Applying emerging genetic methods to wild model systems

Ida-Maria Blåhed
Faculty of Forest Sciences
Department of Wildlife, Fish, and Environmental Studies
Umeå

Doctoral thesis
Swedish University of Agricultural Sciences
Umeå 2019

Acta Universitatis Agriculturae Sueciae
2019:40

Cover: “The challenges of quantifying nature”
(illustration: Sven Norman)

ISSN 1652-6880
ISBN (print version) 978-91-7760-398-6
ISBN (electronic version) 978-91-7760-399-3
© 2019 Ida-Maria Blåhed, Umeå
Print: Arkitektkopia, Umeå 2019

Applying emerging genetic methods to wild model systems

Abstract

Genetic material is an invaluable source of information for assessments of wild populations. By using information derived from genetic markers, individuals can be identified and tracked over time, enabling studies of a wide range of behavioral, ecological, and evolutionary processes. At the population level, estimations of the distribution of genetic variation increase our knowledge about a wide range of population processes and can reveal barriers to gene flow, information that is important to integrate in both conservation- and management plans.

This thesis presents new methods for individual- and population based assessments of deer (Cervidae), with a focus on moose (*Alces alces*). Molecular markers, SNPs (single nucleotide polymorphisms), were developed using two different approaches. First, cross-species amplification was explored between cattle and the five deer species occurring in Sweden; moose, roe deer (*Capreolus capreolus*), red deer (*Cervus elaphus*), fallow deer (*Dama dama*) and reindeer (*Rangifer tarandus*). Next, a reduced representation sequencing approach was chosen for *de-novo* SNP discovery in moose, with the main purpose of finding markers suitable for individual identification. As a result, a moose SNP panel including 86 autosomal, five sex-specific- and five species diagnostic SNPs was developed. Both approaches resulted in SNPs useful for both individual- and population level applications. SNP genotyping was subsequently applied for assessments of the Swedish moose population by conducting a non-invasive capture-mark-recapture study and to explore spatio-temporal genetic patterns in a population of seasonally migratory moose.

These studies show that non-invasive SNP genotyping is useful for estimations of population size and sex-ratio, while also producing information about population structure and genetic variation. Furthermore, by combining genetic- and movement data over time, temporal spatial genetic structures were detected. These temporal structures elucidate different components of moose behavior, such as fidelity to seasonal ranges, which to some extent is maintained between generations. Consequently, the genetic information retrieved in this thesis clearly illustrate the potential of SNP genotyping for assessments of population- and behavioral processes in deer, including non-invasive monitoring.

Keywords: wildlife genetics, genomics, SNP, genotyping, deer, *Alces alces*, cross-species amplification, population assessment, non-invasive, population structure, capture-mark-recapture

Author's address: Ida-Maria Blåhed, SLU, Department of Wildlife, Fish, and Environmental Studies, SE-901 83 Umeå, Sweden

Dedication

To Einar and Emmy

Continuous improvement is better than delayed perfection
Mark Twain

Contents

List of publications

This thesis is based on the work contained in the following papers, referred to by Roman numerals in the text:

I Blåhed, I-M.*, Jansson, J. and Spong G. Identification of informative SNPs in five deer species by cross-species amplification (manuscript)

II Blåhed, I-M.*, Königsson, H., Ericsson, G. and Spong, G. (2018). Discovery of SNPs for individual identification by reduced representation sequencing of moose (*Alces alces*). *Plos One 13 doi:10.1371/ journal. pone.0197364*

III Blåhed, I-M.*, Ericsson, G. and Spong, G. Non-invasive population assessment of moose (*Alces alces*) by SNP genotyping of fecal pellets (under revision)

IV Blåhed, I-M.*, Singh, N. J., Ericsson, G., Norman, A. J. and Spong, G. Spatio-temporal genetic patterns in a large migratory mammal (manuscript)

Paper II is reproduced with the permission of the publisher.

* Corresponding author.

The contribution of Ida-Maria Blåhed (IMB) to the papers included in this thesis was as follows:

I IMB was largely responsible for the analyses and was responsible for writing the manuscript.

II IMB contributed to the experimental design and laboratory work, performed the bioinformatics, analyses and was responsible for writing the manuscript.

III IMB contributed to the experimental design, performed the analyses and was responsible for writing the manuscript.

IV IMB contributed to the experimental design, performed the analyses and was responsible for writing the manuscript.

1 Introduction

1.1 Wildlife genetics[1]

Today, DNA-based information is fundamental for assessments of wild populations. By analyzing and comparing genetic information; kinship and genetic diversity can be estimated and used for a wide range of inferences about individual and population processes. While some of these estimates only can be made with genetic information, parameters such as population size and sex-ratio are more or less challenging to quantify in wild populations using non-genetic approaches. This is mainly explained by elusiveness of many wild animals, often in combination with inaccessible terrain. Moreover, by using modern molecular techniques, genomic information can be obtained from *non-invasive* samples such as feces, hair or saliva. By non-invasive sampling, retrieval of DNA for subsequent analyses is made possible without disturbing or harming the studied animals. This approach also leads to a reduction in both cost and time spent on sampling, i.e. more samples can potentially be obtained with the same amount of resources. Thus, with improving genetic methods, important knowledge about wild populations is generated that can be implemented in conservation and management plans.

1. I have chosen to throughout the thesis use the term “genetics” although “genomics” could be argued as more appropriate. However, “wildlife genetics” is an established term for studies of wild animals using their DNA as a source of information.

1.2 SNP development

1.2.1 A rapidly evolving field

Recent years' increased use of genetic applications in studies of wild populations has been tightly linked to an extremely rapid decrease in DNA sequencing costs. This is explained by the introduction of "next-generation high throughput sequencing" in 2008, which was a technical revolution in the field of genomics. Largely driven by competition between sequencing companies to offer affordable human whole-genome sequencing, the cost for whole genome sequencing has decreased at a high rate for many years. Today, a whole human-sized genome can be sequenced for 10% of the cost in 2011 when this PhD project was initiated (The National Human Genome Research Institute). Hence, the genomic toolbox has become available also for non-model species and research projects with smaller budgets. The advances within the field of DNA sequencing and the subsequent possibility to generate large quantities of genomic data, has allowed the development and use of sophisticated molecular markers, such as single nucleotide polymorphisms (SNPs), also for studies of non-model species (Garvin et al. 2010; Seeb et al. 2011).

1.2.2 SNP discovery

SNPs are single base pair positions that hold variation between individuals, and thus constitute a direct measure of genetic variation in a population. For a variation to be classified as a SNP, the least frequent allele typically has to occur in more than 1% of the population. Due to their co-dominant and primarily bi-allelic nature, interpretations of SNP-calls are relatively straightforward, which make SNP based studies more easily repeatable (Brookes 1999; Garvin et al. 2010). Moreover, SNPs have a wide genomic distribution and can therefore be selected to represent both coding- and non-coding genomic regions as well as linked or non-linked loci. Consequently, SNP-based methods have a wide application range. Nevertheless, there is an extensive development phase preceding the stage when genotypes can be retrieved from the population of interest. Depending on the intended application, different approaches for SNP discovery may be suitable (Garvin et al. 2010). The quality criteria for informative SNPs depend largely on the intended use. For the purpose of identification and separating individuals, both alleles of the SNP should be relatively common among the studied individuals, i.e. SNPs with high minor allele frequencies are desired. If on the other hand, phylogenetic studies or colonisation patterns are of interest, also low frequency alleles may contribute

valuable information. A challenge related to development of SNP panels, regardless of the method, is to capture the distribution of allelic variation in the population of interest. If this fails, the panel will suffer from ascertainment bias (Nielsen and Signorovitch 2003). To avoid such bias it is important to include a sufficient number of individuals, from a geographic distribution that is representative for the intended study. However, biases can arise at different stages of the discovery process and the type of bias and the consequences for downstream analyses will vary between SNP discovery approaches and applications.

The approach taken to identify SNPs in a non-model species depends largely on the time, money and technology available. The large quantity of sequencing data generated has to be quality filtered and screened for putative SNPs, requiring a substantial computational effort. Furthermore, although producing genomic information has become less expensive, the total cost for a complete *de novo* SNP development project can still be substantial.

Cross-species amplification

A short-cut for identifying SNPs in a non-model species is to use commercial SNP panels developed for closely related species (Haynes and Latch 2012; Kaminski et al. 2012; Ogden et al. 2012; Sacks and Louie 2008). Since most research projects have limited budgets, time and sometimes also limited bioinformatics experience, cross-species amplification can be an interesting alternative to *de-novo* sequencing.

Depending on the phylogenetic distance between the target species and the species that the SNPs were developed for, different success rates are to be expected. Normally, shorter divergence time means fewer mismatches between target and source sequences and higher likelihood that the amplification will be successful. Miller et al. (2012) found that less than 5% of amplified loci can be expected to have retained polymorphism when the divergence time between the target and the "source" species is more than three million years. Yet, a few percentage of the large numbers of SNPs included in high-density commercial SNP panels are often enough to allow SNP-based inferences about individuals or populations. Moreover, a specific type of ascertainment bias that is associated with cross-species amplification is the targeting of SNPs in highly conserved genomic regions. This may result in a lack of more recently emerged variations, which in particular can limit the usefulness of the SNPs for phylogenetic inferences (Malhi et al. 2011).

De-novo reduced representation sequencing

Species-specific SNPs that are customized for the considered applications can be obtained by high throughput *de-novo* sequencing. Using this approach, the detection of high quality SNPs is performed in a subset of individuals that represent the population or species of interest well. Again, limiting the level of ascertainment bias is of importance in the SNP discovery stage for accurate inferences of allele frequency distributions in downstream analyses. In order to distinguish true variable nucleotides from sequencing errors, a high read depth (many replicates of the same fragment) is required. Since the total amount of data that can be sequenced simultaneously is limited, there is a trade-off between the number of included individuals and the amount of data from each individual. Included in this trade-off is also the depth and breadth of coverage for each individual. A widely adopted approach for sequencing with a sufficient read depth and genomic coverage while decreasing the total quantity of data is the construction of reduced representation libraries (RRL). This way, novel SNPs can be discovered also in non-model species without an available reference genome (Miller et al. 2007). Prior to RRL high-throughput sequencing, the DNA is digested by one or several restriction enzymes to generate unassociated fragments of DNA of different length. By selecting restriction fragments within a given size range, the quality of the sequencing can be improved, both by increasing the read depth and by reducing bias related to certain fragment lengths. For increased efficiency, samples from different individuals can be individually labelled with barcodes and pooled together during sequencing (Andrews et al. 2016; Baird et al. 2008; Davey et al. 2011). The fragments of DNA produced by sequencing, can, in the absence of a reference genome, be assembled into "stacks" of matching fragments, in which SNPs can be detected.

1.3 Wildlife SNP genotyping

During the last ten years, SNP genotyping has been developed and applied to a wide range of wild, non-model species for conservation-, monitoring, and research purposes (Attard et al. 2018; Harris and Munshi-South 2017; Norman et al. 2013; Ogden et al. 2013; Soderquist et al. 2017; Spitzer et al. 2016) as well as forensics applications (Bylemans et al. 2016; Ogden 2011). The low genotyping error rates and reproducible results associated with SNP genotyping, are features that are desired in all research. Another important factor explaining why many studies of wild animals have chosen SNP genotyping, in particular microfluidic array technology[2], is that this method requires very short DNA

2. A technology that enable PCR reactions in nanoliter volumes.

fragments and has been found to generate reliable genotypes also for low target DNA quantities, which often is true for non-invasively collected DNA (Kraus et al. 2015; Norman and Spong 2015; von Thaden et al. 2017).

1.3.1 Allele frequency distributions

Maintenance of genetic variation is a prerequisite for populations to stay viable and to be able to adapt to changing conditions over time. Hence, measures of genetic variation across individuals, populations or species is fundamental in wildlife genetic assessments. Deviations from expected patterns of allele frequencies, based on evolutionary null models (see Methods section 2.6), reflect that evolutionary processes might have been, and perhaps still are, acting on the distribution of genetic variation among the studied individuals. The population history and behavior aspects of a species are examples of such factors that may distort the distribution of alleles from expectations.

Individual level

By comparing the allele frequency distribution of a set of loci, e.g., a SNP panel, between pairs of individual genotypes, estimations of relatedness can be obtained. Information about kinship is essential for studies of reproductive behavior and can also be an explanatory factor of other behavioral mechanisms (Randall et al. 2007; Shafer et al. 2012; Weinman et al. 2015). Based on relatedness estimates from a large proportion of a population, pedigrees can be reconstructed (Spitzer et al. 2016) which among other things enable estimations of dispersal rates (Norman and Spong 2015) and reproductive success in wild populations. Moreover, pairwise comparisons of allele frequencies allow estimations of genetic distance between individuals, which is informative by themselves or as variables in multivariate analyses.

Population level

One of the most common applications of wildlife genotyping is the exploration of genetic differentiation between groups, i.e. population structuring. For studies of population differentiation, 30-80 SNPs have been proposed as sufficient, depending on the level of differentiation targeted (Morin et al. 2009; Turakulov and Easteal 2003). Substructuring of a population occurs when groups of individuals are physically separated with limited gene flow in between them. Local adaptation, differences in demography, or drift, may then cause the allele frequency distributions to diverge. SNP-based studies of genetic differentiation in wild animals have among other things investigated effects of anthropogenic

disturbance on the genetic diversity of species (Bourgeois et al. 2018; Sovic et al. 2016), interbreeding between farmed and wild animals (Soderquist et al. 2017; Thirstrup et al. 2015), and adaptation of species to changing environments (Harrisson et al. 2017). Moreover, behavioral mechanisms, such as non-random mating and fidelity to areas for reproduction, also influence the distribution of genetic variation, and can give rise to population structure. Such effects on population structure have been possible to expose in wild populations using SNP genotyping (Moore et al. 2017; Radersma et al. 2017).

1.3.2 Individual identification

With an increased number of markers included for genotyping, higher precision can be achieved in terms of separating samples with a high proportion of shared alleles (i.e. close relatives). Even though single SNPs have a much lower discriminatory power compared to other molecular markers such as microsatellites[3], highly accurate individual identification is achieved with a sufficient number of SNPs (2-3 times as many SNPs as microsatellites) (Fernandez et al. 2013; von Thaden et al. 2017). Retrieval of reliable individual genotypes enable genetic "tagging" and tracing of individuals on ecologically relevant time- and spatial scales (Palsboll 1999). For example, population demographic parameters can be obtained by genetic capture-mark-recapture based on individual genotypes obtained from non-invasive samples (Lampa et al. 2015; Petit and Valiere 2006). Consequently, individual based methods constitute a shift from traditional population genetic methods that measure processes over evolutionary time-scales, towards allowing studies of e.g., gene flow and population differentiation over more contemporary time-scales (Jombart et al. 2008; Palsbøll et al. 2013).

1.4 Study species

1.4.1 Deer in Sweden

There are currently four wild species of deer occurring in Sweden, of which three are regarded as native; moose (*Alces alces*), roe deer (*Capreolus capreolus*) and red deer (*Cervus elaphus*). Moose colonised the country after the last glacial period about 10 000 years ago while the colonisation by roe deer and red deer took place 1000 years later. All native deer colonized Scandinavia from the south

3. Microsatellites are repetitive sequences found across the genome and associated with a relatively high mutation rate, leading to a high allelic diversity.

via Denmark. Fallow deer (*Dama dama*) were introduced in the sixteenth century but were mainly held in fenced areas until the nineteenth century, when escaped or released animals started to form wild populations in the south of Sweden (Liberg 2010). A fifth deer species; reindeer (*Rangifer tarandus*), is semi-domesticated in Sweden.

The wild deer populations in Sweden are at present viable with stable or increasing population trends. As for many other wild populations, precise population size estimates are difficult to obtain and hence not available for the different deer in Sweden.

Current distribution ranges differ between the species, the moose has a wide distribution range and can be found throughout Sweden except for on the island of Gotland. Although the densities are highest in southern areas, roe deer occur in all parts of the country except for in alpine areas. The major part of the red deer population occurs in southern forest areas, whereas fallow deer is exclusively found in the south of Sweden. Both red deer and fallow deer are slowly expanding their distribution ranges. Reindeer occur in both forest- and mountain areas in the northern parts of Sweden. As a generalization, all four wild deer co-exist in large parts of southern Sweden, while moose, roe deer and locally also reindeer are sympatric further north.

The diets of the different deer species are overlapping to various degrees; moose are primarily browsers of woody plant parts while the other deer species' diets are mixed with, or dominated by, grass, forbs and leaves.

Thus, potential areas of application for SNPs informative about individual identity and/or allele frequency distributions are studies of browsing behavior, resource competition and estimations of population size by non-invasive capture-mark-recapture.

Moose in focus

The moose is the largest deer (Cervidae) species and is widely distributed throughout the boreal forests of the northern hemisphere. The moose occurring in Sweden, *A. a. alces*, is the same variety that is found in Europe and western Siberia whereas other varieties of moose; *A. alces* spp, occur in North America, eastern Asia and eastern Siberia. The moose in general is considered as a species of "least concern" in the IUCN red list of threatened species (Hundertmark 2016).

Scandinavian moose has been found to harbor a lower genetic variation, and is also differentiated from, Finnish and eastern European moose (Niedzialkowska et al. 2016). These patterns have been suggested to be a result of different colonization patterns, population fluctuations and limited gene flow due to the Baltic Sea. A relatively recent genetic evaluation of the Swedish

moose population found two major genetic groups and support for bottlenecks and recent population expansions within both groups (Wennerstrom et al. 2016).

The number of moose in Sweden is today estimated to about 350 000 individuals, of which 25-30% is harvested yearly (Jägareförbundet 2015; Wallgren 2016). Relative to other deer species in Sweden, the moose population is rather well monitored, including estimations of population trends based on several decades of recorded data from hunters' observations. However, as with many wild species, precise estimations of population demography are difficult to obtain. The large non-random reduction of the population every year, could alter the genetic diversity both on local and regional scales, which may have consequences for the long-term viability of moose (Harris et al. 2002). This is of even more concern in the light of the current rapid changes of the environment (Knell and Martinez-Ruiz 2017).

1.5 Aims of the thesis

The overall aim of this thesis is to develop and apply genetic methods for individual-based assessments of deer, with a focus on moose.

Specific aims:

- To develop and apply species-specific SNP genotyping for individual identification.
- To develop and apply *non-invasive* genetic assessments.
- To assess contemporary population genetic effects of seasonal migration.

2 Methods

2.1 Samples and locations (Papers I to IV)

A substantial part of the moose samples that have been included for analyses in this thesis belongs to the biobank at the department of Wildlife, Fish, and Environmental studies at the Swedish University of Agricultural sciences. Since the 1980's, moose have been equipped with VHF/GPS-collars for the purpose of collecting movement data. While sedated to enable the collaring, biological samples are taken and subsequently frozen and stored in -20°C for future research purposes. From this biobank, whole-blood samples were used in the full genotyping of moose in paper I, for *de-novo* SNP discovery in paper II and genotyping in papers III and IV. The samples of the remaining deer included in the full genotyping in paper I had been extracted from tissue in a previous research project.

In the pilot study of cross-amplification using bovine SNPs (paper I), nine ungulate species were genotyped; moose (*Alces alces*, n=49), roe deer (*Capreolus capreolus*, n=8), reindeer (*Rangifer tarandus*, n=10), fallow deer (*Dama dama*, n=8), red deer (*Cervus elaphus*, n=8), cattle (*Bos Taurus*, n=3), mule deer (*Odocoileus hemionus*, n=1), American bison (*Bison bison*, n=1) and wild boar (*Sus scrofa*, n=3). DNA from cattle, wild boar and four moose were extracted from commercial meat and hunter-harvested animals. Remaining samples had been previously extracted from tissue. The geographical origins of the targeted deer were as far as possible consistent with their Swedish distribution range. In the full genotyping, 10 samples each of moose, roe deer, fallow deer, red deer and eight samples of reindeer were included.

The 34 moose included for SNP discovery (paper II) were selected based on their geographical origin, to cover as much of the Swedish moose distribution range as possible and when available, both male and female samples were

selected from each location. Consequently, 34 samples from Abisko (68º) in the north to the island of Öland (56º) in the south of Sweden were included for sequencing (Fig 1).

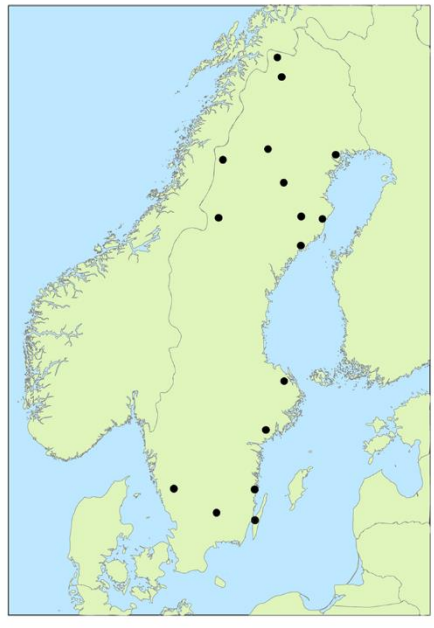

Figure 1. Sampling locations of moose (n = 34) included for *de novo* sequencing including 16 locations (1-3 individuals per location) throughout Sweden.

The sampling in paper III was conducted on the Baltic island of Öland (56°N, 16°E). Fecal pellets were collected non-invasively with an opportunistic approach, in which favorable moose habitats were targeted in order to obtain a sufficient sample size. Fecal pellets were collected throughout the island in 2016 during the period 29 January to 11 April (Fig 2). Also, fecal samples from seven captive moose (from the zoo ”Lycksele djurpark”) were included to enable validation of sex-specific SNPs. Furthermore, blood samples from 23 GPS-equipped moose on Öland were genotyped for validation purposes.

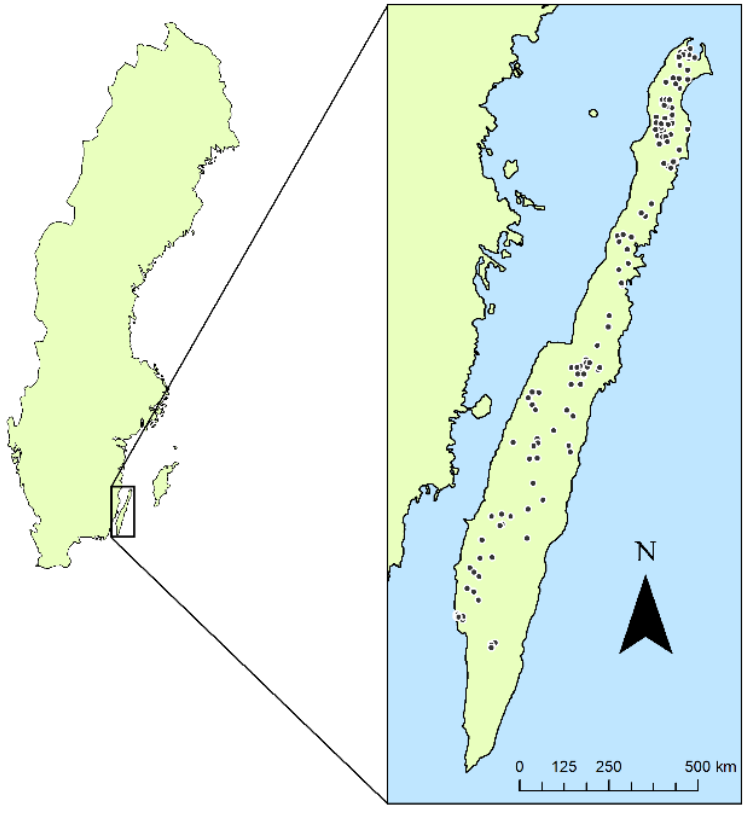

Figure 2. Map of Sweden with the island of Öland enlarged showing sampling locations of the fecal pellets included for genotyping.

In the study presented in paper IV, 42 GPS-equipped female moose were genotyped. The location for the study was the northern Scandinavian mountain range and adjacent Swedish lowland (Fig 3).

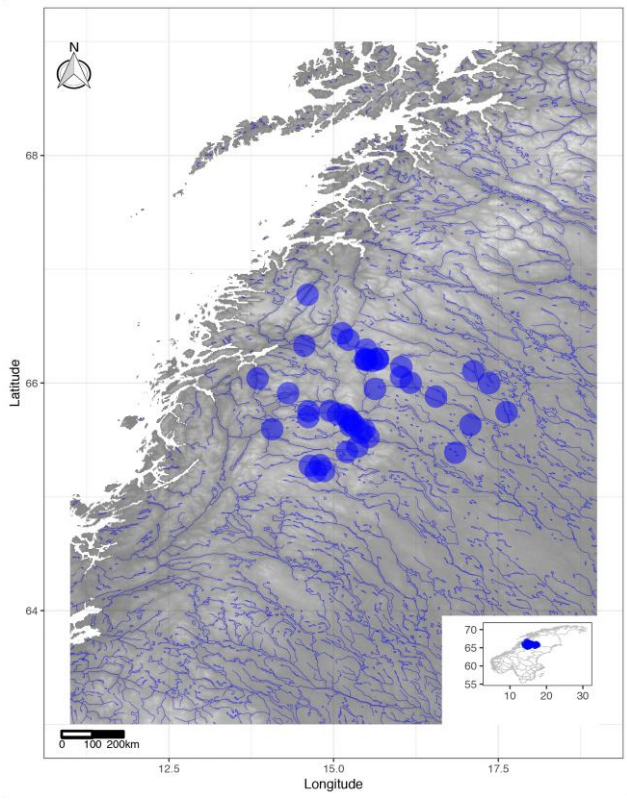

Figure 3. Map showing sampling locations for 42 moose in the northern Scandinavian mountain range and Swedish lowland.

2.2 DNA extractions (Papers I to IV)

All DNA extractions from blood or tissue were automatized using the QIAsymphony SP instrument (Qiagen, Hilden, Germany). Samples were processed with the QIASymphony DSP DNA Mini Kit according to the protocols of the manufacturer. For samples included in high-throughput sequencing, the DNA quantity was estimated with spectrophotometry (260:280 ratio, NanoDrop Technologies, Inc.) and the DNA quality controlled using gel electrophoresis.

Only the surface material from the fecal pellets was included for DNA extractions. This was to target moose epithelial cells and exclude as much of the non-target DNA as possible. Extractions of DNA from fecal samples were done manually with the QIAamp DNA Stool Mini Kit or the QIAamp Fast DNA Stool Mini Kit following the protocols of the manufacturer. All DNA extracts were stored at -20°C until used.

2.3 Genotyping using bovine SNPs (Paper I)

2.3.1 Pilot study

The potential of using cross-species amplification between cattle and the Swedish deer species was explored by selecting 48 SNPs from the commercial Illumina BovineSNP50 Genotyping BeadChip (Fluidigm Corporation, San Francisco, USA). The selected SNPs had in a study by Haynes and Latch (2012) shown polymorphism in the North American deer; black-tailed deer, mule deer *(Odocoileus hemionus* ssp*)* and white-tailed deer *(O. virginianus)*. In total, 93 DNA samples were genotyped on a 96-well array chip by the Fluidigm Biomark platform (Fluidigm Corporation, San Francisco, USA).

2.3.2 Full genotyping

Following the pilot study, a full genotyping with the complete BovineSNP50 BeadChip was conducted. Five deer species; moose, roe deer, red deer, fallow deer and reindeer, were targeted for SNP identification. Forty-eight samples were genotyped by AROS Applied Biotechnology, Denmark.

2.3.3 Evaluation of genotyping

In the pilot study, the genotyping output was manually evaluated, i.e. the performance of each SNP in terms of distinct assignment of samples to genotypes in the Biomark scatter plot plane was controlled. Three no template controls (NTCs) were included in the genotyping. These samples contain water instead of DNA and were included to detect possible primer-dimer binding during the PCR preceding the genotyping. If primer-dimer bindings were detected, all samples clustering in the same area of the scatter plot plane were removed. Also, samples with ambiguous genotype assignment were removed from the clustering algorithm. Evaluations were done with two different SNP amplification thresholds; ≥ 50% and ≥ 75%. Error rates were estimated for SNP duplicates, and calculated as the number of allelic mismatches divided by the total number of amplified alleles.

In the full genotyping, allele calling and basic filtering (samples with SNP call scores < 0.15 were not considered) had been performed prior to delivery of the genotyping data. We applied an amplification rate thresholds of ≥ 90% in the full genotyping but we also evaluated loci with ≥ 75% amplification rate. SNPs with a minor allele frequency (MAF) below 0.08 were removed in order to avoid including false SNPs generated by genotyping errors in the downstream

analyses. By using the cut-off value of 0.08, loci where the minor allele was only found once were excluded for all sample sizes, also for loci with missing data (90% amplification). The risk of not detecting the minor allele, due to the limited number of individuals genotyped at the species level, was estimated based on the allelic proportions under Hardy-Weinberg equilibrium (HWE).

2.4 Moose SNP development (Paper II)

2.4.1 RRL sequencing

In order to reduce the genomic material for sequencing while maintaining a sufficient genomic coverage for discovery of high quality autosomal SNPs, a reduced representation library (RRL) sequencing approach was taken. Thus, extracted genomic DNA was digested with the restriction enzyme Eae I to produce fragments of DNA for sequencing. Eae I was estimated to produce a relatively large amount of fragments (of length 1500 bp) in the chosen size range (300-500 bp) per sample based on an estimated genome size of 3 Gb (Altshuler et al. 2000). We aimed for a read-depth per base of 20X and a genomic coverage of 2%.

The digested DNA was purified using the MinElute Reaction Cleanup Kit (Qiagen) to maximize the DNA yield per unit. Multiplexed library preparations and subsequent high-throughput paired-end sequencing on the Illumina HiSeq™ 2500 platform were performed by the National Genomics Infrastructure (NGI), SciLifeLab, Stockholm.

2.4.2 Autosomal SNP discovery

Following initial quality assessment of the sequencing output, the Stacks pipeline (Catchen et al. 2013) was used for further data trimming and subsequent discovery of autosomal SNPs. The minimum read depth for SNP calling was set to three in order to avoid excluding putative SNPs with low read depth in some of the individuals. The criteria for selection among putative SNPs were; one SNP per read, all three allele combinations present among the sequenced individuals (XX, XY, and YY), data from at least 24 of the 34 individuals and in all sampled locations, minor allele frequency > 0.2 and low linkage between SNPs. SNP assays (customized primer pairs) were developed by the Fluidigm Corporation, San Fransisco, USA. Validation of the SNP assays were done by manual screening of the output from genotyping the sequenced individuals and additional moose.

Evaluations of SNP genotyping

Following a manual control of the genotyping output (same as for the pilot study in paper I), the performance of single SNPs and samples were evaluated. In general, a threshold of 75% amplification rate was applied for both SNPs and samples to be kept for further analyses. Technical sample replicates were included for estimations of genotyping error rates; twelve samples were duplicated and two samples were replicated 10 times. The later was to allow for discrimination between drop-out and drop-in misprinting errors, which cannot be achieved for duplicates only. Error rates were calculated as the number of allelic mismatches divided by the total number of amplified alleles. A mother with twin offspring were included to enable validation of Mendelian inheritance patterns of the SNP alleles.

2.4.3 Discovery of sex-specific- and species diagnostic SNPs

Three sex-specific SNPs were also developed *de novo*. These were designed to be Y-chromosome-specific i.e. only amplifying male samples. This was achieved by PCR (polymerase chain reaction) amplification of the *Sry* region (Lindsay and Belant 2008) in six moose and followed by Sanger sequencing of the PCR products. Moreover, three species diagnostic SNPs were developed for separation of the deer species occurring in Sweden (moose, roe deer, fallow deer, red deer and reindeer). This was done by Sanger sequencing of the 12S mitochondrial region of six individuals of each species (Yang et al. 2014).

All Sanger sequencing was performed on the 3730 xl DNA analyser (Applied Biosystems, Foster City, USA) at the department of Medical Biosciences, Umeå University, Sweden. The sequencing outputs were aligned and manually screened for putative SNPs using the software BioEdit 7.0.5 (Hall, T. Ibis Therapeutics, Carlsbad, USA). Validation of the SNPs was performed by genotyping samples of known sex (both male and female) and species respectively.

2.5 Non-invasive individual identification (Paper III)

Individual genotypes were retrieved from fecal pellets for the purpose of estimating population size and sex-ratio. Because of the lower quality of non-invasively collected DNA, sample duplicates were analyzed to confirm reliable genotypes. Using the package *allelematch* (Galpern et al. 2012) with the R software (R Core Team 2017), genotypes were compared and matched. In manual screenings of the outputs of the analysis, the performance of sample duplicates were controlled in order to find the most appropriate threshold for the

number of allelic mismatches that could be allowed for correct separation of individuals without matching unique genotypes. Non-consistent genotypes (duplicated samples not producing identical results) were discarded in order to avoid "false unique" individuals that could inflate population size estimations.

2.5.1 Population size estimation

Based on uniquely identified genotypes among the non-invasively collected samples, the population size of moose on the island of Öland was estimated using two different capture-mark-recapture (CMR) approaches. Basic assumptions for the population size estimations were that the population was closed (no births and negligible number of deaths or migrants) in space and time and that all individuals were correctly identified. A maximum likelihood estimate was obtained with the R package *capwire* (Pennell et al. 2013) using a model assuming equal capture probabilities. In addition, a Bayesian single sampling session CMR method implemented in the R function *CMRpopsize* (Petit and Valiere 2006) was applied.

2.5.2 Sex identification of non-invasive samples

By genotyping fecal samples from individuals with known sex, estimations of allelic drop-out- (Y-specific allele not amplifying in male) and drop-in (Y-specific allele amplifying in female) rates could be made. Based on these estimations, the probability of falsely determining a male sample as female was calculated (Paper III, Fig 2). The sex determination of non-invasive samples were furthermore validated by comparing the results with matching genotypes from blood samples from moose of known sex. Calculations of sex-ratio was based on uniquely identified individuals.

2.6 Population genetic analyses (Papers I to IV)

Throughout the thesis, estimations of expected and observed heterozygosity have been conducted to quantify the genetic variation of groups of individuals. Based on these estimations, deviations from the population genetic "null model" constituted by the Hardy-Weinberg equilibrium (e.g., random mating, infinite population size, non-overlapping generations, no natural selection, no migration) have been investigated, both during the SNP validation process and in downstream population genetic assessments. These calculations have been performed using the different software Plink (Purcell S et al. 2007), Genepop (Raymond 1995) and GenAlEx (Peakall and Smouse 2006; Peakall and Smouse

2012). For the applications in this thesis, a prerequisite for an informative set of SNPs is the random association of alleles at different loci. Hence, tests of linkage disequilibrium (LD) were performed using Plink and Genepop during SNP selection and subsequent empirical validations of the SNP assays.

As a measure of genetic differentiation between groups of individuals, the Wright's F-statistics were applied in all four papers using GenAlEx or Genpop. Calculations of F_{st} were conducted to measure the proportion of genetic variation explained by between-population variation in relation to within-population variation. The inbreeding coefficient, F, was used in paper III to detect plausible inbreeding. Further analyses of genetic differentiation between groups were performed using different approaches. The Bayesian clustering algorithm in the program STRUCTURE (Pritchard et al. 2000), that use Markov chain Monte Carlo (MCMC) sampling, was applied in papers II and III. In STRUCTURE, population clusters are constructed based on maximal HWE and minimum LD within groups. Succeeding selection of the most likely number of population clusters were based on Evanno's ΔK-method (Evanno et al. 2005). Moreover, a multivariate principal coordinate analysis (PCoA) or the similar principal component analysis (PCA), was performed with GenAlEx and with the R package adegenet (Jombart 2008; Jombart and Ahmed 2011) respectively. One of the two methods was used to visualize similarities or dissimilarities of genotypes in all papers. Another method to describe genetic groups in the data was applied in paper IV, by using the discriminant analysis of principal components (DAPC) (Jombart et al. 2010). Prior to the DAPC, groups are defined by the clustering algorithm *k*-means, which maximize the variance between groups. Thereafter, the groups are described by the DAPC, by finding the largest between-group and smallest within-group variance. Furthermore in paper IV, spatio-temporal genetic structure was explored combining genotypic and movement data using spatial principal component analysis (sPCA) (Jombart et al. 2008). The sPCA uses the product of spatial autocorrelation and variance to find global structures (differentiation of individuals) and local structures (repulsion between nearby individuals). Five different time points were considered for analyses; sampling/GPS collaring, calving, the rut, one summer position, and one winter position.

The SNPs' capacity to accurately separate individuals were evaluated based on calculations of probability of identity (PID) in GenAlEx. PID_{sib} estimates the probability that two samples, with allele frequencies equivalent to what is expected for first order relatives, erroneously are identified as the same individual. Pairwise relatedness was estimated for validation purposes in paper I using the program COANCESTRY (Wang 2011) in the R package related (Pew

et al. 2015) and in paper IV (GenAlEx) to find out if highly related individuals could explain cryptic population structure.

3 Summary of results

3.1 Cross species amplification (Paper I)

In the pilot study, 11 to 14 out of 48 loci amplified in each species. Out of these, one to three showed polymorphism. Five loci amplified in all species but no shared polymorphic loci were found. A locus amplification rate of 50% was associated with high error rates and concluded too low for reliable SNP detection. The amplification rate for cattle was a few percentage lower than expected, but largely explained by a single sample that showed low amplification rate.

In the full genotyping with the BovineSNP50 BeadChip, 22 929 of a total 54 609 loci (42.0%) were successfully amplified (>90%) in all species. At the species level, 36 200 (66.3%) to 38 850 (71.1%) loci were amplified. Between 225-883 polymorphic loci, in a wide minor allele frequency range, were detected in the different species, the fewest in moose and the most in fallow deer (Paper I, Table 2).

No SNPs that were shared among all species were found in the full genotyping. The numbers of shared SNPs between species pairs were also low but more than twice as high between fallow deer and reindeer compared to the other pairs. Roe deer and reindeer were the two species with the highest number of detected SNPs.

A combination of eleven SNPs were found enough to separate first order relatives with high certainty ($PID_{sib} < 0.01$) in all species (Fig 4). In fallow deer, only nine SNPs were required to obtain the same result. The expected heterozygosity was between 0.40-0.46 for all species. A higher variation was found in the observed heterozygosity; from 0.49 in red deer to 0.75 in fallow deer. Yet, because of low sample sizes, all species were found in HWE. The moose samples were separated into one north and one south population group,

similar to previous studies (Fig 5). The F_{st} between these groups was estimated to 0.07. For the other four species, the pattern of differentiation was not as clear. In roe deer, the southernmost individuals (M) were more similar to each other than to the rest. In red deer, the same pattern appeared for the two northernmost individuals (X). Both in fallow deer and reindeer, one individual was found to be highly genetically separated from the rest. For fallow deer, the differentiated sample originated from the county of Uppsala. Among the reindeer samples, one individual ("Jokkmokk") was highly separated from all other genotyped reindeer (Fig 6). Another reindeer sample, retrieved from the county of Västerbotten did also separate from the remaining reindeer that all originated from the county of Norrbotten.

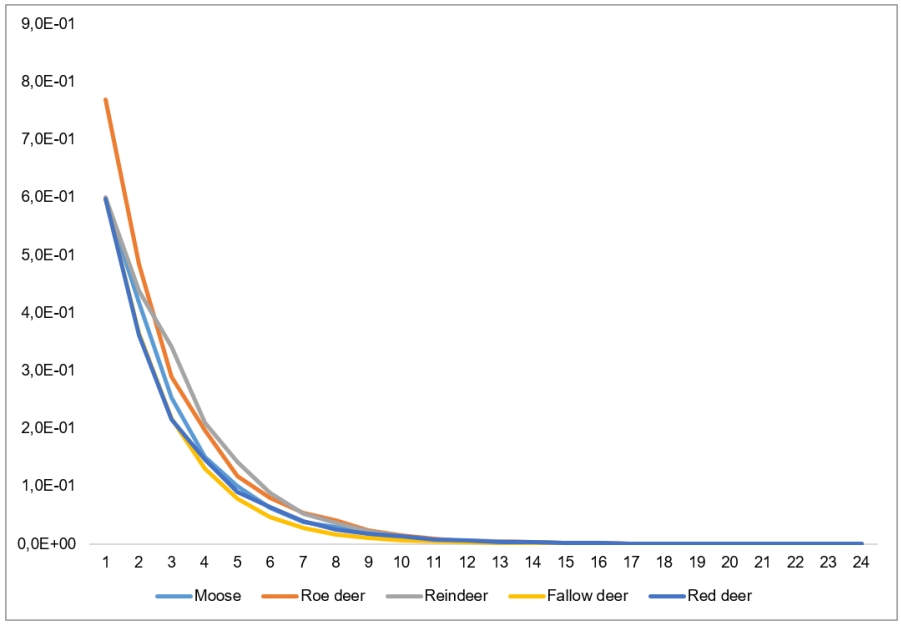

Figure 4. Results of tests for probability of identity for siblings in all Swedish deer species. The x-axis displays the number of loci and the y-axis the probability of identity, which decreases as more SNPs are added. Using 11 SNPs, siblings are separated ($PID_{sib} < 0.01$), in fallow deer nine SNPs are sufficient.

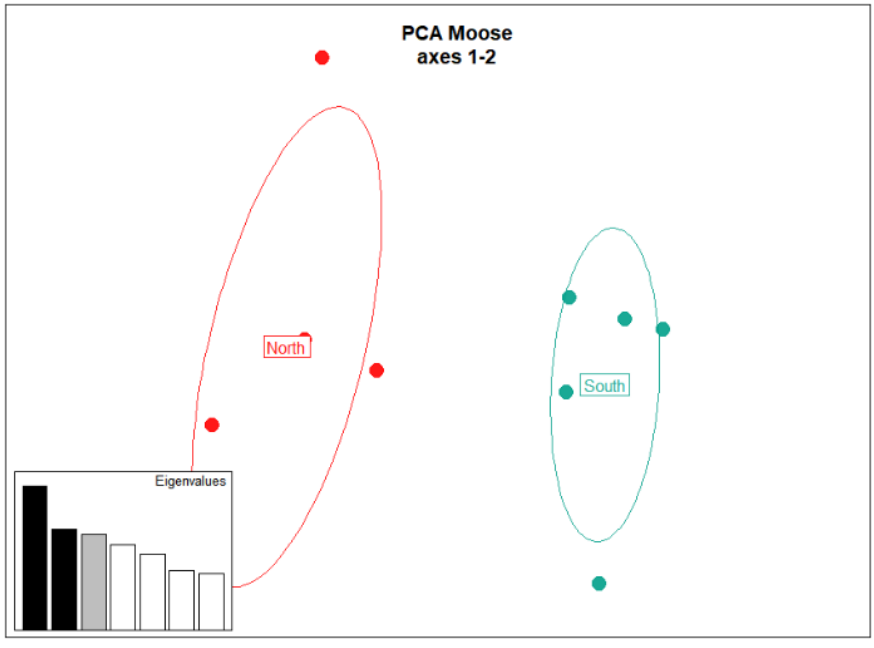

Figure 5. The PCA separated moose in concordance with the previously found North-South substructure for moose in Sweden.

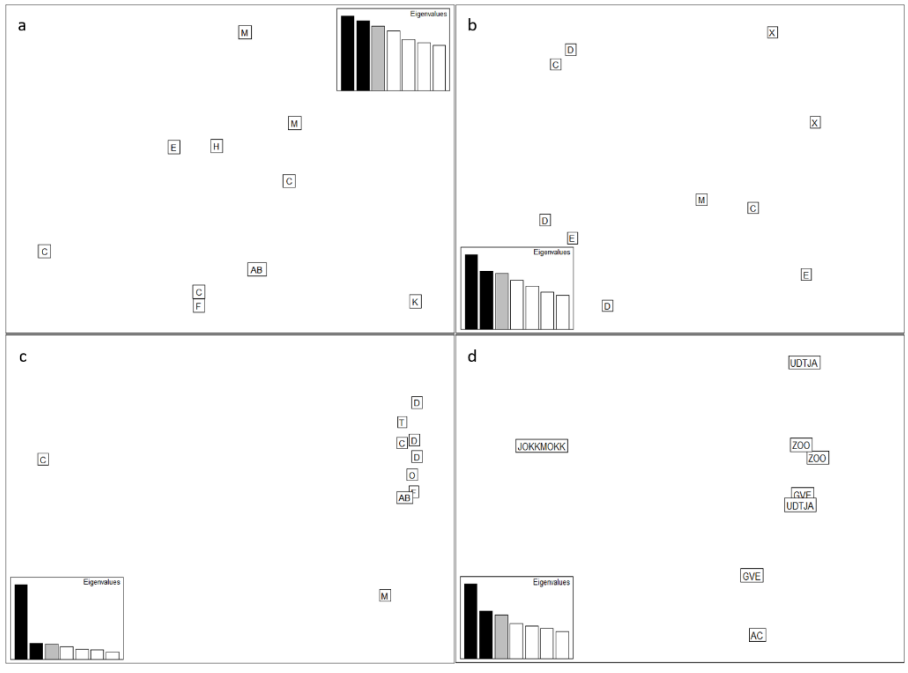

Figure 6. Visualization of PCA results for the genotyped species. a) roe deer (PC 1 = 14.7%, PC 2 = 13.5%), b) red deer (PC 1 = 16.7%, PC 2 = 13.4%), c) fallow deer (PC 1 = 31.8%, PC 2 = 12.3%) and d) reindeer (PC 1 = 23.7%, PC 2 = 16.5%). Letters for single samples represent their origin by name of location or county codes. Abbreviations; AB = Stockholms län, AC =Västerbottens län, BD = Norrbottens län, C = Uppsala län, D = Södermanlands län, E = Östergötlands län, F = Jönköpings län, GVE = Gällivare, H = Kalmar län, K = Blekinge län, M = Skåne län, O = Västra Götalands län, T = Örebro län, X = Gävleborgs län

3.2 De-novo moose SNP discovery (Paper II)

Almost 50 million read pairs of genomic data, between 0.61- 2.16 million per sample, were obtained from high-throughput sequencing. Over 240 000 unique stacks of matching sequences were generated of which almost 50 000 contained at least one SNP. The resulting genomic breadth of coverage was estimated to 1.2%. The filtering for high quality SNPs[4] resulted in the selection of 140 SNPs for assay design. The 86 best performing SNPs were subjected to a final genotyping effort including 59 DNA samples for estimations of error rates and evaluations of their performance in downstream analyses (Fig 7).

4. A high quality SNP has a high amplification rate and well defined clusters across a range of sample qualities.

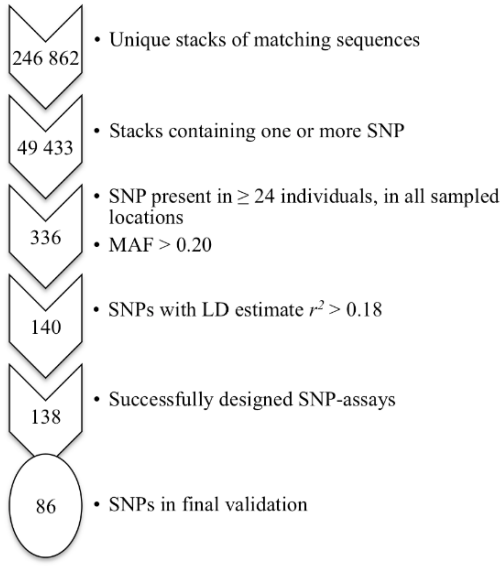

Figure 7. Schematic overview of the selection process of autosomal SNPs for individual identification, starting at nearly 250 000 stacks of matching DNA sequences and resulting in the selection of 86 SNPs for final validation.

The mean overall error rate per locus was estimated to 0.002. The mean expected heterozygosity was 0.48 (*SD* = 0.02) and observed heterozygosity was 0.45 (*SD* = 0.08). Seven SNPs deviated from Hardy-Weinberg expectations after removing close relatives ($r > 0.35$) and four SNPs deviated from HWE based on sequenced individuals only. The mean minor allele frequency of the SNPs was 0.43. Linkage disequilibrium was suggested in 1.5% of the pairwise comparisons at an α-level of 0.01. The probability of identity for first order relatives (PID_{sib}) was below 0.01 for the 10 most informative SNPs. The relatedness coefficients for the known kinships (*r*) (Lynch and Ritland 1999) were estimated to; mother-daughter: 0.47, mother-son: 0.35 and brother-sister (twins) 0.51. Two genetically differentiated clusters, a north and a south, were identified based on the STRUCTURE analysis and following calculations of Evanno's ΔK. The F_{st} between these clusters was estimated to 0.08 (Fig 8).

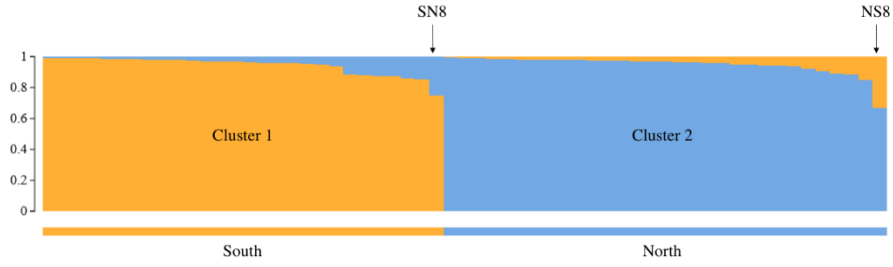

Figure 8. Barplot showing the two clusters suggested by STRUCTURE. The SNP panel separate the 59 moose included in the SNP validation into two genetic clusters. Information about sampling location (south/north) was added to the figure after the analysis to visualize the concordance between assignment of genetic cluster and sampling location. Two individuals, SN8 (south-north 8) and NS8 (north-south 8), are pointed out since they show the most admixture between the two clusters.

Five sex-specific SNPs were included for genotyping (two previously developed), four of these successfully amplified in the final validation run. These four SNPs correctly identified the sex of all genotyped samples (n male = 28, n female = 31), except two samples labelled as female. However, these samples were consistently identified as male in all genotyping runs and thus highly likely mislabeled. Excluding these two samples, the error rate was zero for the sex-specific SNPs.

The species diagnostic SNPs included in the final validation (n=5) correctly identified the species for all 59 samples. Again, the error rate was zero. One SNP appeared variable among roe deer on the Baltic island of Öland, leading to an identical allele pattern to reindeer. For studies in areas were these species co-exist, this species diagnostic SNP should be replaced.

3.3 Non-invasive population assessment (Paper III)

Of 489 fecal pellet samples genotyped, 302 (62%) were successfully genotyped. Eight of these were identified as roe deer by the species diagnostic SNPs. Samples with high genotyping success were associated with fewer allelic mismatches and consequently produced more reliable genotypes (Negative binomial GLM: $\chi^2 = 144.89$, $p < 0.001$). A quality filtering process (Fig 9) in which genotypes were matched and the results manually evaluated, led to a final dataset with reliable genotypes for individual identification consisting of 182 genotypes. The unfiltered mean genotyping error for the autosomal SNPs was estimated to 0.14 (sd = 0.07) and for the quality filtered dataset 0.06 (sd = 0.06). Evaluated based on the uniquely identified genotypes, the PID_{sib} approached

zero for 16 of the most informative SNPs. Five SNPs were found to deviate from HWE ($p < 0.05$) in Öland moose. The mean minor allele frequency for the SNPs was estimated to 0.33.

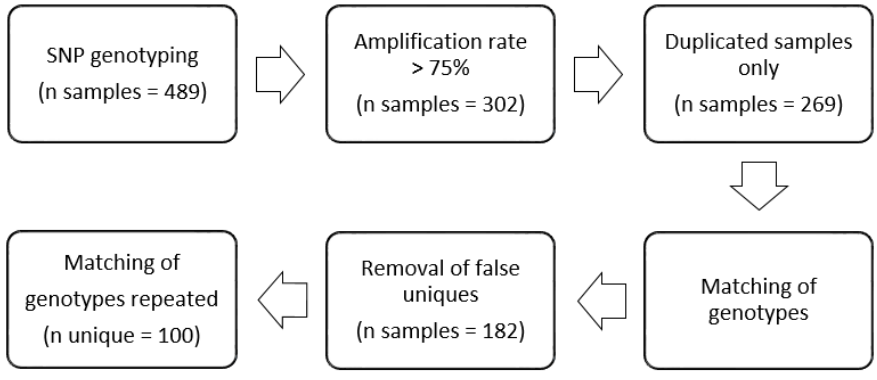

Figure 9. The main steps of the quality filtering process undertaken for the genotyped non-invasive samples.

Out of the 182 reliable genotypes, 100 unique individuals were identified. Six of the unique individuals identified from the non-invasive samples had previously been GPS collared (26%). The maximum likelihood estimate for the population size was 134 individuals, ranging between 115–156 individuals within a 95% confidence interval. The Bayesian approach estimated the number to 135 individuals with a 95% highest posterior density interval (HPD) of 118-156. Seventeen GPS equipped moose occurred on the island at the time of sampling that were not recaptured in the non-invasive data.

Among the 100 unique individuals, 37 males were identified, resulting in a male sex ratio of 37% ±9% (95% CI). The error rate for the sex-specific SNPs was estimated to 0.01. The error rate for the species diagnostic SNPs was 0.02.

Genetic differentiation was found between moose on Öland and the Swedish mainland based on STRUCTURE analysis and PCoA (Fig 10). The F_{st} between moose in these areas was 0.03. The inbreeding coefficient, F, for moose on Öland was estimated to 0.04-0.05. A weaker structure separating moose in the northernmost forest areas of Öland from remaining moose on the island was detected.

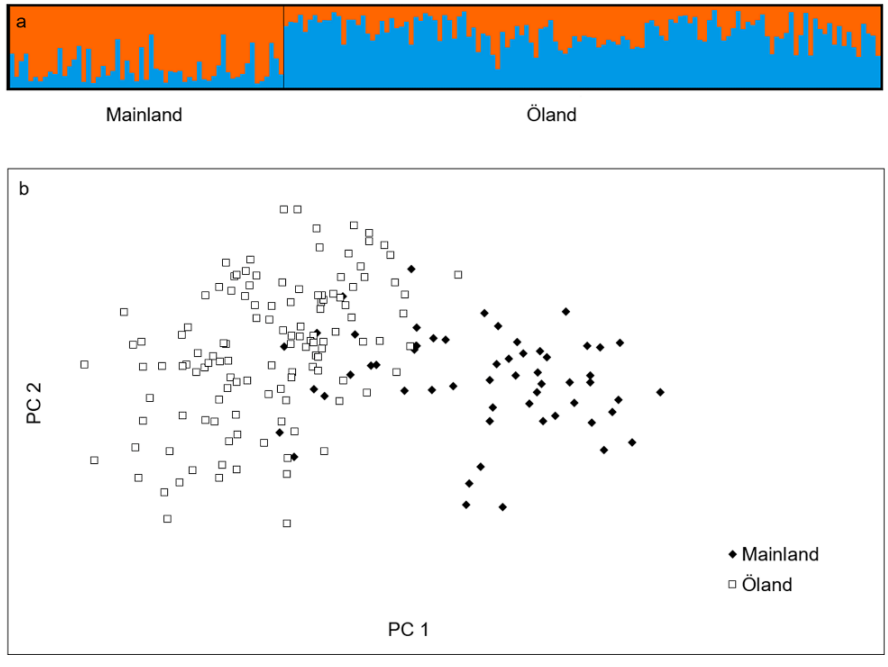

Figure 10. Population differentiation between moose on the Swedish mainland and moose on Öland. (a) Barplot of STRUCTURE output K = 2 separating mainland and Öland moose in two population clusters (cluster one = orange, cluster two = blue). (b) PCoA visualizing the genetic distribution of moose from the mainland (n = 57) and Öland (n = 124). PC 1 and 2 explain 7.54% and 5.41% of the variation respectively.

3.4 Spatio-temporal population structure (Paper IV)

Global structures, indicative of differentiation between spatially separated groups, was detected during the rut (max (**t**) = 0.04, $p = 0.004$), in August (max (**t**) = 0.04, $p = 0.007$) and at the calving positions (max (**t**) = 0.04, $p = 0.027$) but not at the positions for sampling (at moose summer range in early November) nor in January. Two global scores were retained for each significant time point, the spatial autocorrelation (I) associated with the first two scores was for the rut: 0.81 and 0.65, in August: 0.81 and 0.65 and at the calving: 0.77 and 0.72. The strongest structure, most pronounced during the rut and in August, was separating north-western alpine moose from all other moose, which occurred in the south-east alpine and lowland areas. A pattern of a population cline between these two areas appeared form the first global score. Low differentiation between these groups was found based on analyses of merely genetic variation; the F_{st} between the areas was estimated to 0.02 and neither a PCoA nor a DAPC identified this separation. A second global structure pointed out a south-eastern

group in the Ume River valley as different from the rest during the rut and in August. No significant local structures were found. However, a pattern of negative spatial autocorrelation appeared during the rut and was most pronounced within the group in the Ume River valley (Fig 11). This observation was accompanied with a considerably lower p-value ($p = 0.19$) than received for local structures in other time points ($p = 0.70$-1.00). Calculations of pairwise relatedness among the 42 individuals revealed one pair of first order relatives ($r > 0.50$) whereas 2.6% of the pairwise comparisons resulted in an r-value above 0.2%.

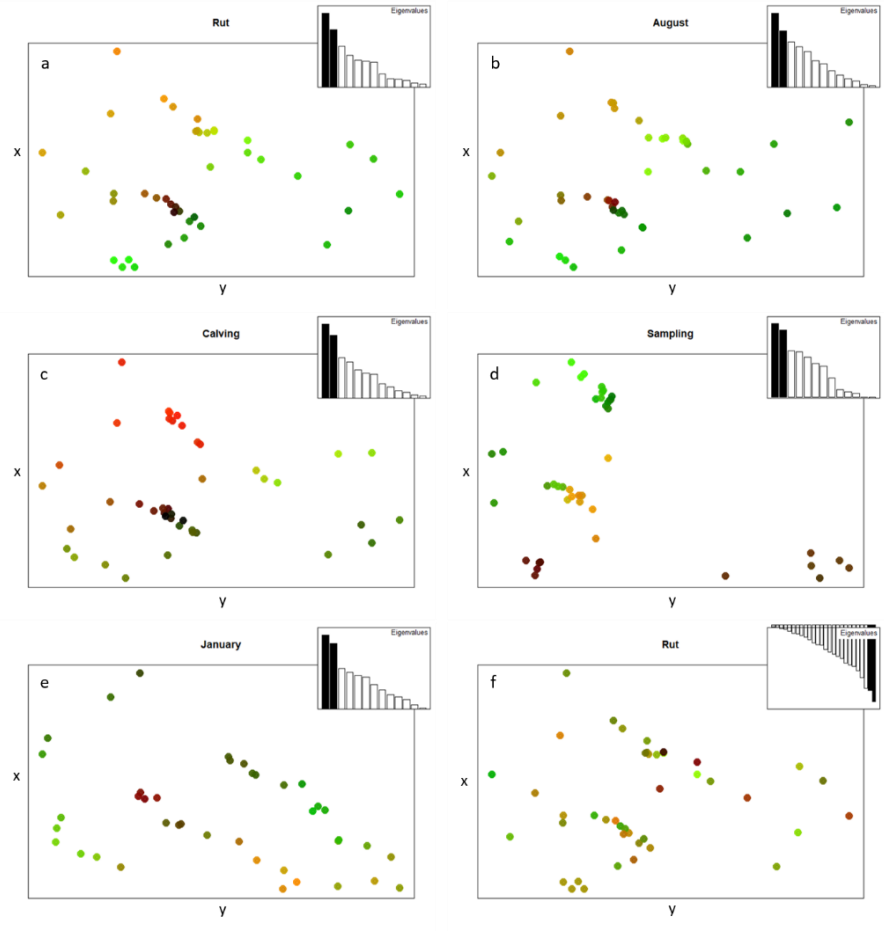

Figure 11. Global structure identified by the sPCA during all time points analyzed (a-e, green and yellow represent most extreme values). Significant global structure were detected during the rut (axis 1 and 2 explaining 40.2 % of the variation), in August (axis 1 and 2 explaining 37.3% of the variation) and during calving (axis 1 and 2 explaining 40.8% of the variation). A general pattern of a population cline between north-western alpine- and all other moose appeared. During the rut and in August, a group of moose in south-east, by the Ume River valley, stands out as different (brown and black). A tendency of a local structure was seen during the rut, most pronounced within the group in the Ume River valley (f).

4 Discussion

4.1 SNP development

4.1.1 Cross species amplification (Paper I)

SNPs were discovered using two different approaches. The reason for the separate approaches was largely explained by the rapid technical development within the field of genomics during the work with this thesis. When this PhD project started in the end of 2011, the cost for *de novo* sequencing was still high enough to warrant alternative methods for SNP discovery. This motivated the pilot study in which deer were genotyped with a selected number of bovine SNPs. Since the selected loci previously had shown polymorphism in other deer species, we anticipated a relatively high proportion of polymorphism. Accordingly, we did obtain a somewhat higher proportion of polymorphic loci compared to the amplification of North American deer by Haynes and Latch (2012). However, the success rate was still low, indicating that genotyping using a considerably higher number of bovine SNPs was going to be required in order to develop useful SNP panels for the different species. The most important conclusion of the pilot study was that a lower amplification rate threshold for the SNPs resulted in higher error rates and hence less reliable SNPs. Based on the literature, different thresholds, sometimes as low as 50%, have been used to define "successful SNP amplification" (Miller et al. 2012). This makes comparisons of success rates between different cross-species amplification studies difficult.

In the full genotyping, the five deer species that are found in Sweden were genotyped on the full Illumina BovineSNP50 Genotyping BeadChip. A similar amplification rate (42%) to that reported for the North American deer was achieved. The amplification rate was comparable between the Swedish species,

indicating that the many SNP binding sites are conserved among the five deer species. The proportion of amplified loci that were found polymorphic varied between the species from 0.6% (moose) to 2.3% (fallow deer).

Based on the fact that low levels of SNPs were shared among species and that no correlation to phylogeny was found (moose, roe deer and reindeer belong to subfamily Capreolinae and fallow deer and reindeer to subfamily Cervinae), the majority of the detected SNPs are likely new mutations that have arisen after the divergence among the Cervidae family. The variation in SNP detection rate between species is also likely reflecting different population histories. Fewest SNPs were found in moose and roe deer. These species have both been through severe bottle necks during the last few hundred years, which probably have caused fixation of alleles and thus loss of polymorphism. The population differentiation in moose is furthermore reflecting the wide spatial distribution of the species and potentially also shows effects of relatively recent processes of population expansions in the different regions. We failed to cover the northern distribution range of roe deer, but based on the 10 samples genotyped, low levels of population structure appeared. Red deer has also been severely suppressed, however, new introductions from Denmark and southern Europe have been made as late as during the twentieth century (Liberg 2010), which probably have counteracted a loss of polymorphism. The number of SNPs found in red deer was accordingly in the mid-range of the numbers found for the rest of the species. The relatively large numbers of SNPs detected in fallow deer might imply that this species have a higher proportion *retained* SNPs compared to the other deer species. The higher level of polymorphisms found for this species may also be explained by introductions of animals from a potentially wide geographical distribution, which also could be reflected in the two individuals that are clearly separated from the rest in the PCoA. Furthermore, the high observed heterozygosity in fallow deer (0.75) could suggest recent interbreeding between previously isolated groups. The second most SNPs were identified in reindeer and the mean heterozygosity level for the semi-domesticated reindeer was similar as for the other deer species (except fallow deer). Consequently, given the low number of individuals genotyped in this study, the genetic variation of reindeer nonetheless seem comparable to other deer species in Sweden. The highly separated reindeer sample originated from a mountain reindeer herding district, whereas the rest of the reindeer from the Norrbotten county belonged to forest reindeer herding districts.

For all five species, a relative low number (9 to 11) of SNPs was enough for separation of first-order relatives. Hence, a "multi-species" SNP panel for individual identification of sympatric deer is one application that is now achievable using the SNPs detected in this study.

4.1.2 De-novo moose SNP discovery (Paper II)

De-novo sequencing of moose for SNP detection was initiated after the pilot study using cross-species amplification proved to result in low numbers of detected SNPs in moose. Based on results from SNP discovery in brown bear (Norman et al. 2013), a reduced representation library sequencing approach was found suitable. In particular since, as for the brown bear, the most important application for a moose SNP panel was going to be individual identification of non-invasively collected samples.

High-throughput sequencing produced different amounts of genomic information for the 34 individuals. This can partly be explained by variable DNA quality retrieved from the whole-blood samples. This furthermore led to that the stack depth for some individuals had to be reduced considerably during SNP discovery since we wanted each SNP to be present in as many individuals as possible in order to detect SNPs informative about individual identity. However, the risk of including false SNPs produced by sequencing errors was kept low since the stacks used for SNP detection included reads from 24-34 individuals.

The validation by genotyping using the discovered SNPs showed that genotypes were produced with low error rates (0.002 per locus). Furthermore, the SNPs were found to perform well for their intendent purpose; to separate genotypes with high certainty. Using ten of the most informative SNPs, first order relatives were separated (PID < 0.01). Seven SNPs were found to deviate from HWE based on genotyping of moose (n=59) throughout Sweden. Wild animal populations rarely meet the assumptions of HWE and hence the deviations were not surprising although it should perhaps be considered for population genetic applications assuming markers in HWE. Surprisingly, the estimated relatedness coefficient (*r*) for one mother-offspring pair was lower (0.35) than expected (~0.50). The *r* value between the calf and its sibling as well as in between the mother and the sibling was close to 0.50. Since the SNPs performed well in all other aspects, this discrepancy could not be explained. The autosomal SNPs assigned moose to a northern and a southern subpopulation in accordance with the sampling locations. The F_{st} between northern and southern moose was estimated to 0.08 which is comparable with the estimation of 0.10 based on microsatellites by Wennerstrom et al. (2016).

The sex-specific- and species diagnostic SNPs developed and validated in this study produced results with high confidence. But caution should be taken in species diagnostic applications including both roe deer and reindeer, since roe deer on the island of Öland shared one allelic variation with reindeer. Replacing this SNP would be the best solution for genotypic applications including both these species, i.e. in areas of northern Sweden. Consequently, the appearance of polymorphism in the species diagnostic SNPs indicates that the number of

individuals included in the discovery process (n = 6) was too low for confirming monomorphic SNPs.

4.1.3 Synthesis (Papers I and II)

Cross-amplification between cattle and deer in this study resulted in the detection of SNPs with a wide minor allele frequency range (0.08-0.5) for all species. The detected SNPs appeared to be informative about the distribution of genetic variation among the individuals genotyped and furthermore showed to separate first order relatives with high certainty. Hence, the application range is broad for the detected SNPs for all five species. As an example, SNPs from all or a selection of species could be combined into a multi-deer SNP panel. In combination with species diagnostic SNPs, such a SNP panel could potentially be used to identify species and individuals from non-invasively collected samples such as saliva from browsed twigs or fecal samples. However, the performance of the bovine SNPs have not yet been verified for low-quality DNA. Nevertheless, based on previous research (Kraus et al. 2015; Norman and Spong 2015; von Thaden et al. 2017) and the non-invasive genotyping in paper III, the SNPs are expected to produce reliable genotypes (on microfluidic platforms) as long as highly degraded DNA and large amounts of PCR-inhibitors are avoided.

For moose, SNPs were developed with two different approaches. The SNPs discovered by *de-novo* sequencing were selected to have a high discriminatory power for individual identification, which was confirmed in empirical evaluations. Interestingly, the bovine SNPs separated individuals almost equally efficiently. This is explained by that 120 of the 225 detected SNPs were associated with a high MAF (> 0.2).

Both SNP panels showed similar results also for analyses of population structure including comparable F_{st} estimates between the northern and the southern moose. However, the bovine SNPs might be expected to be less informative about recent population processes because highly conserved loci were targeted. As a conclusion, the different approaches taken for SNPs detection in moose resulted in two SNP panels that proved to generate comparable inferences from both individual- and population level analyses. However, the validation process for the *de novo* sequenced SNPs was more rigorous including estimations of error rates and relatedness coefficients for mother-offspring-sibling pairs.

4.2 Moose population assessments

4.2.1 Non-invasive population assessment (Paper III)

The objective of this study was to investigate the performance of *de novo* discovered SNPs on non-invasively collected fecal samples. Furthermore, we wanted to apply non-invasive genotyping on highly relevant areas of application, in particular for populations controlled by hunting; estimations of population size and sex-ratio.

Using thoroughly quality filtered genotypes, unique individuals could be confirmed and a population size estimation retrieved. Both the point estimate and the confidence interval was concurrent between the two CMR approaches used and was 134 (maximum likelihood estimate) and 135 (Bayesian estimate), respectively, and in the range of 115-156 individuals. Hence, the estimation appeared as rigid. However, the precision of the population size estimation might have been affected by the large loss of samples during the quality filtering process. In addition to a decreased capture probability, the spatial coverage was negatively affected by the loss of samples, causing a deficiency of samples from the southernmost area of Öland. However, the lacking of southern samples should not have a large effect on the population size estimation, since the major part of moose are found in the northern and central parts of the island.

The sex-specific SNPs gave unambiguous results for all 100 uniquely identified genotypes. As a result, the male sex ratio was estimated to 37% (±9%). The precision of the sex ratio estimate would have been improved with a higher sample size. Moreover, the inclusion of species diagnostic SNPs was found useful, since eight samples of roe deer were identified. This study hence verify that both the species diagnostics- and sex-specific SNPs work well on non-invasive DNA.

One of the strengths with the use of microfluidic platforms for genotyping, is that they have been found suitable for fragmented DNA, which is an expected feature of non-invasively collected DNA. This has furthermore been empirically shown in studies analyzing non-invasive samples (Kraus et al. 2015; Norman and Spong 2015; von Thaden et al. 2017). However, moose winter fecal samples may contain high proportions of phenolic compounds from coniferous trees and shrubs. We propose this as an explanation to the lower than expected sample amplification rate. In accordance with von Thaden et al. (2017), we also found that amplification rate was a predictor of the reliability of genotypes. Consequently, the resulting mean amplification rate for the quality filtered dataset was high; 0.97.

In the analyses of genetic differentiation, a separation between moose on the island of Öland and the Swedish mainland appeared. Also, the level of heterozygosity for moose on the island was significantly lower than for moose on the mainland. Although some gene flow over the sea is occurring, the rate is most likely low. Thus, the lower levels of genetic variation detected among moose on Öland could be explained by a founder effect, i.e. loss of genetic variation due to that a small fraction of the mainland moose population were the founders of this population. Population structure, although weaker, was found also among Öland moose; moose in the northernmost forest area were somewhat differentiated from remaining moose on the island. This could be explained by a stationary behavior among the moose in the northern, more favorable moose habitat, which might create a spatial separation from moose south of this area. Another possibility is that fidelity to rutting areas could be a partial explanation to this pattern of structure.

Despite the constraints associated with retrieving reliable genotypes from non-invasively collected DNA, SNP genotyping of moose winter fecal pellets was found to generate highly valuable information that is difficult to obtain by other means. The benefits of non-invasive sampling, related to animal welfare and the possibility to conduct large scale population genetic assessments motivate further use and improvements of non-invasive genotyping of moose.

4.2.2 Spatio-temporal population structure (Paper IV)

Often, the sampling location determines the geographical origin of samples. This information is subsequently used for inferences from analyses of, e.g., population differentiation and individual assignment. Yet, many species move between different areas throughout the year, which makes defining of "home location" challenging. Moreover, depending on the time point for population genetic assessment, the pattern of genetic variation can be expected to change. By combining information about genetic variation with spatial information for different time points during the year, temporal population structure can be studied.

As expected, temporal variations in spatial genetic structure were detected in moose with flexible migratory behavior. Strongest population differentiation appeared during the rut and late summer and appeared as a population cline between north-western alpine moose and south-eastern moose. The same pattern of differentiation appeared during calving, although in the form of a patchier structure. An additional structure separated a south-eastern group in the Ume River valley from the rest during the rut and in the late summer. The differentiation between north-western alpine and all other moose in the study is

unlikely to be an effect of the topography *per se* for a large mammal such as the moose. A more probable contributing factor to the temporal structures found is that individuals return to the same areas between years, i.e. show site fidelity, to summer home ranges and in particular to areas for reproduction and calving. Since offspring is known to follow their mothers during their first year, migratory routes can be passed on between generations and hence lead to genetic structuring (Cederlund et al. 1987). Moreover, the separation of the group of moose in the Ume River valley might reflect that this area is a highly suitable moose habitat, potentially leading to an increased site fidelity among moose to this area.

Spatial genetic analyses based on the time points for sampling the moose and during winter, did not give rise to significant population structures although patterns of genetic structure could be seen also during these time points. However, the spatial distribution of moose in their winter home ranges was quite different compared to the other time points in the study. During winter, the studied individuals were more spread out in a west-east direction along the river valleys, which resulted in a lower spatial autocorrelation compared to during summer. At the time of sampling, the moose were still in their summer ranges but occurred in more distinct patches compared to the other summer time points.

With the sPCA analysis used, negative autocorrelation can be detected, indicating repulsion of genetically similar individuals or attraction between "individuals from other genetic pools" (Jombart et al. 2008). Although non-significant, the local structures appeared as stronger during the rut compared to the other time points. This effect was most pronounced within the group in the river valley. A speculation is that negative autocorrelation during the rut could be caused by behavioral mechanisms acting against inbreeding.

Since population structure was not found in analyses comparing allele frequencies only, the gene flow among moose in the study area seem to have been historically high. An interesting conclusion is thus that even though spatial genetic structure occur during the reproductive period, this has not caused substructuring of the population. The fact that the studied individuals are part of a relatively large population with a flexible seasonally migratory behavior, probably counteracts the effects of temporal structures. As a conclusion, this study shows that, by combining genetic and spatial data from different biologically valid time points, spatio-temporal genetic patterns can be detected in moose that most likely are caused by a combination of different behavioral and landscape components.

References

Altshuler D, Pollara VJ, Cowles CR, Van Etten WJ, Baldwin J, Linton L, Lander ES (2000) An SNP map of the human genome generated by reduced representation shotgun sequencing Nature 407:513-516

Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA (2016) Harnessing the power of RADseq for ecological and evolutionary genomics Nature Reviews Genetics 17:81 doi:10.1038/nrg.2015.28

https://www.nature.com/articles/nrg.2015.28#supplementary-information

Attard CRM, Beheregaray LB, Moller LM (2018) Genotyping-by-sequencing for estimating relatedness in nonmodel organisms: Avoiding the trap of precise bias Molecular Ecology Resources 18:381-390 doi:10.1111/1755-0998.12739

Baird NA et al. (2008) Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers Plos One 3 doi:10.1371/journal.pone.0003376

Bourgeois S et al. (2018) Single-nucleotide polymorphism discovery and panel characterization in the African forest elephant Ecology and Evolution 8:2207-2217 doi:10.1002/ece3.3854

Brookes AJ (1999) The essence of SNPs Gene 234:177-186 doi:10.1016/s0378-1119(99)00219-x

Bylemans J et al. (2016) Evaluating genetic traceability methods for captive-bred marine fish and their applications in fisheries management and wildlife forensics Aquaculture Environment Interactions 8:131-145 doi:10.3354/aei00164

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics Molecular Ecology 22:3124-3140 doi:10.1111/mec.12354

Cederlund G, Sandegren F, Larsson K (1987) SUMMER MOVEMENTS OF FEMALE MOOSE AND DISPERSAL OF THEIR OFFSPRING Journal of Wildlife Management 51:342-352 doi:10.2307/3801014

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing Nature Reviews Genetics 12:499-510 doi:10.1038/nrg3012

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study Molecular Ecology 14:2611-2620 doi:10.1111/j.1365-294X.2005.02553.x

Fernandez ME et al. (2013) Comparison of the effectiveness of microsatellites and SNP panels for genetic identification, traceability and assessment of parentage in an inbred Angus herd Genetics and Molecular Biology 36:185-U194 doi:10.1590/s1415-47572013000200008

Galpern P, Manseau M, Hettinga P, Smith K, Wilson P (2012) Allelematch: an R package for identifying unique multilocus genotypes where genotyping error and missing data may be present Molecular Ecology Resources 12:771-778 doi:10.1111/j.1755-0998.2012.03137.x

Garvin MR, Saitoh K, Gharrett AJ (2010) Application of single nucleotide polymorphisms to non-model species: a technical review Molecular Ecology Resources 10:915-934 doi:10.1111/j.1755-0998.2010.02891.x

Harris RB, Wall WA, Allendorf FW (2002) Genetic consequences of hunting: what do we know and what should we do? Wildlife Society Bulletin 30:634-643

Harris SE, Munshi-South J (2017) Signatures of positive selection and local adaptation to urbanization in white-footed mice (Peromyscus leucopus) Molecular Ecology 26:6336-6350 doi:10.1111/mec.14369
Harrisson KA et al. (2017) Signatures of polygenic adaptation associated with climate across the range of a threatened fish species with high genetic connectivity Molecular Ecology 26:6253-6269 doi:10.1111/mec.14368
Haynes GD, Latch EK (2012) Identification of Novel Single Nucleotide Polymorphisms (SNPs) in Deer (Odocoileus spp.) Using the BovineSNP50 BeadChip Plos One 7 doi:10.1371/journal.pone.0036536
Hundertmark K (2016) Alces alces. The IUCN Red List of Threatened Species 2016: e.T56003281A22157381. http://dx.doi.org/10.2305/IUCN.UK.20161.RLTS.T56003281A22157381.en. Accessed 2019-05-12
Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers Bioinformatics 24:1403-1405 doi:10.1093/bioinformatics/btn129
Jombart T, Ahmed I (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data Bioinformatics 27:3070-3071 doi:10.1093/bioinformatics/btr521
Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations Bmc Genetics 11 doi:10.1186/1471-2156-11-94
Jombart T, Devillard S, Dufour AB, Pontier D (2008) Revealing cryptic spatial patterns in genetic variability by a new multivariate method Heredity 101:92-103 doi:10.1038/hdy.2008.34
Jägareförbundet S (2015) Vilt & Vetande. Svenska jägareförbundet. Accessed 2019-05-12 2019
Kaminski S, Olech W, Olenski K, Nowak Z, Rusc A (2012) Single nucleotide polymorphisms between two lines of European bison (Bison bonasus) detected by the use of Illumina Bovine 50 K BeadChip Conservation Genetics Resources 4:311-314 doi:10.1007/s12686-011-9535-z
Knell RJ, Martinez-Ruiz C (2017) Selective harvest focused on sexual signal traits can lead to extinction under directional environmental change Proceedings of the Royal Society B-Biological Sciences 284 doi:10.1098/rspb.2017.1788
Kraus RHS et al. (2015) A single-nucleotide polymorphism-based approach for rapid and cost-effective genetic wolf monitoring in Europe based on noninvasively collected samples Molecular Ecology Resources 15:295-305 doi:10.1111/1755-0998.12307
Lampa S, Mihoub JB, Gruber B, Klenke R, Henle K (2015) Non-Invasive Genetic Mark-Recapture as a Means to Study Population Sizes and Marking Behaviour of the Elusive Eurasian Otter (Lutra lutra) Plos One 10 doi:10.1371/journal.pone.0125684
Liberg O, Bergström, R., Kindberg, J. and von Essen, H. (2010) Ungulates and their management in Sweden. In: Apollonio M, Andersen, R. and Putman, R. (ed) European ungulates and their management in the 21st century. Cambridge University Press,
Lindsay AR, Belant JL (2008) A simple and improved PCR-based technique for white-tailed deer (Odocoileus virginianus) sex identification Conservation Genetics 9:443-447 doi:10.1007/s10592-007-9326-y
Lynch M, Ritland K (1999) Estimation of pairwise relatedness with molecular markers Genetics 152:1753-1766
Malhi RS et al. (2011) Genotyping Single Nucleotide Polymorphisms (SNPs) Across Species in Old World Monkeys American Journal of Primatology 73:1031-1040 doi:10.1002/ajp.20969
Miller JM, Kijas JW, Heaton MP, McEwan JC, Coltman DW (2012) Consistent divergence times and allele sharing measured from cross-species application of SNP chips developed for three domestic species Molecular Ecology Resources 12:1145-1150 doi:10.1111/1755-0998.12017
Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA (2007) Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers Genome Research 17:240-248 doi:10.1101/gr.5681207
Moore JS et al. (2017) Genomics and telemetry suggest a role for migration harshness in determining overwintering habitat choice, but not gene flow, in anadromous Arctic Char Molecular Ecology 26:6784-6800 doi:10.1111/mec.14393
Morin PA, Martien KK, Taylor BL (2009) Assessing statistical power of SNPs for population structure and conservation studies Molecular Ecology Resources 9:66-73 doi:10.1111/j.1755-0998.2008.02392.x

Niedzialkowska M et al. (2016) The contemporary genetic pattern of European moose isshaped by postglacial recolonization, bottlenecks, andthe geographical barrier of the Baltic Sea Biological Journal of the Linnean Society 117:879-894 doi:10.1111/bij.12713
Nielsen R, Signorovitch J (2003) Correcting for ascertainment biases when analyzing SNP data: applications to the estimation of linkage disequilibrium Theoretical Population Biology 63:245-255 doi:10.1016/s0040-5809(03)00005-4
Norman AJ, Spong G (2015) Single nucleotide polymorphism-based dispersal estimates using noninvasive sampling Ecology and Evolution 5:3056-3065 doi:10.1002/ece3.1588
Norman AJ, Street NR, Spong G (2013) De Novo SNP Discovery in the Scandinavian Brown Bear (Ursus arctos) Plos One 8 doi:10.1371/journal.pone.0081012
Ogden R (2011) Unlocking the potential of genomic technologies for wildlife forensics Molecular Ecology Resources 11:109-116 doi:10.1111/j.1755-0998.2010.02954.x
Ogden R, Baird J, Senn H, McEwing R (2012) The use of cross-species genome-wide arrays to discover SNP markers for conservation genetics: a case study from Arabian and scimitar-horned oryx Conservation Genetics Resources 4:471-473 doi:10.1007/s12686-011-9577-2
Ogden R et al. (2013) Sturgeon conservation genomics: SNP discovery and validation using RAD sequencing Molecular Ecology 22:3112-3123 doi:10.1111/mec.12234
Palsboll PJ (1999) Genetic tagging: contemporary molecular ecology Biological Journal of the Linnean Society 68:3-22 doi:10.1111/j.1095-8312.1999.tb01155.x
Palsbøll PJ, Zachariah Peery M, Olsen MT, Beissinger SR, Bérubé M (2013) Inferring recent historic abundance from current genetic diversity Molecular Ecology 22:22-40 doi:10.1111/mec.12094
Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research Molecular Ecology Notes 6:288-295 doi:10.1111/j.1471-8286.2005.01155.x
Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update Bioinformatics 28:2537-2539 doi:10.1093/bioinformatics/bts460
Pennell MW, Stansbury CR, Waits LP, Miller CR (2013) Capwire: a R package for estimating population census size from non-invasive genetic sampling Molecular Ecology Resources 13:154-157 doi:10.1111/1755-0998.12019
Petit E, Valiere N (2006) Estimating population size with noninvasive capture-mark-recapture data Conservation Biology 20:1062-1073 doi:10.1111/j.1523-1739.2006.00417.x
Pew J, Muir PH, Wang JL, Frasier TR (2015) related: an R package for analysing pairwise relatedness from codominant molecular markers Molecular Ecology Resources 15:557-561 doi:10.1111/1755-0998.12323
Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data Genetics 155:945-959
Purcell S et al. (2007) PLINK: a toolset for whole-genome association and population-based linkage analysis. American Journal of Human Genetics 81
R Core Team RFfSC (2017) R: A Language and Environment for Statistical Computing
Radersma R, Garroway CJ, Santure AW, de Cauwer I, Farine DR, Slate J, Sheldon BC (2017) Social and spatial effects on genetic variation between foraging flocks in a wild bird population Molecular Ecology 26:5807-5819 doi:10.1111/mec.14291
Randall DA, Pollinger JP, Wayne RK, Tallents LA, Johnson PJ, Macdonald DW (2007) Inbreeding is reduced by female-biased dispersal and mating behavior in Ethiopian wolves Behav Ecol 18:579-589 doi:10.1093/beheco/arm010
Raymond MR, F. (1995) GENEPOP: population genetics software for exact tests and ecumenicism., (version 1.2) edn., J. Heredity 86: 248-249
Sacks BN, Louie S (2008) Using the dog genome to find single nucleotide polymorphisms in red foxes and other distantly related members of the Canidae Molecular Ecology Resources 8:35-49 doi:10.1111/j.1471-8286.2007.01830.x
Seeb JE, Carvalho G, Hauser L, Naish K, Roberts S, Seeb LW (2011) Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms Molecular Ecology Resources 11:1-8 doi:10.1111/j.1755-0998.2010.02979.x
Shafer ABA, Northrup JM, White KS, Boyce MS, Cote SD, Coltman DW (2012) Habitat selection predicts genetic relatedness in an alpine ungulate Ecology 93:1317-1329

Soderquist P et al. (2017) Admixture between released and wild game birds: a changing genetic landscape in European mallards (Anas platyrhynchos) European Journal of Wildlife Research 63 doi:10.1007/s10344-017-1156-8
Sovic MG, Carstens BC, Gibbs HL (2016) Genetic diversity in migratory bats: Results from RADseq data for three tree bat species at an Ohio windfarm Peerj 4 doi:10.7717/peerj.1647
Spitzer R, Norman AJ, Schneider M, Spong G (2016) Estimating population size using single-nucleotide polymorphism-based pedigree data Ecology and Evolution 6:3174-3184 doi:10.1002/ece3.2076
Thirstrup JP et al. (2015) Population genetic structure in farm and feral American mink (Neovison vison) inferred from RAD sequencing-generated single nucleotide polymorphisms Journal of Animal Science 93:3773-3782 doi:10.2527/jas.2015-8996
Turakulov R, Easteal S (2003) Number of SNPS loci needed to detect population structure Human Heredity 55:37-45 doi:10.1159/000071808
Wallgren M (2016) Sverige har världens tätaste älgstam. Skogforsk. https://www.skogforsk.se/kunskap/kunskapsbanken/2016/varldens-tataste-algstam/. Accessed 2019-05-12 2019
Wang JL (2011) COANCESTRY: a program for simulating, estimating and analysing relatedness and inbreeding coefficients Molecular Ecology Resources 11:141-145 doi:10.1111/j.1755-0998.2010.02885.x
Weinman LR, Solomon JW, Rubenstein DR (2015) A comparison of single nucleotide polymorphism and microsatellite markers for analysis of parentage and kinship in a cooperatively breeding bird Molecular Ecology Resources 15:502-511 doi:10.1111/1755-0998.12330
Wennerstrom L, Ryman N, Tison J-L, Hasslow A, Dalen L, Laikre L (2016) Genetic landscape with sharp discontinuities shaped by complex demographic history in moose (Alces alces) Journal of Mammalogy 97:1-13 doi:10.1093/jmammal/gyv146
von Thaden A et al. (2017) Assessing SNP genotyping of noninvasively collected wildlife samples using microfluidic arrays Scientific Reports 7 doi:10.1038/s41598-017-10647-w
Yang L, Tan ZQ, Wang DR, Xue L, Guan MX, Huang TS, Li RH (2014) Species identification through mitochondrial rRNA genetic analysis Scientific Reports 4 doi:10.1038/srep04089

Popular science summary

Access to high quality genetic data is important for studies of wild animals since it contains information that often cannot be obtained by other means. While collecting data from wild species can be challenging, one solution is to conduct studies non-invasively, by collecting and extracting DNA from for example feces, saliva or hair. By using non-invasive approaches, animals are furthermore less negatively affected by the sampling procedure. A well-established method to retrieve useful information from DNA is by using genetic markers that provide measures of genetic variation among groups of individuals. Another useful application of genetic markers is the identification of individuals. With individual-based information, population size estimations can be made from non-invasive sampling efforts, and by tracking known individuals over time, individual behavior and life histories can be studied.

Non-invasive DNA is often of limited quality and quantity, which can make genetic analyses difficult. A method that has shown to be successful is SNP (single nucleotide polymorphisms) genotyping. SNPs are positions in the DNA that vary and thereby can be used as genetic markers. However, these markers should ideally be developed specifically for the species of interest. This has long been an obstacle for SNP applications for non-commercial purposes, since the development implies that large amounts of genomic information has to be retrieved and processed, which requires time, money and computational expertise. However, the technology that generates the genomic information, DNA sequencing, has rapidly improved during the last decade, which has led to that the cost has decreased. As an effect, SNP-genotyping has become more accessible and therefore used more commonly in studies of wild species.

In this thesis, the development of SNPs for the wild deer species occurring in Sweden; moose, roe deer, red deer, fallow deer, and the semi-domesticated reindeer is presented. Two different approaches were taken to identify SNPs. First, the possibility that SNPs developed for distantly related bovines also show variation in deer was explored. Next, genomic information from moose was

retrieved by high throughput sequencing and scanned for new SNPs that could identify individuals and sex and also separate between the different deer species.

The SNPs were quality controlled, validated and subsequently utilized for retrieval of information about different relevant areas of application for wild animals and moose in particular. A large enough number of bovine SNPs were successful in both individual- and population based evaluations in all deer species. Hence, potential areas of application includes non-invasive studies of the deer populations' genetic status and, in combination with SNPs that separate species, the identification of individuals also in areas were different deer species co-exist.

To evaluate the new moose SNPs' performance on non-invasive samples, fecal pellets were analyzed with the purpose of estimating population size and sex ratio of moose on the island of Öland. As a result, the number of moose were found in the range of 115-156 individuals of which 37% were male during the winter of 2016. Moose on Öland also appeared to be somewhat different from moose on the Swedish mainland and a weak genetic differentiation of moose were found within Öland as well.

In the final study, shifts in genetic patterns between seasons were explored by using both GPS- and genetic data from moose that move between summer- and winter home ranges. The study showed that patterns of genetic structure appeared during certain times of the year, such as during the rut and calving, which might be explained by that moose return to the same areas between years, and that this behavior to some extent is transferred between generations. Thus, the combination of movement- and genetic data appeared as a useful strategy to study moose population dynamics.

As a conclusion, the SNP panels developed in this thesis allow both individual- and population based studies of deer occurring in Sweden. By using moose as a model system, this thesis explore how non-invasive genotyping and individually based genetics can be used to greatly improve our understanding of wild deer populations.

Populärvetenskaplig sammanfattning

Genetisk data av hög kvalitet är en ovärderlig tillgång för studier av vilda populationer eftersom denna innehåller information som kan vara svår att utvinna på andra sätt. Att samla in prover från vilda djur är ofta utmanande på grund av att de håller sig undan människan och ofta är utspridda över stora ytor. En lösning är därför att använda ”icke-invasiva” metoder, det vill säga att samla in till exempel spillning, saliv eller hår och utvinna DNA från dessa. Denna metod har även fördelen att djuren som studeras inte påverkas i lika hög grad eftersom de inte behöver hanteras av människor. En väletablerad metod för att erhålla genetisk information från DNA är att använda genetiska markörer. Dessa markörer kan användas för att ta fram ett mått på den genetiska variationen inom en grupp individer eller för att jämföra olika individer eller grupper med varandra. En annan mycket användbar tillämpning av genetiska markörer är identifiering av individer. Genom att skilja ut unika individer ur ett stort antal prover, och sedan undersöka hur ofta samma individ återfinns bland proverna, kan populationsuppskattningar göras. Ett annat användningsområde är att spåra kända individer i realtid för att undersöka till exempel hur släktskap mellan individer påverkar deras beteende.

DNA som insamlats med icke-invasiva metoder är ofta av bristande kvalitet och kvantitet, vilket kan försvåra genetiska analyser. En analysmetod som dock har visat sig fungera bra är så kallad SNP (single nucleotide polymorphism) genotypning. SNP:ar är positioner i DNA som varierar mellan individer eller populationer och när ett antal sådana variationer mäts samtidigt kan de användas för att studera en rad olika frågeställningar, både på individ- och populationsnivå. Dock har det länge varit väldigt resurskrävande att utveckla SNP:ar, eftersom stora mängder genomisk information måste produceras och bearbetas. Under det senaste decenniet har teknologin som används för att ta fram genomisk information, DNA sekvensering, utvecklats enormt vilket har fått till följd att SNP-genotypning blivit tillgängligt och alltmer använt även för studier av vilda arter.

I denna avhandling presenteras utvecklingen av SNP:ar för de hjortarter som finns i vilt tillstånd i Sverige; älg, rådjur, kronhjort och dovhjort, samt den delvis tämjda renen. Två olika tillvägagångsätt användes för att identifiera SNP:ar. Först utforskades huruvida SNP:ar framtagna i ett annat klövdjur; nötkreatur, även visade variation inom de svenska hjortarterna. Därefter utvecklades nya SNP:ar för älg, användbara för att framförallt identifiera individer, kön och art, genom att utforska stora mängder genomisk material. De utvecklade SNP:arna utvärderades noga innan de sedan användes för att inhämta information om olika aspekter relevanta för vilda djur i allmänhet och älgar i synnerhet. Det visade sig att en tillräcklig mängd SNP:ar från nötkreatur kunde användas för att identifiera individer och studera genetiska uppdelningar (populationsstruktur) i samtliga hjortarter. I kombination med SNP:ar som separerar arter, är ett tänkbart användningsområde studier av resursutnyttjande och mellanartskonkurrens i områden där flera hjortarter samexisterar.

För att praktiskt utvärdera SNP-analyser av icke-invasiva prover från älg, identifierades unika individer från spillningsprover, med syfte att uppskatta den Öländska älgstammens populationsstorlek och könskvot. Vinterspillning från älg visade sig vara relativt svåranalyserad, efter en grundlig kvalitetskontroll av det genererade datat kunde en populationsuppskattning göras som resulterade i 115-156 individer varav 37% tjurar under vintern 2016. Den Öländska älgen visade sig till en viss grad annorlunda från älg på fastlandet och en viss genetisk uppdelning kunde även ses inom den Öländska älgstammen.

Den sista studien i avhandlingen handlar om hur genetisk data kan användas till att utforska skiftningar i genetiska strukturer över olika säsonger. Detta studerades inom en del av älgstammen som till hög grad vandrar mellan sommar- och vinterhemområden. I denna studie användes både GPS- och genetisk data från samma individer, vilket visade sig vara en användbar metod för att inhämta information om populationsdynamik hos älg. Mönster av genetisk struktur framkom vid vissa tidpunkter under året, så som under brunst och kalvning. Dessa mönster kan bero på att älgar årligen återkommer till samma områden och att detta beteende i viss mån förs vidare mellan generationer.

Sammanfattningsvis innebär denna avhandling att SNP-baserade metoder för både individ- och populationsbaserade studier av samtliga svenska hjortarter nu är tillgängliga. Potentialen av dessa metoder är framförallt utforskad för älg och visar på ett brett användningsområde och goda förutsättningar för en ökad kunskap om hjortdjuren.

Acknowledgements

I have been lucky to be surrounded by many great people during the work with this thesis. I would first of all like to deeply thank my main supervisor Göran Spong for giving me this opportunity and for your scientific guidance along the way. You have been very supportive and I have appreciated your positive attitude and kindness. Sorry for not keeping up with all those calculations of probabilities over the years!

I am also grateful to my assistant supervisor, Göran Ericsson, thank you for being supportive and creating a positive working environment. Helena, you're a "rock" in the lab and I am thankful for your assistance and expertise (and for finding that data at a critical time point). Joakim, thank you for introducing me to the department many years ago by hiring me as a field assistant. Navinder, thank you for being encouraging and helpful.

Thank you to all fellow PhD students that I have had the pleasure of sharing this journey with. Anita, we shared an office for four years and I'm so happy that you're back, now as a researcher! Julia, I am glad that you became my next office mate, I know you have a bright future as a researcher. Lo, Hanna and Ellinor, life has changed over the years; less socializing with colorful drinks in the sun and more…kids. Anyhow, thanks for being the best and for your friendship over the years. And thank you Lo for your help and support during the final stages of this thesis. Henrik, I am glad that you're still around, thanks for squeezing in social activities in between work! Martijn, thanks for well needed input regarding the thesis work towards the end. Chris, thank you for the music!

I would not have been able to finalize this thesis without the help from my and Erik's families, thank you for taking care of the kids a lot lately. And Erik, thank you for letting me stay focused on work during this spring, I am forever grateful and forever yours. Einar and Emmy, you are without a doubt my greatest and dearest achievements during my time as a PhD student!