UPPSALA
UNIVERSITET

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology 1391

Inferring demographic history and speciation of grouse using whole genome sequences

RADOSLAV KOZMA

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2016

ISSN 1651-6214
ISBN 978-91-554-9627-2
urn:nbn:se:uu:diva-299926

Dissertation presented at Uppsala University to be publicly examined in Lindahlssalen, EBC (Evolutionary Biology Center), Norbyvägen 18B, Uppsala, Friday, 16 September 2016 at 10:00 for the degree of Doctor of Philosophy. The examination will be conducted in English. Faculty examiner: Assistant Professor Dr. Jeff A. Johnson (University of North Texas, Institute of Applied Sciences).

Abstract
Kozma, R. 2016. Inferring demographic history and speciation of grouse using whole genome sequences. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology* 1391. 49 pp. Uppsala: Acta Universitatis Upsaliensis. ISBN 978-91-554-9627-2.

From an ecological perspective, knowledge of demographic history is highly valuable because population size fluctuations can be matched to known climatic events, thereby revealing great insight into a species' reaction to past climate change. This in turn enables us to predict how they might respond to future climate scenarios. Prominently, with the advent of high-throughput sequencing it is now becoming possible to assemble genomes of non-model organisms thereby providing unprecedented resolution to the study of demographic history and speciation. This thesis utilises four species of grouse (Aves, subfamily Tetraoninae) in order to explore the demographic history and speciation within this lineage; the willow grouse, red grouse, rock ptarmigan and the black grouse. I, and my co-authors, begin by reviewing the plethora of methods used to estimate contemporary effective population size (N_e) and demographic history that are available to animal conservation practitioners. We find that their underlying assumptions and necessary input data can bias in their application, and thus we provide a summary of their applicability.

I then use the whole genomes of the black grouse, willow grouse and rock ptarmigan to infer their population dynamics within the last million years. I find three dominant periods that shape their demographic history: early Pleistocene cooling (3-0.9 Mya), the mid-Brunhes event (430 kya) and the last glacial period (110-10 kya). I also find strong signals of local population history – recolonization and subdivision events – affecting their demography. In the subsequent study, I explore the grouse dynamics within the last glacial period in more detail by including more distant samples and using ecological modelling to track habitat distribution changes. I further uncover strong signals of local population history, with multiple fringe populations undergoing severe bottlenecks. I also determine that future climate change is expected to drastically constrict the distribution of the studied grouse.

Lastly, I use whole genome sequencing to uncover 6 highly differentiated regions, containing 7 genes, hinting at their role in adaptation and speciation in three grouse taxa. I also locate a region of low differentiation, containing the *Agouti* pigmentation gene, indicating its role in the grouse plumage coloration.

Keywords: Demographic history, speciation, effective population size, adaptation, willow grouse, red grouse, black grouse, rock ptarmigan, Tetraoninae, conservation genetics, climate change, PSMC, species distribution modelling, FOXP4, Agouti

Radoslav Kozma, Department of Ecology and Genetics, Animal ecology, Norbyvägen 18 D, Uppsala University, SE-752 36 Uppsala, Sweden.

ISSN 1651-6214
ISBN 978-91-554-9627-2
urn:nbn:se:uu:diva-299926 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-299926)

To all the grouse...

Cover illustration by Jonas Nilsson.

List of Papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

I Graham, S., **Kozma**, **R**. and Höglund, J. The utility of effective population size in population management 1: estimating contemporary effective size. *Submitted manuscript*.

II **Kozma**, **R**., Graham, S. and Höglund, J. The utility of effective population size in population management 2: estimating demographic history. *Submitted manuscript*.

III **Kozma**, **R**., Melsted, P., Magnússon, K.P. and Höglund, J. (2016) Looking into the past - the reaction of three grouse species to climate change over the last million years using whole genome sequences. *Molecular Ecology*, 25(2):570-580

IV **Kozma**, **R**., Benito, B.M., Svenning, J-C and Höglund, J. Past and potential future dynamics of three grouse species using ecological and whole genome coalescent modelling. *Manuscript*.

V **Kozma**, **R**. and Höglund, J. Insight into speciation and adaptation in grouse as revealed by whole genome sequencing. *Manuscript*.

Additional publications

In addition to the papers included in the thesis, the author has also published the following papers.

i. Wang, B., Ekblom, R., Castoe, T.A., Jones, E.P., **Kozma**, **R**., Bongcam-Rudloff, E., Pollock, D.D. and Höglund, J. (2012) Transcriptome sequencing of black grouse (*Tetrao tetrix*) for immune gene discovery and microsatellite development. *Open Biology*, 2:120054

ii. Rózsa, J., Strand, T.M., Montadert, M., **Kozma**, **R**. and Höglund, J. (2016) Effects of a range expansion on adaptive and neutral genetic diversity in dispersal limited Hazel grouse (*Bonasa bonasia*) in the French Alps. *Conservation Genetics*. 17(2):401-412

Contents

Abbreviations

N_e	Effective population size
kya	Thousand years ago
Mya	Million years ago
LGM	Last Glacial Maximum [~21 kya]
LIG	Last Inter-Glacial [~130 kya]
MBE	Mid-Brunhes event [~430 kya]
PSMC	Pairwise Sequentially Markovian Coalescent

1. Introduction

1.1. Demographic history

Environmental conditions differ throughout the world and all living organisms inhabiting Earth's various ecosystems are subjected to a plethora of biotic and abiotic challenges. And while almost everywhere on Earth at least some kind of organisms can be found, these unique challenges ensure that there is no single species that can live ubiquitously. These restrictions therefore shape the geographic range of each species. It is the ways in which the ranges of all the species overlap that gives rise to the spatial pattern of biodiversity seen today (1). In addition to changing in space, environmental conditions also change throughout time. Twenty thousand years ago, during the peak of the last glacial period, the vast majority of Scandinavia was covered in a thick layer of permanent ice (2), so the past conditions in this region were drastically different from today. The species however do not stay put, instead they can track their preferred habitat (3) - in this case by surviving further south in Europe (4). This means that species' geographic range moves over time – it can expand and contract depending on the prevailing environmental circumstances. As a consequence, the number of individuals within that species changes as well. Under favourable conditions, populations can experience a 'boom' where population size quickly increases. Conversely, under severely unfavourable conditions population pass through a bottleneck, whereby only few individuals may survive (5).

Tracking these changes in species population size through time reveals the species' demographic history. In ecology this can be a worthwhile endeavour, because if the changes in population size can be aligned back to known climatic events [e.g. periods of cooling/warming], the demographic history can reveal important insight into the reaction of the species to climate change. This in turn can help predict how the species might react to potential future climate scenarios (6).

However, accurate tracking of population size can be very labour intensive and expensive. It is also reliant on past as well as presently continued effort, so this kind of information may be present for well managed and iconic species [e.g. >30 years of Grizzly bear estimation of population size from individual counts within their recovery zone in northern US, (7, 8)], but for the remainder of species accurate estimates of past population size is lacking. It is in cases like these that a biologist can turn to the vast body of theory on the subject of population genetics, which provides an alternative approach to studying past population dynamics.

1.2. Effective population size

The concept of effective population size [N_e] was proposed during the initial theoretical developments of population genetics back in early twentieth century by Sewall Wright (9). In its essence, it measures the force of genetic drift in the population of interest (10). It differs substantially from the census population size. For example, a population consisting of 10 identical clones [i.e. same genetic information], the census population size is 10 but the effective population size is 1, since every individual is the same. From this perspective, it is possible to see the duality of N_e. It captures the amount of genetic diversity within a population and therefore the potential force of genetic drift acting upon it. This is non-trivial information, because there is a general positive trend between the amount of genetic diversity harbouring within a population and the likelihood of its persistence (11–13). However persistence of genetically impoverished populations has been documented (see for example: 13,14). As such, N_e can act as a proxy for the overall genetic health of a population in question. This is further compounded by the fact that with an increase in N_e natural selection becomes better at fixing advantageous alleles and purging deleterious ones (16).

Moreover, the current effective population size is the result of past demographic processes acting upon the population. Humans are a case example. The current census population size is around 7.4 billion, however the effective population size is just a fraction of this - between 1,000-10,000 (17, 18) - reflecting the effect of long term bottleneck during the Pleistocene as well as the out-of-Africa dispersal and subsequent colonisation of Europe and Asia (19, 20). This highlights the intricate link between demographic history and effective population size. With expanding theoretical body of work,

decreasing costs of genome sequencing and ever increasing computational power we are now able to examine past fluctuations in N_e in unprecedented detail (19, 21, 22), which allows us to piece together a clearer picture of the demographic history of populations. All this reveals more insight into the ecology and evolutionary past of species – even ones that are now extinct [e.g. passenger pigeon (23) and woolly mammoth (24)].

The idea of estimating and tracking effective population size is now also gaining relevance to animal conservation and management (25, 26), with many estimation methods being available to practitioners (see *Paper I* and *II*). Within the scope of conservation genetics, N_e provides information on the rate at which populations are losing genetic variation through inbreeding or genetic drift. Populations of small N_e experience the highest rates of genetic drift and may therefore lose genetic variation through drift and inbreeding more quickly than they acquire it through mutation and migration. Consequently, small populations may not have the genetic variation necessary for adaptation when conditions change. This inability to adapt will cause increase in mortality rates and reduction in reproductive rates, which will together reduce N_e even more. The resulting population of smaller N_e will experience even stronger genetic drift and thus the downward spiral continues; the population is trapped in an extinction vortex (27, 28). As such, the amount of genetic variation present in a population [i.e. N_e] will dictate to a large parts its potential for adaptation.

1.3. Adaptation & Speciation

Adaptation can be described as the evolutionary process by which organisms become better at surviving in their environment (29). The genetic basis of adaptation may manifest itself in a variety of different traits; physiological, behavioural, structural or life-history. It is natural selection that drives the evolution of these adaptive traits, whereby different alleles will confer different fitness, and the more different alleles there are, the more scope for selection and adaptation exists. Over time this can lead to the process of speciation, in which enough differences between differently adapted populations build so that they can no longer mate among each other. As such, this process is intricately coupled to the aforementioned concept of effective population size, overall levels of genetic diversity and demographic history of populations (30).

The rise of Next Generation Sequencing [NGS] techniques has now spurred on the 'genomics' era in biology, where high throughput sequencing of whole genomes is becoming readily available. This allows the study of the genetic basis of adaptation at an unprecedented scale and, just as importantly, to be placed within a genomic perspective (31). Ultimately, utilising these tools allows the identification of adaptive genes in different species in the light of their own demographic histories, thus helping to elucidate the process of speciation as a whole.

1.4. Study species

In this thesis I utilise four species of grouse in order to explore the demographic history and speciation within this lineage: willow grouse [also called willow ptarmigan, *Lagopus lagopus lagopus*], red grouse [*Lagopus lagopus scoticus*], rock ptarmigan [*Lagopus muta*] and black grouse [*Tetrao tetrix*]; Fig. 1. All four belong to the Tetraoninae subfamily, with the *Tetrao* lineage splitting from the *Lagopus* lineage approximately 3 Mya, while the rock ptarmigan split from the *Lagopus lagopus* lineage 2-1 Mya (32, 33).

The willow grouse can be found in the subalpine habitat, boreal forests and moorland in the Palearctic and Nearctic (Fig. 2), surviving on a mainly herbivorous diet consisting in large parts of the willow [*Salix*] (34, 35). The red grouse is officially recognised as the subspecies of the willow grouse (34), however it is endemic to the moorlands of UK has thus formed a fully separate breeding population at least since the time the British Isles separated from mainland Europe [~6000 years ago (36, 37)]. No record of gene flow between the British red grouse and Scandinavian willow grouse exists, with even the more proximal Irish and British red grouse showing no gene flow and substantial genetic differentiation (38). The rock ptarmigan is the most cold adapted species of the four, whereby it is a sedentary species that breeds across the open arctic and subarctic habitat (Fig. 2), dominated by grasses, lichens and mosses and selects wintering areas that allow access to the ground vegetation [e.g. windswept ridges and slopes] (35, 39). Where the distributions of rock ptarmigan and willow grouse overlap, rock ptarmigan occurs at higher altitudes (40). Lastly, the black grouse has a broad habitat preference whereby it inhabits boreal forest edges and early stages of forest succession within the Palearctic. Outside the boreal region, it also inhabits moorlands and heaths (Fig. 2) (35).

Figure 1. The study species: a) willow grouse, showing the partial brown summer plumage, © XAlexandraS/Wikimedia Commons/CC-BY-SA-4.0, b) red grouse in year-round brown plumage, © MPF/Wikimedia Commons/CC-BY-SA-3.0, c) rock ptarmigan in all white winter plumage, © Ómar Runólfsson/Wikimedia Commons/CC-BY-SA-2.0 and d) black grouse in year-round black plumage, ©Vnp/Wikimedia Commons/CC-BY-SA-3.0

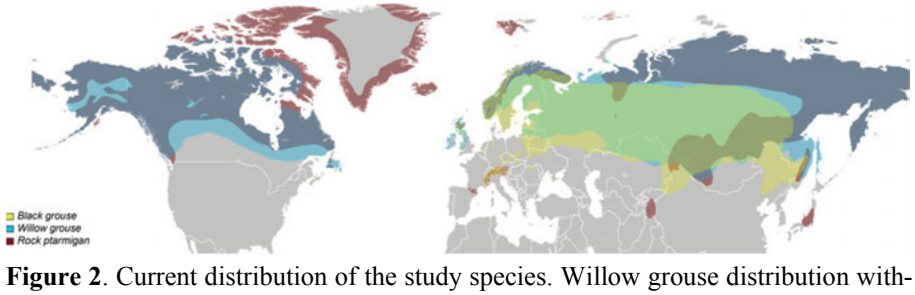

Figure 2. Current distribution of the study species. Willow grouse distribution within the UK and Ireland depicts the red grouse subspecies range. Taken from (41).

The study species also exhibit differences in plumage colouration; the rock ptarmigan and willow grouse have a brown summer plumage [full in females and partial in males] which moults into an all-white winter plumage, while both sexes of the red grouse forgo the white winter plumage instead remaining brown all year round. The black grouse never adopts a white winter plumage, but instead remains black throughout the year (Fig. 1).

Lastly, the species also differ in their mating behaviour. The black grouse has a lekking mating system, where males aggregate to competitively display in order entice females for mating (42). This results in relatively high levels of polygyny and large variance in male mating success [0-8, with up to 25 females per male, (43)]. In the willow grouse and rock ptarmigan the males are territorial and once mated the pairs remain largely monogamous, with 20-30% of males mating polygynously (44, 45).

1.5. Objectives

The overarching aim of the thesis is to explore the methods available to study effective population size and demographic history, apply these methods to investigate the past demographic history of the grouse species and, in the light of these results, utilise NGS techniques to study their adaptation and speciation dynamics.

Specifically, in ***Paper I*** and ***II*** we explore the ways in which contemporary N_e and demographic history can be estimated, respectively. Additionally, we examine how each can be applied to conservation and management.

In ***Paper III***, we utilise whole genome sequencing of three grouse genomes [willow grouse, rock ptarmigan and black grouse] to infer the broad scale pattern of past population dynamics in these species within the last 3 million years.

In ***Paper IV***, we build upon the results of Paper III by combining whole genome modelling with ecological modelling in order to test their concordance as well as to explore their demographic history in more detail during the last glacial period.

Lastly, in ***Paper V*** we resequence the genomes of 34 grouse individuals [willow grouse, red grouse and rock ptarmigan] and by performing a genome-wide F_{ST} outlier test we identify genomic regions of high as well as low differentiation. By studying the gene content within these regions, we examine the dynamics of speciation and adaptation in these grouse species.

2. Methods

2.1. Review of N_e estimation

Papers I and *II* provide a review, so for these papers we have gathered data from publications on: i) methods estimating contemporary N_e from demographic data as well as genetic data [including: standard temporal method, Jorde-Ryman method, MLNe, heterozygous excess, linkage disequilibrium, parentage assignment, molecular coancestry and approximate Bayesian computing] and ii) past N_e fluctuations [including: sequence mismatch analysis, Msvar, Bayesian skyline plots and Pairwise Sequentially Markovian Coalescent [PSMC] method].

2.2. Samples

For the study of broad scale demographic history (*Paper III*), we took advantage of the previously published black grouse genome (46), which is based on a black grouse male sampled in Norway in 2011. Additionally, a willow grouse sample was collected from a wild male in Sweden in 2011 and a rock ptarmigan sample was collected from a wild male in north-eastern Iceland in 2012. For the study of speciation and adaptation (*Paper V*), we resequenced 34 grouse individuals consisting of: 17 willow grouse sampled from Scandinavia between 1995 and 2006, one willow grouse from eastern Russia [Magadan] sampled in 1995, one willow grouse from Alaska, USA sampled in 1995, 9 red grouse individuals sampled in the northern England in 2013 and 6 rock ptarmigan individuals sampled in south-western Greenland in 2007. For the more fine scale demographic analysis (*Paper IV*) we re-used 5 of the individuals sequenced for *Paper V*: one Scandinavian willow grouse, the Russian and Alaskan willow grouse, one red grouse and one rock ptarmigan.

2.3. DNA extraction, sequencing and assembly

In all cases, DNA was extracted using the Qiagen DNeasy Blood & Tissue Kit® following the manufacturer's instructions (Qiagen) and DNA quality was checked on an agarose gel and quantified using the Quibit® Fluorometer. Library preparation was performed using the Illumina TruSeq protocol (www.illumina.com) following the manufacturer's instructions. For *Paper III*, the willow grouse and rock ptarmigan samples were sequenced using the Illumina HiSeq machine at deCODE Gentics Inc. in Reykjavik, Iceland. For *Paper IV* and *V*, the samples were sequenced at the SNP&SEQ technology platform of Uppsala University also using an Illumina HiSeq machine. In all three cases, the resultant reads were 125bp long with target insert size of 350bp. Quality filtering and error correction was performed by Musket [*Paper III*, (47)] and Trimmomatic [*Paper IV* and *V*, (48)] . In *Paper III*, the willow grouse and rock ptarmigan genomes were de-novo assembled using *SOAPdenovo2* (49). The scaffolds were then mapped onto the black grouse genome backbone using the *bwa-mem* alignment algorithm (50), in order to obtain chromosomal sequences. Resultant coverage[1] was 68*x* for the willow grouse and 101*x* for the rock ptarmigan. Heterozygosity rate and mutation rate were then also calculated using these chromosomal alignment sequences. For *Paper IV* and *V*, all properly paired reads that passed quality control were mapped onto the black grouse genome using the *bwa-mem* alignment algorithm (50) as well. Duplicate reads were marked with Picard (http://broadinstitute.github.io/picard/) and local realignment around indels[2] was performed with the GATK IndelRealigner tool (51, 52) producing the final filtered alignment files [in *bam* format]. The resultant mean coverage across all individuals was 28x. For full parameter details see the "Method" section of the respective paper.

2.4. PSMC analysis

This method was used in *Paper III* to infer the demographic history of a species based on a single genome (19), allowing us to go as far back as 3 Mya. Briefly, the method uses the density of heterozygous bases present within a single diploid individual to infer genomic blocks that share the same common ancestor [i.e. no recombination has occurred within these regions].

[1] the average number of reads covering each base-pair of the genome – a parameter indicating the confidence in discovery of a variable sites
[2] INsertions or DELetions within a sequence when aligned to a reference

Consequently, the method reconstructs the time to the most recent common ancestor (TMRCA) across the genome. Changes in TMRCA between these genomic blocks occur if recombination has taken place between them [termed 'transition probability'], where this transition probability is dependent on the mutation rate, recombination rate and the ratio of $N_e(t)/N_0$ [termed $\lambda(t)$ and represents the relative effective population size at state t]. Thus, from the changes in TMRCA and user supplied ratio of mutation rate/recombination rate, the program is able to estimate $\lambda(t)$ across all the coalescent intervals in the past [N_0 is calculated as $\theta/4\mu$, where θ is the 'population mutation rate' – a measure of nucleotide diversity - and μ is the per-generation neutral mutation rate (53)]. Finally, the generation time of the focal species and its mutation rate are needed in order to scale the PSMC output into years. Only the consensus autosomal[3] sequences [in *fastq* format] obtained from the filtered alignment file were used in the analysis. This is because the effective population size of the Z [sex] chromosome is 3/4 that of the autosomes (54), which would negatively affect the inference of demographic history. For full details of the parameters used to run the PSMC analysis, see the "Methods" section in *Paper III*.

2.5. Species Distribution Modelling (SDM) analysis

This method was used in *Paper IV* in addition to the PSMC analysis, in order to track changes in the area and location of the grouse suitable habitat at various time points in the past. Briefly, SDM is a correlative approach that computes habitat suitability maps based on the statistical relationship between the presence of the focal species at a precise point and their associated environmental predictors [usually climate] (55). These habitat suitability maps can then be projected at different times [e.g. known past climatic conditions or future projected climatic conditions] and thereby allow the visualisation and quantification of the changes in distribution over time. As such, it is a purely ecological approach to studying potential demographic changes occurring in the past.

For this method, we downloaded 19 climatic variables from Bioclim (www.worldclim.org/bioclim) for the present conditions, LIG, LGM, projected year 2050 and projected year 2070 [climatic models do not exist for periods prior the LIG, so that is the furthest back in time we can model accu-

[3] All chromosomes apart from the sex chromosome

rately]. To reduce autocorrelation, the 19 variables were ultimately filtered down to 5, which were deemed the most important for the grouse ecology. We then used two datasets to represent presence points: presence records downloaded from GBIF (www.gbif.org) and pseudo-presence points generated from range maps, which were downloaded from Birdlife international (56). We then used a weighted polynomial GLM to build a model to predict suitable habitat for each time frame. From the resultant maps, total suitable habitat area as well as changes in range sizes between time periods were calculated.

2.6. Population genomic analysis

This approach was used in *Paper V* to uncover the speciation and adaptation dynamics between the willow grouse, red grouse and rock ptarmigan. Throughout, all population genetic parameters estimation was based on genotype likelihoods obtained from the filtered alignment files [in *bam* format] created during the assembly process (see section 2.3). First, we used the ANGSD program (57) to calculate allele frequency likelihoods for each taxon separately. Then from these likelihoods, we calculated the 2 dimensional site frequency spectrum for each taxon pair. Subsequently, these were used to estimate the F_{ST}[4] across the genome of each taxon pair in non-overlapping 15kb windows. The F_{ST} scores were further Z-transformed and any window with a ZF_{ST} score of ≥ 6 was considered an outlier. Lastly, we used BLAST (58) to align the outlier windows to the chicken genome in order to identify gene content.

Using ANGSD and the previously calculated allele frequency likelihoods, we also estimated diversity and neutrality test-statistics [pairwise theta [π], Tajima's *D*, Fay and Wu's *H* (59)] for each taxon separately. This was done in order to provide more resolution of the genome architecture within the outlier regions, as well as the non-differentiated genomic regions. By genotyping all individuals together, rather than splitting them into separate taxa as was done previously, we were also able to perform a principal component analysis [PCA] using the ngsTools program (60).

[4] A measure of differentiation, ranging from 0-1. Higher value indicates higher level of differentiation of that region between the two species being compared.

We also reconstructed the phylogeny of all 34 individuals by genotyping them at a random subset of 500 000 autosomal SNPs[5], which were further filtered to ~59 000 in order to remove redundancy due to linkage disequilibrium. A maximum likelihood phylogeny tree was subsequently inferred from these SNPs using the RaxML program (61).

[5] Single Nucleotide Polymorphism – a variable site

3. Results and Discussion

3.1. *Paper I* – Contemporary N_e estimation

There are multiple ways of estimating contemporary N_e, but broadly they can be classified into three categories based on what data they require and how often the population of interest needs to be sampled: i) demographic methods, ii) genetic - temporally separated methods and iii) genetic - single sample methods.

Demographic methods

These approaches work on the idea that the empirical N_e of a population can be calculated by investigating how the population of conservation interest specifically diverges from an ideal population[6] in its demographic parameters. These parameters include: sex ratio, population size and variance in family size. More specifically, any deviance from an equal sex ration will decrease the N_e, because some of the genetic variation held by the more common sex in the parental generation cannot be passed to the offspring generation. This increases the strength of genetic drift. Fluctuations in population size will also affect N_e, whereby the periods of small population size will disproportionally decrease N_e because population bottlenecks remove variation from the population. Conversely, population booms only increase the amount of variation very slowly through the occurrence of novel mutations. Lastly, in real populations individuals differ in their probabilities of having offspring. Some produce disproportionally more than others. The larger the skew, the more variation is lost and thus the N_e decreases. The main consideration when using demographic data is that a real population will most likely differ in more than one of the above mentioned parameters. As such, if the goal is to apply this approach to a population with conservation and management concern, the data will need to be accurately estimated.

[6] Ideal population is a theoretical population that has a specifically assumed set of population parameters and life history traits, wherein genetic drift is the only evolutionary force acting on the population. Therefore there is no selection, mutation or migration.

And if the goal is to monitor this over time, then one needs to be able to detect changes in these parameters between years.

Genetic – temporally separated methods

Since acquiring accurate demographic data can be very impractical and time consuming, focus has shifted to calculating contemporary N_e from genetic data. The temporally separated genetic methods work on the principle that by sampling the population at two time points, the change in allele frequencies can be calculated, which can then be used to infer the strength of genetic drift - and thus quantify N_e. Three main methods have been implemented in population management: 'Standard temporal' method (62, 63), 'Jorde-Ryman' method (64, 65) and 'MLNe' (66). The standard temporal method requires the two samples to be separate by 5-10 generations, which can be very impractical for any long-lived species. And even reliance on museum samples might lead to inaccurate estimations of contemporary N_e due to the limited genetic resolution they offer. The Jorde-Ryman method overcomes this by sampling two subsequent generations; however, for accurate estimation, the method requires around 50 individuals to be genotyped for at least 1000 polymorphic loci, in addition to having information on vital rates of the population. This can also then be rather impractical. Lastly, the MLNe aims to not only estimate N_e but also quantify the effect of migration. The method requires around 200 samples to be separated by 2-4 generations, genotyped at 10-20 polymorphic loci. In addition, the possible source population of migrants needs to be sampled in order to estimate allele frequencies. Taking into account all these considerations, we recommend the MLNe method to be used in the effort of estimating Ne and thus evaluate the genetic health of the population.

Genetic – single sample methods

Waiting for a number of generations between samples can be problematic, especially in long-lived species and as such, methods where only a single temporal sample is required are gaining popularity. By calculating the allele frequencies within a population, these methods produce a variety of parameters which can then in turn be used to estimate N_e. The heterozygosity [the 'heterozygosity excess' method (67)] and linkage disequilibrium [the 'LD' method (68)] are such parameters. However, in both cases, evaluation of the methods reveals that they suffer from a lack of sensitivity and thus can be liable to provide wrong estimates of N_e. The 'estimator by parentage assignment' method [EPA, (69)] allows the estimation of N_e by calculating the

distribution of parent-offspring relationships within a single cohort. As such, the amount of genetic data required is as much as is needed to accurately estimate parent-offspring relationship, which can be as low of 10-20 microsatellite markers. However, a large proportion of a cohort needs to sampled, therefore this method is restricted to populations where cohort sizes remain manageable. Lastly, the 'Approximate Bayesian computing' method offers a Bayesian approach to calculating N_e (70). By calculating 8 summary statistics based on the observed allele frequencies and subsequently running population simulations with known N_e, the method can calculate the likelihood of a particular N_e, given the set of 8 summary statistics. While the method provides a compelling approach to N_e estimation, it has not been thoroughly evaluated and as such caution should be taken.

The overall ramification of the paper is that conservation practitioners have many options when selecting a N_e estimation method, but they have little information on which to base their selection. It is therefore unsurprising that many estimation methods are applied in situations where they are expected to be biased. Additionally, each method has its own advantages and drawbacks and its successful application will depend on many factors pertaining to the population, such as generation time, census size, ease of cohort identification and availability of genetic/genomic resources. With this in mind, *Figure 3* provides a flowcharts that aims to help in selecting the most appropriate contemporary N_e estimation method.

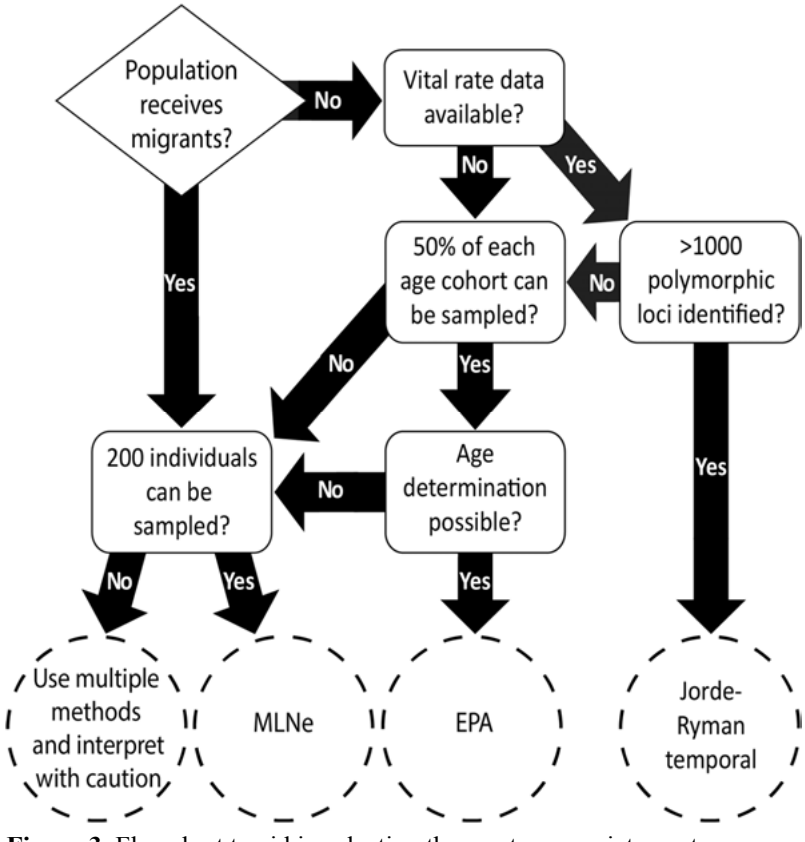

Figure 3. Flowchart to aid in selecting the most appropriate contemporary N_e estimation method.

3.2. *Paper II* – Demographic history estimation

Following from *Paper I*, conservation practitioners now have the option to not only estimate contemporary effective population size but to also study the past fluctuations of N_e, thereby piecing together the species' demographic history. However, the major obstacle is that the variety of methods available require different genetic data inputs and have varying temporal resolutions. As such, the type of questions that can be answered differs between the methods.

We separate the methods into 3 broad categories, wherein each attempts to investigate a different aspect of species' past demography: i) methods that study recent population declines, ii) methods capturing the effect of LGM

[~20 kya to present] and iii) methods detecting long-term fluctuations in N_e [>1 Mya to present].

Recent population declines

To detect and time recent population declines we advocate the use of the Msvar program (71, 72). It is a powerful method that can detect population changes occurring as recently as within the last century. As such, it is a useful method for detecting the anthropogenic influence on the population of interest because over such time scale detailed knowledge of how humans affected the particular habitat may be known. This therefore allows the method to be used for detailed hypothesis testing [e.g. forest exploitation of humans and the subsequent decline in the Bornean orang-utan, *Pongo pygmaeus*, (73)]. Also encouraging is that the method returns accurate and precise estimates when around 50 individuals are genotyped at 10-20 microsatellite loci – numbers that represent a very achievable goal for conservation practitioners today.

Effect of LGM on demography

To study the effect that the LGM and subsequent warming had on a population of interest, we advocate the use of the Extended Bayesian Skyline Plot [EBSP] method (74). This method has high precision in detecting fluctuations in N_e spanning over the last ~30 thousand years, and as such is ideal in capturing the demographic patterns of species throughout the LGM. The scope of the method allows the study of species response to this particular climate change scenario because over such time frame climatic conditions are thought to be main driver of N_e fluctuations. The method is also quite broadly applicable, because accurate results can be obtained with as few as 8 individuals genotyped at ~50 highly polymorphic markers (75) – numbers we feel are again very achievable with current methods such as RAD-sequencing (76).

Long-term N_e fluctuations

To detect long-term species dynamics we recommend the PSMC program (19). The method is able to uncover the demographic history patterns spanning over millions of years, thereby giving unprecedented view into the species' past. As such, it can be used to study the response of species to various climatic conditions dominating the earth at various time points [e.g. the climatically driven population fluctuations of the giant panda, *Ailuropoda melanoleuca*, (77)]. Or, it can be used to investigate whether the species'

history was dominated by recurrent bottlenecks, which can give insight into the species' ability to recover from such a process in the present (23). The main limitation of the method is that it requires a whole genome of [at least] one individual to be sequenced at a minimum of ~20x. This represents a genomic resource that not every animal with a conservation concern has. Encouragingly, the growing number of species with their genomes sequenced and made publicly available provides the opportunity that a genome of a related species may be used to alleviate this problem. The sequencing costs however can still be limiting.

3.3. *Paper III* – PSMC of grouse

The purpose of this paper was to use the PSMC method and whole genome sequences of the willow grouse, rock ptarmigan and black grouse to infer large-scale pattern of demographic history within the last ~3 million years. The general climatic conditions spanning this period are known (2, 78–81), which allowed us to make direct hypotheses:

i) Early Pleistocene cooling (3–0.9 Mya, blue section in *Fig. 4*), which saw the drop in average global temperatures by around 2 °C, should cause a population increase in the more cold adapted species – the willow grouse and rock ptarmigan. The black grouse population size should decline.
ii) The mid-Brunhes event (MBE, 430–110 kya, yellow section in *Fig. 4*), characterized by large climatic oscillations with very warm interglacials and cold glacials will cause recurrent population contractions and expansions, hence, lowering the N_e of all three species.
iii) During the last ice age and LGM (110–10 kya, grey section in *Fig. 4*), the cold temperatures should again favour willow grouse and rock ptarmigan and decrease the N_e of black grouse.

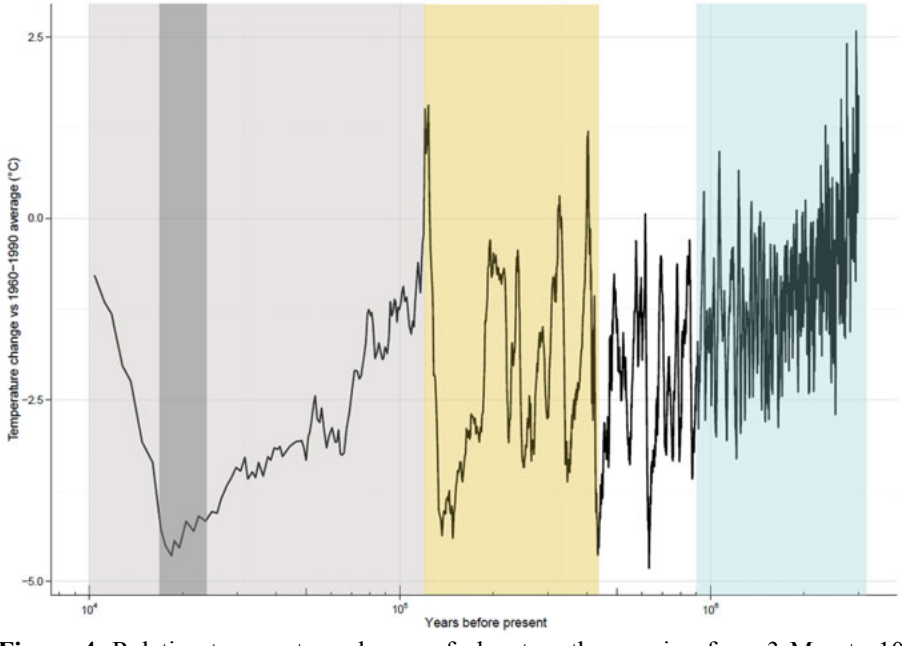

Figure 4. Relative temperature change of planet earth spanning from 3 Mya to 10 kya. Blue: Early Pleistocene cooling, yellow: Mid-Brunhes event and following climatic oscillations, light grey: last ice age, dark grey: LGM. Taken from (41).

The resulting PSMC curves found support for some but not all of the hypotheses. Firstly, we did indeed find that the early Pleistocene cooling favoured the proliferation of the cold adapted grouse [willow grouse and rock ptarmigan], and also decreased the population size of the black grouse (blue regions in *Fig. 5*).

Secondly, the prediction that the mid-Brunhes event would lower the effective population size of all three species was only partially supported by the PSMC trajectories. The predicted pattern was seen in the willow grouse and rock ptarmigan, where both species reached a peak population size following the MBE and are on a sharp decrease entering the last ice age. In comparison, the black grouse experienced a population increase during the 430–110 kya time period and reached a peak N_e at the onset of the last ice age (yellow regions in *Fig. 5*). This may be due to the black grouse not being limited by the cyclic climatic conditions as is the case for the willow grouse and rock ptarmigan. Alternatively, population subdivision can cause an artificial increase in N_e, resulting in a hump in the PSMC trajectory. This scenario would fit with the Saalian glaciation [160-140 kya] where the ice sheet extended as far south as the north shore of the black sea, effectively separating the European and Asian black grouse populations (82).

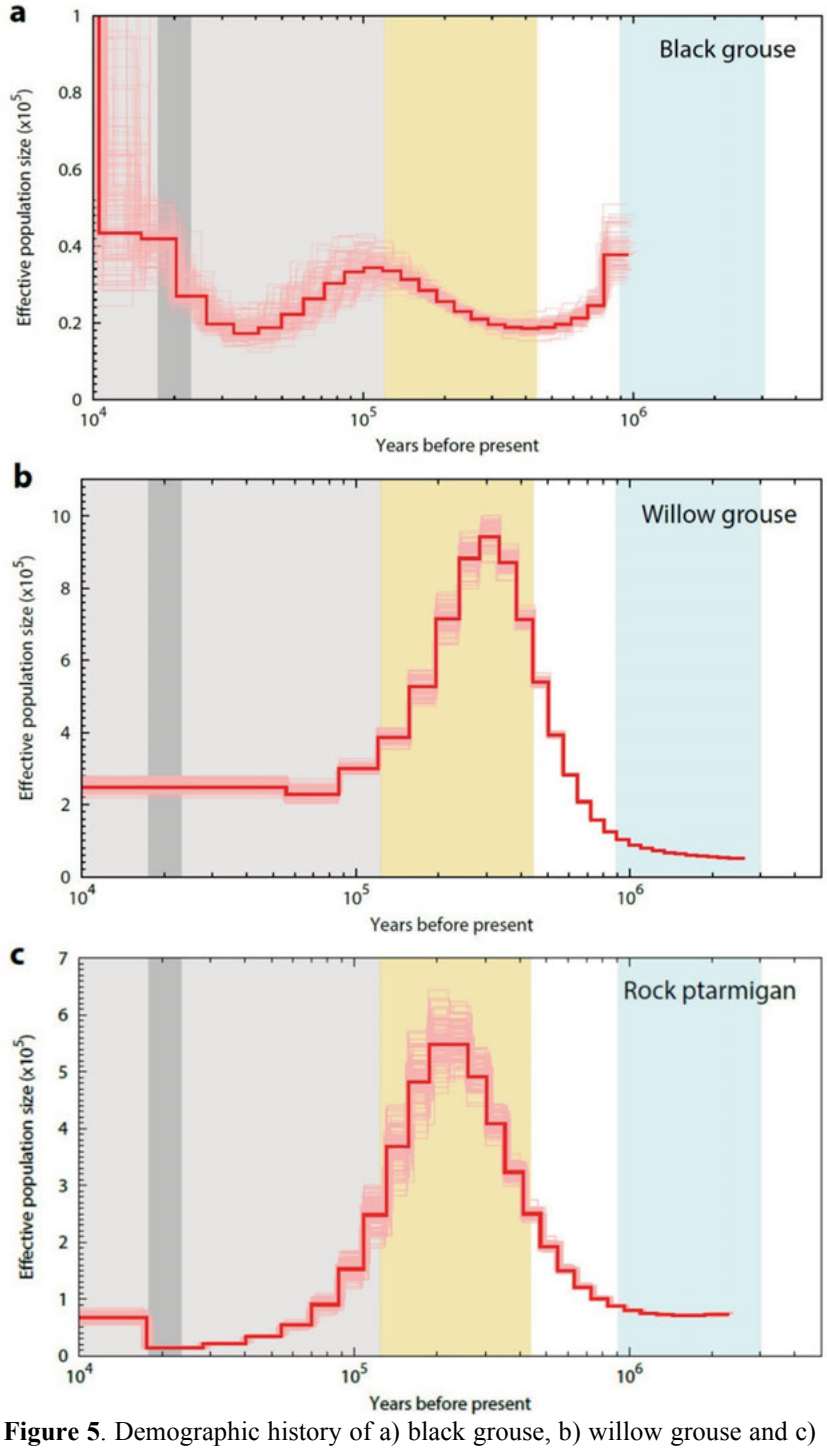

Figure 5. Demographic history of a) black grouse, b) willow grouse and c) rock ptarmigan, spanning from ~3 Mya to 10 kya. Coloured bars correspond to the climatic periods delineated in *Fig. 4*. Taken from (41).

Lastly, the PSMC trajectory during the last glacial period and LGM did not follow our predictions. The willow grouse population decreased and was maintained at a constant, relatively low, size; the rock ptarmigan underwent a severe bottleneck and only increased following the termination of the LGM, while the black grouse population increased already prior to the LGM [starting at ~35 kya]. This most likely indicates that within the last glacial period the PSMC only shows the local population trajectories and not the whole species dynamics. More specifically, the rock ptarmigan was sampled in Iceland, which was covered in ice during the LGM, and as such, the severe bottleneck most likely represents the recolonization of Iceland following its deglaciation. To this effect, our timing of the population boom [~17 kya] coincides with the hypothesised colonisation of Iceland based on nuclear and mitochondrial markers – around 19 kya (83). The willow grouse was sampled in Sweden, and the constant population size seems to indicate that the Scandinavian population experienced latitudinal shifts, but not range expansions. It is also know that the Russian and Scandinavian willow grouse are genetically different (34), thus the recolonization of Scandinavia following deglaciation must have occurred from southern Europe and not Russia. On the contrary, the Scandinavian black grouse population did most likely originate in the Ural Mountains (84). This part of the Palearctic was not as heavily glaciated during the LGM (82), so in agreement with the PSMC trajectories, it seems likely that the population has already expanded following the deglaciation of its eastern range prior the LGM and then extended further into Scandinavia once this region became ice-free (see *Table 1* in *Paper III* for more details about the key events in the grouse history).

Overall, the PSMC method allowed us to investigate population dynamics of these three species in an unprecedented time scale, thereby enabling the testing of various hypotheses. However it also showed that caution should be taken when interpreting the more recent population size changes, as they can be heavily influenced by processes acting on the local population from which the individual was sampled.

3.4. *Paper IV* – SDM of grouse

As a consequence of the results obtained in *Paper* III, in this study we aimed to gain a more detailed understanding the grouse demographic history spanning from the LIG up to the present. To achieve this we sequenced further

grouse individuals [including red grouse from the UK, as well as Russian and Alaskan willow grouse individuals and rock ptarmigan from Greenland] to perform the PSMC analysis on populations with different local histories. Additionally, we performed SDM analysis to track the extent of favourable habitat for the willow grouse, black grouse and rock ptarmigan, which also enabled a more direct comparison between the two approaches [e.g. tracking whether an increase in range size is mirrored by an increase in N_e].

The SDM analysis showed that all three species experienced a contraction in total range size during the very warm LIG (*Figures 6* and *7*). This contraction was drastic for the willow grouse and rock ptarmigan, both of which experienced population subdivision between the European and Asian range extents as well having overall patchy distribution throughout the northern Palearctic (*Figure 6* here, as well as *Fig. 1, 2, S1* and *S2* in *Paper IV*). The conditions were also severe for the black grouse, with population subdivision occurring between the European/West Asian region and East Asian region (*Fig. 3* in *Paper IV*). The cold LGM saw the largest range extent for the cold adapted willow grouse and rock ptarmigan, with both maintaining a well-connected habitat throughout the northern Palearctic. It also saw an increase in range size and overall connectivity in the black grouse, however, in this species it is the current climatic conditions that offer the largest range extent. Lastly, the increasing temperatures over the next half a century are expected to decrease the range size of all three species again, in the case of the willow grouse and rock ptarmigan returning them to similar distributions as seen during the LIG.

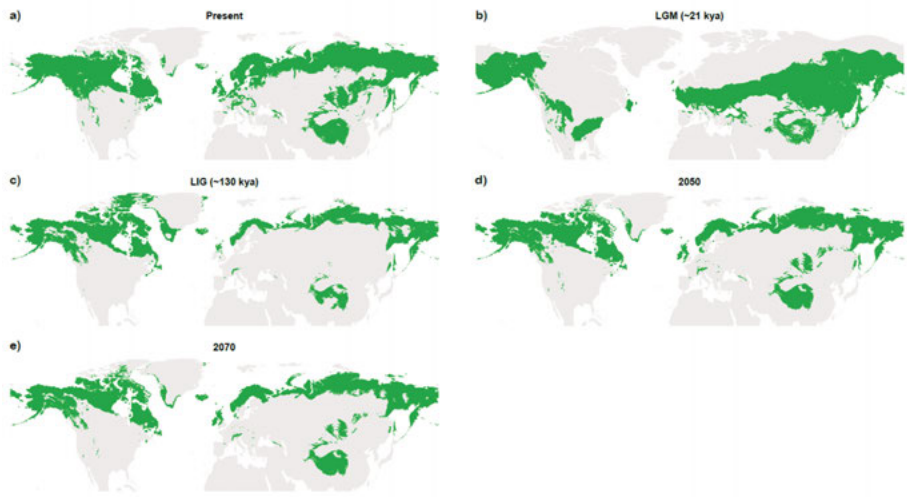

Figure 6. The modelled range of the willow grouse at a) present time, b) Last Glacial Maximum [LGM, ~21 kya], c) Last Inter-Glacial [LIG, ~130 kya], d) projected year 2050 and e) projected year 2070. See *Fig. 1* in *Paper IV* for greater detail.

The extended sampling and resultant PSMC curves did indeed highlight the effect of processes acting on localised populations during the last glacial period (*Figure 8*). The red grouse and willow grouse show the same trajectories as the previously published individual (*Figure 5b*), but the Russian and Alaskan individuals show a second population expansion following the onset of the last ice age. The Alaskan willow grouse then experienced a further bottleneck, while the Russian willow grouse was maintained at large population size. Likewise, the Greenland rock ptarmigan also showed a similar bottleneck as the previously sequenced Icelandic individual (*Figure 5c*).

Four main implications can be determined from this study. Firstly, the extended sampling does illustrate that the PSMC captures lineage specific population dynamics. This is seen in the divergent population size trajectories of willow grouse samples originating in Scandinavia and Britain versus Siberia and Alaska. This result further validates the observed genetic sub-structuring within the *L. lagopus lagopus* clade, where the Russian individuals cluster closer to the North American willow grouse [*L. lagopus muriei* & *L. lagopus alexandrae*] than to the Scandinavian willow grouse (34). Thereby also supporting the conclusions made in *Paper III*.

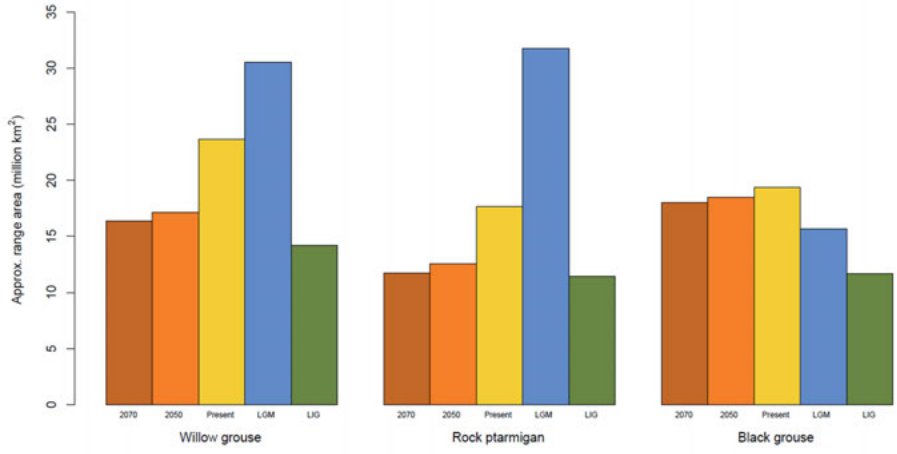

Figure 7. The estimated total range size of the three grouse species modelled across the 5 time periods. For willow grouse and rock ptarmigan, we see the largest range extent during the LGM, while their predicted future ranges will be comparable to their LIG range extent. Black grouse is experiencing the largest range extent at present, but future climate change will shrink its habitat as well.

Secondly, we see a good overlap between the SDM and PSMC methods for the Siberian willow grouse, where the SDM method predicts a steady increase in suitable habitat from the LIG up until the LGM and the PSMC does indeed show an increase in N_e which then remains at a stable high level. It is the discord between the two methods for the remaining samples that points to their special underlying demographic histories. During the LGM, the SDM predicts the largest range size for the willow grouse and rock ptarmigan, however, the PSMC shows large population bottlenecks in all but the Siberian sample. This indicates that the crashes must have been due to due other demographic events rather than unfavourable conditions. Since all but the Siberian individuals are in fringe populations that were either glaciated during the LGM [Greenland, Iceland, Scandinavia and Britain (2, 81)] or were separated from the remainder of their distribution prior to the LGM [Alaska (85)], the population crashes must indicate recolonization of these areas following their deglaciation. The Siberian willow grouse population, which occurs in the centre of the habitat range, did not undergo these recolonization events and thus follows the SDM predictions much more closely.

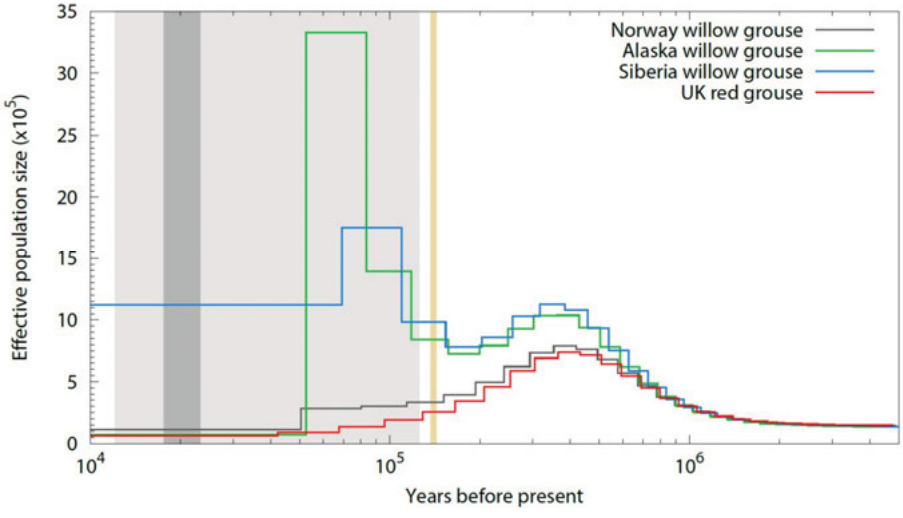

Figure 8. PSMC trajectory of the red grouse and various willow grouse subpopulations. Yellow: LIG, light grey: last glacial period, dark grey: LGM. For the Greenland rock ptarmigan PSMC trajectory, see *Fig. 6* in *Paper IV*.

Thirdly, in the case of the black grouse, the SDM analysis supports the 'subdivision' hypothesis made in *Paper III*, where the subdivision of the black grouse into East Asian and European subpopulations throughout the LIG artificially augments the N_e for the duration of the subdivision (see *Fig. 3* in *Paper IV*).

Lastly, from the SDM results we see that the future climate change will not favour any of the grouse. In fact, the willow grouse and rock ptarmigan are expected to lose about 30% of their current range, returning to similar distributions seen during the LIG. This corroborates the idea within the paleoecology community that the LIG can be seen as an analogue to the future warm climate (86). And while the species are expected to decrease in number as a result, it is important to keep in mind that they did survive throughout the warm LIG. This, however, also highlights the need to prevent alteration to the core habitat which is expected to be crucial for their survival - namely the arctic and boreal forests of the northern Europe and Russia.

3.5. *Paper V* – Speciation and adaptation

After gaining a deeper knowledge of the demographic history, here we set out to sequence the genomes of 34 individuals comprising the willow

grouse, red grouse and rock ptarmigan in order to uncover the genomic architecture of their speciation and the genes involved in adaptation.

By performing an F_{ST} outlier test, we uncovered 6 non-overlapping regions (2 regions for each taxon, *Figure 9*) containing the sequences of 7 genes (*Figure 10*). In the willow grouse the positively selected genes were *CDH7* [on chromosome 2] and *FOXP4* [on chromosome 26]. In the red grouse the genes were *SUN3* [on chromosome 2] and *EDIL3* [on chromosome Z] while in the rock ptarmigan these were *GADD45A* [on chromosome Z], *ROMO1* and *CPNE1* [both on chromosome 20].

The *FOXP4* gene, selected for in the willow grouse, is an intriguing candidate for speciation. It is part of the forkhead box [FOX] group of transcription factors (87), another member of which [*FOXP2*] has been shown to be highly involved in neural development, specifically important for learning of bird song and human speech (88, 89). Importantly, the product of the *FOXP4* gene has been shown to function in similar fashion to *FOXP2* (90). Yet the willow grouse song repertoire and complexity are rather nominal, which raises two potential issues. Firstly, it is perhaps possible that the song repertoire of the willow grouse hides more intricacies than once thought and would thus select for higher call learning capability. Or, it may be the case that *FOXP4* has a more ubiquitous role in bird calls and may be selected for even in species without intricate song repertoires.

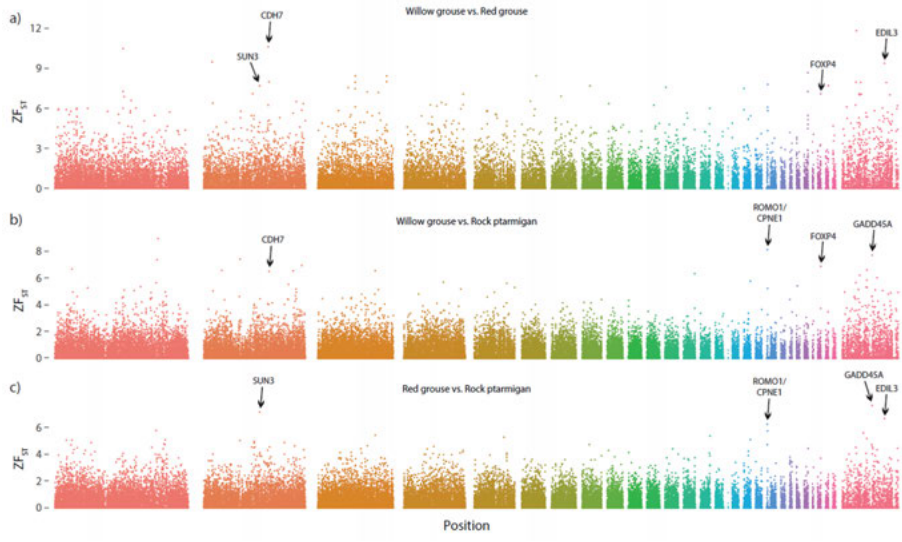

Figure 9. Genome-wide F_{ST} outlier test, showing the three pairwise comparisons of the three study taxa; a) willow grouse vs. red grouse, b) willow grouse vs. rock ptarmigan and c) red grouse vs. rock ptarmigan. The y-axis shows the ZF_{ST} score, where a 15kb window with a score ≥ 6 is deemed an outlier – i.e. differentially selected in one taxon. Each colour represents a different chromosome, with the autosomes arranged 1-28 [left to right] and chromosome Z located on the far right. The genes lying within same outlier window detected in two of the three comparisons are shown. See *Fig. 2* in *Paper V* for greater detail.

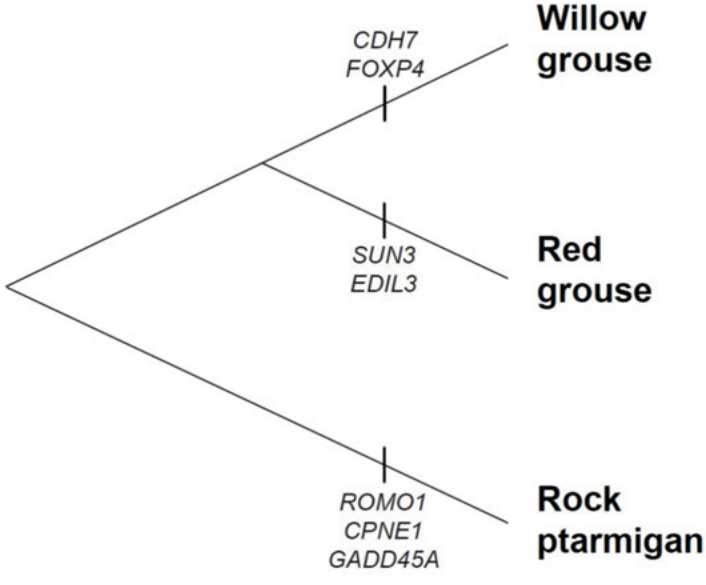

Figure 10. Genes under divergent selection in each of the three grouse, as revealed by the F_{ST} outlier test.

One of the involvement that the *CDH7* gene has been shown to have is in chick limb development (91). Thus, positive selection acting on this gene would most likely indicate a morphological adaptation. However, it has also been found in a host of other pathways [including neural, optic, brachial and olfactory ontogeny], and as such it has a very broad scope of effects (92, 93). Therefore pin-pointing the exact ramification of positive selection acting on this gene will require further research and insight.

Two of the genes selected for in the rock ptarmigan stand out as potential candidates for future research: *ROMO1* and *GADD45A*. Both are involved in response to stress. *ROMO1* is integral in the cellular response following reactive oxygen species [ROS] accumulation (94) and *GADD45A* is expressed following environmental stress (95). Out of the species studied here, the rock ptarmigan inhabits the most extreme environments, so genes that would help cope with the effects of increased stress are likely to be more important for this species than for the other studied grouse. However, experimental evidence for the direct function of these two genes in birds will have be established before their exact effects can be understood.

Taxon specific tests of neutrality found an overall strong signal of population expansion, wherein the genome-wide Tajima's *D* was highly negative and Fay and Wu's *H* was positive. This is in accordance with the results from *Papers III* and *IV*, in which the Scandinavian willow grouse, British red grouse and Greenland rock ptarmigan undergo bottlenecks of varying severity during the LGM and then upon the deglaciation of their habitat increase in population size up until the present.

Lastly, the F_{ST} outlier analysis also revealed an approximately 3.5Mb long region on chromosome 20, containing 30 genes, with considerably lower levels of differentiation between all three taxa, suggesting uniform selection is acting on the whole or parts of the genomic segment in these populations. Right at the 5' start of the region, the *Agouti* [*ASIP*] gene is located. In mice, this gene has been shown to directly affect the coat colouration by binding to the *MC1R* cell surface protein thereby causing an increase in the production of pheomelanin [yellow/red pigment] and down-regulating the production of eumelanin [brown/black pigment] (96, 97). In quail it has also been shown that the yellow phenotype is caused by a mutation > 90kb upstream of the Agouti gene, consequently changing its promoter (98). As discussed previously, pigmentation plays an integral role in the ecology of these grouse

species. Both the willow grouse and rock ptarmigan moult their brown summer plumage into a pure white plumage prior to winter in order to match the predominantly snowy environment. The red grouse on the other hand, living in the generally snow-free British Isles, forgoes this winter moult and instead retains its brown plumage year-round. Because the genomic scan did not show any differentiation in this region between the red grouse and the other two species, this hints to the involvement of *Agouti* in the strongly conserved brown summer plumage, rather than the differentiated white-winter plumage. However, the gene has also been found to be involved in regulation of lipid metabolism in mice and humans (99).This raises the other possibility that the regulation of *Agouti* could be involved in the adaptation to the colder, more open habitats and sedentary lifestyle seen in all three grouse. Again, more knowledge will have to be gathered in the avian systems before this issue can be fully resolved.

4. Conclusions and future perspectives

Over the last decade we have seen an ever expanding theoretical knowledge in population genetics and importantly, this was followed by the establishment of novel methods that allow the application of NGS data to test these theories. This thesis aimed to take advantage of these developments and apply these methodologies to investigate past population history as well speciation and adaptation dynamics in grouse.

In *Paper I* and *II*, we started out by providing a review of the applicability of methods that estimate contemporary N_e and demographic history to animal conservation and management. In the case of contemporary N_e estimation, we find that there is a large degree of variation in the underlying assumptions and the necessary amount of data required for accurate estimation between the methods. This can produce biased results and therefore cause wrong inferences to be made about the studied population. As such, greater care needs to be taken when applying these methods. Ideally, a more thorough examination by running various simulations across all these methods should establish the benchmark requirements that will provide accuracy and precision in estimating contemporary N_e. Furthermore, application of NGS techniques to this field should also be taken into account. The steady move away from microsatellites and into genome sequencing [either whole or reduced] of non-model organisms will provide further potential by the increased resolution that such numerous SNP markers offer (100, 101).

In the case of demographic history, the application of NGS techniques cannot be ignored. With better understanding of the genomic architecture, methods that estimate demographic history can draw upon more and more information. And the increasing number of available genomes and the lowering costs of sequencing means that multiple individuals of a species can be sequenced, thereby providing increased resolution across all time periods. However, the major issue with these methods lies with their interpretation. The observed fluctuations in N_e can be caused by not only changing climatic conditions, but as this thesis and others show (102, 103), demographic pro-

cesses, such as populations sub-structuring, colonisation events and migration will all leave traces in the N_e trajectory. Therefore, the challenge will soon become not about gaining more resolution, but delineating the individual effects of these processes and thereby determining the extent to which they shape the current populations.

In *Paper III* we applied the PSMC approach to study past population dynamics in three grouse species and found three periods within the last 3 million years that affected their demography. These were the early Pleistocene cooling [3 Mya - 0.9 Mya], mid-Brunhes event [~430 kya] and the last glacial period [110 - 10 kya]. This was the first paper to ever show grouse dynamics at such time scales. The population trajectories of each species during the last glacial period however did not go in line with the hypothesised climatic conditions; therefore we proposed that they depict the dynamics of localised populations from which the individuals were sampled.

Subsequently in *Paper IV*, we aimed to gain a deeper of understanding of this dynamic. We sequenced more individuals originating from different parts of the species' distribution range and in addition, we used ecological modelling to track the shifting grouse habitat throughout the LIG and LGM. We did indeed find that within the last glacial period, PSMC depicts local population dynamics, whereby the fringe grouse populations experienced severe bottlenecks – thus illustrating the recolonization of these habitats following their deglaciation. Importantly, we find that in populations that make up the centre of the species distribution, and thus did not undergo such colonisation events, the PSMC and ecological modelling results corroborate each other. Unfortunately, we were not able to get samples from every part of the grouse distribution. We only have rock ptarmigan from Greenland and Iceland, and as such, we lack any mainland population [North America or Eurasia] for comparison. Having samples from these parts would be ideal in studying how different regions responded to the climatic conditions during the last glacial period. With the same aim, it would be valuable to sequence a black grouse from the eastern part of its range as well. This would provide further insight into population subdivision and eventual recolonization history of the species across its range.

To follow on from the past population dynamics, in *Paper V* we investigated the genomic architecture underlying speciation and adaptation in three of the grouse. By performing an F_{ST} outlier test, we identified 2 highly differentiat-

ed regions in each taxon. Among them, these 6 regions contain 7 genes. Moreover, the analysis also revealed a 3.5-Mb long highly undifferentiated region on chromosome 20 - indicating uniform selection acting in this region across all three taxa.

One major limitation in understanding the speciation dynamics in greater details is the fact that 5 out of the 7 genes found in this study to be under divergent selection have no experimental evidence of their function in avian systems. This makes the interpretation of the effect of selection acting on them rather difficult. More work is needed in establishing gene expression patterns and applying functional assays such as knock-out / knock-down in the chicken at least, if not the grouse themselves, to better understand their roles.

Lastly, the extent of recombination within the genomes will have to be quantified. By studying the changes in recombination rate along chromosomes, further insight into the nature of selection acting upon particular regions can be gained (104, 105). If positive selection is indeed responsible for the outlier loci presented in this study, one would expect the suppression of recombination within the same and neighbouring outlier windows to follow. As such, this approach would highly complement the F_{ST} outlier analysis in trying to understand what regions are responsible for the speciation dynamics in these grouse taxa.

5. Sammanfattning på svenska

Ur ett ekologiskt perspektiv är kunskapen om arternas demografiska historia mycket värdefull eftersom fluktuationer i populationens storlek kan matchas till kända klimatförhållanden, som kan ge stor insikt i en arts reaktion på tidigare klimatförändringar. Detta i sin tur kan göra det möjligt för oss att förutsäga hur de skulle reagera på framtida klimatscenarier. Utvecklingen inom DNA-sekvensering och populationsgenetisk teori erbjuder nu stora möjligheter att studera demografisk historia och artbildningsprocesser hos olika organismer i större detalj. Denna avhandling använder fyra arter skogshöns (Aves, underfamilj Tetraoninae] för att undersöka den demografiska historien och artbildningen inom denna grupp av närbesläktade fåglar; dalripa [*Lagopus lagopus lagopus*], moripa [*Lagopus lagopus scotica*], fjällripa [*Lagopus muta*] och orre [*Tetrao tetrix*].

Jag och mina medförfattare, börjar med att gå igenom de olika metoder som används för att uppskatta nutida effektiv populationsstorlek [N_e] och demografisk historia och som är tillgängliga för praktiskt naturvårdsarbete (uppsats I och II). Vi finner att underliggande antaganden och nödvändiga data kan påverka sådana undersökningar. Därför uppmanar vi de som använder dessa metoder att vara försiktiga med valet av vilka metoder de använder. I ett försök att förbättra förståelsen av dessa metoder ger vi en sammanfattning av deras användbarhet.

Sedan använder jag hela genomen av orre, dalripa och fjällripa för att rekonstruera deras populationsdynamik under de senaste årmiljonerna (uppsats III). Jag hittade tre dominerande perioder som formar arternas demografiska historia: tidig Pleistocen nedkylning [3 - 0,9 miljoner år sedan], Mid-Brunhes händelsen [430 tusen år sedan, t.å.s.] och den sista istiden [110 - 10 t.å.s.]. Nedkylningen av jorden under Pleistocen orsakade att orrens antal minskade medan de köldanpassade dalripe och fjällripepopulationerna ökade i antal. Från 430 t.ås. till 130 t.å.s ser vi det motsatta mönstret - ökande orrpopulatiner medan dalripe och fjällripepopulationerna minskade. Varje art reagerade på olika sätt på den kalla temperaturen under den senaste istiden.

Sammantaget så fångade studien storskalig populationsdynamik på artnivå men också indikationer på starka signaler från lokala populationers historia, där återkolonisering och fragmenteringshändelser formar artens nuvarande demografi.

I den efterföljande studien undersöker jag ripdynamiken inom den senaste istiden i mer detalj genom att inkludera mer avlägsna prover och använda ekologisk modellering för att spåra ändringar i arternas utbredning (uppsats IV). Jag hittar här starka signaler från de lokala populationernas historia, där marginella populationer [Storbritannien, Skandinavien, Island, Grönland och Alaska] genomgår kraftiga flaskhalsar. Dessa kommer sig av återkolonisering av områden efter slutet av den senaste istiden. I populationer vars utbredning inte genomgått så omfattande nedisningar, ser vi inte dessa. Dessutom studerar jag också hur framtida klimatförändringar förväntas förändra utbredningen av dessa arter.

Slutligen använder jag helgenomsekvensering av flera individer av dalripa, moripa och fjällripa för att studera artbildningsprocessen i deras genom (uppsats V). Jag upptäckte sex starkt differentierade regioner, innehållande sju gener, som sannolikt kan förklara de fenotypiska skillnaderna mellan dessa arter. Jag hittar också en region med låg differentiering, som innehåller Agouti, en pigment gen, vilket indikerar dess bevarande roll i skogshönsens fjäderdräktsfärg.

6. Acknowledgements

First and foremost, I'd like to acknowledge the grouse that were used for this thesis. I can only hope that this work is useful enough to justify the sacrifice.

And now the people.
Jacob - Not a single sentence would have been written if it wasn't for you. You have been unbelievably caring and gave me the opportunity to learn new things and shape the work as we both see fit. For all these years, I am forever thankful to you. David Sumpter, as my second supervisor, you have also played a large part in making the whole PhD possible, for which I am very grateful.

I would also like to thank all my collaborators and folks who helped with the work presented here. Kristinn and Páll for approaching us and providing the data that took the work beyond where we thought possible. Blas and Jens-Christian for being such good hosts in Aarhus and letting me know all about SDM's. Biao, even though technically we did not write any papers together for this thesis, it is because of all your previous work that I was able to just get started. Ludo, buddy… You have done so much for me that it would be a never-ending list of thank-you's. You are a true friend and I hope I get the chance to repay all the favours! Brian … not only have you read over manuscripts and significantly improved them, but you have also introduced me to things such as Yu Xiang aubergine and Barbarella, which I reckon is even more valuable! Thank you for everything! Ahmed, you were always happy to help me with every single silly question I had about bioinformatics and for that I am deeply grateful! And although I did not have as much lab work as I initially thought, I am forever thankful to Gunilla and Reija for teaching me how to do things properly and not destroy the whole lab! I am also very grateful to the whole Höglund-Laurila lab group for the countless stimulating discussions and positive feedback on my work.

Lastly a huge thank you … to all the lovely folks at Animal Ecology, all my funky and groovy friends at EBC, Uppsala and around the world, my parents and my family who are just the best and to my sweet & sexy Sardinian. Giulia, you have shown unwavering confidence in me and have made me a better human being… You are my sunshine!

7. References

1. Lomolino MV, Riddle BR, Brown JH (2006) *Biogeography* (Sinauer Associates, Inc. Publishers). 3rd Ed.
2. Clark P, Mix A (2002) Ice sheets and sea level of the Last Glacial Maximum. *Quat Sci Rev* 21:1–7.
3. Lawing AM, Polly PD (2011) Pleistocene climate, phylogeny, and climate envelope models: An integrative approach to better understand species' response to climate change. *PLoS One* 6(12). doi:10.1371/journal.pone.0028554.
4. Stewart JR, Lister AM, Barnes I, Dalén L (2010) Refugia revisited: individualistic responses of species in space and time. *Proc Biol Sci* 277(1682):661–671.
5. Moura AE, et al. (2014) Killer whale nuclear genome and mtDNA reveal widespread population bottleneck during the last glacial maximum. *Mol Biol Evol* 31(5):1121–1231.
6. Smith SE, Gregory RD, Anderson BJ, Thomas CD (2013) The past, present and potential future distributions of cold-adapted bird species. *Divers Distrib* 19(3):352–362.
7. Kendall KC, et al. (2009) Demography and Genetic Structure of a Recovering Grizzly Bear Population. *J Wildl Manage* 73(1):3–17.
8. U.S. Fish and Wildlife Service (1993) *Grizzly bear recovery plan* (Missoula, Montana, USA).
9. Wright S (1931) Evolution in Mendelian populations. *Genetics* 16(97):97–159.
10. Charlesworth B (2009) Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. *Nat Rev Genet* 10(3):195–205.
11. Hogg JT, Forbes SH, Steele BM, Luikart G (2006) Genetic rescue of an insular population of large mammals. *Proc Biol Sci* 273(February):1491–1499.
12. Pimm SL, Dollar L, Bass OL (2006) The genetic rescue of the Florida panther. *Anim Conserv* 9:115–122.
13. Johnson JA, Dunn PO (2006) Low genetic variation in the Heath Hen prior to extinction and implications for the conservation of prairie-chicken populations. *Conserv Genet* 7:37–48.
14. Johnson JA, et al. (2009) Long-term survival despite low genetic diversity in the critically endangered Madagascar fish-eagle. *Mol Ecol* 18:54–63.
15. Groombridge JJ, Jones CG, Bruford MW, Nichols R a (2000) Conservation biology: "Ghost" alleles of the Mauritius kestrel. *Nature* 403(6770):616–616.
16. Lanfear R, Kokko H, Eyre-Walker A (2014) Population size and the rate of evolution. *Trends Ecol Evol* 29(1):33–41.
17. Mele M, et al. (2012) Recombination Gives a New Insight in the Effective Population Size and the History of the Old World Human Populations. *Mol Biol Evol* 29(1):25–30.

18. Park L (2011) Effective population size of current human population. *Genet Res (Camb)* 93(02):105–114.
19. Li H, Durbin R (2011) Inference of human population history from individual whole-genome sequences. *Nature* 475(7357):493–6.
20. Reich DE, et al. (2001) Linkage disequilibrium in the human genome. *Nature* 411(6834):199–204.
21. Schiffels S, Durbin R (2014) Inferring human population size and separation history from multiple genome sequences. *Nat Genet* 46(8):919–925.
22. MacLeod IM, Larkin DM, Lewin HA, Hayes BJ, Goddard ME (2013) Inferring demography from runs of homozygosity in whole-genome sequence, with correction for sequence errors. *Mol Biol Evol* 30(9):2209–23.
23. Hung C-M, et al. (2014) Drastic population fluctuations explain the rapid extinction of the passenger pigeon. *Proc Natl Acad Sci U S A* 111(29):10636–10641.
24. Palkopoulou E, et al. (2015) *Complete Genomes Reveal Signatures of Demographic and Genetic Declines in the Woolly Mammoth* doi:10.1016/j.cub.2015.04.007.
25. Palstra FP, Ruzzante DE (2008) Genetic estimates of contemporary effective population size: what can they tell us about the importance of genetic stochasticity for wild population persistence? *Mol Ecol* 17(15):3428–3447.
26. Hare MP, et al. (2011) Understanding and Estimating Effective Population Size for Practical Application in Marine Species Management. *Conserv Biol* 25(3):438–449.
27. Fagan WF, Holmes EE (2006) Quantifying the extinction vortex. *Ecol Lett* 9(1):51–60.
28. Höglund J (2009) *Evolutionary Conservation Genetics* (Oxford University Press).
29. Mitton JB (2002) Adaptation and Natural Selection: Overview. *Encyclopedia of Life Sciences* (John Wiley & Sons, Ltd, Chichester, UK). doi:10.1038/npg.els.0001706.
30. Futuyma DJ (2009) *Evolution* (Sinauer Associates, Inc. Publishers). Second edi.
31. Stapley J, et al. (2010) Adaptation genomics: the next generation. *Trends Ecol Evol* 25(12):705–712.
32. Gutierrez RJ, Barrowclough GF, Groth JG (2000) A classification of the grouse (Aves: Tetroninae) based on mitochondrial DNA sequences. *Wildlife Biol* 6(4):205–212.
33. Drovetski S (2003) Plio-Pleistocene climatic oscilations, Holarctic biogeography and speciation in an avian subfamily. *J Biogeogr* 30(1989):1173–1181.
34. Höglund J, Wang B, Axelsson T, Quintela M (2013) Phylogeography of willow grouse (Lagopus lagopus) in the Arctic: taxonomic discordance as inferred from molecular data. *Biol J Linn Soc* 110(1):77–90.
35. Storch I (2006) *Grouse Status Survey and Conservation Action Plan 2006-2010* (Gland, Switzerland).
36. Ward I, Larcombe P, Lillie M (2006) The dating of Doggerland – post-glacial geochronology of the southern North Sea. *Environ Archaeol* 11(2):207–218.
37. Quintela M, Berlin S, Wang B, Höglund J (2010) Genetic diversity and differentiation among Lagopus lagopus populations in Scandinavia and Scotland: evolutionary significant units confirmed by SNP markers. *Mol Ecol* 19(12):2380–93.

38. McMahon BJ, Johansson MP, Piertney SB, Buckley K, Höglund J (2012) Genetic variation among endangered Irish red grouse (Lagopus lagopus hibernicus) populations: implications for conservation and management. *Conserv Genet* 13(3):639–647.
39. Watson A, Moss R, Rae S (1998) Population dtnamics of Scottish rock ptarmigan cycles. *Ecology* 79(4):1174–1192.
40. Storch I (2000) Conservation Status and Threats to Grouse Worldwide: An Overview. *Wildlife Biol* 6(4):195–204.
41. Kozma R, Melsted P, Magnússon KP, Höglund J (2016) Looking into the past - the reaction of three grouse species to climate change over the last million years using whole genome sequences. *Mol Ecol* 25(2):570–580.
42. Alatalo R V., Höglund J, Lundberg A, Sutherland WJ (1992) Evolution of black grouse leks: female preferences benefit males in larger leks. *Behav Ecol* 3(1):53–59.
43. Höglund J, Alatalo R V., Lundberg A (1990) Copying the mate choice of others? Observations on female black grouse. *Behaviour* 114(1):221–231.
44. Holder K, Montgomerie R (1993) Context and consequences of comb displays by male rock ptarmigan. *Anim Behav* (45):457–470.
45. Hannon SJ, Dobush G (1997) Pairing status of male willow ptarmigan: is polygyny costly to males? *Anim Behav* (53):369–380.
46. Wang B, Ekblom R, Bunikis I, Siitari H, Höglund J (2014) Whole genome sequencing of the black grouse (Tetrao tetrix): reference guided assembly suggests faster-Z and MHC evolution. *BMC Genomics* 15:180.
47. Liu Y, Schröder J, Schmidt B (2013) Musket: a multistage k-mer spectrum-based error corrector for Illumina sequence data. *Bioinformatics* 29(3):308–15.
48. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114–2120.
49. Luo R, et al. (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* 1(1):18.
50. Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv Prepr arXiv13033997* 00(00):3.
51. McKenna A, et al. (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20(9):1297–303.
52. DePristo MA, et al. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43(5):491–8.
53. Watterson GA (1975) On the number of segregating sites in genetical models without recombination. *Theor Popul Biol* 7(2):256–276.
54. Bachtrog D, et al. (2011) Are all sex chromosomes created equal? *Trends Genet* 27(9):350–357.
55. Svenning J-C, Fløjgaard C, Marske K a., Nógues-Bravo D, Normand S (2011) Applications of species distribution modeling to paleobiology. *Quat Sci Rev* 30(21-22):2930–2947.
56. BirdLife International, NatureServe (2014) *Bird species distribution maps of the world* (BirdLife International, Cambridge, UK and NatureServe, Arlington, USA.).
57. Korneliussen T, Albrechtsen A, Nielsen R (2014) ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* 15(1):356.
58. Zhang Z, Schwartz S, Wagner L, Miller W (2000) A Greedy Algorithm for Aligning DNA Sequences. *J Comput Biol* 7(12):203–214.

59. Fay JC, Wu CI (2000) Hitchhiking under positive Darwinian selection. *Genetics* 155(3):1405–13.
60. Fumagalli M, Vieira FG, Linderoth T, Nielsen R (2014) NgsTools: Methods for population genetics analyses from next-generation sequencing data. *Bioinformatics* 30(10):1486–1487.
61. Stamatakis A (2014) RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9):1312–1313.
62. Waples RS (1989) A generalized approach for estimating effective population size from temporal changes in allele frequency. *Genetics* 121(2):379–91.
63. Nei M, Tajima F (1981) Genetic drift and estimation of effective population size. *Genetics* 98:625–640.
64. Jorde PE, Ryman N (1995) Temporal Allele Frequency Change and Estimation of Effective Size in Populations With Overlapping Generations. *Genetics* 139:1077–1090.
65. Waples RS, Yokota M (2007) Temporal estimates of effective population size in species with overlapping generations. *Genetics* 175(1):219–33.
66. Wang JL, Whitlock MC (2003) Estimating effective population size and migration rates from genetic samples over space and time. *Genetics* 163(January):429–446.
67. Pudovkin A, Zaykin D, Hedgecock D (1996) On the potential for estimating the effective number of breeders from heterozygote-excess in progeny. *Genetics* 387:0–4.
68. Hill W (1981) Estimation of effective population size from data on linkage disequilibrium. *Genet Res* 38:209–216.
69. Wang J, Brekke P, Huchard E, Knapp LA, Cowlishaw G (2010) Estimation of parameters of inbreeding and genetic drift in populations with overlapping generations. *Evolution (N Y)* 64(6):1704–1718.
70. Tallmon DA, Koyuk A, Luikart G, Beaumont MA (2008) COMPUTER PROGRAMS: onesamp: a program to estimate effective population size using approximate Bayesian computation. *Mol Ecol Resour* 8(2):299–301.
71. Storz J, Beaumont M (2002) Testing for genetic evidence of population expansion and contraction: an empirical analysis of microsatellite DNA variation using a hierarchical Bayesian model. *Evolution (N Y)* 56(1):154–166.
72. Beaumont M (1999) Detecting population expansion and decline using microsatellites. *Genetics* 153:2013–2029.
73. Goossens B, et al. (2006) Genetic signature of anthropogenic population collapse in orang-utans. *PLoS Biol* 4(2):e25.
74. Heled J, Drummond AJ (2008) Bayesian inference of population size history from multiple loci. *BMC Evol Biol* 8:289.
75. Trucchi E, et al. (2014) King penguin demography since the last glaciation inferred from genome-wide data. *Proc Biol Sci* 281(1787):20140528.
76. Davey JW, et al. (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat Rev Genet* 12(7):499–510.
77. Zhao S, et al. (2013) Whole-genome sequencing of giant pandas provides insights into demographic history and local adaptation. *Nat Genet* 45(1):67–71.
78. Augustin L, et al. (2004) Eight glacial cycles from an Antarctic ice core. *Nature* 429(6992):623–628.
79. Lisiecki LE, Raymo ME (2005) A Pliocene-Pleistocene stack of 57 globally distributed benthic δ 18 O records. *Paleoceanography* 20(1):PA1003.
80. Zachos JC, Dickens GR, Zeebe RE (2008) An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. *Nature* 451(January):279–283.

81. Clark PU, et al. (2009) The Last Glacial Maximum. *Science (80-)* 325(5941):710–714.
82. Svendsen J, et al. (2004) Late Quaternary ice sheet history of northern Eurasia. *Quat Sci Rev* 23(11-13):1229–1271.
83. Holder K, Montgomerie R, Friesen V (1999) A test of the glacial refugium hypothesis using patterns of mitochondrial and nuclear DNA sequence variation in rock ptarmign (Lagopus mutus). *Evolution (N Y)* 53(6):1936–1950.
84. Corrales C, Pavlovska M, Höglund J (2013) Phylogeography and subspecies status of Black Grouse. *J Ornithol* 155(1):13–25.
85. Holder K, Montgomerie R, Friesen VL (2000) Glacial vicariance and historical biogeography of rock ptarmigan (Lagopus mutus) in the Bering region. *Mol Ecol* 9(9):1265–1278.
86. Kukla GJ, et al. (2002) Last Interglacial Climates. *Quat Res* 58(1):2–13.
87. Rousso DL, et al. (2012) Foxp-Mediated Suppression of N-Cadherin Regulates Neuroepithelial Character and Progenitor Maintenance in the CNS. *Neuron* 74:314–330.
88. Miller JE, et al. (2008) Birdsong decreases protein levels of FoxP2, a molecule required for human speech. *J Neurophysiol* 100(4):2015–25.
89. Scharff C, Haesler S (2005) An evolutionary perspective on FoxP2: strictly for the birds? *Curr Opin Neurobiol* 15(6):694–703.
90. Mendoza E, et al. (2015) Differential coexpression of FoxP1, FoxP2, and FoxP4 in the Zebra Finch (Taeniopygia guttata) song system. *J Comp Neurol* 523(9):1318–1340.
91. Kim D, Kang S-S, Jin E-J (2009) Alterations in the temporal expression and function of cadherin-7 inhibit cell migration and condensation during chondrogenesis of chick limb mesenchymal cells in vitro. *J Cell Physiol* 221(1):161–170.
92. Lalani SR, et al. (2006) Spectrum of CHD7 mutations in 110 individuals with CHARGE syndrome and genotype-phenotype correlation. *Am J Hum Genet* 78(2):303–14.
93. Aramaki M, et al. (2007) Embryonic expression profile of chicken CHD7, the ortholog of the causative gene for CHARGE syndrome. *Birth Defects Res Part A Clin Mol Teratol* 79(1):50–57.
94. Chung YM, et al. (2008) Replicative Senescence Induced by Romo1-derived Reactive Oxygen Species. *J Biol Chem* 283(48):33763–33771.
95. Takekawa M, et al. (1998) A Family of Stress-Inducible GADD45-like Proteins Mediate Activation of the Stress-Responsive MTK1/MEKK4 MAPKKK. *Cell* 95(4):521–530.
96. Graham A, Wakamatsu K, Hunt G, Ito S, Thody AJ (1997) Agouti protein inhibits the production of eumelanin and phaeomelanin in the presence and absence of alpha-melanocyte stimulating hormone. *Pigment Cell Res* 10(5):298–303.
97. Steiner CC, Weber JN, Hoekstra HE (2007) Adaptive variation in beach mice produced by two interacting pigmentation genes. *PLoS Biol* 5(9):e219.
98. Nadeau NJ, et al. (2008) Characterization of Japanese quail yellow as a genomic deletion upstream of the avian homolog of the mammalian ASIP (agouti) gene. *Genetics* 178(2):777–86.
99. Xue B, Moustaid-N N, Wilkison WO, Zemel MB (1998) The agouti gene product inhibits lipolysis in human adipocytes via a Ca2+-dependent mechanism. *FASEB J* 12(13):1391–6.
100. Ekblom R, Galindo J (2010) Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity (Edinb)* 107(1):1–15.

101. Shafer AB a., et al. (2015) Genomics and the challenging translation into conservation practice. *Trends Ecol Evol* 30(2):78–87.
102. Mazet O, Rodríguez W, Grusea S, Boitard S, Chikhi L (2016) On the importance of being structured: instantaneous coalescence rates and human evolution—lessons for ancestral population size inference? *Heredity (Edinb)* 116(4):362–371.
103. Mazet O, Rodríguez W, Chikhi L (2015) Demographic inference using genetic data from a single individual: Separating population size variation from population structure. *Theor Popul Biol* 104(104):46–58.
104. Burri R, et al. (2015) Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of Ficedula flycatchers. *Genome Res* 25(11):1656–1665.
105. Nosil P, Feder JL (2012) Genomic divergence during speciation: causes and consequences. *Philos Trans R Soc B Biol Sci* 367(1587):332–342.

Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology 1391

Editor: The Dean of the Faculty of Science and Technology

A doctoral dissertation from the Faculty of Science and Technology, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology. (Prior to January, 2005, the series was published under the title "Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology".)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-299926

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2016