Genomic Studies of Contemporary Processes in Wild Populations

With the Scandinavian Brown Bear as a Model

Anita J. Norman
Faculty of Forestry
Department of Wildlife, Fish and Environmental Studies
Umeå

Doctoral Thesis
Swedish University of Agricultural Sciences
Umeå 2016

Acta Universitatis agriculturae Sueciae
2016:28

Cover: Genomic insights into wild populations with the brown bear in focus
(Art: Liza Le Roux)

ISSN 1652-6880
ISBN (print version) 978-91-576-8558-2
ISBN (electronic version) 978-91-576-8559-9

Print: Arkitektopia, Umeå 2016

Genomic Studies of Contemporary Processes in Wild Populations with the Scandinavian Brown Bear as a Model

Abstract

Genomic tools can greatly facilitate our understanding of wild populations. For the purposes of ecology and conservation, the most pertinent insights into wild populations are those that are contemporary. Much of the genetic-based research on wild populations has been derived from a population genetic framework resulting in historically derived summary statistics. These statistics are undoubtedly useful for understanding things such as effective dispersal and population structuring. However, they provide little indication to processes affecting populations within existing generations. One way to overcome this is to work at the individual level and consolidate the findings to improve understanding at the population level. For individual-based genetic studies, it is essential to be able to identify unique individuals and obtain reliable inferences of relatedness. Molecular markers must therefore possess qualities that make them suitable for identifying individuals and inferring relatedness between them.

This dissertation first describes the development of a set of 96 single nucleotide polymorphisms (SNPs) designed to infer relatedness between individuals in the Scandinavian brown bear population. The SNPs were used to study three contemporary features through relatedness inferences and pedigree reconstruction based on noninvasively collected samples: population size, natal dispersal distances, and fine-scale spatial structuring. These three studies are all based on new methods, one developed by Creel and Rosenblatt (2013) but empirically tested here, and the other two first developed for this dissertation. Using these methods, I successfully identified contemporary characteristics of a wild population. These methods can easily be applied to other species of ecological and conservation interest.

Keywords: single nucleotide polymorphism, SNPs, population size, census, natal dispersal, fine-scale population structure, isolation by distance, continuous population, noninvasive genetic sampling, citizen science

Author's address: Anita Norman, SLU,
Department of Wildlife, Fish and Environmental Studies, SLU
Skogsmarksgränd, 901 83 Umeå, Sweden
E-mail: Anita.Norman@ slu.se

Dedication

To Mom & Dad, with love.

Wherever you go, go with all your heart..
Confucious

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 **Norman AJ**, Street NR, Spong G (2013). *De novo* SNP discovery in the Scandinavian brown bear (*Ursus arctos*). *PLoS ONE 8*(11), e81012.

II Spitzer R, **Norman AJ**, Schneider M, Spong G. (2016) Estimating population size using single nucleotide polymorphism-based pedigree data. *Ecology and Evolution. (In Press)*

III **Norman AJ**, Spong G (2015). Single nucleotide polymorphism-based dispersal estimates using noninvasive sampling. *Ecology and Evolution* 5(15), 3056-3065.

IV **Norman AJ**, Stronen AV, Fuglstad G-A, Ruiz-Gonzales A, Kindberg J, Street NR, Spong G. Landscape relatedness: detecting contemporary, fine-scale spatial structure in wild populations. (*In Review*).

The contribution of Anita J Norman to the papers included in this thesis was as follows:

I Norman was largely responsible for the experimental design, laboratory work, bioinformatics, analyses, and manuscript preparation.

II Norman contributed to the experimental design, analyses and manuscript preparation.

III Norman was largely responsible for the experimental design, genotyping, analyses, and manuscript preparation.

IV Norman came up with the idea for this manuscript and was largely responsible for experimental design, analyses, interpretation, and manuscript preparation.

Abbreviations & Definitions

alleles	the nucleotide variants at a polymorphic locus
bp	base pairs
breeding dispersal	the movement of an individual to a new location for breeding purposes
CMR	capture-mark-recapture
CRE	Creel-Rosenblatt estimator
fitness	an individual's ability to survive and reproduce
F_{ST}	fixation index
gene flow	passing on of genetic variants to areas previously devoid of them
genetic	few molecular markers used in classical population genetic studies
genomic	100s or more genetic markers or whole genome analysis
genotyping	the identification of alleles within individuals
H_E	expected heterozygosity
H_O	observed heterozygosity
homozygosity	the existence of the same allele on both copies of the genome in a diploid organism
HWE	Hardy-Weinberg equilibrium
in silico	performed computationally
INLA	integrated nested Laplace approximation
linkage	SNPs that are inherited together due to their close proximity on the genome
locus/loci	location(s) in a genome
MAF	minor allele frequency
MCMC	Markov chain monte carlo
MOM	method of moments

natal dispersal	the movement of an individual away from the natal area for reproduction
panmictic	randomly mating and fully interacting population
PCR	polymerase chain reaction
philopatric	a behaviour where individuals settle within or close by their natal area
phylogeography	the interspecies and intraspecies geographic distribution
primers	sequence of nucleotides that occur before or after a target locus/loci
r-value	Lynch-Ritland relatedness coefficient
read	computational interpretation of a genomic DNA sequence
SNP	single nucleotide polymorphism

1 Introduction

1.1 The Biodiversity Crisis

Global biological diversity (biodiversity) is declining rapidly as a consequence of anthropogenic influence. Biodiversity, by definition, includes several components: genetics, species, ecosystems and the processes and interactions within (Huston 1994). Thus, biodiversity loss includes everything from the reduction of genetic variation and extinction of species to degradation of ecosystem function. The integrated nature of all three components means that the loss or change of one biological property will often have a ripple effect that permeates and affects other properties. There are several prominent threats to biodiversity including climate change, habitat destruction, species invasions, overexploitation, and environmental toxins (Groom et al. 2006). These threats are so pervasive that we are evidently in the throes of a sixth mass extinction with the current extinction rate estimated at 100 times the natural rate (Ceballos et al. 2015). Besides the direct effect on species and ecosystems, biodiversity loss will also impact humans in ways unprecedented as human life on earth depends on ecosystem functions such as hydrological cycles. It is therefore in our interest to do all we can to mitigate biodiversity loss.

1.2 Importance of Genetic Diversity in Wildlife Populations

1.2.1 Population Size

Genetic diversity operates at the most basic level of biodiversity and is optimally assessed at the population level (Luck et al. 2003). Genetic diversity in a population facilitates adaptation in changing environments. Without genetic variation, populations subjected to changes in their environment will exhibit reduced overall **fitness**. To explain this further, individuals in healthy populations are thought to invest their energy expenditure first in surviving and then reproducing (Stearns 1992). If conditions are sub-optimal, individuals spend more of their energy surviving and less reproducing. Those that successfully reproduce are likely to have genetic variants that are beneficial to

the altered environment allowing them more energy for reproduction. These beneficial gene variants are then passed on to the next generation and thus spreads in the population (Darwin 1859). It is natural selection at play. However, if genetic diversity is low, there is less of a chance that individuals will have these beneficial gene variants. Furthermore, when populations begin to have reduced genetic diversity, they become more prone to the negative effects of deleterious recessive mutations due to increased **homozygosity** (Lande 1988). Consequently, both reproduction and survival may be reduced, thereby decreasing population size. Small population sizes are prone to inbreeding as well as stochastic effects, mainly genetic drift, leading to a further reduction in genetic diversity, exacerbating the problem. This is a well-known phenomenon that is often referred to as the extinction vortex (Gilpin & Soulé 1986). The size of the population is therefore an important issue in conservation.

1.2.2 Dispersal

Dispersal, whether at the individual or population level, plays an important role in maintaining genetic diversity. When environmental changes occur, forcing individuals in populations to put more energy into surviving, one key strategy is for individuals to disperse to a new location where conditions are more favourable for survival and reproduction. This is referred to as **breeding dispersal** (Matthysen 2012). Another more common type of dispersal, **natal dispersal**, occurs when young leave their birth area to reproduce (Matthysen 2012). Natal dispersal strategies are said to have developed for a multitude of reasons, but the leading hypothesis is to avoid mating with kin (Lawson Handley & Perrin 2007). Often populations will have a sex-biased dispersal strategy where one sex will disperse far from the natal site while the other sex remains close leading to a large distance between the two opposite-sexed kin thereby reducing the chance that they will mate (Pusey 1987). Another central explanation for dispersal is to reduce competition amongst kin of opposite sexes (Hamilton & May 1977). Not only is dispersal important for avoiding inbreeding, it is the mechanism behind gene flow (Slatkin 1987). **Gene flow** is the passing on of genetic variants to individuals in areas previously devoid of these genetic variants, thereby increasing genetic diversity and fitness.

1.2.3 Population Structure

When gene flow becomes restricted for some reason, populations will become genetically structured over time. Genes will be maintained within the group of individuals that have access to each other, but not to groups where the access is quite limited or cut off completely. These groups, identified as subpopulations, will become genetically differentiated from each other. The longer the time one subpopulation is isolated from other subpopulations, the more differentiated it becomes as new mutations are introduced but are not passed to other groups. In

addition, genetic drift causes genetic variants to become lost or fixed at random in the population further differentiating the subpopulations from each other. If a population becomes structured due to, for example, habitat fragmentation, there becomes an increased risk that, if isolated enough, smaller subpopulations enter into an extinction vortex. This, in turn, would affect the population as a whole by reducing its overall genetic diversity. Ensuring that subpopulations have connectivity between them is therefore an important conservation priority.

To summarise, genetic diversity is a fundamental concept in conservation biology. Population size, dispersal, and how populations are structured in the landscape are all key issues affecting genetic diversity. Conservation genetics is a scientific field that aims to build knowledge that can be applied to the prevention or reduction of loss of genetic diversity in wild populations. The most direct way to build knowledge about genetic diversity is to study the genetics (genomics) of individuals in a population to assess levels of genetic diversity within the population. However, the genetic profile of individuals can also be used indirectly to understand population processes that can aid in reducing loss of genetic diversity. This can include everything from the species level, including taxonomic delineations, species divergence patterns and how species form in the landscape (**phylogeography**) to the population level, including identifying the extent of inbreeding or outbreeding, identifying dispersal strategies and gene flow, detecting population structure, estimating population and effective population size and, finally, to individual-based analysis such as reproductive success, relatedness with other individuals, individual-based dispersal and migration.

1.3 Noninvasive Sampling

In order to study genetics within wild populations, a DNA sample must be obtained. This can be done noninvasively to avoid negatively affecting individuals under study. However, working with samples collected noninvasively can be challenging. Often, DNA in samples which are collected from the environment and have been exposed to UV radiation, time lapse, high temperatures, moisture, and sources of contamination, become degraded, thereby hampering DNA extraction and analysis (Taberlet & Luikart 1999; Waits & Paetkau 2005). Additionally, collecting noninvasive samples can be logistically challenging. If a population is widespread across the landscape, sampling will necessitate much ground to be covered, requiring many people and much time. Individuals can also be elusive making it difficult to locate their samples, thus requiring expertise and time. This is where citizen science can be of great help: Not only do volunteers become involved in conservation programs, researchers receive the benefit of having more people involved in

collecting samples, which reduces the cost and time that would otherwise be needed.

1.4 Aims

In this dissertation, I present a new genomic tool for study within the Scandinavian brown bear (*Ursus arctos*) population, which offers several advantages over other commonly used tools. Using this tool, I present new methods using noninvasively, citizen-collected samples to further understand contemporary population processes such as population size, dispersal, and spatial structure. While this work is focused on the brown bear, the intention is that these methods can be used for other species to aid in conservation. My main objectives are as follows:

1. Develop a panel of 96 **single nucleotide polymorphism (SNP)** markers useful for inferring relatedness and ascertained throughout the entire Scandinavian brown bear population.
2. Empirically estimate population size based on pedigree reconstruction using a recently developed method by Creel & Rosenblatt (2013).
3. Derive precise estimates of individual natal dispersal distance and mean natal dispersal distances for males and females.
4. Identify contemporary, fine-scale spatial structure, relatedness patterns, and population heterogeneity in continuously distributed populations.

2 Background

Conservation of wild populations is greatly facilitated by an in-depth understanding of the species being conserved including its life history, behaviours, population characteristics and processes, and evolutionary history (Sæther et al. 1996; Frankel 1974). With the exception of a species' evolutionary history, the more contemporary the knowledge obtained is, the better the conservation potential (Palsbøll et al. 2010; Vucetich & Waite 2003). Additionally, regular monitoring of populations of conservation concern is necessary to detect vital changes that may affect their viability (Nichols & Williams 2006). Building contemporary knowledge of populations can be conducted in several ways: direct observation, radio- or GPS-tracking, camera trapping, or through genetic sampling. All have their advantages and disadvantages and the best approach is often to combine two or more of these. The focus in this dissertation is to develop and test methods aimed at building a contemporary understanding of populations through noninvasive genetic sampling, with a focus on individual identification and inference of relatedness between individuals in a population.

The terms genetics and genomics in wildlife studies have much overlap and can therefore be confusing. Genetics often refers to classical population genetics studies that are based on a few molecular markers (Ouborg et al. 2010). The last decade has witnessed the rise of genomics, which refers to studies that use whole genomes or many genome-wide markers (Allendorf et al. 2010). The absolute difference between the two terms is arbitrary and not well-defined. Here, my work is based on more than a few molecular markers (96), which are representative of the whole genome. However, it is not the thousands or tens of thousands that genomics often refers to suggesting that this work falls in the grey area between genetics and genomics. Consequently, I use it interchangeably throughout this dissertation.

2.1 Relatedness

Several questions can be addressed with relatedness estimates and genetically-based pedigree reconstruction. One of the key issues in conservation, particularly for small populations, is the risk that inbreeding results in the reduced biological fitness of a population, otherwise known as inbreeding depression. Inbreeding results from related individuals reproducing, thus it follows that relatedness analyses can provide direct insights into levels of inbreeding. Indeed, simulation and empirical studies have confirmed the importance of relatedness estimates used to detect inbreeding (see Santure *et al.* 2010; Robinson *et al.* 2013; Wang 2015). Through pedigree reconstruction, questions pertaining to population size (e.g. Creel & Rosenblatt 2013), effective population size (e.g. Cronin *et al.* 2009), captive breeding (Russello & Amato 2004; Putnam & Ivy 2014), reproductive success (e.g. Spong *et al.* 2008; Araki *et al.* 2009; Patzenhauerová *et al.* 2013) and mating behaviours (e.g. Serbezov *et al.* 2010; Kanno *et al.* 2011) can also be examined. Assessing natal dispersal is another key process for which knowledge of relatedness is indispensable (e.g. Pardini *et al.* 2001; Spong & Creel 2001; Qi *et al.* 2013). Knowledge of relatedness has also been useful for determination of genetic structuring in populations (e.g. Morin *et al.* 2009; Palsbøll *et al.* 2010). These are just a few of the many examples of how relatedness can provide insights on important conservation questions.

2.2 Single Nucleotide Polymorphisms

Single nucleotide polymorphisms (SNPs) are single DNA nucleotide differences originating from ancestral mutations (Figure 1). They are common throughout the genome and can be assayed and used as a type of molecular marker. SNPs have only recently (the past 10 years or so) been adopted as a marker of choice for many types of studies due to advances in technology. The onset of high-throughput sequencing enabled the discovery of genome wide SNPs, which were previously too time consuming and prohibitively expensive to detect. Since one SNP contains less information (typically only two **alleles**) relative to more allele-rich markers such as microsatellites (also known as single tandem repeats or simple sequence repeats), more SNPs are required to obtain the same or better statistical power as microsatellites (Liu et al. 2005). However, obtaining enough SNPs is no longer an issue. In fact, the genome wide representativeness of SNPs is ideal for many types of studies as it minimizes potential genomic biases.

Besides having high genomic resolution, SNPs have many other advantages. Their biallelic nature (i.e. two alleles) makes **genotyping** (the process of determining the SNP variant(s) present within an individual's genome) easier and less error prone than other markers. They are ideal for genotyping samples with degraded DNA, such as with noninvasively collected samples and ancient DNA, because only short DNA fragments (< 100 **base pairs**) are needed. Furthermore, unlike microsatellites, SNP genotyping is directly comparable across laboratories (Seeb et al. 2011).

For these reasons, SNPs are an ideal choice for studies that aim to infer genetic relatedness among individuals within a population (Tokarska et al. 2009; Hauser et al. 2011). With the careful selection of SNPs, it is quite possible to infer relatedness between individuals with a small panel (e.g. 100) of SNPs rather than thousands to hundreds of thousands that are used in other types of studies (Krawczak 1999). Ensuring that each SNP is contributing information that is independent of other SNPs (e.g. SNPs that are not linked with each other) and that each one provides the maximum amount of information for the population will result in a highly informative panel for inferring relatedness between individuals (Anderson & Garza 2006; Thompson 1975).

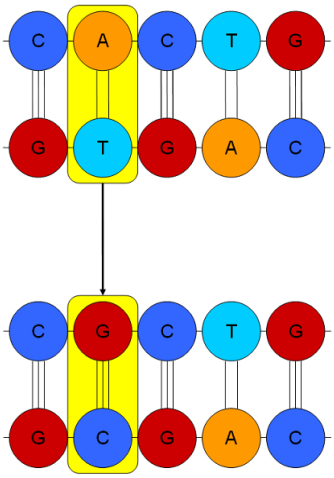

Figure 1 Graphic interpretation of a single nucleotide polymorphism (SNP). Two fragments of a double-stranded genome containing five base pairs, one of which differs between the two fragments and is highlighted in yellow – this is the SNP.

3 Model System and Methods

For the development of new methods, it is advantageous to use an appropriate model system. The Scandinavian brown bear is a system that is well researched, and is sampled at high resolution and monitored regularly. Thus, with this system, there is enough background knowledge to test new methods and to develop methods that provide novel information. Furthermore, the development of a panel of SNPs for the Scandinavian brown bear can prove to be of considerable value for regular monitoring schemes using noninvasive genetic sampling as well as answering research questions that require high levels of discrimination between individuals such as inferring relatedness and pedigree reconstruction.

3.1 Scandinavian Brown Bear (All Papers)

In Europe, two major lineages of brown bear exist: the eastern European lineage and the western European lineage (Taberlet & Bouvet 1994). Since the last glacial maximum, these two lineages colonized Sweden from two different routes: the eastern lineage entered from the northeast through Finland, and the western lineage entered from the south through Denmark (Bray et al. 2013). Contemporary populations continue to display these historical patterns as evidenced through genetics: While some male-mediated hybridization between the two lineages is evident, the eastern European lineage occurs distinctly in the northern part of the country and consists of two subpopulations and the western European lineage occurs in the south central part of Sweden (Manel et al. 2004). The two lineages are separated by a contact zone running through the county of Jämtland (Taberlet et al. 1995) (Figure 2).

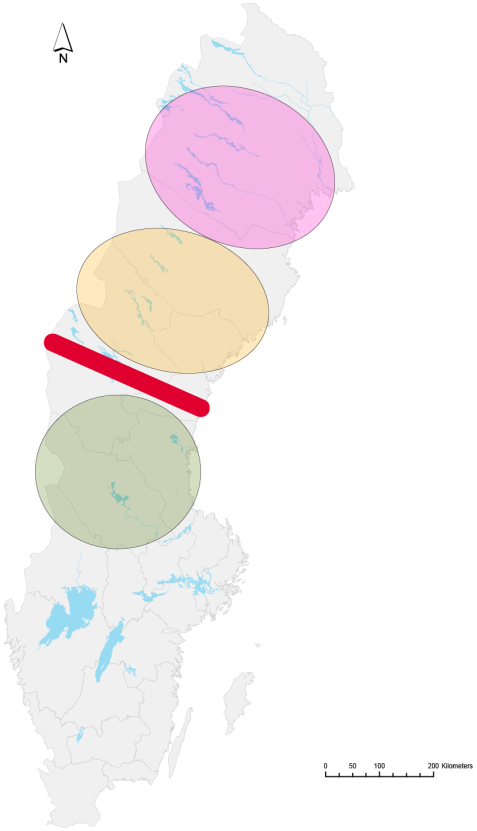

Figure 2 The three genetically identified subpopulations of brown bear in Sweden. The two northern subpopulations belong to the eastern European lineage and the southern subpopulation belongs to the western European lineage from Iberia. The red line shows the contact zone separating the two lineages.

There has been extensive research conducted by the Scandinavian Brown Bear Research Project (SBBRP) in the south central subpopulation (the western European lineage). Consequently, much is known about this subpopulation making it an ideal model system to test new methods. However, it is often the case that the more that is known about a system, the more questions arise. In addition, it is a population of conservation concern as it represents one of the few remaining relics of the western lineage in Europe and has a unique genetic structure (Bray et al. 2013). An ongoing monitoring program is organised by the county administration boards (Länstyrelsen) for estimating population size on a county-by-county basis. Within this program, faecal samples are collected by volunteers who are already out in the field participating in activities such as moose hunting. These samples are sent to laboratories to be analysed. Consequently, population size has been estimated across the country multiple times providing good baseline data for testing a new method of obtaining population size estimates (see Bellemain et al. 2005; Kindberg et al. 2011). Støen et al. (2006) conducted a study looking at dispersal in the south central subpopulation using radio-collared individuals, providing the basis for verifying the results of new dispersal estimation models. Several studies have looked at population substructuring within Sweden (Taberlet & Bouvet 1994; Waits et al. 2000; Manel et al. 2004) leading to the knowledge that three major subpopulations exist (Figure 2) (Manel et al. 2004).

Brown bears are mainly solitary except during mating season and when there is young (Sandell 1989). Mating occurs during the late spring and early summer months and both sexes are considered promiscuous, meaning they mate with several individuals (Bellemain et al. 2006). Only females provide parental care, which typically lasts one to three years (Dahle & Swenson 2003). Natal dispersal is sex-biased with males dispersing at higher rates and further distances than females (Støen et al. 2006). One study identified only 41% of female brown bears as having dispersed whereas the remaining were **philopatric** (Zedrosser et al. 2007).

3.2 SNP Development (Paper I)

The development of the panel of SNPs was performed in several phases (Figure 3). The first involved detailed calculations to determine the optimum sequencing parameters, primarily taking into account estimated genome size, number of individuals to sequence and the total number of sequences. The second phase involved laboratory preparation including DNA extraction, restriction enzyme digestion and clean-up. The third phase was the size selection and sequencing performed by SciLife (Stockholm, Sweden). Fourth was the bioinformatics component where I filtered for target, high quality sequence reads, aligned the reads to each other, detected putative SNPs and applied strict filtering of the putative SNPs to minimize false positives and linkage with other SNPs. This involved aligning the putative SNP reads to a draft genome (A. Janke pers. comm.) that I obtained part way through this project (Figure 4). The fifth phase was the validation, which involved using the selected SNPs to genotype 68 individuals to ensure the SNPs worked. The sixth and final phase involved analysis of the validation genotypes to select the final panel of SNPs with properties that provide high discrimination between individuals including a high overall **minor allele frequency (MAF)** and independently segregating **loci**.

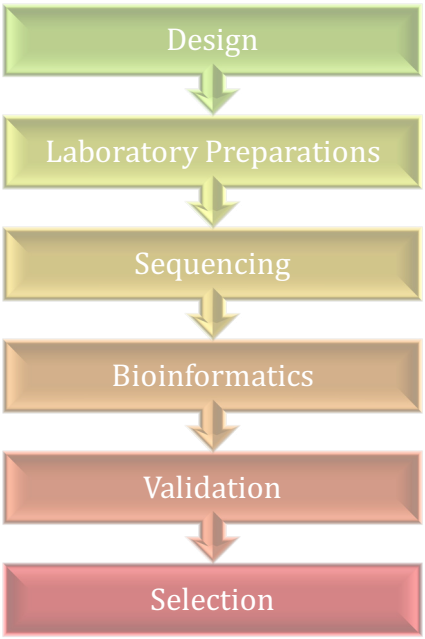

Figure 3 The six phases of developing a panel of SNPs.

Figure 4 Each step of the filtering process and the number of SNPs remaining are shown in sequence. Putative SNPs were identified through Stacks software. The files generated through Stacks were used in the filtering process and are denoted with ST (top row) and were performed in parallel (independently). The orange boxes indicate filtering criteria that were applied using the software blastn and the draft genome assembly.

In addition to the above, I worked to find markers on both the mitochondrial genome and the Y chromosome. For the mitochondrial markers, I designed **primers** from a previously published mitochondrial genome. For the Y-chromosome, I identified primers that had previously been published (Hellborg & Ellegren 2003). I then performed **PCR** to obtain many copies of the full sequence and sent the results for sequencing. Once I received the sequences, I searched for SNP variants and included four mitochondrial variants and two monomorphic Y chromosome markers in the validation step as described above.

3.3 Sample Collection and SNP Genotyping (Papers II to IV)

Collection of samples was organized by the local county administration boards on a county-wide basis for the purpose of conducting a population census.

Faecal samples were collected from Dalarna and Gävleborg counties during a 12-week period in the autumn of 2012 and were sent to Bioforsk (now NIBIO, Svanhovd, Norway) for analysis. Samples collected represented 434 unique individuals. Out of these, one DNA extract per individual was sent to our laboratory for SNP genotyping. With one sample excluded due to duplication, we ended up with 433 unique individuals, of which 412 had spatial data. The official population estimate for this area was 793 (95% CI: 621-1179) (Kindberg & Swenson 2013). Papers II, III and IV used the genotypes from the Dalarna-Gävleborg data.

A second sample collection was organized by the Västerbotten county administration board, which took place in the autumn of 2014. Faecal samples were sent to our laboratory (SLU, Umeå) and were SNP genotyped directly after DNA extraction. A total of 271 individuals were identified and the population size was estimated to be 362 (95% CI: 310-459) (Kindberg & Swenson 2015). Paper II used genotypes from the Västerbotten data.

SNP genotyping was performed on the Biomark system (Fluidigm, San Francisco, USA) using a 96 (samples) x 96 (SNP assays) plate. Results were analysed using the Biomark software (Fluidigm, San Francisco, USA).

3.4 Relatedness (Papers II to IV)

Several methods exist for inferring genetic relatedness between individuals including **method of moments (MOM)** or likelihood-derived coefficients of relatedness, kinship coefficients, estimating portions of the genome that are identical-by-descent, and through reconstructing pedigrees (Robinson et al. 2013). MOM coefficients of relatedness are commonly used and easy to calculate without much computational input like likelihood methods. Several studies have tested the performance of the various coefficient of relatedness methods and the one that performs best in most situations is the MOM method developed by Lynch and Ritland (1999) (Thomas 2005; Csilléry et al. 2006; Robinson et al. 2013).

The Lynch-Ritland coefficient of relatedness (**r-value**) is calculated pairwise between two individuals and can be approximately equated to the proportion of the genome shared between the pair due to common recent descent. As such, it gauges the level of gene sharing that exceeds expected levels by chance given the population allele frequencies. Table 1 shows the r-value associated with the categorical relationship (e.g. parent-offspring). Unrelated individuals can have r-values near zero and below. Negative values indicate a lesser degree of relatedness than what would be expected under **panmictic** conditions, where individuals are well mixed throughout the population and are randomly mating.

Table 1 Classification of relationships and their expected relatedness coefficient value (r-value). The order refers to the number of steps between the pair in their pedigree.

Relationship	r-value	Classification
Parent-Offspring/ Full-Siblings	0.50	First Order
Half-Siblings/ Grandparent-Grandoffspring/	0.25	Second Order
Aunts/Uncles-Nieces/Nephews	0.25	Third Order
Unrelated	0.00	Unrelated

3.5 Pedigree Reconstruction (Papers II and III)

Reconstructed pedigrees were estimated from the SNP genotype data using FRANz software (version 1.9.999 and 2.0; Riester et al. 2009), which incorporates a Markov Chain Monte Carlo (**MCMC**) approach, where probability distributions are computed. Genotyping errors were included in the run for each study area and were calculated as the number of genotyping mismatches divided by the total number of duplicated genotypes. Parameters for maximum number of females and males in the population were also included, however, different calculation methods were used for the two studies: For paper II the maximum number of females and males were derived from a rarefaction analysis combined with the sampled ratio of females to males. Rarefaction analysis was used to mimic scenarios where no prior population estimate was available. For paper III, the official population estimates were combined with the sampled ratio of females to males. Paper III focused only on the resulting parent-offspring pairs that had a posterior probability equal to or greater than 0.95 to ensure highest accuracy in dispersal estimations.

3.6 Dispersal (Paper III)

It is well documented that (sub)populations become more genetically distant with greater geographic distance, even if the (sub)population is continuous (isolation by distance). This can be understood as a function of dispersal since genes can only be carried as far as individuals carrying them move (and reproduce). Evidence of isolation by distance thus indicates that dispersal distances are mostly contained within the geographic region under study. I tested for isolation by distance in the Dalarna-Gävleborg samples using a Mantel test for: all individuals, females only and males only. I identified

mother and offspring pairs from the reconstructed pedigrees. Estimates for dispersal were then calculated as the Euclidean distance between inferred mother and offspring pairs (Figure 5). To further our understanding of population-based dispersal, I calculated the mean dispersal distance for all offspring, females only, and males only. I then used a Wilcoxon Rank Sum Test to determine if the differences between males and females were statistically significant.

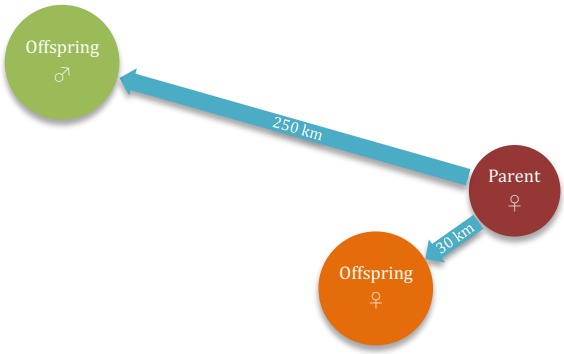

Figure 5 Schematic of dispersal estimation technique.

3.7 Landscape Relatedness (Paper IV)

For the final study, my aim was to identify possible fine-scale population structure within the continuously distributed population of brown bears in Dalarna-Gävleborg. The SNPs for the brown bear were developed to be highly discriminatory for individual identification with high allele frequencies. However, population structure analysis based on genetic differentiation requires molecular markers with medium or low minor allele frequencies to help distinguish geographically localized alleles. Therefore, using these SNPs may pose a risk of obtaining a false negative result (i.e. no population structure detected when population structure actually exists) since the SNPs were ascertained to be highly variable throughout the study area and would thus smooth over any underlying structure. I therefore devised a method that maximizes the utility of the SNPs. Instead of using the SNPs to detect genetic differentiation between areas, I used genetic relatedness and interpolated them across the landscape in order to detect non-uniformity in the spread of relatedness and thereby detecting fine-scale structuring. This was done using a statistical method called **integrated nested Laplace approximations (INLA)**.

The interpolations were done at the full scale of the study area (i.e. all of Dalarna-Gävleborg) referred to as the global area as well as smaller areas referred to as local areas. Heat maps were created to show the levels of relatedness across the landscape with red showing high degrees of relatedness to blue showing low degrees of relatedness.

4 Summary of Results

4.1 SNP Development (Paper I)

4.1.1 Sequencing

The sequencing resulted in 20 billion gigabytes of data. Unfortunately, only 30% was targeted DNA. This was due to highly degraded DNA being broken down into smaller fragments that happened to be in the target size range. Nonetheless, the 30% of targeted sequences were of high quality and were therefore suitable for detecting SNPs. The average **read** depth per individual ranged from five to eight reads, much lower than our estimate of 38. Table 2 shows the estimated parameters used in the design relative to the actual parameters that were obtained through an ***in silico*** restriction enzyme digestion. The estimates and the actual parameters are relatively close thereby confirming that one can estimate these parameters reliably without having prior knowledge of the target species.

Table 2 Estimated genomic parameters versus actual as determined through an in silico digestion with the draft genome assembly.

	Avg Frag Size (bp)	Genome Size (billion)	# Unique Fragments	Read Depth	Genomic Coverage (%)
Estimated	3100	2.4	131,910	38	1.10
Genome	3465	2.3	93,678	53	0.82
% Diff	-11.8	5.1	29.0	-28.3	25.5

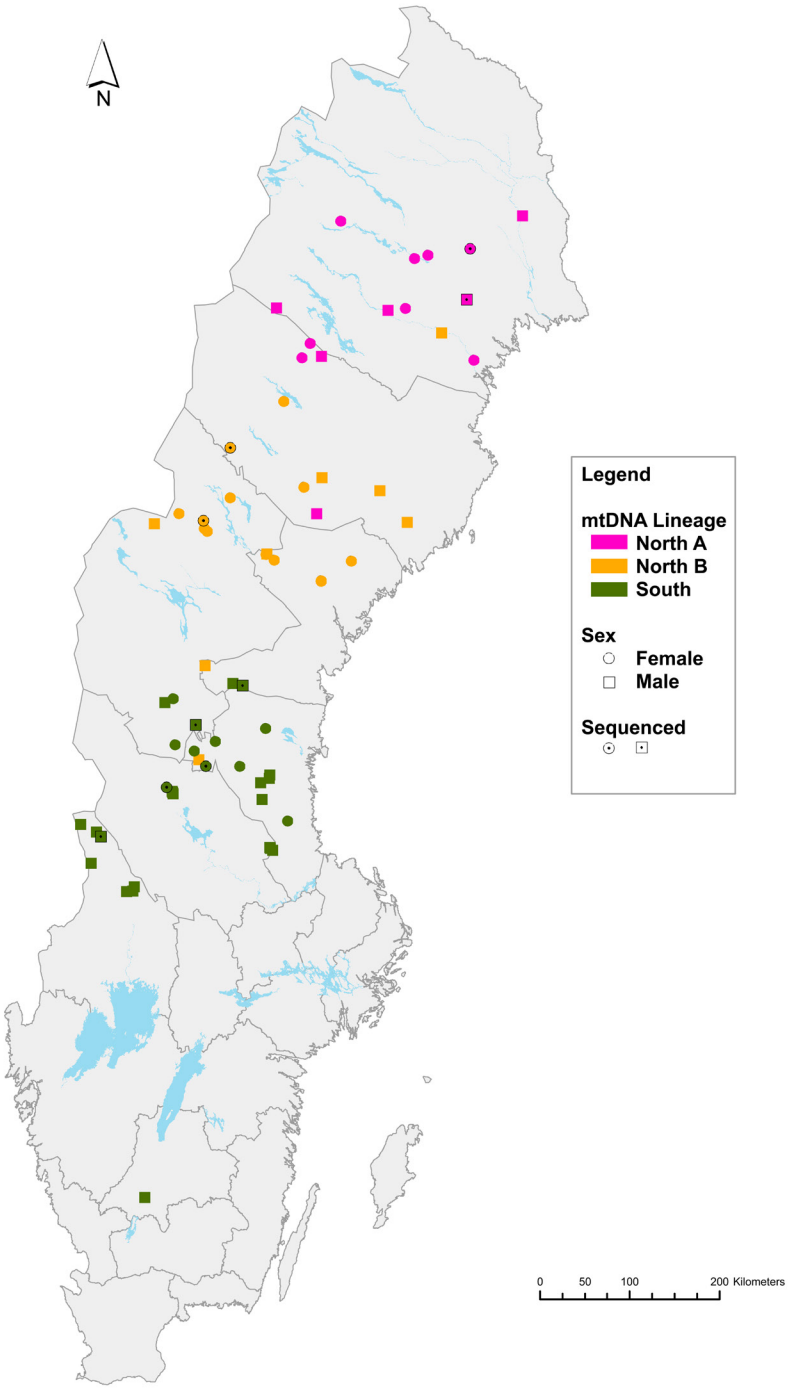

Figure 6 Map of Sweden with the sampling locations of brown bears, both that were sequenced and that were used for SNP validation. Colours indicate which subpopulation the bears belonged to based on their mitochondrial haplotype.

4.1.2 SNP Selection

A total of 96 SNPs were selected for the final SNP panel. The original panel included 87 autosomal SNPs, four diagnostic mitochondrial markers (informative for subpopulation origin), two Y chromosome markers (for sex identification) and three X chromosome SNPs (to help with sex identification). After testing for **linkage** (non-independence) between the SNPs, two SNPs were found to be linked and were subsequently withdrawn from the panel to avoid redundancy. These two SNPs were replaced by two additional Y chromosome SNPs that were identified and provided to us by Bidon et al. (2015). One of these Y chromosome SNPs is polymorphic in the Scandinavian brown bear and can thus be used for paternity analysis.

4.1.3 SNP Validation

The SNPs were validated on a panel of 68 brown bears spanning the geographic range in Sweden (Figure 6). Summary statistics are presented in Table 3. All SNPs were in **Hardy Weinberg equilibrium (HWE)** with the exception of three; two in the southern population (p-value = 0.0217, 0.0087) and one in the northern population (p-value = 0.0162). Since a few SNPs are expected to be out of HWE due to chance alone, I opted to keep these in the SNP panel. Indeed, a later analysis of genotypes from the southern population revealed these same SNPs to be within HWE.

Table 3 Summary statistics for the final panel of SNPs. MAF = Minor Allele Frequency; $\boldsymbol{H_E}$ *and* $\boldsymbol{H_O}$ *= expected and observed heterozygosity respectively;* $\boldsymbol{F_{ST}}$ *is the genetic differentiation calculated based on the three mitochondrial haplotypes (i.e. three subpopulations).*

	MAF	H_E	H_O	Overall F_{ST}
Overall Mean	0.39	0.47	0.44	0.08
Standard Deviation	0.06	0.03	0.07	0.10

4.2 SNP Genotyping (Papers II to IV)

All but one individual aliquot was successfully genotyped from the Dalarna-Gävleborg samples. Of these, I identified 243 females and 190 males. The mean MAF was 0.373 (SD=0.0922). Västerbotten samples were genotyped and 265 individuals were identified of which 136 were females and 129 were males. The mean MAF was 0.366 (SD=0.0915)

4.3 Relatedness (Papers II to IV)

Lynch-Ritland relatedness values were calculated for all pairs of individuals in Dalarna-Gävleborg and Västerbotten. In Dalarna-Gävleborg, there were a total of 93,528 pairwise comparisons resulting in a mean r-value of -0.0023 (SD=0.13). Paper III used a subset including 132 individuals which resulted in a mean r-value of -0.0003 (SD=0.14). Paper IV included only individuals with spatial data associated with them. From 412 individuals, the mean r-value was 0.00 (SD=0.13). Each subset was based on the same r-values calculated from the total numbers of individuals sampled. Paper II used a subset of Dalarna-Gävleborg and Västerbotten individuals including only those identified as first-order relatives based on the pedigrees. This resulted in a total of 294 parent-offspring pairs with a mean r-value of 0.50 (SD=0.10) and 40 full sibling pairs with a mean r-value of 0.53 (SD=0.11).

4.4 Pedigree Reconstruction (Papers II and III)

For the software to estimate correctly which individuals are related and what their relations are, it requires a good estimation of the number of errors that may be present in the genotyping. These were entered as 1.538×10^{-4} for Dalarna-Gävleborg and 0.01 for Västerbotten. The difference in error rates arose due to differences in the quality of DNA extracts used. Additionally, the software needs to have an idea of the ratio of females to males. The maximum number of females to males in Dalarna-Gävleborg and Västerbotten study areas were entered as 538:419 and 249:239 respectively. Mean posterior probabilities for parent-offspring inferences were 0.79 (SD=0.22) and 0.73 (SD=0.23) for Dalarna-Gävleborg and Västerbotten respectively. Figure 7 shows how the relatedness values correlate with the pedigree results from the Dalarna-Gävleborg individuals.

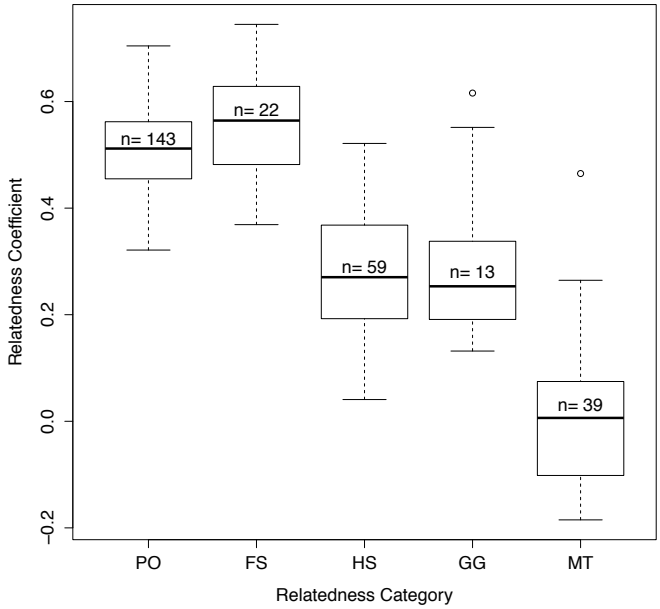

Figure 7 Correlation between the calculated relatedness coefficient and the pedigree categories: parent-offspring (PO), full siblings (FS), half siblings (HS), grandparent-grandoffspring (GG), and mates (MT). PO and FS are considered first-order relatives and are expected to have an r-value around 0.50. HS and GG are second-order relatives with and expected r-value around 0.25. Mates should be unrelated with an expected r-value around 0.00.

4.5 Population Census (Paper II)

The number of samples collected are known, but the number of individuals that were missed are unknown. By reconstructing the pedigrees, we can detect some of these missing samples. Reconstructed pedigrees revealed a total of 115 unsampled individuals from the Dalarna-Gävleborg and 85 from Västerbotten subpopulations. After accounting for mortality, the population estimates were 630 and 408 respectively. These fall within the 95% confidence interval of the official DNA-based capture-recapture estimates (Figure 8).

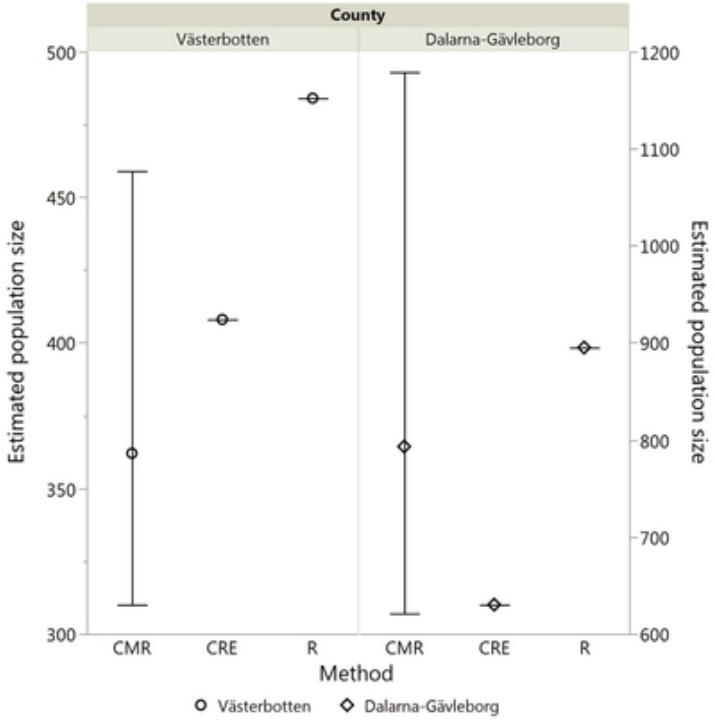

*Figure 8 Comparison of the official population estimates based on capture-mark-recapture (**CMR**; 95% confidence intervals), the pedigree-derived population estimates (**CRE**), and the rarefaction estimates (R).*

4.6 Dispersal (Paper III)

The Mantel test showed significant isolation by distance (P<0.001) for all individuals and for females, but not for males (P=0.080) (Figure 9). This indicates that for females, dispersal distance estimates are likely to be representative of the true dispersal distance. For males, on the other hand, mean dispersal distances are likely to be underestimated. Mean dispersal distances were estimated to be 12.9 km (SD=11.7 km) for females and 33.8 km (SD: 33.9 km) for males (Figure 10). These estimates are significantly different (p-value=0.02) which corroborates previous findings that brown bears exhibit male-biased dispersal.

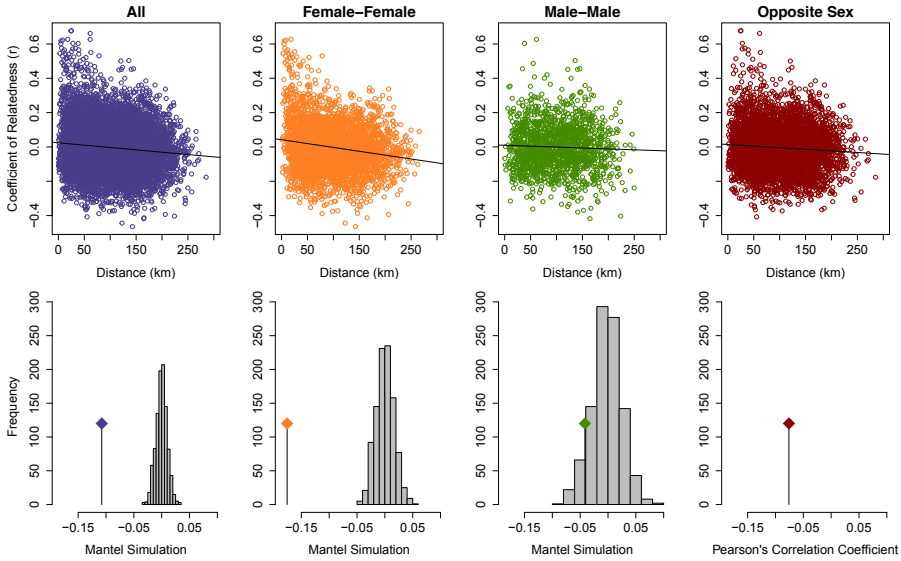

Figure 9 Top row shows scatterplots of relatedness versus Euclidean distance. The steeper the trendline, the greater the chance that isolation by distance (IBD) exists. The second row shows the results of the test for IBD (Mantel). IBD is present if the sample statistic (the vertical line with the coloured triangle on top) does not overlap with the simulated frequency histograms.

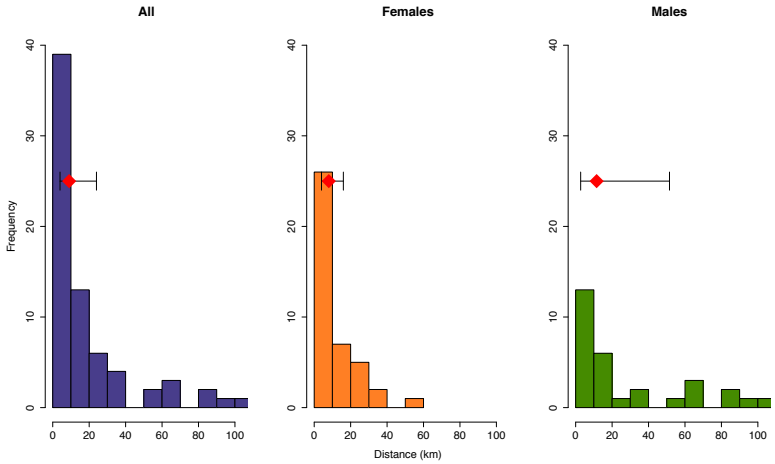

Figure 10 Frequency histogram showing the pedigree-based estimates of dispersal distance for: all mother-offspring pairs (n=63); mother-daughter pairs (n=38); and mother-son pairs (n=25). The red diamonds represent the median distance and the lines extending from the diamonds show the interquartile ranges.

4.7 Landscape Relatedness (Paper IV)

The interpolations of relatedness values across the global area revealed a low degree of relatedness throughout. There were two areas in the northern region where the individuals were significantly less related to the population as a whole than expected by chance alone (Figure 11). Focusing on these areas revealed a high degree of relatedness within (Figure 12). This is strong evidence that these two areas are segregated from the rest of the population. The cause of this population structuring is presently unknown. Furthermore, a comparison of relatedness values across the three local areas suggest that inbreeding may have occurred in at least one of these segregated areas since the mean relatedness value was significantly higher than for the other areas (Table 4).

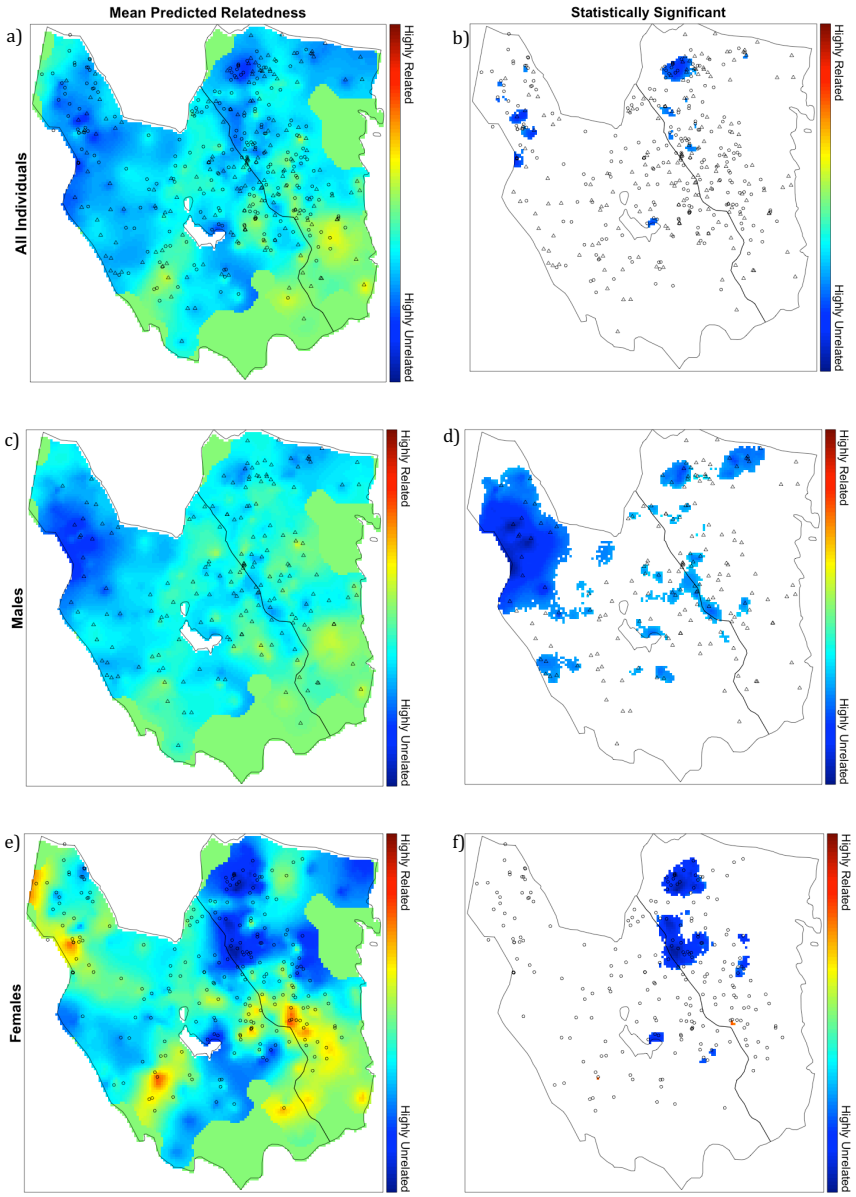

Figure 11 The first column shows interpolations of relatedness for the entire study area for pairwise relatedness of a) all individuals to all individuals, c) males to all, and e) females to all. The second column shows areas of statistical significance, meaning that individuals in these areas are significantly more (if red) or less (if blue) related to the population as a whole than expected by chance.

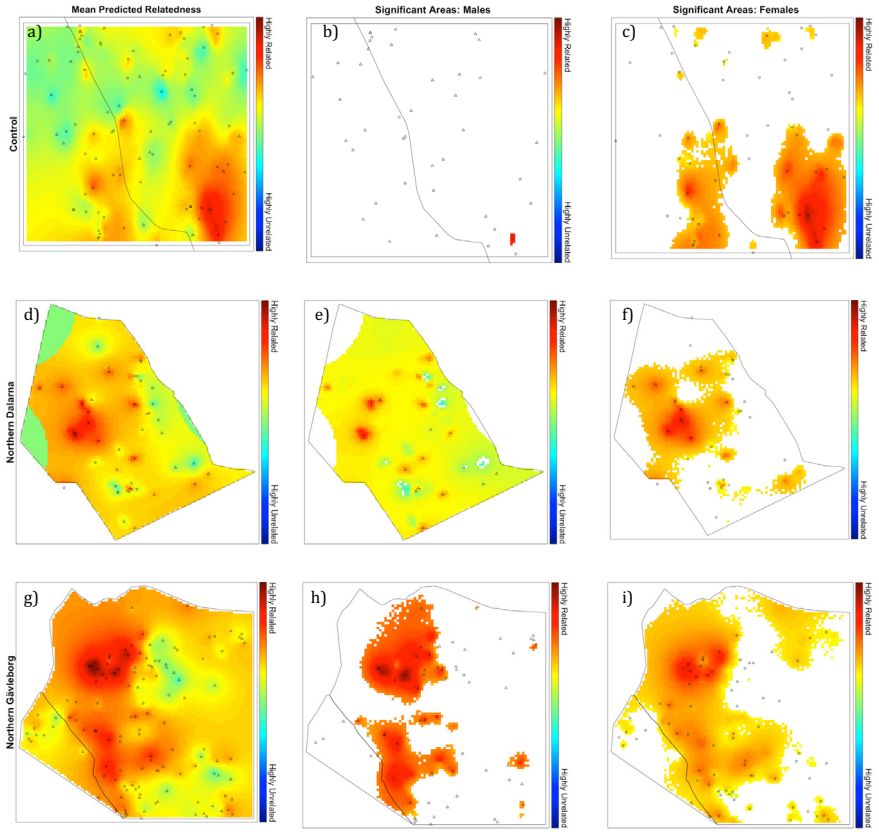

Figure 12 The first column shows the overall interpolation of relatedness for the three local areas: a-c) control area (CA); d-f) northern Dalarna (ND); and g-i) northern Gävleborg (NG). The second and third column shows significant areas of relatedness for males and females respectively

*Table 4 All relatedness values above 0.40 were extracted from each of the local areas. A Wilcoxon Sum Rank Test was conducted to compare the three areas pairwise. Boldfaced text and a * indicates that the two areas differ significantly.*

Area	Mean High Relatedness	N	P-Value Control	NG
Control	0.50	93	--	--
N Dalarna	0.54	46	**0,016***	**0,0079***
N Gävleborg	0.51	131	0,81	--

5 Contributions to Conservation

5.1 Paper I

The panel of SNPs designed for the Scandinavian brown bear enables both fundamental research and monitoring schemes to be conducted with several advantages over the current marker-of-choice, microsatellites. A comparison of the two markers resulted in the SNPs being less expensive, more sensitive, less error-prone and expedetious. This makes the SNPs a highly useful tool for management. In addition, the SNPs can be analysed across different laboratories with the same results unlike microsatellites (Seeb et al. 2011).

The high discriminatory power of the SNP panel leads to a probability of identity that is extremely low (based on 96 SNPs with a mean MAF = 0.40 and unrelated individuals the $P(ID) < 2 \times 10^{-40}$). This strongly suggests that the SNPs will be robust for individual-based studies as well as for inferring relatedness between individuals. This leads to many research possibilities, some of which are illustrated in this dissertation. Others not addressed here include mating patterns, reproductive success, and forensics.

5.2 Paper II

Reliable population size estimates are important for conservation and management planning, but they are difficult to obtain. There are many methods in existence, however, as there are always uncertainties with the estimations, it can be helpful to try new methods. This is what Creel and Rosenblatt (2013) attempted to do with their development of the pedigree reconstruction method. Previously, the pedigree reconstruction method (CRE) of estimating population

size was only tested through simulations (Creel & Rosenblatt 2013). Paper II shows how this method performs in an empirical setting and provides the advantages and disadvantages associated with the method. The results of CRE fell within the confidence limits of the official estimates suggesting that the CRE method is reliable and therefore can be an additional tool for estimating population size. However, a minimum of 40% of the total population is required to be sampled to avoid serious underestimations.

5.3 Paper III

Dispersal is a fundamental process in wild populations. It acts as a mechanism to maintain or increase genetic diversity, avoid inbreeding, spread genes, and is a key element of metapopulation dynamics (Matthysen 2012). It is an important factor in ecological studies and for conservation and management of species and populations. Yet, due to logistics, it is difficult to study, especially for populations that are of conservation concern. These populations are often relics of a larger, previously existing population or contain few individuals. The ability to approach and handle individuals directly may be extremely difficult, harmful to both the animal and handler, or both (Arnemo et al. 2006). This means that studying dispersal in many populations can only be done through observation (logistically prohibitive) or through noninvasive means. This paper empirically shows how dispersal can be assessed using noninvasive genetic sampling combined with citizen science and SNPs for pedigree reconstruction. Citizen volunteers can be a great help with sample collection leading to a wider geographic survey and a shorter timeframe than would otherwise be possible. SNPs have several features that make them an ideal choice for pedigree-based studies. These features include: low error rates; the need for only small DNA fragments; and SNPs are highly sensitive meaning that only very low amounts of DNA are needed to detect a genotype. Even though this method misses long distance dispersers (a problem with most methods), it can nonetheless indicate dispersal strategies in populations, such as sex-biased dispersal, as it did with the brown bear. This was done without ever having affected individual subjects and despite not having life history information. This can be highly useful for other species that are of conservation concern, especially those that are sensitive to human presence and handling or pose a danger to researchers.

5.4 Paper IV

There are many factors that can affect the spatial distribution of a species including habitat heterogeneity (e.g. Contasti et al. 2012), presence or absence of predators and prey (Wright 1950), landscape features such as rivers (Cushman et al. 2006), and anthropogenic influences such as habitat fragmentation (e.g. Husemann et al. 2015). The latter is constantly changing and increasing as the human population continues to grow. It has long been in the interests of conservationists to understand how wild populations are structured in the landscape. Consequently, many methods have been developed to detect substructuring within and between populations (e.g. Pritchard et al. 2000; Guillot et al. 2005; Jombart et al. 2008; Basto et al. 2016). But there are several factors that must be taken into account during the study design including scale, population characteristics, and the ultimate aim of the study. Thus different methods suit different purposes. Currently, most methods are directed towards large-scale metapopulations and are based on genetic differentiation which has been derived from historical and evolutionary processes. Few methods exist for looking at fine-scale structure and even fewer exist for populations that are continuous. Paper IV offers a new method designed to detect fine-scale spatial structuring in a continuous population. This method is especially useful for ecological and conservation studies because it reveals structure that has arisen through contemporary processes. If a large highway was built affecting the latest two generations within a population, this method would detect subsequent structuring. It is therefore useful for monitoring programs. Additionally, it avoids some of the issues that are associated with other methods including sensitivity to related individuals included in the samples, markers under selection or out of Hardy-Weinberg equilibrium and marker-based ascertainment biases.

6 Concluding Remarks

This thesis presents the development of a new tool for study on the Scandinavian brown bear and three empirically tested methods that can be used for study on other populations of conservation concern. As we fast forward through the rapid pace of genomic development and acquire vast amounts of data, my hope is that the process with which I followed to develop the panel of SNPs will be an example of a simplified, targeted technique that requires minimal computation time and expertise. I have taken much care to include in the publication the methods I followed in such detail that it should be reproducible. In scientific literature, writing reproducible methods is what is expected, but unfortunately, in my review of the literature, this was rather uncommon. So far, the SNPs have been highly useful and have proven to be easy to work with, both in the laboratory and in downstream analysis. As the use of SNPs is rather new to the field of conservation, it is necessary that new methods be developed to optimize their use. With this thesis, I have empirically tested a new method for estimating population size, developed a model to estimate natal dispersal distance without the need to have contact with the study individuals, and have developed a new method where I detected novel, fine-scale, contemporary, spatial structure within a continuously distributed subpopulation. Each of these methods can be adopted for most other species where noninvasive sampling is possible. It is my hope that this work will contribute to the conservation of populations and help to minimize biodiversity loss.

References

Allendorf, F.W., Hohenlohe, P.A. & Luikart, G., 2010. Genomics and the future of conservation genetics. *Nature Reviews Genetics*, 11(10), pp.697–709.

Anderson, E.C. & Garza, J.C., 2006. The power of single-nucleotide polymorphisms for large-scale parentage inference. *Genetics*, 172(4), pp.2567–82.

Araki, H., Cooper, B. & Blouin, M.S., 2009. Carry-over effect of captive breeding reduces reproductive fitness of wild-born descendants in the wild. *Biology Letters*, 5(5), pp.621–624.

Arnemo, J.M. et al., 2006. Risk of capture-related mortality in large free-ranging mammals: experiences from Scandinavia. *Wildlife Biology*, 12(1), pp.109–113.

Basto, M.P. et al., 2016. Assessing genetic structure in common but ecologically distinct carnivores: The stone marten and red fox. *PLoS ONE*, 11(1), pp.1–26.

Bellemain, E. et al., 2005. Estimating population size of elusive animals with DNA from hunter-collected feces : four methods for brown bears. *Conservation Biology*, 19(1), pp.150–161.

Bellemain, E. et al., 2006. The dilemma of female mate selection in the brown bear, a species with sexually selected infanticide. *Proceedings of the Royal Society B*, 273(1584), pp.283–91.

Bidon, T. et al., 2015. Genome-wide search identifies 1.9 megabases from the polar bear Y chromosome for evolutionary analyses. *Genome Biology and Evolution*, pp.1–38.

Bray, S.C.E. et al., 2013. Ancient DNA identifies post-glacial recolonisation, not recent bottlenecks, as the primary driver of contemporary mtDNA phylogeography and diversity in Scandinavian brown bears. *Diversity and Distributions*, 19(3), pp.245–256.

Ceballos, G. et al., 2015. Accelerated modern human – induced species losses: entering the sixth mass extinction. *Sciences Advances*, 1(5), pp.1–5.

Contasti, A.L. et al., 2012. Explaining spatial heterogeneity in population dynamics and genetics from spatial variation in resources for a large herbivore. *PLoS ONE*, 7(10), p.e47858.

Creel, S. & Rosenblatt, E., 2013. Using pedigree reconstruction to estimate population size: genotypes are more than individually unique marks. *Ecology and Evolution*, 3(5), pp.1294–1304.
Cronin, M.A. et al., 2009. Genetic variation, relatedness, and effective population size of polar bears (Ursus maritimus) in the southern Beaufort Sea, Alaska. *The Journal of Heredity*, 100(6), pp.681–90.
Csilléry, K. et al., 2006. Performance of marker-based relatedness estimators in natural populations of outbred vertebrates. *Genetics*, 173(4), pp.2091–101.
Cushman, S. a et al., 2006. Gene flow in complex landscapes: testing multiple hypotheses with causal modeling. *The American Naturalist*, 168(4), pp.486–99.
Dahle, B. & Swenson, J.E., 2003. Factors influencing length of maternal care in brown bears (Ursus arctos) and its effect on offspring. *Behavioral Ecology and Sociobiology*, 54(4), pp.352–358.
Darwin, C., 1859. *On the Origin of Species by Means of Natural Selection, Or, The Preservation of Favoured Races in the Struggle for Life*, London: J. Murray.
Frankel, O.H., 1974. Genetic conservation: our evolutionary responsibility. *Genetics*, 78(1), pp.53–65.
Gilpin, M.E. & Soulé, M.E., 1986. Minimum viable populations: processes of species extinction. In M. E. Soulé, ed. *Conservation Biology: The Science of Scarcity and Diversity*. MA: Sinauer, pp. 19–34.
Groom, M.J., Meffe, G.K. & Carroll, C.R., 2006. *Principles of Conservation Biology* Third., Sunderland: Sinauer Associates, Inc.
Guillot, G., Mortier, F. & Estoup, A., 2005. Geneland: a program for landscape genetics. *Molecular Ecology Notes*, 5, pp.712–715.
Hamilton, W.D. & May, R.M., 1977. Dispersal in stable habitats. *Nature*, 269(5629), pp.578–581.
Hauser, L. et al., 2011. An empirical comparison of SNPs and microsatellites for parentage and kinship assignment in a wild sockeye salmon (Oncorhynchus nerka) population. *Molecular Ecology Resources*, 11, pp.150–61.
Hellborg, L. & Ellegren, H., 2003. Y chromosome conserved anchored tagged sequences (YCATS) for the analysis of mammalian male-specific DNA. *Molecular Ecology*, 12(1), pp.283–91.
Husemann, M. et al., 2015. Post-fragmentation population structure in a cooperative breeding Afrotropical cloud forest bird: emergence of a source-sink population network. *Molecular Ecology*, 24(6), pp.1172–87.
Huston, M.A., 1994. *Biological Diversity: The Coexistence of Species on Changing Landscapes*, Cambridge: Cambridge University Press.
Jombart, T. et al., 2008. Revealing cryptic spatial patterns in genetic variability by a new multivariate method. *Heredity*, 101(1), pp.92–103.
Kanno, Y., Vokoun, J.C. & Letcher, B.H., 2011. Sibship reconstruction for inferring mating systems, dispersal and effective population size in headwater brook trout (Salvelinus fontinalis) populations. *Conservation Genetics*, 12(3), pp.619–628.
Kindberg, J. et al., 2011. Estimating population size and trends of the Swedish

brown bear Ursus arctos population. *Wildlife Biology*, 17(2), pp.114–123.
Kindberg, J. & Swenson, J.E., 2013. *Beräkning av björnstammens storlek i Värmland , Dalarnas och Gävleborgs län*,
Kindberg, J. & Swenson, J.E., 2015. *Björnstammens storlek i Västerbotten 2014*,
Krawczak, M., 1999. Informativity assessment for biallelic single nucleotide polymorphisms. *Electrophoresis*, 20(8), pp.1676–81.
Lande, R., 1988. Genetics and demography in biological conservation. *Science*, 241(4872), pp.1455–1460.
Lawson Handley, L.J. & Perrin, N., 2007. Advances in our understanding of mammalian sex-biased dispersal. *Molecular Ecology*, 16(8), pp.1559–78.
Liu, N. et al., 2005. Comparison of single-nucleotide polymorphisms and microsatellites in inference of population structure. *BMC Genetics*, 6(Suppl I), p.S26.
Luck, G.W., Daily, G.C. & Ehrlich, P.R., 2003. Population diversity and ecosystem services. *Trends in Ecology and Evolution*, 18(7), pp.331–336.
Lynch, M. & Ritland, K., 1999. Estimation of pairwise relatedness with molecular markers. *Genetics*, 152(4), pp.1753–66.
Manel, S. et al., 2004. Assumed and inferred spatial structure of populations: the Scandinavian brown bears revisited. *Molecular Ecology*, 13(5), pp.1327–31.
Matthysen, E., 2012. Multicausality of dispersal: a review. In J. Clobert et al., eds. *Dispersal Ecology and Evolution*. Oxford University Press, pp. 3–18.
Morin, P.A., Martien, K.K. & Taylor, B.L., 2009. Assessing statistical power of SNPs for population structure and conservation studies. *Molecular Ecology Resources*, 9(1), pp.66–73.
Nichols, J.D. & Williams, B.K., 2006. Monitoring for conservation. *Trends in Ecology and Evolution*, 21(12), pp.668–673.
Ouborg, N.J. et al., 2010. Conservation genetics in transition to conservation genomics. *Trends in Genetics*, 26(4), pp.177–87.
Palsbøll, P.J., Zachariah Peery, M. & BéRubé, M., 2010. Detecting populations in the “ambiguous” zone: Kinship-based estimation of population structure at low genetic divergence. *Molecular Ecology Resources*, 10(5), pp.797–805.
Pardini, a T. et al., 2001. Sex-biased dispersal of great white sharks. *Nature*, 412(6843), pp.139–40.
Patzenhauerová, H. et al., 2013. Parentage analysis of Ansell’s mole-rat family groups indicates a high reproductive skew despite relatively relaxed ecological constraints on dispersal. *Molecular Ecology*, 22(19), pp.4988–5000.
Pritchard, J.K., Stephens, M. & Donnelly, P., 2000. Inference of population structure using multilocus genotype data. *Genetics*, 155(2), pp.945–59.
Pusey, A.E., 1987. Sex-biased dispersal and inbreeding avoidance in birds and mammals. *Trends in Ecology & Evolution*, 2(10), pp.295–9.
Putnam, A.S. & Ivy, J.A., 2014. Kinship-based management strategies for captive breeding programs when pedigrees are unknown or uncertain. *Journal of Heredity*, 105(3), pp.303–311.
Qi, Y. et al., 2013. Genetic evidence for male-biased dispersal in the Qinghai toad-

headed agamid Phrynocephalus vlangalii and its potential link to individual social interactions. *Ecology and Evolution*, 3(5), pp.1219–1230.
Riester, M., Stadler, P.F. & Klemm, K., 2009. FRANz: reconstruction of wild multi-generation pedigrees. *Bioinformatics*, 25(16), pp.2134–9.
Robinson, S.P., Simmons, L.W. & Kennington, W.J., 2013. Estimating relatedness and inbreeding using molecular markers and pedigrees: the effect of demographic history. *Molecular Ecology*, 22(23), pp.5779–92.
Russello, M.A. & Amato, G., 2004. Ex situ population management in the absence of pedigree information. *Molecular Ecology*, 13(9), pp.2829–2840.
Sæther, A.B. et al., 1996. Life history variation, population processes and priorities in species conservation: toawards a reunion of research paradigms. *Oikos*, 77(2), pp.217–226.
Sandell, M., 1989. *Carnivore Behavior, Ecology, and Evolution* J. L. Gittleman, ed., Boston, MA: Springer US.
Santure, A.W. et al., 2010. On the use of large marker panels to estimate inbreeding and relatedness: empirical and simulation studies of a pedigreed zebra finch population typed at 771 SNPs. *Molecular Ecology*, 19(7), pp.1439–51.
Seeb, J.E. et al., 2011. Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms. *Molecular Ecology Resources*, 11, pp.1–8.
Serbezov, D. et al., 2010. Mating patterns and determinants of individual reproductive success in brown trout (Salmo trutta) revealed by parentage analysis of an entire stream living population. *Molecular Ecology*, 19(15), pp.3193–3205.
Slatkin, M., 1987. Gene flow and the geographic structure of natural populations. *Science*, 236(4803), pp.787–792.
Spong, G. & Creel, S., 2001. Deriving dispersal distances from genetic data. *Proceedings of the Royal Society B*, 268(1485), pp.2571–4.
Spong, G.F. et al., 2008. Factors affecting the reproductive success of dominant male meerkats. *Molecular Ecology*, 17(9), pp.2287–99.
Stearns, S.C., 1992. *The Evolution of Life Histories*, Oxford: Oxford University Press.
Støen, O.-G. et al., 2006. Inversely density-dependent natal dispersal in brown bears Ursus arctos. *Oecologia*, 148(2), pp.356–64.
Taberlet, P. et al., 1995. Localization of a contact zone between two highly divergent mitochondrial DNA lineages of the brown bear Ursus arctos in Scandinavia. *Conservation Biology*, 9(5), pp.1255–1261.
Taberlet, P. & Bouvet, J., 1994. Mitochondrial DNA polymorphism, phylogeography, and conservation genetics of the brown bear Ursus arctos in Europe. *Proceedings of the Royal Society Biological Sciences*, 255(1344), pp.195–200.
Taberlet, P. & Luikart, G., 1999. Non-invasive genetic sampling and individual identification. *Biological Journal of the Linnean Society*, 68(1-2), pp.41–55.
Thomas, S.C., 2005. The estimation of genetic relationships using molecular

markers and their efficiency in estimating heritability in natural populations. *Philosophical Transactions of the Royal Society B*, 360, pp.1457–67.
Thompson, E.A., 1975. The estimation of pairwise relationships. *Annals of Human Genetics*, 39(2), pp.173–188.
Tokarska, M. et al., 2009. Effectiveness of microsatellite and SNP markers for parentage and identity analysis in species with low genetic diversity: the case of European bison. *Heredity*, 103(4), pp.326–32.
Vucetich, J.A. & Waite, T.A., 2003. Spatial patterns of demography and genetic processes across the species' range: null hypotheses for landscape conservation genetics. *Conservation Genetics*, 4(5), pp.639–645.
Waits, L. et al., 2000. Nuclear DNA microsatellite analysis of genetic diversity and gene flow in the Scandinavian brown bear (Ursus arctos). *Molecular Ecology*, 9(4), pp.421–31.
Waits, L.P. & Paetkau, D., 2005. Noninvasive genetic sampling tools for wildlife biologists: a review of applications and recommendations for accurate data collection. *Journal of Wildlife Management*, 69(4), pp.1419–1433.
Wang, J., 2015. Pedigrees or markers: Which are better in estimating relatedness and inbreeding coefficient? *Theoretical Population Biology*, 107, pp.4–13.
Wright, S., 1950. Genetical structure of populations. *Nature*, 166, pp.247–249.
Zedrosser, A. et al., 2007. Should I stay or should I go? Natal dispersal in the brown bear. *Animal Behaviour*, 74(3), pp.369–376.

Acknowledgements

I would like to extend my heartfelt thanks to my supervisors for their guidance, support and endless patience. I have been very fortunate to have supervisors like you. **Nat**, you have been there from the start and have always been open and willing to help in a kind and friendly way. You patiently taught me what I needed to know about the world of bioinformatics and made it easy for me by providing me with everything I needed. But my gratitude extends far beyond this. Your advice during my PhD education in general has always been top notch and you have really helped me to improve my scientific writing. You have encouraged me and believed in me even when I had my doubts. For this and more, I am so grateful. **Astrid**, you first appeared in my life at a bus stop in Italy and have since been an incredible role model. Not only are you a great scientist, you are a great person, always considerate, professional, and humble. You have listened to me with a patient ear and have provided guidance and encouragement when I needed it. I thoroughly enjoyed my research visit with you in Aalborg and I thank you for your hospitality. We have had many interesting scientific discussions and you have helped me to consider things I would have otherwise overlooked. You have also helped me to improve my writing. Thanks for your kindness, patience and for being a great role model. **Jonas**, even though you have only recently become my supervisor, you have been an important part of my PhD journey from the very beginning. I really appreciate all your knowledge and sound advice. Your ongoing support has really been encouraging, especially at times when I may have felt a bit disheartened. You have always been warm, friendly and welcoming and it has meant a lot to me. Thanks for all that and more Jonas! Skål! And last, but certainly not least, **Göran**. From the very beginning, you have been an exceptional supervisor, allowing me independence and yet with your door always open. Your unwavering belief in me has kept me going even when I didn't believe in myself. Our scientific discussions have been interesting and

stimulating and I am so thankful that we could connect in that way. I am very grateful to you for your patience when I maybe didn't deserve it; for your guidance when I wasn't sure; for generously giving me time when you may have been busy; for not just listening to me but hearing; and for your endless support. Most of all, thank you for giving me this amazing opportunity to grow as a person in so many ways. I could not have asked for a better main supervisor. Thank you.

I wish to thank all those that have been involved in my PhD work in some way. Thanks to the ConGenOmics program by the European Science Foundation for supporting me in three different endeavours: a workshop, a spring school and a research visit. I have met so many great people. In particular, I would like to thank Gernot Segelbacher, Ettore Randi, Aaron Shafer and all the steering committee. I would like to thank everyone involved in the Scandinavian Brown Bear Research Project. I enjoyed meeting you all in Thessaloniki and Öster Malma. Thanks especially to Jon Swenson, Sven Brunberg, Andrea Friebe and again to Jonas Kindberg for always being so kind and welcoming. Many thanks to Siv-Grethe, Alexander Kopatz, Julia Schregel, Snorre Hagen and Hans-Geir Eiken for our various visits and discussions. Thanks to Axel Janke and team for their generosity in helping with the SNP project. Thanks to Torsten Nygård Kristensen, Cino Pertoldi and Lisbeth Skjot Moller for hosting me in Aalborg. Many thanks to Håvard Rue and his lab for both hosting me during a research visit and all the stimulating discussions. Thanks to Arne Soderberg and everyone at SciLifeLabs who were involved in the sequencing of brown bear. Thanks to Michael Schneider for all of your support both with regards to samples as well as during conferences an meetings. Many thanks to Patrizia Giangregorio for our collaborations and for a lovely visit here in Umeå. It was a pleasure to work with Aritz Ruiz Gonzales and Geir-Arne Fuglstad. And of course, Helena Königsson, a big thank you for teaching me and all of your support both in and out of the lab in all these years.

I would very much like to thank everyone who has been an important part of my life during these years. Steven Declerk, you have been a great friend and a fantastic role model when it comes to science. Ida-Maria, it has been a pleasure to share the office with you. Anders, Ruth, Ellinor, I have thoroughly enjoyed being in the same lab together and your friendships. Andy and Anouschka, you have been great friends and I have very much enjoyed our activities together, especially the birding activities. Sabrina, Klara and Liza you are amazing and I am so grateful to have your friendship. Edward, Per and Magnus, you have also been great friends and I very much enjoyed our visits.

Robert and Joanna, you have been a joy to work with. All other PhD students and mentors, thank you. I have enjoyed getting to know you all. I would like to extend my thanks to my friends that have been instrumental in keeping me grounded every Saturday. Thanks for that Cheryl, Linn, Marina and Kicki.

Thank you Frank for being my closest friend and partner during these years and for all your support and love. Thanks to my dogs, Nelson and Milo, for your sweetness and for being my furry companions during this journey. A multitude of thanks to my parents for their love, encouragement, and support. What would I do without you? Thanks so much to my sister Sylvia and brother David for your love and support.

Many more thanks are owed to all my friends and family. I couldn't have done this without you.