White-tailed eagles in time and space

Charles Christian Riis Hansen

Faculty of Life and Environmental Sciences
University of Iceland
2021

White-tailed eagles in time and space

Population genetics and the aftermath of severe bottlenecks in *Haliaeetus albicilla*

Charles Christian Riis Hansen

Dissertation submitted in partial fulfillment of a *Philosophiae Doctor* degree in Biology

Supervisor
Snæbjörn Pálsson

PhD Committee
Professor Snæbjörn Pálsson
Professor Agnar Helgason
Professor Tom Gilbert

Opponents
Frank Hailer
Jón Hallsteinn Hallsson

Faculty of Life and Environmental Sciences
School of Engineering and Natural Sciences
University of Iceland
Reykjavik, December 2021

White-tailed eagles in time and space - Population genetics and the aftermath of severe bottlenecks in *Haliaeetus albicilla*
Population genetics in *Haliaeetus albicilla*
Dissertation submitted in partial fulfillment of a *Philosophiae Doctor* degree in Biology

Faculty of Life and Environmental Sciences
School of Engineering and Natural Sciences
University of Iceland
Sturlugata 7
102, Reykjavik
Iceland

Telephone: 525 4000

Bibliographic information:
Charles Christian Riis Hansen, 2021, White-tailed eagles in time and space, PhD dissertation, Faculty of Life and Environmental Sciences, University of Iceland, 214 pp.

Author ORCID: 0000-0003-1554-9840

ISBN 978-9935-9614-7-1

Printing: Háskólaprent, Fálkagata 2, 101, Reykjavík, Iceland
Reykjavik, Iceland, December 2021

Abstract

Genomic population bottlenecks and small population sizes are well described and well-studied areas of biology. However, few studies analyse samples from before and after a population have gone through a genetic bottleneck, and thus studies commonly infer bottleneck effects only using modern samples, modelling, or by comparison with populations that have not gone through bottlenecks. The white-tailed eagle is one of many species which have experienced large reductions in population sizes due to human influences, resulting in extinction in several geographical areas. To study the effects of bottlenecks and genetic variation in white-tailed eagles in Iceland and other countries in the north-western part of the species range, three studies focusing on different aspects of population genomics were conducted. Firstly, the mitogenomic diversity and population history were explored for Iceland and four neighbouring countries, using modern whole mitochondrial genomes, extending knowledge regarding the mitochondrial diversity, and revealing signs of selection maintaining the mitogenomic variation. Secondly, applications for identifying the homogametic sex chromosomes were evaluated. Distinguishing the main chromosomal types is important as they vary in effective population size and may have different evolutionary histories. The third and major study evaluated the impact of a bottleneck and assessed the genomic variation using modern and historic museum samples from Iceland, Greenland and four other countries from the mainland of Europe. A direct effect of bottlenecks was observed on variation within and among countries as well as within and among individuals. Furthermore, the study resolved part of the evolutionary history of the analysed populations.

Útdráttur

Áhrif stofnstærðarsveiflna á erfðabreytileika og hegðun breytileika í litlum stofnum hefur verið vel lýst. Fjölmargar rannsóknir í líffræði hafa verið gerðar á erfðabreytileika stofna en sjaldan hafa sýni verið greind fyrir og eftir slíkar niðursveiflur, sem kallaðar hafa verið erfðafræðilegur flöskuháls, og því eru áhrifin oft greind eftir breytingarnar annað hvort með líkanagerð eða með samanburði við stofna sem hafa ekki farið í gegnum slíkan flöskuháls. Haförninn er ein margra tegunda sem hafa fækkað mikið vegna áhrifa mannsins. Tegundin er útdauð á mörgum svæðum en haferninum fer nú fjölgandi. Til að rannsaka áhrif flöskuhálsa og erfðabreytileika hjá haförnum á Íslandi og í nálægum löndum voru þrjár rannsóknir gerðar á erfðamengjum tegundarinnar. Í fyrsta lagi var erfðabreytileiki og saga stofna á Íslandi og í fjórum nálægum löndum metin með því að greina erfðamengi hvatbera hjá núlifandi örnum. Náttúrulegt val virðist hafa viðhaldið breytileika í hvatberamengjunum innan stofna. Í öðru lagi var árangur fjögurra aðferða til að greina Z-kynlitninginn í erfðamengjum arnanna metinn, en greining á breytileika m.t.t. til meginlitninga er mikilvæg þar sem stofnstærð þeirra er mismunandi og þeir geta átt ólíka þróunarsögu. Þriðja og ítarlegasta rannsóknin var um aðgreiningu í erfðamengjum stofna við Norður Atlantshaf, þ.e. frá Íslandi, Grænlandi og fjórum öðrum löndum á meginlandi Evrópu. Í samanburði á um hundrað ára gömlum sýnum, fengnum úr söfnum, og nýjum sýnum greindust skýr áhrif af flöskuhálsum á breytileika erfðamengjanna, bæði innan og milli stofna. Rannsóknin leiddi einnig í ljós hluta af þróunarsögu þessara ólíku stofna.

I dedicate this work to Ida,

mom, and dad

Table of Contents

List of Figures

List of Tables

List of original papers

The thesis is based on the research conducted on the following papers, two is in review, and one is a manuscript. Hereafter they will be referred to by their numbers:

Paper I: Hansen CCR, Baleka S, Guðjónsdóttir SM, Rasmussen JA, Ballsteros JAC, Hallgrimsson GT, Stefansson RA, von Schmalensee M, Skarphédinsson KH, Labansen AL, Leivits M, Skelmose K, Sonne C, Dietz R, Boetmann D, Eulaers I, Martin MD, Pálsson S (2021). Distinctive mitogenomic lineages within populations of white-tailed eagles (*Haliaeetus albicilla*). Accepted with minor revision, Ornithology.

Paper II: Hansen CCR, Westfall KM, Pálsson S (2021). Evaluation of four methods to identify the homozygotic sex chromosome in small populations. In review with BMC Genomics. Preprint: 10.21203/rs.3.rs-892602/v1

Paper III: Hansen CCR, Rasmussen JA, Ballsteros JAC, Sinding M, Hallgrimsson GT, Stefansson RA, von Schmalensee M, Skarphédinsson KH, Labansen AL, Leivits M, Skelmose K, Sonne C, Dietz R, Boetmann D, Eulaers I, Martin MD, Helgason A, Gilbert MTP, Pálsson S. Genomics of white-tailed eagle (*Haliaeetus albicilla*) in the North-Atlantic islands reveal low diversity and substantial inbreeding in comparison with the mainland populations. In review.

Peer-reviewed paper not included in the thesis:

Hansen et al., The Muskox Lost a Substantial Part of Its Genetic Diversity on Its Long Road to Greenland, Current Biology (2018).

Abbreviations

DNA	Deoxyribonucleic acid
F	Inbreeding coefficient
H	Heterozygosity
H_e	Expected heterozygosity
H_o	Observed heterozygosity
IUCN	International Union of Conservation of Nature
LC	Least Concern
LD	Linkage Disequilibrium
mtDNA	Mitochondrial DNA
N	Census population size
N_e	Effective population size
N_f	Female population size
N_m	Male population size
NUMT	Nuclear Mitochondrial DNA
PAR	Pseudoautosomal region

Acknowledgements

First, I want to thank my supervisor, Snæbjörn Pálsson, without whom none of this would have been possible. I appreciate the hours we have spent both in and out of the university. I will always remember the first time we met, which included a 30-minute meeting about the PhD, and the next six hours spent in a bar.

I thank my committee, Agnar Helgason and Tom Gilbert, who provided both material and intellectual assistance for the project. With this, I thank deCODE genetics for letting me borrow their cluster, help with the sequencing of old samples, and offered me many a lunch. I also want to thank the people both in Agnars and Toms group who helped me, or just had lunch with me during the PhD.

Of course, I want to thank Snæbjörn Palsson´s research group, which has changed a bit over time; but it includes everyone that has been there at one point or another in my time as PhD student; Hildur, Ragnhildur, Kera, Aldís, Jake, Elisabeth, Hafrún, Sólveig. You all made it all less hard when things were not too fun.

Most of Snæbjörns research group is based in Askja, and here there are some other people I want to thank too, for a lot of good times including, talks, coffee and beer to just mention a few activities. I am sure to forget some, but a few I spent more time with than others, who deserve special thanks for making the whole experience a whole lot more fun; Gotje and Hannah.

I want to thank all the co-authors without whom it would have been very hard to get to this point.

Further, I want to thank my girlfriend Ida, who I have missed a lot in this time, but who have supported me along the way, even if it was not easy when living 2,000 km apart. I want to thank my mom Gerd, my best friend Uffe, and my very good friend Kenneth. And lastly, my dad who unfortunately could not see me finish.

Though of course one gets sucked in when doing a PhD, I did notice that there were some other major events in the world in the time it took to form this thesis. Just to mention a few: Brexit, the warmest days registered ever, a pandemic and a volcanic eruption in Iceland.

The project is supported by Research Grant nr 185280-052 from The Icelandic Research Council, the Doctoral student found of the University of Iceland and The University of Iceland Research Fund.

Thank you to my opponents, Frank Hailer and Jón Hallsteinn Hallsson, for agreeing to review my thesis. And lastly, anyone I must have forgotten, that helped or supported me and the project.

1 Introduction

1.1 The study species

Figure 1. Adult white-tailed eagle on a rock in Iceland. Picture by Gunnar Þór Hallgrímsson.

The white-tailed eagle (*Haliaeetus albicilla* Linnaeus, 1758), is also called white-tailed sea eagle (which is the direct Latin translation), grey sea eagle, or just sea eagle, and Haförn in Icelandic (Figure 1). It is a large raptor with a wingspan from 1.8 to 2.5 meters and can weigh up to around 7 kg (Ferguson-Lees and Christie 2001). Being a top predator, it has potential and probable large impact on their ecosystem (Salo et al. 2008, Ritchie et al. 2012), and it has even been mentioned as a flagship species (Langguth et al. 2013, Kitowski et al. 2017). The white-tailed eagle has a large range spread over the Palearctic and Greenland (Figure 2). It is considered as mostly sedentary, but young individuals, as well as certain populations, are known to travel long distances, even up to 2000 kilometres (Literák et al. 2007, Sugimoto and Shiraki 2014). The closest relative to the white-tailed eagle is the bald eagle (*Haliaeetus leucocephalus*) with a split time of around one million years (Mindell et al. 2018), which lives in the Nearctic (Birdlife International 2016) with no overlapping range with the white-tailed eagle. The golden eagle (*Aquila chrysaetos*), a more commonly studied eagle species, has diverged from the white-tailed eagle for about 18 million years (Mindell et al. 2018).

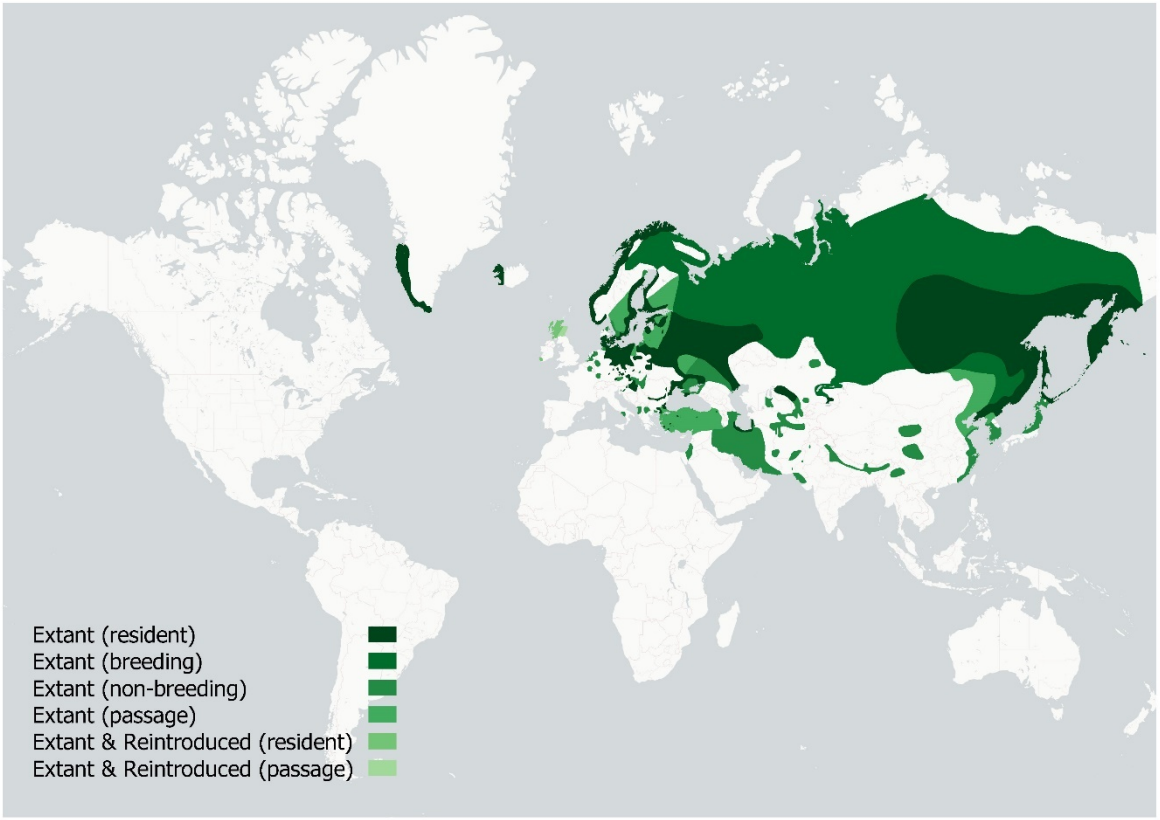


Figure 2. The range of white-tailed eagle. Extant is separated into six categories and include both natural extent and introduced populations. The underlying data is altered from IUCN red list (Birdlife International 2020).

The white-tailed eagle has a life span of 25 years or more (up to 40 in captivity) (Evans et al. 2009, Krüger et al. 2010), and a generation time of 15.6 years (Birdlife International 2020). The world population of mature individuals is estimated to be 20-50 thousand individuals, and up to 20 thousand mature individuals live in Europe alone (Birdlife International 2020). Both in Europe and on the global scale, the species is increasing in numbers, and it is currently categorized as Least Concern (LC) by the International Union of Conservation of Nature (IUCN). However, during the late 19th and 20th century it was decreasing and had become locally extinct in several countries, firstly due to persecution and later due to organic pollutants (Bijleveld 1974, Love and Ball 1979, Helander et al. 1982, 2002; Walker et al. 2009). This study focuses on populations from the islands of Greenland and Iceland and the mainland of northern Europe. Currently, the population in Greenland lives on the southwest coast and consists of around 200 pairs (Boertmann and Bay 2018), in 1950 it was estimated to be only between 50 and 75 pairs (Hansen 1979). In Iceland, the eagles were abundant prior to 1850 and occupied most of the coastline around Iceland. Due to human persecution and fox poisoning, the population plummeted to only around 20 pairs at the beginning of the 20th century, and even though it was conserved by law in 1914, the population didn't start to rise until fox poisoning was banned in 1964 (Petersen 1998, Skarphéðinsson 2013). Today there are probably around 80 pairs (Skarphéðinsson 2013), and the full population is estimated to be roughly 350-400 individuals (Kristinn H. Skarphéðinsson, personal communication). The island populations are considered to be isolated from the mainland populations, and there are no records of migration. This is probably the reason for the previous classification of the Greenlandic population as a

subspecies (*H. a. groenlandicus*) due to their larger size (Salomonsen 1979, Hailer et al. 2007), and further by studies on the mitochondrial DNA that also indicate isolation (Hailer et al. 2007, Langguth et al. 2013). Unlike many other populations the Norwegian population did not suffer from a great reduction in numbers in the last century, and it currently counts around 2,000 breeding pairs (Jais 2020). The Danish population did not only suffer from a reduction in numbers but was extinct altogether at the beginning of the 20th century, but Denmark was recolonized in 1995 and reached 133 breeding pairs in 2020 (Skelmose and Larsen 2021). The Estonian population was before the 19th century a large population with 400-500 breeding pairs (Lõhmus 1998), but this population too suffered and was down to 20 pairs by the end of the 19th century (Randla and Õun 1980), it did however recover and is estimated to 290-330 pairs today (Elts et al. 2019). Despite the efforts of conservation in Iceland and the rise in population size, the fecundity is still only 1/3-1/2 of that in Scandinavia (measured in productivity) (Evans et al. 2009).

Genetically these countries have previously only been studied using the mitochondrial genome and microsatellite markers, and they have been found to differ genetically, but do share some haplotypes and genotypes (Hailer et al. 2007, Honnen et al. 2010, Langguth et al. 2013, Nemesházi et al. 2016). All these areas were likely colonized within the last ~10,000 years, after the deglaciation (Clark and Mix 2002), potentially from one or two suggested glacial refugia (Hailer et al. 2007, Honnen et al. 2010, Langguth et al. 2013).

1.2 Effects of small populations and fluctuation in population sizes (bottlenecks)

Genetic diversity generally reflects a species' effective population size and mutation rate unless it is affected by natural selection. Severe reductions in populations sizes can lead to reduced genetic diversity, a process that is often referred to as a genetic bottleneck, and similar effects can be expected in small island populations (Hartl and Clark 2007). However, it is also easy to imagine scenarios where it can be necessary and important to distinguish between the two constructs: Bottlenecked and being a small island population. One example is that a small island population may be constrained by small carrying capacity (Smith and Smith 2003), whereas a mainland population that has gone through a bottleneck due to some harmful effect though the environment could sustain otherwise a large population. The small island population has most likely experienced effects from a founder event in addition to the environmental constraints when the population first settled on the island (Hartl and Clark 2007).

A genetic bottleneck and a founder event will result in a loss of heterozygosity, loss of variants, distortion of allele frequencies, an increase in linkage disequilibrium (LD) and potentially reduced population connectivity (Hartl and Clark 2007, Nielsen and Slatkin 2013). These consequences can thus cause further differentiation of populations due to genetic drift, and it can cause inbreeding depression in a population (Hartl and Clark 2007, Hedrick and Garcia-Dorado 2016). Loss of heterozygosity and variants will cause less diversity and lower adaption potential in a population (Smith and Smith 2003, Hartl and Clark 2007, Nielsen and Slatkin 2013). Increased LD can also contribute to lower diversity due to selection at linked loci and increased accumulation of harmful mutations in a population (Hartl and Clark 2007, Nielsen and Slatkin 2013). Genetic drift will act more strongly in small populations (no matter if it is a bottleneck population or a small island

population), potentially causing the effects of a bottleneck or founder effect to strengthen or be upheld over time, as the increased effect of genetic drift may override the fate of mutations with weak selection effects. Thus, mutations with slight deleterious effects can accumulate, and new beneficial mutations are less likely to get fixed (Bataillon and Kirkpatrick 2000, Hartl and Clark 2007). Inbreeding depression can follow a genetic bottleneck or founder event as recessive or partially recessive harmful deleterious mutations will more likely be expressed following mating among close relatives, and thus the average individual fitness will decline (Hartl and Clark 2007, Hedrick and Garcia-Dorado 2016). Inbreeding can be quantified in various ways, and I have used three methods: i) The classical method of contrasting expected heterozygosity (H_e), given the allele frequencies within a population under random mating, and the observed heterozygosity (H_o) with the inbreeding coefficient $F=(H_e-H_o)/H_e$ (Nei 1977, Hartl and Clark 2007). ii) Contrasting expected and observed homozygosity across sites on the autosomal sites for an individual *i* as done in PLINK with $F_H = (O_i - E_i)/(L_i - E_i)$, where O is the observed number of homozygous sites, E expected the number of homozygous sites considering the sample frequencies, and L is the number of genotyped SNPs (Purcell et al. 2007). iii) The ratio of the total length of runs of homozygosity to the length of the autosomal genome as $F_{roh} = \Sigma L_{roh}/L_{auto}$ (McQuillan et al. 2008). The coefficient ranges usually from 0 to 1, where F=0 represents complete random mating, and inbreeding causes F to move towards 1 (Nielsen and Slatkin 2013). The inbreeding coefficients F and F_H rising above 0 are dependent on the reference population and can both be due to deviation from random mating within subpopulation (F_{is}) and due to population structures (F_{st}) where subpopulations may differ in frequencies due to drift over time. Negative values can arise under outbreeding or if selection favours heterozygotes. Though inbreeding may be harmful, as it may increase the risk of an effect of recessive deleterious mutations, it does not necessarily lead to inbreeding depression, and it may not have a severe effect on the population if the period of inbreeding is short. It can be problematic to quantify whether there is inbreeding depression in a population, as e.g., a smaller population growth rate in an island population could be caused by environmental constraints, however, it is possible e.g., by assessing variation in fitness or fitness-related traits.

1.3 Population genetics

Population genetics is the study of genetic variation (alleles) in a population. It is concerned with the evolutionary composition and change in populations in the past, present, and future (Nielsen and Slatkin 2013). It is concentrated on analysing genetics through geno- and phenotypes. The offset of population genetics lies at least partly in the work of Gregor Mendel, and his law of inheritance, and was later extended to segregation of alleles in a population with the Hardy-Weinberg principle and further development at the beginning of the 20th century (Hartl and Clark 2007). Population genetics is widely used and is essential in many branches of biology and related sciences such as genetics, genomics, computational biology, evolutionary biology, systematics, natural history, ecology, breeding, forestry, horticulture, conservation, and wildlife management. It is further applied in medicine, law, biotechnology anthropology, molecular biology and more (Hartl and Clark 2007). The use of genetics in conservation is becoming more and more widely used (Hailer 2006, Holderegger et al. 2019), and though it has limitations, in an ever-changing world, with less space for species, due to increase in the human population (Kremer 1993), and the 6th mass extinction ongoing (Barnosky et al. 2011), the use of conservation genetics becomes more

and more important, both to assess genetic adaption potential (including assessment of effective population size), and the conservation of distinct genetic lineages.

1.3.1 Population genetics in the context of conservation

Conservation genetic research is often done by researchers focusing on populations, but with the hope that policymakers will make use of the research when choosing what to conserve and how. This gap between researchers and policymakers is apparent as even IUCN Red List is not currently using genetic factors in the categorisation of threat level for species (Garner et al. 2020). This is an obvious problem, as a general loss of genetic diversity over time is a concern (Hoban et al. 2020, 2021). However, many results from population genetics can be relevant and used in conservation, such as population history and coalescent, diversity, structure, and adaptation, but often especially relevant is loss of heterozygosity and N_e (effective population size), as these can be good and fairly easy to assess measures of use when deciding whether conservation efforts are warranted for a species or population (Hoban et al. 2020).

Genetic studies have revealed that genetic diversity is dependent on population size, however, this diversity does not necessarily reflect the census size (N), but rather it is shaped by sex ratio, past events, variance in the number of offspring, and selection, and thus the effective population size (N_e) is used commonly in population genetics (Frankham 1995b, Charlesworth 2009, Hoban et al. 2021). N_e is the number of individuals in a population that would produce the same amount of genetic drift as in the analysed population (Nielsen and Slatkin 2013). The most extreme cases one can think of with a bias between effective population and census size could be in cloning, where all offspring have the same parent, or in a population with a highly skewed sex ratio e.g., in breeding programs where the effective size will be determined by the sex of lower size, following the equation $N_e=4N_mN_f/(N_m+N_f)$ (Wright 1933, 1939). By analysing more than 100 species Frankham (1995b) estimated that the average rate of N_e to census size was only 0.10-0.11, i.e., only 10%, though there are large differences between species and classes. The ratio N_e/N for the Madagascar fish-eagle, a sister species to the white-tailed eagle, has been estimated to be ~10% (Johnson et al. 2009).

Loss of heterozygosity is the mathematical prediction of how much heterozygosity, and thus adaptation and resilience potential, a population will lose over time, and therefore the term is highly relevant in the context of conservation in any population. The reduction in heterozygosity loss per generation is predicted with $1/(2N_e)$ (Wright 1931, James 1970). In conservation, loss of heterozygosity is sometimes calculated per 100 years, using the equation from Wright (1931): fraction heterozygosity remaining = $[1-(1/2*N_e)]^t$. With t being number of generations in 100 years. Heterozygosity in a population can also be lost or maintained over time due to selection and will thus not follow the equation (e.g., Charlesworth et al. 1993, Pamilo and Pálsson 1998).

Both N_e and expected heterozygosity loss have been suggested as a metric for IUCN Red List assessments (Garner et al. 2020) to assure there is sufficient genetic diversity to avoid inbreeding, and ultimately extinction of populations and species. A proposed threshold for loss of heterozygosity after which management is warranted, and a population could risk extinction, is 5 or 10% (Allendorf and Ryman 2002). Though a reduction of this magnitude may warrant management, there are examples of species going through ancient bottlenecks

but persisting (O'Brien et al. 2017). For N_e, a 50/500 rule was suggested by Ian R Franklin and Michael Soulé in 1980. The rule suggested that an effective population size of a minimum of 50 was needed to prevent inbreeding depression over 5 generations in the wild, and 500 was needed to reduce genetic drift and thus retain evolutionary adaptation potential (Franklin 1980, Soulé 1980). This rule has been suggested to be revised to 100/1000 (Frankham et al. 2014). However, regardless of the chosen threshold being 50/500 or 100/1000, they make up a rule of thumb for the desired effective population size.

In conservation, inbreeding is of concern as stated above due to the increase of homozygous harmful deleterious mutations, and the reduction of adaptation potential (potential for a species or population to adapt to e.g., changing climate), and for small populations management can be warranted (Hedrick and Garcia-Dorado 2016). But while undertaking management and/or genetic rescue, outbreeding depression is a discussed concern (Frankham et al. 2011). A random mating population will have an inbreeding coefficient (F) of zero. Outbreeding is the breeding between individuals from different groups, populations or species and will have an $F<0$ (Ralls et al. 2014). Breeding with a genetically distinct group can be problematic if the outbreeding causes a fitness reduction in the offspring and thus outbreeding depression (Ralls et al. 2014). During outbreeding depression, intermediate genotypes can be formed with lower fitness than the original genotype in either population. An extreme example could be that an allele in two different populations is a fixed homozygous for either variant (i.e., $p(AA)=0$ and $p(AA)=1$), and the heterozygous genotype (Aa) is harmful or either homozygous variant in the populations is causing a difference in adaption, and a heterozygous will be less adapted than the original populations (Frankham 1995a). Outbreeding depression can occur when doing genetic rescue by introducing genetic material from a distant population into a highly adapted or inbred population, causing the population to crash (Frankham et al. 2011). Though outbreeding should be kept in mind, genetic rescue has been shown beneficial in populations, e.g., in the inbred Scandinavian arctic fox (Hasselgren et al. 2018). Examples both for and against genetic rescue are scarce and thus the full effect and scale of outbreeding depression warrants more research, however, the fears of outbreeding depression have been suggested to be exaggerated (Ralls et al. 2018).

1.3.2 Population genetics and different genomic regions

Inferences based on genetic variation, effective population size (N_e) and population structure are dependent on which region of the genome is being analysed i.e., autosomes, mitochondrial DNA, or sex chromosomes, as they differ in ploidy, substitution, and recombination rates, and thus identification of these genome types are also of high importance (Wilson Sayres 2018). Due to these differences, drift and selection pressure will differ between the genomic regions e.g., between haploid mtDNA and autosomes, and due to selection within non-recombining regions due Hill-Robertson effect (Hill and Robertson 1966).

The different properties of the chromosomal regions can lead to different evolutionary trajectories; thus the genetic diversity, N_e, and population structure can vary and may cause different or even contradicting conclusions dependent on which region is being analysed (Shaw 2002, Ballard and Whitlock 2004). Complications about inference but also discovery can further occur as the regions may not follow the expectation, as e.g., the autosomes can contain nuclear mitochondrial DNA (NUMT) (Simone et al. 2011) or haploid sex chromosomes can have pseudoautosomal regions (PAR) (Mangs and Morris 2007). Thus,

when analysing mtDNA, some nucleotides may appear heterozygous, though the mtDNA is haploid, as NUMT's from the autosomal genome may map to the mtDNA, and when trying to isolate the sex chromosome, some regions may appear to be diploid and show recombination between the two different sex chromosomes (i.e., Z and W or X and Y), further complications when analysing and discovering the sex chromosomes arise with non-recombining homologous regions (gametologs), as they will also have different evolutionary histories (Garcia-Moreno and Mindell 2000). These are all problems that can make bioinformatical identification of different genomic regions challenging. However, as described it is important to be aware of which regions are being analysed, and thus methods of easy identification of the different chromosomes are necessary.

1.4 Aim and objectives

The overall aim of this project was to evaluate the effects of bottlenecks and small population sizes on the genetic composition of the white-tailed eagle. Using a combination of temporal and spatial samples, we aimed furthermore to uncover how different genomic patterns within this species have changed over the 20th century.

In **paper I** we explored the whole mitogenomic diversity of contemporary white-tailed eagle samples from Greenland, Iceland, Norway, Denmark, and Estonia, to investigate for spatial diversity, divergence, and evolution between populations, and compare it to previous work only done on a small part of the mitochondrial genome.

The focus of **Paper II** was different from papers I and III, as we took a step back from population genomic analysis and looked at the underlying requirements for doing precise and coherent genomics. In this paper, the objective was to investigate bioinformatical methods of finding the homogametic sex chromosome in draft genomes generated from non-model species. This was especially important when this project started as there was at the time no obvious reference genome with known chromosomes. There were multiple species with known autosomes and Z-chromosomes (very few with the W-chromosome), but all these had a split time with the white-tailed eagle more than 63 million years ago.

In **paper III** the emphasis was shifted back to the white-tailed eagle genomics, and the consequences of small population sizes and bottlenecks. Here we made use of historic and contemporary samples from six different countries, analysing the whole autosomal genome diversity, divergence, and coalescence. The objective was to directly analyse the effects of the bottlenecks and small population sizes, using the temporal samples (that were sampled at the beginning and after the bottleneck), with particular emphasis on the small, bottlenecked population in Iceland.

2 Methods

2.1 Sampling

Contemporary (collected post-1990) and historic (collected between 1885 and 1950) white-tailed eagle samples were obtained from Greenland, Iceland, Norway, and Denmark. In addition samples from three contemporary specimens from Estonia and historic samples from Turkey and Russia were obtained (Table 1). These samples were analysed to a different extent in the three papers presented below.

Blood was sampled from all contemporary Icelandic individuals in an ongoing monitoring project being conducted in Iceland since 2001, led by Kristinn H. Skarphéðinsson at the Natural History Institute of Iceland and kindly provided by him. The contemporary specimens from Estonia, Denmark and Greenland consisted of whole blood and muscle tissue and were provided by the Department of Bioscience, Arctic Research Centre, AU, Roskilde, Denmark (Estonian, Danish and Greenland samples), Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark (Danish samples) and the Greenland Institute of Natural Resources, Nuuk, Greenland (Greenland samples). The contemporary Norwegian individuals were provided as sequences from the collaborator Michael D. Martin from the Department of Natural History, NTNU University Museum, Norwegian University of Science and Technology (NTNU), Trondheim, Norway.

Historic samples all consisted of toepad clippings, taken with disposable sterile scalpel blades. The samples from Greenland, Iceland, Denmark, Turkey, and Norway were from museum samples provided by The Natural History Museum of Denmark; Icelandic Institute of Natural History, and Department of Natural History, NTNU University Museum. Around two-thirds of the Norwegian specimens were sampled by others, the rest I sampled.

Finally, one paper made use of three already published genomes, two from golden eagle with one consisting of 1142 scaffolds and one assembled to chromosome level (GenBank Assembly Accession numbers: GCA_000766835.1 and GCA_900496995.2, respectively), and one chicken assembly that contains both chromosomes and is annotated (GenBank Assembly Accession: GCA_000002315.3).

In **paper I,** 89 individuals from contemporary samples from Greenland, Iceland, Norway, Denmark, and Estonia were used. **Paper II** consisted of a total of 135 contemporary individuals from Iceland, and the three published genomes described above. In **paper III,** 92 historic and contemporary individuals from Greenland, Iceland, Norway, Denmark, Estonia, and Turkey were used (Table 1). Known locations for the individuals used in **papers I** and **III** can be seen in Figure 3, and individual information can be found in the supplementary for **papers I** and **III**.

Table 1. Number of individuals and their geographic origin analysed in the three manuscripts and the thesis chapter "Genetic diversity of the eagles, past, present, future". The column labels "cont" and "hist" refer to contemporary and historic samples.

	Paper I	**Paper II**	**Paper III**		**Thesis analysis**	
	Cont	Cont	Hist	Cont	Hist	Cont
Greenland	12	-	8	12	16	12
Iceland	42	135	2	25	4	42
Norway	21	-	13	12	16	12
Denmark	11	-	5	11	13	12
Estonia	3	-	-	3	-	3
Turkey	-	-	1	-	1	-
Russia	-	-	-	-	1	-

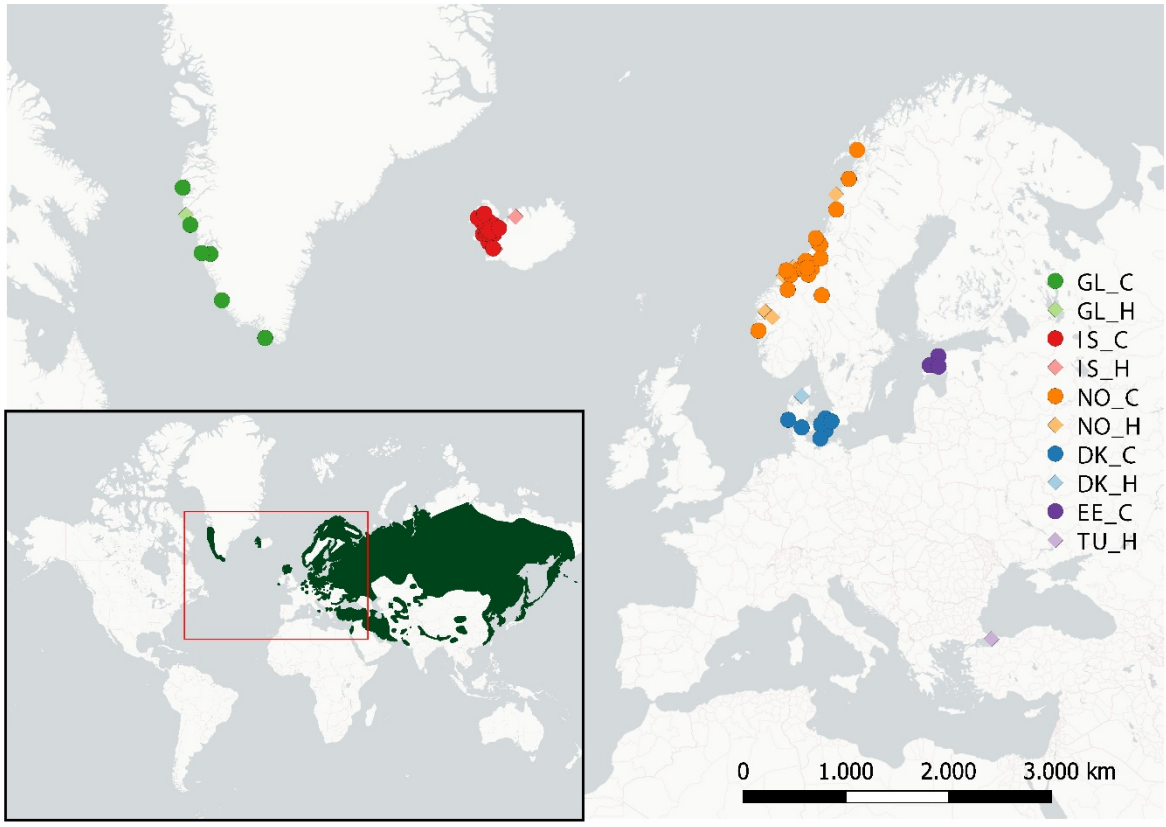


Figure 3. Origin of individuals in papers I, and III. The world map shows the full species range. Dots indicate contemporary samples ("C"), diamonds indicate historic samples ("H"). Abbreviations are GL=Greenland, IS=Iceland, NO=Norway, DK=Denmark, EE=Estonia, TU=Turkey.

2.2 DNA extraction, RADseq and genome sequencing

Contemporary DNA was extracted using Thermo Fisher GeneJET Whole Blood Genomic DNA Purification Mini Kit following the standard protocol (Thermo Fisher 2016) and Macherey-Nagel NucleoMag® 96 Tissue kit (Macherey-Nagel 2014) for tissue, at the University of Iceland. All historic samples except eight Norwegian individuals were processed at the clean laboratory facilities at the Globe Institute at the University of Copenhagen. To minimize cross-contamination from other museum samples they were firstly rinsed using bleach, ethanol, and molecular biology grade water. DNA was then extracted with DNeasy Blood and Tissue Kit Protocol (Qiagen 2006) with modification.

For use in **papers I** and **III,** 21 contemporary and eight historic Norwegian individuals were processed at the NTNU University Museum's standard molecular genetics and dedicated palaeo-genomics laboratories, respectively. Extraction of both contemporary and historic individuals was performed with a Qiagen DNeasy Blood & Tissue kit (Qiagen 2006), using different modifications (see papers I and III).

For all historic samples, both processed in Denmark and Norway, blunt-end Illumina shotgun sequencing libraries were prepared using the Blunt-End-Sigle-Tube (BEST) protocol (Carøe 2017).

Contemporary samples processed at the University of Iceland were sequenced at BGI Hong Kong using DNBseq Normal DNA library construction and DNBseq PE150. Contemporary samples processed in Norway were sequenced over two runs on the Illumina HiSeq 4000 platform at the NTNU Genomics Core Facility and over one run on the Illumina NovaSeq 6000 platform at the University of Oslo Norwegian National Sequencing Centre.

Historic samples processed in Copenhagen were paired-end sequenced on four flow cells with 2x150 bp read length at deCODE Genetics in Iceland using an Illumina NovaSeq 6000. The historic samples processed in Norway were sequenced the same as the contemporary samples sequenced in Norway.

Preparations of the RADseq libraries from the 133 individuals used in **paper II** were done by Kristen M. Westfalls, a former post-doctoral researcher at the University of Iceland. These libraries were sequenced at deCODE genetics.

2.3 Mapping

Before mapping the sequences, in all three manuscripts, the reads were run through FastQC (Babraham Bioinformatics 2010) and AdapterRemoval v2 (Schubert et al. 2016), to check the quality and to remove adapters.

In **paper I** sequences from 89 individuals from Greenland (12), Iceland (42), Norway (21), Denmark (11), and Estonia (3) were mapped to the white-tailed eagle mitochondrial genome of a Korean specimen (NCBI: NC_040858.1) (Kim et al. 2019) with Burrows-Wheeler Aligner (BWA) mem and converted to bam files using samtools view and sort (Li and Durbin

2009, Li et al. 2009). Bam files were transformed to vcf and fasta files using samtools mpileup with quality filters -q 30 and -Q 20, and bcftools call (Li et al. 2009, Li 2011a, b).

In **paper II** the RADseq and two additional high depth shotgun sequenced individuals were also mapped using BWA mem and samtools, but to the golden eagle reference genome assembled to scaffold level (GCA_000766835.1).

Fastq files from 92 individuals from Greenland (12 contemporary and eight historic), Iceland (25 contemporary and two historic), Norway (12 contemporary and 13 historic), Denmark (11 contemporary and five historic), Estonia (three contemporary), and Turkey (one historic) generated for and used in **paper III,** were mapped to the newly published golden eagle (*Aquila chrysaetos*) genome (GCA_900496995.3) using bwa aln, samse and sampe, with the flags -q 15 and -k 1 (Li and Durbin 2009). Though there are recently published white-tailed eagle genomes, choosing the golden eagle had two benefits, firstly it ensured that no populations were closer to the reference than others, which has been shown to potentially cause problems (Gopalakrishnan et al. 2017), secondly, the published golden eagle genome have identified chromosomes, which ensured that only autosomes would be analysed.

2.4 Analysis

As the focus and datasets for the three manuscripts differ, a variety of approaches and analyses were applied.

In **paper I** the mitogenomic patterns in white-tailed eagle were analysed, primarily using R and Rstudio (RStudio Team 2019, R core Team 2020). The R packages pegas (Paradis 2010), hierfstat (Goudet 2005) and MASS (Ripley 1996, Cox and Cox 2008) were used to calculate different measures of molecular diversity and differentiation within and between populations. popART (Bandelt et al. 1999) and BEAST (Bouckaert et al. 2014) were used to construct the genetic relationship between individuals in a network and a tree (including divergence times calculated with six different calibrations) and mapped with respect to the country of origin. Finally, signs of selection along the mtDNA were evaluated using MacDonald and Kreitman test (Egea et al. 2008), and selection for the whole genome per population using Tajima's D (Tajima 1989), Fay and Wu's H (Fay and Wu 2000), and the E-test (Zeng et al. 2006) using Zeng's DH-software.

The aim of **paper II** was to identify the homogametic sex chromosome in a genome with unknown chromosomes using four different approaches and evaluate their accuracy, using the golden eagle reference genome (a male) and sequenced white-tailed eagle genomes as models. The four approaches were: depth, heterozygosity, mapping, and SNP-loading. For the depth, we calculated and compared relative sequencing depth per scaffold for two high-depth individuals, a male and a female, using bedtools (Quinlan and Hall 2010), as the homogametic sex chromosome in birds (Z) is present in one copy in females but two in males. For the heterozygosity approach we started by calling genotypes, again for the high-depth male and female, using Graphtyper (Eggertsson et al. 2017, 2019), and then by looking at heterozygosity per scaffold in windows of 50kb over each scaffold, with the expectation of scaffolds from the Z-chromosome having ¾ of heterozygosity for autosomes when looking at both individuals, and no heterozygosity for the female on the Z-chromosome. The mapping approach was quite simple as the golden eagle reference with unknown parts was

mapped to a distant relative with known chromosomes (the chicken) using LASTZ (Harris 2007). Finally, we looked at SNP-loadings estimated from a PCA calculated using PCangsd (Korneliussen et al. 2014, Meisner and Albrechtsen 2018) in which males and females had been separated on the first axis, and the hypothesis was thus that the sex-chromosome was driving the split. The resulting scaffolds identified as Z-chromosomal in each of the four analyses were then compared to what is known to be the Z-chromosome by mapping the golden eagle reference genome with unknown chromosomes to a newly published golden eagle reference genome with known chromosomes (Z and autosomes).

In **paper III** the focus shifted back to population genomics with the aim of analysing spatiotemporal patterns in the white-tailed eagle, and most analyses were done on called genotypes. Diversity was calculated using VCFtools (Danecek et al. 2011). Structure, admixture, and divergence were examined with PCA, admixture plot, dendrogram based on IBS (Identity-By-Decent), and Nei's G_{st}, using EIGENSOFT (Patterson et al. 2006, Price et al. 2006), ADMIXTURE (Alexander et al. 2009), SNPRelate (Zheng et al. 2012), and vcfR (Knaus and Grünwald 2016), respectively. Autozygosity and bottlenecks were analysed using VCFtools (Danecek et al. 2011) and PLINK (Purcell et al. 2007). Divergence and evolutionary history were further analysed using Treemix (Shriner et al. 2014) with the bald eagle (*Haliaeetus leucocephalus*) as an outgroup. The potential selection was analysed with Tajima's D (Tajima 1989), again using VCFtools. Finally, changes in effective population size back in time was evaluated using stairway Plot v2 (Liu and Fu 2015, 2020) on an SFS (site-frequency-spectrum) calculated directly from the bam files in an attempt to not bias it by using the highly filtered vcf file, using ANGSD realSFS (Li 2011a, Nielsen et al. 2012, Korneliussen et al. 2014).

3 Results and discussion

3.1 Mitogenomics

In **paper I** the mitogenomic differences in white-tailed eagles from Greenland, Iceland, Norway, Denmark, and Estonia were analysed. A split is observed between the countries as they differ in haplotypic composition and/or haplotype frequency, indicating limited or no gene flow between the countries. Two distinct haplogroup lineages that show a split between eastern and western Europe are observed (HG-A and HG-B in Figure 4), which were interpreted as the earlier suggested eastern and western clades covering the range of the white-tailed eagle (Hailer et al. 2007, Langguth et al. 2013). The recently established population in Denmark is found to contain haplotypes from both clades. The structure between and within countries are supported by a Bayesian tree. Two distinct lineages are observed within Greenland, Iceland, and Norway which are shared by the three countries, and which have been upheld for a long time, as shown by the divergence times found in the tree. The patterns in the network, tree, and the negative Fay and Wu's H in Greenland and Iceland, and high positive Zeng's E, indicate that admixture or balancing selection is maintaining the two lineages within countries. Considering the lack of migration and the small populations it seems that the selection has maintained the variation within countries. As the mitochondria are haploid, it is unlikely that selection acting directly on the mitochondria is responsible for the maintenance of this variation. As the mitochondria are inherited with the W-chromosome, any form of balancing selection acting on W and Z may maintain the variation on the mitogenome.

In paper I estimates of split times in the neighbour-joining tree and network were also estimated. Six different calibrations were applied which gave varying estimates, however, conclusions were primarily made from the estimates based on the standard rate of 2.1% divergence in birds for cytochrome B, as it is widely used, the results are in line with earlier estimates, and the results fit at least to some extent with the known history of ice coverage for the analysed countries, and refugia theory for the white-tailed eagle during the last glaciation.

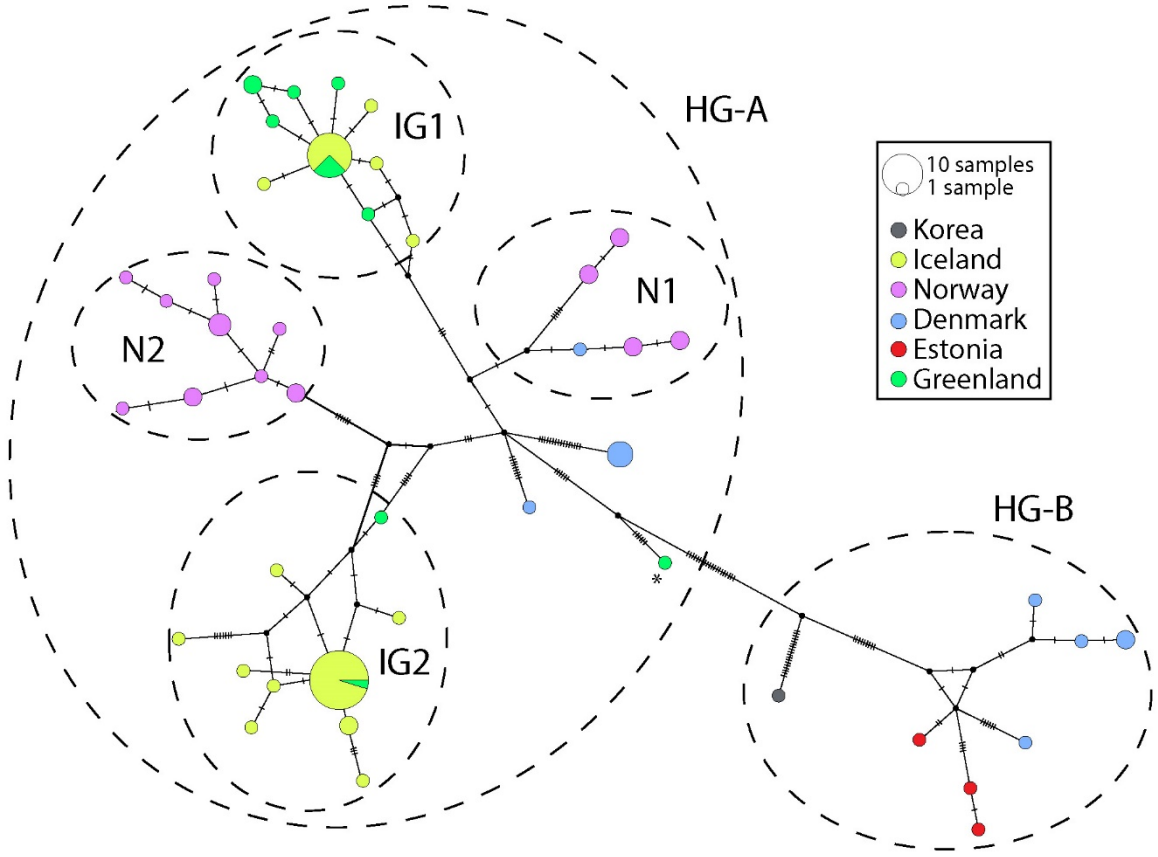


Figure 4. Median-joining network presenting the divergence between mitochondrial genomes in white-tailed eagles from the five different countries sampled between 1990-2018 compared to the Korean reference specimen. The size of the pies represents the number of individuals that share the same haplotype. Marks on the lines refer to the number of sites that differ among the connected haplotypes. HG-A and HG-B correspond to the haplogroups A and B from Hailer et al (2007). IG1, IG2, N1 and N2 are Iceland-Greenland group 1, Iceland-Greenland group 2, Norway group 1 and Norway group 2, respectively. An individual from Greenland displaying signs of being a chimaera or incorrect sequence is marked with an asterisk.

3.2 Sex chromosome

The aim of **paper II** was to evaluate four approaches to identify the homogametic sex chromosome (Z-chromosome in birds). Three of the four were able to discover a high fraction of the homogametic sex chromosome in the analysed scaffolds assembled genome. Of these three, LASTZ mapping performed best (Figure 5), both assigning the Z-chromosome correctly, but also in having the least false positives and negatives (i.e., autosomal scaffolds looking like Z-chromosome and Z-chromosomal scaffolds looking like autosomes, respectively). The depth approach found second-most, and SNP-loadings third most (Figure 5), with false positives and negatives following the same pattern.

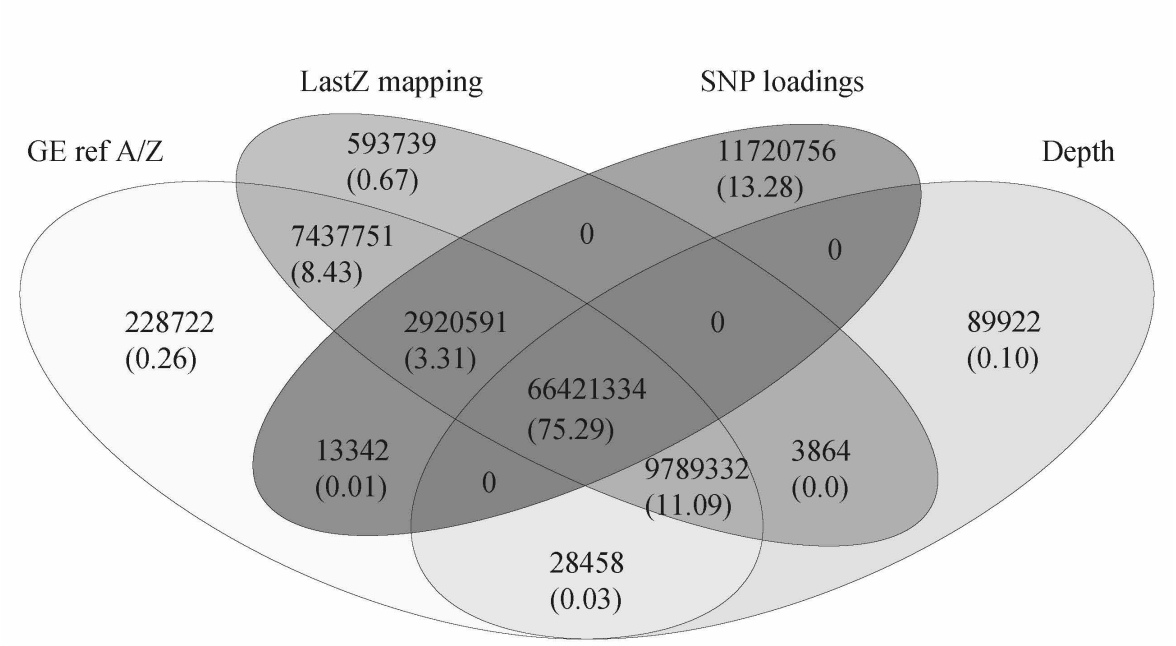


Figure 5. Venn diagram summarizing the size of scaffolds in bases identified as Z-chromosome with the three different analyses: mapping, depth, and SNP-loadings. The Z-chromosomal scaffolds were assigned by mapping the genome with scaffolds to the genome with known chromosomes. Values in parentheses represent percentage size compared to the size of the known Z-chromosome. Note that the percentage found by mapping the golden eagle scaffold assembly to the golden eagle genome is only 98.42%.

Mapping to a distantly related species can be problematic due to chromosomal changes, however, this did not seem to be a big concern here, though it should be considered that the sex chromosome in birds may be well conserved compared to other classes. The scaffolds not identified as Z-chromosomal in the depth analysis was especially concentrated on four scaffolds and are likely due to pseudoautosomal regions (PAR) on the Z-chromosome, or potentially gametologs. The scaffolds not identified as Z-chromosomal in the SNP-loading approach, are probably due to them exhibiting very few differences between the male and female, and thus they do not show the high expected loading, combined with no precise distinction of the signal for Z-chromosomal and autosomal scaffolds. Also, here the three of the four scaffolds from the depth analysis didn't identify as Z-chromosome in the SNP-loading analysis, which supports that they are not different in the two sexes, i.e., they could contain PAR or gametologs. Finally, the heterozygosity approach was not able to confidently distinguish the Z-chromosomal scaffolds. There are probably numerous reasons in combination with one another, but we especially assigned it to the presumed low effective population size in Iceland (Skarphéðinsson 2013), meaning autosomes and Z-chromosomes will not show a big difference, and further again PAR and gametologs can make the Z-chromosomal scaffolds in the female look like autosomes. And finally, it could be a difference in the sex chromosomes between the reference (golden eagle) and applied sequenced genomes (white-tailed eagle).

3.3 Genetic analysis of the autosomes

In **paper III** the population genetics of the eagles was investigated using the autosomes of spatiotemporally distributed samples. Unlike the mitochondrial genomes, the autosome show differentiation between the countries as expected by geography, with samples within countries mostly making up monophyletic groups, the exception being in Denmark where the historic and contemporary samples do not make up a monophyletic group (Figure 6). This can be explained by the extinction of the Danish population at the turn of the last century and its reestablishment at its end. To this pattern, we also find that the island and the mainland populations make up two separate monophyletic groups (Figure 6). When only considering the contemporary samples, clear signs of reduced diversity, and increased inbreeding in the island populations are seen. Two of the three applied methods for estimating inbreeding (F_H and F_{roh}), found consistent and high amounts of inbreeding in the island samples (this was the case for both temporal samples), with the contemporary Icelandic sample having the highest mean inbreeding of 0.39 (F_H). Further, increased drift in the island populations is seen, all of which is in line with the founder event that must have happened when the islands were colonized and are supported in the Treemix analysis, in which a migration edge shows large migration from Norway to the islands. This is supported by signs of ancient bottlenecks in both the island populations when looking at ROH. Iceland further also shows signs of a recent bottleneck.

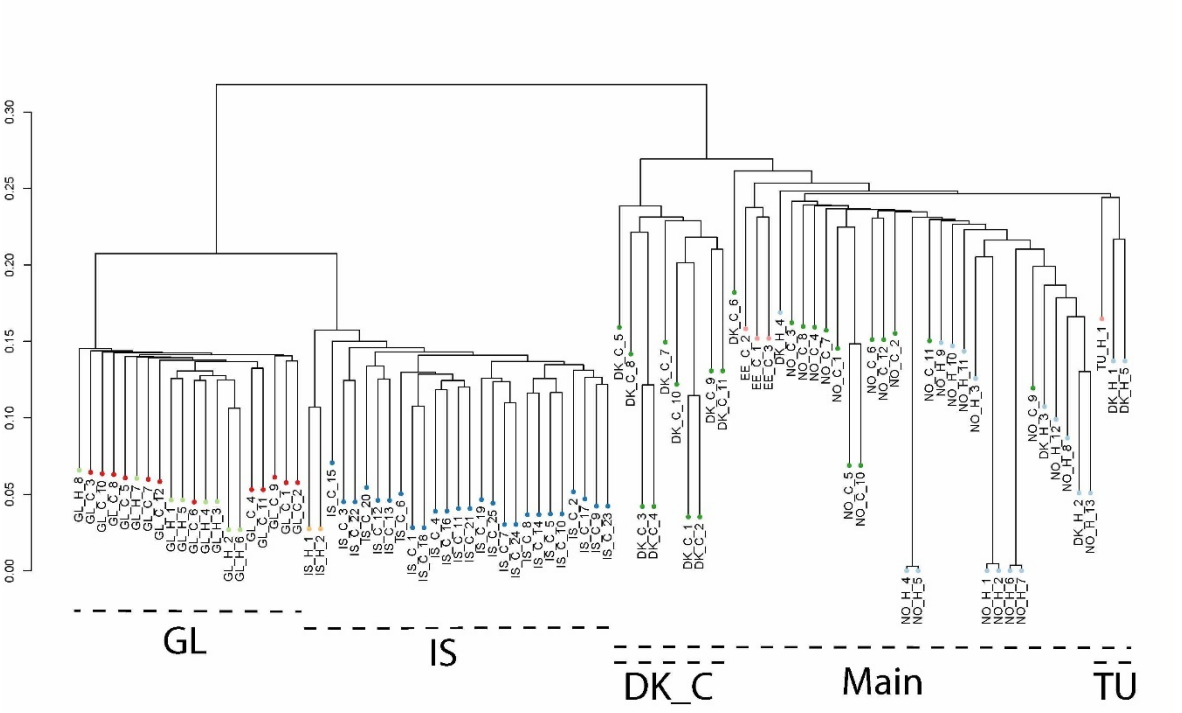


Figure 6. Dendrogram displaying Identity-By-Decent (IBD) between all samples. "C" refers to contemporary samples and "H" to historic samples. Two major clades are found, one for Iceland and Greenland, which then split in two (GL and IS), and one for the mainland populations (Main), which firstly split between all but one contemporary Danish individual (DK_C) and the rest. Then secondly the Estonian sample and the last contemporary Danish individual split of (in Main). Turkey and two historic Danish individuals make up one branch (TU), and all the Norwegian historic, as well as contemporary samples, make up a cluster together with two historic Danish specimens (in Main).

Using stairway plot v2 analysis (Liu and Fu 2020), we look at the effective population size back in time and see that there has been a general and substantial reduction in population size over time and that the current day effective population sizes are extremely low, with the island populations being the lowest (Figure 7). When applying the higher mutation rate, the effective population is 3 (1-21) for Greenland, 3 (1-15) for Iceland, 8 (1-54) for Norway, and 6 (1-38) for Denmark, with 2.5-97.5 confident interval in parentheses (Figure 7). The lower mutation rate gives slightly higher but very comparable estimates. The stairway plot analysis reveals some very sudden drops in population for all analysed populations, which can indicate a population split. When looking at these for the higher mutation rate a split 5-8,000 years ago is found (Figure 8), which could very well be the split of these populations and their colonization in their respective countries after the ice retracted with the last glacial period (Clark and Mix 2002). A drop in population size for all populations is also found 35-55,000 years BP, which could be when the species was split into the eastern and western clade driven by refugia during the last glacial period as earlier suggested (Hailer et al. 2007, Langguth et al. 2013). This is supported by Ponnikas et al. (2013), which found patterns of a rapid population expansion in the eastern clade no more than 30,000 years ago, using microsatellites. Times for these two events are not in line with what we found in paper I, which could have multiple reasons. Firstly, as stated earlier the time estimates from paper I are very uncertain, they are further based on only one marker, and we find that the lineages have been subject to balancing selection, which will inflate the times we found. Also, the time estimates in paper III are unsure due to the lack of an accurate mutation rate. Finally, two drops are seen which happened very recently, one ~1,000 years BP, and one during the last couple of hundred years (Figure 8). The former could be due to human expansion and their population growth, e.g., coinciding with the settlement of Iceland about 1100 years ago, and the latter due to human persecution and toxic pollutions.

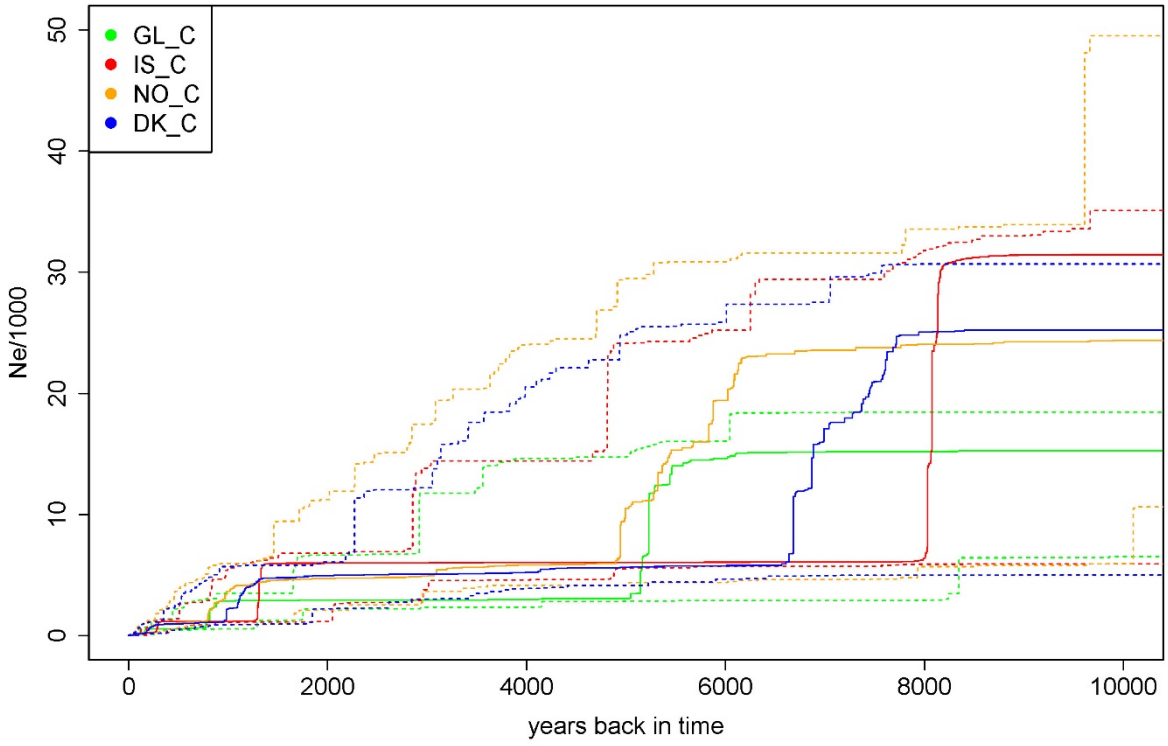


Figure 7. Changes in effective population sizes over time. The sizes were estimated with the stairway plot v2 method for the contemporary samples from Greenland (GL_C, green), Iceland (IS_C, red), Norway

(NO_C, yellow) and Denmark (DK_C, blue). The x-axis display years back in time, y-axis displays the effective population size divided by 1000. A mutation rate from collared flycatcher, 2.3e-9, was used to scale the Ne as it is the only known mutation rate for birds (Smeds et al. 2016). A generation time of 15.6 years was used as reported by IUCN red list (Birdlife International 2020).

The temporal samples did not reveal great differences over time, except in Denmark where the current population reflects a newly established population. We do however see signs in Iceland that the recent bottleneck has had an impact, as we find signs of increased genetic drift in the contemporary samples, and a distinction between the historic and contemporary samples.

4 Genetic diversity of the eagles past, present and future

In **Paper III** the heterozygosity per analysed SNPs was calculated, in this chapter the estimates are reassessed by adding more individuals with lower sequencing depth and the expected loss of heterozygosity in the future is further estimated.

4.1.1 Methods

As the heterozygosity calculated in paper III is not a true reflection of the genome-wide heterozygosity but only the analysed SNPs, I have estimated the heterozygosity for each sample using ANGSD which is based on genotype likelihoods (Nielsen et al. 2012, Korneliussen et al. 2014) rather than called genotypes and can thus use lower depth data than what is preferred for called genotypes. As samples sequenced at lower depth were usable here, more samples were included than in the previous analysis, but they were processed to bam files as described in paper III. The number of individuals used per sample was: Greenland 12 contemporary and 16 historic, Iceland 42 contemporary and four historic, Norway 12 contemporary and 16 historic, Denmark 12 contemporary and 13 historic, Estonia three contemporary, Turkey one historic, and finally one historic individual from Russia. Using ANGSD, firstly site allele frequency likelihood files (SAF) were calculated from all these specimens individually on the autosomes only, with the settings gl 1, doSaf 1, minMapQ 30, minQ 20 and uniqueOnly 1. From these output files and using the ANGSD realSFS program with standard settings but the modification tole 1e-15 I calculated per individual SFSs, and from that individual heterozygosity was estimated as the fraction of heterozygotes over all sites.

As past and present N_e is calculated for four populations in paper III (Figure 7 herein), we can use the estimated effective population sizes to calculate the expected percentage of remaining heterozygosity in a hundred years using the Wrights equation (1931) presented in the introduction, assuming Hardy-Weinberg i.e., that the proportion of heterozygotes is equal to the expected heterozygosity. I calculate the expected percentage of remaining heterozygosity using the generation time of 15.6 years, and for both the mutation rate of 2.3e-9 and 1e-9 used in paper III.

These calculations of past, present, and future heterozygosity are used to both state what the development of diversity has been in these populations in the time period for which we have temporal samples, but also to estimate what will happen in the future for these populations, and thus it can be decided whether further conservation efforts are needed for these populations.

4.1.2 Results and discussion

Examining the present diversity, I find that the island populations display the lowest heterozygosity, compared to the mainland, and further in the contemporary mainland samples, there is more diversity in Estonia and Denmark than in Norway, which is however known to be a large population that has not gone through a bottleneck in recent time (Figure 8). I don't have mainland contemporary samples from outside the northwest corner of Europe but comparing these to the historic samples from Russia and Turkey, we find here

some of the highest heterozygosity (though there is only one individual analysed from each country, Figure 8). These signals of lower heterozygosity, partly in the northwest corner of Europe, but even more so on the islands, is in line with the idea of a refugium to the south of the ice edge during the last glacial period (Hailer et al. 2007, Langguth et al. 2013) and the spread from there possibly through founder events, firstly north, and then to the islands. And though the Norwegian population has been referred to as a major stronghold of individuals (Hailer et al. 2006), recent studies have shown greater diversity in central Europe (Literák et al. 2007).

The historic samples display the same signs as the contemporary, with the islands having the lowest diversity, then Norway, Denmark, and then Turkey and Russia (Figure 8). Comparing the mean values between the countries with temporal samples revealed a great reduction in heterozygosity in all countries from the sampling of the historic to the contemporary samples (Figure 8). We found that the reduction per country was as follows: Greenland 57.5%, Iceland 48.5%, Norway 32.3%, and Denmark 40.3%. Thus, the islands show the greatest reduction, and Norway has the least reduction, and Denmark falls intermediate. These results clearly show, even for the Norwegian population which is not known to have gone through a bottleneck, that there has been a major reduction in diversity over the last ~70-130 years.

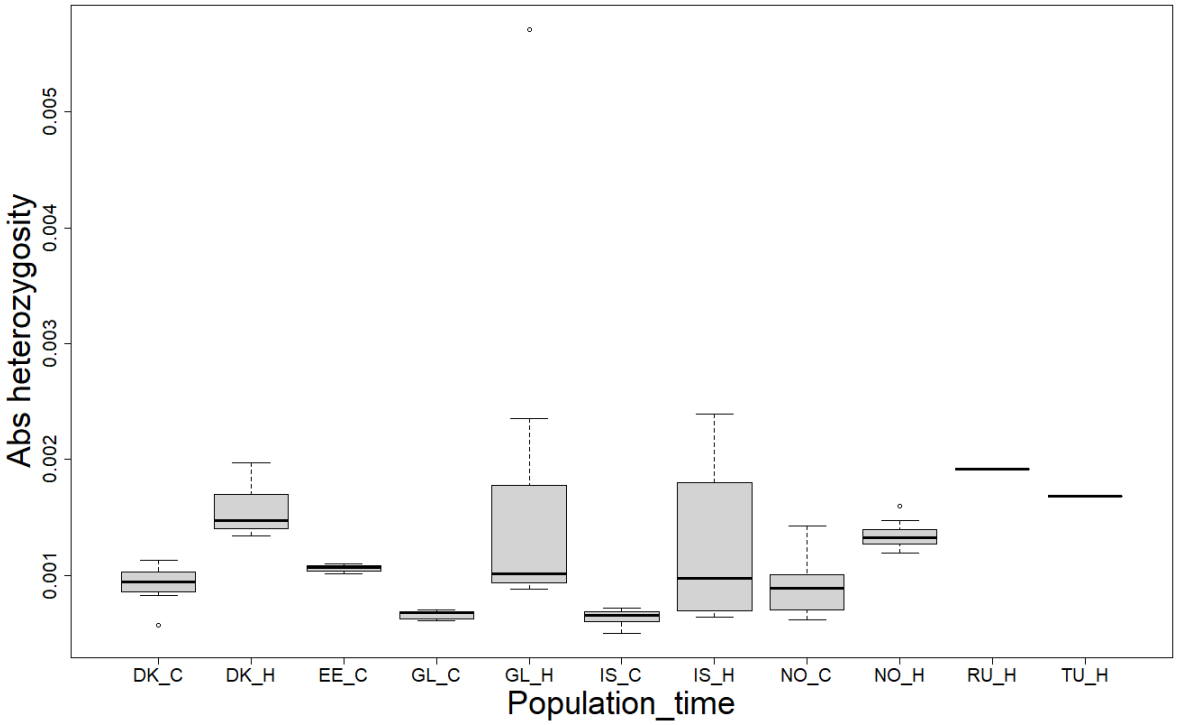


Figure 8. Absolute observed heterozygosity per contemporary and historic sample. The estimates are calculated per individual as the proportion of heterozygous sites over all sites. "Population_time" refers to country and whether it is contemporary ("C") or historic ("H") samples. DK=Denmark, EE=Estonia, GL=Greenland, IS=Iceland, NO=Norway, RU=Russia, and TU=Turkey.

A great reduction in diversity has occurred from the past to the present (Figure 8), however, it is also relevant to attempt to calculate what the genetic diversity will be in the future. And thus, using the results from the stairway analysis in paper III, and the loss of heterozygosity equation from the introduction I find that all four populations from the stairway analysis, Greenland, Iceland, Norway, and Denmark, will most likely see quite a reduction in heterozygosity over the next hundred years (Table 2, under the expectation of constant

population size) unless conservation efforts are made to keep more of the genetic diversity. Only looking at the mean I find that the island populations may have as little as ~30% of the heterozygosity they have today left, and even the large Norwegian population could lose 34% of its heterozygosity (when using a mutation rate of 2.3e-9, Table 1). If the eagle has a lower mutation rate, I still predict a potentially quite substantial reduction in heterozygosity over the next 100 years.

Table 2. Fraction of heterozygosity remaining in a hundred years. Calculated for the four populations for which the effective population size back in time (stairway plot) was calculated in paper III. 2.3e-9 and 1e-9 refer to the mutation rate used to calculate the stairway plot in paper III. Numbers outside the parentheses are calculated from the mean effective population, in parentheses are calculated from the 2.5-97.5% effective population size. The stairway plot was only calculated from the contemporary samples, and thus GL=Greenland, IS=Iceland, NO=Norway, and DK=Denmark.

	Mut. rate 2.3e-9	Mut. rate 1e-9
GL	0.310 (0.011-0.856)	0.661 (0.011-0.937)
IS	0.310 (0.011-0.804)	0.572 (0.011-0.932)
NO	0.661 (0.011-0.942)	0.825 (0.310-0.972)
DK	0.572 (0.011-0.918)	0.815 (0.158-0.962)

5 Conclusions and future perspectives

In conclusion, we can to some extent resolve the genetic relationships for the northwest corner of the white-tailed eagle species range, both for the mitochondrial DNA and the nuclear genome (autosomes). The mitochondrial and nuclear genome, however, do not agree on the genetic structure, as two distinct genetic lineages found in and shared between Greenland, Iceland, and Norway seems to be upheld in the mtDNA (potentially by selection), whereas the autosomes show a structure that follows the expectation when considering geography, in which individuals are closest related to individuals from their own country of origin (**papers I, and III**).

In **paper II**, four methods of identifying the homogametic sex chromosomes in the small eagle population in Iceland were analysed and three methods were found to be successful. The best method was reference mapping, but also analysing the difference in depth performed well. It also became apparent, that it was not an easy task to identify the whole sex chromosome, and that PAR and gametologs can play a role and blur the picture.

Finally, we find in **paper III** that Greenland and Iceland have been highly affected by having small population sizes and being isolated for probably a long time. However, it is also apparent that all the analysed populations have very small and decreasing effective population sizes. Though the bottlenecks may have existed for a short period considering the relatively long generation time of the eagles, and thus it may not have had a severe effect, we do see clear signs of the bottleneck that happened in Iceland, displaying higher inbreeding, lower diversity, and increased drift.

5.1 Future research

Especially relevant for future research of the white-tailed eagle would be to analyse a larger number of samples from a wider range of the distribution. So far very few studies have looked at the full east/west range of the eagles, and none using whole genomes. This would help in the resolution of reconstructing the colonisation routes, and further analysis of the peculiar signal presented here for the mitochondrial genome. Wider use and analysis of historic samples seem very relevant, as all temporal analysis herein showed a great reduction in diversity, and it is unknown if this is a general trend for the white-tailed eagle. Further comparison of historic and contemporary samples is also warranted to identify potential adaptations or gene differentiation over time. This could also help in the identification of deleterious mutations on the eagles, or regions with extreme low diversity.

A different avenue of research would be to connect phenotypic and fitness-related traits, e.g., in connection to fecundity, to the genomics of the eagles, and the temporal change herein.

As the identification of homogametic sex chromosomal scaffolds, using bioinformatic approaches, did not identify the whole Z-chromosome, and there were both false positives and negatives, further optimisation of these methods are needed, and automation would make it more applicable in a time where we are not running low of sequencing data, but rather people to analyse it. Development of approaches to find the heterogametic sex chromosome too are also needed.

5.2 Prospects of the white-tailed eagle

All the analysed populations in this study display low diversity. The contemporary Icelandic display the lowest mean heterozygosity of just 0.00063, followed by the contemporary Greenland sample (0.00066), which means they have as low diversity as the vulnerable Andean condor and endangered eastern mountain gorilla. The contemporary samples from Denmark and Estonia shows an amount of heterozygosity (0.0009 and 0.0010, respectively) at the levels of chimpanzee and the giant panda, and lower than the critically endangered Californian condor (Abascal et al. 2016, Robinson et al. 2021). The contemporary samples analysed here displays some of the lowest diversities reported for birds (Chung et al. 2015, Figuet et al. 2016, Gelabert et al. 2020, Robinson et al. 2021). Further, populations with temporal samples show a great reduction in diversity and evidence of increased drift. It is also shown that the effective population sizes have been reduced over a long period in Greenland, Iceland, Norway, and Denmark. And the current day effective population size is way below both 500, and 50, and thus the populations cannot be expected to retain evolutionary adaptive potential and the chance of inbreeding depression seems very present. The estimated reduction in heterozygosity ranges from 32-57% over the last 70-130 years, which is an enormous reduction and far exceeds the suggested reduction of 5-10% before the population could risk extinction (Allendorf and Ryman 2002).

Further, I find that estimates of future heterozygosity all show an equally large or potentially larger reduction as seen to date, and thus also exceeds the maximum reduction suggested for viable populations by Allendorf and Ryman (2002). Thus, the effective population sizes seem to have little chance of rising, in the current state.

The idea of the populations being less than optimal is supported for Iceland when looking at its productivity and fledged brood size, which was 0.49 and 1.34, respectively, for the period 1980-2002 (Evans et al. 2009), which as well is well below the estimated 5-year average minimum for a good status of breeding performance of white-tailed eagles of >1 and >1.64, respectively (Helander et al. 2013).

If translocation is considered, also the risk of outbreeding depression needs to be considered and further assessed, especially in the isolated island populations which may harbour unique phenotypic traits such as larger size (Salomonsen 1979), and may not fulfil one of the suggested criteria to avoid outbreeding, i.e., having had gene flow in the last 500 years (Frankham et al. 2011). However, even so, considering the past and expected heterozygosity development it is hard to imagine that the populations would not benefit from further conservation efforts.

References

Abascal, F., A. Corvelo, F. Cruz, J. L. Villanueva-Cañas, A. Vlasova, M. Marcet-Houben, B. Martínez-Cruz, J. Y. Cheng, P. Prieto, V. Quesada, J. Quilez, et al. (2016). Extreme genomic erosion after recurrent demographic bottlenecks in the highly endangered Iberian lynx. Genome Biology 17:251.

Alexander, D. H., J. Novembre, and K. Lange (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Research 19:1655–1664.

Allendorf, F. W., and N. Ryman (2002). The role of genetics in population viability analysis. In Population viability analysis. University of Chicago Press, pp. 50–85.

Babraham Bioinformatics (2010). FastQC. [Online.] Available at http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Ballard, J. W. O., and M. C. Whitlock (2004). The incomplete natural history of mitochondria. Molecular Ecology 13:729–744.

Bandelt, H. J., P. Forster, and A. Rohl (1999). Median-joining networks for inferring intraspecific phylogenies. Molecular Biology and Evolution 16:37–48.

Barnosky, A. D., N. Matzke, S. Tomiya, G. O. U. Wogan, B. Swartz, T. B. Quental, C. Marshall, J. L. McGuire, E. L. Lindsey, K. C. Maguire, B. Mersey, and E. A. Ferrer (2011). Has the Earth's sixth mass extinction already arrived? Nature 471:51–57.

Bataillon, T., and M. Kirkpatrick (2000). Inbreeding depression due to mildly deleterious mutations in finite populations: Size does matter. Genetical Research 75:75–81.

Bijleveld, M. (1974). Birds of Prey in Europe. Macmillan Education UK, London.

Birdlife International (2020). Haliaeetus albicilla. The IUCN Red List of Threatened Species 2020 e.T2269513.

Boertmann, D., and C. Bay (2018). The Greenland red-list 2018.

Bouckaert, R., J. Heled, D. Kühnert, T. Vaughan, C. H. Wu, D. Xie, M. A. Suchard, A. Rambaut, and A. J. Drummond (2014). BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. PLoS Computational Biology 10:1–6.

Carøe, C. (2017). BEST protocol 2.0. 1–10.

Charlesworth, B. (2009). Fundamental concepts in genetics: Effective population size and patterns of molecular evolution and variation. Nature Reviews Genetics 10:195–205.

Charlesworth, B., M. T. Morgan, and D. Charlesworth (1993). The effect of deleterious mutations on neutral molecular variation. Genetics 134:1289–1303.

Chung, O., S. Jin, Y. S. Cho, J. Lim, H. Kim, S. Jho, H. M. Kim, J. H. Jun, H. J. Lee, A. Chon, J. Ko, et al. (2015). The first whole genome and transcriptome of the cinereous vulture reveals adaptation in the gastric and immune defense systems and possible

convergent evolution between the Old and New World vultures. Genome Biology 16:1–11.

Clark, P. U., and A. C. Mix (2002). Ice sheets and sea level of the Last Glacial Maximum. Quaternary Science Reviews 21:1–7.

Cox, M. A. A., and T. F. Cox (2008). Multidimensional Scaling. In Handbook of Data Visualization. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 315–347.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, and R. Durbin (2011). The variant call format and VCFtools. Bioinformatics 27:2156–2158.

Egea, R., S. Casillas, and A. Barbadilla (2008). Standard and generalized McDonald-Kreitman test: a website to detect selection by comparing different classes of DNA sites. Nucleic Acids Research 36:W157–W162.

Eggertsson, H. P., H. Jonsson, S. Kristmundsdottir, E. Hjartarson, B. Kehr, G. Masson, F. Zink, K. E. Hjorleifsson, A. Jonasdottir, A. Jonasdottir, I. Jonsdottir, et al. (2017). Graphtyper enables population-scale genotyping using pangenome graphs. Nature Genetics 49:1654–1660.

Eggertsson, H. P., S. Kristmundsdottir, D. Beyter, H. Jonsson, A. Skuladottir, M. T. Hardarson, D. F. Gudbjartsson, K. Stefansson, B. V. Halldorsson, and P. Melsted (2019). GraphTyper2 enables population-scale genotyping of structural variation using pangenome graphs. Nature Communications 10:1–8.

Elts, J., A. Leito, M. Leivits, L. Luigujõe, R. Nellis, M. Ots, I. Tammekänd, and Ü. Väli (2019). Eesti lindude staatus, pesitsusaegne ja talvine arvukus 2013-2017. Hirunda 32:1–29.

Evans, R. J., J. D. Wilson, A. Amar, A. Douse, A. MacLennan, N. Ratcliffe, and D. P. Whitfield (2009). Growth and demography of a re-introduced population of White-tailed Eagles Haliaeetus albicilla. Ibis 151:244–254.

Fay, J. C., and C. I. Wu (2000). Hitchhiking under positive Darwinian selection. Genetics 155:1405–1413.

Ferguson-Lees, J., and D. A. Christie (2001). Raptors of the world. Houghton Mifflin Company, New York.

Figuet, E., B. Nabholz, M. Bonneau, E. Mas Carrio, K. Nadachowska-Brzyska, H. Ellegren, and N. Galtier (2016). Life History Traits, Protein Evolution, and the Nearly Neutral Theory in Amniotes. Molecular Biology and Evolution 33:1517–1527.

Frankham, R. (1995a). Conservation Genetics. Annu. Rev. Genetics 29:305–27.

Frankham, R. (1995b). Effective population size/adult population size ratios in wildlife: a review. Genetics Research 66:95–107.

Frankham, R., J. D. Ballou, M. D. B. Eldridge, R. C. Lacy, K. Ralls, and M. R. Dudash (2011). Predicting the Probability of Outbreeding Depression. Conservation Biology 25:465–475.

Frankham, R., C. J. A. Bradshaw, and B. W. Brook (2014). Genetics in conservation management: Revised recommendations for the 50/500 rules, Red List criteria and population viability analyses. Biological Conservation 170:56–63.

Franklin, I. R. (1980). Evolutionary Changes in Small Populations. In Conservation Biology: An Evolutionary-Ecological Perspective (M. E. Soulé and B. M. Wilcox, Editors). Sinauer, Sunderland, pp. 135–149.

Garcia-Moreno, J., and D. P. Mindell (2000). Rooting a phylogeny with homologous genes on opposite sex chromosomes (gametologs): A case study using avian CHD. Molecular Biology and Evolution 17:1826–1832.

Garner, B. A., S. Hoban, and G. Luikart (2020). IUCN Red List and the value of integrating genetics. Conservation Genetics 21:795–801.

Gelabert, P., M. Sandoval-Velasco, A. Serres, M. de Manuel, P. Renom, A. Margaryan, J. Stiller, T. de-Dios, Q. Fang, S. Feng, S. Mañosa, et al. (2020). Evolutionary History, Genomic Adaptation to Toxic Diet, and Extinction of the Carolina Parakeet. Current Biology 30:108-114.e5.

Gopalakrishnan, S., J. A. Samaniego Castruita, M. H. S. Sinding, L. F. K. Kuderna, J. Räikkönen, B. Petersen, T. Sicheritz-Ponten, G. Larson, L. Orlando, T. Marques-Bonet, A. J. Hansen, et al. (2017). The wolf reference genome sequence (Canis lupus lupus) and its implications for Canis spp. population genomics. BMC Genomics 18:1–11.

Goudet, J. (2005). HIERFSTAT, a package for R to compute and test hierarchical F -statistics. Molecular Ecology Notes 5:184–186.

Hailer, F. (2006). Conservation Genetics of the White-Tailed Eagle.

Hailer, F., B. Helander, A. O. Folkestad, S. A. Ganusevich, S. Garstad, P. Hauff, C. Koren, V. B. Masterov, T. Nygård, J. A. Rudnick, S. Shiraki, et al. (2007). Phylogeography of the white-tailed eagle, a generalist with large dispersal capacity. Journal of Biogeography 34:1193–1206.

Hailer, F., B. Helander, A. O. Folkestad, S. A. Ganusevich, S. Garstad, P. Hauff, C. Koren, T. Nygard, V. Volke, C. Vila, and H. Ellegren (2006). Bottlenecked but long-lived: high genetic diversity retained in white-tailed eagles upon recovery from population decline. Biology Letters 2:316–319.

Hansen, K. (1979). Status over bestanden af Havørn *Haliaeetus albicilla groenlandicus* Brehm i Grønland i årene 1972-74. Dansk Ornitologisk Forenings Tidsskrift 73:107–130.

Harris, R. S. (2007). Improved pairwise alignmen of genomic DNA. Ph.D. Thesis. [Online.] Available at http://scholar.google.com/scholar?q=related:J_7kcJuUwCoJ:scholar.google.com/&hl

=en&num=20&as_sdt=0,5%5Cnhttp://www.bx.psu.edu/~rsharris/rsharris_phd_thesis_2007.pdf.

Hartl, D. L., and A. G. Clark (2007). Principles of Population Genetics. 4th edition. Sinauer Associates Inc. Publishers.

Hasselgren, M., A. Angerbjörn, N. E. Eide, R. Erlandsson, Ø. Flagstad, A. Landa, J. Wallén, and K. Norén (2018). Genetic rescue in an inbred Arctic fox (Vulpes lagopus) population. Proceedings of the Royal Society B: Biological Sciences 285.

Hedrick, P. W., and A. Garcia-Dorado (2016). Understanding Inbreeding Depression, Purging, and Genetic Rescue. Trends in Ecology & Evolution 31:940–952.

Helander, B. B., M. Olsson, L. Reutergårdh, and L. Reutergardh (1982). Residue levels of organochlorine and mercury compounds in unhatched eggs and the relationships to breeding success in white-tailed sea eagles Haliaeetus albicilla in Sweden. Ecography 5:349–366.

Helander, B., A. Bignert, C. Herrmann, and T. Stjernberg (2013). White-tail eagle productivity. HELCOM Core Indicator Report.

Helander, B., A. Olsson, A. Bignert, L. Asplund, and K. Litzén (2002). The Role of DDE, PCB, Coplanar PCB and Eggshell Parameters for Reproduction in the White-tailed Sea Eagle (Haliaeetus albicilla) in Sweden. AMBIO: A Journal of the Human Environment 31:386–403.

Hill, W. G., and A. Robertson (1966). The effect of linkage on limits to artificial selection. Genetical Research 8:269–294.

Hoban, S., M. Bruford, J. D. Urban, M. Lopes-fernandes, M. Heuertz, P. A. Hohenlohe, I. Paz-vinas, P. Sjögren-gulve, G. Segelbacher, C. Vernesi, S. Aitken, et al. (2020). Genetic diversity targets and indicators in the CBD post-2020 Global Biodiversity Framework must be improved. Biological Conservation 248:108654.

Hoban, S., I. Paz-Vinas, S. Aitken, L. D. Bertola, M. F. Breed, M. W. Bruford, W. C. Funk, C. E. Grueber, M. Heuertz, P. Hohenlohe, M. E. Hunter, et al. (2021). Effective population size remains a suitable, pragmatic indicator of genetic diversity for all species, including forest trees. Biological Conservation 253:2020–2022.

Holderegger, R., N. Balkenhol, J. Bolliger, J. O. Engler, F. Gugerli, A. Hochkirch, C. Nowak, G. Segelbacher, A. Widmer, and F. E. Zachos (2019). Conservation genetics: Linking science with practice. Molecular Ecology 28:3848–3856.

Honnen, A.-C. C., F. Hailer, N. Kenntner, I. Literák, L. Dubská, and F. E. ZACHOS (2010). Mitochondrial DNA and nuclear microsatellites reveal high diversity and genetic structure in an avian top predator, the white-tailed sea eagle, in central Europe. Biological Journal of the Linnean Society 99:727–737.

Jais, M. (2020). White-tailed Eagle. [Online.] Available at http://europeanraptors.org/white-tailed-eagle/.

James, J. W. (1970). The founder effect and response to artificial selection. Genetical Research 16:241–250.

Johnson, J. A., R. E. Tingay, M. Culver, F. Hailer, M. L. Clarke, and D. P. Mindell (2009). Long-term survival despite low genetic diversity in the critically endangered Madagascar fish-eagle. Molecular Ecology 18:54–63.

Kim, J. A., S. G. Kang, H. S. Jeon, J. H. Jeon, J. H. Jang, S. Kim, and J. An (2019). Complete mitogenomes of two Accipitridae, Haliaeetus albicilla, and Pernis ptilorhynchus. Mitochondrial DNA Part B: Resources 4:392–393.

Kitowski, I., D. Jakubas, D. Wiącek, and A. Sujak (2017). Concentrations of lead and other elements in the liver of the white-tailed eagle (Haliaeetus albicilla), a European flagship species, wintering in Eastern Poland. Ambio:1–17.

Knaus, B. J., and N. J. Grünwald (2016). vcfr: A package to manipulate and visualize variant call format data in r. Molecular Ecology Resources:44–53.

Korneliussen, T. S., A. Albrechtsen, and R. Nielsen (2014). ANGSD: Analysis of next generation sequencing data. BMC bioinformatics 15:356.

Kremer, M. (1993). Population Growth and Technological Change: One Million B.C. to 1990. The Quarterly Journal of Economics 108:681–716.

Krüger, O., T. Grünkorn, and B. Struwe-Juhl (2010). The return of the white-tailed eagle (Haliaeetus albicilla) to northern Germany: Modelling the past to predict the future. Biological Conservation 143:710–721.

Langguth, T., A.-C. C. Honnen, F. Hailer, T. Mizera, S. Skoric, Ü. Väli, and F. E. Zachos (2013). Genetic structure and phylogeography of a European flagship species, the white-tailed sea eagle Haliaeetus albicilla. Journal of Avian Biology 44:263–271.

Li, H. (2011a). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27:2987–2993.

Li, H. (2011b). Improving SNP discovery by base alignment quality. Bioinformatics 27:1157–1158.

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

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079.

Literák, I., V. Mrlík, A. Hovorková, P. Mikulíček, J. Lengyel, K. Št'astný, J. Cepák, and L. Dubská (2007). Origin and genetic structure of white-tailed sea eagles (Haliaeetus albicilla) in the Czech Republic: an analysis of breeding distribution, ringing data and DNA microsatellites. European Journal of Wildlife Research 53:195–203.

Liu, X., and Y. X. Fu (2020). Stairway Plot 2: demographic history inference with folded SNP frequency spectra. Genome Biology 21:1–9.

Liu, X., and Y.-X. Fu (2015). Exploring population size changes using SNP frequency spectra. Nature genetics 47:555–9.

Lõhmus, A. (1998). Eesti kotkaste argipäev. Eesti Loodus 5/6:210–213.

Love, J. A. A., and M. E. E. Ball (1979). White-tailed sea eagle Haliaeetus albicilla reintroduction to the Isle of Rhum, Scotland, 1975-1977. Biological Conservation 16:23–30.

Macherey-Nagel (2014). Genomic DNA from Tissue, User manual, NucleoMag® 96 Tissue. [Online.] Available at https://www.takarabio.com/assets/documents/User Manual/NucleoMag 96 Tissue Genomic DNA Purification User Manual (PT5008-1)_Rev_03.pdf.

Mangs, A. H., and B. Morris (2007). The Human Pseudoautosomal Region (PAR): Origin, Function and Future. Current Genomics 8:129–136.

McQuillan, R., A. L. Leutenegger, R. Abdel-Rahman, C. S. Franklin, M. Pericic, L. Barac-Lauc, N. Smolej-Narancic, B. Janicijevic, O. Polasek, A. Tenesa, A. K. MacLeod, et al. (2008). Runs of Homozygosity in European Populations. American Journal of Human Genetics 83:359–372.

Meisner, J., and A. Albrechtsen (2018). Inferring Population Structure and Admixture Proportions in Low-Depth NGS Data. Genetics 210:719–731.

Mindell, D. P., J. J. Fuchs, and J. A. Johnson (2018). Phylogeny, Taxonomy, and Geographic Diversity of Diurnal Raptors: Falconiformes, Accipitriformes, and Cathartiformes. In Birds of Prey: Biology and conservation in the XXI century (J. H. Sarasola, J. M. Grande and J. J. Negro, Editors). Springer International Publishing, Cham, pp. 1–522.

Nei, M. (1977). F-statistics and analysis of gene diversity in subdivided populations. Ann. Hum. Gent. 41:225–233.

Nemesházi, E., S. Kövér, F. E. Zachos, Z. Horváth, G. Tihanyi, A. Mórocz, T. Mikuska, I. Hám, I. Literák, S. Ponnikas, T. Mizera, and K. Szabó (2016). Natural and anthropogenic influences on the population structure of white-tailed eagles in the Carpathian Basin and central Europe. Journal of Avian Biology 47:795–805.

Nielsen, R., T. Korneliussen, A. Albrechtsen, Y. Li, and J. Wang (2012). SNP calling, genotype calling, and sample allele frequency estimation from new-generation sequencing data. PLoS ONE 7(7):e37558.

Nielsen, R., and M. Slatkin (2013). An Introduction to Population Genetics: Theory and Applications. Sinauer Associates, Inc. .

O'Brien, S. J., W. E. Johnson, C. A. Driscoll, P. Dobrynin, and L. Marker (2017). Conservation Genetics of the Cheetah: Lessons Learned and New Opportunities. Journal of Heredity 108:671–677.

Pamilo, P., and S. Pálsson (1998). Associative overdominance, heterozygosity and fitness. Heredity 81:381–389.

Paradis, E. (2010). pegas: an R package for population genetics with an integrated-modular approach. Bioinformatics 26:419–420.

Patterson, N., A. L. Price, and D. Reich (2006). Population Structure and Eigenanalysis. PLoS Genetics 2:e190.

Petersen, Æ. (1998). Haförn. In Íslenskir Fuglar (S. Harðardóttir, Editor). Vaka-Helgafell hf., Reykjavik, pp. 120–123.

Ponnikas, S., L. Kvist, T. Ollila, T. Stjernberg, and M. Orell (2013). Genetic structure of an endangered raptor at individual and population levels. Conservation Genetics 14:1135–1147.

Price, A. L., N. J. Patterson, R. M. Plenge, M. E. Weinblatt, N. A. Shadick, and D. Reich (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nature Genetics 38:904–909.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, and P. C. Sham (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. The American Journal of Human Genetics 81:559–575.

Qiagen (2006). DNeasy ® Blood & Tissue Handbook. [Online.] Available at http://www.qiagen.com/products/genomicdnastabilizationpurification/dneasytissuesystem/dneasybloodtissuekit.aspx#Tabs=t0.

Quinlan, A. R., and I. M. Hall (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842.

Ralls, K., J. D. Ballou, M. R. Dudash, M. D. B. Eldridge, C. B. Fenster, R. C. Lacy, P. Sunnucks, and R. Frankham (2018). Call for a Paradigm Shift in the Genetic Management of Fragmented Populations. Conservation Letters 11:1–6.

Ralls, K., R. Frankham, and J. D. Ballou (2014). Inbreeding and Outbreeding ☆. Reference Module in Life Sciences:1–10.

Randla, T., and A. Õun (1980). Kaljukotkas ja merikotkas Eestis 1970-ndail aastail. Eesti Loodus 4:512–515.

Ripley, B. D. (1996). Pattern Recognition and Neural Networks. Cambridge University Press.

Ritchie, E. G., B. Elmhagen, A. S. Glen, M. Letnic, G. Ludwig, and R. A. McDonald (2012). Ecosystem restoration with teeth: What role for predators? Trends in Ecology and Evolution 27:265–271.

Robinson, J. A., R. C. K. Bowie, O. Dudchenko, E. L. Aiden, S. L. Hendrickson, C. C. Steiner, O. A. Ryder, D. P. Mindell, and J. D. Wall (2021). Genome-wide diversity in

the California condor tracks its prehistoric abundance and decline. Current Biology 31:2939-2946.e5.

Salo, P., M. Nordström, R. L. Thomson, and E. Korpimäki (2008). Risk induced by a native top predator reduces alien mink movements. Journal of Animal Ecology 77:1092–1098.

Salomonsen, F. (1979). Ornithological and ecological studies in SW Greenland. Meddelelser om Grønland 204.

Schubert, M., S. Lindgreen, and L. Orlando (2016). AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Research Notes 9:88.

Shaw, K. L. (2002). Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: What mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. Proceedings of the National Academy of Sciences of the United States of America 99:16122–16127.

Shriner, D., F. Tekola-Ayele, A. Adeyemo, and C. N. Rotimi (2014). Genome-wide genotype and sequence-based reconstruction of the 140,000 year history of modern human ancestry. Scientific Reports 4:6055.

Simone, D., F. M. Calabrese, M. Lang, G. Gasparre, and M. Attimonelli (2011). The reference human nuclear mitochondrial sequences compilation validated and implemented on the UCSC genome browser. BMC Genomics 12.

Skarphéðinsson, K. H. (2013). Haförninn. Fuglavernd (Fuglaverndarfélag íslands), Reykjavik.

Skelmose, K., and Ole. F. Larsen (2021). Projekt Ørn – Årsrapport 2020, DOF BirdLife Danmark.

Smeds, L., A. Qvarnström, and H. Ellegren (2016). Direct estimate of the rate of germline mutation in a bird. Genome Research 26:1211–1218.

Smith, R. L., and T. M. Smith (2003). Elements of Ecology. Fifth edit. Benjamin Cummings.

Soulé, M. E. (1980). Thresholds for Survival: Maintaining Fitness and Evolutionary Potential. In Conservation Biology: An Evolutionary-Ecological Perspective (M. E. Soulé and B. M. Wilcox, Editors). Sinauer, Sunderland.

Sugimoto, T., and S. Shiraki (2014). Origin of a White-tailed Eagle Wintering on Tsushima Island, Japan, Established by Mitochondrial and Nuclear DNA Analyses. Ornithological Science 13:117–120.

Tajima, F. (1989). Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. Genetics 123:585–595.

Thermo Fisher (2016). Thermo Scientific GeneJET Genomic DNA Purification Kit #K0721, #K0722. [Online.] Available at https://www.thermofisher.com/document-connect/document-connect.html?url=https%3A%2F%2Fassets.thermofisher.com%2FTFS-

Assets%2FLSG%2Fmanuals%2FMAN0012667_GeneJET_Whole_Blood_Genomic_DNA_Purification_Mini_Kit_UG.pdf&title=VXNlciBHdWlkZTogR2VuZUpFVCBXaG9sZSBC.

Walker, L. A., S. J. Beith, A. J. Lawlor, C. Moeckel, M. G. Peréira, E. D. Potter, and R. F. Shore (2009). Persistent Organic Pollutants (POPs) and inorganic elements in predatory bird livers and eggs 2007 to 2009: a Predatory Bird Monitoring Scheme (PBMS) Report.

Wilson Sayres, M. A. (2018). Genetic diversity on the sex chromosomes. Genome Biology and Evolution 10:1064–1078.

Wright, S. (1931). Evolution in mendelian populations. Genetics 16:97–159.

Wright, S. (1933). Inbreeding and Homozygosis. Proceedings of the National Academy of Sciences 19:411–420.

Wright, S. (1939). Statistical Genetics in Relation to Evolution. Herman & Cie, Paris, France.

Zeng, K., Y.-X. Fu, S. Shi, and C.-I. Wu (2006). Statistical Tests for Detecting Positive Selection by Utilizing High-Frequency Variants. Genetics 174:1431–1439.

Zheng, X., D. Levine, J. Shen, S. M. Gogarten, C. Laurie, and B. S. Weir (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics 28:3326–3328.

Paper I

Paper I

Distinctive mitogenomic lineages within populations of white-tailed eagles (*Haliaeetus albicilla*)

Accepted with minor review, Ornithology

Author contribution: Conceived the idea: CCRH and SP; Performed the experiments: CCRH, SB, SMG, JAR, JACB, MDM and SP; Analysed the data: CCRH, SB (BEAST and dating analysis), and SP; wrote the paper: CCRH, SB, MDM and SP; Commented on the paper: SMG, JAR, JACB, GTH, RAS, MvS, KHS, ALL, ML, CS, RD, KS, DB and IE; Contributed substantial materials, resources or funding: CCRH, GTH, RAS, MvS, KHS, ALL, ML, CS, RD, KS, DB, IE, MDM and SP.

Distinctive mitogenomic lineages within populations of white-tailed eagles (*Haliaeetus albicilla*)

Charles Christian Riis Hansen[1*], Sina Baleka[1], Sólveig Magnea Guðjónsdóttir[1], Jacob Agerbo Rasmussen[2,3], Jesus Adrian Chimal Ballesteros[2,4], Gunnar Thor Hallgrimsson[1], Robert A. Stefansson[5], Menja von Schmalensee[5], Kristinn Haukur Skarphédinsson[6], Aili Lage Labansen[7], Madis Leivits[8], Kim Skelmose[9], Christian Sonne[10], Rune Dietz[10], David Boetmann[10], Igor Eulaers[10], Michael D. Martin[2], and Snæbjörn Pálsson[1]

[1]Faculty of Life and Environmental Sciences, University of Iceland, Reykjavik, Iceland

[2]Department of Natural History, NTNU University Museum, Norwegian University of Science and Technology (NTNU), Trondheim, Norway

[3]Section for EvoGenomics, The Globe Institute, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

[4]Natural History Museum, University of Oslo, Oslo, Norway

[5]West-Iceland Nature Research Centre, 340 Stykkisholmur, Iceland

[6]Icelandic Institute of Natural History, Iceland

[7]Greenland Institute of Natural Resources, 3900 Nuuk, Greenland

[8]Chair of Clinical Veterinary Medicine, Institute of Veterinary Medicine and Animal Sciences, Estonian University of Life Sciences, 51006, Tartu, Estonia

[9]Natural History Museum of Denmark, University of Copenhagen, 1350 Copenhagen, Denmark.

[10]Department of Bioscience, Aarhus University, 4000 Roskilde, Denmark

Corresponding author: CCRH, ccr3@hi.is. Sæmundargata 18, 211. 102 Reykjavik.

Acknowledgments

We thank BirdLife Denmark (DOF BirdLife) for letting us use the samples they have gathered and also the Natural History Museum of Denmark for providing samples.

Funding statement: The study was supported by Research Grant nr 185280-052 from The Icelandic Research Council, the Doctoral student fund of the University of Iceland and The University of Iceland Research fund, and an NTNU Onsager Fellowship award to MDM. For the Norwegian samples; NovaSeq 6000 sequencing services were provided by the Norwegian Sequencing Centre, a national technology platform hosted by the University of Oslo and supported by the "Functional Genomics" and "Infrastructure" programs of the Research Council of Norway and the Southeastern Regional Health Authorities. Illumina HiSeq 4000 sequencing was performed by the NTNU Genomics Core Facility (GCF), which is funded by the Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology (NTNU), and the Central Norway Regional Health Authority. Some analyses were performed on resources provided by the National Infrastructure for High Performance Computing and Data Storage in Norway (UNINETT Sigma2).

Ethics statement: This research was conducted in compliance with the Guidelines to the Use of Wild Birds in Research.

Author contributions: Conceived the idea: CCRH and SP; Performed the experiments: CCRH, SB, SMG, JAR, JACB, MDM and SP; Analysed the data: CCRH, SB, SMG and SP; wrote the paper: CCRH, SB, MDM and SP; Corrected the paper: SMG, JAR, JACB, GTH, RAS, MvS, KHS, ALL, ML, CS, RD, KS, DB and IE; Contributed substantial materials, resources or funding: CCRH, GTH, RAS, MvS, KHS, ALL, ML, CS, RD, KS, DB, IE, MDM and SP.

Data Depository: The DNA-sequences mapped to the Korean specimen has been submitted to DRYAD and will be released when the manuscript is published at https://doi.org/10.5061/dryad.stqjq2c32.

Abstract

Using whole mitochondrial DNA sequences from 89 white-tailed eagles (*Haliaeetus albicilla*) sampled from Iceland, Greenland, Norway, Denmark and Estonia between 1990-2018, we investigate the mitogenomic variation within and between countries. We show that there is a substantial population differentiation between the countries, reflecting similar major phylogeographic patterns obtained previously for the control region of the mitochondria, which suggested two main refugia during the last glacial period. Distinct mitogenomic lineages are observed within countries with divergence times exceeding the end of the last glacial period of the Ice Age. Deviations from neutrality indicate that these lineages have been maintained by natural selection and there is an excess of segregating amino acids in comparison with number of fixations suggesting a large load of deleterious mutations. The maintenance of the distinct mitogenic lineages within countries inflates our estimates of divergence times.

Keywords: Phylogeography, selection, divergence time, conservation

Lay summary

- Whole mitochondrial genomes were used to examine the population genetics of white-tailed eagles
- Large genetic differences potentially upheld by selection within populations were identified
- Potential time of population splits during the last Ice age were identified

Introduction

The white-tailed eagle (*Haliaeetus albicilla*) is a top predator distributed over most of northern Eurasia, from Greenland and Iceland in the west to Japan in the east, and as far south as North Africa (BirdLife International 2016). Phylogeographic studies based on sequences of the mitochondrial control region and microsatellites have described a major split between European and Asian populations, with a large contact zone (Hailer et al. 2007, Honnen et al. 2010, Langguth et al. 2013). Two mitochondrial haplotypes have been observed in the control region in samples from Greenland and Iceland, which are only separated by a single mutation, one which is unique to the two islands and the other which is common in northern Europe, where more general variation was also observed (Hailer et al. 2007). The low diversity in the island populations may reflect both a possible bottleneck during colonization of these islands and small population sizes. The number of white-tailed eagles plummeted in Iceland in the late 19th century as in many other countries in Europe (Langguth et al. 2013, Treinys et al. 2016). The population in Iceland experienced a reduction from ~150 pairs in the mid-nineteenth century to ~20 pairs or less in the early 20th century, and despite a conservation law passed in 1915, the population did not increase in numbers until after 1967, following a ban on fox poisoning in 1964 (Petersen 1998, Skarphéðinsson 2013). As with other European populations, the Icelandic population has increased in number to about 80 pairs (Skarphéðinsson 2013). The Greenlandic population is larger than the Icelandic population, with 150-200 breeding pairs (Greenland Institute of Natural Resources 2020), however the Greenlandic population also experienced a reduction in the 20th century and was estimated to 50-75 pairs around 1950 (Hansen 1979). The mainland population in Europe consists of approximately 17,900-24,500 pairs (BirdLife International 2016), with the Norwegian population being the largest of all European countries, counting about 2,000 breeding pairs (Jais 2020). With the population in mainland Europe expanding after successful conservation programs, white-tailed eagles recolonized Denmark in 1995 after being extinct for more than 50

years, and in 2016, 61 breeding pairs were recorded (Kongeaastien 2020). In Estonia the population decreased from 400-500 breeding pairs (Lõhmus 1998) to about 20 pairs in the late 19th century, having years without successful breeding (Randla and Õun 1980). Currently, the Estonian white-tailed eagle breeding population is estimated to be 290-330 pairs with a strong positive trend (Elts et al. 2019)

White-tailed eagles are to a large extent sedentary and display philopatry (Hailer et al. 2007) even though they have been shown to migrate for more than 2,000 kilometers outside the breeding season (Ueta et al. 1998, Bragin et al. 2018). Between mainland Europe, Iceland, and Greenland there are no documented migrants. The mitochondrial DNA (mtDNA) diversity and an earlier classification of white-tailed eagles in Greenland (and potentially Iceland) as a subspecies, primarily due to their large size (Salomonsen 1979) suggests that the two island populations have been isolated for centuries. Comparison among the mtDNA sequences can be used to date the colonization of the island populations and to determine their divergence. However, such time estimates and other predictions including effective population size based on the neutral theory of molecular variation can be affected by natural selection (Zink and Barrowclough 2008). Furthermore, application of a molecular clock rate obtained from comparison between species may also lead to an overestimate of the divergence time within species (Ho et al. 2007).

The small Icelandic and potentially other white-tailed eagle populations are predicted to have little variation in the haploid, maternally inherited mtDNA, and this variation can be further reduced by directional selection either due to background selection (Charlesworth et al. 1993) or genetic hitchhiking (Smith and Haigh 1974) due to lack of recombination. In bird populations, it is expected that such linked effect is augmented due to linkage of mtDNA with the W-chromosome through shared maternal transmission, thus selection on the W-chromosome will also affect the mitochondrial DNA. Accordingly, lower variation has been observed in avian mtDNA than in mammalian species, which do not have this linkage with the heterogametic sex chromosome

(Berlin et al. 2007). In birds, however, Berlin et al. (2007) also observed a higher neutrality index (NI) than in mammals. The neutrality index (NI) quantifies the direction of the selection, a positive NI index indicates a negative selection or increased mutational load, where more deleterious mutations are segregating within-species than would be expected by the number of fixed differences between species, possibly due to reduced efficacy of selection in small populations (Lynch 1996, Neiman and Taylor 2009).

Here we investigate the genetic variation using complete mtDNA genomes of white-tailed eagles from Iceland and Greenland in comparison with the much larger mainland population in Norway, Estonia and the recently established population in Denmark. We evaluate signals of selection, the impact of different population sizes and recent bottlenecks, and investigate the split between the island populations of Iceland and Greenland from the mainland populations with the aim to reconstruct the population history and origin of these populations of white-tailed eagles based on mtDNA lineages.

Material and methods

Samples and sequencing

Samples (blood or muscle) were obtained from 89 individuals from five countries; Greenland (n=12; 1990-2017), Iceland (n=42; 2003-2011), Norway (n=21; 2001-2015), Denmark (n=11; 2015-2018) and Estonia (n=3; 2015), sampled during 1990-2018. An overview of sampling sites and year of sampling for all individuals studied is presented in Figure 1 (and in more details in Table S1). Blood samples from Iceland were collected from 42 chicks from Northwest Iceland, from 27 nest territories (ranging in latitude from N 64.38 to 65.87 degrees and in longitude from E 23.51 to 21.33), in an ongoing monitoring of the white-tailed eagles in Iceland (led by the Icelandic Institute of Natural History). Three to ten mL of blood were extracted from each chick. The blood was stored in EDTA buffer at -20 °C until DNA extraction. Whole blood and muscle tissue samples

149 from the Estonian, Danish and Greenland individuals were stored at -20 °C until DNA extraction,
150 and were provided by the Department of Bioscience, Arctic Research Centre, AU, Roskilde,
151 Denmark (Estonian, Danish and Greenland samples), Natural History Museum of Denmark,
152 University of Copenhagen, Copenhagen, Denmark (Danish samples) and the Greenland Institute
153 of Natural Resources, Nuuk, Greenland (Greenland samples).

154 DNA from the 68 individuals from Iceland, Greenland, Denmark and Estonia was extracted using
155 Thermo Fisher GeneJET Whole Blood Genomics DNA Purification Mini Kit following the standard
156 protocol (Thermo Fisher 2016) for blood samples and Macherey-Nagel NucleoMag® 96 Tissue kit
157 (Macherey-Nagel 2014) for tissue, at the University of Iceland. The concentration of the DNA was
158 > 20 ng μL^{-1} estimated using the NanoDrop 1000 (Thermo Fisher). The 68 samples were
159 sequenced at BGI Hong Kong using DNBseq Normal DNA library construction and DNBseq PE150.

160 DNA from the 21 Norwegian individuals was processed at the NTNU University Museum's
161 standard molecular genetics laboratories, and extractions were performed with a Qiagen DNeasy
162 Blood & Tissue kit (Qiagen 2006). The manufacturer's protocol was followed except that the
163 amount of proteinase K used in the lysis step was doubled, and the lysis step incubation at 56°C
164 was extended to 15 hours. The DNA solutions were incubated at 37°C for 10 minutes prior to
165 elution. A blank control extraction was performed alongside all tissue extractions. The DNA
166 extracts were sheared to a mean fragment size of 350 bp using a Covaris LE220 focused
167 ultrasonicator instrument, and then short fragments were removed via size-selection using SPRI
168 beads prepared as in Rohland & Reich (2012). Blunt-end Illumina library preparation was
169 conducted on the 21 Norwegian DNA extractions using the Blunt-End-Sigle-Tube (BEST) protocol
170 (Carøe et al. 2018), during which customized blunt-end adapters (Kircher et al. 2012) were ligated
171 to the genomic DNA fragments. Sample-specific, dual-indexing barcodes were incorporated into
172 each library using custom, indexed primers during the indexing PCR. Library indexing and
173 amplification was carried out in 100 µL PCR reactions with 7 µL library template, 0.25 mM each

dNTP, 0.25 µM forward primer, 0.25 µM reverse primer, 1 µL Herculase II Fusion DNA polymerase, 20 µl 5X Herculase II Reaction Buffer, and 65.8 µL molecular biology H_2O. For each library, an optimal number of indexing PCR cycles was determined using quantitative PCR. The indexing PCRs were performed under the following conditions: 95°C for 3 min, 13-21 cycles of 95°C for 20 sec, 60°C for 20 sec, 72°C for 40 sec, with a final extension of 72°C for 5 min. The amplified libraries were purified using SPRI beads (Rohland and Reich 2012) prior to elution into EB buffer (Qiagen). Purified, indexed libraries were pooled and sequenced over two runs on the Illumina HiSeq 4000 platform at the NTNU Genomics Core Facility and over one run on the Illumina NovaSeq 6000 platform at the University of Oslo Norwegian National Sequencing Centre.

Mapping

Quality of the 89 individual fastq files was checked with FastQC (Babraham Bioinformatics 2010), and then run through AdapterRemoval v2 (Schubert et al. 2016) to remove potential adapters when appropriate, using standard setting with the following customizing arguments: --collapse and --trimns. Individuals were mapped to the white-tailed eagle mitochondrial genome of a Korean specimen (NCBI: NC_040858.1) (Kim et al. 2019) with bwa mem and converted to bam files using samtools view and sort (Li and Durbin 2009, Li et al. 2009). Bam files were transformed to vcf and fasta files using samtools mpileup with quality filters -q 30 and -Q 20, and bcftools call (Li et al. 2009). The Korean reference sequence includes a nuclear mitochondrial DNA segment (NUMT) in the 12s rRNA from position 492-538 (both included) which was removed from the alignment, resulting in total sequence length of 17,461 bp. Quality, depth, ratio of the common allele and binomial probabilities for all segregating sites were evaluated from the vcf file. A recently published mitochondrial genome of a German specimen (NCBI: MN356434.1) was retrieved from genebank and compared to our sequences (Feng et al. 2020) but it was not included in subsequent analyses unless specifically stated.

Data Analysis

Unless otherwise stated, R and Rstudio were used for all calculations (version 3.6.3 and 1.2.5033, respectively) (RStudio Team 2019, R core Team 2020). Due to the small sample size from Estonia, it was omitted from the statistical test described below assessing differences in variation within countries. Molecular diversity per country was summarized based on: number of haplotypes, segregating sites, haplotype diversity (Nei and Tajima 1981), nucleotide diversity (π) (Nei 1987) using the pegas package (Paradis 2010) and haplotype richness (to account for difference in sample sizes) using the hierfstat package (Goudet 2005) which uses the rarefaction function, with standard settings and the smallest sample size of 11 (Paradis 2010). Differences in genetic diversity among the countries were summarized with three estimators of the population parameter Theta ($\theta = N_e u$), which may vary in sensitivity to detect population bottlenecks and selection, using pegas (Paradis 2010). The estimators are: Theta Pi (θ_π), based on nucleotide diversity (Tajima 1996), Theta K (θ_K), using the expected number of alleles (Ewens 1972), and Theta S (θ_S), based on the number of segregating sites (Watterson 1975, Tajima 1989).

To summarize the relations between the sequences and their geographic origin, a median-joining network was drawn for all individuals using popART with standard settings (Bandelt et al. 1999, Leigh and Bryant 2015). A maximum likelihood tree was calculated using PhyML v. 3.3.3 (Guindon and Gascuel 2003), including red kite (*Milvus milvus*) as outgroup using the GTR+I+G model, based on the likelihood estimates from jModelTest v. 2.1.10 (Darriba et al. 2012). As the maximum-likelihood approach was not able to completely resolve the topology a Bayesian approach was conducted using BEAST v. 2.6.2 (Bouckaert et al. 2014). The dataset was partitioned into 17 sets (one for each gene and rRNA, one containing all tRNAs, and one containing non-coding sequences) and bModelTest v. 1.2.1 (Bouckaert and Drummond 2017) to average over substitution models applying the default transition-transversion split option, and uncorrelated lognormal relaxed clock and constant populations size. However, this analysis did not converge for several partitions mainly due to the tree likelihood. To investigate this, we reduced the

226 complexity of our analysis and ran BEAST again, this time on the unpartitioned dataset, applying a
227 strict clock and a HKY substitution model with estimated frequencies and constant population
228 size. We ran the analysis 10 times, each time for 20 million generations, and visualized the output
229 using FigTree v. 1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/). Convergence was visually
230 inspected using Tracer v. 1.7.1 (Rambaut et al. 2018) with all effective sample size (ESS) values
231 >200. This resulted in the recovery of six possible topologies (Figure S3).

232 Estimation of divergence times were again conducted in BEAST v. 2.6.2. All analyses were run
233 using the Bayesian skyline model (Drummond et al. 2005), bModelTest v. 1.2.1 (Bouckaert and
234 Drummond 2017) applying the default transition-transversion split option, and a strict clock in
235 accordance with the low diversity of our intraspecies data set. The most frequent tree from our
236 initial Bayesian analysis (Figure S3A), described above, was used as a starting tree, fixing the
237 topology throughout the runs. The MCMC chain was run each time for 20 Mio generations and
238 the first 10% of each run were discarded as burn-in. Again, convergence was visually inspected in
239 Tracer v. 1.7.1 (Rambaut et al. 2018) with all ESS values >200.

240 Since no fossil calibrations are available and substitution rates in birds are subject to discussion,
241 four different approaches (giving 6 scenarios) were used to get a set of differently calibrated
242 analyses (also see Table S2). Firstly (Scenario 1), we used the widely adopted standard rate of 2.1
243 % divergence in birds for Cytochrome B (CYTB) (Shields and Wilson 1987, Paxinos et al. 2002, Weir
244 and Schluter 2008). The rate was applied for CYTB and for the rest of the mtDNA it was estimated
245 for 17 partitions: one for each gene and one for each rRNA, as well as one each for all tRNAs and
246 one for all non-coding regions. The rate for each partition was inferred by its relative nucleotide
247 diversity, calculated in pegas (Paradis 2010) in comparison to CYTB. Secondly, we used body mass
248 related rate (Nabholz et al. 2016), assuming a body mass of 5 kg (Helander et al. 2007, Isomursu
249 et al. 2018). The analysis was run using 3rd codon positions only and fossil calibration sets 2
250 (Scenario 2; see Table S2 for an overview over all six scenarios) and 4 (Scenario 3), respectively

(Nabholz et al. 2016). Nabholz et al. (2016) use four different sets of fossil calibrations to derive a body mass related substitution rate for birds. They abandon two calibration sets (1 & 3) since known divergence dates within the avian phylogeny could not be estimated correctly when using these calibration sets. Calibrations sets 2 and 4 both contain the following fossil calibrations: split between Neognathae and Paleognathae (86.5 – 66 MYR); Anseriformes – Galliformes (free (no upper boundary - 66 MYR); Sphenisciformes – Procellariiformes (free – 60.5 MYR); Coraciidae – Alcedinidae (free – 51.57 MYR); Apodidiae – Trochilidae (free -51 MYR) and Psittaciformes – Passeriformes (65.5 – 53.5 MYR). In addition to that, calibration set 4 contains the split between Oscines and Suboscines (34 – 28 MYR). Although body mass seems to be a more accurate predictor for substitution rates when using 3rd codon positions (Nabholz et al. 2016), the analysis was additionally run with all coding positions included, again using fossil calibration sets 2 (Scenario 4) and 4 (hardly different from Scenario 4, so not added as different scenario), respectively. Thirdly, we used rates calculated for Accipitriformes from Arcones et al. (2019) as substitution rates for the different mitochondrial genes (Scenario 5). Finally (Scenarios 6), we inferred rates using the divergence time between common buzzard (*Buteo buteo*) and white-tailed eagle of 12.25 MYR (Mindell et al. 2018). The common buzzard was used as it is the closest available relative with an associated divergence date to the white-tailed eagle. The mean pairwise Tamura-Nei distances were calculated for each of the 17 partitions, between the sequences and common buzzard using MEGA version X (Kumar et al. 2018).

Partitions of the molecular variance among and within populations (AMOVA) were summarized based on haplotype frequencies (F_{ST}) and evolutionary distances between sequences (Φ_{ST}), as well as between any pairs of samples (Excoffier et al. 1992, Tamura and Nei 1993) using the hierfstat (Goudet 2005) and pegas (Paradis 2010) packages in R. The significance of the F_{ST} and Φ_{ST} values were estimated by permutation of sequences among samples 1000 times. A Kruskal's Non-metric

multidimensional scaling (MDS) plot was calculated based on the pairwise Φ_{ST} distances, using the MASS package (Ripley 1996, Cox and Cox 2008).

Signs of selection on the mitochondrial variation were investigated by calculating the Neutrality Index (NI) (Rand and Kann 1996) and with comparisons based on different estimates of the population parameter θ (see below). The Neutrality Index was calculated using the standard and generalized MacDonald and Kreitman test (with the program at http://mkt.uab.es) (Egea et al. 2008) using the vertebrate mitochondrial code for all 13 mitochondrial protein coding genes concatenated (with ND6 reversed), as well as for each gene for our complete data set and separately for the different countries. The ratio of segregating synonymous and non-synonymous variants within the populations was compared with the number of corresponding differentiations from black kite (*Milvus migrans*, NC_038195) as closest related species with an available complete mitochondrial sequence. Results were obtained for both uncorrected NI as well as corrected for distance by Jukes & Cantor (Jukes and Cantor 1969) as implemented on the website. When at least one of the cells of the McDonald-Kreitman 2x2 contingency table contained a count of zero, a count of one was added to each cell. Differences in the estimates of the population parameter theta were tested with Tajima's D (Tajima 1989), Fay and Wu's H (Fay and Wu 2000), and the E-test (Zeng et al. 2006) which may also reveal changes in population sizes, were calculated using Zeng's DH-software (http://zeng-lab.group.shef.ac.uk) and tested by coalescence simulations using the ms-program assuming constant population sizes (Hudson 2002).

Results

A total of 124 variable sites defining 43 mitogenomic haplotypes were found in the 88 individuals from Iceland, Norway, Denmark, Estonia, and Greenland (one individual from Greenland was excluded from statistical analysis, see below). 48,495,446 reads were mapped to the individuals with a length of 257 bases each, after filtering 45,303,601 were left (93,4%). Only five individuals

had missing data, with a maximum of 49 missing bp, i.e., the breadth (coverage) was minimum 0.997, and 84 individuals had a breadth of 1. The mean depth ranged from 12.7 to 7894.84 (mean = 857.3). The control region was separately analysed and had a mean depth between 9.9 and 6774,34 (mean = 800), and thus slightly but not substantially lower. Overall quality, depth, ratio of the common allele and binomial probability of segregating sites in each individual within each country, summarized from the filtered vcf file, show a good support for the segregating sites and that they behave as haploid markers (Figure S1), with no evidence of sample contamination or index hopping. Inspection of highly variable sites within countries were no exception, as found for eleven sites separating the two main mtDNA lineages in Iceland (Table S3). Nucleotide diversity along the mitochondrial genome was well distributed (Table S4). The small population in Iceland has the lowest haplotype richness and displays lowest haplotype diversity (0.7), but the population from Greenland show the lowest nucleotide diversity ($3.61*10^{-4}$) (Table 1). The large population in Norway has the highest haplotype richness and diversity, while the recently established population in Denmark has the highest nucleotide diversity. The population parameters θ_S based on segregating sites and nucleotide diversity are also higher in Denmark than for the other countries. θ_K, which unlike nucleotide diversity is sensitive to changes in number of alleles and thus bottlenecks, was lowest in Iceland and the newly established population in Denmark.

Both the median-joining network (Figure 2) and the Bayesian tree (Figure 3) reveal two major haplogroups in our data, one with all Estonian and about half the Danish individuals (HG-B), the other with Norway, Iceland, Greenland, and the rest of the Danish individuals (HG-A). The haplotype detected in the three Danish individuals clustering in HG-A were identical to the haplotype from Germany (MN356434.1). The Korean reference specimen forms a third and a distinct branch (Figure 2) that shows greater similarity to the branch with Estonian and half of the

Danish individuals (HG-B) than to the others (HG-A) (Figure S2). One sequence from Greenland (GL-3) clustered far from the rest of the sequences from Greenland in the network, but in the Bayesian analyses, this sequence clustered within the group IG-1. An inspection of the sequence showed signs of it being a chimaera, as 17 of 25 variable sites in the sequence covering most of the sequence i.e., from position 1 to 14597 were shared by other individuals from Greenland (in group HG-A), whereas eight variable sites from the control region (CR) were shared with the distinct lineage (HG-B) observed in individuals from Denmark, Estonia, and Korea. The individual was marked with an asterisk (Figure 2 and 3), but removed from statistical analysis, except for the neutrality index which was based on the coding sequence. Six topologies were found with the Bayesian analysis, they all show identical grouping of individuals but differ in the sequence of splits which were poorly supported (Figure S3). The most frequent topology (Figure 3) was observed four times. One topology was observed twice and four topologies only once. Two distinct clusters were found within each country in HG-A. These clusters were separated by several mutations (Figure 2 and 3) and contain highly similar haplotypes (N1 and IG1 in one, and N2 and IG2 in the other). Iceland and Greenland share closely related or similar haplotypes that differ in frequency at the different clusters (IG1 and IG2). The topology indicates no batch effect even though the data was obtained from various sources and sequenced at two different facilities. A network based on the part of the control region analysed by Hailer et al. (Hailer et al. 2007), and for the whole region excluding the control region, both present the same overall pattern (except for the individual from Greenland displaying signs of being a chimaera), although the CR-fragment does not provide as high a resolution as the entire mitogenome (Figure S4, S5 and Table S1).

The earliest divergence of haplogroup HG-B from haplogroup HG-A (Figure 3) occurred on average 130K years ago. A second major split distinguishes the Danish individuals (except one) from the Norwegian, Icelandic, and Greenlandic eagles (57K years ago). The third major split divides one

third of the Norwegian individuals, one third of the Icelandic individuals, and all but two Greenlandic individuals into one clade (N1 and IG1), and the remaining in the other clade (N2 and IG2) (51K years ago). In both of these two last clades, the Norwegian individuals make up separate monophyletic clades (with one Danish individual), splitting from the Icelandic and Greenlandic individuals around 30 and 37 thousand years ago. All these referred times are, as in Figure 3, assuming standard mutation rate.

Split times for all the six methods used (three methods were used for body mass), and for the three major splits mentioned above are displayed in Figures S6-S8. The most recent splits obtained with the different methods range from around 336K (237-437K) to 51K (36-68K) years ago. For the root node, the mean split times range from 847K (621K-1.086 Mio) to 130K (94-169K) years ago for the oldest and youngest, respectively.

The AMOVA (Excoffier et al. 1992) analysis reveals that a large proportion of the variation in genetic distances is due to differences between populations from the different countries (Φ_{ST} = 0.56, $p<0.01$). This proportion was considerably larger than the one obtained with the partition based on haplotypes only (F_{ST} = 0.079, $p<0.01$) and reflects the unique lineages found in most samples (Figure 2).

Accordingly, pairwise Φ_{ST} were larger than F_{ST} values (Table 2) and showed significant differences between all countries, excluding the small sample of the Estonian population. The positions based on the pairwise Φ_{ST} distances in the MDS-plot reflect well the original genetic distances (with a stress of 2.5%) and show relation to geographic maps (Figure S9), although the distance between Greenland and Iceland is large despite close similarity of the sequences from the two clades IG1 and IG2 in Figure 2. The overall distance between the two samples can be explained by the different proportions of Icelandic and Greenland haplotypes in the two clusters (IG1 has 13 Icelandic and 10 Greenlandic individuals, IG2 has 29 Icelandic and 2 Greenlandic individuals).

We found an excess of segregating amino acids in comparison with the fixed differences between the white-tailed eagle and the black kite, indicating negative selection (NI>1) (Table 3), with no stop codons detected. This excess of nonsynonymous polymorphisms is especially driven by high NI values for the COX complex (COX 1 and 2) and for CYTB when all individuals are pooled. High significant NI-values in the COX complex for all populations are found, primarily in COX2 in Iceland, Greenland and Norway, but COX1 in DK. Iceland and Greenland show significant negative selection in ND1, Norway and Denmark display significant negative selection in CYTB (Table 3). The Fu and Way's H test statistic and especially Zeng's E-test support this (Table 4); there are high frequency variants segregating within each population which are located at the internal branches within the two distinct lineages in each country, each with closely related haplotypes, i.e., the tree has a greater height than expected based on its overall length (Figure 3). An evaluation of the impact of sample size in Iceland revealed similar results (Table S5). None of the Tajima's D values were significant, but interestingly, the test statistic is positive for all countries except Greenland, indicating there are less rare alleles than expected (Table 4).

Discussion

Population structure

The mtDNA genomes of white-tailed eagles sampled from Greenland, Iceland, Norway, Denmark, and Estonia show difference in haplotype frequencies, occurrence of different haplotypes in west and east Europe, and thus differentiation between the countries (Φ_{ST}), which are all in line with earlier findings on the mitochondrial DNA in white-tailed eagles based on a fragment of the control region (Hailer et al. 2007, Honnen et al. 2010). The study by Hailer et al. (2007) revealed a recent genetic differentiation between the populations from Iceland and Greenland, where the frequency of two haplotypes differed between the two islands. One of the haplotypes was endemic to the islands and the other common in northern Europe. Even though Hailer et al. (2007) observed large spatial overlap of CR haplotypes on the mainland, when constructing a

401 neighbour-joining tree they detected a split between western and eastern Europe. Here we also
402 observe the recent divergence of the two island populations and a split between western Europe
403 and the samples from Estonia in eastern Europe, yet the recently established population from
404 Denmark shows an admixture of the two groups. The sample from Denmark shares one haplotype
405 with the Norwegian sample, suggesting migration from Norway, another haplotype is identical to
406 one from Germany and five others share a similar haplotype to the Estonian samples, suggesting a
407 settlement from the east. However, further sampling, and comparison with the surrounding
408 countries is needed to confirm geneflow, and the full population structure.
409
410 ## Divergence of populations
411 Despite the similarities between the main geographic patterns in our study and the previous
412 studies, a more complex pattern is observed based on the entire mitochondrial genomes than the
413 control region. Within each country, except in the small sample set from Estonia, two distinct
414 mitogenomic lineages are observed and are being upheld, potentially through selection. The
415 sequences within the two separate lineages in Iceland and Greenland are similar to each other
416 but have diverged from the Norwegian haplotypes in the same clade for about 30,000-37,000
417 years ago, as estimated by the standard mutation rate for CYTB or around the last glacial
418 maximum c.a. 25,000 years ago, when the three countries were covered by glaciers (Clark and Mix
419 2002). It is interesting that the split happened during or before the last glacial maximum, i.e.,
420 prior to the colonization of the three countries. The time to the most recent common ancestor
421 (TMRCA) for the two distinct lineages within Greenland, Iceland and Norway was older or about
422 ~50,000 years (CI: 115,000 – 11,700 years ago) or during the last glaciation, suggesting that they
423 originated from two refugia during the last glacial period. The dating to the TMRCA for all
424 countries occurred close to onset of the last glaciation assuming the standard rate (Figure S6),
425 130,000 years (mean standard rate).

426 The other methods of timing based on more ancient divergence times between species i.e., on
427 body size (Nabholz et al. 2016) and the one from Arcones et al. (2019) gave more ancient splits
428 than based on the standard rate for CYTB (Weir and Schluter 2008). This difference could result
429 from molecular clock rates obtained from comparisons between species, which may overestimate
430 the divergence times within species as they do not consider the influence of population dynamics
431 such as selection (Ho et al. 2005, 2007; Zink and Barrowclough 2008). The standard rate may be
432 less affected, and our estimates come close to the earlier estimates (Hailer et al. 2007), however
433 the large divergence within populations, which seems to be maintained by selection, as discussed
434 below, will also affect the time of divergence.

435 Different refugia during the last glacial period have been suggested as the main factor driving the
436 formation of the western and eastern clades in the white-tailed eagle (Hailer et al. 2007, Langguth
437 et al. 2013), and the estimates here from the standard rate are close to these earlier suggested
438 split times. The exact location of the refugia is unknown, but such refugia driving population splits
439 are also seen in other birds with a large east/west range on the northern hemisphere (Ruokonen
440 et al. 2004, Lovette 2005). More extensive geographic sampling could further refine theories of
441 the location of the refugia and the colonization history of Greenland and northern Europe.

442

443 ## Selection

444 As the mitochondria contain haploid, non-recombining genomes, selection is predicted to reduce
445 the molecular variation within populations due to background selection against deleterious
446 mutations (Charlesworth et al. 1993) or due to hitchhiking with positively favored mutations
447 (Smith and Haigh 1974). These effects are expected to be even stronger in birds due to linkage
448 with the W-chromosome as they are inherited together (Berlin et al. 2007).

449 Here we observe a different outcome, high variation within populations with distinct lineages and
450 an excess of high frequency variants as revealed by the negative Fay and Wu's H in Iceland and

Greenland, which is in line with Zeng's E, displaying long internal branches (i.e., the deep split within the countries). The significantly low H and high E values could indicate balancing selecting for the two lineages, with IG1 and NO1 in one, and IG2 and NO2 in the other (Zeng et al. 2006). However, balancing selection is unlikely to be directed toward the mitochondrial genome, but could rather act on the W chromosome, due to the linkage. And be driven by its interaction with the Z-chromosome e.g., in the pseudoautosomal region (PAR) or non-recombining homologs (Xu and Zhou 2020) and then indirectly maintain different lineages in the mitochondria. And as balancing selection will result in larger sequence variation within populations, our estimates of TMRCA within populations are overestimated as well as the time since the populations diverged from each other (Zink and Barrowclough 2008, Cooper et al. 2015). Further evidence for the effect of selection was seen with the neutrality index > 1, i.e., higher ratio of polymorphic amino acids within the white-tailed eagles in comparison to the ratio of fixed amino acid replacements relative to a related species, which suggests that deleterious mutations are segregating within the mitochondrial genome of the eagles so the variation may not reflect neutral processes.

A sympatric occurrence of distinct genetic lineages from a deep gene tree as seen here, has been suggested earlier for the white-tailed eagle (Honnen et al. 2010), and though rarely seen it has been suggested for other species, e.g., Adelie penguin, mallard, snow goose, blue tit, and orcas (Avise et al. 1987, Dizon et al. 1992, Avise and Walker 1998). Although this dichotomy may represent an admixture of different clades, two different colonization events, high ancestral variation or incomplete linage sorting, the maintenance of these distinct clades within the small populations of white-tailed eagles in Greenland and Iceland, and in the larger population in Norway, indicate that natural selection has played a role. Nuclear inserts of mitochondrial DNA (NUMTs) could have led to such a pattern, but inspection of the variable sites separating the lineages within populations did not reveal any such evidence.

Conclusion

Here we confirm the suggested western and eastern clades of mitochondrial variation in white-tailed eagle and the main phylogeographic pattern in Greenland and Northern-Europe. Surprisingly, two distinct lineages are found within each country. These two lineages in Greenland, Iceland and Norway seem to have been upheld by natural selection and may reflect divergence in different refugia during the last glacial period. In Denmark, the two lineages derive from the eastern and western clades and reflect a recent recolonization from both areas. Wider geographic sampling is warranted for a complete phylogeography of Europe and to rebuild the natural history of the species. Though useful for conservation efforts an addition of other genomic markers is needed to assess the diversity within populations, including assessments of demography, effective population size and divergence. Furthermore, to explore the signal of selection an analysis of the sex-chromosomes Z and W could reveal whether balancing selection acting on Z and W can extend over to the mitogenome and explain the coexistence of the two lineages within countries.

References

Arcones, A., R. Ponti, and D. R. Vieites (2019). Mitochondrial substitution rates estimation for molecular clock analyses in modern birds based on full mitochondrial genomes. bioRxiv:11.

Avise, J. C., J. Arnold, R. M. Ball, E. Bermingham, T. Lamb, J. E. Neigel, C. A. Reeb, and N. C. Saunders (1987). Intraspecific Phylogeography: The Mitochondrial DNA Bridge Between Population Genetics and Systematics. Annual Review of Ecology and Systematics 18:489–522.

Avise, J. C., and D. E. Walker (1998). Pleistocene phylogeographic effects on avian populations and the speciation process. Proceedings of the Royal Society of London. Series B: Biological Sciences 265:457–463.

Babraham Bioinformatics (2010). FastQC. [Online.] Available at http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

503 Bandelt, H. J., P. Forster, and A. Rohl (1999). Median-joining networks for inferring intraspecific
504 phylogenies. Molecular Biology and Evolution 16:37–48.
505 Berlin, S., D. Tomaras, and B. Charlesworth (2007). Low mitochondrial variability in birds may
506 indicate Hill–Robertson effects on the W chromosome. Heredity 99:389–396.
507 BirdLife International (2016). Haliaeetus albicilla. The IUCN Red List of Threatened Species 2016
508 e.T2269513.
509 Bouckaert, R., J. Heled, D. Kühnert, T. Vaughan, C. H. Wu, D. Xie, M. A. Suchard, A. Rambaut, and
510 A. J. Drummond (2014). BEAST 2: A Software Platform for Bayesian Evolutionary Analysis.
511 PLoS Computational Biology 10:1–6.
512 Bouckaert, R. R., and A. J. Drummond (2017). bModelTest: Bayesian phylogenetic site model
513 averaging and model comparison. BMC Evolutionary Biology 17:1–11.
514 Bragin, E. A., S. A. Poessel, M. J. Lanzone, and T. E. Katzner (2018). Post-fledging movements and
515 habitat associations of White-tailed Sea Eagles (Haliaeetus albicilla) in Central Asia. The
516 Wilson Journal of Ornithology 130:784–788.
517 Carøe, C., S. Gopalakrishnan, L. Vinner, S. S. T. Mak, M. H. S. Sinding, J. A. Samaniego, N. Wales, T.
518 Sicheritz-Pontén, and M. T. P. Gilbert (2018). Single-tube library preparation for degraded
519 <scp>DNA</scp>. Methods in Ecology and Evolution 9:410–419.
520 Charlesworth, B., M. T. Morgan, and D. Charlesworth (1993). The effect of deleterious mutations
521 on neutral molecular variation. Genetics 134:1289–1303.
522 Clark, P. U., and A. C. Mix (2002). Ice sheets and sea level of the Last Glacial Maximum.
523 Quaternary Science Reviews 21:1–7.
524 Cooper, B. S., C. R. Burrus, C. Ji, M. W. Hahn, and K. L. Montooth (2015). Similar efficacies of
525 selection shape mitochondrial and nuclear genes in both Drosophila melanogaster and
526 Homo sapiens. G3: Genes, Genomes, Genetics 5:2165–2176.
527 Cox, M. A. A., and T. F. Cox (2008). Multidimensional Scaling. In Handbook of Data Visualization.
528 Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 315–347.
529 Darriba, D., G. L. Taboada, R. Doallo, and D. Posada (2012). JModelTest 2: More models, new
530 heuristics and parallel computing. Nature Methods 9:772.
531 Dizon, A. E., C. Lockyer, W. F. Perrin, D. P. Demaster, and J. Sisson (1992). Rethinking the Stock
532 Concept: A Phylogeographic Approach. Conservation Biology 6:24–36.
533 Drummond, A. J., A. Rambaut, B. Shapiro, and O. G. Pybus (2005). Bayesian coalescent inference
534 of past population dynamics from molecular sequences. Molecular Biology and Evolution
535 22:1185–1192.

536 Egea, R., S. Casillas, and A. Barbadilla (2008). Standard and generalized McDonald-Kreitman test: a
537 website to detect selection by comparing different classes of DNA sites. Nucleic Acids
538 Research 36:W157–W162.
539 Elts, J., A. Leito, M. Leivits, L. Luigujõe, R. Nellis, M. Ots, I. Tammekänd, and Ü. Väli (2019). Eesti
540 lindude staatus, pesitsusaegne ja talvine arvukus 2013-2017. Hirunda 32:1–29.
541 Ewens, W. J. (1972). The sampling theory of selectively neutral alleles. Theoretical Population
542 Biology 3:87–112.
543 Excoffier, L., P. E. Smouse, and J. M. Quattro (1992). Analysis of molecular variance inferred from
544 metric distances among DNA haplotypes: Application to human mitochondrial DNA
545 restriction data. Genetics 131:479–491.
546 Fay, J. C., and C. I. Wu (2000). Hitchhiking under positive Darwinian selection. Genetics 155:1405–
547 1413.
548 Feng, S., J. Stiller, Y. Deng, J. Armstrong, Q. Fang, A. H. Reeve, D. Xie, G. Chen, C. Guo, B. C.
549 Faircloth, B. Petersen, et al. (2020). Dense sampling of bird diversity increases power of
550 comparative genomics. Nature 587:252–257.
551 Goudet, J. (2005). HIERFSTAT, a package for R to compute and test hierarchical F -statistics.
552 Molecular Ecology Notes 5:184–186.
553 Greenland Institute of Natural Resources (2020). White-tailed Eagle. [Online.] Available at
554 https://natur.gl/arter/white-tailed-eagle/?lang=en.
555 Guindon, S., and O. Gascuel (2003). A Simple, Fast, and Accurate Algorithm to Estimate Large
556 Phylogenies by Maximum Likelihood. Systematic Biology 52:696–704.
557 Hailer, F., B. Helander, A. O. Folkestad, S. A. Ganusevich, S. Garstad, P. Hauff, C. Koren, V. B.
558 Masterov, T. Nygård, J. A. Rudnick, S. Shiraki, et al. (2007). Phylogeography of the white-
559 tailed eagle, a generalist with large dispersal capacity. Journal of Biogeography 34:1193–
560 1206.
561 Hansen, K. (1979). Status over bestanden af Havørn *Haliaeetus albicilla groenlandicus* Brehm i
562 Grønland i årene 1972-74. Dansk Ornitologisk Forenings Tidsskrift 73:107–130.
563 Helander, B., F. Hailer, and C. Vilà (2007). Morphological and genetic sex identification of white-
564 tailed eagle Haliaeetus albicilla nestlings. Journal of Ornithology 148:435–442.
565 Ho, S. Y. W., M. J. Phillips, A. Cooper, and A. J. Drummond (2005). Time dependency of molecular
566 rate estimates and systematic overestimation of recent divergence times. Molecular Biology
567 and Evolution 22:1561–1568.

568 Ho, S. Y. W., B. Shapiro, M. J. Phillips, A. Cooper, and A. J. Drummond (2007). Evidence for time
569 dependency of molecular rate estimates. *Systematic Biology*.
570 Honnen, A.-C. C., F. Hailer, N. Kenntner, I. Literák, L. Dubská, and F. E. ZACHOS (2010).
571 Mitochondrial DNA and nuclear microsatellites reveal high diversity and genetic structure in
572 an avian top predator, the white-tailed sea eagle, in central Europe. Biological Journal of the
573 Linnean Society 99:727–737.
574 Hudson, R. R. (2002). Generating samples under a Wright-Fisher neutral model of genetic
575 variation. Bioinformatics 18:337–338.
576 Isomursu, M., J. Koivusaari, T. Stjernberg, V. Hirvelä-Koski, and E.-R. Venäläinen (2018). Lead
577 poisoning and other human-related factors cause significant mortality in white-tailed eagles.
578 Ambio 47:858–868.
579 Jais, M. (2020). White-tailed Eagle. [Online.] Available at http://europeanraptors.org/white-tailed-
580 eagle/.
581 Jukes, T. H., and C. R. Cantor (1969). Evolution of Protein Molecules. In Mammalian Protein
582 Metabolism (H. N. Munro, Editor). Elsevier, pp. 21–132.
583 Kim, J. A., S. G. Kang, H. S. Jeon, J. H. Jeon, J. H. Jang, S. Kim, and J. An (2019). Complete
584 mitogenomes of two Accipitridae, Haliaeetus albicilla, and Pernis ptilorhynchus.
585 Mitochondrial DNA Part B: Resources 4:392–393.
586 Kircher, M., S. Sawyer, and M. Meyer (2012). Double indexing overcomes inaccuracies in multiplex
587 sequencing on the Illumina platform. Nucleic Acids Research 40:e3–e3.
588 Kongeaastien (2020). The white-tailed eagle. [Online.] Available at
589 http://kongeaastien.dk/en/content/white-tailed-eagle.
590 Kumar, S., G. Stecher, M. Li, C. Knyaz, and K. Tamura (2018). MEGA X: Molecular evolutionary
591 genetics analysis across computing platforms. Molecular Biology and Evolution 35:1547–
592 1549.
593 Langguth, T., A.-C. C. Honnen, F. Hailer, T. Mizera, S. Skoric, Ü. Väli, and F. E. Zachos (2013).
594 Genetic structure and phylogeography of a European flagship species, the white-tailed sea
595 eagle Haliaeetus albicilla. Journal of Avian Biology 44:263–271.
596 Leigh, J. W., and D. Bryant (2015). <scp>popart</scp> : full-feature software for haplotype
597 network construction. Methods in Ecology and Evolution 6:1110–1116.
598 Li, H., and R. Durbin (2009). Fast and accurate short read alignment with Burrows-Wheeler
599 transform. Bioinformatics 25:1754–1760.

600 Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R.
601 Durbin (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics
602 25:2078–2079.
603 Lõhmus, A. (1998). Eesti kotkaste argipäev. Eesti Loodus 5/6:210–213.
604 Lovette, I. J. (2005). Glacial cycles and the tempo of avian speciation. Trends in Ecology and
605 Evolution 20:57–59.
606 Lynch, M. (1996). Mutation accumulation in transfer RNAs: Molecular evidence for Muller's
607 ratchet in mitochondrial genomes. Molecular Biology and Evolution 13:209–220.
608 Macherey-Nagel (2014). Genomic DNA from Tissue, User manual, NucleoMag® 96 Tissue.
609 [Online.] Available at https://www.takarabio.com/assets/documents/User
610 Manual/NucleoMag 96 Tissue Genomic DNA Purification User Manual (PT5008-
611 1)_Rev_03.pdf.
612 Mindell, D. P., J. J. Fuchs, and J. A. Johnson (2018). Phylogeny, Taxonomy, and Geographic
613 Diversity of Diurnal Raptors: Falconiformes, Accipitriformes, and Cathartiformes. In Birds of
614 Prey: Biology and conservation in the XXI century (J. H. Sarasola, J. M. Grande and J. J. Negro,
615 Editors). Springer International Publishing, Cham, pp. 1–522.
616 Nabholz, B., R. Lanfear, and J. Fuchs (2016). Body mass-corrected molecular rate for bird
617 mitochondrial DNA. Molecular ecology 25:4438–4449.
618 Nei, M. (1987). Molecular evolutionary genetics. Columbia University Press, New York.
619 Nei, M., and F. Tajima (1981). DNA polymorphism detectable by restriction endonucleases.
620 Genetics 97:145–163.
621 Neiman, M., and D. R. Taylor (2009). The causes of mutation accumulation in mitochondrial
622 genomes. Proceedings of the Royal Society B: Biological Sciences 276:1201–1209.
623 Paradis, E. (2010). pegas: an R package for population genetics with an integrated-modular
624 approach. Bioinformatics 26:419–420.
625 Paxinos, E. E., H. F. James, S. L. Olson, M. D. Sorenson, J. Jackson, and R. C. Fleischer (2002).
626 mtDNA from fossils reveals a radiation of Hawaiian geese recently derived from the Canada
627 goose (Branta canadensis). Proceedings of the National Academy of Sciences of the United
628 States of America 99:1399–1404.
629 Petersen, Æ. (1998). Haförn. In Íslenskir Fuglar (S. Harðardóttir, Editor). Vaka-Helgafell hf.,
630 Reykjavik, pp. 120–123.

631 Qiagen (2006). DNeasy ® Blood & Tissue Handbook. [Online.] Available at
632 http://www.qiagen.com/products/genomicdnastabilizationpurification/dneasytissuesystem/
633 dneasybloodtissuekit.aspx#Tabs=t0.
634 R core Team (2020). R: A language and environment for statistical computing. [Online.] Available
635 at https://www.r-project.org/.
636 Rambaut, A., A. J. Drummond, D. Xie, G. Baele, and M. A. Suchard (2018). Posterior summarization
637 in Bayesian phylogenetics using Tracer 1.7. Systematic Biology 67:901–904.
638 Rand, D. M., and L. M. Kann (1996). Excess amino acid polymorphism in mitochondrial DNA:
639 Contrasts among genes from Drosophila, mice, and humans. Molecular Biology and
640 Evolution 13:735–748.
641 Randla, T., and A. Õun (1980). Kaljukotkas ja merikotkas Eestis 1970-ndail aastail. Eesti Loodus
642 4:512–515.
643 Ripley, B. D. (1996). Pattern Recognition and Neural Networks. Cambridge University Press.
644 Rohland, N., and D. Reich (2012). Cost-effective, high-throughput DNA sequencing libraries for
645 multiplexed target capture. Genome Research 22:939–946.
646 RStudio Team (2019). RStudio: Integrated Development for R. [Online.] Available at
647 http://www.rstudio.com/.
648 Ruokonen, M., L. Kvist, T. Aarvak, J. Markkola, V. v. Morozov, I. J. Øien, E. E. Syroechkovsky, P.
649 Tolvanen, and J. Lumme (2004). Population genetic structure and conservation of the lesser
650 white-fronted goose Anser erythropus. Conservation Genetics 5:501–512.
651 Salomonsen, F. (1979). Ornithological and ecological studies in SW Greenland. Meddelelser om
652 Grønland 204.
653 Schubert, M., S. Lindgreen, and L. Orlando (2016). AdapterRemoval v2: rapid adapter trimming,
654 identification, and read merging. BMC Research Notes 9:88.
655 Shields, G. F., and A. C. Wilson (1987). Calibration of mitochondrial DNA evolution in geese.
656 Journal of Molecular Evolution 24:212–217.
657 Skarphéðinsson, K. H. (2013). Haförninn. Fuglavernd (Fuglaverndarfélag íslands), Reykjavik.
658 Smith, J. M., and J. Haigh (1974). The hitch-hiking effect of a favourable gene. Genetical Research
659 23:23–35.
660 Tajima, F. (1989). Statistical Method for Testing the Neutral Mutation Hypothesis by DNA
661 Polymorphism. Genetics 123:585–595.

662 Tajima, F. (1996). The amount of DNA polymorphism maintained in a finite population when the
663 neutral mutation rate varies among sites. Genetics 143:1457–1465.
664 Tamura, K., and M. Nei (1993). Estimation of the number of nucleotide substitutions in the control
665 region of mitochondrial DNA in humans and chimpanzees. Molecular Biology and Evolution.
666 https://doi.org/10.1093/oxfordjournals.molbev.a040023
667 Thermo Fisher (2016). Thermo Scientific GeneJET Genomic DNA Purification Kit #K0721, #K0722.
668 [Online.] Available at https://www.thermofisher.com/document-connect/document-
669 connect.html?url=https%3A%2F%2Fassets.thermofisher.com%2FTFS-
670 Assets%2FLSG%2Fmanuals%2FMAN0012667_GeneJET_Whole_Blood_Genomic_DNA_Purific
671 ation_Mini_Kit_UG.pdf&title=VXNlciBHdWlkZTogR2VuZUpFVCBXaG9sZSBC.
672 Treinys, R., D. Dementavičius, S. Rumbutis, S. Švažas, D. Butkauskas, A. Sruoga, and M. Dagys
673 (2016). Settlement, habitat preference, reproduction, and genetic diversity in recovering the
674 white-tailed eagle Haliaeetus albicilla population. Journal of Ornithology 157:311–323.
675 Ueta, M., F. Sato, E. G. G. Lobkov, and N. Mita (1998). Migration route of White-tailed Sea Eagles
676 Haliaeetus albicillain northeastern Asia. Ibis 140:684–686.
677 Watterson, G. A. (1975). On the number of segregating sites in genetical models without
678 recombination. Theoretical Population Biology 7:256–276.
679 Weir, J. T., and D. Schluter (2008). Calibrating the avian molecular clock. Molecular Ecology
680 17:2321–2328.
681 Xu, L., and Q. Zhou (2020). The female-specific W chromosomes of birds have conserved gene
682 contents but are not feminized. Genes 11:1–14.
683 Zeng, K., Y.-X. Fu, S. Shi, and C.-I. Wu (2006). Statistical Tests for Detecting Positive Selection by
684 Utilizing High-Frequency Variants. Genetics 174:1431–1439.
685 Zink, R. M., and G. F. Barrowclough (2008). Mitochondrial DNA under siege in avian
686 phylogeography. Molecular Ecology 17:2107–2121.
687

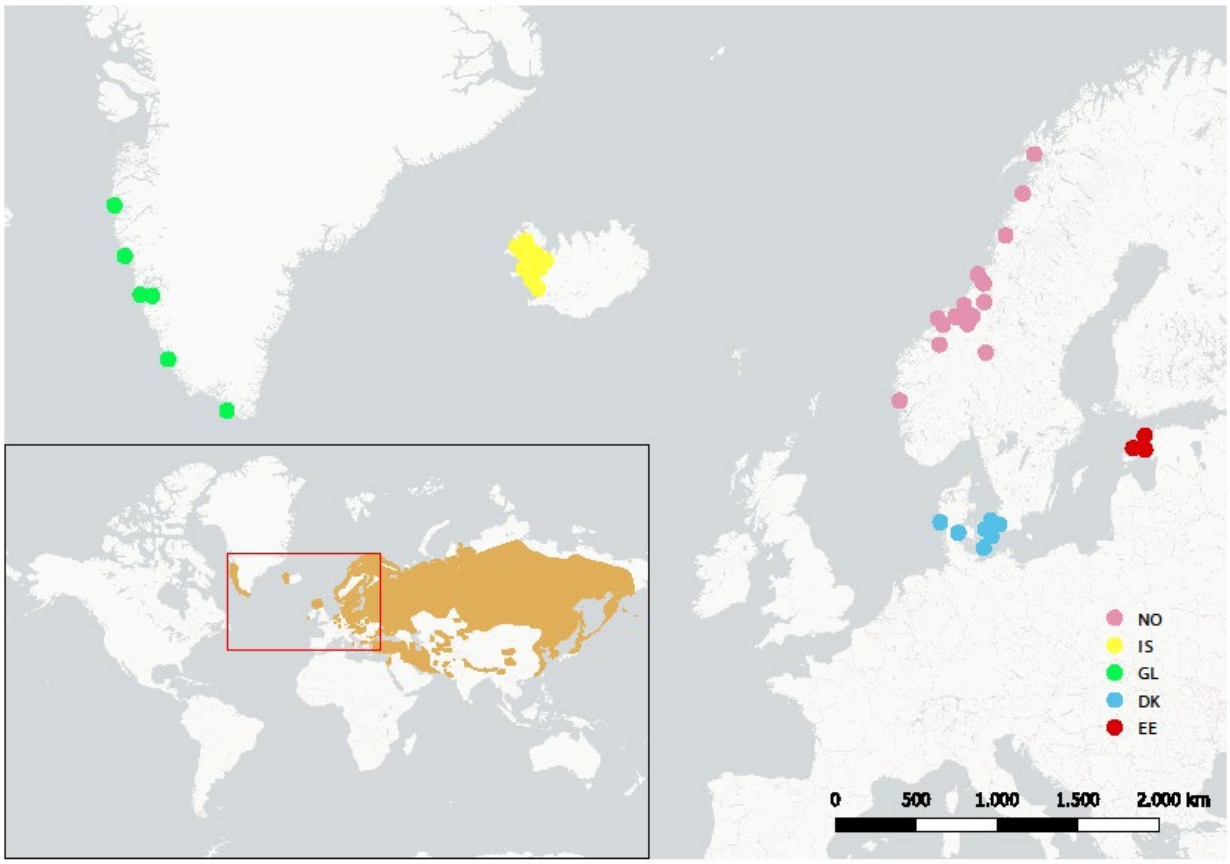


Figure 1. *Map of distribution and sample origin of the white-tailed eagles from five different countries analysed here. Small map in black box displays the world and the white-tailed eagle distribution in orange (without regards to where they breed and only use as passage etc.) (Birdlife International, 2020). The red box is the section of the world that makes up the larger map. The dots represent the known sample origin for the samples analysed. Some dots represent more than one sample, and some dots are overlapping. Greenland (GL) is in green (n=12), Iceland (IS) in yellow (n=42), Norway (NO) in pink (n=21), Denmark (DK) in blue (n=11), Estonia (EE) in red (n=3).*

Table 1. *Molecular diversity per country and overall for white-tailed eagle sampled between 1990 and 2018. Sample size (n), number of haplotypes (nh), haplotype richness (HR), haplotype diversity (h), nucleotide diversity (**π**), genomic diversity Theta Pi (θ_π) = nucleotide diversity (**π**) *17461 bp (17,461 is the length of the analysed mtDNA)* (Tajima, 1996), *Theta K (θ_K)* (Ewens, 1972), *Theta S (θ_S)* (F. Tajima, 1989; Watterson, 1975), *census size as reported and referenced in introduction (overall just report the estimate for Europe), numbers in parentheses are standard deviations. Standard deviation for θ_K were obtained with bootstrap. The individual from Greenland displaying signs of being a chimaera is excluded here.*

Country	n	nh	HR	h	π	Θ_π	Θ_K	Θ_S	Census
Greenland	11	8	5.68	0.93 (0.05)	3.61e-4 (2.0e-4)	6.31 (3.66)	11.69 (1.99)	6.83 (3.01)	150-200
Iceland	42	14	5.29	0.73 (0.06)	4.82e-4 (2.5e-4)	8.54 (4.47)	6.94 (1.04)	7.43 (2.45)	80
Norway	21	13	8.76	0.95 (0.01)	5.79e-4 (3.0e-4)	10.1 (5.37)	13.57 (1.98)	8.33 (3.11)	2,000

Denmark	11	7	7.00	0.87 (0.08)	1.83e-3 (9.7e-4)	31.97 (17.07)	6.15 (1.58)	24.58 (9.97)	61
Estonia	3	3	3.00	1.00 (0.19)	2.6e-4 (2.2e-4)	4.67 (3.90)	NA	4.66 (3.12)	290-330
Overall	88	43	8.4	0.92 (0.02)	1.08e-3 (5.3e-4)	18.97 (9.39)	32.57 (2.21)	24.5 (6.39)	17,900-24,500

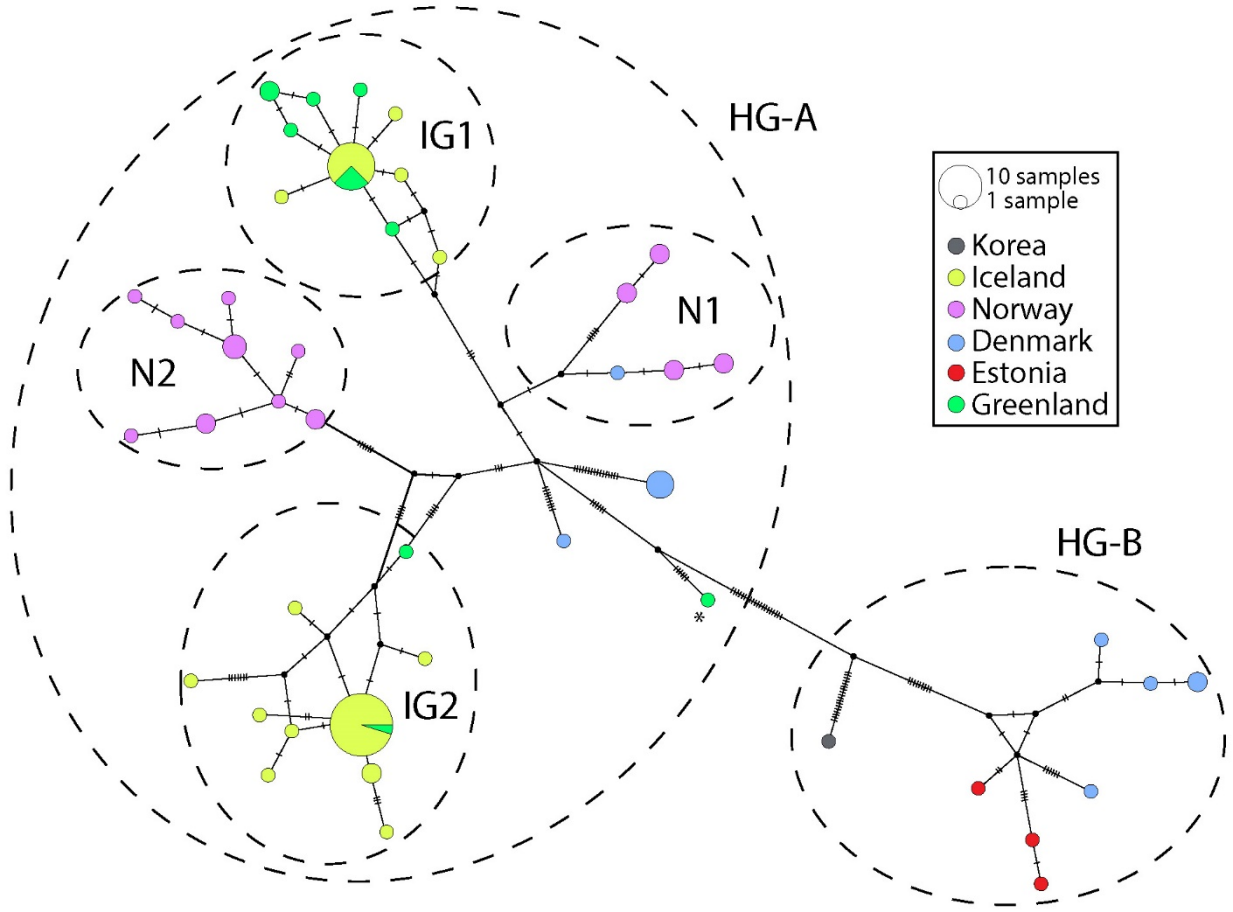
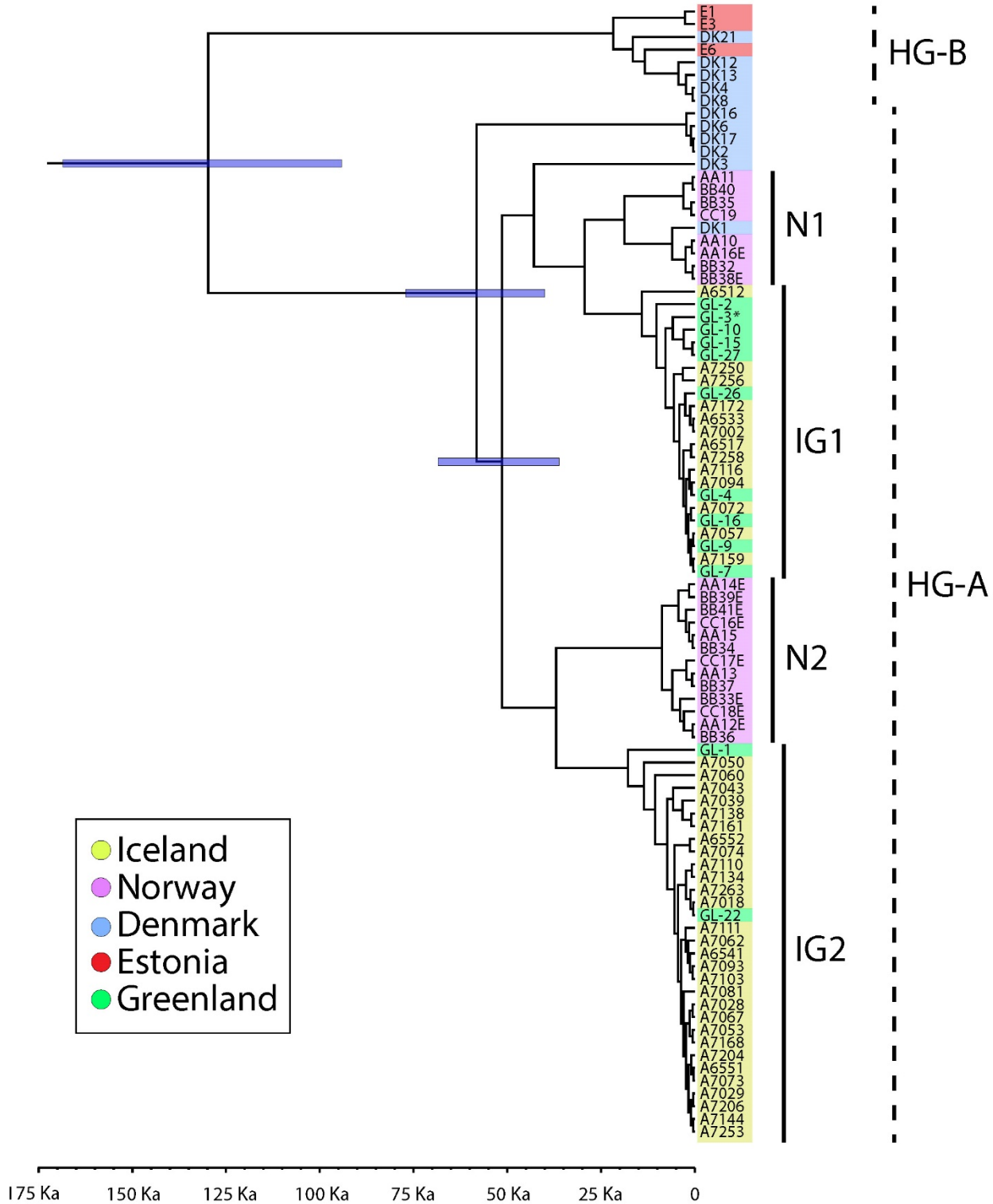


Figure 2. *Median-joining network presenting the divergence between mitochondrial genomes in white-tailed eagles from the five different countries sampled between 1990-2018 compared to the Korean reference specimen. The size of the pies represents the number of individuals that share the same haplotype. Marks on the lines refer to number of sites that differ among the connected haplotypes. HG-A and HG-B corresponds to the haplogroups A and B from Hailer et al* (2007) *(Figure S4). IG1, IG2, N1 and N2 are Iceland-Greenland group 1, Iceland-Greenland group 2, Norway group 1 and Norway group 2, respectively. The individual from Greenland displaying signs of being a chimaera is marked with an asterisk.*



719 ***Figure 3.*** *Phylogenetic tree based on the mitochondrial genomes in white-tailed eagles from five countries collected*
720 *between 1990 and 2018. The tree was reconstructed using BEAST. Datings are shown on the X-axis, from today and back*
721 *in time. The datings are based on the standard rate (2.1%), and error bars for the three major splits are marked in blue.*
722 *HG-A and HG-B corresponds to the haplogroups A and B from Hailer et al* (2007) *(Figure 2 and S4). IG1, IG2, N1 and N2*
723 *are Iceland-Greenland group 1, Iceland-Greenland group 2, Norway group 1 and Norway group 2, respectively. The*
724 *individual from Greenland displaying signs of being a chimaera is marked with asterisk.*

725

Table 2. *Pairwise comparison between five populations of white-tailed eagle sampled between 1990 and 2018. Φst below the diagonal, and pairwise Fst above the diagonal. Asterisks *, ** and *** indicates p-values: 0.01 <= P < 0.05, 0.001 <= P < 0.01 and P < 0.001, respectively. IS = Iceland, NO = Norway, DK = Denmark, EE = Estonia and GL = Greenland.*

	IS	NO	DK	EE	GL
Iceland		0.083***	0.109***	0.085	0.041*
Norway	0.408***		0.043***	0.012	0.026**
Denmark	0.560***	0.453***		0.038	0.049*
Estonia	0.940***	0.935***	0.273		0.018
Greenland	0.397***	0.459***	0.308*	0.961**	

Table 3. *Neutrality index (Rand & Kann, 1996). Bold = significant, italic = to all values in the 2x2 contingency table a count of 1 was added, since at least one cell contained a zero and a calculation of the neutrality index would not have been possible otherwise. All = all samples, JC = JC corrected. IS = Iceland, GL = Greenland, NO = Norway, DK = Denmark. Bases per gene; ATP6: 683, ATP8: 167, COX1: 1550, COX2: 683, COX3: 783, CYTB: 1142, ND1: 977, ND2: 1038, ND3: 351, ND4: 1377, ND4L: 296, ND5: 1817, ND6: 518.*

	All	All – JC	IS	IS – JC	GL	GL – JC	NO	NO – JC	DK	DK – JC
All genes	**2.3**	**3.0**	1.8	2.4	2.2	2.9	2.3	**3.1**	1.9	**2.5**
ATP6	*1.1*	*1.3*	*4.3*	*5.3*	*4.3*	*5.3*	*4.3*	*5.3*	*1.1*	*1.3*
ATP8	*0.5*	*0.5*	*0.9*	*1.1*	*0.9*	*1.1*	*0.5*	*0.5*	*0.9*	*1.1*
COX1	**8.4**	**11.0**	*7.1*	***9.3***	*7.1*	***9.3***	*7.1*	***9.3***	**12.6**	**16.4**
COX2	**45.6**	**60.8**	***29.0***	***38.6***	***29.0***	***38.6***	**22.8**	**30.4**	*4.8*	*6.4*
COX3	*3.8*	*4.8*	*7.6*	***9.7***	*7.6*	***9.7***	*5.0*	*6.4*	*7.8*	***10.0***
CYTB	**14.8**	**22.2**	*3.4*	*4.6*	*3.4*	*4.6*	**15.4**	**21.0**	**21.8**	**29.7**
ND1	4.0	5.4	***10.9***	***14.8***	***10.9***	***14.8***	*5.4*	*7.4*	6.0	8.1
ND2	0.8	1.1	*3.2*	*4.2*	*3.2*	*4.2*	*3.2*	*4.2*	0.8	1.1
ND3	1.2	1.7	*1.9*	*2.4*	*1.9*	*2.4*	*1.9*	*2.4*	3.7	5.2
ND4	3.2	4.3	*6.4*	*8.3*	*6.4*	*8.3*	***12.6***	***16.4***	*2.1*	*2.8*
ND4L	*0.4*	*0.6*	*3.5*	*4.6*	*3.5*	*4.6*	*3.5*	*4.6*	*0.4*	*0.5*
ND5	1.7	2.3	1.1	1.6	1.7	2.3	0.7	0.9	2.8	**3.7**
ND6	1.0	1.3	*3.8*	*5.3*	*3.8*	*5.3*	*3.8*	*5.3*	1.0	1.3
Complex I (ND)	1.4	1.9	1.3	1.7	1.8	2.4	1.3	1.7	1.4	1.8
Complex IV (COX)	**11.2**	**14.6**	**12.5**	**16.3**	**12.5**	**16.3**	**7.5**	**9.8**	**6.3**	**8.2**
Complex V (ATPase)	*0.5*	*0.6*	*2.7*	*3.3*	*2.7*	*3.3*	*1.3*	*1.6*	*0.7*	*0.8*

739 ***Table 4.*** *Tests of selection for white-tailed eagle samples between 1990 and 2018. Tajima's D, the normalized Fay and*
740 *Wu's H, and the E-test. 10.000 coalescence simulations with the number of segregating sites fixed. P-values as*
741 *proportions of number of simulations resulting in equal or lower values than the observed statistic.*

Region	Tajima's D	Fay and Wu's H	E-test
Greenland	-0.46 (0.343)	-1.99 (0.035)	1.68 (0.968)
Iceland	0.92 (0.863)	-2.66 (0.017)	3.22 (0.999)
Norway	0.85 (0.849)	-1.19 (0.091)	1.92 (0.979)
Denmark	1.47 (0.961)	-0.19 (0.268)	1.59 (0.966)

742

743

744

Supplementary information

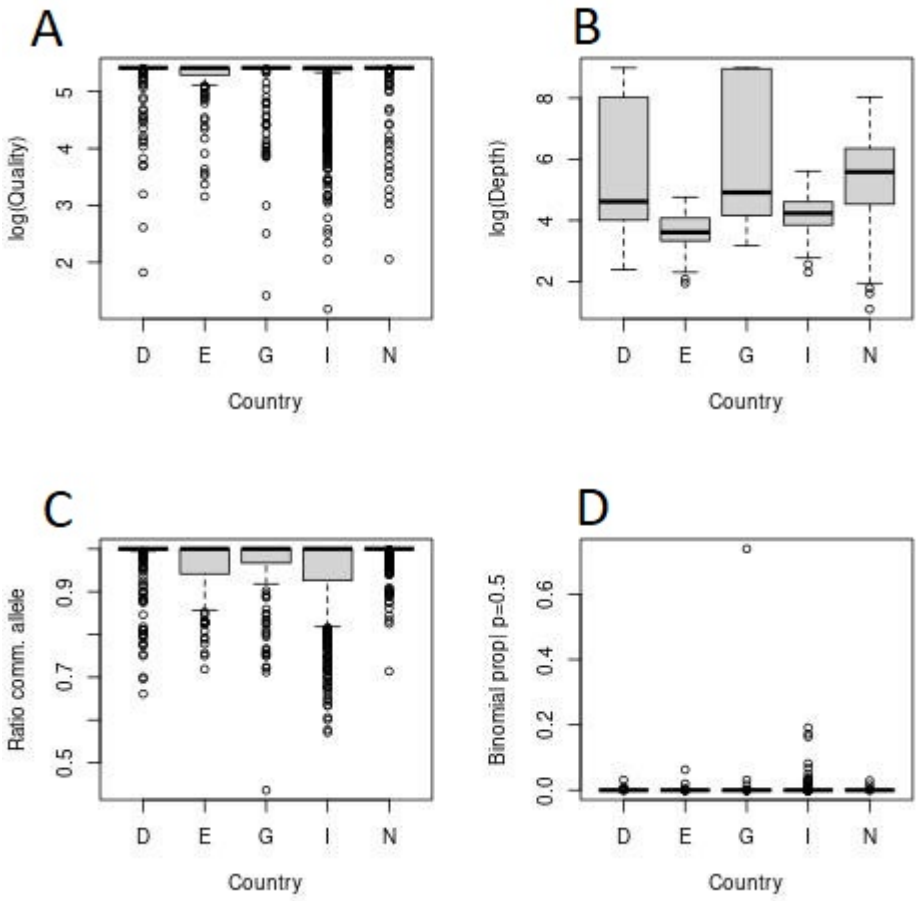


Figure S1. Assessment of the filtered mtDNA sequences from each country. Labels on the X-axis are D=Denmark, E=Estonia, G=Greenland, I=Iceland, N=Norway. Y-axis presents in A) log of the quality per position, B) log of the sequencing depth per position, C) proportion of the common allele, and D) the binomial probability of segregating sites assuming equal proportions. The statistics were summarized from the filtered vcf genotype file, plotted for all positions and for all individuals.

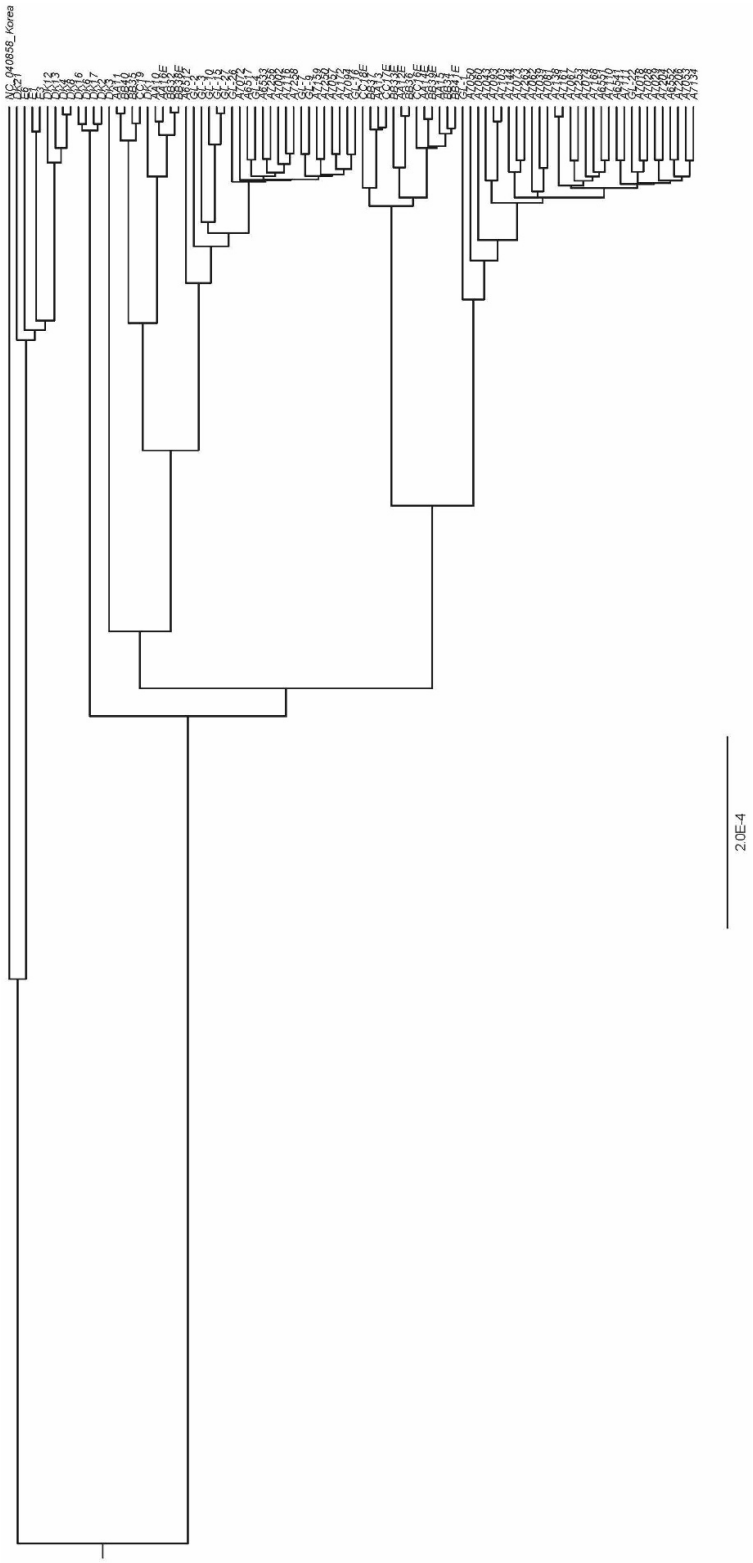


Figure S2. Topology of the mtDNA from white-tailed individuals from Greenland, Iceland, Norway, Denmark and Estonia including a Korean individual, which was used as reference, and obtained from genebank (NCBI Reference Sequence: NC_040858.1). Bayesian phylogeny tree was constructed using BEAST. Scale bar are in units of number of mutations per site.

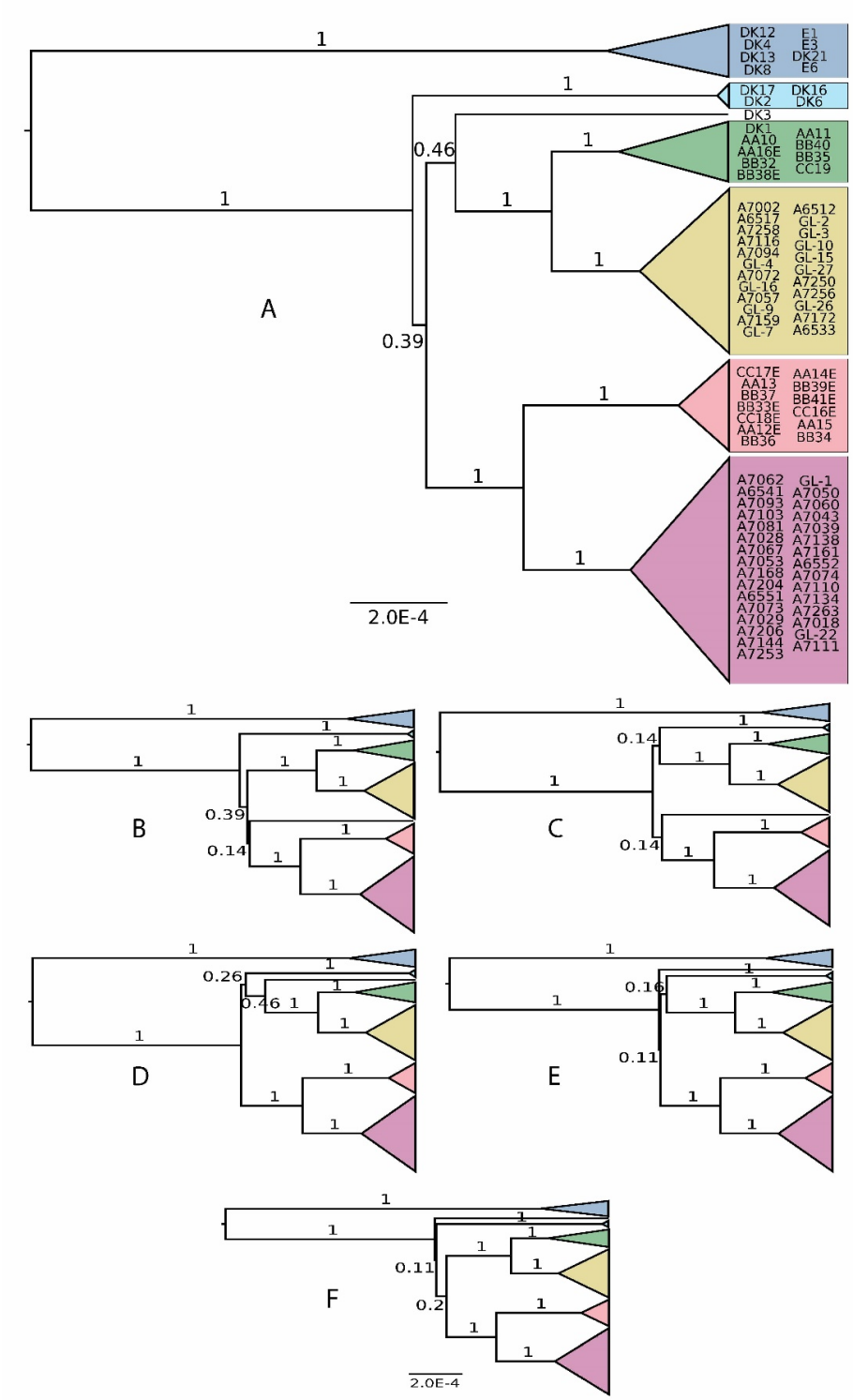


Figure S3. Topology of all white-tailed eagle individuals, included in the study, i.e., samples from Greenland, Iceland, Norway, Denmark, and Estonia, done to resolve the most occurring topology. Six different topologies were recovered with the Bayesian approach using the unpartitioned dataset. One tree (A) occurred four times, and thereby most often, and was used in all MCMC divergence time analyses. Tree E occurred twice. The rest only once. Numbers next to splits are bootstrap values. The different colors are clades consistent between topologies. Scale bar are in units of number of mutations per site.

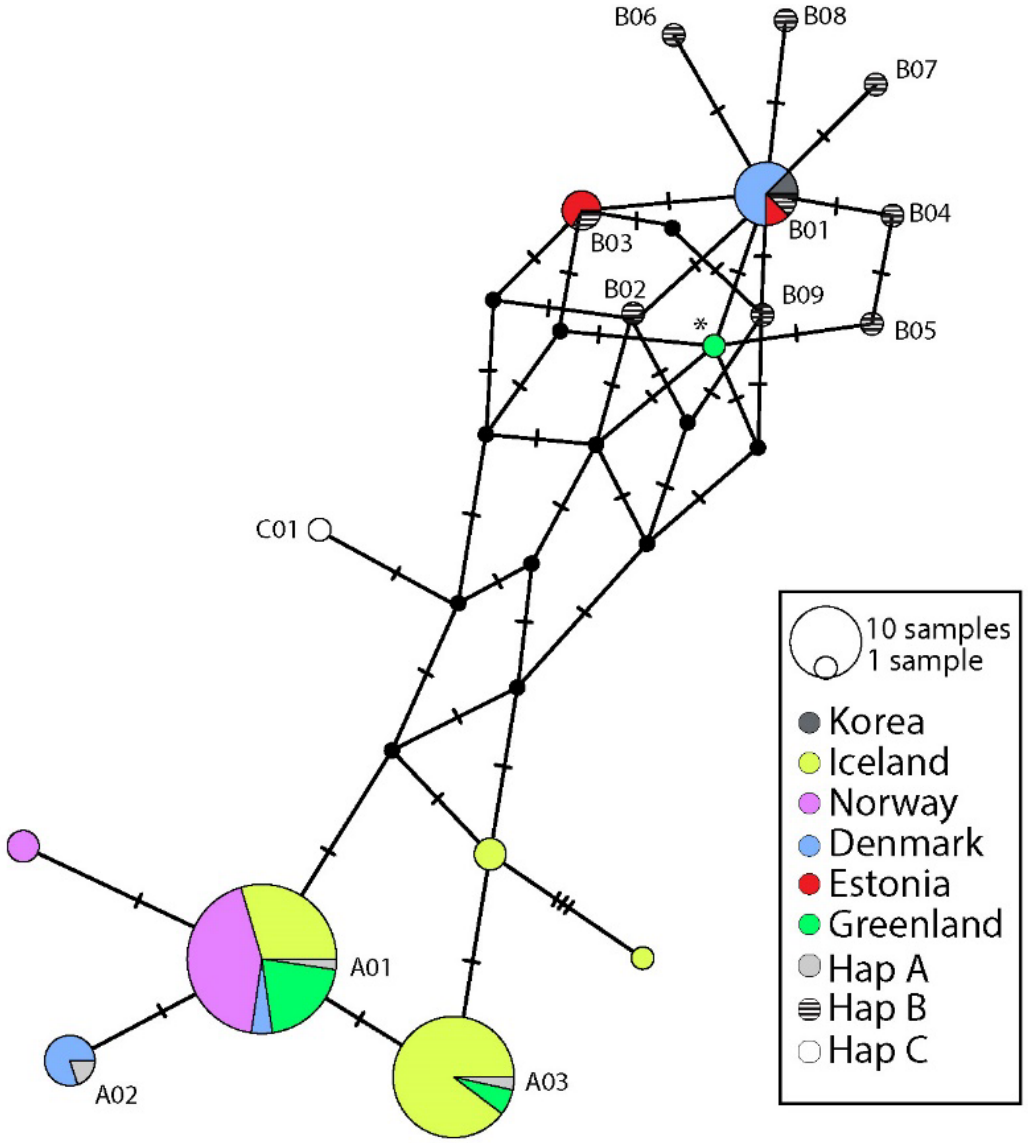


Figure S4. Median-joining network for the control region (CR) of the individuals used in this paper from Greenland, Iceland, Norway, Denmark and Estonia, and the reference individual from Korea. Compared to the former suggested haplotypes from Hailer et al. (Hailer et al., 2007). *It is found that the overall patters for the individuals analysed in this study are in line with what was found in Hailer et al., except here three Icelandic individuals do not fall in the A01 or A03 clade, and one Greenlandic individual also are quite far from A01 and A03. We do not find any individuals that are close to Haplotype C, however as no individuals from Sweden were included this was expected. The position of individuals here, can be compared to their corresponding position in Figure 2 in the main manuscript by looking in Table 1. From Table 1 it is also clear that our division of individuals into haplogroups A and B are completely in line with the findings for the CR region. The individual from Greenland displaying signs of being a chimaera is marked with asterisk.*

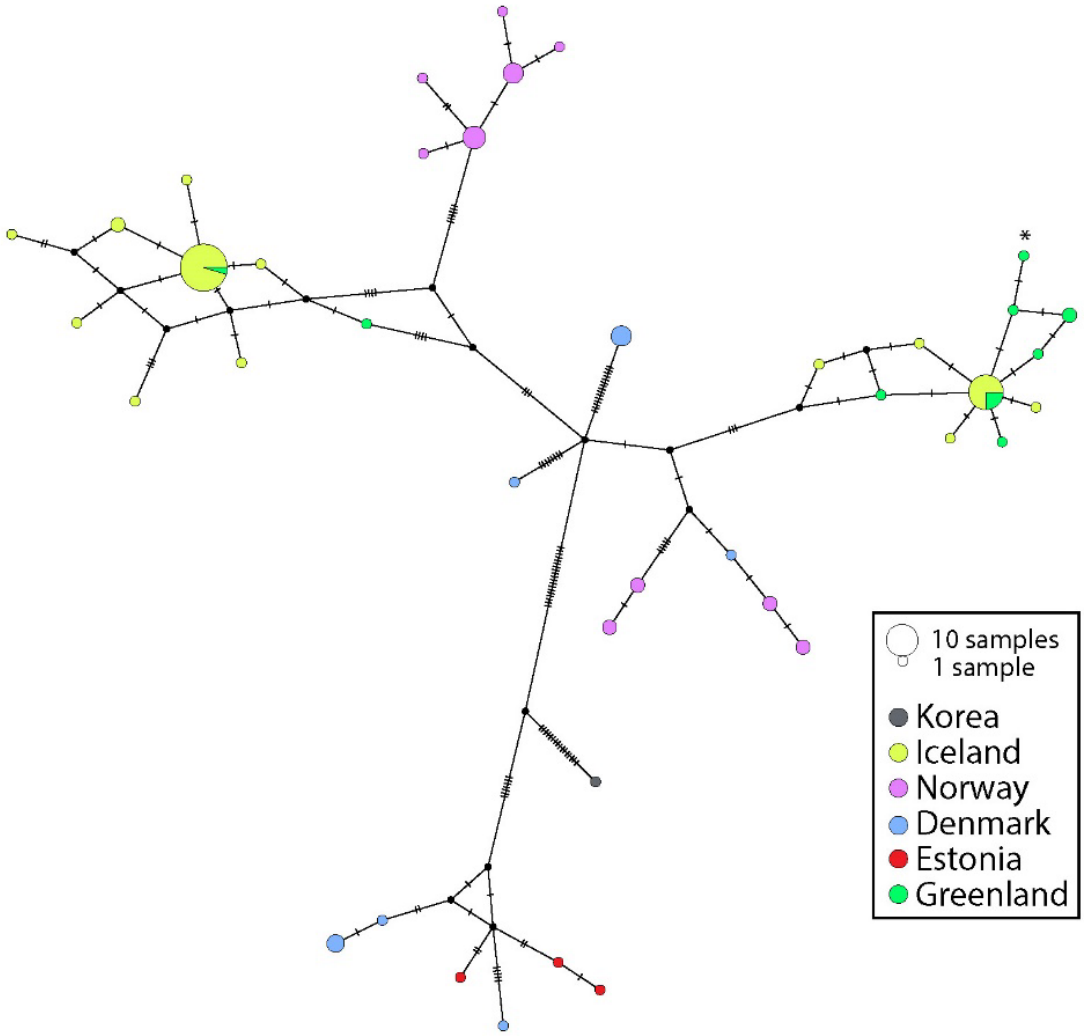


Figure S5. Median-joining network for the analysed mtDNA region, excluding the control region (CR) of the individuals used in this paper from Greenland, Iceland, Norway, Denmark and Estonia, and the reference individual from Korea. This was done to make sure that the results from Figure 2 holds without the control region. A few individuals were collapsed into a single haplotype when the CR was excluded. The biggest difference is that the Greenlandic individual that are located alone in Figure 2 in group HG-A, and alone in Figure S4, is here put with most other Greenlandic individuals in the right cluster that contains Greenlandic individuals, i.e., is the splitting of this individual from the rest caused by the CR region. The individual from Greenland displaying signs being a chimaera is marked with asterisk.

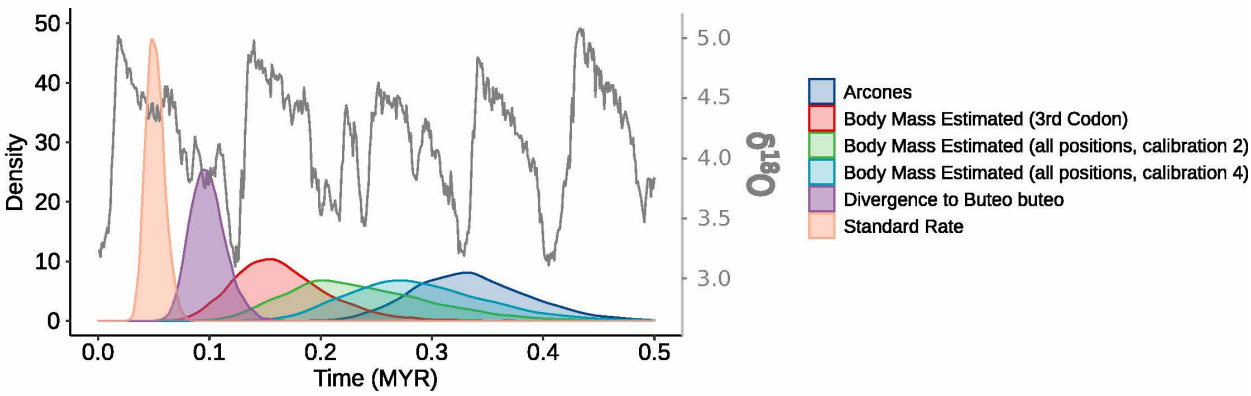


Figure S6. Split times of split three (latest) as in Figure 2, using the different mutation rates. Standard rate (2.1%) is the fastest and Arcones, Ponti, & Vieites (2019) *are the slowest. Overlayed with $\delta^{18}O$ from Lisiecki and Raymo* (2005) *where high numbers are equivalent to low temperatures. Colors represent the different mutation rates, as the described in the paper. X-axis is the time back in time in million years (MYR). The Y-axis is the density of the confidence interval.*

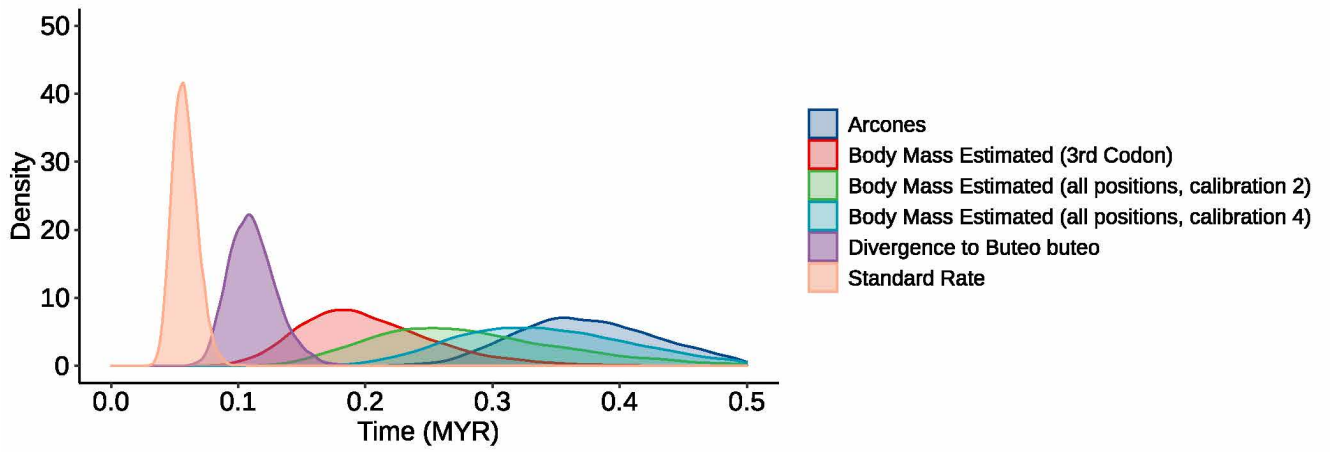


Figure S7. Split times of split two in Figure 2, using the different mutation rates. Standard rate (2.1%) is the fastest and Arcones (2019) *are the slowest. Colors represent the different mutation rates, as the described in the paper. X-axis is the time back in time in million years (MYR). The Y-axis is the density of the confidence interval.*

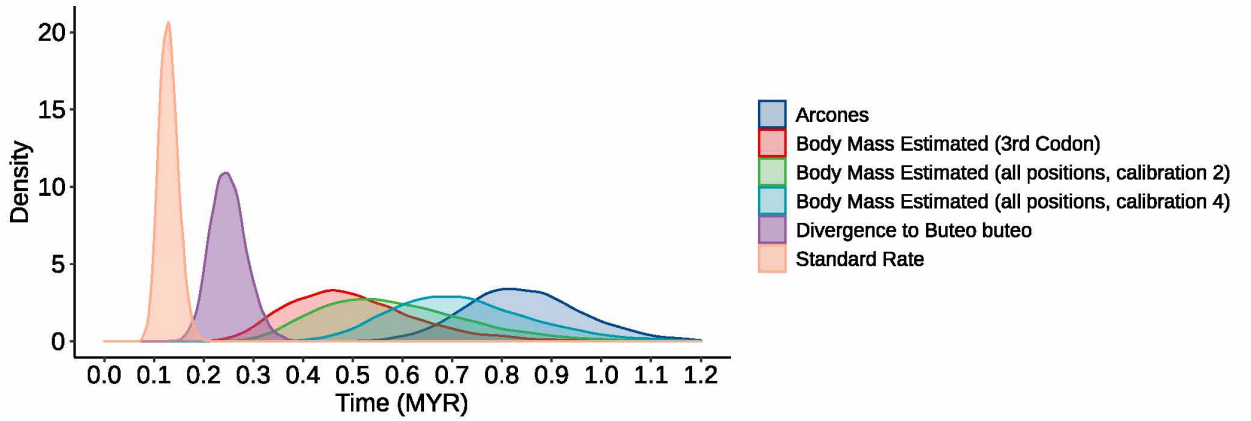


Figure S8. Split time for the root split in Figure 2, using the different mutation rates. Standard rate (2.1%) is the fastest and Arcones (2019) *are the slowest. Colors represent the different mutation rates, as the described in the paper. X-axis is the time back in time in million years (MYR). The Y-axis is the density of the confidence interval.*

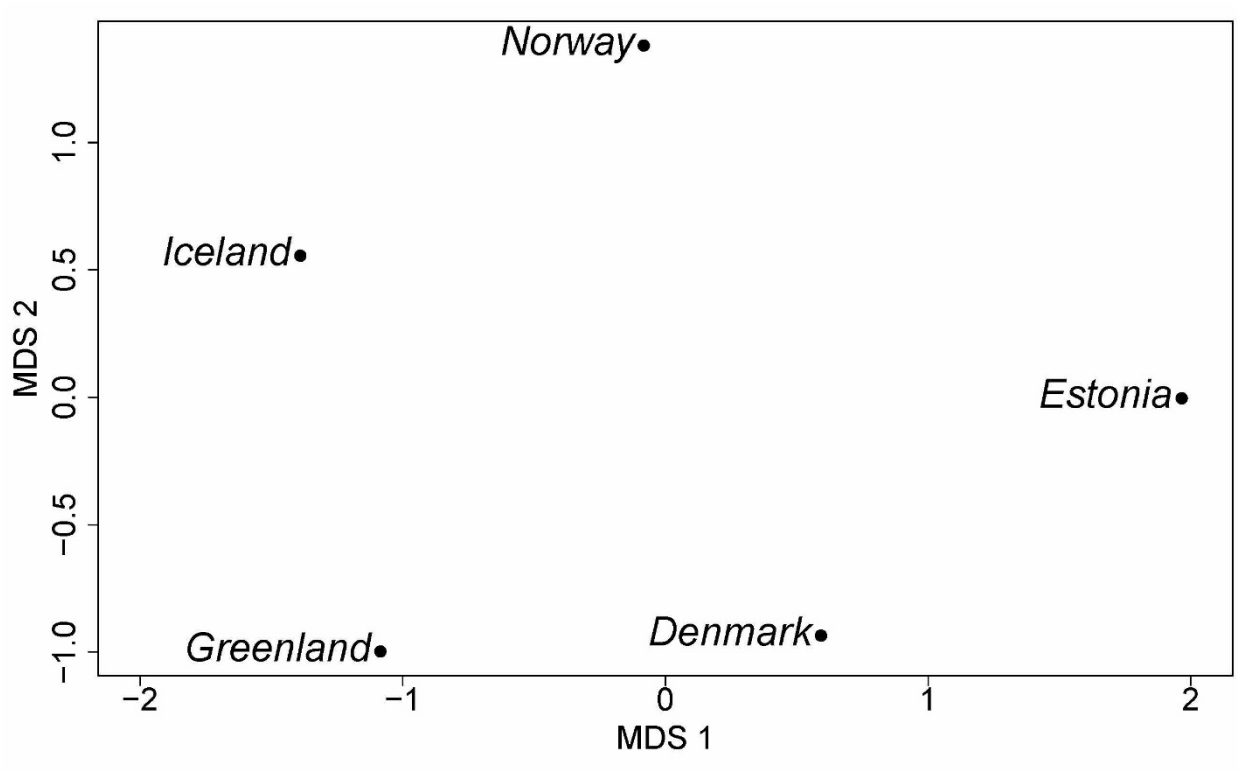


Figure S9. Multidimensional Scale (MDS) plot based on the Φ_{st} values, calculated between the five population, Greenland, Iceland, Norway, Denmark and Estonia. The stress for the MDS is 2.5%.

Table S1. Overview of samples, country, sampling location, sample type, age of individual at sampling, nest site number (when relevant), sampling year, which haplogroup it is located in the network (Figure 2), and where it fits when compared to the haplotypes in Hailer et al. (2007) *(Figure S4).*

#	Sample	Country	Location	Sample type	Age	Nest site number	Sample year	Haplo-group	Haplotype in Hailer et al.
1	A6512	Iceland	Faxaflói	Blood EDTA	Nestling	112	2003	IG1	A01
2	A6517	Iceland	Faxaflói	Blood EDTA	Nestling	120	2003	IG1	A01
3	A6533	Iceland	N-Breiðafjörður	Blood EDTA	Nestling	262	2003	IG1	A01

4	A6541	Iceland	S-Breiðafjörður	Blood EDTA	Nestling	206	2003	IG2	A03
5	A6551	Iceland	Vestfirðir	Blood EDTA	Nestling	312	2003	IG2	A03
6	A6552	Iceland	N-Breiðafjörður	Blood EDTA	Nestling	268	2003	IG2	A03
7	A7002	Iceland	S-Breiðafjörður	Blood EDTA	Nestling	211_31	2004	IG1	A01
8	A7018	Iceland	S-Breiðafjörður	Blood EDTA	Nestling	222	2004	IG2	A03
9	A7028	Iceland	Húnaflói	Blood EDTA	Nestling	410	2004	IG2	A03
10	A7029	Iceland	S-Breiðafjörður	Blood EDTA	Nestling	227	2004	IG2	A03
11	A7039	Iceland	S-Breiðafjörður	Blood EDTA	Nestling	222	2005	IG2	A03
12	A7043	Iceland	S-Breiðafjörður	Blood EDTA	Nestling	224	2005	IG2	A03
13	A7050	Iceland	N-Breiðafjörður	Blood EDTA	Nestling	268	2005	IG2	-
14	A7053	Iceland	N-Breiðafjörður	Blood EDTA	Nestling	275b	2005	IG2	A03

15	A7057	Iceland	N-Breiðafjörður	Blood EDTA	Nestling	278	2005	IG1	A01
16	A7060	Iceland	Vestfirðir	Blood EDTA	Nestling	410	2005	IG2	A03
17	A7062	Iceland	Húnaflói	Blood EDTA	Nestling	410	2006	IG2	A03
18	A7067	Iceland	S-Breiðafjörður	Blood EDTA	Nestling	214	2006	IG2	A03
19	A7072	Iceland	S-Breiðafjörður	Blood EDTA	Nestling	211_31	2006	IG1	A01
20	A7073	Iceland	S-Breiðafjörður	Blood EDTA	Nestling	252	2006	IG2	A03
21	A7074	Iceland	N-Breiðafjörður	Blood EDTA	Nestling	272	2007	IG2	A03
22	A7081	Iceland	N-Breiðafjörður	Blood EDTA	Nestling	312	2007	IG2	A03
23	A7093	Iceland	S-Breiðafjörður	Blood EDTA	Nestling	215	2007	IG2	-
24	A7094	Iceland	S-Breiðafjörður	Blood EDTA	Nestling	216	2007	IG1	A01
25	A7103	Iceland	Faxaflói	Blood EDTA	Nestling	124	2004	IG2	-
26	A7110	Iceland	Faxaflói	Blood EDTA	Nestling	124	2005	IG2	A03

27	A7111	Iceland	Faxaflói	Blood EDTA	Nestling	124	2005	IG2	A03
28	A7116	Iceland	Faxaflói	Blood EDTA	Nestling	128	2006	IG1	A01
29	A7134	Iceland	S-Breiðafjörður	Blood EDTA	Nestling	226	2008	IG2	A03
30	A7138	Iceland	S-Breiðafjörður	Blood EDTA	Nestling	213	2008	IG2	A03
31	A7144	Iceland	S-Breiðafjörður	Blood EDTA	Nestling	229	2008	IG2	A03
32	A7159	Iceland	Húnaflói	Blood EDTA	Nestling	306	2008	IG1	A01
33	A7161	Iceland	Vestfirðir	Blood EDTA	Nestling	273	2009	IG2	A03
34	A7168	Iceland	N-Breiðafjörður	Blood EDTA	Nestling	225b	2009	IG2	A03
35	A7172	Iceland	S-Breiðafjörður	Blood EDTA	Nestling	215	2009	IG1	A01
36	A7204	Iceland	Húnaflói	Blood EDTA	Nestling	410	2009	IG2	A03
37	A7206	Iceland	S-Breiðafjörður	Blood EDTA	Nestling	228	2009	IG2	A03
38	A7250	Iceland	Faxaflói	Blood EDTA	Nestling	124	2011	IG1	A01

39	A7253	Iceland	N-Breiðafjörður	Blood EDTA	Nestling	278	2011	IG2	A03
40	A7256	Iceland	Vestfirðir	Blood EDTA	Nestling	306	2011	IG1	A01
41	A7258	Iceland	Vestfirðir	Blood EDTA	Nestling	211_31	2011	IG1	A01
42	A7263	Iceland	S-Breiðafjörður	Blood EDTA	Nestling	214	2011	IG2	A03
43	AA10	Norway	Ertvågøy	Toepad	Post nestling	na	2007	N1	A01
44	AA11	Norway	Elgskaret	Toepad	Post nestling	na	2013	N1	A01
45	AA12E	Norway	Namsos	Toepad	Post nestling	na	2014	N2	-
46	AA13	Norway	Snåsamoen	Toepad	Post nestling	na	2014	N2	A01
47	AA14E	Norway	Muleskole	Toepad	Post nestling	na	2015	N2	A01
48	AA15	Norway	Otterøya	Toepad	Post nestling	na	2010	N2	A01
49	AA16E	Norway	Hannkistelen	Toepad	Post nestling	na	2012	N1	A01
50	BB32	Norway	Bodå	Toepad	Post nestling	na	2008	N1	A01
51	BB33E	Norway	Slåttavika	Toepad	Post nestling	na	2008	N2	A01
52	BB34	Norway	Håmanntjønna	Toepad	Post nestling	na	2009	N2	A01
53	BB35	Norway	Tettila	Toepad	Post nestling	na	2010	N1	A01

54	BB36	Norway	Frøvarp, Elvalandet	Toepad	Post nestling	na	2011	N2	-
55	BB37	Norway	Vikdalen	Toepad	Post nestling	na	2011	N2	A01
56	BB38E	Norway	Verma	Toepad	Post nestling	na	2013	N2	A01
57	BB39E	Norway	Smøla	Toepad	Post nestling	na	2010	N2	A01
58	BB40	Norway	Olsvik	Toepad	Post nestling	na	2012	N1	A01
59	BB41E	Norway	Risvær	Toepad	Post nestling	na	2012	N2	A01
60	CC16E	Norway	Vikhammer	Toepad	Post nestling	na	2001	N2	A01
61	CC17E	Norway	Åstfjorden	Toepad	Post nestling	na	2004	N2	A01
62	CC18E	Norway	Laugsjøen	Toepad	Post nestling	na	2005	N2	A01
63	CC19	Norway	Hesthåggån	Toepad	Post nestling	na	2009	N1	A01
64	DK12	Denmark	Filsø, Jutland	Full blood	Nestling	na	2015	HG-B	B01
65	DK13	Denmark	Filsø, Jutland	Full blood	Nestling	na	2015	HG-B	B01
66	DK16	Denmark	Hyllekrog, Lolland	Full blood	Nestling	na	2015	HG-A	A02
67	DK17	Denmark	Hyllekrog, Lolland	Full blood	Nestling	na	2015	HG-A	A02
68	DK21	Denmark	Kastrup, Copenhagen	Skin/muscle	Adult	na	2015	HG-B	B01

69	DK1	Denmark	Præstø, Zealand	Skin/muscle	Adult	na	2016	N1	A01
70	DK2	Denmark	Skudelev, Sealand	Skin/muscle	Adult	na	2016	HG-A	A02
71	DK3	Denmark	Sorø, Zealand	Skin/muscle	Adult	na	2018	HG-A	A01
72	DK4	Denmark	Klejs, Zealand	Skin/muscle	Adult	na	2018	HG-B	B01
73	DK6	Denmark	na	Skin/muscle	Adult	na	2016	HG-A	A02
74	DK8	Denmark	Haderslev, Jutland	Full blood	Nestling	na	2015	HG-B	B01
75	E1	Estonia	Spithami, Lääne	Full blood	Nestling	na	2015	HG-B	B03
76	E3	Estonia	Harju, Lääne	Full blood	Nestling	na	2015	HG-B	B03
77	E6	Estonia	Kiili, Lääne	Full blood	Nestling	na	2015	HG-B	B01
78	GL-10	Greenland	Nuuk	Full blood	na	na	2011	IG1	A01
79	GL-15	Greenland	na	Skin/muscle	Adult	na	>2013	IG1	A01
80	GL-16	Greenland	Nanortalik	Skin/muscle	Adult	na	2017	IG1	A01
81	GL-1	Greenland	na	Full blood	na	na	>1990	IG2	A03
82	GL-22	Greenland	na	Skin/muscle	na	na	1999	IG2	A03
83	GL-26	Greenland	Paamiut	Skin/muscle	na	na	2013	IG1	A01
84	GL-27	Greenland	Maniitsoq	Skin/muscle	na	na	2013	IG1	A01

85	GL-2	Greenland	Sisimiut	Full blood	Immature	na	1990	IG1	A01
86	GL-3	Greenland	Nuuk	Full blood	Adult	na	1993	HG-A	-
87	GL-4	Greenland	Ameralik	Full blood	Adult	na	1992	IG-1	A01
88	GL-7	Greenland	Nanortalik	Full blood	Immature	na	1996	IG1	A01
89	GL-9	Greenland	Nuuk	Full blood	na	na	1996	IG1	A01

51

52

Table 2. Overview over the six different scenarios and the subsequent substitution rates (in substitution per site per million years) used for divergence time estimation in BEAST. The body mass related substitution rate when using all coding positions was very similar for both calibrations sets. Therefore, only one was added as scenario 4 (thus "X" was not used). (References: Arcones et al., 2019; Nabholz, Lanfear, & Fuchs, 2016).

	Scenario 1	**Scenario 2**	**Scenario 3**	**Scenario 4**	**X**	**Scenario 5**	**Scenario 6**
Partition	**Standard Rate**	**Body mass related rate (Nabholz et al., 2016)**				**Rates for Accipitrifromes (Arcones et al., 2019)**	**Divergence To *Buteo buteo***
		3rd Codon Positions Calibration Set 2	**3rd Codon Positions Calibration Set 4**	All coding positions Calibration Set 2	All coding positions Calibration Set 4		
CYTB	0.0105	0.0084 (unpartitioned)	0.0079 (unpartitioned)	0.0023 (unpartitioned)	0.0022 (unpartitioned)	0.001242	0.0049
COX1	0.0082					0.001298	0.0041
COX2	0.0185					0.001051	0.0041
COX3	0.0089					0.000949	0.0037
ND1	0.0025					0.001842	0.0053
ND2	0.0042					0.002267	0.0057
ND3	0.0189					0.001576	0.0057
ND4	0.0025					0.001813	0.0045
ND4L	0.0154					0.001434	0.0061
ND5	0.0178					0.001451	0.0049
ND6	0.0145					0.002565	0.0069
ATP6	0.006					0.002284	0.0082
ATP8	0.0124					0.002641	0.0057
12S rRNA	estimated	not included	not included	not included	not included	0.001105	estimated
16S rRNA	estimated	not included	not included	not included	not included	0.000364	estimated

all tRNAs	estimated	not included	not included	not included	not included	estimated	estimated
Non-coding	estimated	not included	not included	not included	not included	estimated	estimated

Table S3. Assessment of quality (qual), depth, ratio of the common allele (comm.ratio) and binomial probability of high frequency segregating sites assuming equal proportions, per position (pos) within Iceland causing the major split in the network and tree (Figure 2 and 3).

pos	qual	depth	comm.ratio	pbinom
1857	224.559	91.103	0.991	0.000
1886	224.490	90.000	0.991	0.000
7144	225.007	92.552	0.988	0.000
7490	223.904	84.552	0.990	0.000
9394	224.145	86.828	0.992	0.000
12599	150.435	66.643	0.832	0.005
12899	150.221	87.250	0.842	0.001
13227	145.388	59.179	0.828	0.004
13283	164.364	54.429	0.831	0.009
13926	138.411	95.429	0.828	0.002
15481	224.283	91.655	0.994	0.000

Table S4. Nucleotide diversity and length in base pairs for each of the 13 genes analysed in this paper on the mitochondrial DNA of the white-tailed eagle, for all samples from Greenland, Iceland, Norway, Denmark and Estonia pooled. It is seen that diversity is spread all along the chromosome.

mtDNA gene	nucleotide diversity	length in base pairs
ND1	0,0003	978
ND2	0,0005	1039
COX1	0,0010	1551
COX2	0,0022	684
ATP8	0,0015	168
ATP6	0,0007	684
COX3	0,0011	784
ND3	0,0023	351
ND4L	0,0019	297
ND4	0,0003	1378
ND5	0,0022	1818
CYTB	0,0013	1143
ND6	0,0018	519

Table S5 Tests of selection for white-tailed eagle subset of samples from Iceland. Originally with 25 individual (table 4), here subset for comparison of strength of result, with 12 and 21 individuals. Tajima's D, the normalized Fay and Wu's H, and the E-test. 10.000 coalescence simulations. P-values in parentheses as proportions of number of simulations resulting in equal or lower values than the observed statistic.

Subset from Iceland #	Tajima's D	Fay and Wu's H	E-test
12	0.757 (0.751)	-1.698 (0.072)	2.38 (0.992)
21	0.861 (0.833)	-2.03 (0.045)	2.72 (0.994)

Arcones, A., Ponti, R., & Vieites, D. R. (2019). Mitochondrial substitution rates estimation for molecular clock analyses in modern birds based on full mitochondrial genomes. *BioRxiv*, 11. doi: 10.1101/855833

Hailer, F., Helander, B., Folkestad, A. O., Ganusevich, S. A., Garstad, S., Hauff, P., … Vilà, C. (2007). Phylogeography of the white-tailed eagle, a generalist with large dispersal capacity. *Journal of Biogeography*, *34*(7), 1193–1206. doi: 10.1111/j.1365-2699.2007.01697.x

Lisiecki, L. E., & Raymo, M. E. (2005). A Pliocene-Pleistocene stack of 57 globally distributed benthic δ 18O records. *Paleoceanography*, *20*(1), 1–17. doi: 10.1029/2004PA001071

Nabholz, B., Lanfear, R., & Fuchs, J. (2016). Body mass-corrected molecular rate for bird mitochondrial DNA. *Molecular Ecology*, *25*(18), 4438–4449. doi: 10.1111/mec.13780

Paper II

Paper II

Evaluation of four methods to identify the homogametic sex chromosome in small populations

In review with BMC Genomics

Author contribution: CCRH and SP designed the study; KMW prepared the RADseq libraries; CCRH and SP analyzed the data; CCRH and SP wrote the paper; KMW commented on the paper.

Evaluation of four methods to identify the homozygotic sex chromosome in small populations

Charles Christian Riis Hansen[1*], Kristen M. Westfall[1,2], and Snæbjörn Pálsson[1,3]

[1]Department of Life and Environmental Sciences, University of Iceland, Reykjavik, Iceland.

[2]Current: Fisheries and Oceans Canada, Pacific Biological Station, Nanaimo, BC, Canada.

[3]Senior Author

*Correspondence: CCRH

Abstract

Background

Whole genomes are commonly assembled into a collection of scaffolds and often lack annotations of autosomes, sex chromosomes, and organelle genomes (i.e., mitochondrial and chloroplast). As these chromosome types differ in effective population size and can have highly disparate evolutionary histories, it is imperative to take this information into account when analysing genomic variation. Here we assessed the accuracy of four methods for identifying the homogametic sex chromosome in a small population using two whole genome sequences (WGS) and 133 RAD sequences of white-tailed eagles (*Haliaeetus albicilla*): i) difference in read depth per scaffold in a male and a female, ii) heterozygosity per scaffold in a male and a female, iii) mapping to the reference genome of a related species (chicken) with annotated sex chromosomes, and iv) analysis of SNP-loadings from a principal components analysis (PCA), based on the low-depth RADseq data.

Results

The best performing approach was reference mapping (method iii), which identified 98.12% of the expected homogametic sex chromosome (Z). Read depth per scaffold (method i) identified 86.41% of the homogametic sex chromosome with few false positives. SNP-loading scores (method iv) identified 78.6% of the Z-chromosome and had a false positive discovery rate of more than 10%. Heterozygosity per scaffold (method ii) did not provide clear results due to a lack of diversity in both the Z and autosomal chromosomes, and potential interference from the heterogametic sex chromosome (W). The evaluation of these methods also revealed 10 Mb of putative PAR and gametologous regions.

Conclusion

Identification of the homogametic sex chromosome in a small population is best accomplished by reference mapping or examining differences in read depth between sexes.

Keywords

Homogametic sex chromosome, population genetics, non-model organisms, white-tailed eagle

Background

Inferences about genetic variation, effective population size and population structure from genomic data in species that have heteromorphic sex chromosomes are dependent on their correct identification and other markers from different genomic regions, i.e., autosomes and the plastid genomes. As these different genomic regions typically have different ploidy numbers, substitution rates, and recombination rates, it follows that they will also be variably affected by genetic drift and selection [1]. Such information can be imperative for successful conservation management based on genetic variation and evolutionary studies. It should though be noted that several species lack heteromorphic sex chromosomes, such as the majority of vertebrate species, including most fish, amphibians, and reptiles [2]. Annotating genomic regions can be accomplished either from a high-quality reference genome of the same species, a closely related species, or from the more computationally intensive and time-consuming method of de novo assembly. Here, we use genomic data from the white-tailed eagle mapped to a golden eagle reference genome to determine which scaffolds belong to the Z and autosomal chromosomes.

Using a different reference genome from the study species is frequently done [3–6] when chromosomal information is lacking and can be beneficial as references of the same species can affect genomic coverage, number of variants discovered, and heterozygosity estimates [7]. However, there can be drawbacks as related species can differ e.g., in genome size, synteny and other chromosomal rearrangements, or even lack sex chromosomes altogether [8]. Mapping to a closely related genome could also lead to mis-identification of a sequence that is sex-linked in the reference but not in the focal species or missing the sex chromosome content present in the focal that is not in the reference. However, birds are characterized by evolutionary stable chromosomes with rather little variation in genome size compared to other groups [9, 10].

Like many non-model species, the white-tailed eagle (*Haliaeetus albicilla*) lacks a well annotated genome. The specimens studied here come from a small and geographically isolated population of

white-tailed eagles in Iceland, which currently consists of 80 breeding pairs. The population is recovering slowly from a severe bottleneck in population size during the 19th-20th centuries, when the number of breeding pairs were about 20 for more than 50 years [11] and is thus expected to have little genetic variation. The golden eagle (*Aquila chrysaetos*) and the white-tailed eagle are large raptors with a wide distribution in the northern hemisphere [12, 13]. Currently there are four genome assemblies available for the golden eagle, consisting of 142 (size: 1,233.7 Mb, N50: 46.9 Mb); 1,142 (size: 1,192.7 Mb, N50: 9.2 Mb); 35,366 (size: 1,196.0 Mb, N50: 0.11 Mb); and 42,881 (size: 1,548.4 Mb, N50: 1.7 Mb) scaffolds, where only the first has scaffolds assigned to chromosomes [14]. Only three fragmented genomes exist for the white-tailed eagle (consisting of 50,905 scaffolds with the size: 1,133.5 Mb, and N50: 0.05 Mb; 35,313 scaffolds with the size: 1,196.5 Mb and N50: 0.12 Mb; and 6,418 scaffolds with the size: 1,222.6 Mb and N50: 4.5 Mb), with no annotated chromosomes [10]. The mitochondrial genomes of both the white-tailed and golden eagle have been identified [14, 15]. The Z-chromosome has been identified in golden eagle (88.2 Mb) and it is large in comparison with Z chromosomes in other birds which have been identified (ranging from 37.9 to 195.3 Mb [16–18]) but similar in size to chicken (*Gallus gallus*, 82.5 Mb [19]). Resolving the chromosomal composition of the white-tailed eagle genome will facilitate research on the genetics and evolutionary history of the species and for other eagle species. Furthermore, assessing the accuracy of methods for identifying the homozygotic sex chromosome may facilitate annotation of genome assemblies in other species characterized by small population sizes. Here we evaluate the success of four methods to identify the Z-chromosome in the small population of white-tailed eagles in Iceland: 1) sequencing depth, 2) patterns of heterozygosity in high-depth whole genome sequence data from one male and one female, 3) mapping the golden eagle scaffold reference genome to the chicken genome, and 4) a PCA of genotypes from low-depth RAD-sequencing data from 133 white-tailed eagles. Our hypothesis is that the use of heterozygosity will be least successful as it will be reduced in the small population.

A recent review describes various methods for identifying sex chromosomes [20]. When template DNA molecules from a genome are sequenced randomly, it is expected that equivalent chromosomal classes will have similar average sequencing depths, and thus the depth can be used to identify different parts of the genome. For example, mitochondrial DNA is expected to have relatively high read depth, due to greater per-cell copy number than the nuclear chromosomes (this also applies to repeated regions). In addition, the sex chromosome found in the homogametic sex (ZZ or XX) is expected to have double the sequencing depth obtained from the heterogametic sex (ZW or XY), in species with differentiated sex chromosomes, as in birds and mammals [21, 22], but not in species with little differentiation between sex chromosomes such as in several fish species [23, 24]. Thus, for example, identification of the Z (and X) chromosome through depth filtering has been successfully applied to flycatchers [25] and humans [26], and depth is also partly used in programmes for discovering the sex chromosomes [27–29].

Sex differences in heterozygosity can also be used to assess which scaffolds belong to the homogametic sex chromosome e.g., [30]. For any given set of individuals from the same population, the Z-chromosome is expected to have fewer heterozygous positions in females (ZW) than in males (ZZ), whereas autosomal scaffolds are expected to have a similar number of heterozygous positions in both sexes. However, several factors can limit the discriminatory power of heterozygosity to identify Z scaffolds when comparing males and females. First, the difference between the sexes will be reduced for scaffolds containing pseudoautosomal (PAR) and gametologous regions (conserved but non-recombining homologous regions). A study on PAR-regions in birds have shown large variation in the size and divergence of W- and Z-chromosomes across species [31], furthermore Xu and Zhou [21] showed that the W-chromosome has retained its gene function in birds better than the Y-chromosome in mammals and that the proportion of gametologs can be high. Moreover, long runs of homozygosity affecting Z scaffolds in males and autosomal scaffolds in both sexes, due to inbreeding or small population size, can mask the expected pattern of sex differences in

heterozygosity. This is expected to be a marked feature of the white-tailed eagles analysed in this study and have a negative impact on how useful the heterozygosity is in identifying the Z-chromosome.

Another approach is to map scaffolds from an incompletely assembled reference genome to a more fully annotated genome from a "closely" related species. Such mapping can be done with several available programs e.g., LASTZ [32], LAST [33] and YASS [34]. The accuracy of chromosomal locations of scaffolds obtained from this approach depends on the evolutionary distance between the two reference genomes, which can differ due to chromosomal translocations, transposed regions, and repetitive regions [35, 36], sometimes even in closely related species [37]. Thus, this method may be only applicable for taxa with relatively stable genomes such as mammals and birds, though some groups of birds have also recently been shown to have dynamic sex chromosomes [38].

In a PCA of genotypes from all scaffolds i.e., belonging to both autosomes and sex-chromosomes, it is possible that one or more principal components (PCs) split males and females, due to sex specific markers on the sex chromosomes, i.e., on W or to markers on Z given a double weight in females. It therefore follows that a PCA could be used to identify scaffolds belonging to sex chromosomes, or alternatively to any sex specific markers, much in the same way as for population or group differentiation. Methods based on sex specific markers have been developed [39, 40] to identify the W and have been commonly used in PCR to diagnose sexes [41]. We tested this by examining the loadings of SNPs from a PCA based on low-depth RAD-sequencing data from 133 white-tailed eagles (Figure S1) - to assess if they contribute to separation along a specific principal axis [42] by sex.

We show that sex differences in sequencing depth and mapping to a more complete reference genome from a related species provide the most effective means to identify Z chromosome scaffolds in the white-tailed eagles. However, the approaches based on the PCA, and heterozygosity

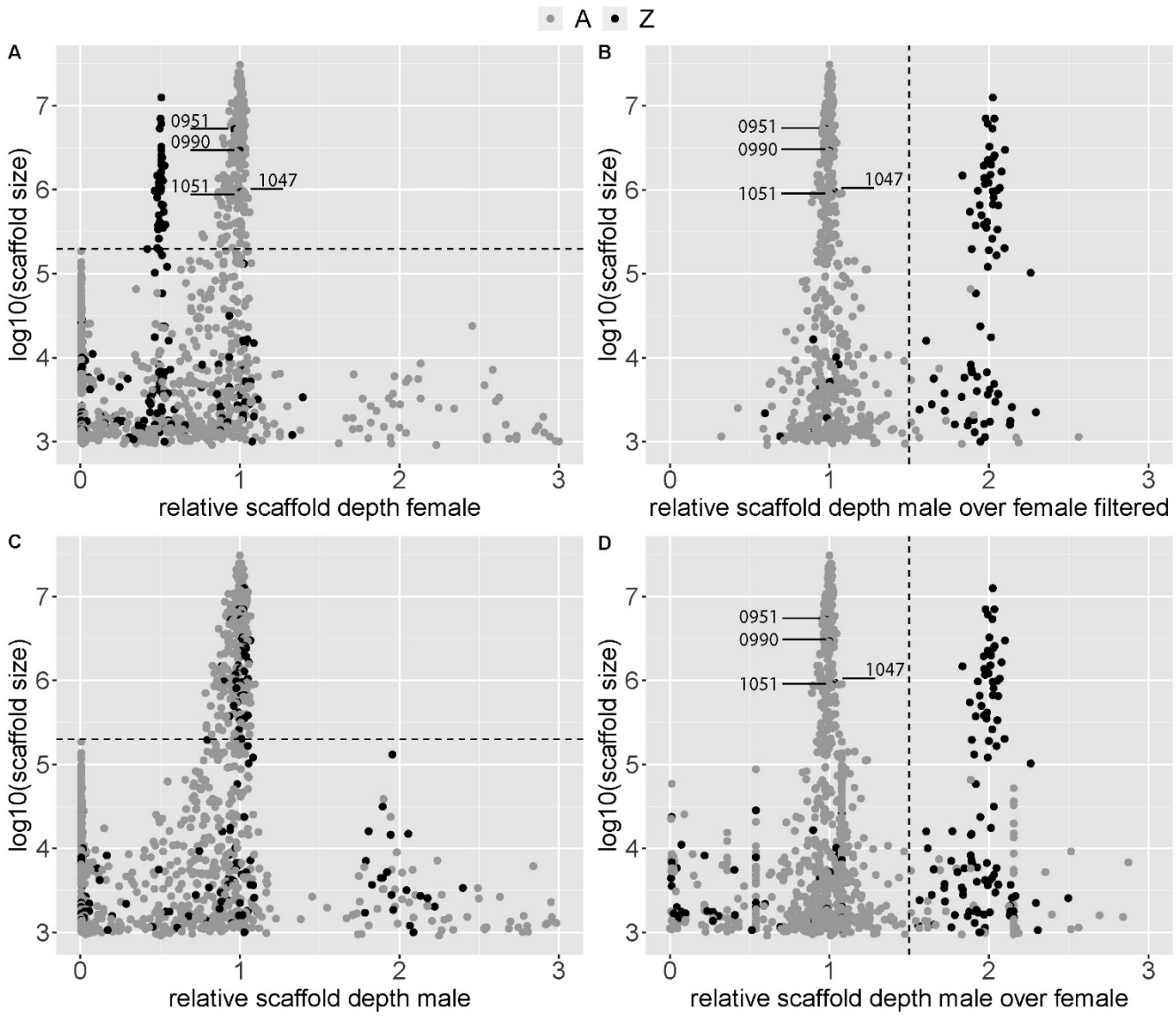
provide valuable additional information and shed light on some key challenges faced by researchers working with genomic data from species with partially assembled reference genomes.

Results

To assess the accuracy of the four approaches used to identify the Z-chromosomal scaffolds (depth, Heterozygosity, mapping, and PCA), the reference "scaffold-assembled" golden eagle genome was mapped to a newly released "chromosome-assembled" golden eagle genome, to know the position in the genome of the scaffolds. This was used as a baseline ("truth") when evaluating the methods (Figure 5, Table 2, and method section).

Depth. The overall modes of depth, 195x for the female and 181x for the male, were used to estimate the relative sequence depth for each position on each scaffold. A clear bimodal distribution of the depths was observed after discarding the shortest scaffolds (<198,789 bases, log10 < 5.29) (Figure 1, Figure S2, Figure S3) and a good distinction of the expected values for the Z-chromosome (0.5) and the autosomes (1) for the female was observed (Figure 1A and S2). As also expected, this was not observed for the male, but a few Z scaffolds had a ratio of 2 suggesting occurrence of paralogous regions (Figure 1C). After the removal of the short scaffolds, 257 scaffolds out of the 1141 scaffolds remained, but covering 98.9 % of the full genome in the chromosome-assembled golden eagle genome, which was used as baseline. In the female, 36 scaffolds comprising ~75.2 Mb had a relative depth close to 0.5 (from 0.466 to 0.533), all from the golden eagle Z-chromosome. In comparison, 211 scaffolds (1.0947 Gb) had a relative depth around 1 (from 0.764 to 1.062), whereof 207 were autosomal. The remaining four scaffolds (NW_011950951.1, NW_011950990.1, NW_011951047.1 and NW_011951051.1) mapped to the Z chromosome, comprising ~10 Mb or 0.91% of the scaffolds identified as autosomes (see Table 2 and Table S1 for all numbers).

159 The expected male to female ratio (r_{mf}) of sequence depth is 1 for autosomal and 2 for Z scaffolds.
160 Implementation of r_{mf} for the scaffolds revealed an even clearer split between the Z and the
161 autosomes (Figure 1B), particularly after removing the primarily small scaffolds with relative depth
162 outside the credible range of 0.25-1.5 in either the male or female. This left 618 scaffolds that
163 accounted for 99.53% of the total sequence (Figure 1D). Thereof 93 had $r_{mf} > 1.5$, consistent with
164 the expected depth of Z scaffolds. Of these, 79 (76.2 Mb) identified as Z and 14 (0.09 Mb) as
165 autosomal chromosomes in the golden eagle genome. We observed 525 scaffolds with $r_{mf} <= 1.5$,
166 consistent with the expected depth of autosomes. Of these, 512 scaffolds (1,100.7 Mb) identified
167 as autosomes and 13 (10.05 Mb) as Z in the golden eagle genome (four of these 13 were also
168 scaffolds NW_011950951.1, NW_011950990.1, NW_011951047.1 and NW_011951051.1).



169

170 *Figure 1. Relative sequencing depth of scaffolds in a female and a male white-tailed eagle. Relative scaffold depth was*
171 *estimated as mode of scaffold depth / overall genomic depth, which was 195 for the female and 181 for the male. The*
172 *shading of the dots, representing scaffolds, refer to whether they map to the Z or autosomal (A) chromosomes in the*
173 *golden eagle genome with known chromosomes. A) Relative depth in the female. B) The male to female ratio (r_{mf}) of*
174 *relative scaffold depth after filtering (removing scaffolds with relative depth outside the range of 0.25-1.5 in either the*
175 *male or female). C) Relative depth in the male. D) The male to female ratio (r_{mf}) of relative depth for all scaffolds. In A and*
176 *C the dashed line represents the scaffold size threshold value of 198,789 bases (log10 5.29). In A and C, points lower than*
177 *the threshold value of 198,789 bases displayed high variation for relative depth (Figure S2). Scaffolds below the threshold*
178 *in A and B make up 1.1% of data, only 0.0071% is below the threshold and above a relative depth of 3. Dashed line in B*
179 *and D is 1.5, which is right between expectation for autosomal (1) and Z chromosomes scaffolds (2). "0951", "0990",*

"1047", and "1051", in A, B, and D, refers to the scaffolds NW_011950951.1, NW_011950990.1, NW_011951047.1 and NW_011951051.1.

Heterozygosity. Only 32% of scaffolds (365 of 1,141), covering 97.5% of the genome, had at least one heterozygous genotype after filtering in either of the two individuals, with slightly fewer in the female (288) than in the male (300). The majority of the scaffolds with no heterozygous sites mapped to the Z (80% in the female, corresponding to 30% of the Z chromosome; 77% in the male, covering 23% of Z). The Z had generally fewer heterozygous sites after filtering (Table 1, Supplement Figures 3 and 4), but a majority of the autosomal scaffolds lack heterozygous sites (67%, 1.1% in size). Furthermore, there were more autosomal scaffolds than Z's. Seventy-seven scaffolds (52.5 Mb, ranging from 1.5-5,565 kb) had no heterozygous genotypes in the female but a minimum of one heterozygous genotype in the male and ten of those scaffolds (10.1 Mb) mapped to the Z-chromosome in the golden eagle genome. Aside the larger fraction of the Z scaffolds which had no variation on Z, about 62% of the Z-chromosome in the female had also considerably fewer heterozygous sites than the male (supplement Figure 3), but some show autosomal levels of heterozygosity in the female (separately marked in Figure 2A). Four of these scaffolds also exhibited autosomal levels of depth in the female (Figure 1) and two of those scaffolds ("NW_011950951.1" "NW_011950990.1") in the female had the highest number of heterozygous sites (1823, 5568), followed by NW_011951047.1 which had 450 sites.

Table 1. Information about heterozygosity for a female and male. Heterozygosity for each of the male and female for scaffolds that map to the A and Z in the golden eagle genome with known chromosomes. Numbers of heterozygous (hets.) sites, scaffolds and windows of size 50,000 bases. Total number of scaffolds and 50k windows were 1,141 and 23,585 respectively.

	Female Z	Female A	Male Z	Male A
Proportion of heterozygous sites before filtering	0.00534	0.00067	0.00050	0.00065
Proportion of heterozygous sites after filtering	0.00010	0.00018	0.00007	0.00019
Scaffolds with no heterozygous sites	134 (80%)	720 (74%)	130 (77%)	712 (73%)
Size of scaffolds with no heterozygous sites (kb)	26,625 (31%)	55,018 (5%)	20,010 (23%)	65,907 (6%)
Scaffolds with heterozygous sites	34	254	38	262
Heterozygous sites per window (50kb) (median)	0	1	0	2
Standard deviation per window (50kb)	43.0	12.3	8.1	12.2
Coefficient of dispersion (CD)	360.6	16.1	20.2	15.6
Windows with no heterozygous site	1,398	10,264	1,267	9,857
Windows with heterozygous sites	304	11,619	435	12,026

The four Z chromosomal scaffolds that had a male-like pattern of autosomal depth and heterozygosity in the female were further analysed in windows of 50Kb, as heterozygous sites can be restricted to small parts of the scaffold (Figure S6). An examination of the number of filtered heterozygous sites per 50Kb window in these four scaffolds in the female showed that NW_011950951.1, NW_011950990.1 consisted of either 1 or 2 continuous regions, whereas the other two were more fragmented.

The average heterozygosity per scaffold, prior to filtering, was >10-fold higher in the female than the male for the Z-chromosome (Table 1), and several scaffolds were even higher (Figure 2B). The

filtering removed most of this excess heterozygosity in the female (Figure 2C, D and E). As the pattern of excess heterozygosity in the female was primarily seen in Z rather than autosomal scaffolds, we postulate that these instances might represent the mapping of diverged homologous reads from the W chromosome.

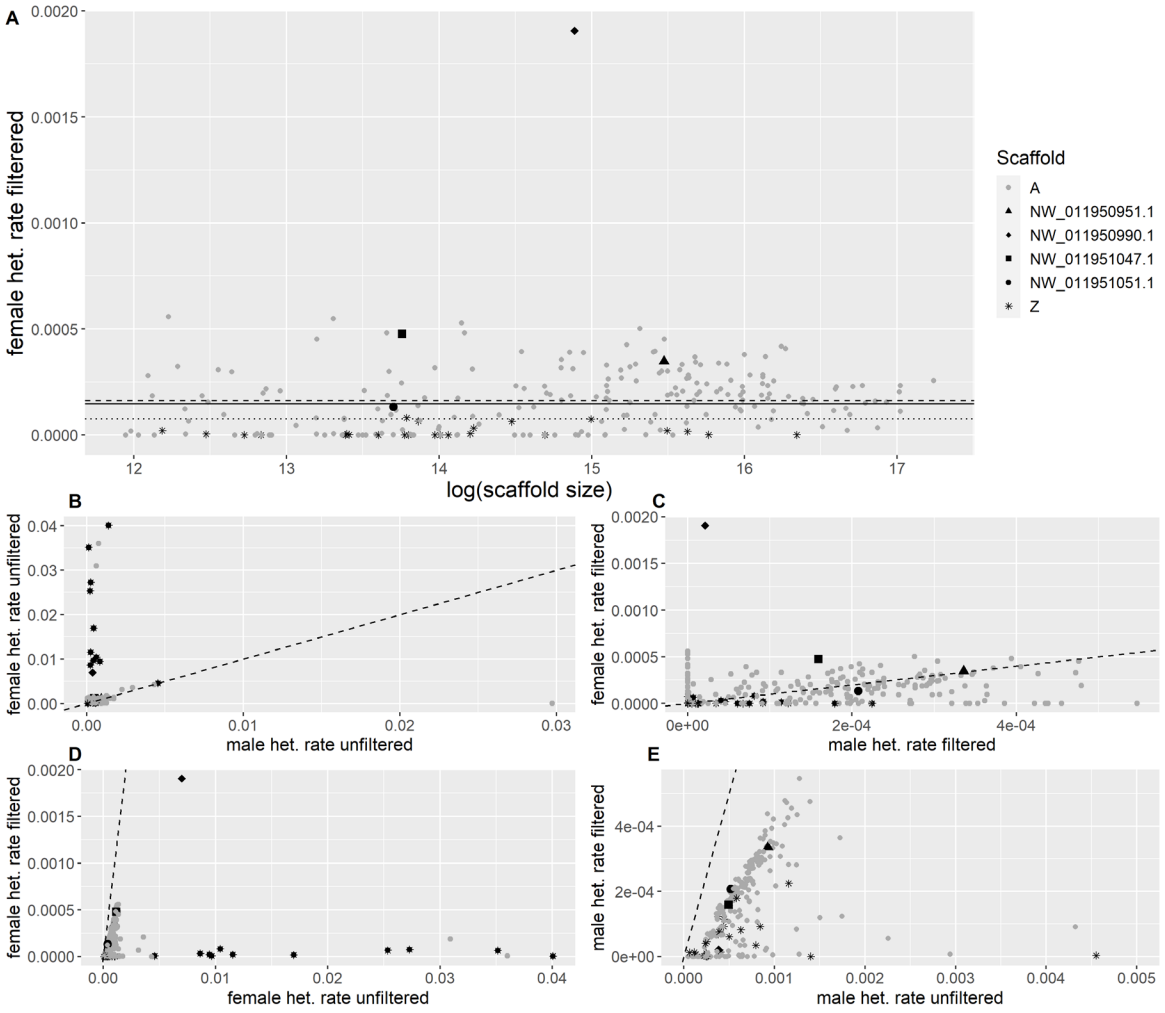


Figure 2. Heterozygosity rate of scaffolds mapped to autosomes and the Z-chromosome in white tailed eagle. Heterozygosity rate as number of heterozygous positions divided by length. A) filtered heterozygosity rate of the female. The dashed and dotted lines represent the mean filtered heterozygosity rate for autosomal and Z scaffolds, respectively. The full line is for all scaffolds. B) unfiltered heterozygosity rate for the female plotted against the male. C) filtered heterozygosity rate for the female plotted against the male. D) filtered versus unfiltered heterozygosity rate in the female. E) filtered versus unfiltered heterozygosity rate in the male. In all plots shape and color reflect scaffold type; grey dot is autosomal; black star Z-chromosomal; triangle, diamond, square and black dot are scaffolds NW_011950951.1,

NW_011950990.1, NW_011951047.1, NW_011951051.1, respectively, which all show high heterozygosity in the female and have a relative depth as being autosomal. In B through E, dashed lines represent the identity line (slope=1).

Overall, the distributions of heterozygous sites per window was similar for the male and the female and almost half of the windows had no heterozygosity (49% in the female and 47% in the male). When the windows were grouped by Z and autosomes, a difference between the sexes was observed for the Z-chromosome (Table 1 and Figure S4 and S5). As expected, there was a higher proportion of windows on Z with no heterozygous sites in the female (82%) than in the male (74%) ($P = 6.111*10^{-8}$, Fishers exact test). However, the 10 most variable 50kb windows in the female, with rate of heterozygous sites ranging from 0.17-1.73% all came from the scaffold NW_011950990.1 which map to Z. The window in the male with the largest rate of heterozygous sites had 0.15%. This difference in the distribution of heterozygosity per 50 kb windows on the Z chromosome per sex is also reflected in the average number and standard deviation of heterozygous genotypes per window, which was larger in the female Z (5.1 and 43) than in the male Z (3.2 and 8.1), whereas no differences were observed in these descriptive statistics for the autosomes. This means that the distribution of heterozygous genotypes was more clumped for Z in the female (Coefficient of dispersion, CD=360.5) than in the male (20.2) and the autosomes of both sexes (~16) (Table 1).

Mapping

Mapping the 1141 scaffolds from the golden eagle scaffold assembly to the chicken genome, using LASTZ, resulted in 110 scaffolds (86.5 Mb) correctly assigned to the Z-chromosome, and 940 scaffolds correctly assigned to autosomes, according to the comparison of mapping to the golden eagle chromosome-assembled genome. On the other hand, 33 scaffolds (0.59 Mb, amounting to 0.69% of the total length of scaffolds) were wrongly assigned to the Z-chromosome, and 58 scaffolds (0.27 Mb, 0.024%) were wrongly assigned to autosomes (Table 2).

PCA

The analysis of the loadings of 164,952 SNPs from the PCA analysis (Figure S1), based on 133 RADseq individuals with an average sequencing depth per site of 2.25 per individual, was limited to the 280 scaffolds (40 Z and 240 autosomal) that had more than 50 SNPs (accounting for 98.3% of the genome). We calculated the 95% range of SNP-loadings on PC1 (i.e., the quantiles 0.025 and 0.975) in our attempt to identify scaffolds belonging to the Z, using a threshold (0.1006) that corresponds to 3 standard deviations above the mean (Figures 3A and 3B, Table 2). Of the scaffolds included in this analysis, 28 (78%) scaffolds from the Z-chromosome were above this threshold, accounting for 69.3 Mb (83.6% of the total length of Z scaffolds used in this analysis). In contrast, only 9 (3.75%) of the autosomal scaffolds were above the threshold, amounting to 11.7 Mb (1.1% of the total length of autosomal scaffolds used in this analysis). Thus, the range of PC1 loadings provides some discriminatory power to distinguish Z from autosomal scaffolds.

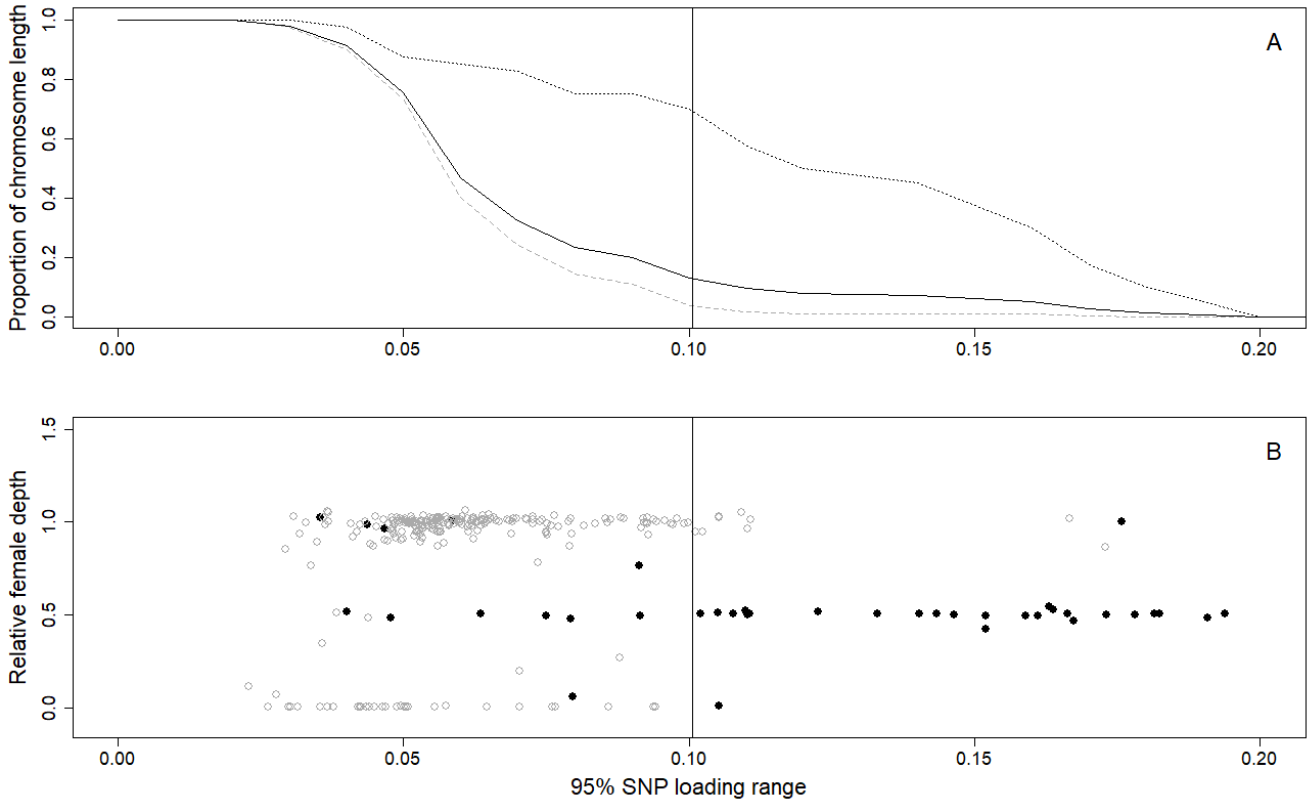


Figure 3. Discrimination of the sex-chromosomes with respect to SNP-loading and genomic information. A) Proportion of the genome plotted against 95% range of SNP-loading values for PC1. Dotted black line presents the proportions for the Z-chromosomal scaffolds, dashed grey line the values from the autosomal and full black line the values for all scaffolds pooled. B) Relative scaffold depth for the female plotted against the 95% range in SNP-loadings. Open grey circles refer to autosomal scaffolds and full black dots to Z-chromosomal scaffolds. The vertical line, in both plots, represents 3 SDs above the mean. For legibility, the upper value on the y axis was set to 1.5 in panel B. Two scaffolds, one autosomal and one Z-chromosomal, had relative depth greater than 1.5 (both >15), with a SNP loading range around 0,025 and 0,15, respectively.

Comparison of the four methods.

Table 2. Classification of scaffolds identified as Z or autosomal scaffolds. Classification for each of the approaches: depth, heterozygosity, LASTZ, and SNP-loading analysis. The identification was found by comparison to the golden eagle genome bAquChr1.2 (GCA_900496995.2) with known chromosomes. Results for the different methods are given in a) for total size of scaffolds (bp), and in b) for the number of scaffolds, missing is compared to the golden eagle scaffold assembly.

		Depth		Heterozygosity		LASTZ		SNP-loading	
a)		Z	A	Z*	A	Z	A	Z	A
	Z	76,239,124	10,056,095	-	60,214,856	86,569,008	270,522	69,355,267	13,642,226
	A	93,786	*1,100,765,118*	-	1,050,885,219	597,603	1,105,305,943	11,720,756	1,078,283,284
	Total		*1,187,154,123*		*1,159,757,217*		*1,192,725,744*		*1,173,001,533*
	Missing		*5,571,621*		*29,104,198*		*0*		*19,714,211*
	Z	79	13	-	34	110	58	28	12
b)	A	14	512	-	254	33	941	9	231
	Total		618		*365*		*1,141*		*280*
	# NA		*523*		*776*		*0*		*861*

**values not assigned due to lack of heterozygosity on the Z chromosome*

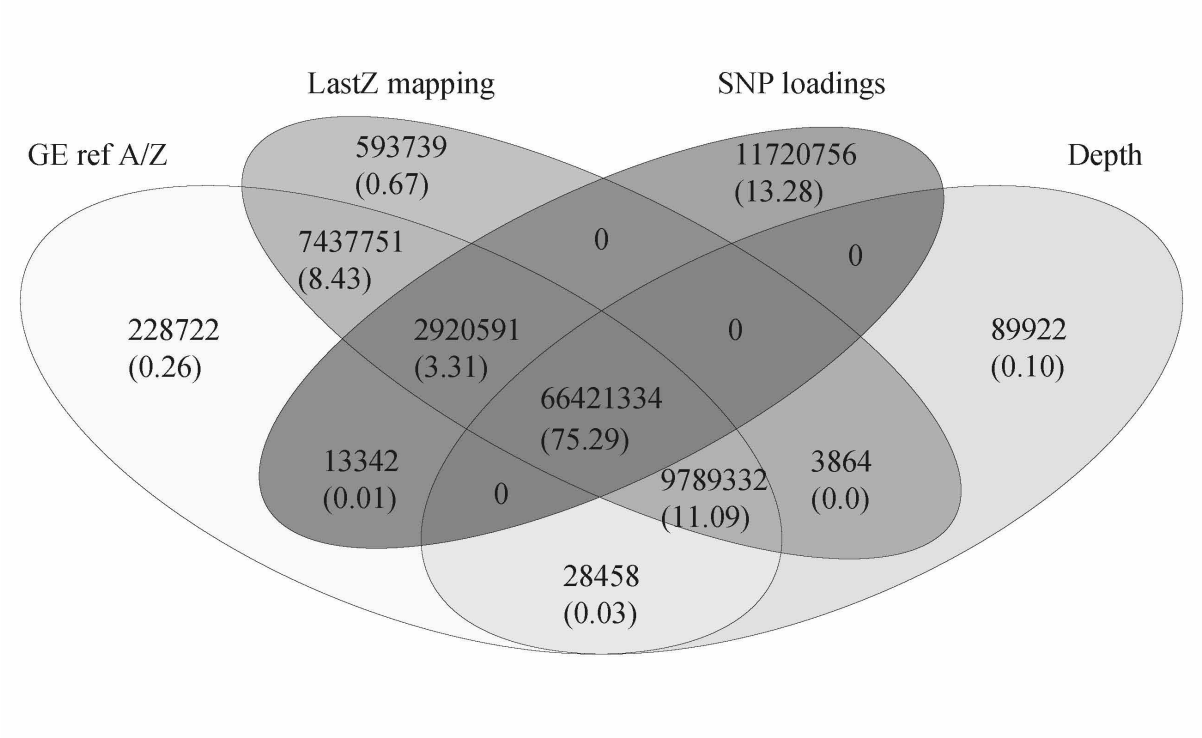


Figure 4. Venn diagram summarizing the size of scaffolds in bases identified as Z-chromosome with the three different analyses: mapping, depth and SNP-loadings. The Z-chromosomal scaffolds were assigned by mapping the genome with scaffolds to the genome with known chromosomes. Values in parentheses represent percentage size compared to the size of the known Z-chromosome. Notice that the percentage found by mapping the golden eagle scaffold assembly to the golden eagle genome is only 98.42%.

Using chromosome assignments obtained by mapping the golden eagle scaffold assembly to the golden eagle genome with assigned chromosomes, the most successful method was mapping to the chicken genome, finding 98.12% of the expected size (Table 2, Figure 4). In second place was the depth analysis with 86.41% and, in third, the SNP-loading with 78.61%. Heterozygosity was poorly suited to find Z-chromosomal scaffolds as a large fraction of scaffolds had no variation, and some Z-chromosomal scaffolds were found to be highly variable in the female (likely due to the mapping of reads that belong to the W chromosome). Depth, mapping to the chicken and SNP-loading all found false positives, i.e., autosomal scaffolds that were categorised as Z-chromosomal scaffolds (0.09, 0.59 and 11.72 Mb, respectively). All approaches resulted in false negatives i.e., Z-

296 chromosomal scaffolds categorised as autosomal (Table 2), but least with mapping to the chicken
297 (0.27 Mb), whereas depth, heterozygosity, and SNP-loading had 10.05, 60.21 and 13.64 Mb of false
298 negatives, respectively. Forty-five very short Z-chromosomal scaffolds (with a total length of 0.22
299 Mb) were not found by any analysis, and were only found when the golden eagle scaffold assembly
300 was mapped to the golden eagle with known chromosomes. Mapping of the golden eagle scaffold
301 assembly to the golden eagle with assembled chromosomes revealed 98.42% of the whole known
302 Z-chromosome (Table 2, Figure 4). Though the goal of the study was to evaluate the approaches
303 separately, a combined analysis (Figure 4) where at least two of three approaches (e.g., depth,
304 mapping to the chicken, and SNP-loading) were compared, detected between 75.29-86.29% of the
305 size of the Z-chromosome of the golden eagle genome, and only the approach combining depth
306 and mapping to the chicken found false positives, which was less than <0.01% of the size of the
307 golden eagle Z-chromosome.

308

309 ## Discussion

310 Three of the four methods evaluated in this study; the relative depth, mapping to chicken, and SNP-
311 loadings, were able to detect a high fraction of the Z-chromosome of the white-tailed eagle that
312 had been mapped on the golden eagle scaffold assembly. The success of the methods varied as
313 they may be affected differently by the small population size of the study species. The approaches
314 applying heterozygosity and PCA are expected to be more affected by a small population because
315 they analyse genomic and population diversity, whereas depth and mapping are expected to be
316 less affected by the low diversity in a small population.

317 The mapping of contigs to genome sequences from a distantly related species such as golden eagle
318 to chicken can be problematic due to architectural changes such as translocations and inversions.
319 Minor mismatches, e.g., transposable elements and mutations, may further impact the success of
320 finding the Z-chromosome. However, sex chromosomes may be well preserved in birds e.g., Xu and

Zhou [21], and this effect seems to be minimal in the case of mapping the golden eagle scaffold assembly to the chicken with a split time >80 million years [6].

The Z scaffolds that were not detected using the SNP-loading approach are likely due to parts of the Z-chromosome that lack variation, or that share homologous regions in the distinct sex chromosomes and do not contribute to the difference between the sexes in the PCA-plot. The PCA approach found few false positives, possibly due to the lack of a precise distinction between the range of loadings observed for the autosomal and Z-chromosomal scaffolds. Considering the information from the mapping it is clear that the Z-scaffolds have higher impact, as most false positives were just above the threshold of three SDs (i.e., 0.10 95% SNP loading range), and only two autosomal scaffolds were larger than ~0.11 comprising only a total size of 1.73 Mb, or 14% of the false positives. The SNP-loading approach also found false negatives (Table 2) and we feel this deserves further research.

Here, the approach of looking at all scaffolds in a single PCA was used, but this could potentially be optimized by using sliding windows [43] to identify signals different from the overall population signal. However, this also requires diversity on the homogametic sex chromosome in males compared to females, which may be lacking in small populations such as in the Icelandic white-tailed eagle.

Inspection of the heterozygosity for all scaffolds revealed that it is difficult to distinguish between autosomal and Z-chromosomal scaffolds without any prior knowledge. However, there was a difference in the average heterozygosity between autosomal and Z-chromosomal scaffolds, especially in the female. Small populations, such as the white-tailed eagles in Iceland [11], have reduced heterozygosity and long runs of homozygosity were observed on the Z-chromosome and the autosomes, making it more difficult to distinguish among the chromosomal types. Furthermore, there is a clear overlap in scaffolds with some heterozygosity which might belong to PAR and non-homologus regions, e.g., due to inversions, on the Z- and W-chromosomes. PAR and the

nonrecombining homologous regions, could explain some deviations in the prediction of the Z-chromosome in the SNP-loading analysis but these regions are probably small, and thus won't display the signal of an autosome in the depth analysis. Although genome wide information from a single individual can provide assessment of variation within populations, it can be biased due to missing chromosomal fragments and thus the overall success of the method. However, the two high depth individuals here show no clear indication of such deviation, as we obtain most of the Z-chromosome in the analysis.

The relative depth analysis revealed 86.41% of the expected size of the Z-chromosome and found few false positives. Four scaffolds were noted as false negatives in one of the two depth analysis. These four scaffolds (NW_011950951.1, NW_011950990.1, NW_011951047.1, and NW_011951051.1) make up about 10 Mb and show the highest heterozygosity of all Z-chromosomal scaffolds after filtering; their levels are comparable or even higher than observed for the autosomal scaffolds. Three of the four scaffolds showed low 95% SNP-loading ranges (all around 0.05), unlike the scaffolds contributing to the separation of the sexes. One scaffold (NW_011950990.1) had a very high 95% SNP-loading range and very high heterozygosity. This signal in these four Z scaffolds, and position at the end of the Z-chromosome supports that they belong to the pseudo-autosomal regions (PAR) as seen in other birds [31, 44]. In birds, PAR vary greatly in size from just a few Mb to more than 60 Mb [31]. Alternatively, they could represent non-recombining homologous regions (gametologs) [21, 45] which can be expected to have even higher heterozygosity in females than within the recombining Z-chromosomes in the homogametic males or the autosomes, because such regions could have evolved independently for millions of years. Two of the four scaffolds mentioned above, NW_011950990.1 and NW_011951051.1, display a higher heterozygosity ratio in the female compared to the male (17, and 2.5 times higher, respectively), as expected for gametologous regions, whereas the other two NW_011950951.1, and NW_011951047.1, may present PARs, as they display a ratio close to one between the sexes (1.08

371 and 0.78, respectively). A fully annotated genome of the white-tailed eagle would provide further
372 information about these gametologous regions within the Z- and W-chromosome.

373 Although depth analysis has shown to be a promising method to identify sex chromosomes [25,
374 46], it is not error free. Scaffolds belonging to the Z-chromosome can have a depth as high as
375 autosomes, as variance in depth can be large in small scaffolds which may be poorly sampled due
376 to low variation, or the scaffolds include regions from both Z- and W-chromosomes i.e., gametologs
377 and the PAR regions. Here the best approach for identifying the homogametic sex chromosome
378 was mapping to a reference with annotated homogametic sex chromosome. . To identify the Z-
379 chromosome, a combination of the mapping with at least one other analysis is recommended as it
380 may result in fewer potential false positives and negatives. Further, it should be noted that the
381 methods used here maybe more applicable in taxa with relative stable sex chromosomes, such as
382 mammals and birds [21, 22], but less effective in taxa such as fish where the sex chromosomes can
383 be less differentiated [23, 24].

384 The dynamic nature of the Z-chromosome (e.g., songbirds [38, 49]) and potential deviations in
385 synteny may introduce errors into assemblies of two species, however, there is significant and
386 relevant justification for doing so. Firstly, the approach using a different reference from the study
387 species has successfully been employed in other studies[3–5]. Secondly, the use of a different
388 reference species has been recommended because using references of the same species has been
389 shown to potentially introduce errors in the analyses [7]. Finally, using our novel white-tailed
390 sequences against the golden eagle assemblies, with one scaffold assembled, and one chromosome
391 assembled, made it possible to evaluate the precision of these approaches to a greater extent. This
392 study highlights potential problems when trying to identify the homogametic sex chromosome that
393 are specific to small populations, which bears importance for the conservation of species at risk.

394 Even though all known eukaryote species may soon be sequenced [50], it will still be a long time
395 before all parts of their chromosomes have been identified. Thus, it is important to further explore

these different methods and how they depend on sequence variation and scaffold sizes, as variation in the different chromosomes will differ due to different effective population sizes and evolutionary histories.

Conclusion

The best performing approach for identifying the homogametic sex chromosome in a small island population of white-tailed eagle was reference mapping to a related species. The second-best approach was analysis read depth per scaffold, and thirdly, SNP-loading in PCA. Identification using genomic diversity approaches; SNP-loading and heterozygotic differences between sexes are potentially affected by the small population size and a recent population bottleneck.

Evaluation of these methods are highly relevant as genomic regions vary in effective population size and can have different evolutionary histories. Furthermore, the use of a different reference genome to the study species is still a widely used approach, which has several upsides.

Methods

Sample collection, laboratory work and sequencing

Blood samples were collected from white-tailed eagle chicks as a part of an ongoing monitoring program in Iceland since 2001 by the Natural History Institute of Iceland. The sex of the chicks was determined in the field based on tarsus thickness and weight [51]. Three to ten mL of blood was extracted from each chick. The blood was stored in EDTA buffer at -20 degrees Celsius until DNA extraction.

DNA from blood samples of 135 chicks was extracted using the ThermoFisher GeneJET Whole Blood Genomics DNA Purification Mini Kit following the standard protocol [52]. DNA concentration was

estimated using the NanoDrop 1000 and run on 0.7% agarose gels to evaluate the fragment size. Samples with concentration higher than 60 ng/µl were selected for library preparation and sequencing. The 133 of 135 extracts were double digest restriction-site associated DNA sequenced (RADseq) on the Illumina HiSeq2500 (see supplementary text 1 for full description).

A male and female white-tailed eagle were selected for high-depth whole genome shotgun sequencing with two lanes each on an Illumina HiSeqX. Library preparation and sequencing was done at deCODE genetics, using the TruSeq Nano sample preparation method [53].

Two reference assemblies from male golden eagles (ZZ), one in 1142 scaffolds and one assembled to chromosome level (GenBank Assembly Accession numbers: GCA_000766835.1 and GCA_900496995.2, respectively), and a female chicken assembly (ZW) (GenBank Assembly Accession: GCA_000002315.3) were downloaded from NCBI and used in the analysis [14, 54].

Sequence cleaning and mapping

The white-tailed eagle RADseq data was demultiplexed, sorting sequence reads into individual files, both for forward and reverse sequences using the command 'process_radtags' in Stacks version 1.47 [55, 56]. Default settings were used for the RADseq data, applying the option "r" to rescue barcodes and RAD-tags.

After demultiplexing, FastQC [57] was run for quality control. For the RADseq data, an excess of specific sequences (kmers) were removed using AdapterRemoval v2 (version 2.2.2) [58]. The high depth shotgun sequenced individuals were tested in the same way but found no excess of kmers.

The Burrows-Wheeler Aligner (BWA) mem and SAMtools version 0.7.17-r1188 and 1.7, respectively [59, 60] were used to process RADseq and high depth shotgun data and map reads to the golden eagle scaffold assembly of 1142 scaffolds with no identified chromosomes (GCA_000766835.1) [14] using default settings in both instances.

Four different approaches to find the Z-chromosome - Depth, Heterozygosity, Mapping and SNP-loadings

Four different approaches were used to identify scaffolds in the white-tailed eagle genome belonging to the Z-chromosome, by comparison with the golden eagle scaffold assembly with no chromosomes (GCA_000766835.1). An assembly consisting of 1,141 assembled scaffolds, excluding mtDNA, and a total of 1,192,725,744 bp, ranging in size from 913 to 30,727,332 bp with a median of 5,587 bp, and average length of 1,045,334 bp (SD 3,203,066 bp). An overview of the methods is presented in Figure 5 and the data used in each analysis is available in supplementary Table S1.

Depth. For the high-depth white-tailed eagle sequencing data, the average autosomal sequencing depth was estimated for the male and female separately, as the mode of the number of mapped reads per position across all scaffolds based on results from the command “bedtools coverage” from Bedtools v2.18.2 [61]. Using these averages, 195 for the female and 181 for the male, the relative sequencing depth was calculated for each position in each scaffold for both individuals. The per-scaffold relative sequence depth was then estimated for the female and male separately, as the mode across positions. Positions in autosomal scaffolds are expected to have a relative depth of 1 in both sexes, whereas Z-chromosomal scaffolds are expected to have a relative depth of 0.5 in females and 1 in males. As the estimate of relative depth may be less reliable for smaller scaffolds, the dependency of the relative mode depth due to scaffold size was analysed by calculating the variance in the depths per interval of scaffold sizes, transformed to a log scale. The distribution of the proportions of scaffolds at each interval was summarized with a cumulative percentage curve. In addition, the depth per scaffold was evaluated by comparing the per-scaffold relative sequencing depth between the two individuals: male over female. Scaffolds with a relative sequencing depth below 0.25 and above 1.5 were removed (corresponding to 523 scaffolds, and 0.47% of the genome). This ratio is expected to be around two for Z-chromosomal scaffolds and one for the

autosomal scaffolds, as the male has two copies of Z and the female one. Thus, a cut-off was set at 1.5.

Heterozygosity. Sex differences in heterozygosity were assessed by comparing numbers of heterozygous sites per scaffold based on genotypes of the high-depth white-tailed eagle male and female, called using Graphtyper [62, 63] with default settings. The variation on the Z-chromosome is expected to be ¾ of the autosomes and it should be restricted to the male, except for the PAR and non-recombining homologous regions. As scaffolds vary in length and may include short variable regions, the variation was also analysed per 50 kb window. Genotypes were filtered for quality using vcftools and bcftools version 0.1.15 and 1.7, respectively [64, 65] before counting, using minimum GQ score 20, minimum Q score 1000, missingness 1 (both individuals had to have a valid genotype at the site), mapping quality equal to 60 (MQ), and only biallelic sites. Two additional criteria were applied to remove sites with likely spurious heterozygous genotypes. First, heterozygous genotypes where the number of mapped reads deviated significantly from the mode depth of the scaffold, based on a two-sided Poisson test ($P < 0.01$) were excluded. Second, we used a binomial test to assess whether the proportion of reads in heterozygous genotypes, either in the male or the female, deviated from the 50/50 expectation, using $P < 0.05$ as the exclusion threshold.

Mapping. In order to assign the short reads from the white-tailed eagle to chromosomes, the 1142 scaffolds from the golden eagle scaffold assembly (which the white-tailed eagle genome had been mapped on) were mapped to the chicken genome, which has assigned chromosomes, using LASTZ [32]. Standard settings were used with the following modifications: ambiguous=iupac, gfextend, chain, gapped. Scaffolds in the golden eagle which mapped better to the Z-chromosome than any other chromosome, measured as most bases mapped, were deemed to belong to the golden eagle Z-chromosome.

SNP-loadings. A PCA analysis of 133 low-depth RAD sequenced white-tailed eagle individuals was constructed using PCangsd version 1.0 [66], an extension of ANGSD [67], as described below. A clear

split between males and females was observed along the first principal component (PC) (Figure S1). Loadings obtained with PCangsd were used to identify which parts of the scaffolds induced the split, with the "-selection" option [66] and with sites passing the following filters: a minimum 25% of individuals had to have valid genotypes, only unique mapping sites, base quality minimum 20, mapping quality minimum 30, SNP p-value 1e-6. ANGSD uses genotype likelihoods to tackle the restrictions of low depth [67, 68]. To assess which scaffolds contributed to the split on the first axis (PC1), a 95% range of loading values for all SNPs per scaffold was calculated using R [69] and compared between scaffolds with more than 50 SNPs. The distributions of the range of loading values were summarized with accumulation curves, combined for all scaffolds, and separately based on the results obtained by the mapping on the autosomes and Z chromosome. Scaffolds were assigned to the Z-chromosome or autosomes depending on whether the range-values were above or below a threshold of three standard deviations from the mean (covering ~99% of a normally distributed variable).

Comparison of the four methods. To evaluate how well the four approaches performed, the golden eagle scaffold assembly (GCA_000766835.1) was mapped to a golden eagle genome with known chromosomes (GCA_900496995.2) using LASTZ with the same settings and cut-off as described previously. In the results, the outcome of this mapping was used as the true chromosome identity of the 1141 scaffolds that was used to assess the accuracy of our four different approaches to identify Z chromosome scaffolds (Figure 5 and Table 2). A total of 168 scaffolds were assigned to the Z-chromosome, with a total length of 86,839,530 bp (mean = 516,902, sd = 1,509,132, and median = 5,236), which is slightly smaller than the Z-chromosome in the newly released genome of 88,216,475 bp (GenBank Assembly Accession: GCA_900496995.2). The autosomal loci mapped to 973 scaffolds of a size of 1,105,886,214 bp (mean = 1,136,574, sd = 3,403,676, and median = 5,674). The overlap of these four methods was summarized with the R-package VennDiagram [70].

Summary of the data and further statistical analyses, if otherwise not stated was done using R.

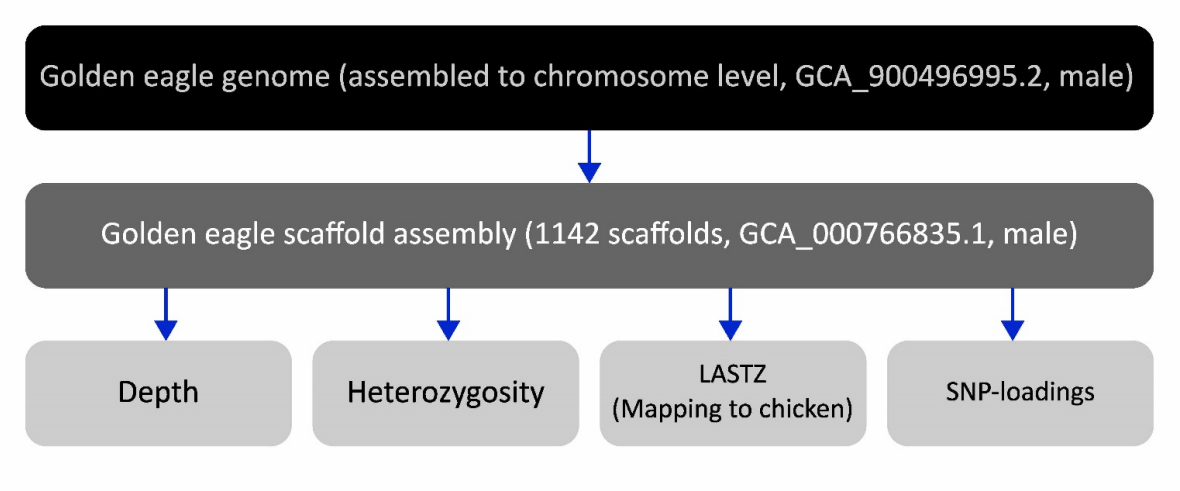


Figure 5. Schematic overview of the methods used to identify the Z-chromosome in a scaffold assembled genome. The golden eagle genome referred to in the dark grey box represents the reference in which we are attempting to identify scaffolds belonging to the Z-chromosome. The golden eagle genome in the black bar is the genome with known chromosomes, used to identify which scaffolds in the dark grey boxed genome probably belong to Z-chromosome (and autosomes) – to use as a reference. The light grey boxes are the four approaches we tested to find the scaffolds belonging to the Z-chromosome: 1) Depth: analysis of difference in sequencing depth between scaffolds in a high depth whole genome sequenced white-tailed eagle female. 2) Heterozygosity: analysis of the difference in heterozygosity per scaffold a high depth whole genome sequenced white-tailed eagle male and female. 3) LASTZ: mapping of the golden eagle reference genome to the chicken genome using LASTZ. 4) SNP-loadings: analysis of SNP-loadings for principal components splitting the sexes, in 133 RADseq white-tailed eagle individuals.

Declaration

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The raw dataset supporting the conclusions of this article is available in the DRYAD data repository https://doi.org/10.5061/dryad.v9s4mw6vs.

Further, the analysed dataset supporting the conclusions of this article is included in the supplementary.

Competing interests

The authors declare no competing interests.

Funding

The study was supported by Research Grant nr 185280-052 from The Icelandic Research Council, the Doctoral student fund of the University of Iceland and The University of Iceland Research fund.

Author contribution

CCRH and SP designed the study; KMW prepared the RADseq libraries; CCRH and SP analyzed the data; CCRH, KMW and SP wrote the paper.

Acknowledgement

Thanks to Gunnar T. Hallgrimsson from Department of Life and Environmental Sciences, University of Iceland, Menja von Schmalensee and Robert A. Stefansson from West-Iceland Centre of Natural History, and Kristinn Haukur Skarphédinsson from the Icelandic Institute of Natural History for sample collection.

Thanks to Jonas Meisner from section for Computational and RNA Biology, University of Copenhagen, for great help with PCAngsd.

Thanks to Agnar Helgason from deCODE genetics for guidance in analysis, writing, and access to facilities at deCODE genetics. Further, thanks to deCODE genetics for sequencing of the two white-tailed genomes.

Authors' information

Charles Christian Riis Hansen[1*], Kristen M. Westfall[1,2], and Snæbjörn Pálsson[1,3]

[1]Department of Life and Environmental Sciences, University of Iceland, Reykjavik, Iceland.

[2]Current: Fisheries and Oceans Canada, Pacific Biological Station, Nanaimo, BC, Canada.

[3]Senior Author

*Correspondence: CCRH

References

1. Hill WG, Robertson A. The effect of linkage on limits to artificial selection. Genetical Research. 1966;8:269–94.

2. Charlesworth D, Mank JE. The birds and the bees and the flowers and the trees: Lessons from genetic mapping of sex determination in plants and animals. Genetics. 2010;186:9–31.

3. de Manuel M, Barnett R, Sandoval-Velasco M, Yamaguchi N, Garrett Vieira F, Zepeda Mendoza ML, et al. The evolutionary history of extinct and living lions. Proceedings of the National Academy of Sciences. 2020;117:10927–34.

4. Pedersen CET, Albrechtsen A, Etter PD, Johnson EA, Orlando L, Chikhi L, et al. A southern African origin and cryptic structure in the highly mobile plains zebra. Nature Ecology and Evolution. 2018;2:491–8.

5. Pečnerová P, Garcia-Erill G, Liu X, Nursyifa C, Waples RK, Santander CG, et al. High genetic diversity and low differentiation reflect the ecological versatility of the African leopard. Current Biology. 2021; February. https://doi.org/10.1016/j.cub.2021.01.064.

6. Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, et al. Phylogenomic analyses data of the avian phylogenomics project. GigaScience. 2015;4:1–9.

7. Gopalakrishnan S, Samaniego Castruita JA, Sinding MHS, Kuderna LFK, Räikkönen J, Petersen B, et al. The wolf reference genome sequence (Canis lupus lupus) and its implications for Canis spp. population genomics. BMC Genomics. 2017;18:1–11.

8. Boulesteix M, Weiss M, Biémont C. Differences in genome size between closely related species: The Drosophila melanogaster species subgroup. Molecular Biology and Evolution. 2006;23:162–7.

9. Ellegren H. Evolutionary stasis: the stable chromosomes of birds. Trends in Ecology and Evolution. 2010;25:283–91.

10. Zhang G, Li C, Li Q, Li BB, Larkin DM, Lee C, et al. Comparative genomics reveals insights into avian genome evolution and adaptation. Science. 2014;346:1311–20.

11. Skarphéðinsson KH. Haförninn. Reykjavik: Fuglavernd (Fuglaverndarfélag íslands); 2013.

12. BirdLife International. Aquila chrysaetos. The IUCN Red List of Threatened Species 2016. 2016;e.T2269606.

13. BirdLife International. Haliaeetus albicilla. The IUCN Red List of Threatened Species 2016. 2016;e.T2269513.

14. Doyle JM, Katzner TE, Bloom PH, Ji Y, Wijayawardena BK, DeWoody JA. The genome sequence of a widespread apex predator, the golden eagle (Aquila chrysaetos). PLoS ONE. 2014;9:20–2.

15. Kim JA, Kang SG, Jeon HS, Jeon JH, Jang JH, Kim S, et al. Complete mitogenomes of two Accipitridae, Haliaeetus albicilla, and Pernis ptilorhynchus. Mitochondrial DNA Part B: Resources. 2019;4:392–3.

16. Damas J, O'Connor R, Farré M, Lenis VPE, Martell HJ, Mandawala A, et al. Upgrading short-read animal genome assemblies to chromosome level using comparative genomics and a universal probe set. Genome Research. 2017;27:875–84.

17. Warren WC, Clayton DF, Ellegren H, Arnold AP, Hillier LW, Künstner A, et al. The genome of a songbird. Nature. 2010;464:757–62.

18. Sigeman H, Ponnikas S, Chauhan P, Dierickx E, Brooke M de L, Hansson B. Repeated sex chromosome evolution in vertebrates supported by expanded avian sex chromosomes. Proc R Soc B. 2019;:20192051.

19. Bellott DW, Skaletsky H, Pyntikova T, Mardis ER, Graves T, Kremitzki C, et al. Convergent evolution of chicken Z and human X chromosomes by expansion and gene acquisition. Nature. 2010;466:612–6.

20. Palmer DH, Rogers TF, Dean R, Wright AE. How to identify sex chromosomes and their turnover. Molecular Ecology. 2019; September:1–16.

21. Xu L, Zhou Q. The female-specific W chromosomes of birds have conserved gene contents but are not feminized. Genes. 2020;11:1–14.

22. Graves JAM. Sex chromosome specialization and degeneration in mammals. Cell. 2006;124:901–14.

23. Kitano J, Peichel CL. Turnover of sex chromosomes and speciation in fishes. Environmental Biology of Fishes. 2012;94:549–58.

24. Kikuchi K, Hamaguchi S. Novel sex-determining genes in fish and sex chromosome evolution. Developmental Dynamics. 2013;242:339–53.

25. Nadachowska-Brzyska K, Burri R, Ellegren H. Footprints of adaptive evolution revealed by whole Z chromosomes haplotypes in flycatchers. Molecular Ecology. 2019;:mec.15021.

26. Webster TH, Couse M, Grande BM, Karlins E, Phung TN, Richmond PA, et al. Identifying, understanding, and correcting technical artifacts on the sex chromosomes in next-generation sequencing data. GigaScience. 2019;8:1–11.

27. Feron R, Pan Q, Wen M, Imarazene B, Jouanno E, Anderson J, et al. RADSex: A computational workflow to study sex determination using restriction site-associated DNA sequencing data. Molecular Ecology Resources. 2021;21:1715–31.

28. Rangavittal S, Stopa N, Tomaszkiewicz M, Sahlin K, Makova KD, Medvedev P. DiscoverY: a classifier for identifying Y chromosome sequences in male assemblies. BMC Genomics. 2019;20:641.

29. Nursyifa C, Brüniche-Olsen A, Garcia Erill G, Heller R, Albrechtsen A. Joint identification of sex and sex-linked scaffolds in non-model organisms using low depth sequencing data. Molecular Ecology Resources. 2021;:0–3.

30. Brelsford A, Lavanchy G, Sermier R, Rausch A, Perrin N. Identifying homomorphic sex chromosomes from wild-caught adults with limited genomic resources. Molecular Ecology Resources. 2017;17:752–9.

31. Zhou Q, Zhang J, Bachtrog D, An N, Huang Q, Jarvis ED, et al. Complex evolutionary trajectories of sex chromosomes across bird taxa. Science. 2014;346.

32. Harris RS. Improved pairwise alignmen of genomic DNA. Ph.D. Thesis. The Pennsylvania State University; 2007.

33. Kiełbasa SM, Wan R, Sato K, Horton P, Frith MC. Adaptive seeds tame genomic sequence comparison. Genome Research. 2011;21:487–93.

34. Noé L, Kucherov G. YASS: Enhancing the sensitivity of DNA similarity search. Nucleic Acids Research. 2005;33 SUPPL. 2:540–3.

35. Sætre G-P, Ravinet M. Evolutionary Genetics. 1st edition. Oxford University Press; 2019.

36. Jobling M, Hollox E, Hurles M, Kivisild T, Tyler-Smith C. Human evolutionary genetics. 2nd edition. Garland Science, Taylor & Francis Group, LLC; 2014.

37. Hooper DM, Price TD. Chromosomal inversion differences correlate with range overlap in passerine birds. Nature Ecology and Evolution. 2017;1:1526–34.

38. Xu L, Auer G, Peona V, Suh A, Deng Y, Feng S, et al. Dynamic evolutionary history and gene content of sex chromosomes across diverse songbirds. Nature Ecology and Evolution. 2019;3:834–44.

39. Fowler BLS, Buonaccorsi VP. Genomic characterization of sex-identification markers in Sebastes carnatus and Sebastes chrysomelas rockfishes. Molecular Ecology. 2016;25:2165–75.

40. Gamble T, Zarkower D. Identification of sex-specific molecular markers using restriction site-associated DNA sequencing. Molecular Ecology Resources. 2014;14:902–13.

41. Fridolfsson A-K, Ellegren H. A Simple and Universal Method for Molecular Sexing of Non-Ratite Birds. Journal of Avian Biology. 1999;30:116–21.

42. Sim SC, van Deynze A, Stoffel K, Douches DS, Zarka D, Ganal MW, et al. High-Density SNP Genotyping of Tomato (Solanum lycopersicum L.) Reveals Patterns of Genetic Variation Due to Breeding. PLoS ONE. 2012;7:1–18.

654 43. Li H, Ralph P. Local PCA Shows How the Effect of Population. Genetics. 2019;211 January:289–
655 304.

656 44. Otto SP, Pannell JR, Peichel CL, Ashman TL, Charlesworth D, Chippindale AK, et al. About PAR:
657 The distinct evolutionary dynamics of the pseudoautosomal region. Trends in Genetics.
658 2011;27:358–67.

659 45. Smeds L, Warmuth V, Bolivar P, Uebbing S, Burri R, Suh A, et al. Evolutionary analysis of the
660 female-specific avian W chromosome. Nature Communications. 2015;6.

661 46. Huylmans AK, Toups MA, MacOn A, Gammerdinger WJ, Vicoso B. Sex-biased gene expression
662 and dosage compensation on the artemia franciscana Z-chromosome. Genome Biology and
663 Evolution. 2019;11:1033–44.

664 47. Trukhina A V., Smirnov AF. Problems of Birds Sex Determination. Natural Science.
665 2014;06:1232–40.

666 48. Graves JAM. Avian sex, sex chromosomes, and dosage compensation in the age of genomics.
667 Chromosome Research. 2014;22:45–57.

668 49. Sigeman H, Ponnikas S, Hansson B. Whole-genome analysis across 10 songbird families within
669 Sylvioidea reveals a novel autosome–sex chromosome fusion. Biology Letters. 2020;16.

670 50. Lewin HA, Robinson GE, Kress WJ, Baker WJ, Coddington J, Crandall KA, et al. Earth
671 BioGenome Project: Sequencing life for the future of life. Proceedings of the National Academy of
672 Sciences. 2018;115:4325–33.

673 51. Helander B, Hailer F, Vilà C. Morphological and genetic sex identification of white-tailed eagle
674 Haliaeetus albicilla nestlings. Journal of Ornithology. 2007;148:435–42.

675 52. Thermo Fisher. Thermo Scientific GeneJET Genomic DNA Purification Kit #K0721, #K0722.
676 2016; October:1–8. https://www.thermofisher.com/document-connect/document-
677 connect.html?url=https%3A%2F%2Fassets.thermofisher.com%2FTFS-
678 Assets%2FLSG%2Fmanuals%2FMAN0012667_GeneJET_Whole_Blood_Genomic_DNA_Purification
679 _Mini_Kit_UG.pdf&title=VXNlciBHdWlkZTogR2VuZUpFVCBXaG9sZSBC. Accessed 14 Feb 2020.

680 53. Illumina. TruSeq® Nano DNA Library Prep.
681 https://support.illumina.com/content/dam/illumina-
682 support/documents/documentation/chemistry_documentation/samplepreps_truseq/truseqnano
683 dna/truseq-nano-dna-library-prep-guide-15041110-d.pdf. 2015.

684 54. Hillier LW, Miller W, Birney E, Warren W, Hardison RC, Ponting CP, et al. Sequence and
685 comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution.
686 Nature. 2004;432:695–716.

687 55. Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. Stacks: an analysis tool set for
688 population genomics. Molecular ecology. 2013;22:3124–40.

689 56. Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH, De Koning D-J. Stacks:
690 Building and genotyping loci de novo from short-read sequences. Genes|Genomes|Genetics.
691 2011;1:171–82.

692 57. Babraham Bioinformatics. FastQC. 2010.
693 http://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Accessed 16 Jun 2016.

694 58. Schubert M, Lindgreen S, Orlando L. AdapterRemoval v2: rapid adapter trimming,
695 identification, and read merging. BMC Research Notes. 2016;9:88.

696 59. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform.
697 Bioinformatics. 2009;25:1754–60.

698 60. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map
699 format and SAMtools. Bioinformatics. 2009;25:2078–9.

700 61. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features.
701 Bioinformatics. 2010;26:841–2.

702 62. Eggertsson HP, Jonsson H, Kristmundsdottir S, Hjartarson E, Kehr B, Masson G, et al.
703 Graphtyper enables population-scale genotyping using pangenome graphs. Nature Genetics.
704 2017;49:1654–60.

705 63. Eggertsson HP, Kristmundsdottir S, Beyter D, Jonsson H, Skuladottir A, Hardarson MT, et al.
706 GraphTyper2 enables population-scale genotyping of structural variation using pangenome
707 graphs. Nature Communications. 2019;10:1–8.

708 64. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format
709 and VCFtools. Bioinformatics. 2011;27:2156–8.

710 65. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and
711 population genetical parameter estimation from sequencing data. Bioinformatics. 2011;27:2987–
712 93.

713 66. Meisner J, Albrechtsen A. Inferring Population Structure and Admixture Proportions in Low-
714 Depth NGS Data. Genetics. 2018;210:719–31.

715 67. Korneliussen TS, Albrechtsen A, Nielsen R. ANGSD: Analysis of next generation sequencing
716 data. BMC bioinformatics. 2014;15:356.

717 68. da Fonseca RR, Albrechtsen A, Themudo GE, Ramos-Madrigal J, Sibbesen JA, Maretty L, et al.
718 Next-generation biology: Sequencing and data analysis approaches for non-model organisms.
719 Marine Genomics. 2016;30:3–13.

720 69. R core Team. R: A language and environment for statistical computing. 2020.

721 70. Chen H, Boutros PC. VennDiagram: a package for the generation of highly-customizable Venn
722 and Euler diagrams in R. BMC Bioinformatics. 2011;12:35.

723

724

Supplementary

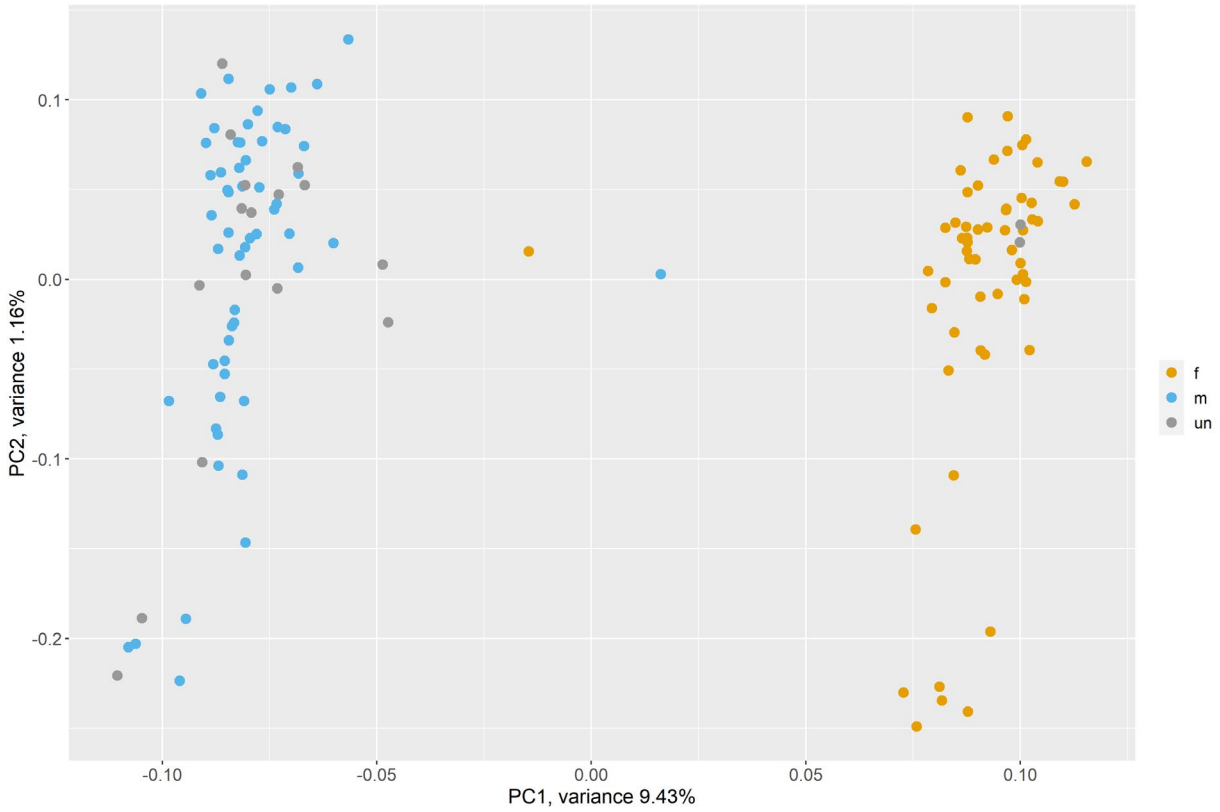


Figure S1. PCA made with PCangsd (Meisner & Albrechtsen, 2018) displaying genetic split between males and females for 133 individuals based on 164,952 SNPs in the Icelandic population. Orange=females (f), blue=males (m), grey=unknown (un), morphologically sexed in the field.

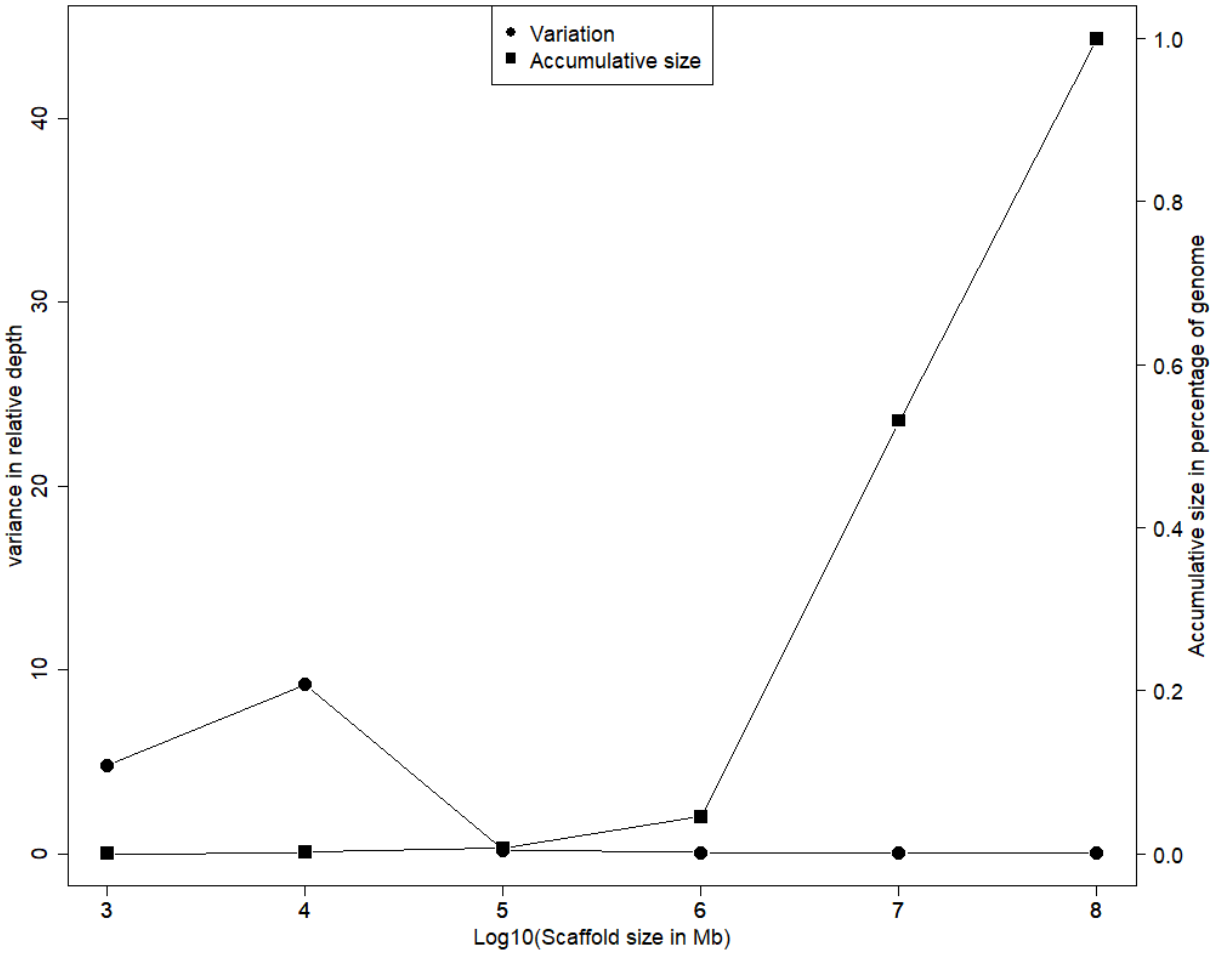
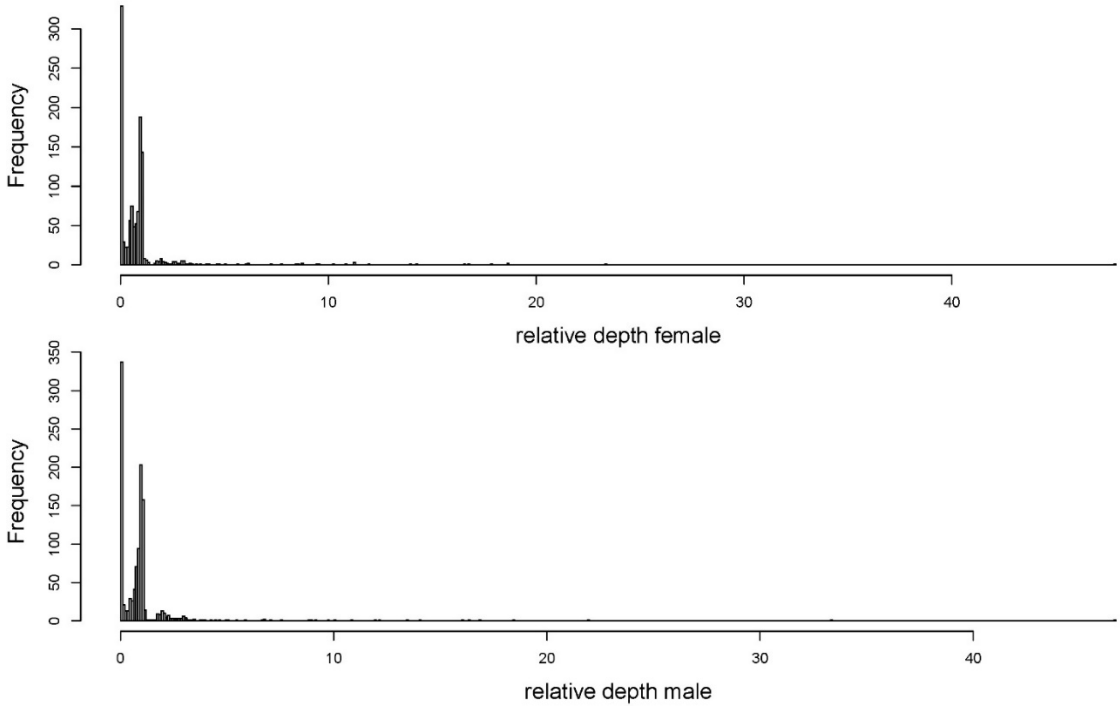
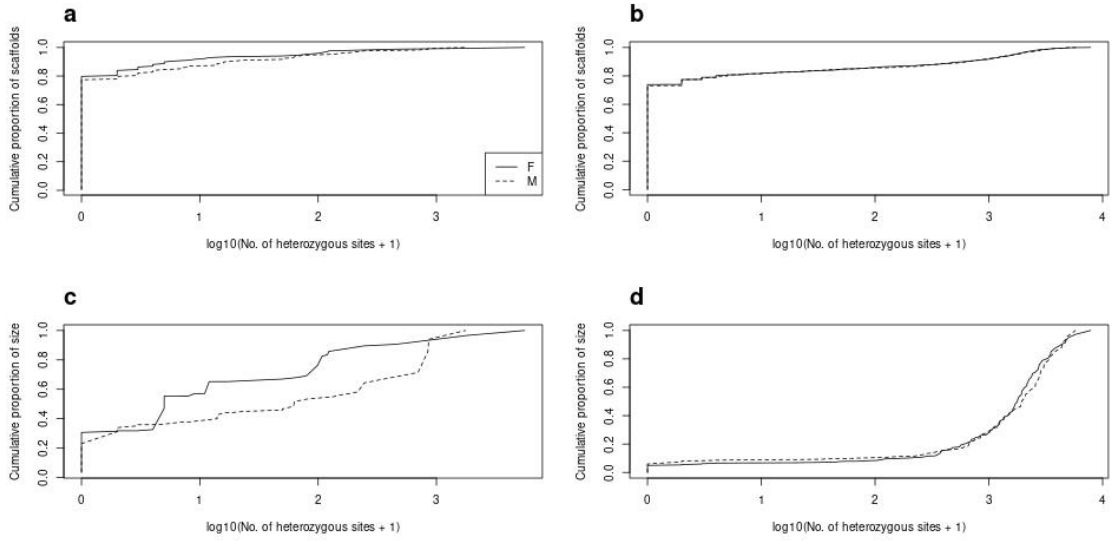


Figure S2. Variation in relative depth in comparison with scaffold size and proportion of the genome. Left y axis with dots show the variance in relative depth between scaffolds per log value, right y axis and square points show the accumulative proportion of the genome.



14 *Figure S3. Relative mode depth for the female (top) and male (bottom) for all scaffolds.*

15



16

Figure S4. Cumulative proportion of heterozygotic sites and of their chromosomes per scaffold, in the genome of the high depth white-tailed eagle female and male. Proportion of scaffolds on the Z-chromosome (a) and autosomes (b) and proportion of total size of the Z chromosome (c) and autosomes (d). F: female, M: male.

17

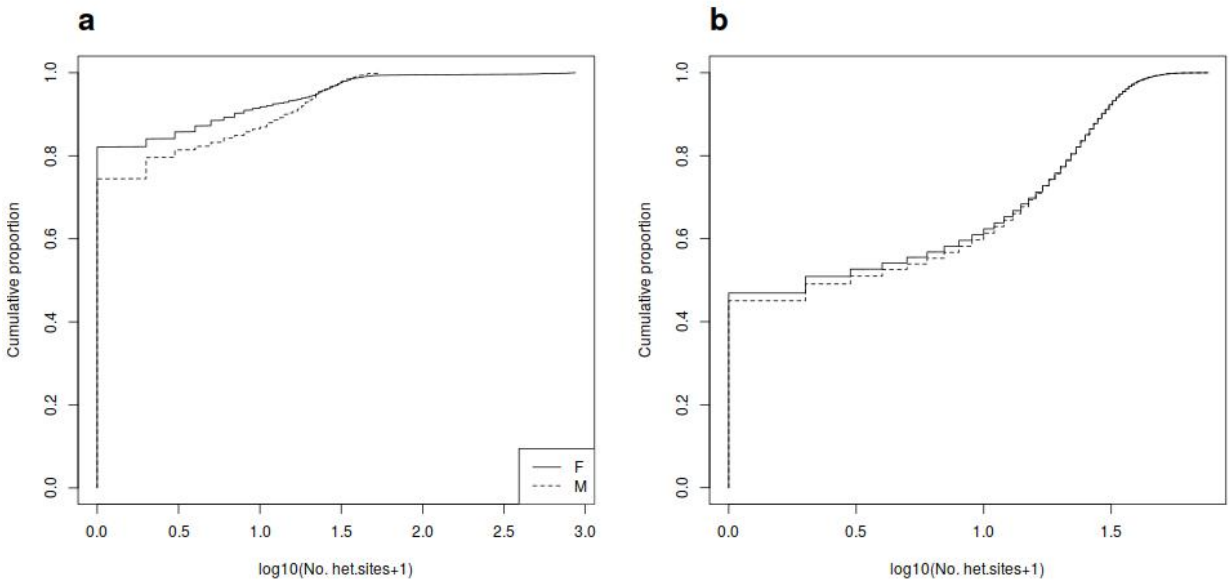


Figure S5. Cumulative proportion of heterozygotic sites in windows of a size of 50 kb in the genome of the two white-tailed eagles. a) Z-chromosome, b) autosomal chromosomes. F: female, M: male.

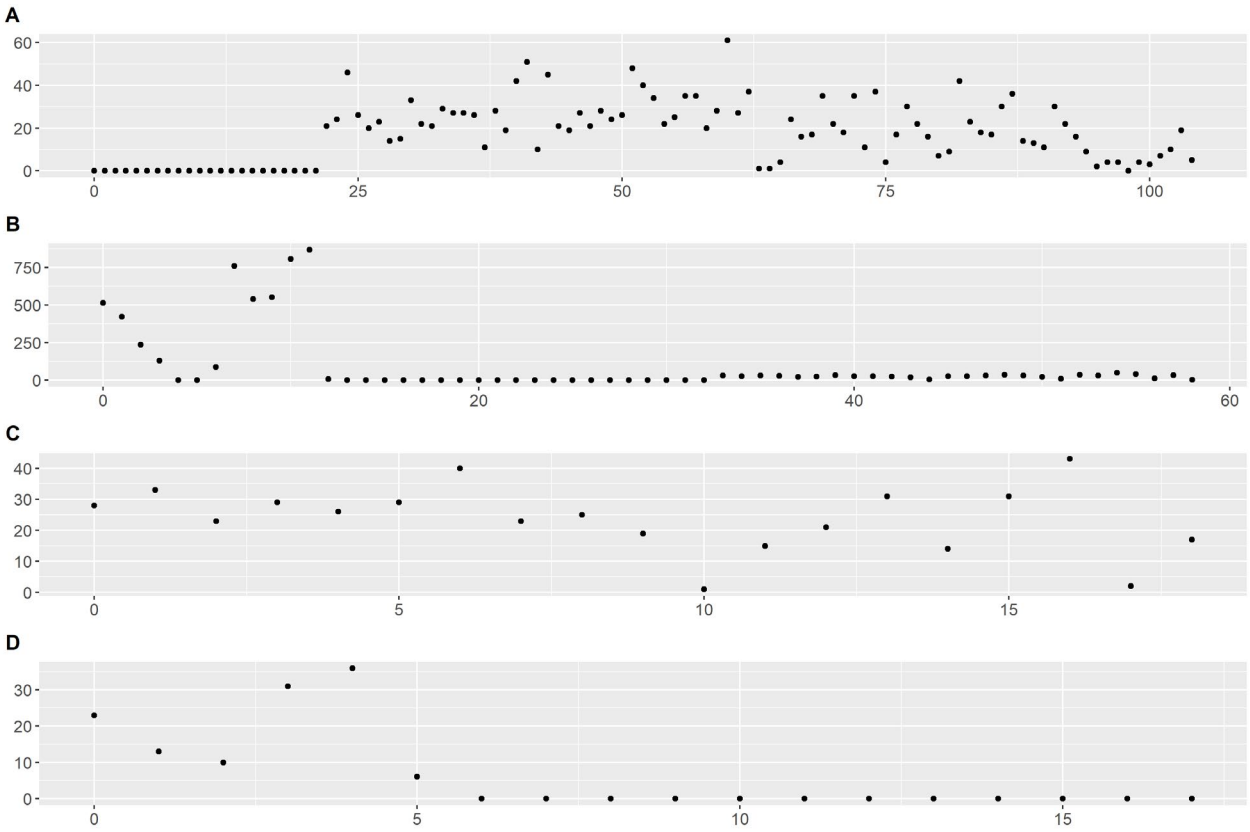


Figure S6. Heterozygosity per window (50 Kb windows) along the scaffold for A) NW_011950951.1, B) NW_011950990.1, C) NW_011951047.1, D) NW_011951051.1.

Supplementary Table 1

See separate file.

Table S1. Full raw data file, containing all values used to refer results in the paper. "golden_contig": Scaffold name in golden eagle scaffold assembled genome; "chicken_contig_best": Chromsome name in chicken genome that the "golden_contig" mapped best to; "mapped_bases_best": Number of bases of "golden_contig" that mapped to the reported chicken chromosome; "mapped_reads_best": Number of continues reads of "golden_contig" that mapped to the reported chicken chromosome; "chicken_contig_secondbest": Chromsome name in chicken genome that the "golden_contig" mapped second best to; "mode_M_WZ": Mode of sequencing depth of high depth male of "golden_contig"; "mode_F_XA": Mode of sequencing depth of high depth female of "golden_contig"; "sd_M_WZ": Standard deviation of mode of sequencing depth of high depth male of "golden_contig"; "sd_F_XA": Standard deviation of mode of sequencing depth of high depth female of "golden_contig"; "length_golden_contig: Length (number of bases) of "golden_contig"; "X5pctile": SNP loading of the 5.

percentil of "golden_contig"; "X95pctile": SNP loading of the 95. percentil of "golden_contig"; "het_m_all": Heterozygoes sites with no filtering in the high depth male in the "golden_contig"; "het_f_all": Heterozygoes sites with no filtering in the high depth female in the "golden_contig"; "het_m_filtered": Heterozygoes sites with filtering as described in the paper in the high depth male in the "golden_contig"; "het_f_filtered": Heterozygoes sites with filtering as described in the paper in the high depth female in the "golden_contig"; "TrueChromosome": Chromsome name in golden eagle chromosome assembled genome that the "golden_contig" mapped to; "TrueChromosomeZorA": Separation of the chromosomes in "TrueChromosome" into autosomes and Z-chromosome.

Supplement text 1 - ddRAD library preparation and sequencing

The 133 samples were prepared for double digest restriction-site associated DNA sequencing (ddRADseq) using modified protocols from Elshire et al. [61] and Peterson et al. [62]. Total genomic DNA (100-500 ng) was sequentially digested using the restriction endonucleases Sau3AI (1U) and ApeKI (2U), respectively, each for four hours at manufacturer (NEB) recommended temperatures in NEB Buffer 4. Digested DNA (100 ng) was ligated to adapters (sequences in Elshire et al. [61]) containing unique combinatorial barcodes (16 unique 5 bp barcodes for ApeKI adapters and five unique 6 bp barcodes for Sau3AI adapters) for each individual (barcode and adapter sequences in Supplementary Information S1) using T4 DNA ligase (NEB) in supplied buffer at 21°C for four hours. Ligation reactions contained a 6:1 molar excess of adapter to fragmented DNA, calculated using the mean fragment size determined from an agarose gel. Ligated DNA was pooled and purified using magnetic beads (Macherey-Nagel NGS clean-up and size selection) following the manufacturers protocol. Size selection of ligated DNA fragments was performed on a Pippin Prep (Sage Science) with 2% ethidium-free agarose gels and external size standard. The narrow range setting included a mean fragment size of 350 bp ± 18 bp. The eluate was split among eight PCR reactions and amplified using the primers and PCR conditions as in Elshire et al. [61]. Each PCR reaction had a total volume of 25 µL containing; 1x OneTaq Master Mix with Standard Buffer (NEB), 0.5 mM each primer, and 8 µL template DNA. PCR products were pooled and purified using magnetic beads before quantification using a SYBR Gold

fluorometric assay (protocol in Supplementary Information S2). The library was prepared for sequencing following manufacturer's instructions with a final concentration of 38 nM. The library was sequenced on an Illumina HiSeq2500 using the Illumina TruSeq kit (2x125bp) following the manufacturer's instructions. The sequencing was done on one lane and obtained 303 million unambiguous PE reads.

Paper III

Paper III

Genomics of white-tailed eagle (*Haliaeetus albicilla*) in the North-Atlantic islands reveal low diversity and substantial inbreeding in comparison with the mainland populations

In review.

Author contribution: Conceived the idea: CCRH, and SP; Performed the experiments: CCRH, JAR, JACB, MS, MDM and SP; Analysed the data: CCRH, and SP; wrote the paper: CCRH and SP; Commented on the paper: JAR, JACB, MS, GTH, RAS, MvS, KHS, ALL, ML, CS, RD, KS, DB and IE, MDM, AH, MTPG; Contributed substantial materials, resources or funding: CCRH, GTH, RAS, MvS, KHS, ALL, ML, CS, RD, KS, DB, IE, MDM, AH, MTPG and SP.

Genomics of white-tailed eagle (*Haliaeetus albicilla*) in the North-Atlantic islands reveal low diversity and substantial inbreeding in comparison with the mainland populations

Charles Christian Riis Hansen[1*], Jacob Agerbo Rasmussen[2,3], Jesus Adrian Chimal Ballesteros[2,4], Mikkel-Holger S. Sinding[3], Gunnar T. Hallgrimsson[1], Robert A. Stefansson[5], Menja von Schmalensee[5], Kristinn Haukur Skarphédinsson[6], Aili Lage Labansen[7], Madis Leivits[8], Christian Sonne[9], Rune Dietz[9], Kim Skelmose[10], David Boertmann[9], Igor Eulaers[9], Michael D. Martin[2], Agnar Helgason[11, 12], M. Thomas P. Gilbert[2,3] and Snæbjörn Pálsson[1]

[1]Department of Life and Environmental Sciences, University of Iceland, Reykjavik, Iceland

[2]Department of Natural History, NTNU University Museum, Norwegian University of Science and Technology (NTNU), Trondheim, Norway

[3]Center for Evolutionary Hologenomics, The Globe Institute, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

[4]Natural History Museum, University of Oslo, Oslo, Norway

[5]West-Iceland Nature Research Centre, Iceland

[6]Icelandic Institute of Natural History, Iceland

[7]Greenland Institute of Natural Resources, Nuuk, Greenland

[8]Chair of Clinical Veterinary Medicine, Institute of Veterinary Medicine and Animal Sciences, Estonian University of Life Sciences, 51006, Tartu, Estonia

[9]Department of Ecoscience, Aarhus University, Roskilde, Denmark

[10]Natural History Museum of Denmark, University of Copenhagen, 1350 Copenhagen, Denmark.

[11]Department of Anthropology, University of Iceland, Reykjavik, Iceland.

[12]deCODE genetics, Reykjavik, Iceland.

Senior Author: SP

Lead author: CCRH

*Correspondence: CCRH, ccr3@hi.is.

Abstract

Using whole genome shotgun sequences from 92 white-tailed eagles (*Haliaeetus albicilla*) sampled from Greenland, Iceland, Norway, Denmark, Estonia, and Turkey between 1885-1950 and after 1990, we investigate the genomic variation within countries over time, and between countries. Clear genetic differentiation is observed between samples from the different countries, with the largest differences between the island and mainland populations, and indications that the island populations share the most recent ancestry with the Norwegian population. We find signs of strong inbreeding in the island populations. Further, temporal differences are observed in some populations, for example, replacement of the Danish gene pool following its population's extinction in the early 20th century, as well as a change in the genetic diversity of the Icelandic population following a severe bottleneck during the last century, all of which could warrant a further conservation effort in Iceland. More generally, all populations show a decline in effective population size, which may have been shaped by I) distinct refugia during the last glacial period, II) population divergence following the colonization of the deglaciated areas ~10,000 years ago, III) human population expansion and e.g., settlement in Iceland ~1,100 years ago, and IV) human persecution and toxic pollutants during the last two centuries.

Introduction

The white-tailed eagle (*Haliaeetus albicilla*, Linnaeus, 1758) is a large raptor whose range spans the Palearctic and Greenland. It is currently categorised as a least-concern species by the International Union for Conservation of Nature (IUCN) and its population size is growing (Birdlife International 2020). However, during the 19th and 20th century the white-tailed eagle experienced severe population bottlenecks and became locally extinct in several countries in western Europe e.g., Denmark and the British Isles (Love and Ball 1979, Ehmsen et al. 2011, Langguth et al. 2013, Treinys et al. 2016). The population declines derived primarily from human persecution (Bijleveld 1974, Love and Ball 1979), although later during the 20th century, were due to toxic effects of

organochlorines and neurotoxins (Helander et al. 1982, 2002; Skarphéðinsson 2003). Habitat destruction or occupancy of sites by humans due to growing human population during the last millennia (Kremer 1993) and its impact may have further restricted the population sizes of eagles via e.g., settlement of coastal sites and islands in the North Atlantic. For example, Iceland and SW-Greenland were settled in the late 9th century (Batt et al. 2015) and around the year 1000 AC, respectively, by Norse people (Jackson et al. 2018).

Fortunately, successful conservation programmes that were introduced in the late 20th century have helped to restore population sizes, as well as measures to reduce harmful substances in the environment such as the persistent organic pollutants (POPs) as outlined in the Stockholm Convention in 2001 (https://www.epa.gov/international-cooperation/persistent-organic-pollutants-global-issue-global-response). Furthermore, urbanisation in the twentieth century may also have favoured the re-establishment of eagles in less populated coastal territories.

Overall, the population bottlenecks are expected to have lowered the genetic diversity within populations and increased the genetic differentiation among groups due to genetic drift, which may have been exacerbated by reduced connectivity among populations. This reduction in genetic variation and population sizes is also expected to have reduced the populations' potential to adapt to environmental changes and, to have decreased the efficacy of selection to purge deleterious mutations from the gene pools (Hoban et al. 2020). Recent evidence has shown that genetic diversity has been decreasing within many wild species (Leigh et al. 2019) and concerns have been raised that this has been neglected in management and conservation policy (Hoban et al. 2021). However, prior studies on the white-tailed eagle based on the mitochondrial genome or at microsatellites loci have not reported any such effect, and authors argued that the bottlenecks undergone by the white-tailed eagles were relatively short in comparison to their lifespan, and thus the effects will have been minimal (Hailer et al. 2006, 2007). It remains to be seen if this lack of apparent bottleneck effect is also reflected in the nuclear genome.

With regards to the white-tailed eagle's population structure, previous studies based on the mitochondrial control region (Hailer et al. 2007) and the whole mitogenome (Hansen et al., 2021 in review) have revealed two or potentially three major genetic clusters within the species range, which were shaped by refugia during the last glacial periods of the Ice age (Hailer et al. 2007, Honnen et al. 2010, Langguth et al. 2013). The mitogenomic studies reported little variation in the populations in Greenland and Iceland and suggested a shared recent matrilineal origin with populations in north-western Europe (Hailer et al. 2007). The population in Estonia had high diversity, as it harboured variation from the two main lineages (from east and west) within the species range (Hailer et al. 2007). A recent analysis of the mitogenome revealed a more complex pattern, with polyphyletic lineages in Iceland, Greenland, and Norway, and the recently established population in Denmark showed signs of admixture between the two main clusters (Hansen et al. 2021 in review).

The white-tailed eagle is primarily sedentary, with no examples of migrants between Greenland, Iceland, and the mainland (Lyngs 2003, Birdlife International 2020). The population in Greenland, earlier classified as a subspecies due to its large body size (Salomonsen 1979, Hailer et al. 2007), inhabits the southwest coast. The number of breeding pairs there has increased in recent decades, from 50-75 pairs around the middle of the 20th century (Hansen 1979), to around 200 pairs today (Boertmann and Bay 2018). The population in Iceland was distributed along the coastline before 1850, after which the population plummeted to around 20 pairs when it was conserved in 1914. The population started to increase following a ban on fox poisoning introduced in 1964 (as the eagles scavenged on poisoned carcasses) (Petersen 1998, Skarphéðinsson 2013), although at a relatively slow rate, and today contains about 80 pairs (Skarphéðinsson 2013). The Norwegian population is the largest in Europe and consists of around 2,000 breeding pairs (Jais 2020). As in several other countries in Europe, the Danish population went extinct during the beginning of the 20th century but re-established in 1995, most likely from expanding neighbouring areas and by 2020

numbered 133 breeding pairs (Skelmose and Larsen 2021). In Estonia, although there was a large population prior to the 19th century, consisting of ca. 400-500 breeding pairs (Lõhmus 1998), by the end of the 19th century, it had declined to only 20 pairs (Randla and Õun 1980). Today however it has recovered to an estimated 290-330 pairs (Elts et al. 2019).

In this study, genomic variation in historic (up to 130 years old) and contemporary samples from the two isolated island populations in Greenland and Iceland are studied and compared with samples from the mainland populations in Norway and Denmark, and contemporary samples from Estonia. In addition, one historical specimen from Turkey was included for reference as an outlier. We specifically evaluate the impact of population size and bottlenecks on the Iceland and Greenland populations in comparison with the large mainland population in Norway, the recently established population in Denmark, the population from Estonia, and the historic samples. The historic samples come from the onset of the reported reduction in population size. Moreover, we aimed to use our data to examine the population structure, the difference between the island and the mainland population both with respect to diversity and inbreeding, the demography, and the origin of the white-tailed eagles in the North Atlantic populations. The comparison of the contemporary and historic samples allows further assessment of any genomic changes over time during the 20th century, firstly whether the diversity has been reduced within populations due to increased genetic drift and whether it has led to increased differentiation among populations.

Material and Methods

Tissue was obtained from 92 specimens: 63 contemporary and 29 historic, from six different countries. These included 12 contemporary and eight historic individuals from Greenland, 25 contemporary and two historic individuals from Iceland, 12 contemporary and 13 historic individuals from Norway, 11 contemporary and five historic individuals from Denmark, three

127 contemporary individuals from Estonia, and one historic individual from Turkey (Figure 1). The
128 historic specimens were sampled between 1885 and 1950 (all but the two Icelandic individuals were
129 sampled prior to 1937), while all contemporary individuals were sampled post-1990 (full individual
130 information is presented in Table 1).

131 Muscle tissue and whole blood from contemporary samples from Estonia, Denmark, and Greenland
132 (Table 1) were stored at -20 °C until DNA extraction and were provided by the Department of
133 Ecoscience, Arctic Research Centre, AU, Roskilde, Denmark (Estonian, Danish, and Greenland
134 samples), Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark
135 (Danish samples) and the Greenland Institute of Natural Resources, Nuuk, Greenland (Greenland
136 samples). Whole blood samples from contemporary samples from Iceland were collected in an
137 ongoing monitoring project of the white-tailed eagle in Iceland (led by the Icelandic Institute of
138 Natural History) and stored in EDTA at -20 °C until DNA extraction. Whole genome shotgun DNA
139 sequences from twelve Norwegian individuals were provided by the Department of Natural History,
140 University Museum, Norwegian University of Science and Technology (NTNU), Trondheim, Norway.
141 DNA extraction, library building, and sequencing of all contemporary samples are described in
142 Hansen et al (2021, in review).

143 Historic samples consisting of toepad clippings, taken with disposable sterile scalpel blades, from
144 museum samples provided by The Natural History Museum of Denmark, University of Copenhagen,
145 Denmark; Icelandic Institute of Natural History, Reykjavik, Iceland, and Department of Natural
146 History, NTNU University Museum, Norwegian University of Science and Technology (NTNU),
147 Trondheim, Norway.

148 Historic samples from Greenland, Iceland, Denmark, Turkey, and five of the thirteen Norwegian
149 specimens were processed at the clean laboratory facilities at the Globe Institute at the University
150 of Copenhagen. Firstly, to prevent cross-contamination from other museum specimens, the
151 samples were cleaned with a dilute bleach solution (ca. 5% commercial strength), then rinsed with

70% ethanol followed by molecular biology grade water performed using a proteinase-based lysis-buffer according to Gilbert et al. (2008). Each sample was added 300 µL lysing buffer including 20 µL proteinase K and incubated for 3 hours. The supernatant was purified by combining 720 µL binding buffer modified as in Allentoft et al. (2015), with 80 µL sample lysate, vortexed and centrifuged through a Monarch® DNA Cleanup Column (New England Biolabs Inc., Beverly, Massachusetts, USA). The binding step was repeated 3 times after which the column was washed with 800 µL PE buffer, from where the DNA eluded into 21.5 µL EBT buffer. Throughout the entire process, only LoBind Eppendorf tubes were used.

The remaining eight Norwegian historic specimens were processed at the Norwegian University of Science and Technology (NTNU) University Museum's dedicated palaeo-genomics laboratory. For these, the genomic DNA extractions were performed with a Qiagen DNeasy Blood & Tissue kit. The manufacturer's protocol was used except that the amount of proteinase K was doubled, and the lysis step incubation at 56°C was extended to 15 hours. The DNA solutions were incubated at 37°C for 10 minutes prior to elution.

For all historic samples, blunt-end Illumina shotgun sequencing libraries were prepared using the BEST protocol (Carøe et al. 2018). In both of the aDNA laboratories, extraction and library blanks were also included to monitor for contamination.

Indexed libraries from historic samples from Greenland, Iceland, Denmark, Turkey and five Norwegian specimens processed at the University of Copenhagen were paired-end sequenced on four flow cells with 2x150 bp read length at deCODE Genetics in Iceland using an Illumina NovaSeq 6000.

The purified and indexed libraries for the eight Norwegian specimens processed at NTNU were pooled and paired-end sequenced over two runs on the Illumina HiSeq 4000 platform at the NTNU

175 Genomics Core Facility, and over one run on an Illumina NovaSeq 6000 at the University of Oslo
176 Norwegian National Sequencing Centre.

177 Fastq file quality of all samples was checked using FastQC (Babraham Bioinformatics 2010), then
178 run through AdapterRemoval v2 using standard-setting, but providing adapter sequences for
179 samples, and using the arguments --collapse and –trimns (Schubert et al. 2016). The fastq files were
180 mapped to the golden eagle (*Aquila chrysaetos*) genome (GCA_900496995.3) using bwa aln, samse,
181 and sampe, with the flags -q 15 and -k 1 (Li and Durbin 2009). Although a white-tailed eagle genome
182 is available, the golden eagle was deliberately chosen as the reference to minimize the potential of
183 mapping biases derived from the fact that the available white-tailed eagle genome is not equally
184 related to all populations studied here (the published white-tailed eagle genomes come from
185 Greenland, UK, and Germany), thus might introduce errors in the analyses (Gopalakrishnan et al.
186 2017). A further benefit of aligning to the golden eagle genome is that it has been assembled to
187 chromosome level completeness and annotated, thus enabling us to both identify and exclude sex
188 chromosomes as needed in some of the downstream analyses, and identify the genes present in
189 regions under selection. Picard (Broad Institute 2020) was used to remove duplicate reads. To
190 identify likely damaged bases the base quality score was rescaled with mapDamage 2.0 (Jónsson et
191 al. 2013). Genotypes were called using GraphTyper2 (Eggertsson et al. 2019) with standard settings.
192 The VCF file for the 92 individuals was filtered using VCFtools, BCFtools, and VCF-annotate; SNPs
193 had to have a minor allele count of one, quality of 1000, genotype quality 20, mapping quality 30,
194 max missingness of 25%, and an allelic heterozygosity balance between 0.2 and 0.8. Individuals had
195 to have a sequencing depth of eight for a particular SNP to have it called. Only the known autosomes
196 1-26 (LR606181.1-LR606206.1) from the golden eagle genome were analysed in this study (these
197 make up 84.43% of the full published genome).

198 The Ts/Tv ratio was examined using VCFtools TsTv-summary (Danecek et al. 2011), to evaluate
199 whether a bias is observed in the historic samples and whether the transitions should be excluded.

MapDamage 2.0 (Jónsson et al. 2013) was used to investigate nucleotide misincorporation between the historic and contemporary samples, as historic and ancient samples are expected to have G-to-A and C-to-T substitutions at the 3'- and 5'end due to post-mortem DNA damage (Jónsson et al. 2013).

As nest origin was known for the contemporary Icelandic specimens, to reduce sample relatedness all specimens selected originated from different nests. To estimate relatedness KING (Manichaikul et al. 2010) was run with the settings --unrelated and --degree 3. No pair of individuals were found to be related at third degree or higher, and thus all 92 individuals were kept.

Diversity within each sample was evaluated based on the observed and expected heterozygosity calculated on a per-individual basis using VCFtools het (Danecek et al. 2011), and observed and expected heterozygosities per SNP per population using VCFtools hardy (Danecek et al. 2011). Euclidean distances within populations were also calculated as 1-identity by state proportion (IBS distance) obtained with SNPrelate in R (Zheng et al. 2012). The heterozygosity along the genome for each population, obtained with the VCFtools hardy was also inspected visually to assess whether its distribution varied over the whole genome.

To look for signals of inbreeding in each population the coefficient F were calculated by comparing the observed and expected estimates per loci, averaged over the genome (from vcftools --hardy, only in populations with a sample size with five or more to ensure adequate power when calculating expected heterozygosity) as F_{IS} (Nei 1977), and based on the proportions of heterozygous sites (F_H) (from vcftools --het) as in plink (Purcell et al. 2007), and by runs of homozygosity (ROH) using plink homozyg (F_{ROH}) (Purcell et al. 2007). The settings for ROH were: homozyg-window-snp 10, homozyg-window-missing 10, homozyg-window-het 1, homozyg-snp 30, homozyg-kb 500, homozyg-gap 1000, homozyg-density 200. F_{ROH} is defined as the sum of ROH length for an individual divided by the total length of the autosomes (McQuillan et al. 2008). It has been shown that F_H and especially F_{ROH} reflect well the true inbreeding (Kardos et al. 2015, Forutan et al. 2018).

The relationship between genotypes was evaluated with a dendrogram based on Identity-By-State (IBS) generated using SNPRelate (Zheng et al. 2012). The dendrogram was calculated from the VCF file with an applied minor allele frequency (MAF) filter of 0.05 in R using SNPRelate (Zheng et al. 2012) with standard settings.

The population structure, admixture, and divergence were further examined with a principal component analysis (PCA) using EIGENSOFT (Patterson et al. 2006, Price et al. 2006), admixture plot using ADMIXTURE v 1.3.0 (Alexander et al. 2009), Weir and Cockerham's F_{ST} (Weir and Cockerham 1984) with VCFtools (Danecek et al. 2011), and pairwise Euclidean distances using SNPRelate (Zheng et al. 2012). For these five analyses, a MAF filter of 0.05 was again applied to the data. To generate the PCA, EIGENSOFT was used with standard settings, except the option numoutlieriter was set to 0, and it was run with lsqproject. For lsqproject, 61 individuals were used to make the initial PCA which the remaining 31 individuals were projected on. The 61 individuals had a maximum missingness of 13%, which meant all sample groups (country/time), except the historic Icelandic, were represented in the initial PCA. 591,698 SNPs were used in the analysis. The ADMIXTURE analysis was run with 100-fold cross-validation (-cv = 100) and with 100 iterations for K1-20. The program was considered to have converged to a given K-value if delta was below 10^{-4} between five iterations in a row. Weighted F_{ST} was calculated using standard settings in VCFtools. Pairwise Euclidean distances were calculated as 1-identity by state proportion (IBS distance) obtained with SNPrelate in R. The difference in IBS distances between temporal samples within countries was tested with a Wilcoxon test (Sokal and Rohlf 2012). Net IBS distances were calculated separately between the historic and the contemporary samples as $D=d_{ij}-(d_{ii}+d_{jj})/2$ (Nei and Li 1979), where d_{ij} is the average distance between samples i and j and d_{ii} and d_{jj} are the average distances within samples. The probability to observe the observed outcome (P-values) was estimated by permuting the distances 100 times among each pair of samples, the p-values were adjusted using sequential Bonferroni.

Following the analyses of structure and divergence, potential selection or deviation from neutral equilibrium was examined by estimating Tajima's D (Tajima 1989) per population, both for the contemporary and historic samples, using VCFtools (TajimaD) in windows of 50K, and calculating the mean and standard deviations. Further, signs of selection were investigated by looking for differentiation along the genome, thus F_{ST} was calculated along the genome in windows of 100,000 bases, using VCFtools. The comparison was done separately between the contemporary and for the temporal samples from only Greenland, Iceland, and Norway. The Estonian and Turkey samples were not analysed as they only have one temporal sample, Denmark was not analysed as it is known that the population in Denmark went extinct in the time between the contemporary and historic samples used here.

The evolutionary histories of the populations were further analysed with Treemix (Shriner et al. 2014), using the bald eagle (*Haliaeetus leucocephalus*) as an outgroup. The bald eagle reference genome (NCBI: GCF_000737465.1) was mapped to the golden eagle reference using bwa mem and bcftools mpileup. BCFtools were used to call genotypes and to filter it as the other genomes in this study, including MAF 0.05. Treemix was run with the options: bootstrap, noss, global, se, and k 500. It was run with 0 to 5 migrations events (m), with 50 iterations per migration event. Treemix was evaluated by looking at the likelihood and proportion of explained variance per m.

Finally, to evaluate the changes in effective population size over time Stairway Plot v2 (Liu and Fu 2015, 2020) was used. Due to the small sizes of the historic samples, only the contemporary samples from Greenland, Iceland, Norway, and Denmark were analysed in this case. Options ninput was set at 200 and pct_training at 0.67, and generation time was set for 15.6 years as reported by IUCN (Birdlife International 2020). Stairway plot was run using a mutation rate of 2.3e-9 per site per year, obtained from the collared flycatcher, the only reported whole genome mutation rate for birds (Smeds et al. 2016). As larger animals can be expected to have a lower mutation rate than smaller animals, possibly due to larger generation time or differences in population sizes (Lynch

2010), the stairway analysis was also run with a mutation rate of $1*10^{-9}$ per site per year. The stairway plot was derived from a site frequency spectrum (SFS) of less filtered data to include as many sites as possible. The SFS was calculated per population directly from the bam files including only the autosomes, calculating first the SAF (site allele frequency likelihood) file using ANGSD with the settings doSaf 1, gl 1, minMapQ 30, minQ 20 and uniqueOnly 1, and then ANGSD realSFS with the settings nsites 100,000,000, bootstrap 10 and tole 1e-15 (Li 2011, Nielsen et al. 2012, Korneliussen et al. 2014). For the contemporary Norwegian population, one individual was removed prior to calculating the SFS due to high amount of missing data, and thus only 11 contemporary Norwegian individuals were used in this analysis. Number of all sites (fixed and variable) used per population in the stairway plot analysis were: Greenland 1,037,550,747; Iceland 1,040,669,524; Norway 1,044,147,099; and Denmark 1,037,504,781.

Results

Quality of sequences

The ratios of transitions to transversions (Ts/Tv) were similar for the contemporary (2.89) and historic individuals (2.95), and both combined (2.94). As low damage was observed with the MapDamage recalculated quality scores (see below), and as we applied a minimum depth filter of 8, we elected to keep both transitions and transversions for further analysis. After filtering, the dataset included 780,604 total SNPs and 671,200 SNPs with minor allele frequency (MAF) > 0.05. Mean depth including missing sites per individual over the 780,604 SNPs was 16.2 for the contemporary specimens (ranging from 4.4 and 41.27), and 10.3 for the historic specimen (ranging from 3.9 to 35.3) (Table S1). The total number of SNPs and mean depth (without missing sites) for the analysed SNPs per individual, with individual heritage information for the 92 individuals are summarised in Table S1. The historic samples had, as expected due to post-mortem damage, slightly more substitutions than the contemporary samples for both the 3' and 5' ends which could lead to overestimates of diversity but it is limited just to the very end of the reads. At the 3´ ends the

historic samples had a mean substitution rate from 0.031 at the first site and 0.024 at the tenth site, for the contemporary samples the corresponding numbers were 0.017 at the first site and 0.023 at the tenth. The same pattern was seen at the 5´ ends for the first and tenth sites respectively, in the historic sample (0.031 and 0.023) and the contemporary sample (0.015 and 0.023) (Figure S1). The substitution rate is increased by 0.01 on average for the 5 bases at the end, so its impact on the overall heterozygosity is small or about 0.1% (e.g., expected to be 2*0.01*0.99 *10/150) due to those errors.

Diversity within samples

The island populations (Greenland and Iceland), have substantially lower diversity than the mainland populations: the observed heterozygosity per SNP for the island populations ranges from 0.17 to 0.19, except in the small historic Icelandic sample with 0.32, but they are around 0.27 in the mainland populations (Table 1). The observed heterozygosity per individual was 0.16 to 0.17, and in the mainland populations ranges from 0.22 to 0.28 (Wilcox exact test p-value < $2.2*10^{-16}$, Figure 2). A similar difference between individuals within the island and mainland populations was observed, the average IBS distance between individuals being ~0.14 vs. ~0.23 for the island and mainland populations, respectively (Table 1). The lowest diversity in the contemporary populations is found in Iceland (0.140 and 0.166, for IBS, and heterozygosity, respectively, Table 1). Similar mean proportions of heterozygous sites are observed in contemporary and historical samples from the same country when looking at (Figure 2). The largest change was observed within Iceland, though not significant, where it decreased from 0.173 (sd=0.019) to 0.161 (sd=0.014). Heterozygosities per site showed a temporal reduction in Denmark, Iceland, and Greenland (Table 1). All average IBS-distances were larger within the contemporary samples than within the historic samples indicating larger differences between individuals possibly as a result of drift or increased inbreeding. Furthermore, a comparison of the ranks of the genetic distances between individuals within the

temporal samples showed fewer differences within the historic samples except for the Icelandic samples (Wilcoxon test, p-value: GL = 1.659e-06, IS = 0.085, NO = 1.261e-09, DK= 0.017).

Inbreeding

Variation at each locus within populations showed little deviation between the observed and expected values or deviation from random mating, and thus no evidence of inbreeding (F_{IS}). However, when considering the proportion of expected heterozygous sites per individual in comparison to all individuals was similar (0.264-0.288) for all populations and sample ages, and as the corresponding observed values differs, the F_H values for the island populations indicate large inbreeding (ranging from 0.349-0.396), with contemporary Icelandic population having the highest levels of mean inbreeding, with individual inbreeding ranging from 0.310 to 0.548, and somewhat higher mean inbreeding than among the historic Icelandic specimens. Little or no inbreeding was found in the mainland populations (ranging from -0.22-.17), however, a couple of individuals displayed inbreeding (especially in the small contemporary Danish population) (Figure 2).

F_{ROH} displayed the same pattern of inbreeding as F_H with an R-value of 0.959 (Figure S2) when individuals with >50% missing data were excluded, as this was shown to affect the ROH length (Figure S6). The removal of individuals with >50% missing values left 78 individuals. The difference in values between F_{ROH} and F_H ranged from -0.11 to 0.15 (calculated as F_{ROH}-F_H) with sd=0.08. The relationship between observed heterozygosity based on heterozygotic sites, and depth of coverage, and missingness was examined to ensure this was not causing any errors and found no to a weak correlation (-0.08 and 0,33, respectively, Figure S3 and S4). Inspection of mean heterozygosity per site for each chromosome followed the overall mean pattern for the populations (Figure S5). The observed heterozygosity per site for each population followed the same picture as the overall heterozygosity.

Runs of homozygosity (ROH) analyses show that both contemporary island populations have elevated levels of ROH, as does the historic population from Greenland (Figure 3). A few of the historic individuals from all countries, and one contemporary Norwegian individual display extreme values and a large spread of ROH, which is probably due to missing data, rather than actual long ROH. This is supported by a comparison of the lengths of ROH with nucleotide missingness per individual, as it is found that missingness above ~0.5 is causing the ROH length to be inflated, in populations that were otherwise showing lower ROH lengths, and the two had a correlation of 0.57 (Figure S6). Correlation between ROH and depth of coverage was also checked, and only a weak correlation was found (-0.23, Figure S7). Further, both contemporary and historic samples from Greenland show signs of having experienced an older bottleneck as they have many ROH segments and long segments. The same is the case for the contemporary Icelandic sample, but it also shows recent consanguinity, as they are located below the diagonal relationship between the number and length of segments. The contemporary Danish sample also appears to show signs of recent consanguinity. Most historic specimens from Norway and most historic specimens from Denmark as well as the contemporary Norwegian and Estonian samples all show low levels of number and length of ROH segments, ranging from 28-160 and 25-265 Mb, respectively (Figure 3).

Genealogy

A tree or a genealogy based on identity-by-descent (IBD) shows two major clades which are further divided into two subclades (which split further, Figure 4). The two major clades include Iceland and Greenland in one and the mainland populations in the other. In the clade with Iceland and Greenland, we find a complete differentiation between Greenland and Iceland. For Iceland, we also see a complete split between the contemporary samples and the two historic individuals. In the Greenland cluster there is slightly more mixing between the two temporal groups, but still separation to some extent. An analysis of temporal samples in Iceland and Greenland, separately,

375 including only genotypes present in all individuals resulted in the same split (29,896 and 7,053 SNPs
376 were used, respectively). The mainland cluster splits into two subclades, all Danish contemporary
377 individuals except one are in one clade, and the rest of the populations in another clade. One
378 divergent Danish contemporary individual (DK_C_6) clusters with the Estonian sample. The
379 Estonian samples and the Turkish specimen are found on a divergent branch within the mainland
380 cluster but share a common ancestor with two Danish historic individuals. The remaining Danish
381 historical individuals are mixed in a cluster containing all the historic and contemporary Norwegian
382 individuals, which do not display a particular pattern (Figure 4).
383
384 *Population structure*

385 Assessing the divergence of samples from the different countries, based on F_{ST}, net IBS-distances
386 between the contemporary and the historic samples separately, as well as the principal component
387 analysis (PCA, described below), revealed clear differentiation. In accordance with the tree in Figure
388 4, the largest contemporary F_{ST} values are between the island and mainland samples (0.24-0.39,
389 Table 2). The F_{ST} for between the contemporary mainland samples are smaller (0.057-0.099) than
390 between the island populations (0.24). The temporal comparisons within countries display lower
391 differentiation, with the largest temporal difference found in Iceland and Denmark (F_{ST}, IBS-distance
392 with p-value in parentheses: Greenland 0.007, 0.002 (0); Iceland 0.096, 0.033 (0.01); Norway 0.014,
393 0.010 (0); and Denmark 0.069, 0.028 (0)). The distance metrics for the historic populations overall
394 follow the same pattern as the contemporary except for F_{ST} for the small Icelandic population which
395 shows the same distance to Norway and Denmark, as it does to Greenland (~0.190), possibly due
396 to the small size of the historic Icelandic sample. Both pairwise F_{ST} and IBS distances show larger
397 differentiation between the contemporary samples than between the historic samples.
398

399 The EIGENSOFT PCA analysis, based on 616,182 SNPs, resulted in a clear split between the mainland
400 and the two island populations, Greenland, and Iceland, on the first PCA-axis (explaining 45.3% of
401 the total variation), but separated the two on the second PCA-axis (explaining 15%) (Figure 5). The
402 contemporary Danish population (except one individual also deviating in the tree DK_C_6) is
403 distinct from the rest of the mainland, and a distinction is observed between the contemporary and
404 historic samples, both in Iceland and Greenland are found (more pronounced in Iceland) (Figure 5),
405 where the contemporary samples deviate more from the centre. The three PC axis 3-5 (explaining
406 10.2 to 4.9% of the total variation) did not reveal any geographical patterns but did underline the
407 temporal difference within countries, except in Norway, where the contemporary and historic
408 samples still had a large overlap.
409
410 All runs of admixture (*K*=2-15) (Figure 6, Figure S8-S15) converged per the criteria of the delta being
411 below 10^{-4} for five iterations in a row. *K*=2 (Figure 6a) separates the island and the mainland
412 populations (as also seen previously for the distances, along the first PCA-axis and by the major split
413 in the dendrogram), at *K*=3 (Figure 6b) Norway separates from the rest of the mainland populations.
414 For *K*=3-6 (Figure 6C-5E), further splits are observed: the historic Icelandic individuals get a separate
415 signature from the contemporary Icelandic (due to drift in the contemporary population), the
416 contemporary Danish population seem to be a mixed population (perhaps containing something
417 from an unsampled origin) and contains one individual with a signature like the Estonian individuals.
418 The historic Danish and Turkish individuals display the same signature. At *K*=7 (Figure 6F) and above
419 only more substructures within populations (and times) are found (Figure S8-S15).
420 *Deviation from equilibrium*
421 A large majority of the Tajima's D values in all populations are positive and skewed to the left,
422 especially for the island populations and the historic Danish sample (Table 3 and Figure S16). A large
423 fraction of the Tajima's D values in the island populations, and especially in Iceland, exceeds two

424 standard deviations, indicating significant values, and the large proportion of positive Tajima D
425 values reflects larger mean nucleotide diversities than expected based on the number of
426 segregating sites.

427

428 The differentiation between populations is spread all along the genome as summarized with F_{ST} for
429 each SNP (Figure S17). The most extreme values are found between the two temporal populations
430 from Iceland, and between the contemporary populations from Iceland and Greenland (Figure S17).

431

432 *Demographics*

433 The trees obtained with replicated runs with Treemix for different migration edges resulted in
434 similar outcomes. The tree with the largest number of migration events (m=5) showed the highest
435 proportion of explained variance (Figure 7 and S18). The tree displays a shallow structure between
436 the mainland populations with the Estonian population diverging earliest (Figure 4). The temporal
437 population pairs are most closely related to each other in the tree. The island populations, Iceland,
438 and Greenland have experienced the most drift as seen from the x-axis drift parameter. The island
439 populations are most closely related to the historic Turkish individual but that might rather reflect
440 shared ancestral variation among unsampled and even extinct populations within southern and
441 western Europe than any recent migration as discussed below. Four of the five listed migration
442 events are coming from bald eagle into the two Norwegian and two Danish populations, with the
443 strongest signal into the contemporary Danish population, which is probably also a signal from
444 ancestral polymorphism as this cannot be explained by any hybridization. The fifth and strongest
445 migration signal is from Norway to the common ancestor of the island populations (Figure 7). The
446 proportion of explained variance and likelihood was also checked for migration edges (m) 6-8.
447 However, as the mean proportion of explained variance only increased with 0.0001 per migration
448 edge and the mean likelihood did not increase substantially, these results were not included.

All populations examined with Stairway plot show a general loss in effective population size from 10.000 years ago until today (Figure 8), but there are two especially large drops. The first big drop happened 5-8,000 years ago for all populations when looking at the mean effective population size, the most recent large drop happened around 1,000 years ago for all populations. The same decline pattern is also seen over the last 500,000 years (Figure S19 and S20), with a large drop in all population 35-55,000 years ago when looking at the mean effective population size, and the oldest drops exceeding 200,000 years, spanning several hundred thousand of years, depending which population is analysed. Lowering the mutation rate to $1*10^{-9}$ per site per year expands the fluctuations events inversely (e.g., by the product of 2.9) (Figure S21 and S22). The most recent drop in population size is pushed to ~2-3,000 years ago, the second ~12-18,000 years BP, the third ~100,000 years ago, and then the oldest and largest found 500-900,000 years ago, all considering just the mean effective population size. Large 2.5-97.5% confident intervals (CI) are found back in time (dashed lines), but the patterns are similar for the different populations, and the current day estimates have very low CI. The current day estimates of the effective population sizes obtained with the higher mutation rate with 95% confidence intervals are for the different countries the following: Greenland 3 (1-21), Iceland 3 (1-15), Norway 8 (1-54) and Denmark 6 (1-38). Slightly higher estimates were obtained with the lower mutation rate: Greenland 8 (1-50), Iceland: 6 (1-46), Norway: 17 (3-115), Denmark: 16 (2-85). The CI increased slightly for all populations (Figure S21 and S22).

Discussion

Overall, the contemporary nuclear genome samples from Greenland, Iceland, Norway, Denmark, and Estonia are differentiated by country, and with most of the historic samples being closest related to the contemporary samples from the same country. For both heterozygosity and diversity,

473 we observed that the island populations have lower diversity than the mainland. Greenland,
474 Iceland, Norway, and Denmark all show a considerable population reduction over time until today.
475
476 The contemporary nuclear genome samples from Greenland, Iceland, Norway, Denmark, and
477 Estonia display differentiation between the countries as expected by their geography, i.e., the
478 island populations are genetically more similar to each other than they are to the mainland
479 populations, and each country makes up a monophyletic group. This is in contrast to what has been
480 found in the mitochondrial DNA analysis for the populations in Greenland, Iceland, and Norway
481 hitherto, which display two polyphyletic clades each with lineages from the three countries (Hailer
482 et al. 2007; Hansen et al., 2021, in review). The mitochondrial variation deviates from neutral
483 expectation (Hansen et al. 2021, in review), and maybe affected by selection on the W-chromosome
484 due to shared inheritance and linkage disequilibrium in birds between the W-chromosome and
485 mitochondrial DNA.
486
487 Inbreeding was examined using three different approaches, F_{IS}, F_H, and F_{ROH}. Deviation of the
488 observed values depends on the reference population. In the case where individuals were
489 compared solely to their own populations no clear deviation from the expected population
490 variation is observed and here the F_{IS} does not reveal any deviation from Hardy-Weinberg or
491 inbreeding, and neither did F_H when calculated separately for each population. However, when
492 comparing each individual with the variation within the total sample a clear difference is observed.
493 Here the inbreeding (F_H) for the island populations are large unlike for the mainland populations,
494 where the former have much lower observed proportions of heterozygous sites than the mainland
495 populations. The two coefficients based on the individual genomic patterns F_H and F_{ROH} gave similar
496 outcomes and showed substantial inbreeding in the island population, and they have been shown
497 to reflect well the “true” inbreeding (Kardos et al. 2015, Forutan et al. 2018). Thus, the

498 contemporary island populations show clear signs of reduced diversity, potential substantial and
499 increasing inbreeding, and increased drift compared to the mainland populations. Larger distances
500 or diversity among individuals are observed in contemporary and more inbred populations with less
501 observed heterozygosity than in the historic populations. Such changes could result from larger
502 scatter due to fewer heterozygous sites in the island populations than in the mainland and longer
503 runs of homozygosity. Similarly, genetic variance has been found to increase with inbreeding where
504 genetic factors will segregate among different lines in linkage disequilibrium (Wang et al. 1998),
505 contrary to the well-established experimentally and theoretically result that a population
506 bottleneck reduces gene diversity and genetic heterozygosity (Crow and Kimura 1970). Iceland
507 shows signs of recent and more historic bottlenecks, which could be both due to the founder event
508 in Iceland as well as the recent known bottleneck in the 19th and 20th centuries (Skarphéðinsson
509 2003). Greenland too shows signs of an ancient bottleneck that could also be due to the founder
510 event, but less so for a recent bottleneck. Analyses of the genetic variation reveal stronger signs of
511 ancient bottlenecks and lower diversity in the island populations than in the mainland populations.
512 The founder event is supported by the analysis on migration, though it revealed that the islands
513 share the most recent ancestry with the historic Turkish individual, but have had substantial input
514 from Norway. Previous studies suggest eastern and western refugia for the eagles during the last
515 glacial period and that the population in Greenland and Iceland originated from north-western
516 Europe (Hailer et al. 2007). The Turkish individual could be a representative of southern and
517 western Europe variants and could reflect unexplored phylogeographic patterns shaped by refugia
518 during the last glacial period of ice age which may have been preserved in the isolated island
519 populations and with an admixture of the ancestors of the Norwegian populations. Wider sampling
520 from western and southern Europe is warranted to clarify this and the phylogeographic patterns in
521 Europe. The signal of migration from the bald eagle into the white-tailed eagle could reflect shared
522 ancestral polymorphism between the species (Pritchard et al. 2000).

In general, the temporal samples within countries are similar, except for the samples from Denmark and Iceland. In Denmark, we find a difference between the historic and contemporary samples in the PCA and the admixture composition. This is not surprising as the population went extinct in Denmark and has since been recolonised (Skelmose and Larsen 2021), and thus the two temporal populations do not necessarily share a direct recent common ancestor, which is supported by their position in the tree and the composition in the admixture plot. Iceland also displays a difference, however here the population did not go extinct, but did go through a strong recent bottleneck (Petersen 1998, Skarphéðinsson 2013), and thus we find a difference for the contemporary and historic samples both in the PCA, admixture, and dendrogram. These analyses also indicate that the contemporary samples have experienced drift as larger distances are observed between the contemporary populations than between the historic samples. Furthermore, less heterozygosity and higher inbreeding are observed in the contemporary Icelandic sample compared to the historic sample. Although this difference may be biased due to the small historic sample size, missing data, and overestimation of heterozygosity due to post-mortem damage, the signal is consistent for the different assessments of diversity and is also supported by the effective population size analysis that shows a continues to fall up until today. The post-mortem damage was just restricted to the 5 bp at the ends of reads and had only a minor potential effect on the heterozygosity. Thus, even though the white-tailed eagle is long-lived, and the recent bottleneck is moderately short (ca. 150 years) it has affected the variation within the Icelandic population.

A large reduction in population size was estimated in comparison to the ancient population sizes for all analysed populations i.e., Greenland, Iceland, Norway, and Denmark, which indicate an overall reduction of population size and introduction of the population structure. This is also supported by the positive Tajima's D, especially for the island populations as lack of rare alleles can indicate population contraction. The oldest drop in population size happened 2-400,000 or up to

900,000 years before present (BP), for the mutation rates of $2.3*10^{-9}$ and $1*10^{-9}$ per site per year, respectively. However, three other drops happened in more recent times. I) The first happened 35-55,000 years BP, using a mutation rate of $2.3*10^{-9}$ per site per year or around 100,000 years BP with a mutation rate of $1*10^{-9}$ per site per year, both estimates are within or at the onset of the last glacial period (Lisiecki and Raymo 2005), and thus this could show a drop in the overall species effective size or a split in the species, driven by the species being isolated in two (or three) refugia, which has been suggested (Hailer et al. 2007, Langguth et al. 2013; Hansen et al., 2021 in review). II) The drop that happened for all populations ~5-8,000 years BP (or 12-18,000 years BP with the $1*10^{-9}$ per site per year mutation rate) could indicate the establishment of the different populations which is found in the PCA, ADMIXTURE and IBD analysis, as the drop is found soon after the end of the last glacial period within these countries (Clark and Mix 2002). The oldest of these estimates in the stairway analysis is still when the countries were covered by ice but considering the CI and difference between countries, this still seems plausible. And III) the latest drop in population size 1,000 (or 2-3,000 years ago) could be due to human expansion in northern Europe (Kremer 1993), including settlement in Iceland around year 871 (Batt et al. 2015), and white-tailed eagle bones have been found in human settlements and may thus have been hunted by humans (Price et al. 2018). The genome-wide mutation rate in the white-tailed eagles or similar species could provide a better estimate of these datings.

Finally, other than these three major drops in population size, we also see a drop in effective population size for all populations during the last centuries, which could well be an anthropogenic effect following the industrial revolution, including human persecution in the 19th century and organic toxic pollutants known to have had a detrimental effect on reproductive success in the eagles during the 20th century (Bijleveld 1974, Love and Ball 1979, Helander et al. 1982, 2002; Walker et al. 2009).

573 The overall pattern of the effective population sizes follows roughly the main known climatic and
574 anthropogenic influences, but the confidence intervals are large for the historic estimates, and
575 there is uncertainty too as the true mutation rate for white-tailed eagle is not known. The estimates
576 for the current effective population sizes are extremely low for all populations (< 50) and below the
577 effective population size of 50, the threshold value which has been suggested for populations to
578 avoid inbreeding depression in the short term (Franklin 1980, Soulé 1980). Similar estimates have
579 been observed for other species e.g., the Madagascar fish-eagle which ratio of effective population
580 size to populations census size (N_e/N_c) is about 10%, which follow the general rule despite variation
581 among taxa (Frankham 1995), as the CI here puts the populations not far from this (with Norway at
582 the lowest reaching 2.5-5% of the census size with the largest CI, and Iceland 3.5% considering adult
583 birds). The small current-day population sizes could lead to several unfavourable scenarios for the
584 populations; they could not be able to effectively purge deleterious mutation, and beneficial
585 mutations have a higher risk of being lost due to drift (Nielsen and Slatkin 2013). Further, the small
586 population size may make them less able to adapt per the "500 rule", which has been suggested as
587 the sufficient minimum to retain evolutionary potential (Franklin 1980, Soulé 1980) e.g., in case of
588 climate, habitat, or prey/predator change. Our analysis further revealed a reduction in
589 heterozygosity, an increasing inbreeding, and an upsurge in drift during the 20th century for the
590 small Icelandic population suggesting its existence may be at risk and it may suffer from inbreeding
591 depression (Hartl and Clark 2007, Nielsen and Slatkin 2013). Although the Icelandic population has
592 been recovering for the last 40 years, where the number of breeding pairs per year has increased
593 from 20 pairs to about 80, the reproductive rate of the Icelandic is low, only 0.5 chicks per pair per
594 year (Skarphéðinsson 2003, 2013; Evans et al. 2009) or about one-third of the rate in Sweden
595 (Helander et al. 2013). And though the heterozygosity loss in Iceland is small, a loss of just 5-10%
596 over 100 years has been suggested to cause a risk of population extinction (Allendorf and Ryman
597 2002), though there are examples of species going through ancient bottlenecks but persisting
598 (O'Brien et al. 2017). All these results suggest that an increased conservation effort can become

necessary, in all the analysed populations, but especially in the small, isolated Icelandic population. Further work linking the genetic variation to variation in fitness-related traits could show whether the eagle populations in Iceland and Greenland do suffer from inbreeding depression and whether admixture of genetic variants from mainland populations should be considered.

Acknowledgement

The study was supported by Research Grant nr 185280-052 from The Icelandic Research Council, the Doctoral student found of the University of Iceland and The University of Iceland Research Fund.

Thanks to deCODE genetics for providing sequencing and computational resources.

References

Alexander, D. H., J. Novembre, and K. Lange (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Research 19:1655–1664.

Allendorf, F. W., and N. Ryman (2002). The role of genetics in population viability analysis. In Population viability analysis. University of Chicago Press, pp. 50–85.

Allentoft, M. E., M. Sikora, K. G. Sjögren, S. Rasmussen, M. Rasmussen, J. Stenderup, P. B. Damgaard, H. Schroeder, T. Ahlström, L. Vinner, A. S. Malaspinas, et al. (2015). Population genomics of Bronze Age Eurasia. Nature 522:167–172.

Babraham Bioinformatics (2010). FastQC. [Online.] Available at http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Batt, C. M., M. M. E. Schmid, and O. Vésteinsson (2015). Constructing chronologies in Viking Age Iceland: Increasing dating resolution using Bayesian approaches. Journal of Archaeological Science 62:161–174.

Bijleveld, M. (1974). Birds of Prey in Europe. Macmillan Education UK, London.

Birdlife International (2020). Haliaeetus albicilla. The IUCN Red List of Threatened Species 2020 e.T2269513.

Boertmann, D., and C. Bay (2018). The Greenland red-list 2018.

Broad Institute (2020). Picard - https://broadinstitute.github.io/picard/. *https://broadinstitute.github.io/picard/* .

Carøe, C., S. Gopalakrishnan, L. Vinner, S. S. T. Mak, M. H. S. Sinding, J. A. Samaniego, N. Wales, T. Sicheritz-Pontén, and M. T. P. Gilbert (2018). Single-tube library preparation for degraded <scp>DNA</scp>. Methods in Ecology and Evolution 9:410–419.

631 Clark, P. U., and A. C. Mix (2002). Ice sheets and sea level of the Last Glacial Maximum.
632 Quaternary Science Reviews 21:1–7.

633 Crow, J. F., and M. Kimura (1970). An introduction to population genetics theory. Harper and Row,
634 New York.

635 Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G.
636 Lunter, G. T. Marth, S. T. Sherry, G. McVean, and R. Durbin (2011). The variant call format
637 and VCFtools. Bioinformatics 27:2156–2158.

638 Eggertsson, H. P., S. Kristmundsdottir, D. Beyter, H. Jonsson, A. Skuladottir, M. T. Hardarson, D. F.
639 Gudbjartsson, K. Stefansson, B. V. Halldorsson, and P. Melsted (2019). GraphTyper2 enables
640 population-scale genotyping of structural variation using pangenome graphs. Nature
641 Communications 10:1–8.

642 Ehmsen, E., L. Pedersen, H. Meltofte, T. Clausen, and & Timme Nyegaard (2011). The occurrence
643 and reestablishment of White-tailed Eagle and Golden Eagle as breeding birds in Denmark.

644 Elts, J., A. Leito, M. Leivits, L. Luigujõe, R. Nellis, M. Ots, I. Tammekänd, and Ü. Väli (2019). Eesti
645 lindude staatus, pesitsusaegne ja talvine arvukus 2013-2017. Hirunda 32:1–29.

646 Evans, R. J., J. D. Wilson, A. Amar, A. Douse, A. MacLennan, N. Ratcliffe, and D. P. Whitfield (2009).
647 Growth and demography of a re-introduced population of White-tailed Eagles Haliaeetus
648 albicilla. Ibis 151:244–254.

649 Forutan, M., S. Ansari Mahyari, C. Baes, N. Melzer, F. S. Schenkel, and M. Sargolzaei (2018).
650 Inbreeding and runs of homozygosity before and after genomic selection in North American
651 Holstein cattle. BMC Genomics 19:1–12.

652 Frankham, R. (1995). Effective population size/adult population size ratios in wildlife: a review.
653 Genetics Research 66:95–107.

654 Franklin, I. R. (1980). Evolutionary Changes in Small Populations. In Conservation Biology: An
655 Evolutionary-Ecological Perspective (M. E. Soulé and B. M. Wilcox, Editors). Sinauer,
656 Sunderland, pp. 135–149.

657 Gilbert, M. T. P., D. I. Drautz, A. M. Lesk, S. Y. W. Ho, J. Qi, A. Ratan, C. H. Hsu, A. Sher, L. Dalén, A.
658 Götherström, L. P. Tomsho, et al. (2008). Intraspecific phylogenetic analysis of Siberian
659 woolly mammoths using complete mitochondrial genomes. Proceedings of the National
660 Academy of Sciences of the United States of America 105:8327–8332.

661 Gopalakrishnan, S., J. A. Samaniego Castruita, M. H. S. Sinding, L. F. K. Kuderna, J. Räikkönen, B.
662 Petersen, T. Sicheritz-Ponten, G. Larson, L. Orlando, T. Marques-Bonet, A. J. Hansen, et al.
663 (2017). The wolf reference genome sequence (Canis lupus lupus) and its implications for
664 Canis spp. population genomics. BMC Genomics 18:1–11.

665 Hailer, F., B. Helander, A. O. Folkestad, S. A. Ganusevich, S. Garstad, P. Hauff, C. Koren, V. B.
666 Masterov, T. Nygård, J. A. Rudnick, S. Shiraki, et al. (2007). Phylogeography of the white-
667 tailed eagle, a generalist with large dispersal capacity. Journal of Biogeography 34:1193–
668 1206.

669 Hailer, F., B. Helander, A. O. Folkestad, S. A. Ganusevich, S. Garstad, P. Hauff, C. Koren, T. Nygard,
670 V. Volke, C. Vila, and H. Ellegren (2006). Bottlenecked but long-lived: high genetic diversity
671 retained in white-tailed eagles upon recovery from population decline. Biology Letters
672 2:316–319.

673 Hansen, C. C. R., Baleka, S., Guðjónsdóttir, S. M., Rasmussen, J. A., Ballesteros, J. A. C.,
674 Hallgrimsson, G. T., Stefansson, R. A., von Schmalensee, M., Sharphédinsson, K. H.,
675 Labansen, A. L., Leivits, M., Skelmose, K., Sonne, C., Dietz, R., Boertmann, D., Eulaers, I.,
676 Martin, M. D., Pálsson, S (2021). Distinctive mitogenomic lineages within populations of
677 white-tailed eagles (Haliaeetus albicilla). Ornithology accepted.

678 Hansen, K. (1979). Status over bestanden af Havørn *Haliaeetus albicilla groenlandicus* Brehm i
679 Grønland i årene 1972-74. Dansk Ornitologisk Forenings Tidsskrift 73:107–130.

680 Hartl, D. L., and A. G. Clark (2007). Principles of Population Genetics. 4th edition. Sinauer
681 Associates Inc. Publishers.

682 Helander, B. B., M. Olsson, L. Reutergårdh, and L. Reutergardh (1982). Residue levels of
683 organochlorine and mercury compounds in unhatched eggs and the relationships to
684 breeding success in white-tailed sea eagles Haliaeetus albicilla in Sweden. Ecography 5:349–
685 366.

686 Helander, B., A. Bignert, C. Herrmann, and T. Stjernberg (2013). White-tail eagle productivity.
687 HELCOM Core Indicator Report.

688 Helander, B., A. Olsson, A. Bignert, L. Asplund, and K. Litzén (2002). The Role of DDE, PCB,
689 Coplanar PCB and Eggshell Parameters for Reproduction in the White-tailed Sea Eagle
690 (Haliaeetus albicilla) in Sweden. AMBIO: A Journal of the Human Environment 31:386–403.

691 Hoban, S., M. Bruford, J. D. Urban, M. Lopes-fernandes, M. Heuertz, P. A. Hohenlohe, I. Paz-vinas,
692 P. Sjögren-gulve, G. Segelbacher, C. Vernesi, S. Aitken, et al. (2020). Genetic diversity targets
693 and indicators in the CBD post-2020 Global Biodiversity Framework must be improved.
694 Biological Conservation 248:108654.

695 Hoban, S., C. D. Campbell, J. M. da Silva, R. Ekblom, W. C. Funk, B. A. Garner, J. A. Godoy, F.
696 Kershaw, A. J. MacDonald, J. Mergeay, M. Minter, et al. (2021). Genetic diversity is
697 considered important but interpreted narrowly in country reports to the Convention on
698 Biological Diversity: Current actions and indicators are insufficient. Biological Conservation
699 261:109233.

700 Honnen, A.-C. C., F. Hailer, N. Kenntner, I. Literák, L. Dubská, and F. E. ZACHOS (2010).
701 Mitochondrial DNA and nuclear microsatellites reveal high diversity and genetic structure in
702 an avian top predator, the white-tailed sea eagle, in central Europe. Biological Journal of the
703 Linnean Society 99:727–737.

704 Jackson, R., J. Arneborg, A. Dugmore, C. Madsen, T. McGovern, K. Smiarowski, and R. Streeter
705 (2018). Disequilibrium, Adaptation, and the Norse Settlement of Greenland. Human Ecology
706 46:665–684.

707 Jais, M. (2020). White-tailed Eagle. [Online.] Available at http://europeanraptors.org/white-tailed-
708 eagle/.

709 Jónsson, H., A. Ginolhac, M. Schubert, P. L. F. Johnson, and L. Orlando (2013). mapDamage2.0:
710 fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics
711 29:1682–1684.

712 Kardos, M., G. Luikart, and F. W. Allendorf (2015). Measuring individual inbreeding in the age of
713 genomics: Marker-based measures are better than pedigrees. Heredity 115:63–72.

714 Korneliussen, T. S., A. Albrechtsen, and R. Nielsen (2014). ANGSD: Analysis of next generation
715 sequencing data. BMC bioinformatics 15:356.

716 Kremer, M. (1993). Population Growth and Technological Change: One Million B.C. to 1990. The
717 Quarterly Journal of Economics 108:681–716.

718 Langguth, T., A.-C. C. Honnen, F. Hailer, T. Mizera, S. Skoric, Ü. Väli, and F. E. Zachos (2013).
719 Genetic structure and phylogeography of a European flagship species, the white-tailed sea
720 eagle Haliaeetus albicilla. Journal of Avian Biology 44:263–271.

721 Leigh, D. M., A. P. Hendry, E. Vázquez-Domínguez, and V. L. Friesen (2019). Estimated six per cent
722 loss of genetic variation in wild populations since the industrial revolution. Evolutionary
723 Applications 12:1505–1512.

724 Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and
725 population genetical parameter estimation from sequencing data. Bioinformatics 27:2987–
726 2993.

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

729 Lisiecki, L. E., and M. E. Raymo (2005). A Pliocene-Pleistocene stack of 57 globally distributed
730 benthic δ 18O records. Paleoceanography 20:1–17.

731 Liu, X., and Y. X. Fu (2020). Stairway Plot 2: demographic history inference with folded SNP
732 frequency spectra. Genome Biology 21:1–9.

733 Liu, X., and Y.-X. Fu (2015). Exploring population size changes using SNP frequency spectra. Nature
734 genetics 47:555–9.

735 Lõhmus, A. (1998). Eesti kotkaste argipäev. Eesti Loodus 5/6:210–213.

736 Love, J. A. A., and M. E. E. Ball (1979). White-tailed sea eagle Haliaeetus albicilla reintroduction to
737 the Isle of Rhum, Scotland, 1975-1977. Biological Conservation 16:23–30.

738 Lynch, M. (2010). Evolution of the mutation rate Michael Lynch. Trends Genet 26:345–352.

739 Lyngs, P. (2003). Migration and winter ranges of birds in Greenland - An analysis of ringing
740 recoveries. Dansk Ornitologisk Forenings Tidsskrift 97:1–167.

741 Manichaikul, A., J. C. Mychaleckyj, S. S. Rich, K. Daly, M. Sale, and W.-M. Chen (2010). Robust
742 relationship inference in genome-wide association studies. Bioinformatics 26:2867–2873.

743 McQuillan, R., A. L. Leutenegger, R. Abdel-Rahman, C. S. Franklin, M. Pericic, L. Barac-Lauc, N.
744 Smolej-Narancic, B. Janicijevic, O. Polasek, A. Tenesa, A. K. MacLeod, et al. (2008). Runs of
745 Homozygosity in European Populations. American Journal of Human Genetics 83:359–372.

746 Nei, M. (1977). F-statistics and analysis of gene diversity in subdivided populations. Ann. Hum.
747 Gent. 41:225–233.

748 Nei, M., and W. H. Li (1979). Mathematical model for studying genetic variation in terms of
749 restriction endonucleases. Proceedings of the National Academy of Sciences of the United
750 States of America 76:5269–5273.

751 Nielsen, R., T. Korneliussen, A. Albrechtsen, Y. Li, and J. Wang (2012). SNP calling, genotype
752 calling, and sample allele frequency estimation from new-generation sequencing data. PLoS
753 ONE 7(7):e37558.

754 Nielsen, R., and M. Slatkin (2013). An Introduction to Population Genetics: Theory and
755 Applications. Sinauer Associates, Inc. .

O'Brien, S. J., W. E. Johnson, C. A. Driscoll, P. Dobrynin, and L. Marker (2017). Conservation Genetics of the Cheetah: Lessons Learned and New Opportunities. Journal of Heredity 108:671–677.

Patterson, N., A. L. Price, and D. Reich (2006). Population Structure and Eigenanalysis. PLoS Genetics 2:e190.

Petersen, Æ. (1998). Haförn. In Íslenskir Fuglar (S. Harðardóttir, Editor). Vaka-Helgafell hf., Reykjavik, pp. 120–123.

Price, A. L., N. J. Patterson, R. M. Plenge, M. E. Weinblatt, N. A. Shadick, and D. Reich (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nature Genetics 38:904–909.

Price, T. D., K. Ritchie, K. J. Gron, A. B. Gebauer, and J. Nielsen (2018). Asnæs Havnemark: a late Mesolithic Ertebølle coastal site in western Sjælland, Denmark. Danish Journal of Archaeology 00:1–22.

Pritchard, J. K., M. Stephens, and P. Donnelly (2000). Inference of population structure using multilocus genotype data. Genetics 155:945–959.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, and P. C. Sham (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. The American Journal of Human Genetics 81:559–575.

Randla, T., and A. Õun (1980). Kaljukotkas ja merikotkas Eestis 1970-ndail aastail. Eesti Loodus 4:512–515.

Salomonsen, F. (1979). Ornithological and ecological studies in SW Greenland. Meddelelser om Grønland 204.

Schubert, M., S. Lindgreen, and L. Orlando (2016). AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Research Notes 9:88.

Shriner, D., F. Tekola-Ayele, A. Adeyemo, and C. N. Rotimi (2014). Genome-wide genotype and sequence-based reconstruction of the 140,000 year history of modern human ancestry. Scientific Reports 4:6055.

Skarphéðinsson, K. (2003). Sea eagles in Iceland: Population trends and reproduction. In Sea eagle 2000. Proceedings from an intemational conference at Björkö, Sweden, 13-17 September 2000. (B. O. Helander, M. Marquiss and W. Bowerman, Editors). Swedish Society for Nature Conservation/SNF & Åtta.45 Tryckeri AB. Stockholm, Sweden, pp. 31–38.

Skarphéðinsson, K. H. (2013). Haförninn. Fuglavernd (Fuglaverndarfélag íslands), Reykjavik.

Skelmose, K., and Ole. F. Larsen (2021). Projekt Ørn – Årsrapport 2020, DOF BirdLife Danmark.

Smeds, L., A. Qvarnström, and H. Ellegren (2016). Direct estimate of the rate of germline mutation in a bird. Genome Research 26:1211–1218.

Sokal, R. R., and F. J. Rohlf (2012). Biometry: The Principles and Practice of Statistics in Biological Research. W. H. Freeman, New York.

Soulé, M. E. (1980). Thresholds for Survival: Maintaining Fitness and Evolutionary Potential. In Conservation Biology: An Evolutionary-Ecological Perspective (M. E. Soulé and B. M. Wilcox, Editors). Sinauer, Sunderland.

797 Tajima, F. (1989). Statistical Method for Testing the Neutral Mutation Hypothesis by DNA
798 Polymorphism. Genetics 123:585–595.

799 Treinys, R., D. Dementavičius, S. Rumbutis, S. Švažas, D. Butkauskas, A. Sruoga, and M. Dagys
800 (2016). Settlement, habitat preference, reproduction, and genetic diversity in recovering the
801 white-tailed eagle Haliaeetus albicilla population. Journal of Ornithology 157:311–323.

802 Walker, L. A., S. J. Beith, A. J. Lawlor, C. Moeckel, M. G. Peréira, E. D. Potter, and R. F. Shore
803 (2009). Persistent Organic Pollutants (POPs) and inorganic elements in predatory bird livers
804 and eggs 2007 to 2009: a Predatory Bird Monitoring Scheme (PBMS) Report.

805 Wang, J., A. Caballero, and W. G. Hill (1998). The effect of linkage disequilibrium and deviation
806 from Hardy-Weinberg proportions on the changes in genetic variance with bottlenecking.
807 Heredity 81:174–186.

808 Weir, B., and C. C. Cockerham (1984). Estimating F-Statistics for the Analysis of Population
809 Structure Author (s): B . S . Weir and C . Clark Cockerham. Evolution 38:1358–1370.

810 Zheng, X., D. Levine, J. Shen, S. M. Gogarten, C. Laurie, and B. S. Weir (2012). A high-performance
811 computing toolset for relatedness and principal component analysis of SNP data.
812 Bioinformatics 28:3326–3328.

813

814

815 Data accessibility

816 Data has been submitted to DRYAD and will be released when the manuscript is published at

817 https://doi.org/10.5061/dryad.fqz612jt8

818 Author contribution

819 Conceived the idea: CCRH, and SP; Performed the experiments: CCRH, JAR, JACB, MHSS, MDM, and

820 SP; Analysed the data: CCRH, and SP; wrote the paper: CCRH and SP; Commented on the paper:

821 JAR, JACB, MHSS, GTH, RAS, MvS, KHS, ALL, ML, CS, RD, KS, DB and IE, MDM, AH, MTPG; Contributed

822 substantial materials, resources or funding: CCRH, GTH, RAS, MvS, KHS, ALL, ML, CS, RD, KS, DB, IE,

823 MDM, AH, MTPG and SP.

824 Competing interests

825 The authors declare no competing financial interests.

826

Tables and figures

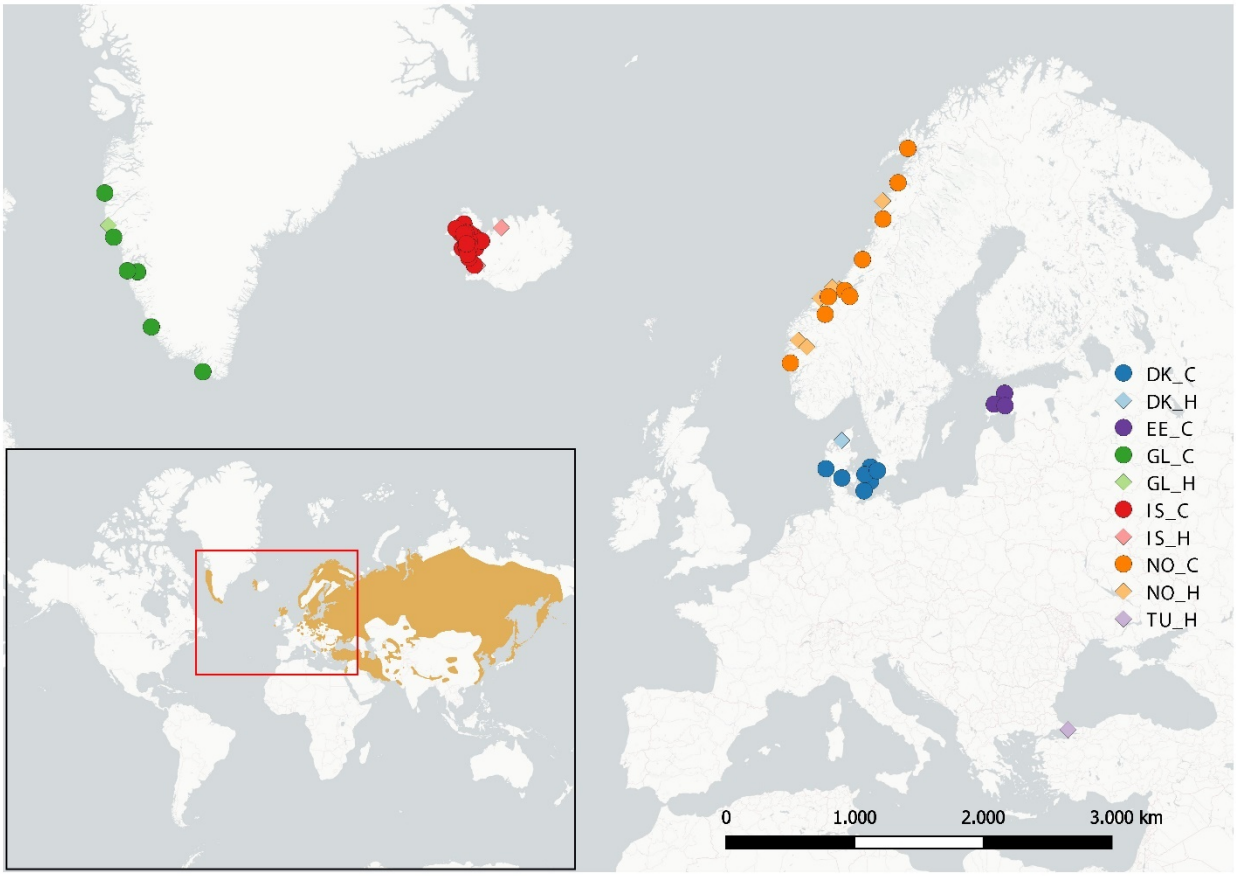


Figure 1. Maps of the sample sites and the species range. Locations of known sampling sites for the 92 white-tailed eagle individuals are marked with specific colours per country. Dots represent contemporary samples, and diamonds represent historic samples. The map in the corner shows the species distribution in orange. The red square is the part that makes up the larger map. GL_C=contemporary Greenland, GL_H=historic Greenland, IS_C=contemporary Iceland, IS_H=historic Iceland, NO_C=contemporary Norway, NO_H=historic Norway, DK_C=contemporary Denmark, DK_H=historic Denmark, EE_C=contemporary Estonia, TU_H=historic Turkey.

Table 1. Molecular diversity per country and overall, for contemporary (C) and historic (H) samples. Sample size (N). Expected and observed heterozygosity (H_E and H_O, respectively) calculated per SNP per population with variance in parentheses, and per population inbreeding coefficient (F_{IS}), and mean IBS-distance between individuals.

Country	Temporal sample	N	H_e	H_o	F_{IS}	Mean IBS distance
Greenland	C	12	0.163 (0.041)	0.177 (0.059)	-0.089	0.142
	H	8	0.162 (0.048)	0.189 (0.086)	-0.166	0.133
Iceland	C	25	0.158 (0.040)	0.166 (0.051)	-0.05	0.140
	H	2	0.164 (0.126)	0.317 (0.493)	NA	0.107
Norway	C	12	0.277 (0.034)	0.297 (0.056)	-0.072	0.238
	H	13	0.269 (0.040)	0.298 (0.074)	-0.106	0.214
Denmark	C	11	0.266 (0.037)	0.278 (0.057)	-0.044	0.233
	H	5	0.277 (0.080)	0.373 (0.235)	-0.351	0.218

Estonia	C	3	0.274 (0.059)	0.337 (0.137)	NA	0.236
Turkey	H	1	0.266 (0.195)	0.533 (0.782)	NA	-
Overall		92	0.268 (0.003)	0.213 (0.011)	0.207	0.265

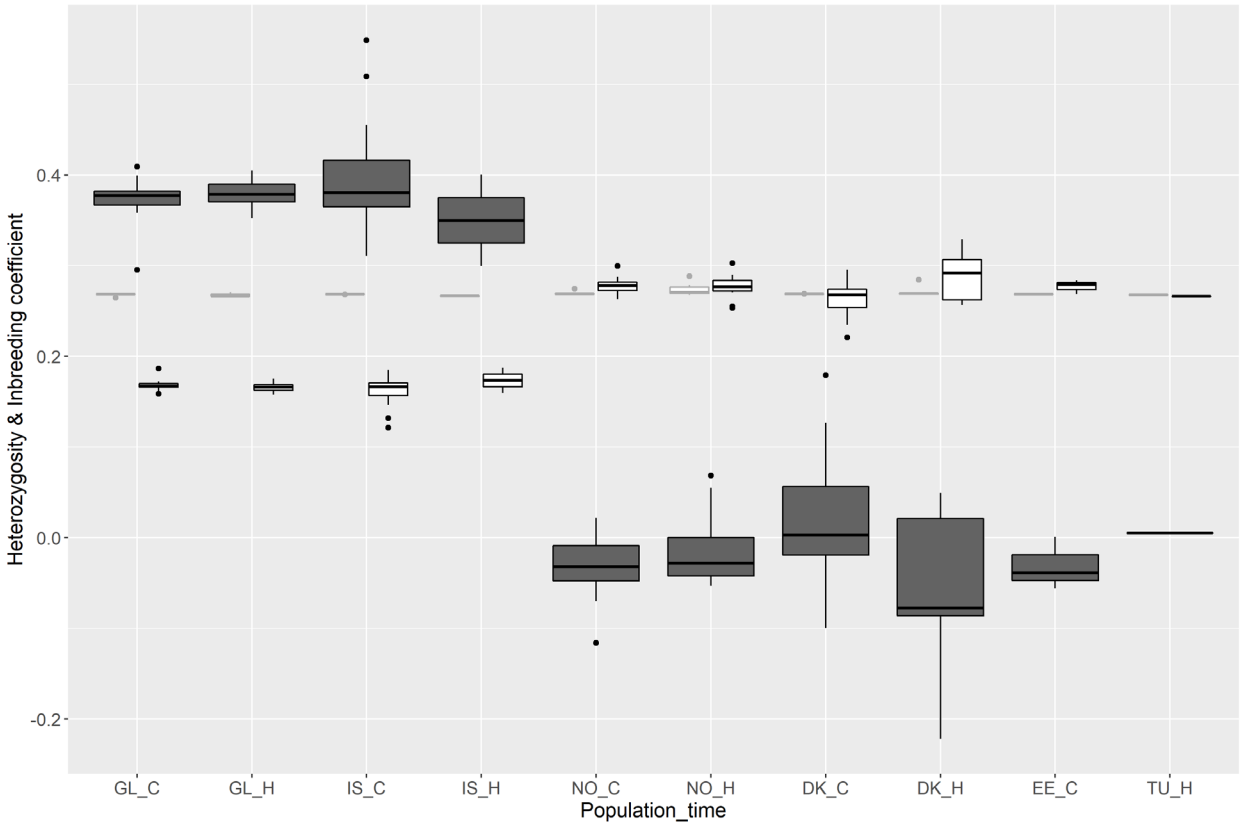


Figure 2. Deviation from random mating within samples. Three boxes are shown for each temporal sample per country: Narrow boxes present expected (shaded grey) and observed (black not filled) heterozygosity per individual. Wide boxes present the inbreeding coefficient (dark grey) per individual "C" refers to contemporary samples and "H" to historic samples. DK=Denmark, EE=Estonia, GL=Greenland, IS=Iceland, NO=Norway, TU=Turkey.

Figure 3. Runs of homozygosity (ROH) for all 92 individuals. The x-axis displays the length of ROHs in megabases (Mb), the y-axis show number of ROH segments. Colours indicate country and temporal category (Pop_Time). C: contemporary samples, H: historic samples. DK=Denmark, EE=Estonia, GL=Greenland, IS=Iceland, NO=Norway, TU=Turkey.

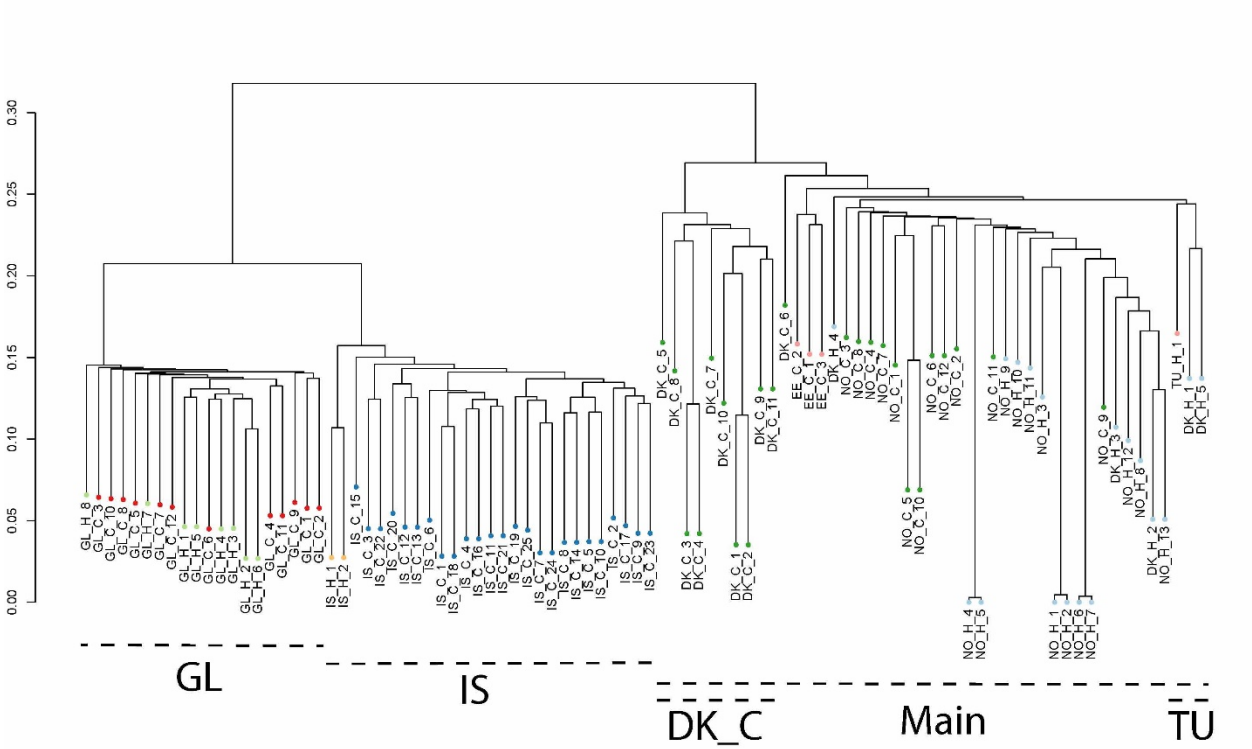


Figure 4. Dendrogram displaying identity-by-descent (IBD) between all samples. Abbreviations refer to a country or the area of origin and the temporal samples GL: Greenland, IS: Iceland, Main: mainland samples, NO: Norway, DK: Denmark, EE: Estonia, Tu: Turkey, C: contemporary samples, and H: historic samples.

Table 2. Mean F_{ST} above the horizontal line, and Identity by State distances below (IBS), with p-value above the diagonal. Left only comparison of contemporary samples, right only comparison of historic. P-value, given for the IBS-distances above the diagonal are based on 100 permutations. Abbreviations refer to country: GL: Greenland, IS: Iceland, NO: Norway, DK: Denmark, EE: Estonia, Tu: Turkey.

Contemporary

F_{st}	GL	IS	NO	DK	EE
GL	NA				
IS	0.246	NA			
NO	0.316	0.351	NA		
DK	0.348	0.382	0.099	NA	
EE	0.367	0.395	0.057	0.08	NA
IBS					
GL	NA	0	0	0	0
IS	0.071	NA	0	0	0
NO	0.129	0.135	NA	0	0.01
DK	0.148	0.153	0.040	NA	0.01
EE	0.132	0.138	0.024	0.034	NA

Historic

F_{st}	GL	IS	NO	DK	TU
GL	NA				
IS	0.190	NA			
NO	0.296	0.190	NA		
DK	0.334	0.192	0.039	NA	
TU	0.397	0.397	0.046	0.020	NA
IBS					
GL	NA	0.02	0	0.0	NA
IS	0.059	NA	0.04	0.04	NA
NO	0.121	0.123	NA	0.03	NA
DK	0.114	0.114	0.014	NA	NA
TU	0.165	0.185	0.033	0.019	NA

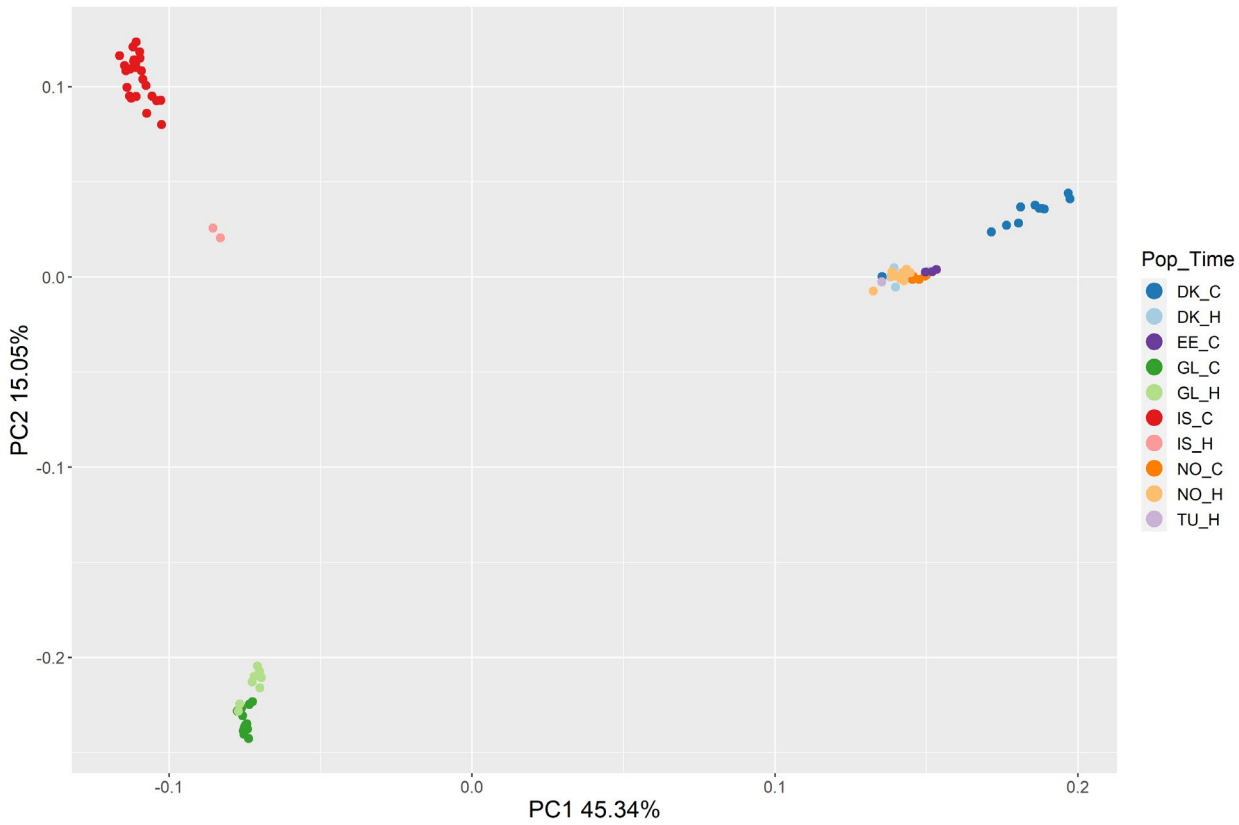


Figure 5. Clustering of white-tailed eagle individuals from contemporary and historic samples based on principal component analysis (PCA) of the genomic variation. The calculation was based on 616,182 SNPs using EIGENSOFT. Percentages given in the axis labels refer to the amount of variation explained by the respective axis. Colours indicate country and temporal category (Pop_Time). C: contemporary samples, H: historic samples. DK=Denmark, EE=Estonia, GL=Greenland, IS=Iceland, NO=Norway, TU=Turkey.

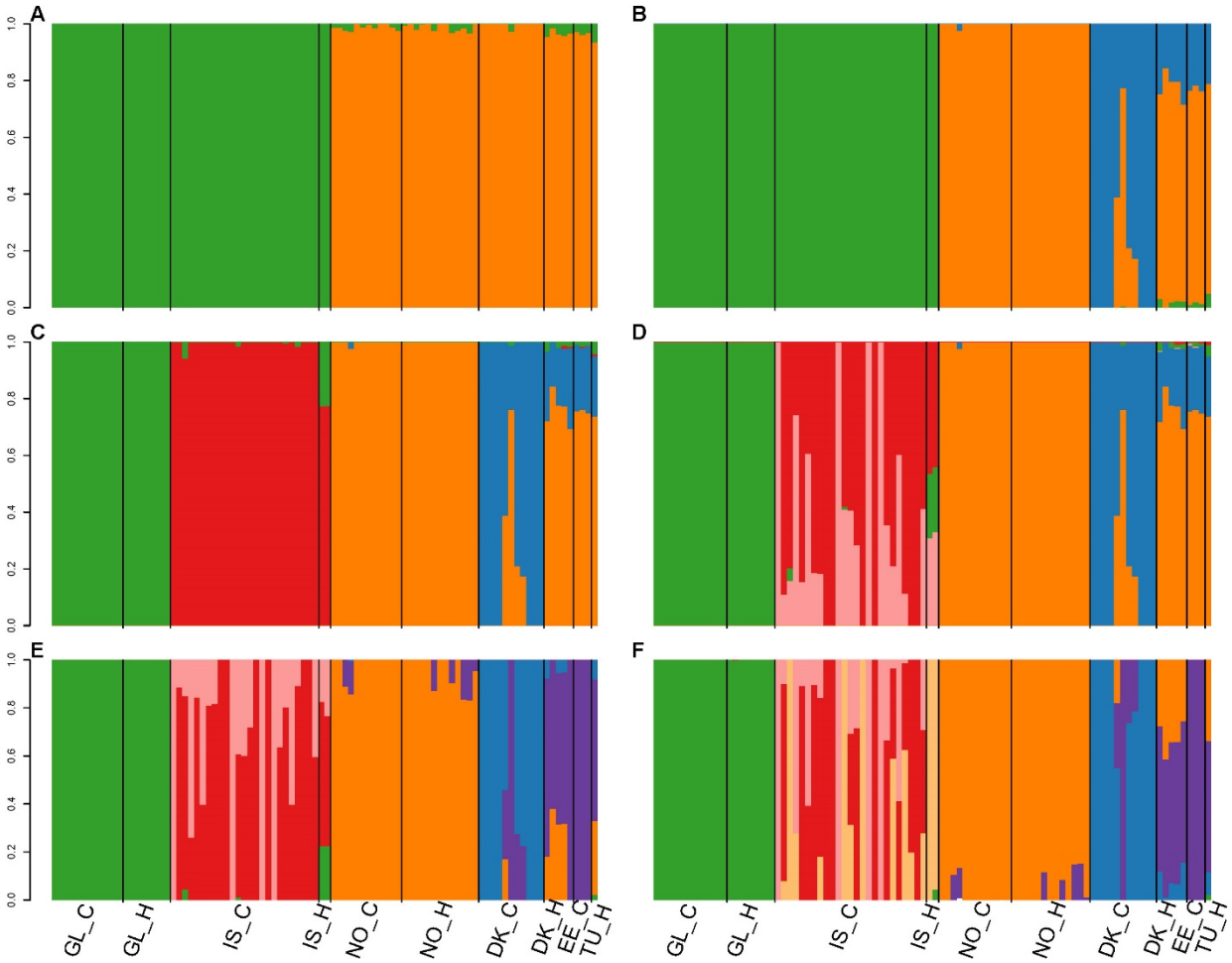


Figure 6. Admixture plots were obtained with ADMIXTURE for a different number of clusters (K). A) K=2, B) K=3, C) K=4, D) K=5, E) K=6, F) K=7. The value on the y-axis shows the assignment probability of each individual to the genetic cluster (K_i). The first two letters in the labels refer to the countries GL = Greenland, IS = Iceland, NO = Norway, DK = Denmark, EE = Estonia, TU = Turkey, and the last to the temporal sample: C: contemporary and H: historic. Admixture analysis K=2-15 converged, from K=7 and above only further substructures within populations are found (Figure S8-S15).

Table 3. Results from the Tajima D test. Mean Tajima's D (TajD) and sd in parentheses, and number of SNPs used in Tajima's D analysis (N_SNPs). Skewness statistics (gl1) of Tajima's D values along the genome (Skewness). Proportion of sites along the genome being less than -2, above 0, or above 2.

Country	Temporal sample	N_SNPs	TajD mean (sd)	Skewness (g1)	P(< 2 , > 0, or > 2)
Greenland	C	360580	0.99 (1.32)	-0.55	0.06, 0.79, 0.29
	H	304840	0.96 (1.02)	-0.43	0.06, 0.81,0.21
Iceland	C	363636	1.49 (1.42)	-0.61	0.07,0.85,0.45
	H	53493	1.53 (0.69)	-1.77	0.37,0.96,0.56
Norway	C	635004	0.87 (0.60)	-0.29	0.01,0.91,0.03
	H	577071	1.1 (0.56)	-0.28	0.01,0.96,0.05
Denmark	C	605382	0.88 (0.65)	-0.29	0.01,0.91,0.04
	H	422581	1.02 (0.45)	-0.44	0.01,0.91,0.04
Estonia	C	461948	0.26 (0.60)	0.04	0.01,0.69,0.01

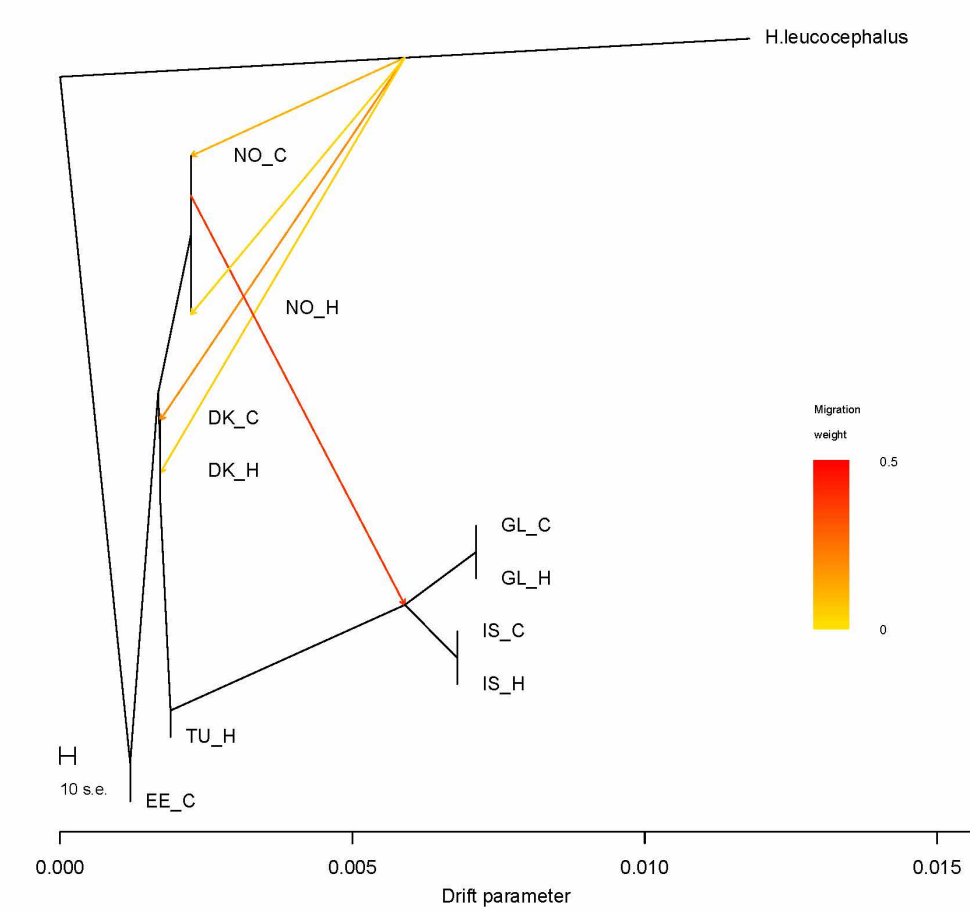


Figure 7. Population structure as well as migration patterns between populations, as characterized by Treemix. The analysis was calculated for zero to five migration events (m), the strongest support was found for m = 5 (Figure S18) as it explained the most variation.

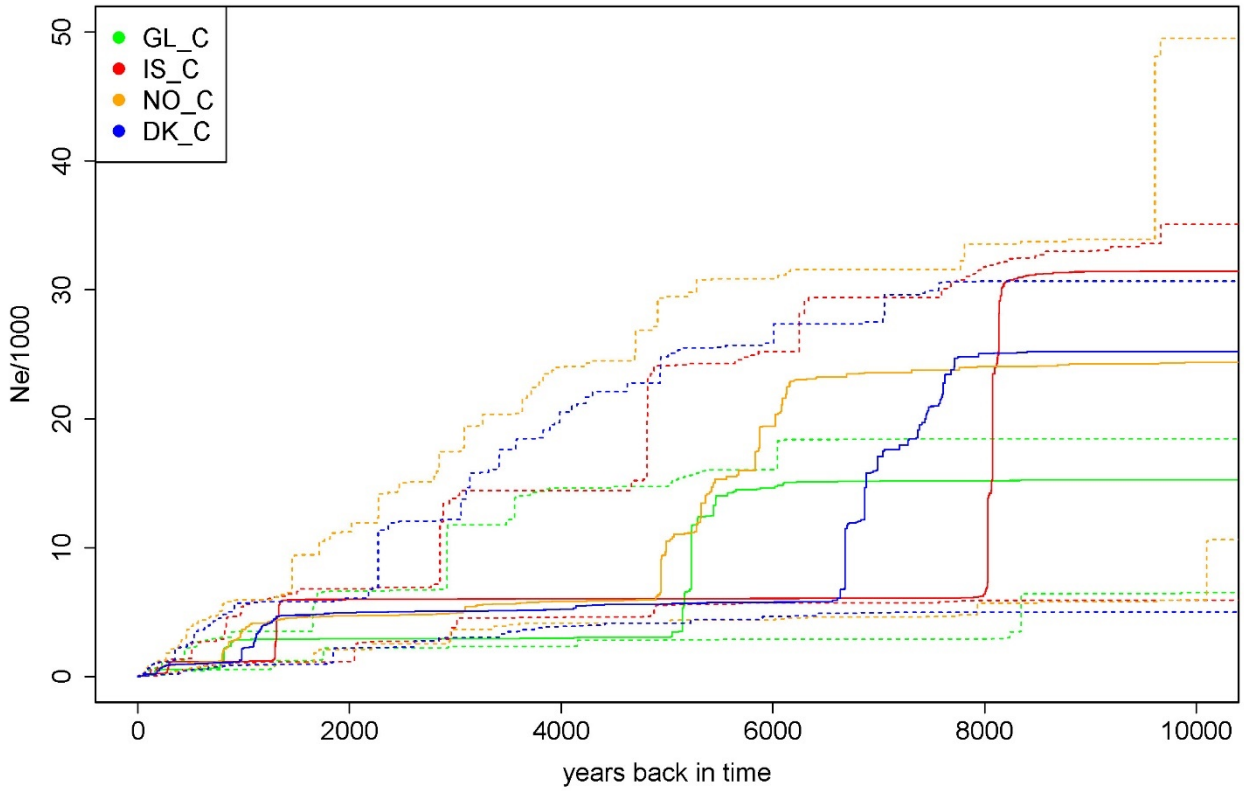


Figure 8. Stairway plot v2 for the contemporary samples from Greenland (GL_C, green), Iceland (IS_C, red), Norway (NO_C, yellow), and Denmark (DK_C, blue). The x-axis display years back in time, y-axis displays the effective population size divided by 1000. Full lines are mean, dashed lines are 2.5-97.5% confident interval (CI). Figure S19 displays until 100,000 years back in time. A mutation rate from collared flycatcher, 2.3e-9 per site per year, was used to scale the Ne as it is the only known mutation rate for birds (Smeds et al. 2016). *A generation time of 15.6 years was used as reported by IUCN red list* (Birdlife International 2020).

Supplementary

Table S1. Sample information. Table displaying sample information for the 92 individuals included in the project. Time indicates if the specimen is contemporary (Con) or historic (His). Name is specimen name. Dendro name is the individual name in the dendrogram. CO is country code: GL=Greenland, IS=Iceland, NO=Norway, DK=Denmark, EE=Estonia, TU=Turkey. Location (Loc) refers to the location in the country (some are unknown), abbreviation: Faxa=Faxafloi, Nbreida=North Breidafjordur, Sbreida=South breidafjordur, Vestf=Westfjords, Huna=Húnaflói. Year is the year of sampling. Sex indicates the sex of the individual. Maturity indicates the age of the individual when sampled, immature=< indicated that the individual is older than a nestling but could be less than a full adult. Sample type indicates the type of tissue taken. Sequencing (Seq) indicates where it was sequenced: BGI=BGI Genomics, NTNU=Norwegian University of Science, deCODE=deCODE genetics in Iceland. Mean depth all is mean depth per individual over all 780,604 SNPs used in the data set including missing sites. Mean depth is mean depth per individual only for sites present in the individual. Frac. missing SNPs is the fraction of missing SNPs per individual.

Time	name	Dendro name	CO	Loc	Year	Sex	Maturity	Sample type	Seq	Mean depth all	Mean depth	Frac. missing SNPs
Con	A6512	IS_C_1	IS	Faxaflói	2003	f	nestling	Blood EDTA	BGI	17.05	17.15	0.013
Con	A6517	IS_C_2	IS	Faxaflói	2003	f	nestling	Blood EDTA	BGI	17.55	17.63	0.011
Con	A6533	IS_C_3	IS	N-Breidafjordur	2003	m	nestling	Blood EDTA	BGI	16.14	16.28	0.020
Con	A6541	IS_C_4	IS	S-Breidafjordur	2003	f	nestling	Blood EDTA	BGI	14.43	14.73	0.044
Con	A6551	IS_C_5	IS	Vestfirdir	2003	m	nestling	Blood EDTA	BGI	16.85	16.96	0.015
Con	A6552	IS_C_6	IS	N-Breidafjordur	2003	f	nestling	Blood EDTA	BGI	14.27	14.58	0.046
Con	A7002	IS_C_7	IS	S-Breidafjordur	2004	-	nestling	Blood EDTA	BGI	17.24	17.33	0.012
Con	A7028	IS_C_8	IS	Húnafló i	2004	m	nestling	Blood EDTA	BGI	17.39	17.47	0.011
Con	A7029	IS_C_9	IS	S-Beidafjordur	2004	f	nestling	Blood EDTA	BGI	16.37	16.49	0.017
Con	A7043.2	IS_C_10	IS	S-Breidafjordur	2005	m	nestling	Blood EDTA	BGI	17.02	17.12	0.014
Con	A7053	IS_C_11	IS	N-Breidafjordur	2005	f	nestling	Blood EDTA	BGI	16.01	16.15	0.021
Con	A7057	IS_C_12	IS	N-Breidafjordur	2005	-	nestling	Blood EDTA	BGI	14.96	15.21	0.036
Con	A7067	IS_C_13	IS	S-Breidafjordur	2006	m	nestling	Blood EDTA	BGI	17.89	17.95	0.010

Con	A7073	IS_C_14	IS	S-Breidafjordur	2006	-	nestling	Blood EDTA	BGI	15.65	15.83	0.026
Con	A7074	IS_C_15	IS	N-Breidafjordur	2007	-	nestling	Blood EDTA	BGI	17.63	17.71	0.010
Con	A7093	IS_C_16	IS	S-breidafjordur	2007	-	nestling	Blood EDTA	BGI	16.53	16.64	0.016
Con	A7094	IS_C_17	IS	S-Breidafjordur	2007	-	nestling	Blood EDTA	BGI	17.27	17.34	0.012
Con	A7103	IS_C_18	IS	Faxaflói	2004	f	nestling	Blood EDTA	BGI	14.60	14.99	0.051
Con	A7116	IS_C_19	IS	Faxaflói	2006	-	nestling	Blood EDTA	BGI	17.29	17.37	0.012
Con	A7134	IS_C_20	IS	S-Breidafjordur	2008	-	nestling	Blood EDTA	BGI	17.11	17.21	0.014
Con	A7138	IS_C_21	IS	S-Breidafjordur	2008	f	nestling	Blood EDTA	BGI	17.29	17.37	0.012
Con	A7159	IS_C_22	IS	Húnafló i	2008	m	nestling	Blood EDTA	BGI	17.38	17.47	0.013
Con	A7161	IS_C_23	IS	Vestfirdir	2009	f	nestling	Blood EDTA	BGI	18.41	18.46	0.008
Con	A7168	IS_C_24	IS	N-Breidafjordur	2009	m	nestling	Blood EDTA	BGI	17.77	17.85	0.011
Con	A7206	IS_C_25	IS	S-Breidafjordur	2009	m	nestling	Blood EDTA	BGI	13.79	14.20	0.061
Con	AA10	NO_C_1	NO	Ertvågøy	2007	f	post nestling	Toepad	NMTU	14.60	14.94	0.049
Con	BB32	NO_C_4	NO	Bodå	2008	-	post nestling	Toepad	NMTU	38.52	12.51	0.010
Con	BB34	NO_C_5	NO	Håmantjønna	2009	m	post nestling	Toepad	NMTU	11.15	41.29	0.211
Con	BB35	NO_C_6	NO	Tettila	2010	f	post nestling	Toepad	NMTU	10.17	38.66	0.274
Con	BB36	NO_C_7	NO	Frøvarp, Elvalandet	2011	m	post nestling	Toepad	NMTU	9.10	12.46	0.370
Con	BB37	NO_C_8	NO	Vikdalen	2011	-	post nestling	Toepad	NMTU	17.44	11.83	0.036
Con	BB38	NO_C_9	NO	Verma	2013	f	post nestling	Toepad	NMTU	4.42	11.25	0.847
Con	AA11	NO_C_2	NO	Elgskaret	2013	f	post nestling	Toepad	NMTU	11.66	17.75	0.144
Con	AA13	NO_C_3	NO	Snåsamoen	2014	m	post nestling	Toepad	NMTU	41.27	10.69	0.001

Con	BB40	NO_C_10	NO	Olsvik	2012	f	post nestling	Toepad	NMTU	11.48	13.70	0.275
Con	BB41	NO_C_11	NO	Risvær	2012	m	post nestling	Toepad	NMTU	8.75	12.22	0.477
Con	CC19	NO_C_12	NO	Hesthåggån	2009	f	post nestling	Toepad	NMTU	17.70	18.00	0.036
Con	DK1	DK_C_5	DK	Præstø, Zealand	2016	-	adult	Skin/muscle	BGI	16.08	18.13	0.023
Con	DK2	DK_C_6	DK	Skudelev, Sealand	2016	-	adult	Skin/muscle	BGI	18.23	15.73	0.010
Con	DK3	DK_C_7	DK	Sorø, Zealand	2018	-	adult	Skin/muscle	BGI	13.00	17.34	0.105
Con	DK4	DK_C_8	DK	Klejs, Zealand	2018	-	adult	Skin/muscle	BGI	15.39	14.59	0.058
Con	DK6	DK_C_9	DK	-	2016	-	adult	Skin/muscle	BGI	16.78	16.23	0.028
Con	DK8	DK_C_10	DK	Hadersl ev, Jutland	2015	f	nestling	Full blood	BGI	14.22	14.07	0.050
Con	DK12	DK_C_11	DK	Filsø, Jutland	2015	m	nestling	Full blood	BGI	18.07	18.29	0.010
Con	DK13	DK_C_2	DK	Filsø, Jutland	2015	m	nestling	Full blood	BGI	15.54	13.74	0.028
Con	DK16	DK_C_3	DK	Hyllekr og, Lolland	2015	m	nestling	Full blood	BGI	17.24	15.90	0.014
Con	DK17	DK_C_4	DK	Hyllekr og, Lolland	2015	f	nestling	Full blood	BGI	14.28	17.01	0.048
Con	DK21	DK_C_6	DK	Kastrup , CPH	2015	-	adult	Skin/muscle	BGI	13.27	14.56	0.107
Con	E1	EE_C_1	ES	Spitha mi, Lääne	2015	f	nestling	Full blood	BGI	14.45	11.24	0.047
Con	E3	EE_C_2	ES	Harju, Lääne	2015	m	nestling	Full blood	BGI	16.95	9.46	0.016
Con	E6	EE_C_3	ES	Kiili, Lääne	2015	m	nestling	Full blood	BGI	17.02	9.56	0.015
Con	GL-1	GL_C_4	GL	-	>1990	-	-	Full blood	BGI	17.49	26.42	0.013
Con	GL-2	GL_C_8	GL	Sisimiut	1990	f	immature	Full blood	BGI	17.94	10.67	0.012
Con	GL-3	GL_C_9	GL	Nuuk	1993	f	adult	Full blood	BGI	14.88	14.76	0.037
Con	GL-4	GL_C_10	GL	Amerali k	1992	f	adult	Full blood	BGI	17.46	17.06	0.014
Con	GL-7	GL_C_11	GL	Nanorta lik	1996	f	immature	Full blood	BGI	16.46	17.12	0.019
Con	GL-9	GL_C_12	GL	Nuuk	1996	-	-	Full blood	BGI	16.58	17.88	0.018
Con	GL-10	GL_C_1	GL	Nuuk	2011	-	-	Full blood	BGI	17.71	14.99	0.019
Con	GL-15	GL_C_2	GL	-	>2013	-	adult	Full blood	BGI	14.70	17.08	0.040

Con	GL-16	GL_C_3	GL	Nanortalik	2017	-	adult	Full blood	BGI	16.98	17.58	0.013
Con	GL-22	GL_C_5	GL	-	1999	-	-	Skin/muscle	BGI	14.09	14.71	0.080
Con	GL-26	GL_C_6	GL	Paamiut	2013	-	-	Skin/muscle	BGI	14.30	15.64	0.164
Con	GL-27	GL_C_7	GL	Maniitsoq	2013	-	-	Skin/muscle	BGI	16.47	16.98	0.052
His	DK.H-2	DK_H_3	DK	-	1907	-	immature=<	Toepad	deCODE	5.96	18.01	0.784
His	DK.H-3	DK_H_4	DK	-	1900	-	immature=<	Toepad	deCODE	26.40	15.14	0.004
His	DK.H-7	DK_H_5	DK	Hjelm	1898	m	immature=<	Toepad	deCODE	8.69	17.56	0.409
His	DK.H-12	DK_H_1	DK	Oxhold, Jylland	1897	f	immature=<	Toepad	deCODE	9.80	16.59	0.297
His	DK.H-14	DK_H_2	DK	-	1909	-	immature=<	Toepad	deCODE	4.10	16.71	0.938
His	GL.H-1	GL_H_3	GL	-	1919	-	immature=<	Toepad	deCODE	6.26	10.38	0.717
His	GL.H-2	GL_H_5	GL	-	1885	-	immature=<	Toepad	deCODE	10.59	9.57	0.238
His	GL.H-3	GL_H_6	GL	Nuuk	1898	-	immature=<	Toepad	deCODE	8.93	9.84	0.398
His	GL.H-4	GL_H_7	GL	-	1920	m	immature=<	Toepad	deCODE	35.34	10.58	0.002
His	GL.H-8	GL_H_8	GL	Kangâmiut	1919	f	immature=<	Toepad	deCODE	28.73	11.93	0.003
His	GL.H-14	GL_H_1	GL	-	1906	-	immature=<	Toepad	deCODE	8.08	11.23	0.480
His	GL.H-15	GL_H_2	GL	-	1892	-	immature=<	Toepad	deCODE	4.40	35.35	0.900
His	GL.H-20	GL_H_4	GL	-	1919	-	immature=<	Toepad	deCODE	8.07	28.75	0.481
His	IS.H-4	IS_H_1	IS	Hvalfjördur, SW-Iceland	1950	-	immature	Toepad	deCODE	5.53	9.78	0.781
His	IS.H-6	IS_H_2	IS	N-Iceland	<1946	-	immature	Toepad	deCODE	5.71	9.56	0.785
His	K9	NO_H_8	NO	Åstfjorden	1916	f	immature=<	Toepad	Mike / NMTU	4.87	11.30	0.873
His	K7	NO_H_6	NO	Åstfjorden	1916	f	immature=<	Toepad	Mike / NMTU	14.45	9.89	0.052
His	K8	NO_H_7	NO	Åstfjorden	1916	f	immature=<	Toepad	Mike / NMTU	5.27	10.98	0.840
His	K5	NO_H_4	NO	Hitra	1904	m	immature=<	Toepad	Mike / NMTU	9.78	11.31	0.291
His	K6	NO_H_5	NO	Hitra	1904	m	immature=<	Toepad	Mike / NMTU	7.00	10.05	0.625

His	K1	NO_H_1	NO	Rødøy	1925	f	immature=<	Toepad	Mike / NMTU	9.70	14.78	0.307
His	K2	NO_H_2	NO	Rødøy	1925	f	immature=<	Toepad	Mike / NMTU	5.76	9.54	0.789
His	K4	NO_H_3	NO	Hitra	1902	m	immature=<	Toepad	Mike / NMTU	8.08	9.78	0.508
His	NO.H-2	NO_H_9	NO	Kristian sund	1922	f	immature=<	Toepad	deCODE	11.60	12.46	0.165
His	NO.H-4	NO_H_10	NO	Kristian sund	1922	f	immature=<	Toepad	deCODE	13.45	13.93	0.082
His	NO.H-5	NO_H_11	NO	Kristian sund	1922	m	immature=<	Toepad	deCODE	9.40	10.98	0.326
His	NO.H-6	NO_H_12	NO	Sundfjord	1937	m	immature=<	Toepad	deCODE	4.73	9.37	0.893
His	NO.H-7	NO_H_13	NO	Sognefjord	1936	m	immature=<	Toepad	deCODE	3.92	9.87	0.945
His	TU.H-1	TU_H_1	TU	Istanbul	1934	f	adult	Toepad	deCODE	14.02	14.43	0.072

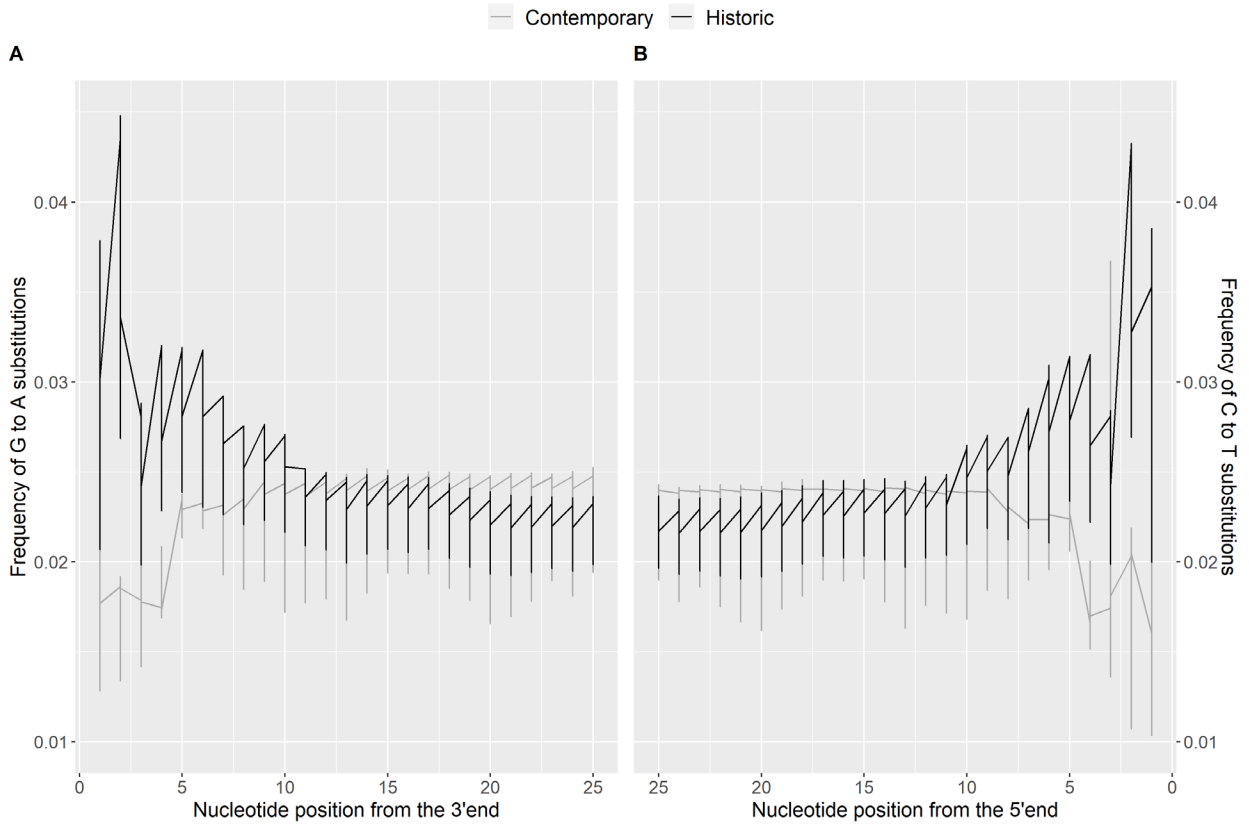


Figure S1. Frequency of mean nucleotide misincorporation over all individuals, for contemporary and historic samples from the 3'end (A) and 5'end (B).

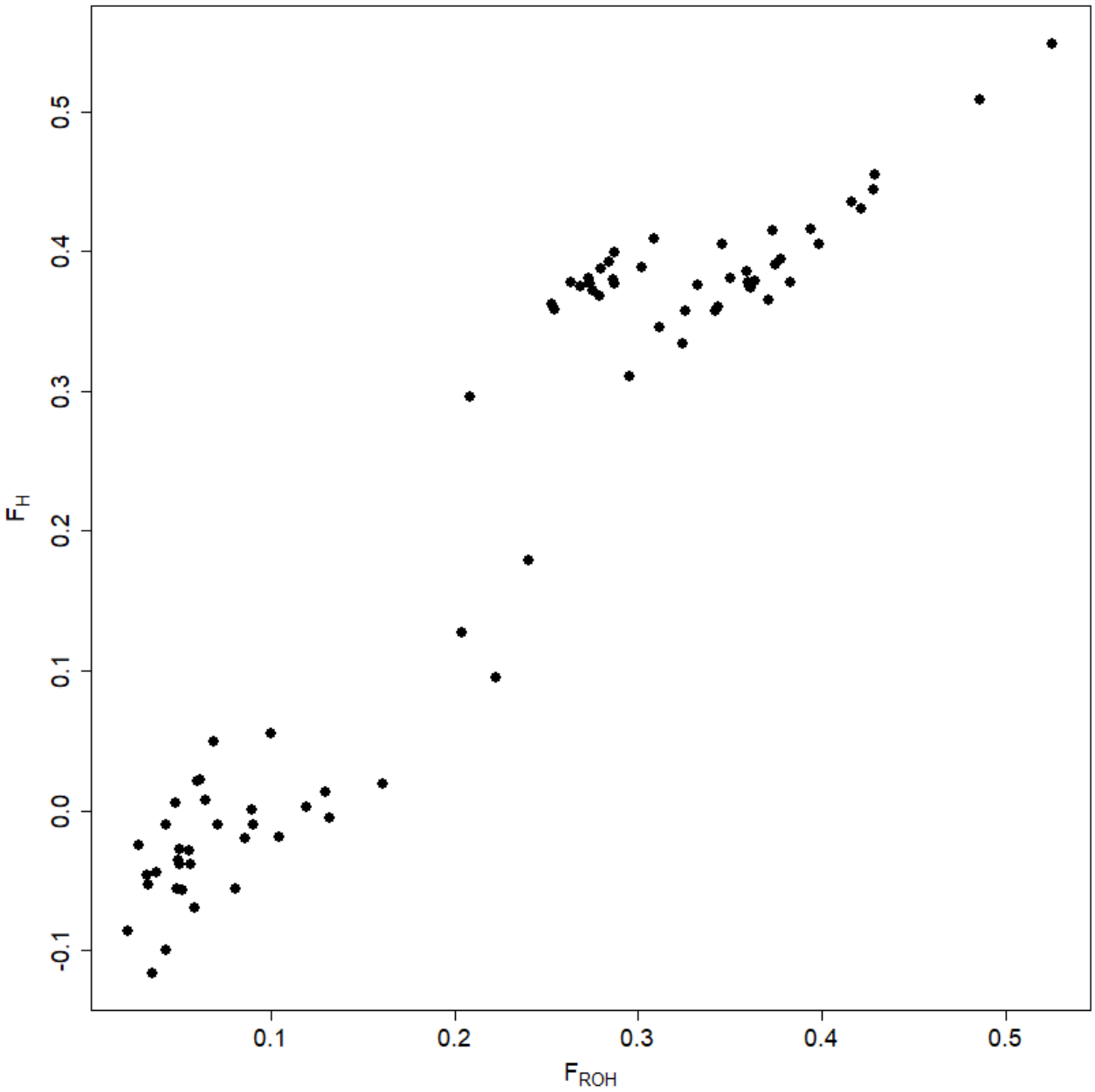


Figure S2. Relationship between inbreeding coefficients F_H and F_{ROH}. Individuals with missing data <50% are included. Pearsons correlation coefficient (r) = 0.95.

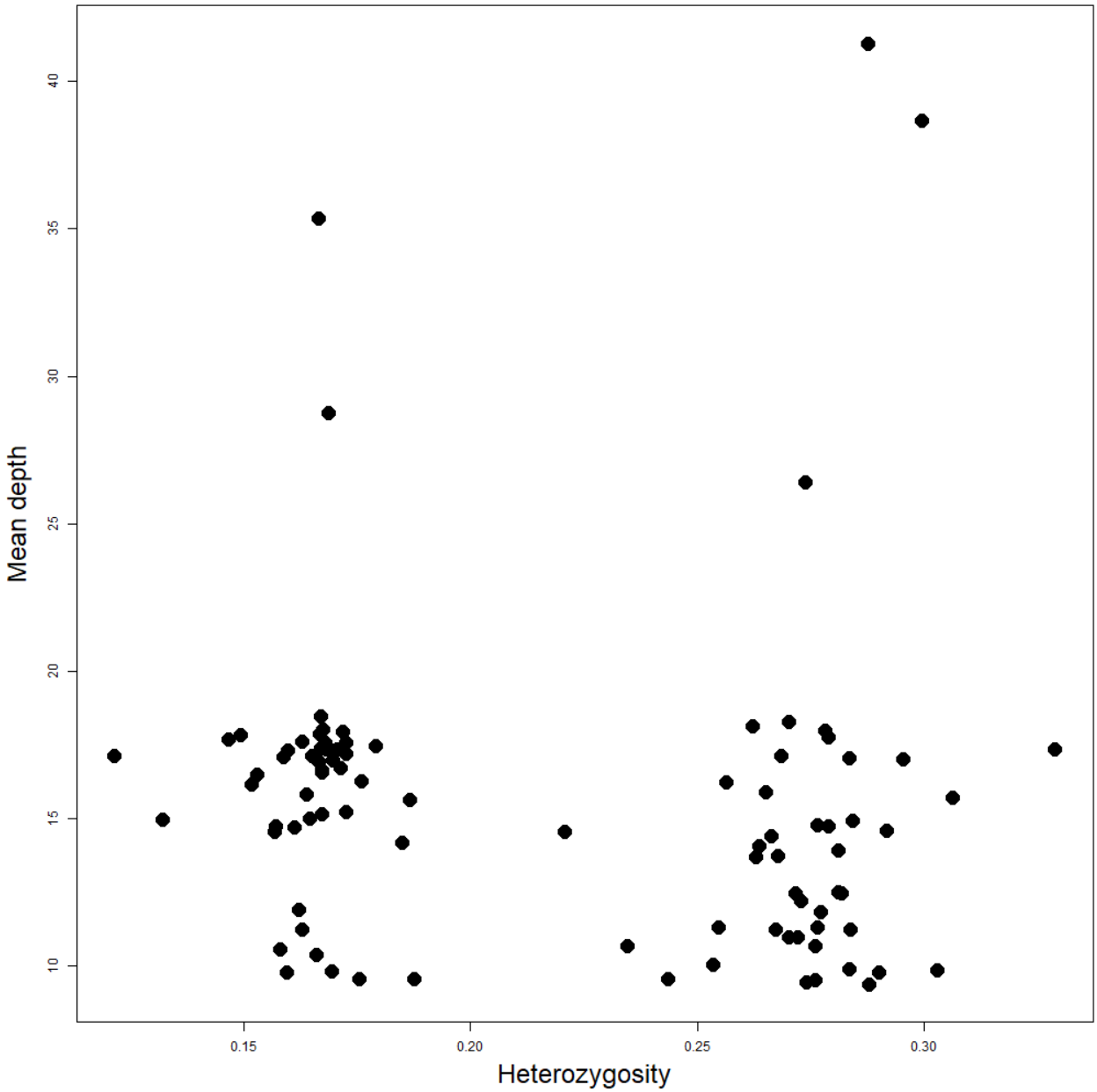


Figure S3. Relationship between mean sequencing depth with no missing sites and observed heterozygosity for all 92 samples. No correlation is found (-.0.08).

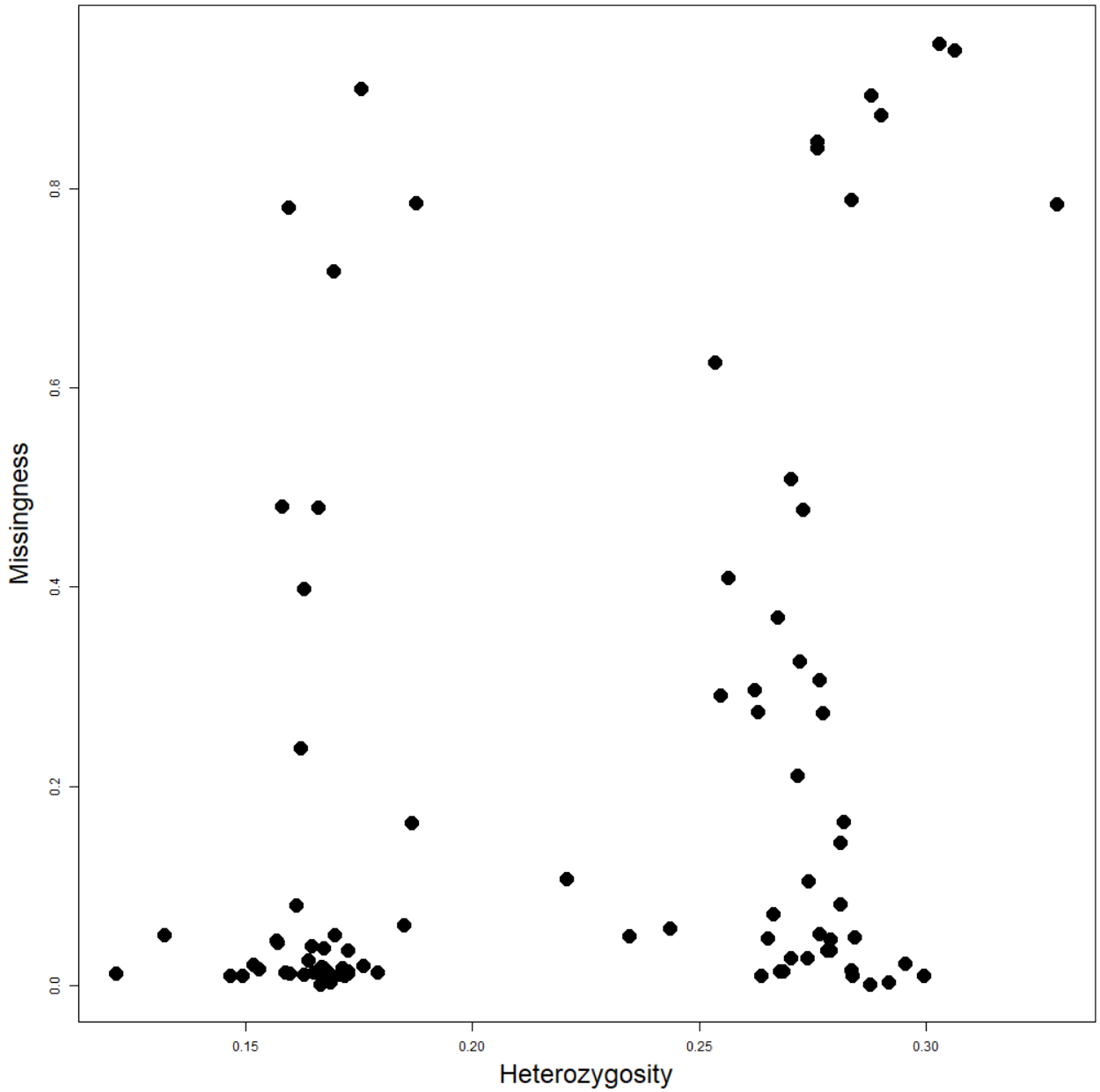


Figure S4. Relationship between proportion of missing sites and observed heterozygosity for all 92 samples. A weak correlation is found (0.33).

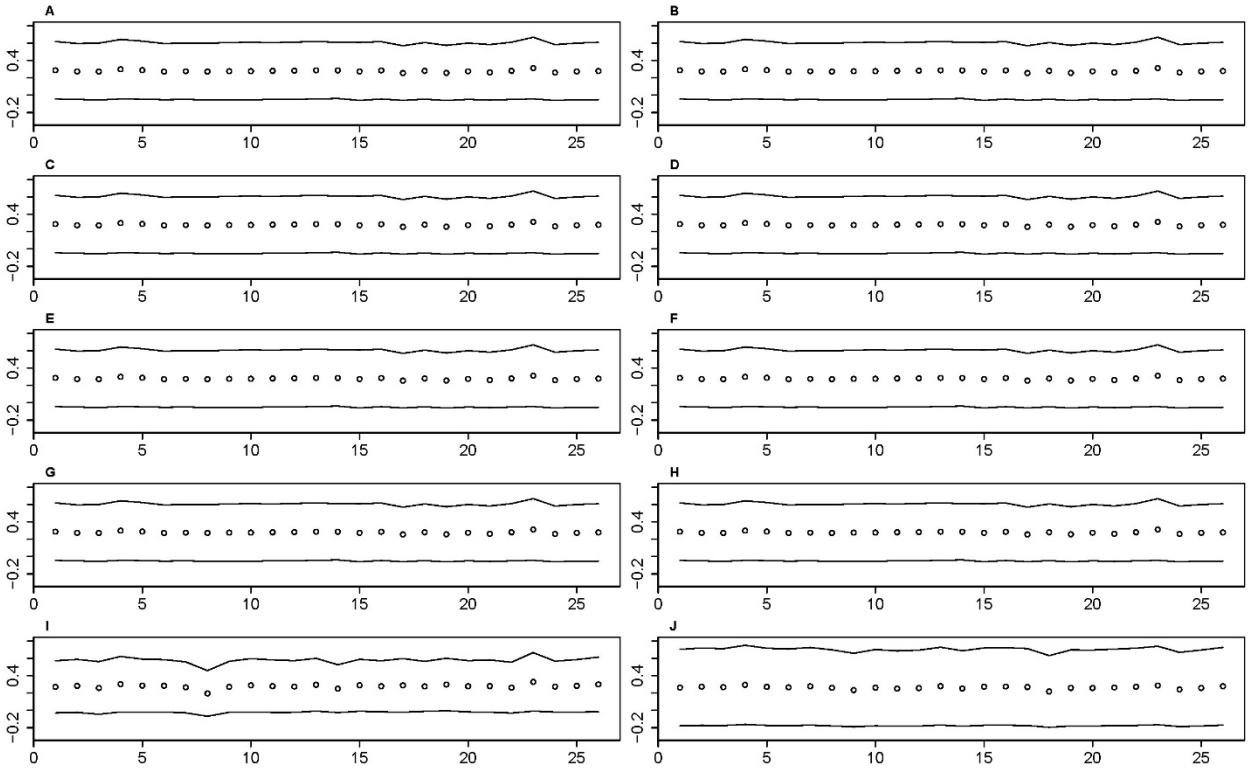


Figure S5. Mean observed heterozygosity per site as mean and sd per chromosome per population. Mean are dots and sd lines. A) contemporary Greenland, B) historic Greenland, C) contemporary Icelandic, D) historic Icelandic, E) contemporary Norwegian, F) historic Norwegian, G) contemporary Danish, H) historic Danish, I) contemporary Danish, J) historic Turkish.

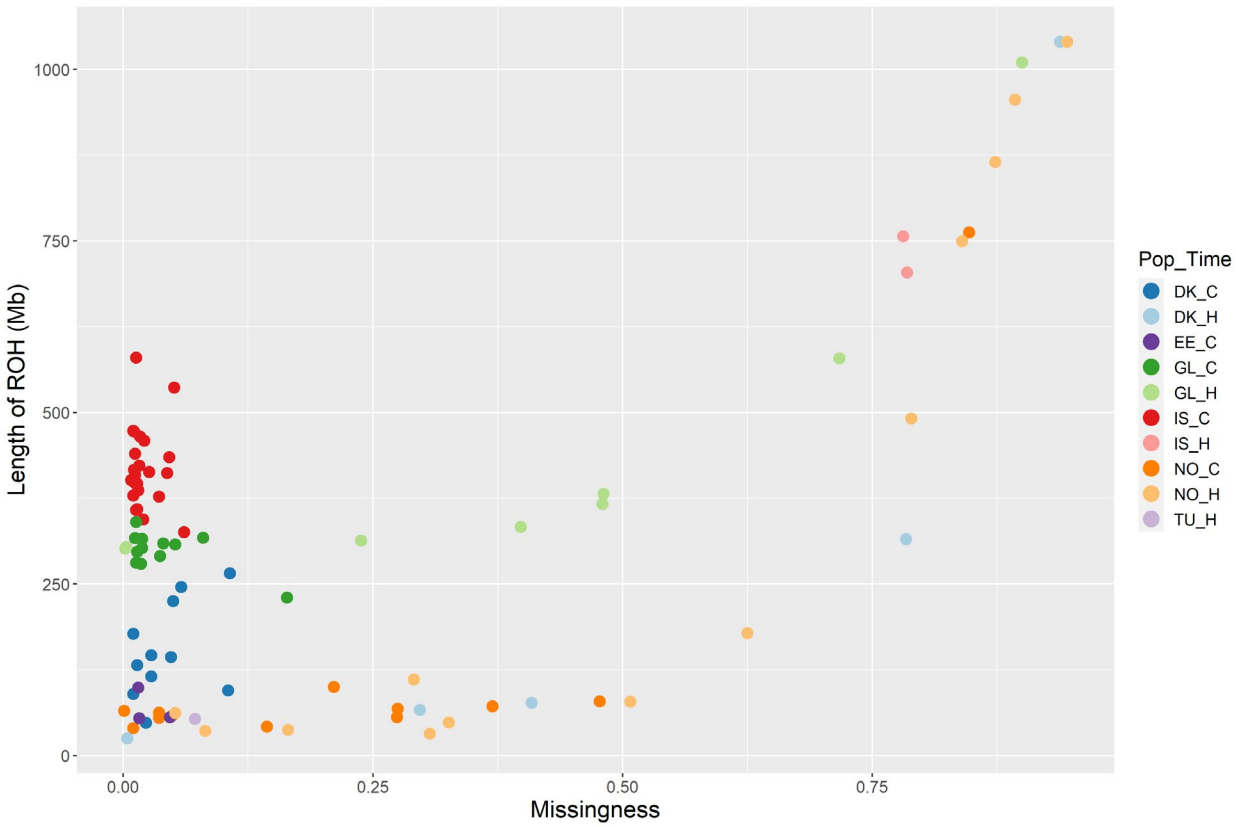


Figure S6. Runs Of Homozygosity (ROH) against missingness. The x-axis displays the length of ROH against nucleotide missingness on the y-axis per individual for the 10 different groups of white-tailed eagles, which is differently coloured. An increase in length of ROH compared to the rest of the group is seen for individuals with missingness over ~0.5. The individuals with extreme missingness above 0.5 are the same individuals displaying an extreme signal for ROH in figure 3.

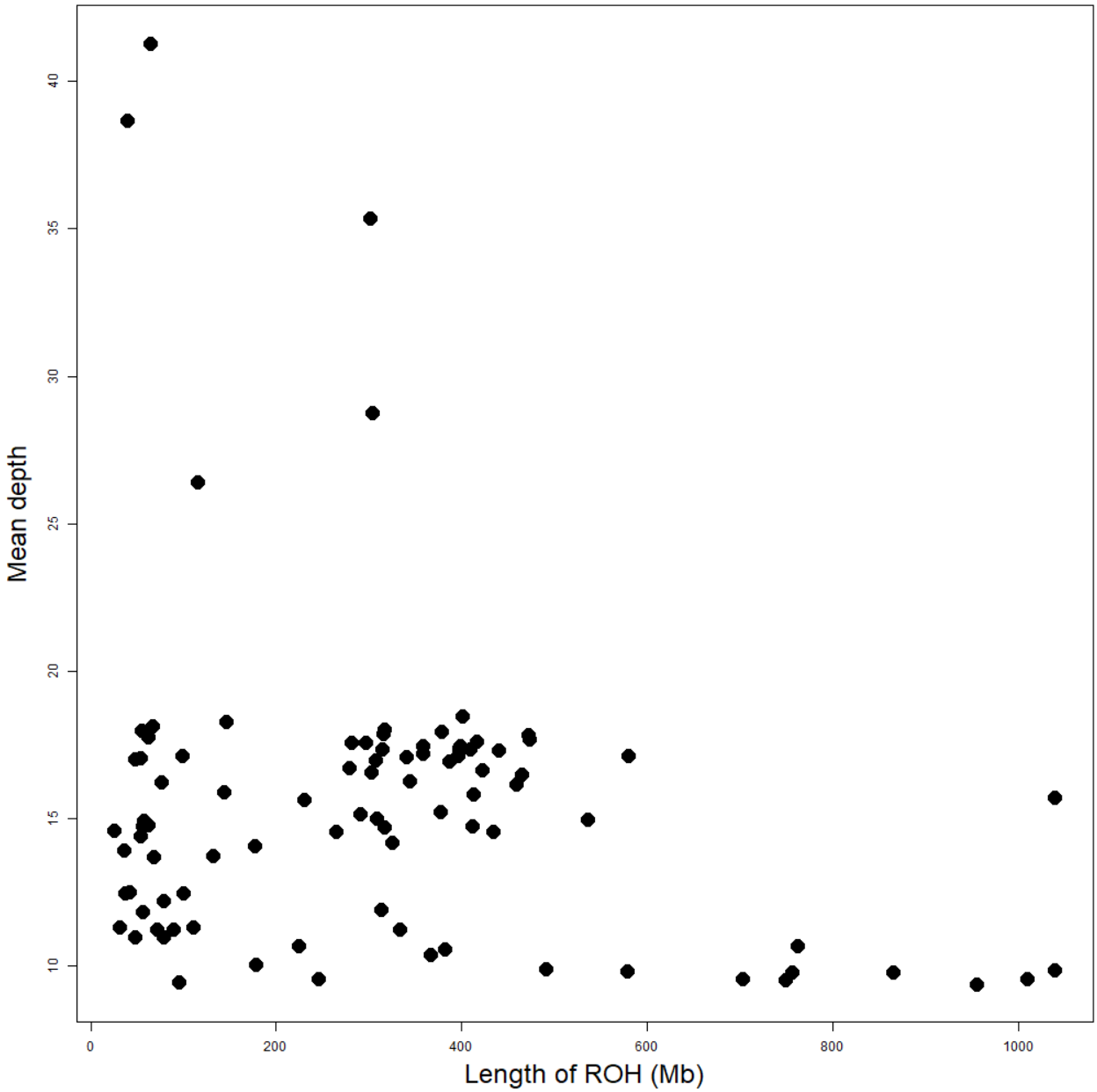


Figure S7. Mean sequencing depth without missing sites against Runs of Homozygosity (ROH). A weak correlation is found (-0.23).

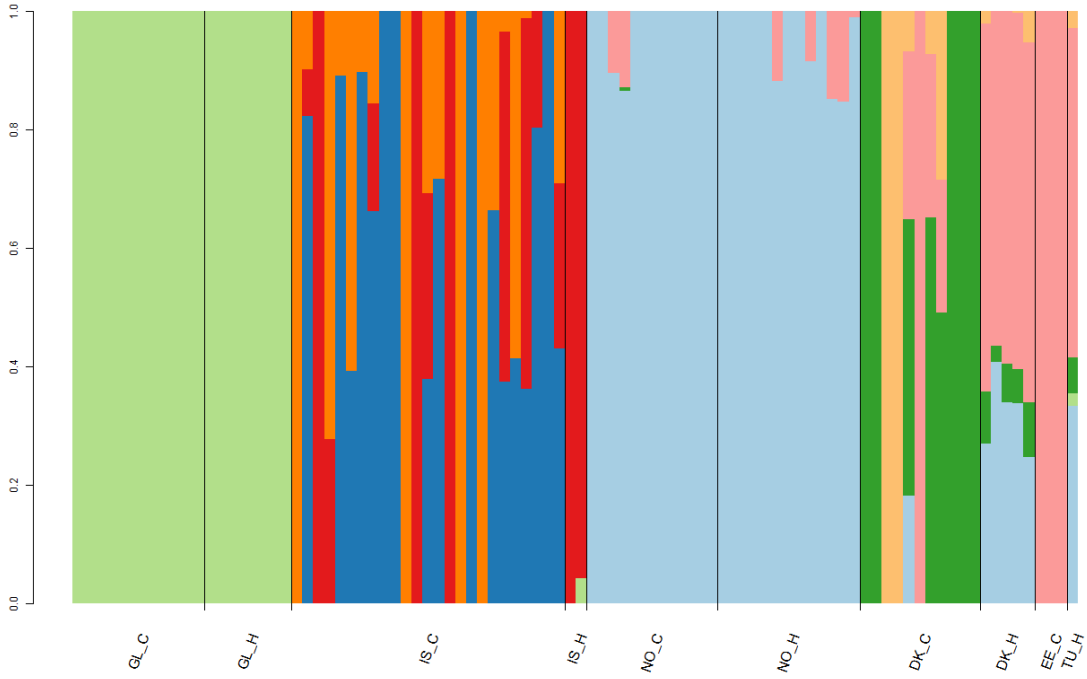


Figure S8. Admixture K8. It is seen that there is just further structure found within populations. GL_C=contemporary Greenland, GL_H=historic Greenland, IS_C=contemporary Iceland, IS_C=historic Iceland, NO_C=contemporary Norway, NO_H=historic Norway, DK_C=contemporary Denmark, DK_H=historic Denmark, EE_C=contemporary Estonia, TU_H=historic Turkey.

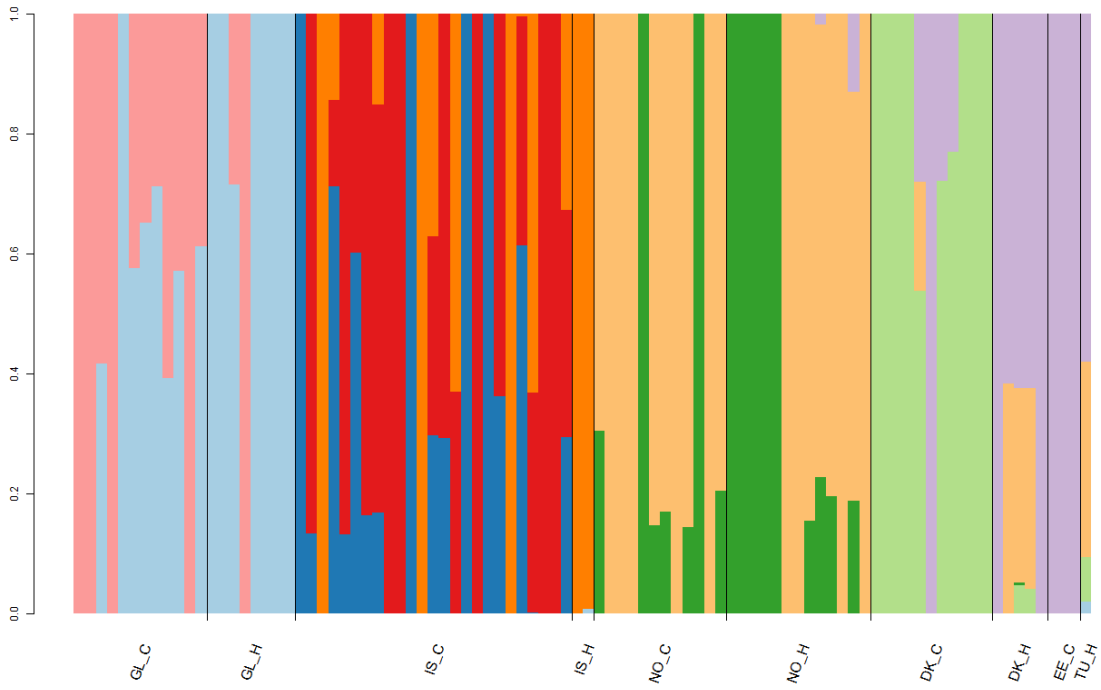


Figure S9. Admixture K9. It is seen that there is just further structure found within populations. GL_C=contemporary Greenland, GL_H=historic Greenland, IS_C=contemporary Iceland, IS_C=historic Iceland, NO_C=contemporary Norway, NO_H=historic Norway, DK_C=contemporary Denmark, DK_H=historic Denmark, EE_C=contemporary Estonia, TU_H=historic Turkey.

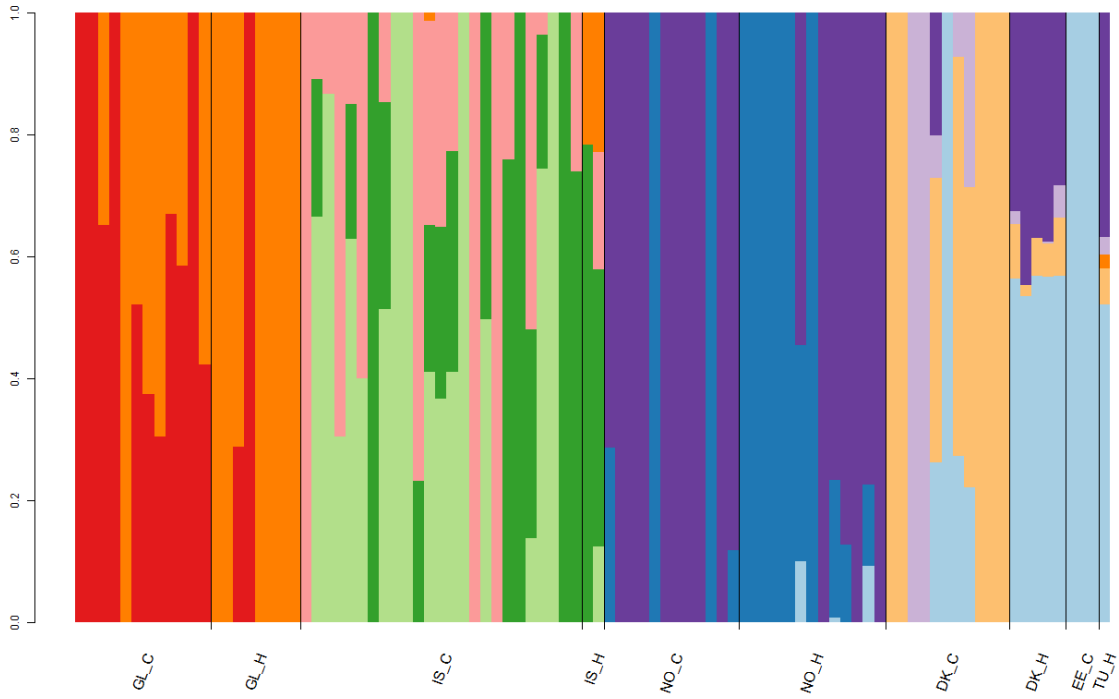


Figure S10. Admixture K10. It is seen that there is just further structure found within populations. GL_C=contemporary Greenland, GL_H=historic Greenland, IS_C=contemporary Iceland, IS_C=historic Iceland, NO_C=contemporary Norway, NO_H=historic Norway, DK_C=contemporary Denmark, DK_H=historic Denmark, EE_C=contemporary Estonia, TU_H=historic Turkey.

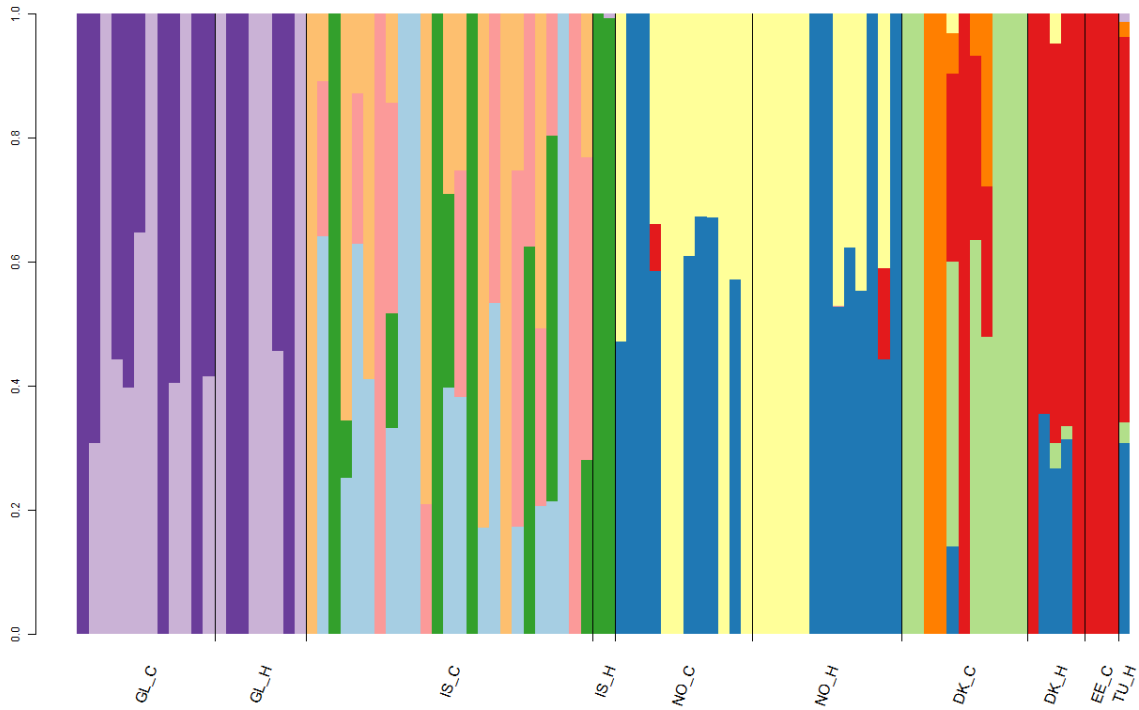


Figure S11. Admixture K11. It is seen that there is just further structure found within populations. GL_C=contemporary Greenland, GL_H=historic Greenland, IS_C=contemporary Iceland, IS_C=historic Iceland, NO_C=contemporary Norway, NO_H=historic Norway, DK_C=contemporary Denmark, DK_H=historic Denmark, EE_C=contemporary Estonia, TU_H=historic Turkey.

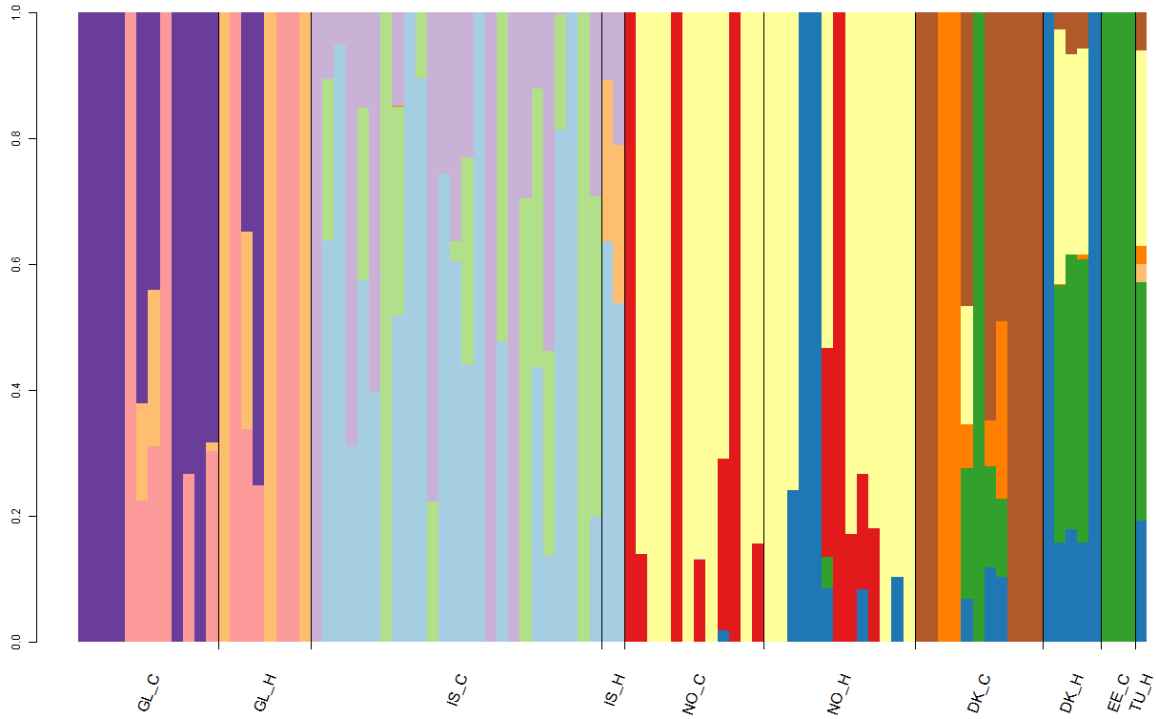


Figure S12. Admixture K12. It is seen that there is just further structure found within populations. GL_C=contemporary Greenland, GL_H=historic Greenland, IS_C=contemporary Iceland, IS_C=historic Iceland, NO_C=contemporary Norway, NO_H=historic Norway, DK_C=contemporary Denmark, DK_H=historic Denmark, EE_C=contemporary Estonia, TU_H=historic Turkey.

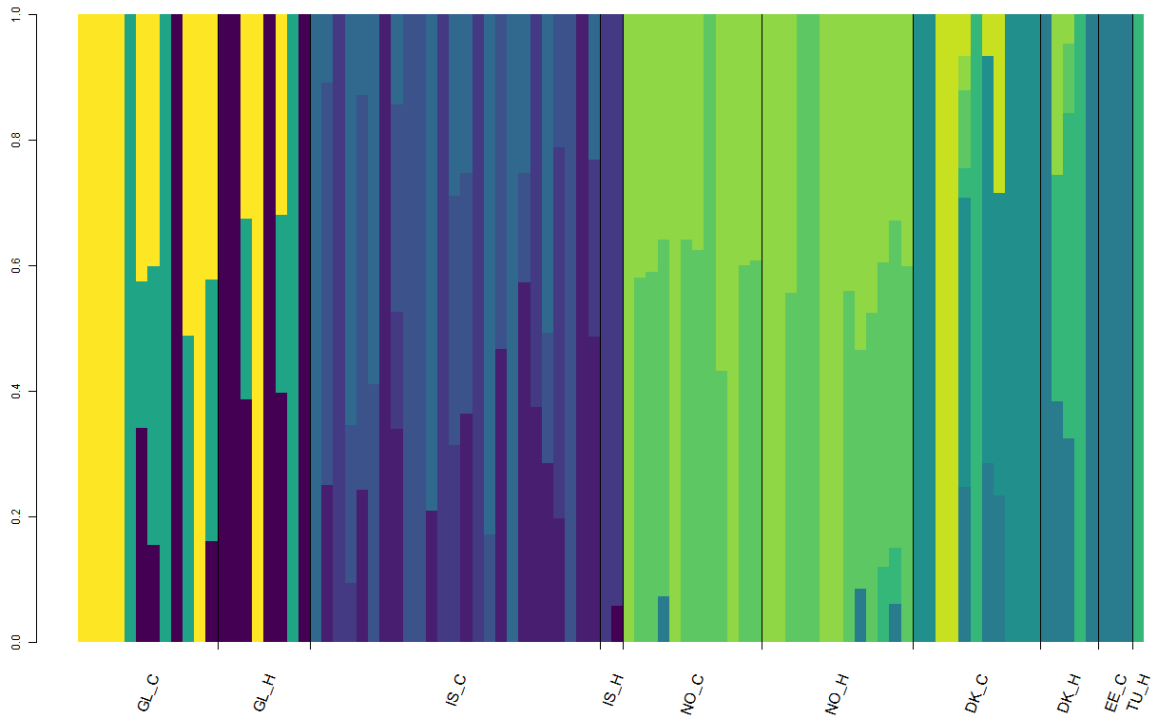


Figure S13. Admixture K13. It is seen that there is just further structure found within populations. GL_C=contemporary Greenland, GL_H=historic Greenland, IS_C=contemporary Iceland, IS_C=historic Iceland, NO_C=contemporary Norway, NO_H=historic Norway, DK_C=contemporary Denmark, DK_H=historic Denmark, EE_C=contemporary Estonia, TU_H=historic Turkey.

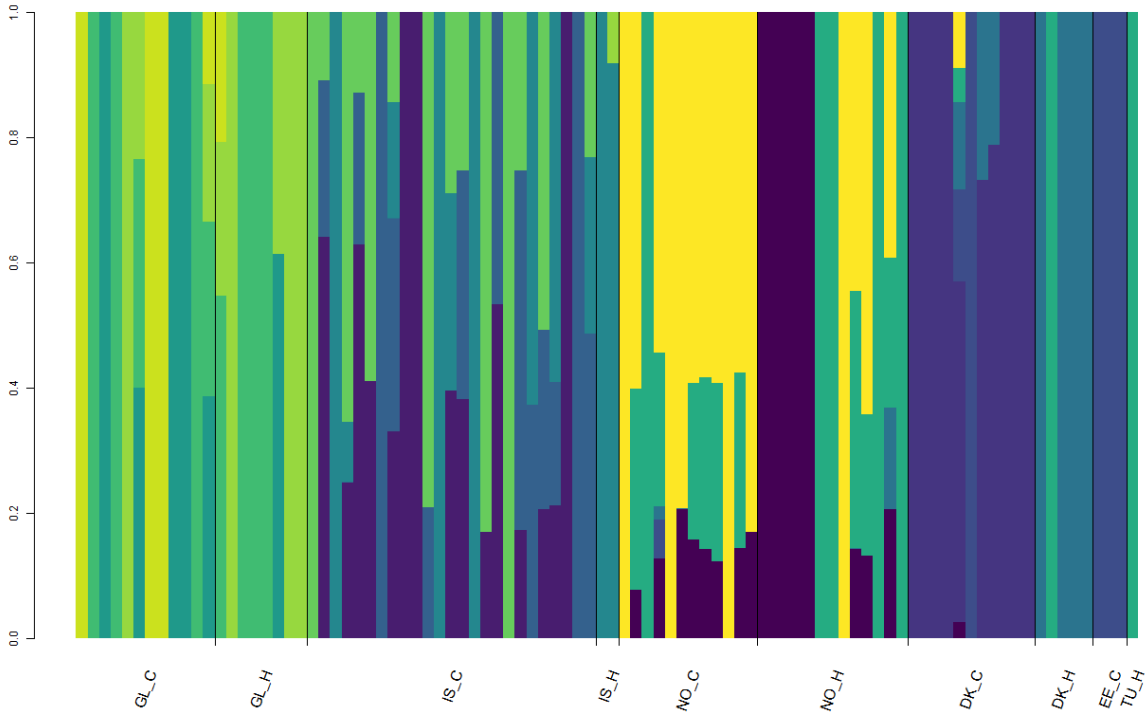


Figure S14. Admixture K14. It is seen that there is just further structure found within populations. GL_C=contemporary Greenland, GL_H=historic Greenland, IS_C=contemporary Iceland, IS_C=historic Iceland, NO_C=contemporary Norway, NO_H=historic Norway, DK_C=contemporary Denmark, DK_H=historic Denmark, EE_C=contemporary Estonia, TU_H=historic Turkey.

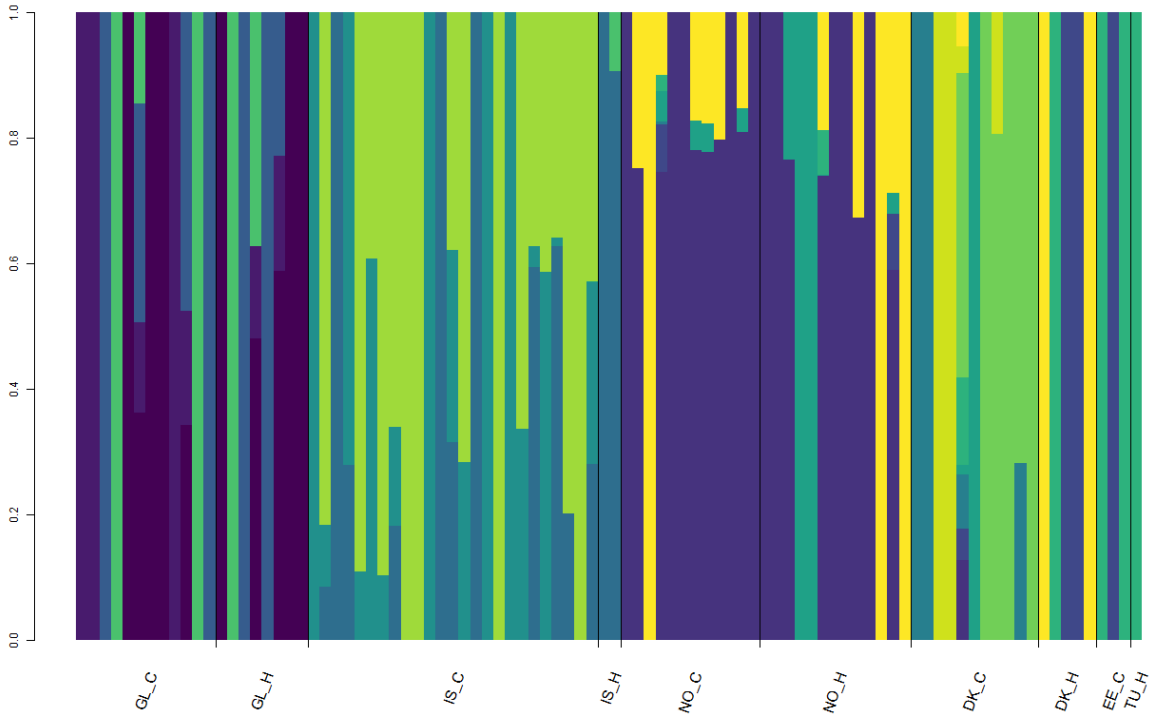


Figure S15. Admixture K15. It is seen that there is just further structure found within populations. GL_C=contemporary Greenland, GL_H=historic Greenland, IS_C=contemporary Iceland, IS_C=historic Iceland, NO_C=contemporary Norway, NO_H=historic Norway, DK_C=contemporary Denmark, DK_H=historic Denmark, EE_C=contemporary Estonia, TU_H=historic Turkey.

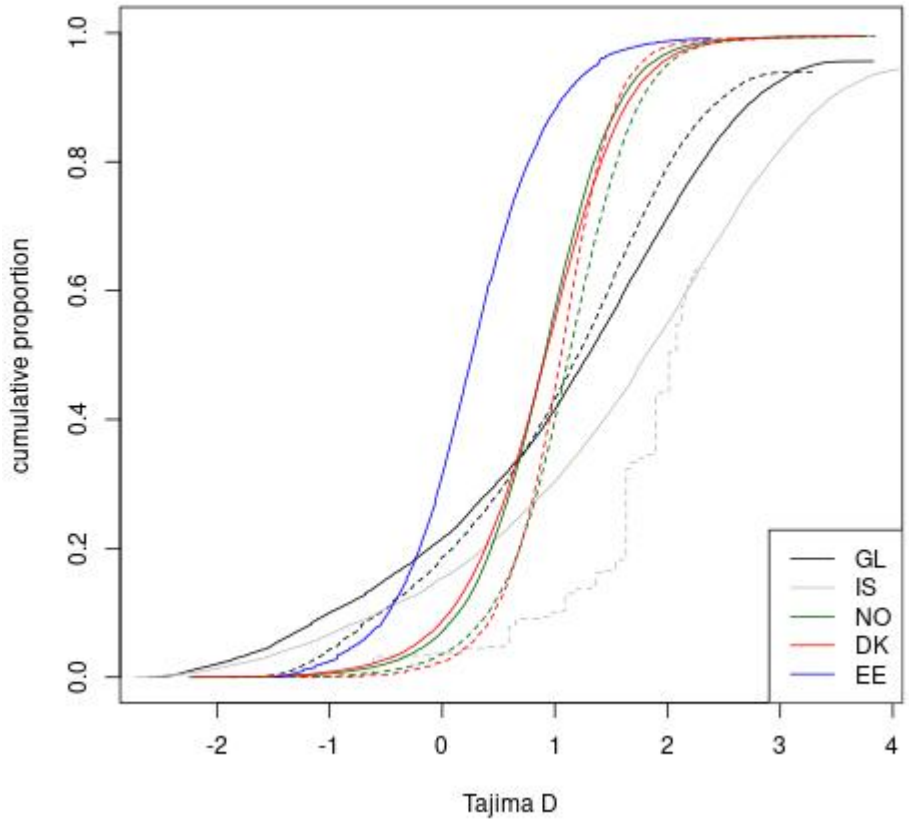


Figure S16. Accumulative Tajima's D values for Greenland (GL), Iceland (IS), Norway (NO), Denmark (DK), and Estonia (EE). Historic populations are in dotted lines.

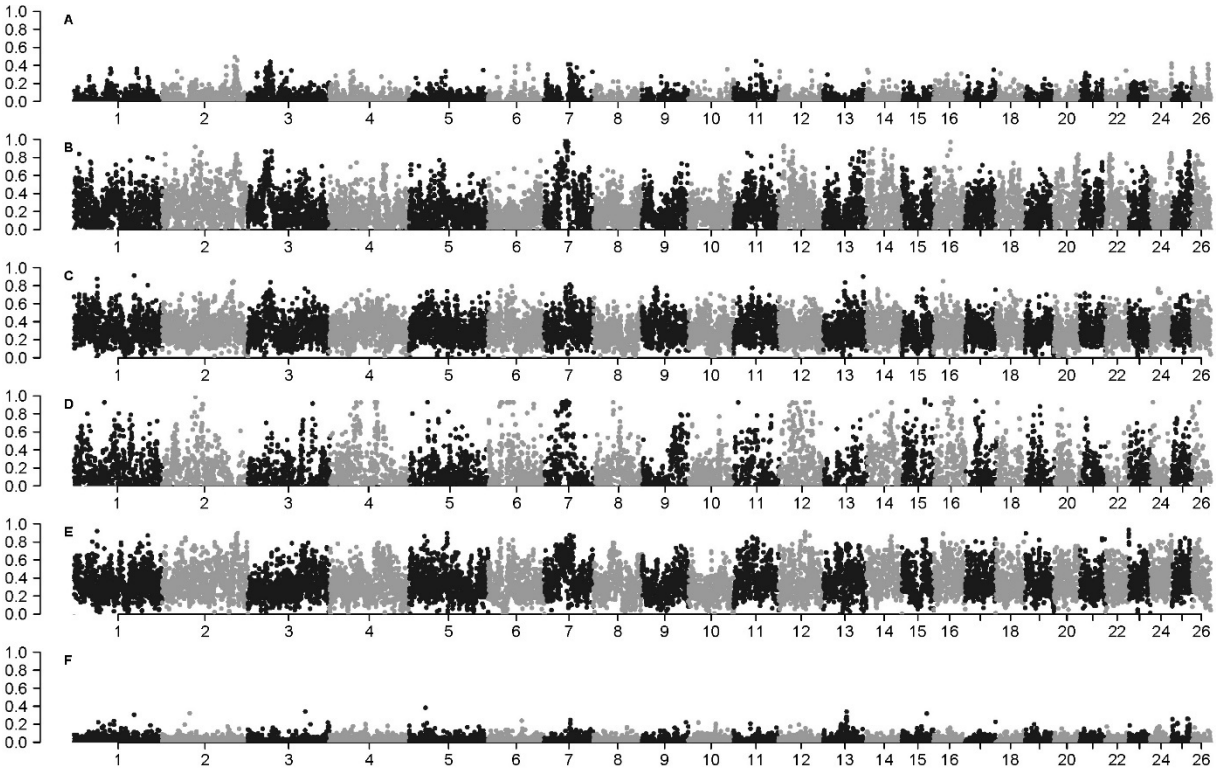


Figure S17. Weighted F_{ST} in windows of 100K bp along the genome for the populations; A) contemporary and historic Greenland, B) contemporary Greenland and contemporary Iceland, C) contemporary Greenland and contemporary Norway, D) contemporary and historic Iceland, E) contemporary Iceland and contemporary Norway, F) contemporary and historic Norway. The most extreme F_{ST} values, both between temporal populations and between countries are found between contemporary and historic Iceland (D), and contemporary Iceland vs contemporary Greenland (B).

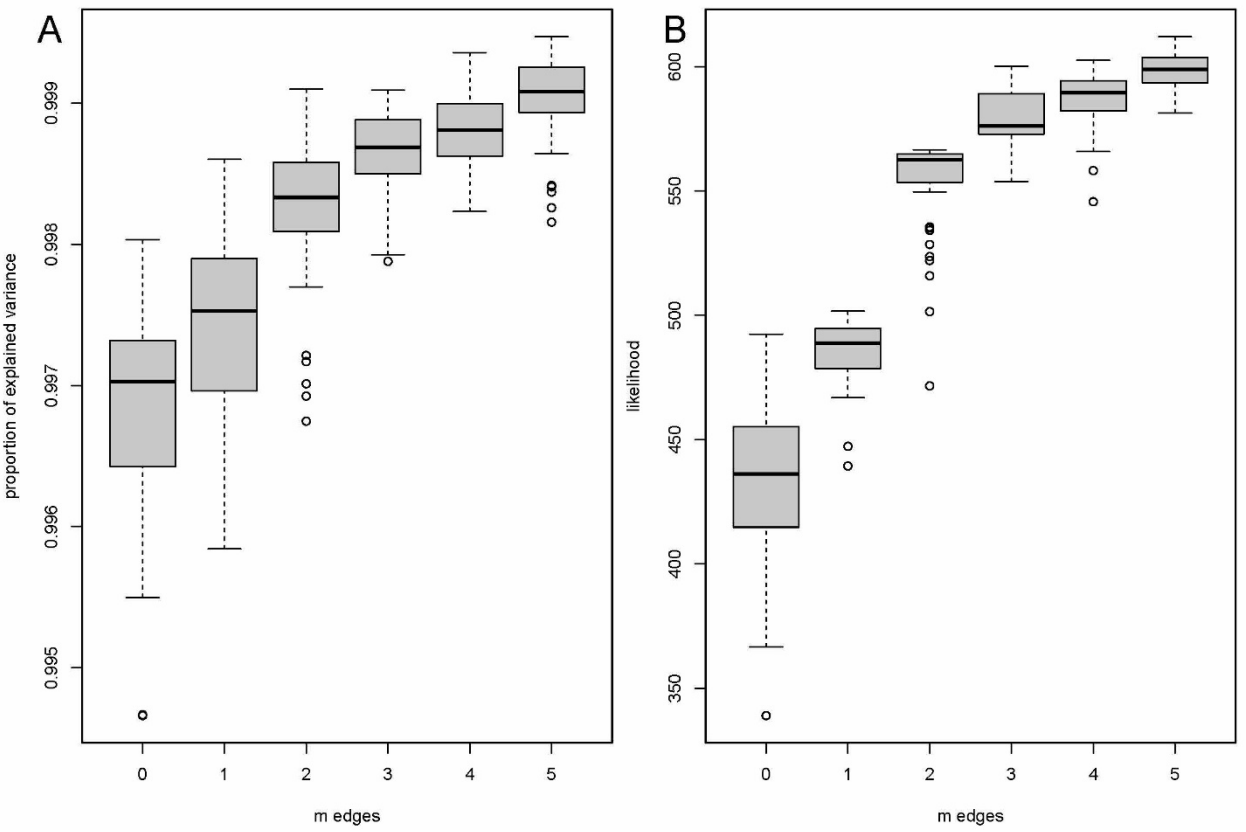


Figure S18. Treemix statistics for migrations edges (m) 0-5. A) proportion of explained variance, with the highest value in m 5. B) likelihood per m.

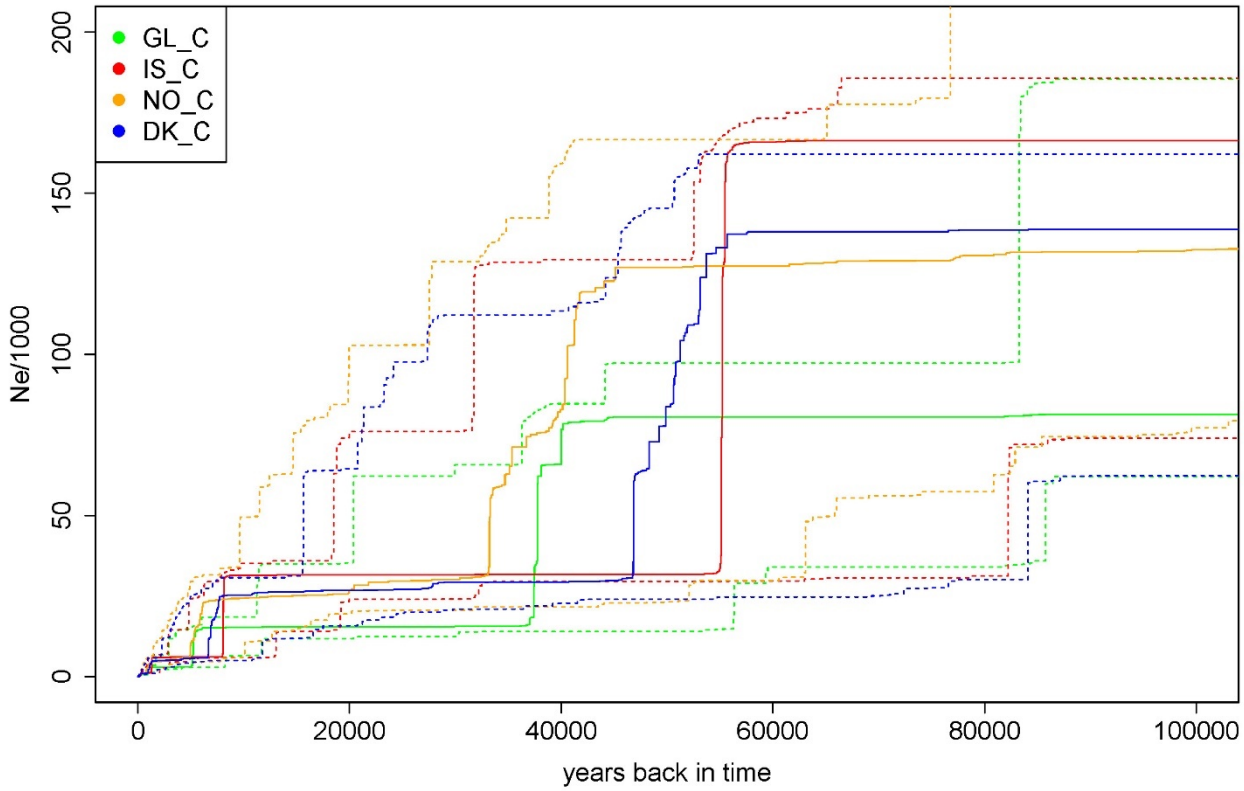


Figure S19. Stairway plot for the last 100,000 years. Stairway plot v2 for contemporary samples from Greenland (GL_C, green), Iceland (IS_C, red), Norway (NO_C, yellow), and Denmark (DK_C, blue). All populations are seen to have reduced population size from 100,000 years ago till today. Very large 2.5-97.5% confident intervals are found (dashed lines), especially for the two island populations in the present day, and for all back in time. The x-axis is years back in time from 0 to 100,000 years. Y-axis is the effective population size divided by 1000. Generation time of 15.6 years and mutation rate of $2.3*10^{-9}$ per site per year.

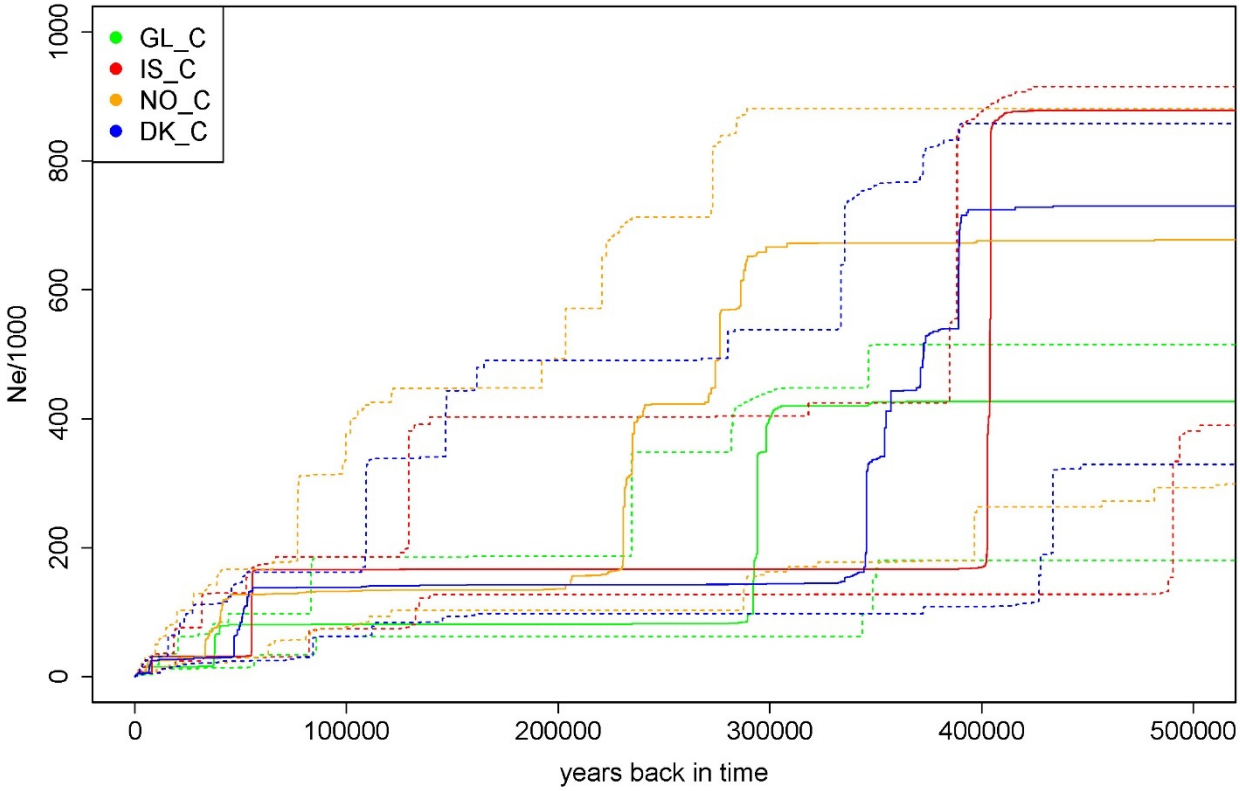


Figure S20. Stairway plot for the last 500,000 years. Stairway plot v2 for contemporary samples from Greenland (GL_C, green), Iceland (IS_C, red), Norway (NO_C, yellow), and Denmark (DK_C, blue). All populations are seen to have reduced population size from 100,000 years ago till today. Very large 2.5-97.5% confident intervals are found (dashed lines), especially for the two island populations in present day, and for all back in time. The x-axis is years back in time from 0 to 100,000 years. Y-axis is the effective population size divided by 1000. Generation time of 15.6 years and mutation rate of $2.3*10^{-9}$ per site per year.

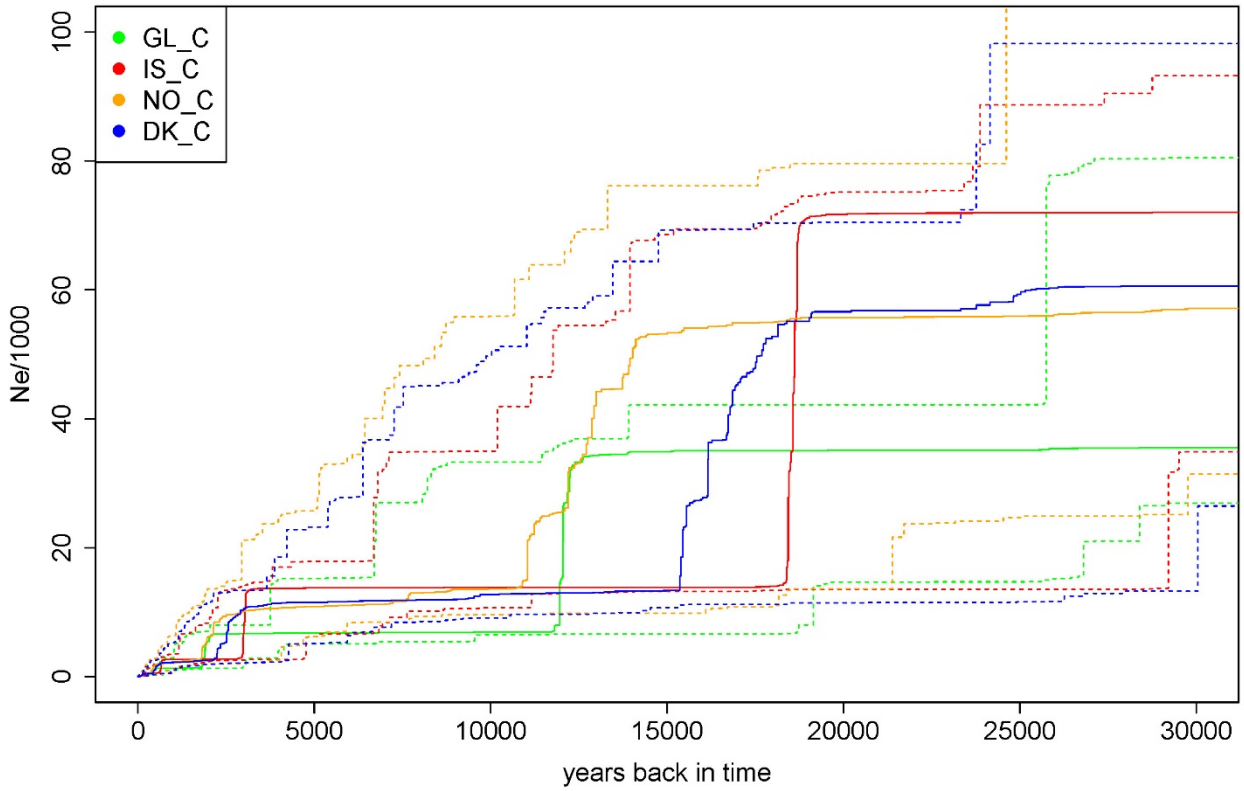


Figure S21. Stairway plot for the last 30,000 years. Stairway plot v2 for contemporary samples from Greenland (GL_C, green), Iceland (IS_C, red), Norway (NO_C, yellow), and Denmark (DK_C, blue), with a mutation rate of $1*10^{-9}$ per site per year and a generation time of 15.6. All populations are seen to have reduced population size from 10,000 years ago till today. The x-axis is years back in time from 0 to 10,000 years. Y-axis is the effective population size divided by 1000.

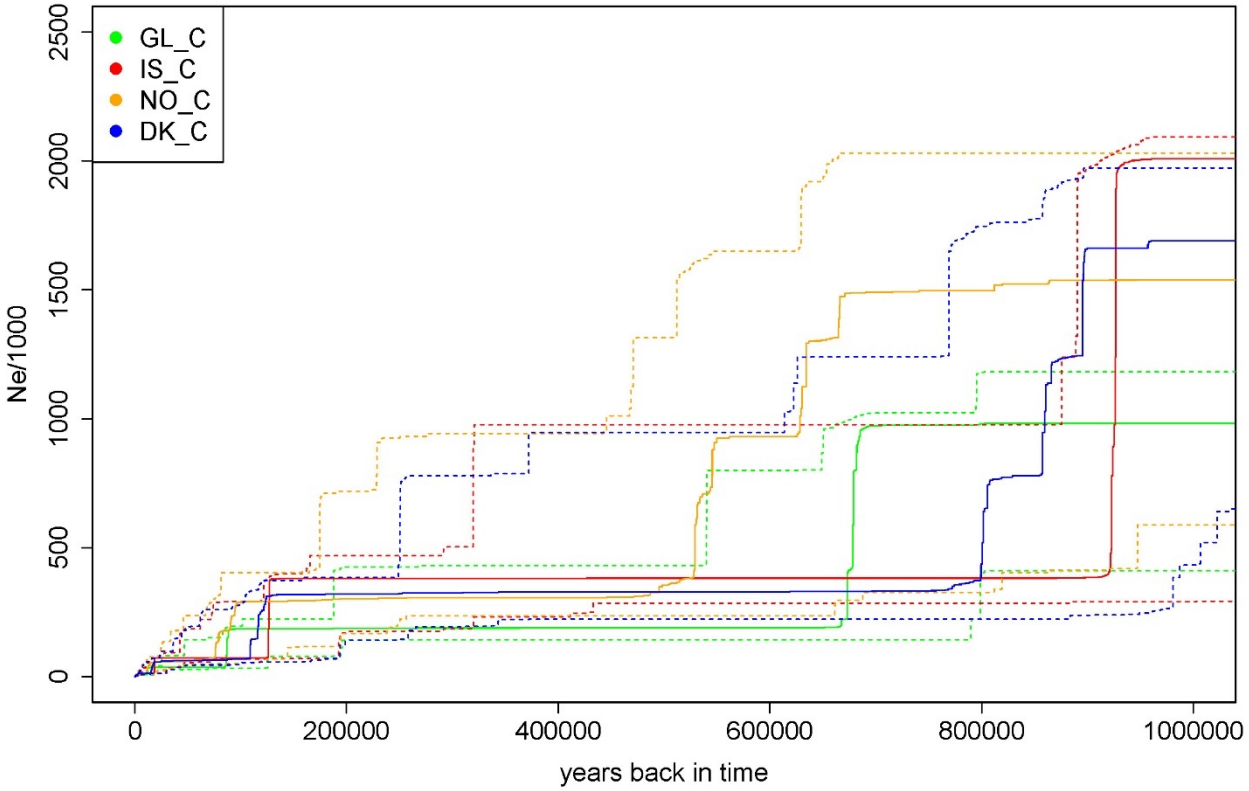


Figure S22. Stairway plot for the last 1,000,000 years. Stairway plot v2 for contemporary samples from Greenland (GL_C, green), Iceland (IS_C, red), Norway (NO_C, yellow) and Denmark (DK_C, blue), with a mutation rate of $1*10^{-9}$ per site per year and a generation time of 15.6. All populations are seen to have reduced population size from 10,000 years ago till today. The x-axis is years back in time from 0 to 200,000 years. Y-axis is the effective population size divided by 1000.