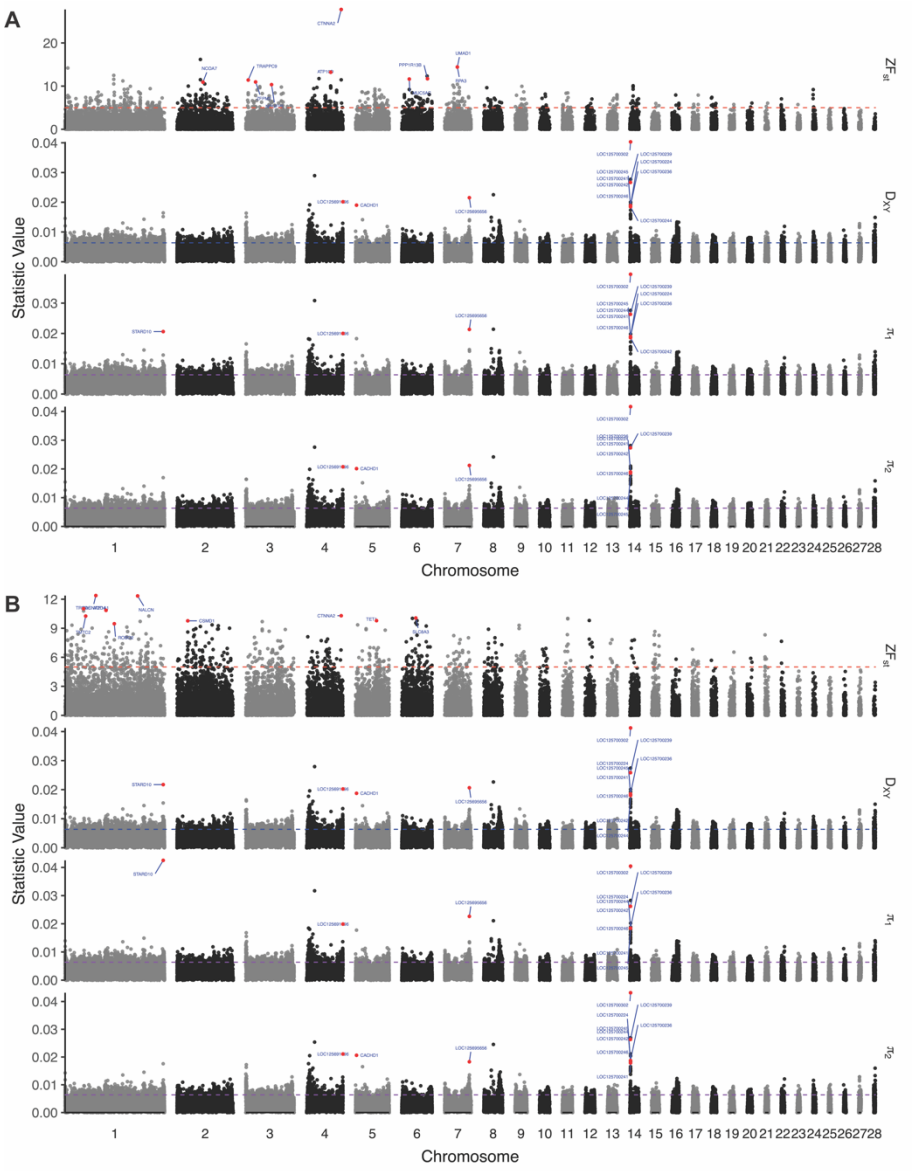
Supplementary Figure 4.4 Heterozygosity and inbreeding coefficients of Icelandic rock ptarmigans from 2007 to 2018

(A | Heterozygosity; B | Inbreeding coefficient). Each group is coloured by collecting year, the white points represent mean values for each sampling year (2007, n = 27; 2010, n = 23; 2013, n = 26; 2018, n = 23).



Supplementary Figure 4.5 Selection analysis based on the π ratio and F_{st} between different groups

(A | The comparisons between trough years, i.e. 2007 and 2013; B | The comparisons between peak years, i.e. 2010 and 2018; 2007, n = 27; 2010, n = 23; 2013, n = 26; 2018, n = 23). The x axis represents π ratio (log2 transformed), the y axis indicates Fst value, and each point represents a window. The coloured regions are the top 5% regions selected.



Supplementary Figure 4.6 Genome wide patterns of differentiation (F_{st} and d_{xy}) and genetic diversity (π) of different groups

(A | Patterns of the comparison between trough years, i.e. 2007 and 2013, and peak years, i.e. 2010 and 2018; B | Patterns of the comparison between 2007 and 2018). The statistical values for the first 28 chromosomes are plotted using 10 Kb sliding windows, with dashed lines indicating the thresholds for each analysis (ZFst at 5 and the top 5% for others). Windows with top 10 values and predicted functions avaliable in NCBI were labelled for each analysis. The panels from top to bottom display ZFst, dxy, and π values for each population. Here, π_1 represents the former population in each comparison, i.e. trough years and 2007, while π_2 indicates the latter population in each comparison, i.e. peak years and 2018.