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)
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

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>List of Publications</strong></td>
<td>8</td>
</tr>
<tr>
<td><strong>Abbreviations &amp; Definitions</strong></td>
<td>10</td>
</tr>
<tr>
<td><strong>1 Introduction</strong></td>
<td>13</td>
</tr>
<tr>
<td>1.1 The Biodiversity Crisis</td>
<td>13</td>
</tr>
<tr>
<td>1.2 Importance of Genetic Diversity in Wildlife Populations</td>
<td>13</td>
</tr>
<tr>
<td>1.2.1 Population Size</td>
<td>13</td>
</tr>
<tr>
<td>1.2.2 Dispersal</td>
<td>14</td>
</tr>
<tr>
<td>1.2.3 Population Structure</td>
<td>14</td>
</tr>
<tr>
<td>1.3 Noninvasive Sampling</td>
<td>15</td>
</tr>
<tr>
<td>1.4 Aims</td>
<td>16</td>
</tr>
<tr>
<td><strong>2 Background</strong></td>
<td>17</td>
</tr>
<tr>
<td>2.1 Relatedness</td>
<td>18</td>
</tr>
<tr>
<td>2.2 Single Nucleotide Polymorphisms</td>
<td>18</td>
</tr>
<tr>
<td><strong>3 Model System and Methods</strong></td>
<td>20</td>
</tr>
<tr>
<td>3.1 Scandinavian Brown Bear (All Papers)</td>
<td>20</td>
</tr>
<tr>
<td>3.2 SNP Development (Paper I)</td>
<td>22</td>
</tr>
<tr>
<td>3.3 Sample Collection and SNP Genotyping (Papers II to IV)</td>
<td>23</td>
</tr>
<tr>
<td>3.4 Relatedness (Papers II to IV)</td>
<td>24</td>
</tr>
<tr>
<td>3.5 Pedigree Reconstruction (Papers II and III)</td>
<td>25</td>
</tr>
<tr>
<td>3.6 Dispersal (Paper III)</td>
<td>25</td>
</tr>
<tr>
<td>3.7 Landscape Relatedness (Paper IV)</td>
<td>26</td>
</tr>
<tr>
<td><strong>4 Summary of Results</strong></td>
<td>28</td>
</tr>
<tr>
<td>4.1 SNP Development (Paper I)</td>
<td>28</td>
</tr>
<tr>
<td>4.1.1 Sequencing</td>
<td>28</td>
</tr>
<tr>
<td>4.1.2 SNP Selection</td>
<td>30</td>
</tr>
<tr>
<td>4.1.3 SNP Validation</td>
<td>30</td>
</tr>
<tr>
<td>4.2 SNP Genotyping (Papers II to IV)</td>
<td>30</td>
</tr>
<tr>
<td>4.3 Relatedness (Papers II to IV)</td>
<td>31</td>
</tr>
<tr>
<td>4.4 Pedigree Reconstruction (Papers II and III)</td>
<td>31</td>
</tr>
<tr>
<td>4.5 Population Census (Paper II)</td>
<td>32</td>
</tr>
<tr>
<td>4.6 Dispersal (Paper III)</td>
<td>33</td>
</tr>
<tr>
<td>4.7 Landscape Relatedness (Paper IV)</td>
<td>35</td>
</tr>
</tbody>
</table>
5 Contributions to Conservation
   5.1 Paper I
   5.2 Paper II
   5.3 Paper III
   5.4 Paper IV

6 Concluding Remarks

References

Acknowledgements
List of Publications

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


Papers I-III are reproduced with the permission of the publishers.
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
<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>natal dispersal</td>
<td>the movement of an individual away from the natal area for reproduction</td>
</tr>
<tr>
<td>panmictic</td>
<td>randomly mating and fully interacting population</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>philopatric</td>
<td>a behavior where individuals settle within or close by their natal area</td>
</tr>
<tr>
<td>phylogeography</td>
<td>the interspecies and intraspecies geographic distribution</td>
</tr>
<tr>
<td>primers</td>
<td>sequence of nucleotides that occur before or after a target locus/loci</td>
</tr>
<tr>
<td>r-value</td>
<td>Lynch-Ritland relatedness coefficient</td>
</tr>
<tr>
<td>read</td>
<td>computational interpretation of a genomic DNA sequence</td>
</tr>
<tr>
<td>SNP</td>
<td>single nucleotide polymorphism</td>
</tr>
</tbody>
</table>
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).
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).

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.
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.
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.
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

<table>
<thead>
<tr>
<th>Relationship</th>
<th>r-value</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parent-Offspring/Full-Siblings</td>
<td>0.50</td>
<td>First Order</td>
</tr>
<tr>
<td>Half-Siblings/Grandparent-Grandoffspring/</td>
<td>0.25</td>
<td>Second Order</td>
</tr>
<tr>
<td>Aunts/Uncles-Nieces/Nephews</td>
<td>0.25</td>
<td>Third Order</td>
</tr>
<tr>
<td>Unrelated</td>
<td>0.00</td>
<td>Unrelated</td>
</tr>
</tbody>
</table>

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.
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.](image)

### 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>
<thead>
<tr>
<th>Avg Frag Size (bp)</th>
<th>Genome Size (billion)</th>
<th># Unique Fragments</th>
<th>Read Depth</th>
<th>Genomic Coverage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimated</td>
<td>3100</td>
<td>2.4</td>
<td>131,910</td>
<td>38</td>
</tr>
<tr>
<td>Genome</td>
<td>3465</td>
<td>2.3</td>
<td>93,678</td>
<td>53</td>
</tr>
<tr>
<td>% Diff</td>
<td>-11.8</td>
<td>5.1</td>
<td>29.0</td>
<td>-28.3</td>
</tr>
</tbody>
</table>

Table 2 Estimated genomic parameters versus actual as determined through an in silico digestion with the draft genome assembly.
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; $H_E$ and $H_O$ = expected and observed heterozygosity respectively; $F_{ST}$ is the genetic differentiation calculated based on the three mitochondrial haplotypes (i.e. three subpopulations).

<table>
<thead>
<tr>
<th></th>
<th>MAF</th>
<th>$H_E$</th>
<th>$H_O$</th>
<th>Overall $F_{ST}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall Mean</td>
<td>0.39</td>
<td>0.47</td>
<td>0.44</td>
<td>0.08</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>0.06</td>
<td>0.03</td>
<td>0.07</td>
<td>0.10</td>
</tr>
</tbody>
</table>

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 $1.538 \times 10^{-4}$ 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 an 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.

<table>
<thead>
<tr>
<th>Area</th>
<th>Mean High Relatedness</th>
<th>N</th>
<th>P-Value Control</th>
<th>P-Value NG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>0.50</td>
<td>93</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>N Dalarna</td>
<td>0.54</td>
<td>46</td>
<td>0.016*</td>
<td>0.0079*</td>
</tr>
<tr>
<td>N Gävleborg</td>
<td>0.51</td>
<td>131</td>
<td>0.81</td>
<td>--</td>
</tr>
</tbody>
</table>
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 expeditious. 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 \( P(\text{ID}) < 2 \times 10^{-40} \)). 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 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.
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


Kindberg, J. et al., 2011. Estimating population size and trends of the Swedish...


Qi, Y. et al., 2013. Genetic evidence for male-biased dispersal in the Qinghai toad-


markers and their efficiency in estimating heritability in natural populations. 


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 and 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.
De Novo SNP Discovery in the Scandinavian Brown Bear (Ursus arctos)

Anita J. Norman¹, Nathaniel R. Street², Göran Spong¹

¹ Department of Wildlife, Fish and Environmental Studies, Molecular Ecology Research Group, Swedish University of Agricultural Sciences, Umeå, Sweden,
² Department of Plant Physiology, Umeå Plant Science Centre, Umeå University, Umeå, Sweden

Abstract

Information about relatedness between individuals in wild populations is advantageous when studying evolutionary, behavioural and ecological processes. Genomic data can be used to determine relatedness between individuals either when no prior knowledge exists or to confirm suspected relatedness. Here we present a set of 96 SNPs suitable for inferring relatedness for brown bears (Ursus arctos) within Scandinavia. We sequenced reduced representation libraries from nine individuals throughout the geographic range. With consensus reads containing putative SNPs, we applied strict filtering criteria with the aim of finding only high-quality, highly-informative SNPs. We tested 150 putative SNPs of which 96% were validated on a panel of 68 individuals. Ninety-six of the validated SNPs with the highest minor allele frequency were selected. The final SNP panel includes four mitochondrial markers, two monomorphic Y-chromosome sex-determination markers, three X-chromosome SNPs and 87 autosomal SNPs. From our validation sample panel, we identified two previously known parent-offspring dyads with reasonable accuracy. This panel of SNPs is a promising tool for inferring relatedness in the brown bear population in Scandinavia.


Introduction

Genomic data are useful for understanding wild populations, particularly for wide-ranging and elusive species like the brown bear (Ursus arctos). Among many uses, genomic markers can help determine genetic relatedness between individuals in a population, which is key for determining many evolutionary, behavioural or ecological processes [1]. For example, although maternity can often be reliably inferred based on behavioural patterns alone (cf. [2]), assigning paternity is typically more problematic. This is the case for some species that appear to have a monogamous mating system when observed in the wild, yet genetic analyses reveal extra-pair paternity as being common [3]. Detecting paternity can help determine, for example, factors affecting reproductive success (e.g. [4]). In addition, relatedness measures can be used to detect hybridization events or identify introgression zones (e.g. [5,6]). Detecting inbreeding can be critical for small or reintroduced populations that are prone to inbreeding depression [7]. Genetic relatedness can also be used to measure gene flow and uncover dispersal patterns [8]. As such, the use of high quality genomic markers can enhance our understanding of biological processes in wild systems as shown by relatedness studies on Ursus species (e.g. [9–12]).

Resolving relationships in wild populations can be challenging [13] and is typically reliant upon high quality markers with high genomic resolution [14]. Insufficient genomic resolution (either too few markers or unequal representation throughout the genome) can result in inflated genotypic variances and, thus, lower confidence making relatedness inferences problematic. Single nucleotide polymorphisms (SNPs) occur frequently throughout the genome rendering them suitable for analyses requiring high genomic resolution. In addition, some marker types (e.g. microsatellites) are error prone and suffer from technical artifacts such as null alleles. Erroneous genotypes can cause significant biases in genetic monitoring [15]. The bi-allelic nature of SNPs leads to simplified genotyping that is less erroneous [16].

Until recently, genome-wide SNP marker development was prohibitively expensive and time-consuming. With the advent of next-generation sequencing (NGS) technologies, SNP development has become more accessible. Correspondingly, SNPs are increasingly being utilized in studies of non-model organisms (e.g. [17–22]). For example, Miller et. al. [23]
SNP Discovery in Scandinavian Brown Bear

developed a set of 100 SNPs for polar bear (Ursus maritimus) and brown bear to investigate phylogenetic history. However, processing the vast amount of data generated by NGS technologies has become a significant challenge due to the large demand for bioinformatics expertise, computational load and data storage infrastructure [24]. Therefore, methodologies that reduce the necessary amount of data and computational complexity within the limits of the study can simplify the complex downstream analyses and reduce demands on infrastructure. For example, application of a reduced representation libraries approach (RRL) for SNP discovery [25] considerably decreases the amount of sequencing data required while simultaneously allowing for high genomic resolution. Advances to the RRL methodology have recently been developed that further increase its utility (e.g. [26]).

An informative SNP panel is one in which each SNP maximizes the differences in allelic representation across individuals within a population when compared to all other SNPs. Hence, SNPs with higher minor allele frequencies (MAF) and that are not in linkage with each other are more informative for relatedness inference [27]. The number of SNPs required for making reliable relatedness estimates has been debated (e.g. [28,29]). However, depending on population characteristics, sample size and level of marker informativeness, there is evidence that relatedness inferences can be reliably inferred using a minimum of 60 SNPs [27]. NGS-based methods have enabled detection of thousands to hundreds of thousands of SNPs (depending on the species, proportion of the genome sequenced and read depth), representing orders of magnitude greater than what is required for relatedness studies. Thus, data reduction through the use of RRL approaches and the application of highly stringent filtering criteria to retain only the highest quality, informative SNPs is particularly relevant.

Historically, over-hunting and habitat fragmentation have negatively affected many brown bear populations, a trend that led to the loss of much of the historical geographic range in Europe [30]. Currently, Scandinavia is among the few regions where the brown bear population is increasing [31]. Maternally-inherited mitochondrial DNA from the control region have shown that the brown bear population in Sweden and Norway consists of two distinct lineages with more than 7% differentiation between them; the eastern European lineage situated in the north, and the western European lineage situated in the south-central part of Sweden [30]. The southern population is of particular conservation interest since it is one of the few relic populations of the western European lineage [30].

In this study, we developed de novo a set of 96 high quality SNPs by applying an NGS-based RRL approach with an ascertainment panel of brown bears across the geographic range in Scandinavia. A SNP-chip was designed primarily to facilitate relatedness studies, although it can be useful for a wider range of studies. In addition to autosomal SNPs and Y-chromosome sex-determination markers, we included mitochondrial (mtDNA) and X-chromosome SNPs to further facilitate determination of parental ancestry. Our approach to reducing data complexity allowed for efficient and simplified ascertainment of a medium-throughput panel of highly informative SNPs.

Materials and Methods

Sample Collection and DNA Extraction

We obtained 68 samples from the National Veterinary Institute (Statens veterinärmedicinska anstalt (SVA), Uppsala, Sweden) from bears deceased either through a licensed hunt or that were found dead through other causes (e.g. natural mortality, vehicle/train collisions). No bears were killed for the purpose of this study or for other research endeavours. Samples were obtained with full consent by SVA. Samples were chosen to represent an even sex ratio and the full geographic range of brown bear throughout Sweden (Figure 1). The majority of samples (n=60) were collected from muscle tissue, while others were from liver (n=10) and skin (n=2). Samples were collected between 2000 and 2012 and, except for liver, were stored in ethanol prior to DNA extraction. Liver samples were kept frozen at -20°C.

DNA was extracted using the QIAxasympy SP and the QIAxasympy DNA kit (Qiagen, Hilden, Germany) according to manufacturer’s instructions. Nucleotide quantity and purity were assessed using a spectrophotometer (NanoDrop, Thermo Fisher Scientific, Massachusetts, USA). DNA quality for extractions used for sequencing was visually assessed by gel electrophoresis using the Kodak Electrophoresis Documentation and Analysis System 120 (Eastman Kodak Company, Rochester, USA).

DNA Sequencing

We used a combination of targeted and anonymous sequencing approaches to identify SNPs that are informative of parental lineage and that are autosomal respectively. The targeted approach involved mitochondria and the Y-chromosome, while the anonymous approach involved high-throughput sequencing of reduced representation libraries.

Mitochondrial DNA. Four primer pairs were designed from the published mitochondrial genome (NCBI Accession # EU497665.1) of a European brown bear using Primer3 [32]. Each primer set was designed to amplify a product of approximately 500 base pairs (bp) (Table 1).

Each of ten samples were PCR amplified in a total reaction volume of 20 μl consisting of 2.5 μl 10-40 ng/μl DNA, 0.5 μl 10 mM MgCl₂, 0.5 μl 2.5 mM dNTP’s, 2.0 μl 0.1X Taq buffer*, 1.6 μl 2.0mM MgCl₂* and 0.17 μl Taq DNA polymerase* (“Fermentas Taq DNA Polymerase – native”). The optimized PCR conditions for all primer pairs include 1 cycle of 94°C for 3 min; 20 cycles of 94°C for 20 s, 50°C for 30 s; and 72°C for 5 min. PCR amplification was confirmed through gel electrophoresis. The remainder of the product was Sanger sequenced by Medicinsk och klinisk genetik (Nonhards Universitetssjukhus, Umeå, Sweden) on a 3730 xl DNA analyzer (Applied Biosystems, Foster City, USA). Sequences were aligned using BioEdit (v 7.0.9; Tom Hall, ibis Biosciences, Carlsbad, USA) and manually screened to identify SNPs.
Figure 1. Brown Bear Sampling Locations. Points represent sampling locations for each individual used in the validation genotyping (n=68). They are graphically presented to indicate the mitochondrial-based lineage the individual belongs to, whether it is female or male, and the nine individuals that were initially sequenced.

doi: 10.1371/journal.pone.0081012.g001
The total expected number of base pairs for all four products measured C-value (2.75 pg; [25]) relative to the dog (2.80 pg) whose genome size is approximately 2.4 Gbp (giga base pairs) based on the mitochondrial genome #EU497665.1 (NCBI, Bethesda, USA) for library construction and sequencing. We digested 0.5 μg of each DNA sample (liver) individually for 16 hours with BglII (Fermentas, Vilnius, Lithuania) according to manufacturer’s instructions. To remove the activated enzyme, samples were purified using the MinElute Reaction Cleanup kit (Qiagen, Hilden, Germany) resulting in 2x100 bp paired-end reads with insert sizes ranging from zero to 500 bp (mean 249.01 ± 130.06). Sequence data has been submitted to the NCBI Sequence Read Archive (SRA) under the study accession number SRP023544 (http://www.ncbi.nlm.nih.gov/sra/?term=srp023544).

Quality Filtering and Alignment

Sequenced reads were demultiplexed using the barcode_splitter option of the FASTX Toolkit (v 0.0.13; https://hannonlab.cshl.edu/fastx_toolkit/) and adapters removed with cutadapt (v 0.9.3; [36]). Reads were trimmed to 100 bp, and quality filtered using the FASTX Toolkit trimmer and quality_filter options respectively using the settings: q 10, p 70.

Reduced Representation and High-Throughput Sequencing. To determine an appropriate balance between genomic coverage and read depth, we performed preliminary calculations for developing a reduced representation library using BglII (A/GATCT) restriction enzyme, based on [25]. Our calculations were based on two assumptions: That the average fragment length resulting from a BglII digest of the brown bear genome is similar to that of the human genome (~3,100 bp; [25]) and that the genome size of the brown bear is approximately 2.4 Gbp (giga base pairs) based on the measured C-value (2.75 pg; [34]) relative to the dog (2.80 pg) whose genome size is approximately 2.5 Gbp [35]. With this, we estimated that we could obtain a genomic coverage of ~1% with a read depth of ~40X for each sample if we used all genomic fragments between 100 and 700 bp after a BglII digest. After the sequencing was conducted, a draft genome assembly was made available to us (pers. comm. Axel Janke, Senckenberg Institute, Germany) which we used to perform an in silico digestion with BglII to test the above assumptions.

Y-Chromosome. Four published Y-chromosome primer pairs were selected (DBY3, DBY5, DBY8 and SMCY7; [33]). The total expected number of base pairs for all four products was 1,550.

Each of 12 samples from males were PCR amplified in a total reaction volume of 20 μl consisting of 2.5 μl of 10X Taq DNA polymerase – native). The PCR conditions were optimized and the resulting products measured and sequenced on one lane of Illumina HiSeq2000 sequencing. After the sequencing was conducted, a draft genome assembly was made available to us (pers. comm. Axel Janke, Senckenberg Institute, Germany) which we used to perform an in silico digestion with BglII to test the above assumptions.

Table 1. Mitochondrial DNA primer pairs for brown bear (Ursus arctos) based on the mitochondrial genome #EU497665.1 (NCBI, Bethesda, USA).

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Primer Position</th>
<th>Product Length</th>
<th>H/L</th>
<th>Primer Sequence (5’-3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urs_mtDNA.40H</td>
<td>4542</td>
<td>546</td>
<td>H</td>
<td>GCAATTATGCAGACAGAGATT</td>
</tr>
<tr>
<td>Urs_mtDNA.50L</td>
<td>5087</td>
<td>546</td>
<td>L</td>
<td>GCAATGTGATTACGGTTGAT</td>
</tr>
<tr>
<td>Urs_mtDNA.66H</td>
<td>6661</td>
<td>501</td>
<td>H</td>
<td>GCACTACAGGACAGCTCTTT</td>
</tr>
<tr>
<td>Urs_mtDNA.71L</td>
<td>7181</td>
<td>501</td>
<td>L</td>
<td>CCTGTGGGATAGCAATGAT</td>
</tr>
<tr>
<td>Urs_mtDNA.84H</td>
<td>9422</td>
<td>499</td>
<td>H</td>
<td>GTTCCTGGTAGCTGTGATCT</td>
</tr>
<tr>
<td>Urs_mtDNA.96L</td>
<td>9920</td>
<td>499</td>
<td>L</td>
<td>ACACCCGAGTACAGAGAT</td>
</tr>
<tr>
<td>Urs_mtDNA.134H</td>
<td>13497</td>
<td>499</td>
<td>H</td>
<td>CTTCTGTCCTACCACCGAT</td>
</tr>
<tr>
<td>Urs_mtDNA.160L</td>
<td>16009</td>
<td>508</td>
<td>L</td>
<td>AAAATGACATTGTGCTGAGG</td>
</tr>
<tr>
<td>Urs_mtDNA.166L</td>
<td>16611</td>
<td>529</td>
<td>L</td>
<td>GGACGACAGGACAGCTACAG</td>
</tr>
</tbody>
</table>

* Markers from these sequences included in final SNP set
** Position according to accession # EU497665.1 (NCBI)
*** Includes primers
§ Indicates (H) and light (L) strands

doi: 10.1371/journal.pone.0081012.001
SNP Discovery in Scandinavian Brown Bear

Babraham Bioinformatics, Cambridge, UK) both before and after filtering. After quality filtering, paired reads were synchronized and reads not containing the cut site (GATCT) were removed using customized python scripts. The remaining reads were used as input for analysis and SNP detection using Stacks (v. 0.9995, [37]) with the settings: m 2, M 3, n 1, t and H. Consensus reads generated by Stacks were aligned to the draft genome (see above) using Bowtie 2 (v 2.0.0, [38]) with the settings: q, X 700.

SNP Calling and Validation

The results from Stacks were imported into a custom MySQL (Oracle Corporation, Redwood City, USA) database where, in combination with python scripts, putative SNPs were filtered to remove ones of low quality (Figure 2). First, we ensured that only one SNP could exist on any given read and that the SNP must be at least 20 nt (nucleotides) from the 5' end and at least 35 nt from the 3' end of the read. The rationale behind the one SNP per read was to both reduce the number of pseudo SNPs resulting from sequencing error and to eliminate any hypervariable sequences. We required that the SNP be located in the middle of the read to ensure that adequate flanking sequences remained on either side for subsequent SNP assay development. We removed any SNP that appeared in less than three individuals and that did not contain all three genotypes (i.e. aa, ab, bb). This was to allow us to choose higher quality SNPs with greater representation across the individuals. Homology searches against the reference genome draft assembly were then performed using Blastn (NCBI, Bethesda, USA). Since our aim was to develop a 96-well chip, we could afford to be strict in our filtering, therefore we reduced the minimum 99% identity (allowing for one mismatch assumed to be the SNP) and no gaps. We removed reads that aligned multiple times to ensure that we would not end up with pseudo SNPs due to paralogous sequences. Likewise, we chose only SNPs that aligned to scaffolds with no other SNP to minimize linkage between SNPs due to close physical vicinity. Finally, SNPs were manually screened to ensure exclusion of those with homopolymers in the flanking region as well as for adequate allelic representation. A total of 150 SNPs were selected, assays were developed (Fluidigm Corporation, San Francisco, USA) and then used to genotype 68 brown bear samples using the Fluidigm Biomark.

SNP-Chip Development

Genotyped individuals were analyzed to determine both relatedness using the Graphical Representation of Relationship errors approach [39] and the number of subpopulations using multidimensional scaling (MDS), both as described in [40] and implemented using the R programming language [41]. Where dyads were represented as outliers on a boxplot (0.95 CI) indicating either very close or very distant relatedness, we removed one of the pairs in subsequent analyses. The remaining samples were grouped into three subpopulations based on mitochondrial lineage as confirmed by the MDS analysis. To identify the most informative and highest quality SNPs, we calculated descriptive statistics on the validated SNPs including minor allele frequency (MAF), expected and observed heterozygosity ($H_e$ and $H_o$), and using Genepop v. 4.2 [42,43], Hardy-Weinberg Equilibrium and $F_{ST}$. SNPs were selected for the final SNP-chip (96 SNPs) if they were among those with the highest MAF. It is important to note that depending on time since divergence, ascertainment bias may affect the utility of these SNPs in brown bear populations outside of Scandinavia. These SNPs were analyzed for linkage disequilibrium with $D^*$ statistics in the R packages genetics v 1.3.8 (http://cran.r-project.org/web/packages/genetics/index.html) and LDheatmap v 0.99 [44]. The 90 nuclear SNPs (excluding mitochondrial and Y-chromosome markers) are published in dsSNP through NCBI (Bethesda, USA) with ss numbers from 778679577 to 778079666.

Relatedness Analyses

To determine how informative the final set of SNPs would be in assessing relatedness, we conducted two additional analyses. First, we incorporated all autosomal SNPs (n=87) and unrelated individuals (n=50) and ran Structure [45–48] using a burnin of 100,000 and MCMC reps of 500,000 with 20 iterations each of K = 2 to K = 5 and default settings. Second, we calculated the Lynch and Ritland relatedness estimator (r) [49] using Coancestry [50] with all individuals (n=68) to identify dyads with possible first-order relatedness (i.e. parents or full siblings). To exclude possible parent-offspring relationships of all dyads with r > 0.40, we used a customized python script to calculate the number of alleles shared at all loci excluding those dyads that have at least one locus where no alleles are shared. Our sample panel consisted of two known parent-offspring dyads.

Results and Discussion

DNA Sequencing

A total of 2015 and 1489 bp were sequenced in the mtDNA and the Y-chromosome respectively. Sequencing of the RRLs generated ~20 Gbp of data from nine samples. One sample failed to sequence for unknown reasons. After quality filtering and removal of reads not containing the restriction cut site, approximately 30 million paired reads (32%) remained. We suspect that the low retention rate is a result of unintended sequencing of degraded DNA as indicated by the gel visualization of the restriction digest. However, 92% of retained reads (unpaired) aligned to the draft genome suggesting that the sequence data used in downstream analysis was of high quality. This is promising for sequencing projects that are dependent on low quality DNA (e.g. ancient DNA or environmental DNA).

We utilized a draft genome assembly to test our assumptions regarding the cut frequency of the BglII enzyme and the genome size of the brown bear. Table 2 shows the results of an in silico digestion of the draft genome (for which the genome size estimate of 2.4 Gbp is in accordance with an independent estimate by Miller et. al. [23]) using the BglII restriction enzyme in comparison with our preliminary calculations. The differences for the two assumptions were minor (5% and 12% respectively), confirming the appropriateness of our approach
in ascertainning the right balance between genomic coverage and read depth.

**SNP Calling**

A total of 57 haploid SNPs were identified from the mtDNA sequences. Fifty-four of these separated the two major

---

**Table 2.** Comparison of the estimated and actual genomic calculations for a BglII restriction digest of the *Ursus arctos* draft genome.

<table>
<thead>
<tr>
<th></th>
<th>Avg Frag Size (d)</th>
<th>Genome Size (G)</th>
<th># Unique Fragments (D)*</th>
<th>Read Depth Per Individual</th>
<th>Max Genomic Coverage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Predicted</strong></td>
<td>3465</td>
<td>2,777,069,268</td>
<td>93,678</td>
<td>53</td>
<td>0.82</td>
</tr>
<tr>
<td><strong>Genome</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>* Diff**</td>
<td></td>
<td>Percent Diff</td>
</tr>
<tr>
<td><strong>Estimated</strong></td>
<td>3465</td>
<td>2,777,069,268</td>
<td>93,678</td>
<td>53</td>
<td>0.82</td>
</tr>
<tr>
<td><strong>Genome</strong></td>
<td></td>
<td></td>
<td>* Diff**</td>
<td>11.8</td>
<td>25.5</td>
</tr>
</tbody>
</table>

* Includes only genomic fragments between 100 and 750 bp

** Unpublished data (Pers. comm. Axel Janke, Senckenberg Institute, Germany)

** A minus sign indicates underestimates

---

**Figure 2.** Filtering Criteria Applied to Putative SNPs. 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. The first four filtering criteria (FC) were applied in parallel as each file contained different information. The orange boxes indicate filtering criteria that were applied using the software blastn and the draft genome assembly.

doi: 10.1371/journal.pone.0081012.g002

---

**SNP Calling**

A total of 57 haploid SNPs were identified from the mtDNA sequences. Fifty-four of these separated the two major haplotypes.
SNP Discovery in Scandinavian Brown Bear

maternal haplotypes that distinguish the eastern European lineage from the western European lineage. We chose four of these mtDNA SNPs for lineage identification in the final SNP-chip.

The Y-chromosome sequences showed no variable sites, concurring with the theory that mammalian Y-chromosomes have low levels of nucleotide diversity [51]. While this does not allow for enhanced resolution of data on paternal lineage, it remains useful for sex-determination. We therefore developed two monomorphic “SNPs” based on the Y-chromosome sequences by designing assays around one non-variable nucleotide.

A total of 1.4M stacks (i.e. consensus sequences generated from sets of co-aligning reads representing restriction products) were created with aligned reads both within and among individuals (n=9). Of these, 105k (14%) contained at least one putative SNP, although this is likely to be an overestimate of the true number of SNPs due to presence of sequencing errors. Mean read depth within each individual ranged from 3X to 8X and is likely to be underestimated due to duplicate stacks resulting from the use of stringent parameters. Although we estimated an expected read depth of 40X, only 32% of the sequence data generated was utilized for creating stacks and, as such, expected depth was reduced to approximately 12X per individual. Nevertheless, read depth was sufficient to reliably call SNPs as shown by our validation results below.

After the initial filtering criteria (i.e. one SNP per read, SNP located in the middle of the read, and representation of all three haplotypes) were applied, 4,612 putative SNPs remained. These SNPs were then further reduced to 1,162 after application of the additional filtering criteria using the draft genome assembly. Figure 2 depicts the filtering process in more detail.

SNP Validation

Our final panel of putative SNPs included 144 nuclear SNPs, four mtDNA markers and two Y-chromosome sex-determination markers. We used a panel of 68 individuals from throughout Sweden (including the initially sequenced individuals) for validation. A total of 144 of the 150 SNPs (96%) produced good results. Of the six that failed, one gave no signal, one was monomorphic, and the remaining four did not pass the control checks. Since we ran two chips with 96 SNPs each, we effectively ran 96 of the SNPs twice as a control. In addition, we included both negative (water in place of DNA) and positive controls. The positive controls included duplication of some of the samples including those that were originally sequenced. The working SNPs passed all of the control checks and we did not detect a single error (error rate < 0.001). Figure 3a shows two representative scatterplots of successful SNPs.

The four mitochondrial markers and two Y-chromosome markers segregated according to expectations. Since mitochondria are haploid, there is no possibility for heterozygotes to exist. As expected, scatterplots of these SNPs display only two distinct clusters each representing one of the two possible alleles (Figure 3b). Similarly, since the Y-chromosome markers were intentionally monomorphic, there should be no possibility for either heterozygotes or a second allele and clusters should contain only male samples. The scatterplots indeed show only one cluster (one allele) and contained male samples as verified with demographic data (Figure 3c).

While it has been documented that there are several subpopulations within the northern population [52], analysis of our original mitochondrial sequences from 10 individuals identified only two haplotypes representing the northern and southern populations. However, the genotyped individuals in the validation run revealed a third haplotype indicating maternal-based substructure within the northern population in concordance with [52]. With our data, we therefore recognize three mitochondrial-based haplotypes: the North A (ABAA), North B (AAAA), and South (BBBB) with Ua03mt, Ua04mt, Ua05mt, Ua07mt markers respectively.

SNP-Chip Development

To reduce the 144 working SNPs to the 96 represented on the chip, we included four mtDNA markers, two Y-chromosome markers and subsequently selected the autosomal SNPs with the highest minor allele frequency (MAF) (valid for the Scandinavian population) and that demonstrated a clear divergence of clusters in the scatterplot. These 96 SNPs were further analyzed for MAF (mean= 0.39), H and HWD and FST (Table S1). After removing outliers (n=18) based on close or distant relatedness (see methods) and sorting into subpopulations by mitochondrial lineage, seven SNPs remained significant, but only within one of the three subpopulations for Hardy-Weinberg disequilibrium (HWD). A linkage disequilibrium analysis (Figure 4) revealed that two pairs of SNPs were linked (D' = 0.9998, 0.9411 respectively). It is likely that there are more pairs that are in high linkage disequilibrium, as would be expected when there are less chromosomes than SNPs [28]. However, further investigation using the draft genome assembly and the pairs of SNPs with high D' values revealed that these two pairs of SNPs were found to be in close proximity to each other on neighboring scaffolds, thereby confirming linkage. However, the integrated fluid circuit of the Fluidigm Biomark (Fluidigm Corporation, San Francisco, USA) is not prespotted allowing for easy replacement of individual SNP assays by any lab operating the Biomark. Future configurations thus allow for the replacement of one SNP per linked pair with an unlinked SNP.

We determined that three of the 96 SNPs most likely occur on the X-chromosome. In all three cases, all male samples (n=36) were homozygous for the same allele whereas female samples (n=32) were either homozygous or heterozygous. The chance of a Type I error (i.e. all 36 males appearing as homozygous by chance) in inducing loci that are on the X-chromosome with 36 male samples and a MAF of 0.31 (our lowest MAF for X-chromosome SNP) is one in 535 million. We therefore feel confident in stating that these SNPs occur on the X-chromosome. These SNPs will be advantageous when determining parentage by allowing additional exclusion power in cases where alleles are not in concordance with putative parent-offspring pairs.
Figure 3. SNP Scatterplots. Scatterplots generated by the Biomark system (Fluidigm, San Francisco, USA) showing allelic clustering based on fluorescence for a) autosomal SNPs, b) mitochondrial haplotype markers, and c) Y-chromosome monomorphic sex-determination markers with male fluorescence.

doi: 10.1371/journal.pone.0081012.g003
Relatedness

The population structure analysis based on Structure [45–48] resulted in K = 3 subpopulations as the most likely scenario (Figure 5) based on the ln probability of data being the lowest of all Ks as described in the documentation. The three mitochondrial haplotypes (North A, North B and South) are well matched to the three autosomal-based subpopulations with only six individuals of 50 having mismatching haplotypes when compared with the individuals' major population assignment for Structure results.

To assess the performance of the SNP set to determine a minimum of first-order relatedness (i.e. parent-offspring or full siblings; $r=0.50$), we calculated the Lynch-Ritland [49]
relatedness estimator \( (r) \). We filtered for all dyads whose values were greater than 0.40 \( (n=18; \text{Table 3}) \). Among these were two known parent-offspring dyads from our sample set \( (r = 0.6176, 0.4922) \). In addition, we were able to exclude parent-offspring relatedness in one dyad \( (r = 0.5725) \) based on X-chromosome data. Out of these 18 dyads with possible first-order relatedness, we could exclude 14 as parent-offspring as a result of them having one or more loci with 0 shared alleles (parent-offspring dyads will always have at least one allele in common by descent at all loci). The two known parent-offspring dyads are included in the four that cannot be excluded as parent-offspring thereby confirming accurate genotyping. As we genotype more individuals in a subpopulation, we will obtain more accurate MAFs, which will allow for more precise estimates of relatedness. These results indicate that the SNP set holds much promise for relatedness analyses.

Figure 5. Inferred Population Structure. Inferred population structure based on autosomal SNPs using Structure \( (\text{burnin period 100,000 cycles; 500,000 MCMC reps}) \) with \( K=3 \), sorted by \( Q \). The coloured circles below each bar represent the mitochondrial haplotype. The colour was chosen based on the bar plot colours where the majority of the mitochondrial lineage (North A, North B, South) is found.

doi: 10.1371/journal.pone.0081012.g005

Table 3. Pairwise relatedness estimates using the Lynch-Ritland \( r \) estimator \[49\] on all validation samples \( (n=68) \) where \( r \geq 0.40 \), thus indicating possible first order relationships.

<table>
<thead>
<tr>
<th>Ind_1</th>
<th>Ind_2</th>
<th>Sex</th>
<th>Hap</th>
<th>( r_{xy} )</th>
<th>2.5</th>
<th>97.5</th>
<th>P-O</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ua12</td>
<td>Ua32</td>
<td>M-M</td>
<td>NA</td>
<td>0.44</td>
<td>0.24</td>
<td>0.60</td>
<td>Excluded</td>
</tr>
<tr>
<td>Ua13</td>
<td>Ua37</td>
<td>M-F</td>
<td>S</td>
<td>0.44</td>
<td>0.24</td>
<td>0.60</td>
<td>Excluded</td>
</tr>
<tr>
<td>Ua13</td>
<td>Ua72</td>
<td>M-M</td>
<td>S</td>
<td>0.41</td>
<td>0.18</td>
<td>0.66</td>
<td>Excluded</td>
</tr>
<tr>
<td>Ua19</td>
<td>Ua65</td>
<td>F-M</td>
<td>NA</td>
<td>0.57</td>
<td>0.45</td>
<td>0.73</td>
<td>Excluded</td>
</tr>
<tr>
<td>Ua23</td>
<td>Ua93</td>
<td>M-M</td>
<td>S</td>
<td>0.48</td>
<td>0.36</td>
<td>0.59</td>
<td>Possible</td>
</tr>
<tr>
<td>Ua41</td>
<td>Ua93</td>
<td>M-M</td>
<td>S</td>
<td>0.41</td>
<td>0.22</td>
<td>0.57</td>
<td>Excluded</td>
</tr>
<tr>
<td>Ua42</td>
<td>Ua46</td>
<td>M-F</td>
<td>S</td>
<td>0.46</td>
<td>0.32</td>
<td>0.59</td>
<td>Possible</td>
</tr>
<tr>
<td>Ua43</td>
<td>Ua73</td>
<td>M-M</td>
<td>S</td>
<td>0.41</td>
<td>0.24</td>
<td>0.55</td>
<td>Excluded</td>
</tr>
<tr>
<td>Ua50</td>
<td>Ua51</td>
<td>M-M</td>
<td>NB</td>
<td>0.54</td>
<td>0.33</td>
<td>0.67</td>
<td>Excluded</td>
</tr>
<tr>
<td>Ua51</td>
<td>Ua93</td>
<td>M-M</td>
<td>NA</td>
<td>0.44</td>
<td>0.28</td>
<td>0.59</td>
<td>Excluded</td>
</tr>
<tr>
<td>Ua62</td>
<td>Ua19</td>
<td>M-M</td>
<td>NB</td>
<td>0.43</td>
<td>0.26</td>
<td>0.57</td>
<td>Excluded</td>
</tr>
<tr>
<td>Ua85</td>
<td>Ua100</td>
<td>M-F</td>
<td>NA</td>
<td>0.49</td>
<td>0.33</td>
<td>0.65</td>
<td>Excluded</td>
</tr>
<tr>
<td>Ua88</td>
<td>Ua98</td>
<td>F-F</td>
<td>NB</td>
<td>0.55</td>
<td>0.42</td>
<td>0.69</td>
<td>Excluded</td>
</tr>
<tr>
<td>Ua88</td>
<td>Ua97</td>
<td>F-F</td>
<td>NB</td>
<td>0.44</td>
<td>0.24</td>
<td>0.63</td>
<td>Excluded</td>
</tr>
<tr>
<td>Ua96</td>
<td>Ua97</td>
<td>F-F</td>
<td>NB</td>
<td>0.42</td>
<td>0.22</td>
<td>0.57</td>
<td>Excluded</td>
</tr>
<tr>
<td>Ua98</td>
<td>Ua96</td>
<td>F-F</td>
<td>NB</td>
<td>0.40</td>
<td>0.23</td>
<td>0.56</td>
<td>Excluded</td>
</tr>
<tr>
<td>Ua98</td>
<td>Ua96</td>
<td>F-F</td>
<td>NB</td>
<td>0.42</td>
<td>0.31</td>
<td>0.75</td>
<td>Known</td>
</tr>
<tr>
<td>Ua100</td>
<td>Ua101</td>
<td>F-M</td>
<td>NA</td>
<td>0.46</td>
<td>0.36</td>
<td>0.60</td>
<td>Known</td>
</tr>
</tbody>
</table>

Sex refers to whether the individual is male \( (M) \) or female \( (F) \). Hap refers to the mitochondria haplotype \( (\text{North A = NA, North B = NB, South } = S) \) of both individuals in the pairs \( \text{none of the pairs differed} \). P-O indicates possible parent-offspring dyads determined by identifying pairs that share at least one allele at every locus and additionally in one pair \( (\text{Ua19 and Ua85}) \) through analysis of the X-chromosome. The two known parent-offspring dyads were confirmed by the presence of at least one shared allele at every locus.

doi: 10.1371/journal.pone.0081012.t003
Conclusions

We present a new panel of 96 SNPs suitable for assaying the Scandinavian brown bear for relatedness and other ecological and evolutionary analyses. Through application of an NGS based RRL approach, we successfully reduced the computational power required to the extent that most analyses were performed on a standard-specification personal computer. This was made possible by eliminating sequences (within the limits of the study) that did not meet strict quality control (eg. inclusion of cut site, questionable quality of putative SNP) and avoiding the often problematic, computationally demanding and error-prone step of sequence assembly through the use of one restriction enzyme. While some applications require a greater number of SNPs, other applications may actually become disadvantaged by too much information. This is likely the case for relatedness studies and thus allowed us the freedom to rapidly decrease the amount of data we analyzed.

This SNP-chip holds much promise for conservation of the Scandinavian brown bear, particularly for the southern population, which is one of the few relic western European populations. There are many potential uses for this SNP-chip including the use of relatedness estimates to monitor the genetic health, identify mating patterns and reproductive success, and track individual movements. It can also be useful for estimating population size based on individual identification, detecting hybridization events between the northern and southern populations, and confirming paternity in possible multiple paternity events or cases of infanticide.

Supporting Information

Table S1. dbSNP submitted SNP (ss#) numbers and descriptive statistics for autosomal SNPs.

Acknowledgements

We thank Helena Königsson for her support in the lab. We thank Arne Söderberg (SVA) for providing us with samples. We thank Axel Janke and Björn Hallström for providing us with the draft genome. We thank Jonas Kindberg and the Scandinavian Brown Bear Project for input and logistical support. We thank Max Käller and Anna Skillermo for their support in library preparation and sequencing. We would like to acknowledge support from Science for Life Laboratory, the National Genomics Infrastructure, and Uppmax for providing assistance in massive parallel sequencing and computational infrastructure. We thank Matthew A. Cronin for helpful comments resulting in the improvement of the manuscript.

Author Contributions

Conceived and designed the experiments: AJN NRS GS. Analyzed the data: AJN GS. Wrote the manuscript: AJN NRS GS. Bioinformatics: AJN NRS GJ.

References

SNP Discovery in Scandinavian Brown Bear


Estimating population size using single nucleotide polymorphism-based pedigree data

Robert Spitzer¹, Anita J. Norman³, Michael Schneider³ & Göran Spong²⁴

¹Wildlife Ecology Group
²Molecular Ecology Group
Department of Wildlife, Fish and Environmental Studies
Swedish University of Agricultural Sciences
SE-901 83 Umeå
Sweden
³Västerbotten County Administration
SE-901 86 Umeå
Sweden

⁴Forestry and Environmental Resources
College of Natural Resources
North Carolina State University
Raleigh 27695
USA

Keywords: population estimate, pedigree reconstruction, rarefaction, single nucleotide polymorphism, SNP, non-invasive sampling, brown bear

Corresponding Author:
Robert Spitzer
Wildlife Ecology Group
Department of Wildlife, Fish and Environmental Studies
Swedish University of Agricultural Sciences
SE-901 83 Umeå
Sweden
Robert.Spitzer@slu.se
Abstract

Reliable population estimates are an important aspect of sustainable wildlife management and conservation but can be difficult to obtain for rare and elusive species. Here we test a new census method based on pedigree reconstruction recently developed by Creel and Rosenblatt (2013). Using a panel of 96 single nucleotide polymorphisms (SNPs) we genotyped fecal samples from two Swedish brown bear populations for pedigree reconstruction. Based on 433 genotypes from central Sweden (CS) and 265 from northern Sweden (NS), the population estimates ($N = 630$ for CS, $N = 408$ for NS) fell within the 95% CI of the official estimates. The precision and accuracy improved with increasing sampling intensity. Like genetic capture-mark-recapture methods, this method can be applied to data from a single sampling session. Pedigree reconstruction combined with noninvasive genetic sampling may thus augment population estimates, particularly for rare and elusive species for which sampling may be challenging.

Introduction

Estimates of the population size and its fluctuations are often fundamental for understanding ecological, behavioral or genetic processes (Ojaveer et al., 2004, Valderrama et al., 2013, Dochtermann and Peacock, 2013) and practically indispensable for management and conservation (Katzner et al., 2011). This includes estimates of both effective and true population size, where the former is usually based on genetic data and models, while the latter typically use some form of census data, sometimes genetic. For example, such population size and trend estimates help identify particular factors that drive population dynamics, and are hence critical for modeling the future of a population under different management scenarios (Lewellen and Vessey, 1998). Moreover, estimates of true population size and trend are the basis for adaptive harvest quotas (Wilson and Delahay, 2001) as well as identifying populations under threat of becoming endangered or extinct (Vié et al., 2009). However, reliable estimates are difficult to obtain. This is especially true for rare and elusive species, which are frequently of high conservation concern (Rolland et al., 2011). Large carnivores are no exception to this (Creel and Rosenblatt, 2013, Kindberg et al., 2009) as they are generally solitary and cryptic, in addition to occurring at low densities and across large home ranges. For several ecological, economic and societal reasons, large carnivores in particular receive a disproportional amount of attention from research, conservation and management. For example, carnivores may strongly affect the ecosystems they occupy, by changing the behavior of other carnivores and by direct and indirect effects on prey, which can lead to downstream effects on primary production (Creel, 2007). In some areas, carnivores pose a threat to humans or come into direct conflict with human husbandry practices causing economic losses. For these, and more reasons, a range of remote or non-invasive methods have therefore been employed to study carnivores (Jackson et al., 2006, Kojola et al., 2014). An increasingly popular and cost efficient approach is to use non-invasive genetic sampling for assessment of the number of individuals in a population (e.g. Sugimoto et al., 2012, Stansbury et al., 2014, Mowry et al., 2011) often by collecting fecal samples during other management activities or by citizen volunteers (e.g. Kindberg et al., 2011).

Genetic data may also be used to assess a population’s effective population size, an important parameter especially for small populations at risk of inbreeding or genetic drift. The framework of population genetics provides several way of inferring the effective population size, but the estimators suffer from being slow to respond to recent events, instead showing historic averages (Palsboll et al., 2013). To obtain more contemporary estimates, genetic data can be also used to
derive demographic data used for calculating the current effective population size (e.g. Creel, 2002). For many studies in ecology, however, the actual population size is a more important parameter to know than the underlying effective population size. Also in conservation, much focus has been placed on the effective population size. Yet, as pointed out by Lande (1988), the drivers of extinction are primarily habitat loss and overharvest, not lack of genetic variation. So while it is informative to know the effective population size, once its relationship to the actual population size has been determined, its continuous monitoring may be less important than knowing the actual population size. This is equally true for critically endangered populations at the verge of extinction as for larger populations not acutely threatened. Here the actual population size is typically what management operates on when setting targets for quotas, dispersal events or the population size and distribution.

Most statistical methods for estimating population size rely on multiple sampling events, known as capture-mark-recapture (CMR) techniques which are comprehensively discussed by Krebs (1999) and Sutherland (2006). A distinct disadvantage of classical CMR methods lies in the circumstance that the physical capture, particularly of large predators, is often impractical, costly and potentially harmful to both sides (Muñoz-Igualada et al., 2008, Mowat et al., 1994, Logan et al., 1999). In addition, differences in catchability resulting from trap-shy or trap-happy individuals could introduce systematic trapping bias. Such differences in personality traits (Sih and Bell, 2008) have been documented for many species, including badgers (Tuyttens et al., 1999), stoats (King et al., 2003) or rabbits (Sunnucks, 1998).

Newer methods, such as camera trapping, have largely made classical CMR approaches obsolete in studies of large animals. But many cameras are needed to reach reasonable detection probabilities (and cameras are sometimes removed or destroyed by humans or other animals). Even more problematic is that relatively few species are reliably individually identifiable from photographs. In contrast, an individual’s genotype is a unique and permanent mark. Non-invasively collected DNA samples (e.g. from feces or hair) in combination with molecular techniques offer another non-invasive alternative (Kohn and Wayne, 1997, Swenson et al., 2011, Taberlet et al., 1999, Waits and Paetkau, 2005). In direct genetic census methods the genotype simply becomes a ‘molecular tag’ (Schwartz et al., 2007) which replaces traditional means of identification like ear marks or leg bands. Genotypes can thus be used as molecular tags in a CMR framework. But genetic data contain more information than just individual genotypes, such as information on pedigree structures in the population. From such information, unsampled individuals could potentially be inferred by their genetic fingerprint and included into the population estimates.

In 2013, Creel and Rosenblatt suggested a new, pedigree-based estimator for total population size. They evaluated the performance of their method through simulations parameterized with demographic data of African lions (Panthera leo) from Zambia. The method, henceforth referred to as the Creel-Rosenblatt estimator (CRE), incorporates the sum of sampled individuals ($N_s$), number of breeders ($B_s$), number of individuals inferred from pedigree reconstruction ($N_{in}$), and the estimated number of individuals that did not breed nor were sampled (rendering them invisible to pedigree reconstruction) into the population estimate. As such, it purports to increase the precision of genetically based population estimates.
As other genetically based CMR methods, the CRE requires only one sampling event (although multiple sampling events are also possible). This makes it a useful extension to the suite of tools available to estimate population sizes under circumstances where repeated sampling is difficult. In addition to being a novel approach for estimating population size, pedigree reconstruction can be used to investigate population structure (Pemberton, 2008, Calboli et al., 2008), mating behavior (Pemberton et al., 1992) or dispersal (Norman and Spong, 2015). The ideal genetic marker for pedigree reconstruction should provide high genomic resolution and be geared towards providing reliable relatedness estimates (Creel and Rosenblatt, 2013). Single nucleotide polymorphisms (SNPs) have proven to be a powerful tool for studying genetic variation in populations (Brumfield et al., 2003, Morin et al., 2004). Compared with microsatellites, another type of frequently used genetic marker, SNPs offer lower error rates from mistyping and allelic dropout (Morin and McCarthy, 2007, Norman et al., 2013). They are also reproducible across laboratories and are cheaper, allowing for higher genomic resolution within a given economic frame (Anderson and Garza, 2006). Because only short intact sequences, typically 50-70 bp, of DNA are required for successful amplification, SNPs are especially suitable when working with degraded DNA, as is usually the case with non-invasively obtained samples (Morin et al., 2004). Here, we use a panel of 96 SNPs recently developed for studying relatedness in the Scandinavian brown bear (Ursus arctos, Fig. 1) population (Norman et al., 2013). We reconstructed pedigrees based on hunter-collected feces and apply the CRE method to estimate the size of the brown bear populations in the Swedish counties of Dalarna, Gävleborg and Västerbotten. Already existing population estimates for the brown bear in these areas (Kindberg and Swenson, 2013, Kindberg and Swenson, 2015) provide us with a benchmark that can be used to empirically assess the performance of the CRE outside of a simulation environment making this study system appropriate. For further comparison we also performed rarefaction analyses to estimate population size. This constitutes the first time that the estimator is applied to empirical data, as we were unable to find any reference describing the application of this method in the Web of Science™ publication database, with the last search completed on 17 November 2015.

**Materials and methods**

**Study area and sample collection**

The two study areas in central and northern Sweden encompassed the Swedish counties of Dalarna and Gävleborg (ca. 46,300 km²) and Västerbotten (ca. 55,200 km²), respectively. To the west these areas are delimited by the Scandinavian mountain range and to the east by the Baltic Sea (Fig. 2). The southern border of Dalarna-Gävleborg also demarcates the approximate southern limit of the brown bear distribution in Sweden. Dalarna-Gävleborg is home to an estimated number of 793 bears, 95% CI [621, 1179] (Kindberg and Swenson, 2013). In 2014, the Västerbotten population was estimated to be 362 bears, 95% CI [310, 459] (Kindberg and Swenson, 2015). Studies of maternally-inherited mitochondrial DNA (mtDNA) have shown that brown bears in Sweden belong to two genetically distinct lineages with approximately 7% differentiation between them (Taberlet and Bouvet, 1994). The western lineage, found in south-central Sweden, originated from the Iberian refugium during the last ice age (today’s France and Spain), whereas the eastern lineage, found throughout northern Sweden, can be traced to Karelia in Russia (Taberlet and Bouvet, 1994). At present the two lineages remain largely separated around a well-documented contact zone at the height of Östersund in central Sweden (Taberlet et al., 1995). Population monitoring could be especially important in case of the western mtDNA haplotype (southern population) which is only found in Europe whereas the eastern haplotype is
also prevalent in Asia and North America (Hirata et al., 2013, Korsten et al., 2009, Saarma et al., 2007, Waits et al., 2000).

Fecal samples were collected by volunteers, predominantly moose (Alces alces)-hunters, following the protocol of Bellemain et al. (2005) and Kindberg et al. (2011) during the periods of August-October 2012 in Dalarna-Gävleborg and August-December 2014 in Västerbotten. Volunteers recorded collection date and coordinates of the sample location and mailed this information together with their samples to the county administrations (in the case of the Dalarna-Gävleborg collection) or in Västerbotten directly to the Molecular Ecology Group at the Swedish University of Agricultural Sciences (SLU) in Umeå. Upon arrival samples were stored in 70% ethanol solution at -20°C as recommended by Frantzen et al. (1998).

**Molecular analysis**

DNA extraction from the Dalarna-Gävleborg samples was carried out by Bioforsk, Norway (Hagen and Aarnes, 2013) following procedures described by Schregel et al. (2012). In Västerbotten, DNA extraction was performed at SLU using a QIAsymphony SP (Qiagen; Hilden, Germany) robot according to the manufacturer’s instructions. SNPs were genotyped on a Fluidigm Biomark™ (Fluidigm Corporation; San Francisco, USA) using the 96 SNP panel developed by Norman et al. (2013). Since its first publication the panel has undergone slight modifications (e.g. two linked SNPs were substituted with Y-chromosome SNPs) and now consists of 85 autosomal SNPs, four mtDNA SNPs as well as four Y-chromosome and three X-chromosome markers for sex determination (Norman and Spong, 2015). Each run included negative controls with water in place of DNA. The genotype clusters assigned by the Biomark software were manually screened and loci of questionable cluster affiliation were invalidated and removed from subsequent analyses. Species and sex were assigned according to the following criteria:

- **bear** = mtDNA SNP calls ≥ 3
- **male** = Y-chromosome SNP calls ≥ 3
- **female** = Y-chromosome SNP calls = 0 and X-chromosome SNP calls ≥ 2

The above criteria were designed to avoid possible misidentification of poorly amplified male samples as females. In males, Y and X markers occur in equal proportion. The requirement that at least two out of three X-markers had to amplify for samples to be called female makes it extremely unlikely that such a sample was a male that had not amplified for the Y markers. As we only included samples that had amplified for more than 70 loci, the risk of having none of four Y markers but two of the X amplify for a male is 1x10^-4.

**Pedigree reconstruction**

To reconstruct pedigrees we used FRANz software version 2.0.0 (Riester et al., 2009) which uses Markov Chain Monte Carlo (MCMC) simulation for estimating the statistical confidence of parentage inference. The software requires specifying an approximate maximum number of females and males (Nfmax and Nmmax) to avoid an empty pedigree due to convergence of the Markov Chain to a very high number of individuals (Riester et al., 2009). We used the estimates from rarefaction analysis and the sex ratio present in the genotyped samples to set Nfmax/Nmmax to 538/419 (Dalarna-Gävleborg) and 249/239 (Västerbotten) respectively.
Typing errors were empirically determined to $1.538 \times 10^{-4}$ for Dalarna-Gävleborg and 0.01 for Västerbotten. The error rates from the two areas differ. This is because samples from Dalarna-Gävleborg held the best available quality extract from each individual successfully genotyped with microsatellites at Bioforsk, whereas the error rate for the Västerbotten samples includes all samples that passed the amplification threshold for SNP genotyping. Since microsatellite genotyping requires much higher quality DNA the error rate of such samples becomes much lower. The maximum likelihood pedigrees produced by FRANz identify the putative sire and dam of sampled individuals. We further verified the FRANz reconstructed pedigrees by calculating the Lynch-Ritland relatedness coefficient ($r$) (Lynch and Ritland, 1999) for all identified parent-offspring (PO) and full-sibling (FS) pairs using COANCESTRY version 1.0.1.2. (Wang, 2011). We chose the Lynch-Ritland relatedness coefficient because it has been found to have the lowest rate of misclassification and lower overall variance compared to other pairwise relatedness estimators (Csillery et al., 2006, Stone and Björklund, 2001).

Population estimates

Rarefaction, also referred to as accumulation-curve method, has traditionally been used to estimate species diversity in an area by plotting the cumulative number of newly recorded species against the total number sampled (Colwell and Coddington, 1994). The same underlying logic can be applied for estimating population size by substituting the species count with the number of unique individuals/genotypes. As suggested by Kohn et al. (1999) a curve defined by the equation $y = ax / (b+x)$ was fitted to our data. In this model $y$ equals the number of unique genotypes, $x$ corresponds to the number of samples (genotyped feces), $b$ is the rate of decline in the slope and the asymptote $a$ represents the estimated population size (Bellemain et al., 2005). We calculated the parameters $a$ and $b$ through nonlinear iterative regression using the statistical software package JMP Pro version 11.0.0 (SAS Institute). To account for the variance caused by the order in which samples are drawn we repeated this process 100 times with random iterations of the genotype sampling order and used the mean of the resulting asymptotes as the rarefaction population estimate.

For the pedigree-based population estimates we followed the recommendations by Creel and Rosenblatt (2013) and specified the number of individuals sampled ($N_s$) as the number of individual genotypes, known breeders ($B_d$) as those individuals who had progeny in the pedigree and inferred individuals ($N_{iu}$) as the missing parent in known parent-offspring dyads. However, assuming that each missing parent in the dyads constitutes a new individual would most likely cause an overestimation because brown bear males are known to mate with several females and vice versa (Steyaert et al., 2012). For example, an inferred sire may be the missing father in more than one of the mother-offspring dyads. Therefore, we used the improbable scenario in which the number of inferred individuals ($N_{iu}$) equals the number of dyads in the pedigree only for approximating an upper bound of the population estimate. For a more realistic estimate that accounts for multiple parentages, we first screened all parent-offspring dyads in the pedigree for individuals with several offspring. If pairwise comparisons of the Lynch-Ritland relatedness suggested full-siblings ($r \sim 0.5$) among those offspring we inferred only one new individual (the missing parent) from these dyads. For the remaining cases we used a different approach where we assumed that the likelihood of sampling each sex was equal: we determined the ratio of all the known individual dams to the known individual sires in the pedigree and then used this ratio to infer the missing counterparts from the individual dams and sires in the pedigree dyads. In this
way, the ratio of dams to sires with the inferred individuals included remains the same as it was in the original pedigree.

Another problem pointed out by Creel and Rosenblatt (2013) is the circumstance that there is no way to ascertain how many of the inferred individuals are actually still alive at the time of the estimate. To account for mortality among inferred individuals we assumed them to be at the typical breeding age of ~5 years (Swenson et al., 2001) and applied the age-specific annual mortality rates as reported in Nilsson (2013) of 7.2% to inferred dams and 11.6% to sires respectively.

Finally, we assessed the accuracy of the CRE results by comparing them to the official population estimates (Kindberg and Swenson, 2013, Kindberg and Swenson, 2015) and to the results of rarefaction analysis. Because there is currently no method to assign confidence limits to the CRE population estimates, we estimated an upper and a lower bound. For the lower bound, we simply used the count of sampled genotypes. For the upper bound, we treated $N_o$ as equal to the number of dyads in the pedigree and assumed zero mortality among the inferred individuals. To test the performance of the CRE at different sampling intensities we used the data from Västerbotten due to the higher sampling coverage (approximately 73% of the population included in the sample) compared to only 55% in Dalarna-Gävleborg. Varying the sampling intensity from 10% to 60% of the official population estimate, we applied the CRE to ten replicates of samples randomly drawn in correspondence with each sampling intensity level. Not having 100% of the population included in the sample is a common limitation in studies based on field data, but validation of simulation results using empirical data may still reveal strengths and weaknesses.

We assumed that the pedigree would become less complete (contain fewer parent-offspring pairs) the further away the sampling occurred from the core sampling frame. This is because breeding individuals in peripheral areas might have moved beyond the borders of the sampling area and therefore may have been missed during the sample collection. In both areas, Dalarna-Gävleborg and Västerbotten, the only true population border is the Baltic Sea to the east. To the north and the south, bears occur beyond the borders of the sampling areas. Most interesting is the border to the west, formed by the Scandinavian mountain range. Because mountain terrain can be difficult to access and because moose hunting is of lower intensity at higher altitudes, the sampling effort by volunteers was lower there than in other areas. If the pedigree were to show similar levels of incompleteness along the western border compared with the “open” borders to the north and the south, it could indicate that bears in the mountains were missed in the sampling. If this were the case it would lead to an underestimation of the population size. If, on the other hand, the mountains form a true border like the Baltic Sea, then the pedigree should be equally complete in both these locations.

From the coordinate data that was provided along with the fecal samples, the median centers of all known locations for an individual were calculated using R (R Development Core Team, 2008). We considered the median to be less biased than the mean because of its lower sensitivity to outliers. Inferring home ranges from the locations of fecal samples is prone to errors but Bellemain et al. (2005) reported that the majority of fecal sites fall inside the home range or within 10 km of it. We determined the center point of the sampling area as the median center of
all individual locations using the GIS package ArcMap version 10.2.2 (ESRI, 2014). We then sampled the individuals closest to the center point and the four borders (North, South, East & West) respectively at sample sizes of \(n = 100\) for Dalarna-Gävleborg and \(n = 70\) in Västerbotten. The number of samples for Västerbotten had to be lower to avoid overlap because fewer individuals in total were available to sample from. In a second step we also sampled males and females separately (Dalarna-Gävleborg, \(n = 50\); Västerbotten, \(n = 30\)) to investigate if there are detectable differences between mother-daughter and father-son dyads.

To test for differences in the completeness of the pedigree we used Pearson’s Chi-square test for homogeneity of proportions with the proportions corresponding to the number of parent-offspring pairs in the pedigree per number of sampled individuals. To further test whether there is a spatial effect on parent-offspring pairs in sex-separated pedigrees, we also sampled males and females randomly across the whole sampling area (Dalarna-Gävleborg, \(n = 50\); Västerbotten, \(n = 30\)).

Results
We successfully genotyped 433 individuals (243 females, 190 males) for Dalarna-Gävleborg and 265 individuals (136 females, 129 males) for Västerbotten. Rarefaction analysis was based on 873 samples from Dalarna-Gävleborg and 677 from Västerbotten. The maximum frequency at which an individual occurred in the sample was 19 for Dalarna-Gävleborg and 16 for Västerbotten respectively. The amplitudes of the curves fitted to the rarefaction data suggested population sizes of \(N = 895\) (Dalarna-Gävleborg) and \(N = 484\) (Västerbotten).

Table 1 summarizes the results of the FRANz reconstructed pedigrees. The proportions of breeders in both samples (0.37 Dalarna-Gävleborg, 0.42 Västerbotten) are not significantly different (\(z = -1.42, p = 0.16\)) and the ratios of dams to sires are also very similar. As shown in Fig. 3 the mean pairwise Lynch-Ritland relatedness coefficient \((r)\) for PO dyads and FS pairs did not significantly differ from the expected value of \(r = 0.5\) for first-order relatives (PO: \(t(292) = 0.33, p = 0.74\); FS: \(t(39) = 1.60, p = 0.12\)) which corroborates the reconstructed pedigrees.

In Dalarna-Gävleborg we inferred 6 sires and 4 dams directly from full-sibling relationships among the known parent-offspring dyads. An additional 52 sires and 65 dams were inferred using the dam to sire ratio approach. After correcting for mortality, the total number of inferred individuals (\(N_{in}\)) (i.e. those missing from the pedigree) was 115. For Västerbotten, screening the parent-offspring dyads for full-siblings yielded 4 sires and 3 dams whereas the ratio method suggested a further 41 sires and 45 dams resulting in \(N_{in} = 85\) after mortality correction. Therefore, applying CRE to these numbers (\(N_i\) and \(B_i\) from Tab. 1 and \(N_{in}\)) resulted in estimated population sizes of \(N = 630\) for Dalarna-Gävleborg and \(N = 408\) for Västerbotten. Comparison with official bear population estimates shows that both CRE results fall within the 95% CI of official estimates (Fig. 4). Using the genotype count as a measure for minimum population size and number of dyads for the estimate of \(N_{in}\) under the assumption of no mortality, the lower and upper bounds for the CRE correspond to 433 and 728 in Dalarna-Gävleborg and to 265 and 476 in Västerbotten.

Testing for effects of sampling intensity with our empirical data, we found a similar pattern as Creel and Rosenblatt (2013) did in their simulations (Fig. 5). At a sampling intensity of 10% the
coefficient of variation (CV) for the different CRE estimates was 7% and the percentage difference to the official population size estimate 157%; at a sampling intensity of 60% both CV and percentage difference decreased to 3%. This suggests that CRE population estimates increase in both precision and accuracy with increasing sampling intensity.

Our tests for edge effects of sampling boundaries revealed no significant differences in completeness of the pedigree between the core area and four peripheral border areas in Dalarna-Gävleborg, $\chi^2(4) = 7.05, p = 0.134$ or Västerbotten, $\chi^2(4) = 1.97, p = 0.74$.

When males and females were sampled separately, significantly more mother-daughter than father-son dyads were found in Dalarna-Gävleborg, $\chi^2(9) = 62.79, p < 0.0001$ (Fig. 6). When the same number of males and females ($n = 50$) were sampled randomly across the whole area there was no significant difference between the proportions of mother-daughter and father-son dyads per sampled individuals, $z = -0.521, p = 0.602$. In Västerbotten, the difference in proportions between male and female dyads was not significant, $\chi^2(9) = 15.28, p = 0.083$. The two-sample $z$-test for proportions when females and males ($n = 30$) were sampled randomly across the whole area was also not significant, $z = -0.645, p = 0.52$.

**Discussion**

In this study, we applied the recently developed Creel-Rosenblatt estimator (CRE), a pedigree reconstruction-based method, to estimate the size of two fractions of the Swedish brown bear population. SNP genotypes obtained from non-invasively collected fecal samples were used to reconstruct pedigrees from which we were able to infer the presence of additional individuals which otherwise would have remained undetected. Compared to a simple count of detected genotypes the CRE increased the population estimates by 45% for Dalarna-Gävleborg and 54% for Västerbotten. The circumstance that reliable population estimates were available prior to this study provided an excellent opportunity for testing this new census method because it allowed for verification of the results. Our pedigree reconstruction-based population estimates of the CRE fell within the confidence limits of the most recent official estimates. This is an indication that the method provides a potential alternative to traditional CMR approaches with the added benefit that it can be employed using data from a single sampling event. While their percentage relative precision (PRP), a measure which relates a population estimate to its 95% confidence limits (Sutherland, 2006), are actually quite good for the study of natural populations (21% in Västerbotten and 35% in Dalarna-Gävleborg) their confidence limits are nevertheless wide. It is therefore difficult to ascertain how close our estimates really are to the true population size. However, the CRE results appear to be further corroborated by the rarefaction results. Simulations by Valière (2002) have shown that the model we used to extrapolate the rarefaction curves has a tendency to overestimate population size if the sampling effort is high. As shown in Figure 4 the rarefaction results indeed exceed both the CMR and CRE estimates and more so in Västerbotten where the sampling intensity was higher. This suggests that the CRE results are actually close to the true figure.

As previously demonstrated by Creel and Rosenblatt (2013), we found that the method works best if the sampling intensity exceeds ~40%. At lower sampling intensities, the estimator tends to severely underestimate the population size. This can be explained by the fact that small sample sizes usually do not contain many parent-offspring pairs which severely restricts the pedigree
reconstruction. At much higher sampling intensities (e.g. >80%) hardly any information is gained over a simple count (see Creel & Rosenblatt 2013). Moreover, the risk of overestimation also increases. If, in the extreme case, 100% of individuals were sampled and no reliable information of mortality was available, the CRE would severely overestimate the size of the population because individuals (albeit dead) would still be inferred from the pedigree (Creel and Rosenblatt, 2013). The CRE is therefore best suited for non-cyclical species with generational overlap and either low or well documented mortality rates. The published mortality rates for Swedish brown bears are likely to be accurate since natural mortality of adults is rare in comparison to hunting or traffic accidents which are closely monitored (Mörner et al., 2005). Gross overestimation due to unknown mortality rates can also be avoided by comparing the CRE estimates to those obtained from rarefaction analysis on the same data because the slope and amplitude of the rarefaction curve provide good approximations of population size and the proportion sampled. Contrary to our expectation we found no significant differences in the completeness of the pedigree when sampling individuals from the peripheries of the study area compared to the center regions. This suggests that many individuals roamed widely with frequent crossings in and out of the study area. Nor did we see a more incomplete pedigree in the west where mountains and low human population density result in a lower sampling effort, suggesting that a similar proportion of individuals are sampled in this area too. Indeed, all peripheral areas were similar, including the hard border to the east, the Baltic Sea. The circumstance that more mother-daughter than father-son dyads were found when separately sampling the same number of individuals from both sexes within a specific area, is expected as brown bears show female philopatry (Blanchard and Knight, 1991, Saarma and Kojola, 2007, Støen et al., 2005).

For well-studied populations that are regularly sampled, the CRE offers no immediate advantage over established CMR methods in terms of estimating population size. However, if sampling occurs over a number of years, the required sampling effort to maintain a desired sampling coverage should be considerably reduced as genotyped individuals accumulate. In simulations by Creel and Rosenblatt (2013) the proportion of the population that had to be sampled typically dropped to ≤ 20% within three years. The CRE could therefore prove to be useful in situations where budgetary or logistic constraints make repeated, large scale sampling events unrealistic; for example in remote regions or developing countries.

In their simulations, Creel and Rosenblatt (2013) tracked all individuals throughout the simulated period of 15 years which means they consistently had accurate information about parent-offspring relationships and mortalities. Based on these data, they were able to infer individuals (as the missing parent in parent-offspring dyads) without error. In pedigrees reconstructed from empirical genetic field data, this inference is less straightforward especially for species with non-monogamous mating behavior. Inferring the correct number of missing sires or dams continues to be a major challenge of the CRE method. The additional information provided by the Lynch-Ritland relatedness coefficient ($r$) helped to improve the resolution of the pedigree by enabling us to detect full-siblings among the parent-offspring dyads which then allowed for correct inference of the missing sire or dam. Based on the values of $r$ we suspect that there are several half-siblings which share the invisible parent. Unfortunately it is not sufficient to infer the missing parent as one individual in these cases because the coefficient only captures the degree of relatedness and not the specific relationship. Half-siblings share on average approximately 25% of alleles but the same is true for grandparent-grandoffspring and avuncular relationships.
Thus, the true relationship between two individuals can usually not be inferred from their degree of relatedness alone.

To further improve the inferences from the pedigree, information about the age of the sampled individuals is needed. If genetic relatedness can be combined with age in the analysis, the most probable relationships are easily determined. We recommend keeping track of each genotyped individual from the date it was first recorded. Even if the true age remains unknown, a minimum age can be assigned and over the course of several sampling periods, individuals can at least be compared on the basis of age relative to one another. This would considerably improve the accuracy of the pedigrees, particularly with regard to the directionality in putative parent-offspring dyads (Kopps et al., 2015) and thereby help to refine the CRE population estimates. Using the ratio of known dams and sires for inference of individuals may not fully reflect reality but given the restrictions of the data, we have shown that it results in a credible estimate.

Concurrent with the simulation results of Creel and Rosenblatt (2013), and using empirical data, we show that accurate estimates of total population size are possible from reconstructed pedigrees. The estimator is limited by the resolution of the pedigree and potentially unknown mortality rates. It therefore works best in long-lived species with lots of generational overlap and is further helped if “first seen” records are kept to give rough estimates of age. This makes the method particularly appealing for recurring sampling in the same population.

Acknowledgements
We thank the numerous volunteers who participated in the sample collection and our lab technician, Helena Königsson, for carrying out the genotyping. Jonas Kindberg from the Scandinavian Brown Bear Research Project provided valuable insights into the official population estimates. We also thank the reviewers for helpful comments that improved the manuscript.

References

CREEL, S. & ROSENBLETT, E. 2013. Using pedigree reconstruction to estimate population size: genotypes are more than individually unique marks. Ecology and Evolution, 3, 1294-1304.


### Tables

*Table 1.* Key characteristics of the reconstructed pedigrees showing the number of individuals with both parents identified (triads), one parent identified (dyads) or no identified parent. $N_s$ denotes the number of directly sampled individuals (number of genotypes) and $B_s$ corresponds to known breeders (individuals with at least one offspring in the pedigree).

<table>
<thead>
<tr>
<th></th>
<th>Dalarna-Gävleborg</th>
<th>Västerbotten</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_s$</td>
<td>433</td>
<td>265</td>
</tr>
<tr>
<td>Number of triads</td>
<td>65</td>
<td>37</td>
</tr>
<tr>
<td>Number of dyads</td>
<td>170</td>
<td>123</td>
</tr>
<tr>
<td>Number with ‘no parent’</td>
<td>198</td>
<td>105</td>
</tr>
<tr>
<td>$B_s$</td>
<td>159</td>
<td>112</td>
</tr>
<tr>
<td>Ratio of dams : sires</td>
<td>1.30</td>
<td>1.20</td>
</tr>
</tbody>
</table>
Figures

**Figure 1.** A female Scandinavian brown bear with cubs. Source: Nyhetsbyrå

**Figure 2.** Map showing the location of the study areas (blue) within Sweden (red).
**Figure 3.** First-order relatives in the reconstructed pedigrees correspond well with the expected value ($r = 0.5$) for the Lynch-Ritland relatedness coefficient.

**Figure 4.** The pedigree reconstruction-based population estimates of the Creel-Rosenblatt estimator (CRE) fall within the 95% confidence limits of the official estimates based on multiple capture-mark-recapture (CMR) techniques. Rarefaction analysis (R) using the extrapolation model suggested by Kohn et al. (1999) resulted in higher estimates.
Figure 5. The precision and accuracy of the CRE improved with increasing sampling intensity. The y-axis is on a log scale to show the changes in variance of the population estimates at different sampling intensities (precision) and their distance from the true value (accuracy) in correct proportions. The dashed line denotes the official population estimate for Västerbotten ($N = 362$) which was assumed to be the true population size (100%). The filled square represents the full set of genotypes ($n = 265$) for Västerbotten which corresponds to a sampling intensity of 73%.

Figure 6. Mother-daughter (light grey) and father-son dyads (dark grey) in the pedigree per sampled individuals ($n = 50$ in Dalarna-Gävleborg, $n = 30$ in Västerbotten) in the core and along the peripheral boundaries of the counties. Except for the core of Dalarna-Gävleborg there are generally more mother-daughter than father-son dyads in each area, indicating female philopatry.
Abstract
Quantifying dispersal within wild populations is an important but challenging task. Here we present a method to estimate contemporary, individual-based dispersal distance from noninvasively collected samples using a specialized panel of 96 SNPs (single nucleotide polymorphisms). One main issue in conducting dispersal studies is the requirement for a high-sampling resolution at a geographic scale appropriate for capturing the majority of dispersal events. In this study, fecal samples of brown bear (*Ursus arctos*) were collected by volunteer citizens, resulting in a high sampling resolution spanning over 45,000 km² in Gävleborg and Dalarna counties in Sweden. SNP genotypes were obtained for unique individuals sampled (*n* = 433) and subsequently used to reconstruct pedigrees. A Mantel test for isolation by distance suggests that the sampling scale was appropriate for females but not for males, which are known to disperse long distances. Euclidean distance was estimated between mother and offspring pairs identified through the reconstructed pedigrees. The mean dispersal distance was 12.9 km (SE 3.2) and 33.8 km (SE 6.8) for females and males, respectively. These results were significantly different (Wilcoxon’s rank-sum test: *P*-value = 0.02) and are in agreement with the previously identified pattern of male-biased dispersal. Our results illustrate the potential of using a combination of noninvasively collected samples at high resolution and specialized SNPs for pedigree-based dispersal models.

Introduction
Knowledge of dispersal patterns in wild populations can benefit research and conservation efforts, but dispersal is notoriously difficult to study (Dieckmann et al. 1999; Nathan 2001; Trakhtenbrot et al. 2005; Driscoll et al. 2014). This is especially true for sensitive, wide-ranging, and elusive species. Several empirical methods have been used to study dispersal, including CMR (capture–mark–recapture), radio-tracking, and genetics (Nathan et al. 2003; Broquet and Petit 2009; Baguette et al. 2012). However, each method has its limitations. For example, CMR methods risk missing long-distance dispersers due to a limited sampling scope (Koenig et al. 1996) and typically require direct handling of individuals possibly affecting their behavior and even survival (Kock et al. 1987). Radio-tracking captures long-distance dispersers that other methods miss (Koenig et al. 1996) and reveals fine-scale details of movement pathways and timing of departure and arrival. However, it requires expensive and highly specialized equipment as well as the need to capture and handle individuals, making it difficult to generate a large enough sample.

Genetic methods have the advantage that samples can be obtained noninvasively (Lawson Handley and Perrin 2007) and contain information that projects beyond the sampled individual (e.g., kinship). But there are many practical issues with genetic methods including, but not limited to, sampling a large enough proportion of the population, obtaining high-quality DNA from noninvasively collected samples, unknown age of individuals, unknown directionality of PO (parent–offspring) relations, assessing whether dispersal has occurred at the time of sampling, and establishing accurate pre- and post-
dispersal locations. Moreover, many genetic dispersal models have been developed in a population genetics framework (Wright 1943; Waser and Strobeck 1998; Gandon and Rousset 1999; Rousset 2001), where stringent assumptions of ideal populations and results that reflect historic population averages severely limit the usefulness of such models for contemporary processes (Sugg et al. 1996; Palßbøll et al. 2013). A lack of genetic resolution has largely prevented alternative approaches for all but the most intensively studied populations, where a combination of observational and genetic inferences has allowed for the reconstruction of accurate pedigrees (e.g., Pemberton 2008; Spong et al. 2008). But if some or all of these issues could be resolved, genetic techniques can be quite effective for measuring dispersal (Nathan 2001; Baguette et al. 2012).

SNPs (Single nucleotide polymorphisms) are suitable for many types of studies as they offer high genomic resolution, reproducibility across laboratories, ease of allelic assignment, and, relative to microsatellites, a reduction in erroneous results due to mistyping and allelic dropout (Anderson and Garza 2006). However, SNPs have only recently been added to the molecular toolbox due to the recent and rapid advancement of sequencing technology. As any one SNP has low statistical power compared to a multiallelic microsatellite, many more SNPs are necessary, but with today’s technology, finding many genomewide SNPs is no harder than finding a few. Choice of molecular marker should be weighed according to the biological question being asked as they afford different properties. SNPs are useful for identifying individuals and inferring relatedness given the right characteristics (Glover et al. 2010). For example, SNPs that have high minor allele frequencies, where both alleles are common within the population of interest and which are unlinked to all other SNPs, tend to be most informative for individual identification and relatedness inference (Anderson and Garza 2006).

One approach to estimating individual-based dispersal distances using molecular markers is through inference of relatedness between individuals (e.g., Spong and Creel 2001; Rollins et al. 2012), in particular mother–offspring pairs. With knowledge of individual locations, measuring the geographic distance between mother and offspring will give an estimate of dispersal distance. However, identifying mother–offspring pairs with molecular markers alone is not sufficient due to the uncertainty of directionality (i.e., it is not directly apparent which individual is the parent and which is the offspring when the relationship is assessed using molecular markers and in the absence of demographic data). As this is an essential component for estimating natal dispersal distance, at least for triads that include a male, one must take it a step further. One way to resolve this is to attempt to reconstruct the pedigree and thus reveal the directionality of the relationship. To do this, it is critical to obtain enough samples as the higher the proportion of individuals sampled, the more complete the pedigree will be (Pemberton 2008). With a high sampling resolution, the possibility of identifying PO triads (i.e., both parents and offspring) becomes greater. These triads provide higher confidence in determining directionality through allele sharing alone as offspring will share at least one allele that is identical by descent with both mother and father at every locus.

The requirement for a large proportion of samples from the population of interest can make sampling effort both time- and cost-intensive, not to mention logistically challenging. However, with a combination of noninvasive sampling and citizen participation, it is possible to achieve a high sampling resolution in a timely, cost-effective way that can be made logistically feasible. Noninvasive sampling is concentrated around locating sources of DNA, which can be found in feces, fur, feathers, saliva, and urine among others eliminating the need to interact with the study subjects (Taberlet and Luikart 1999; Taberlet et al. 1999; Waits and Paetkau 2005). Engaging local citizens who are willing to volunteer to collect samples can be advantageous as it considerably reduces costs and collection time when there are many participants. (Bonney et al. 2009; Devictor et al. 2010; Dickinson et al. 2010). An added benefit is that citizens can be knowledgeable about locating and identifying samples, thereby enhancing collection success. For over a decade, Sweden has successfully engaged citizen volunteers to help collect samples on multiple occasions from feces left by the brown bear (Ursus arctos; Fig. 1) (see Bellenbaum et al. 2005). Resampling has enabled monitoring of the same population over time and has revealed population growth and declines in certain counties within Sweden (Kindberg et al. 2011). Moreover, the data generated from these collections are useful for many other applications such as identifying and tracking individuals (Kindberg et al. 2011), assessing gene flow patterns, and detecting population substructuring (Schregel et al. 2012; Koptz et al. 2014).

Here we use a recently developed SNP panel containing 96 SNPs derived from the Scandinavian brown bear (see Norman et al. 2013) to estimate dispersal distance in the Swedish south-central population of brown bear. The SNP panel was developed for inferring relatedness between individuals, making it suitable for estimating individual-based (direct) dispersal. This study uses SNP genotyping on noninvasive samples collected by citizens to estimate dispersal distances through pedigree reconstruction.
Materials and Methods

Sample collection

Samples of brown bear feces were collected in a twelve-week period between August and October 2012 in the counties of Dalarna and Gävleborg, which consists of the majority of the south-central Swedish population representing the western European lineage (Taberlet and Bouvet 1994). Volunteers, mainly moose hunters, opportunistically collected feces and sent the samples along with the coordinates of the sample location to the county administration board (Länstyrelsen, Sweden). This sample collection was performed following the same protocol described in Bellemain et al. (2005) and Kindberg et al. (2011).

DNA extraction and SNP genotyping

Samples were sent to Bioforsk, Norway, for DNA extraction. Details of the sample storage and DNA extraction procedure can be found in Schregel et al. (2012). Once unique individuals were identified, one aliquot per individual was sent to our laboratory in Umeå, Sweden, for SNP genotyping.

Single nucleotide polymorphism genotyping was performed on the Fluidigm Biomark using the SNP panel as described in Norman et al. (2013) with a slight alteration: Two of the SNPs that were found to be linked (snp163 and snp171 from Norman et al. 2013) were removed and replaced with two Y-chromosome SNPs (Bidon et al. 2015). We manually screened the genotype clusters by the Biomark software and removed any loci with ambiguous cluster affiliation from further analyses. Negative controls (i.e., water in place of DNA) were included in each run. Samples that were close to the negative control were deemed “No Calls”. Duplicates (n = 91) and triplicates (n = 10) of samples were included for the estimation of genotyping error. Allelic dropout was calculated from heterozygote loci as recommended by Broquet and Petit (2004).

Sex of each sample was determined through both the Y-chromosome and X-chromosome markers. If the sample appeared in the cluster for each Y-chromosome marker, it was recorded as a male. If the sample was a “No Call”, it was considered to be a female. Any sample outside the cluster, but not at the origin (i.e., where the negative controls are located), was invalidated. The Y-chromosome determination of sex was then validated through three X-chromosome SNPs by ensuring that any male had only one allele at each X-chromosome marker, hence appearing as a homozygote, and a female was confirmed if it had at least one heterozygote genotype on the X-chromosome. Likewise, mitochondrial haplotype for each sample was determined by allelic state for each of the four diagnostic mitochondrial markers.

Sample locations

Our first step in determining natal dispersal distance was to estimate home range centers for each individual using fecal sample locations. As many individuals’ home ranges overlap, using the center-to-center distances will provide an estimate of even short-distance dispersers. As our sample locations are based on fecal sites, we rely on the assumption that the fecal sites are within the home range. A previous study by Bellemain et al. (2005) within the same area showed that 80% of the fecal sites were found within the home range (estimated as 95% MCP) and those that were outside the home range were within 10 km of the home range. These results suggest that the fecal sites are most likely to be representative of the home range and those that are not are likely to be close by, thereby keeping the margin of error low. For individuals with multiple samples, the estimate of home range centers should be more accurate than for those with just one. For this study, we have calculated the center points as the median center using the R package “aspace” version 3.2 (Bui et al. 2012) for individuals with two or more sample locations (n = 138). Those with one location were maintained as is (n = 275). The median center was chosen due to its insensitivity to outliers, which can be indicative of an individual leaving his/her home range temporarily.

Pedigree reconstruction

Pedigrees were reconstructed using FRANz software version 1.9.999 (Riester et al. 2009) with maximum number
filters were used in subsequent analyses.

Estimation of relatedness

Lynch–Ritland relatedness coefficients (Lynch and Ritland 1999) were calculated for each pair to further assess relatedness between individuals using the R package “related” version 0.8 (Pew et al. 2014). The Lynch–Ritland relatedness coefficient was chosen as it has been shown to perform better than other relatedness estimators (Thomas 2005; Cailley et al. 2006). The reconstructed pedigrees were screened for relatedness categories as follows: PO, FS (full siblings), HS (half siblings), GG (grandparent–grandchild), and mates and plotted against the coefficient of relatedness using R (R Development Core 2013).

Isolation by distance

To determine whether the sample scope would be large enough to capture the majority of dispersal events, we tested for isolation by distance (IBD). Pairwise Euclidean distances were calculated with the median centers for all pairs of sampled individuals using Pythagorean theorem. To detect IBD, a Mantel test was run for only those pairs of putative parents in the pedigree reconstruction results. Euclidean distance and Lynch–Ritland relatedness coefficient matrices were input into man tel.randtest in the R package “adegenet” version 1.4-2 (Jombart and Ahmed 2011). Three categories were computed: (1) all pairwise putative parents; (2) female–female pairs only; and (3) male–male pairs only. Additionally, a Pearson’s product-moment correlation was calculated for these three categories as well as for pairs of the opposite sex.

Estimation of natal dispersal distance

The pedigrees were assessed to detect possible cubs based on three factors: (1) each of the cubs has a full sibling; (2) the full siblings were in the same geographic location; and (3) this geographic location was within 1 km of their mother. Individuals identified as cubs were subsequently removed from natal dispersal distance analysis as they have not yet dispersed. Natal dispersal distances were calculated for all remaining offspring with a known mother as identified in the reconstructed pedigrees. Finally, a Wilcoxon rank-sum test was applied to female and male dispersal distances to determine whether there was a significant difference using R (R Development Core 2013).

Biases

Spatial and logistical limitations may cause biases. As we are using noninvasively collected samples from a portion of the population that is continuous beyond the area sampled, we will inevitably miss some dispersal events and particularly long-distance events. As brown bear exhibits male-biased dispersal (Swenson et al. 1998; McLellan and Hovey 2001; Proctor et al. 2004; Staen et al. 2006), missing these long-distance dispersal events will underestimate distances for males in particular. Likewise, through noninvasive sampling alone, there is no current method to determine the age of individuals. As juveniles disperse between the ages of 2–5 years (Staen et al. 2006), there will likely be individuals accounted for that have not yet dispersed, leading to a possible underestimation of distances. Finally, deviations from true home range centers may lead to slight under- or overestimations.

Results

SNP genotyping

We successfully genotyped 433 individuals from the 434 uniquely identified individual DNA extracts we received at 96 SNP loci. One was unsuccessful due to probable contamination and was therefore removed from all further analyses. Within all heterozygote SNP loci \( n = 7825 \) excluding the haploid SNPs (Y-chromosome and mtDNA), we identified three probable genotyping errors resulting in an error rate of 0.00038. There were 134 (0.36%) autosomal genotypes that were invalidated due to inability to resolve which cluster it belonged to. Thus, the call rate for all SNPs excluding the Y-chromosome was 0.9965. Mean minor allele frequency for autosomal SNPs was 0.37. We identified 243 females and 190 males through the Y-chromosome and X-chromosome markers. All of the individuals shared the same mitochondrial haplotype that is representative of the southern Swedish population (see Norman et al. 2013) with the exception of seven males, six of which have the haplotype common to the middle population and one to the northern population indicating possible long-distance dispersal.
Sample locations

Of the 433 genotyped individuals, we had coordinate data for 412 individuals. Mean and maximum number of samples collected per individual were 2.20 (SD: 2.59) and 19, respectively. Overall mean distance between sample sites of the same individual was 15.6 km (SD: 11.4). Table 1 shows the frequency distribution of number of samples per individual. A map showing the median centers for individuals with multiple samples and single point locations for those individuals with one sample is shown in Figure 2.

<table>
<thead>
<tr>
<th>Number of samples</th>
<th>Number of individuals</th>
<th>Percentage of all individuals</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>275</td>
<td>66.4</td>
</tr>
<tr>
<td>2</td>
<td>45</td>
<td>10.9</td>
</tr>
<tr>
<td>3</td>
<td>25</td>
<td>6.0</td>
</tr>
<tr>
<td>4</td>
<td>24</td>
<td>5.8</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>2.4</td>
</tr>
<tr>
<td>6-10</td>
<td>25</td>
<td>6.0</td>
</tr>
<tr>
<td>11-15</td>
<td>7</td>
<td>1.7</td>
</tr>
<tr>
<td>&gt;=16</td>
<td>3</td>
<td>0.7</td>
</tr>
</tbody>
</table>

Pedigree reconstruction

Of the 433 individuals, FRANz identified two parents for 65 individuals, one parent for 172 individuals, and no parents for 196 individuals. From those with at least one parent identified, the posterior probability was greater than or equal to 0.95 for 82 individuals: 60 triads (both parents identified) and 22 dyads (one parent identified). The total number of unique individuals comprising these triads and dyads is 149. In total, these triads and dyads make up 28 disjoint pedigrees ranging in size from 2 to 13 individuals (mean: 5.36; SD: 3.54) and spanning two to three generations.

Estimation of relatedness

The Lynch–Ritland relatedness coefficient \((r)\) was calculated for all pairs within the sampled individuals \((N = 93,528\) pairwise comparisons). The results were then subset for all pairs of individuals contained within the pedigrees \((N = 11,027)\). Mean relatedness was \(-0.0023\) (SD: 0.1270) and \(-0.0003\) (SD: 0.1424) within all sampled individuals and the pedigreed individuals, respectively. Figure 3 shows the categorical relationships (PO, FS, HS, GG, and MT) and their associated \(r\)-estimates of the pedigreed pairs \((N = 132)\). These results fall into the scope of what can be expected for each relatedness category, indicating that the pedigrees and \(r\)-estimates are in agreement with one another. There are two outliers, one in a grandparent-grandchild (GG) pair and one in a MT (mated pair). Both appear at the upper end of the \(r\)-scale, which can be indicative of pairs with unusually high levels of inbreeding (in the GG pair) and mates who are closely related (in the MT pair). Both outliers are therefore retained in subsequent analyses.

Isolation by distance

Euclidean distance between all pairs of sampled individuals based on the median centers for those with multiple locations resulted in a mean of 100.6 km and SD of 53.3. Isolation by distance was significant for all putative parent pairs \((N = 9870)\; Mantel correlation: \(-0.11; P\)-value < 0.001) and female–female pairs \((N = 3655)\; Mantel correlation: \(-0.18; P\)-value < 0.001) and nonsignificant for male–male pairs \((N = 1485)\; Mantel correlation: \(-0.042; P\)-value = 0.080) (Fig. 4). Additionally, the Pearson’s correlation test (a statistic that is comparable to the Mantel test with the type of data used in this study) was applied to all categories above and additionally to pairs of the opposite sex for which a Mantel test could not be applied due to its asymmetrical nature. Pearson’s correlation for all categories is as follows: all: \(-0.11 (N = 9870)\);
Natal dispersal

From the pedigree analysis, of the 82 offspring with at least one parent identified, 71 included the mother. Of these 71 offspring, eight were identified as cubs (see Materials and Methods for identification technique) and subsequently removed from the natal dispersal distance analyses, leaving 63 mother–offspring pairs.

Natal dispersal distances ranged from 0 to 53 km (mean: 12.9; SD: 11.7 km) for females and 1 to 103 km (mean: 33.8; SD: 33.9 km) for males (Table 2; Fig. 5). A Wilcoxon rank-sum test (Mann–Whitney test) indicates a significant difference between female and male dispersal distances with a 0.05 significance level (Wilcoxon’s rank-sum test; $W = 3930$; $P$-value $= 0.02$).

Discussion

In this study, we estimated natal dispersal distances for brown bear using noninvasively collected samples and a set of 96 SNPs. We determined whether our sample scope would be large enough to capture the majority of dispersal events through a test for IBD. A significant result for females suggests that this is the case, whereas a non-significant result for males suggests that we are missing some of the long-distance dispersal events for males. Indeed, for a comparison with previous estimates from the same population where radio-collars were used, Steen et al. (2006) report similar female distances as our study, but longer male distances. Thus, while our estimates for females are likely to be representative of the true distances, the estimates for males are missing long distances. However, while male estimates are biased toward the shorter distances, we nevertheless detect a significant difference between female and male dispersal estimates with male dispersing further.

Steen et al. (2006) limited distances to those beyond the mother’s home range and, in some cases, to only those that were beyond the mean distance possibly leading to an upward bias. Contrarily, we opted to include all ranges of distances only excluding individuals that, based on their pedigrees, are highly likely to still be in the care of their mother. We chose to include short-distance dispersers as it can reveal population features that would otherwise be missed including kin and nonkin interactions as well as fine-scale details of philopatry such as sex ratio and variations in distances from the natal area. Natal dispersal is defined as the movement of progeny from the birthplace (the natal area) to the area where it reproduces (the breeding area) for various taxa (see Greenwood 1980; Broquet and Petit 2009; Matthysen 2012). For many small mammals in particular, the distance between natal and breeding areas can be measured as the distance between the population where the individual was born and the population where the individual reproduces (e.g., Centeno-Cuadros et al. 2011; Dey et al. 2013). Where dispersal is measured between discrete populations, rates and distances, once detected, are relatively easy to quantify. Contrarily, dispersal events for large mammals are often considered at the population scale where individuals disperse within a population as well as to neighboring populations as with the brown bear. At this scale, unless an individual remains in the direct vicinity of its mother, dispersal rates can be difficult to ascertain as it begs the question: What is a disperser and how is it distinguished from a nondisperser? Sometimes, arbitrary distance thresholds based on life-history parameters are used to make this distinction (Broquet and Petit 2009). However, given the definition, an individual can have dispersed very short distances if it has reproduced and is largely independent of its mother. Very short-distance dispersers can have a considerable effect on conservation issues such as inbreeding and population genetic structure (Greenwood 1980; Eisehardt et al. 2013). We therefore opted to include all natal dispersal distances to
appropriately describe dispersal patterns. It is worth noting that this does not preclude individuals from being considered philopatric.

The use of noninvasively collected samples enabled us to obtain information about the population without disturbing or interacting with the individuals in the study. While there are some limitations to using noninvasively collected samples such as a lack of demographic information and a limited sampling scope, the advantages make it worthwhile in comparison with other methods. Studying dispersal in large carnivores such as the brown bear is difficult as the animals are elusive, highly mobile, and potentially dangerous to researchers. Not only that, but they are sensitive to the mere presence of humans. A study by Ordiz et al. (2013) showed that just the scent of a human nearby affected the behavior of the brown bear for up to 2 days afterward. Other methods, such as tracking radio-collared individuals, require individuals to be captured through sedation. Capturing individuals is in itself challenging as it is expensive often requiring the use

Table 2. Dispersal distance estimates showing the N (number of individuals), the median, mean, SE (standard error), and maximum distance for all individuals, females only, and males only. Results from a previous study by Støen et al. (2006) showing mean and SE of brown bear dispersal distances estimated from the same population as our study, but having used different methods, are shown in the final column.

<table>
<thead>
<tr>
<th>N</th>
<th>Median</th>
<th>Mean ± SE</th>
<th>Max</th>
<th>Previous estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>All offspring</td>
<td>63</td>
<td>11</td>
<td>21.2 ± 3.2</td>
<td>103</td>
</tr>
<tr>
<td>Female offspring</td>
<td>38</td>
<td>9</td>
<td>12.9 ± 1.9</td>
<td>53</td>
</tr>
<tr>
<td>Male offspring</td>
<td>25</td>
<td>14</td>
<td>33.8 ± 6.8</td>
<td>103</td>
</tr>
</tbody>
</table>

*From Støen et al. (2006).
of a helicopter, ethical permits, and the presence of a veterinarian. However, the main concern is the negative consequences on the individuals captured with the worst case scenario being death (Arnemo et al. 2006). In comparison, the use of noninvasively collected samples is ideal. This is true for other large carnivores, but also for many species, large and small, which are sensitive to capture and handling or difficult to detect.

Since the advent of high-throughput sequencing, the use of SNPs in studies of wild populations has been on the rise. This study further exemplifies the advantages of SNPs over other molecular markers. For high confidence, a high sampling resolution combined with a highly informative panel of molecular markers with low error rate is recommended (Pemberton 2008). In this study, more than 50% of the population was sampled leading to a high chance of finding enough individuals within a pedigree to obtain pedigree links and to detect triads, thereby resolving the issue with directionality. Additionally, the panel of SNPs was designed to be most informative for inferring relatedness within the population under study. With a mean minor allele frequency $>0.37$, the cumulative power of the SNPs to distinguish between individuals is high with a probability of identity below $6 \times 10^{-24}$. Furthermore, with one genotyping error for every 2600 loci, the chance of a false-positive relationship appearing is minimal.

**Conclusion**

In this study, we have shown that it is possible to estimate natal dispersal distance in a wild population without any interaction with the individuals included in the study or any behavioral or life-history data. Despite a potential bias toward short-distance dispersers, particularly for males, the large sampling scope enabled us to detect significant male-biased dispersal and IBD in females. Two key factors contributed to this achievement. One is the high sampling resolution made possible by citizen science. It would have otherwise been challenging to obtain such a high sampling resolution in the short amount of time required. Additionally, as the citizens volunteered their time, the cost was kept low. The second key factor is that we used a highly informative SNP panel that was carefully designed for inferring relatedness in this particular population. As public databases are rapidly acquiring genetic data for wild species, the cost and time required to develop a SNP panel in other wild species will be less of a hindrance than it has been in the recent past. In addition, the bioinformatics involved in developing a SNP panel is less cumbersome than for many other applications, such as whole-genome sequencing, yet the value of it for a species of conservation concern is great.

**Acknowledgments**

We are grateful to all volunteers who collected samples used in this study and to Lånestyrelsen for organizing the collection. We thank Bioforsk, Norway, and Michael Schneider at the Västerbotten county administration board for providing us with DNA extracts and data associated with the samples. We thank our laboratory technician, Helena Königsson, who performed the genotyping. We thank Astrid Vik Strønen and Andrew Allen for their helpful comments on the manuscript. We thank the Scandinavian Brown Bear Research Project for their support (Jonas Kindberg and Jon Swenson). We thank Jarrod Hadfield and three anonymous reviewers for their comments on a previous version of this manuscript which resulted in many improvements.
Conflict of Interest
None declared.

References


Landscape relatedness: detecting contemporary fine-scale spatial structure in wild populations

Anita J. Norman¹, Astrid V. Stronen², Geir-Arne Fuglstad³, Aritz Ruiz-Gonzalez⁴,⁵, Jonas Kindberg¹, Nathaniel R. Street⁶, Göran Spong⁷

¹Department of Fish, Wildlife and Environmental Studies, Swedish University of Agricultural Sciences, Umeå, Sweden
²Section of Biology and Environmental Science, Department of Chemistry and Bioscience, Aalborg University, Aalborg, Denmark
³Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim, Norway
⁴Department of Zoology and Animal Cell Biology, University of the Basque Country, UPV/EHU, Vitoria-Gasteiz, Spain
⁵Systematics, Biogeography and Population Dynamics Research Group, Lascaray Research Center, University of the Basque Country, UPV/EHU, Vitoria-Gasteiz, Spain
⁶Umeå Plant Science Centre, Department of Plant Physiology, Umeå University, Umeå, Sweden
⁷Fisheries, Wildlife and Conservation Biology Program, Department of Forestry and Environmental Resources, North Carolina State University, Raleigh, NC, U.S.A.

Corresponding Author:

Anita J Norman
Email: anita.norman@slu.se
Phone: +46 76 148 7131
Fax: +46 90 786 8162
Abstract

Context. Methods for detecting contemporary, fine-scale population genetic structure in continuous populations are scarce, yet they are vital for ecological and conservation studies. It is particularly important under a changing landscape.

Objectives. Here we present a novel, spatially explicit method that we call landscape relatedness (LandRel). With this method, we aim to detect fine-scale population structure that exists at present and that is sensitive to temporal changes in the landscape.

Methods. We interpolate spatially determined relatedness values based on SNP genotypes across the landscape. Interpolations are calculated using the Bayesian inference approach Integrated nested Laplace approximation (INLA). We empirically tested this method on a continuous population of brown bears spanning two counties in Sweden.

Results. Two areas were identified as differentiated from the remaining population. Further analysis suggests that inbreeding has occurred in at least one of these areas.

Conclusions. LandRel enabled us to identify fine-scale structuring in the population that was not previously known. These results will help direct future research efforts, conservation action and aid in the management of the Swedish brown bear. LandRel thus offers an approach for detecting population structure with a focus on contemporary, fine-scale analysis of continuous populations.

Keywords: kinship-based methods; continuous population; non-invasive genetic sampling; SNPs; INLA; Ursus arctos; brown bear; population genetic structure; inbreeding; Sweden
Introduction
Knowledge of contemporary spatial structuring of populations is an important basis for ecological studies and it facilitates conservation of a species (Bossart and Prowell 1998; Palsbøll 1999). With the recent advent of high-throughput technology, studies of spatial structuring have been increasingly geared towards genetic structure. Population genetic structure can be studied either using indirect methods such as Wright’s $F_{ST}$ (Wright 1950) or with direct methods which focus on migrants (Broquet et al. 2009; Palsbøll et al. 2010). Indirect methods have been used for many decades and are still common today. However, they suffer from some critical issues that have been widely addressed in the literature, perhaps foremost the assumption that populations are in equilibrium (see Bossart & Prowell 1998; Beerli & Felsenstein 1999; Whitlock & McCauley 1999; Peery et al. 2008; Broquet et al. 2009; Lowe & Allendorf 2010).

Alternatively, direct methods are largely based on assignment tests where individuals are ‘assigned’ to a subpopulation that is most fitting to their genotype (Manel et al. 2005). Kinship-based methods are a type of assignment test that identifies and locates highly related individuals as inferred through molecular markers (Broquet et al. 2009; Palsbøll et al. 2010). Other direct methods include multidimensional scaling (e.g., PCA) or spatial autocorrelation and are also based on allele frequencies. These methods feature advantages over indirect methods, but each has its own drawbacks. One common limitation is that many direct methods are modeled after the island population model (Latter 1973) and thus designed only for discrete populations. Additionally, many of these methods readily identify subpopulations that are highly differentiated (i.e., with little gene flow between them), but often fail to detect structure in subpopulations with high connectivity (Manel et al. 2005; Saenz-Agudelo et al. 2009; Lowe and Allendorf 2010). Here, we present a new method based on relatedness levels across the landscape. This method is designed to detect contemporary, fine-scale spatial structure in a continuous population thereby offering a complementary method to the existing set of direct methods.

Limitations of Assignment-Based Methods
While assignment-based methods often reveal structuring in a population, assumptions such as random sampling, Hardy-Weinberg equilibrium, and marker neutrality are common (Saenz-Agudelo et al. 2009). Violations of the assumption of marker neutrality is confounded by the variations in selective pressures on different loci, thus influencing estimations of population differentiation (Whitlock and McCauley 1999; Waples and Gaggiotti 2006). Even loci assumed neutral can be subjected to the same selective pressures through linkage. Moreover, ascertainment of genetic markers, their representativeness of the whole genome, and total number of loci can have effects on resulting estimates (e.g., Helyar et al. 2010). Markers ascertained either outside the population of interest or using too few individuals to properly assess population-level allele frequencies can bias estimates of structure. Likewise, marker characteristics such as high polymorphism may mask true underlying structure (Putman and Carbone 2014). For one of the most commonly used methods, cluster-based modeling, Putman & Carbone (2014) recommend a minimum of 50 loci for reliable results. However, if the aim is to detect fine-scale structure, many more than 50 would be required (Peery et al. 2008), thus limiting studies to
broad-scale structure. Furthermore, assignment methods often require a priori knowledge of source populations (Waples and Gaggiotti 2006). Finally, two confounding factors in many assignment-based methods is the presence of related individuals (Manel et al. 2005; Putman and Carbone 2014) and isolation by distance (IBD) (Blair et al. 2012; Meirmans 2012; Cushman et al. 2014), which can result in family structure or allelic clines being confused with population structure.

**Kinship-Based Methods**

Several direct methods exist for estimating structure by using kinship to estimate migration rates (Palsbøll 1999; Peery et al. 2008; Saenz-Agudelo et al. 2009). Parentage methods provide accurate estimates of connectivity between two subpopulations when connectivity between subpopulations is high, which is the opposite of other assignment-based methods, making the two approaches complementary (Saenz-Agudelo et al. 2009). However, sufficient sampling coverage of candidate parents can be challenging. Other kinship methods include capture-mark-recapture to detect migration rates (Vitalis 2002), mixture analysis, which extends the capture-mark-recapture method to include hierarchical structure (Fontanillas et al. 2004), or a combination of these two methods (Broquet et al. 2009). Dharmarajan et al. (2014) used pairwise comparisons of the coefficient of relatedness \( r_{XY} \) and conducted a permutation test to identify patches with significantly high kin structure. Iacchei et al. (2013) performed a similar test, but compared the differences of within-population kinship to between-population kinship based on sampling sites. Økland et al. (2010) derived a clustering algorithm using genetic relatedness.

In contrast to other assignment methods, kinship-based methods are useful in detecting fine-scale structuring (Saenz-Agudelo et al. 2009; Palsbøll et al. 2010). Another key advantage is that they estimate contemporary population structure based on the existing generations at the time of sampling (Palsbøll et al. 2010), whereas non-kinship assignment methods based on genetic divergence only results in estimates of structure that has accumulated over time. These estimates are therefore averages of processes occurring in the past tens to thousands of generations. Thus, kinship-based methods provide reliable estimates of contemporary structure in discrete populations.

**The Landscape Relatedness (LandRel) Method**

Our landscape relatedness (LandRel) method is based on pairwise relatedness values interpolated over the landscape. This method is designed for detecting fine-scale structure in continuous populations, but can theoretically be used for discrete populations. Because it uses continuous relatedness values, there is no distinction of kin categories (e.g. parent-offspring, full-siblings, etc.) thereby differentiating it from kinship-based methods. LandRel relies on the assumption that within a perfectly homogenous landscape in which a population is randomly mating, the mean relatedness throughout the landscape will be uniform. Any divergence from these two criteria will result in an heterogenous distribution of relatedness. Consider, for example, that a fenced highway was built splitting an otherwise panmictic population in two. Individuals born near the road will no longer disperse in random directions, but will disperse alongside or away from the road on both sides. This will lead to a sorting of individuals regarding relatedness. LandRel seeks to detect divergence from landscape homogeneity.
and random mating behavior by testing for uniform distribution of relatedness throughout the landscape.

LandRel shares many of the advantages of kinship-based methods such as the ability to detect fine-scale structuring. However, unlike kinship-based methods, LandRel can be used for continuous populations and does not require a priori knowledge of source populations. Furthermore, since it is based on relatedness, there is no need to remove highly related individuals, which is necessary for some kinship-based methods (e.g. Saenz-Agudelo et al. 2009) and most other assignment methods. Additionally, kinship-based methods, particularly parentage methods, rely heavily on high quality genotyping to infer kinship. LandRel does not require knowledge of kinship making it more tolerant of genotyping errors. Furthermore, LandRel is spatially explicit thereby elucidating where in the landscape structure exists.

**Study System**

The Scandinavian brown bear (*Ursus arctos*) consists of three genetically distinct subpopulations with some connectivity between them (Manel et al. 2004; Norman et al. 2013). The two northernmost subpopulations originate from the eastern European lineage while the southernmost subpopulation is a relic of the western European lineage (Taberlet and Bouvet 1994). The Scandinavian population size has been steadily increasing (Kindberg et al. 2011) since a hunting-induced bottleneck reduced the size to approximately 130 individuals (Swenson et al. 1995). However, the latest population estimate revealed a steep decline in just a few years (Kindberg and Swenson 2014) prompting Artdatabanken (Uppsala, Sweden) to change the national status to Near Threatened. Furthermore, based on samples from up to 29 years ago, Tallmon et al. (2004) found that the southernmost population had a low effective population size and low immigration and recommended that this population be monitored for signs of inbreeding.

This study aims to empirically test LandRel on the southernmost subpopulation of the Scandinavian brown bear. This population is currently estimated at approximately 791 individuals (Kindberg and Swenson 2014), which has not changed significantly since the previous census (Kindberg et al. 2009) and is subjected to quota-based hunting. To our knowledge, no structuring within this subpopulation has previously been identified and therefore we assume that it is a panmictic population. However, since the brown bear exhibits male-biased dispersal and female philopatry (Støen et al. 2006; Zedrosser et al. 2007; Norman and Spong 2015), we expect some structure among females. Additionally, we assume the landscape to be homogenous for the purposes of testing the method. Our expectations are thus that males will show uniform levels of relatedness throughout the landscape while females will have areas with significantly high relatedness akin to core areas.
Methods

Sample Collection and Genetic Processing
Samples were collected in the autumn of 2012 in the neighbouring counties of Dalarna and Gävleborg in central Sweden (Figure 1). For the purposes of conducting a census, the Swedish County Administration Board (Länstyrelsen) organized citizen volunteers to collect faeces opportunistically, place them in vials and send them with the GPS coordinates to Bioforsk (now NIBIO) (Norway) for DNA extraction. Further details of the sampling procedure can be found in Kindberg et al. (2011) and the DNA extraction protocol and individual identification in Schregel et al. (2012). DNA extracts from uniquely identified individuals were then sent to our laboratory.

DNA extracts from all sampled individuals were SNP genotyped at 96 loci as described in Norman & Spong (2015). The panel of SNPs included four Y-chromosome markers, three X-chromosome SNPs, four diagnostic mitochondrial markers and 85 autosomal SNPs. SNPs were ascertainment de novo in brown bears across the geographic range in Sweden and included the population being analysed in this study (Norman et al. 2013). SNPs were selected to be highly discriminatory with characteristics such as high minor allele frequency and low levels of linkage between SNPs (Norman et al. 2013).

Individuals with multiple sample locations were analysed using the median-centre of all locations to estimate the most probable home range core. All other individuals were analysed using their single GPS point. Further details of the median-centre analysis can be found in Norman & Spong (2015).

Population Structure with PCA
We first used our genotype data to perform a principal components analysis (PCA) to discern how a commonly used and statistically rigourous (Patterson et al. 2006) method for analyzing population structure performed with our data. We used the “adegenet” package version 1.4-2 (Jombart et al. 2013) implemented in R (R Development Core 2013) and chose a two-dimensional analysis. Here a matrix of genotype distances is created and subsequently analysed for genetic differentiation based on eigenvalues. If there is clear population structure, points separate into distinguishable clouds. To help determine if there are true clusters, we tested the resulting values in principal component 1 to identify if it was unimodal (i.e. normally distributed), suggesting one cluster, or multimodal, suggesting two or more clusters using a quantile-to-quantile plot (qqplot) implemented in R (R Development Core 2013).

Relatedness Estimates
Estimates of relatedness between pairs of individuals were derived using Lynch-Ritland coefficient of relatedness (Lynch and Ritland 1999) with the R package “related” version 0.8 (R Development Core 2013; Pew et al. 2014). Relatedness values (r-values) reflect the proportion of genome that is identical by descent between two individuals. First order relatives (i.e. parent-offspring and full siblings) share approximately 50% of their genome and have an r-value around 0.50. Second-order relatives (i.e. half-siblings and grandparent-grandoffspring) have an average r-value around 0.25 and those that are unrelated have an r-value around 0.00. If a pair is more unrelated than expected under panmictic
Interpolations

Interpolations of relatedness across the study area were conducted using Integrated Nested Laplace Approximations (Rue et al. 2009) with the package INLA (version 0.014) implemented in R (R Development Core 2013). INLA is an approximate Bayesian inference designed for structured latent Gaussian models, a type of additive regression model. INLA uses direct numerical integration to approximate marginal posterior densities as does the Markov Chain Monte Carlo method, but with much less computational time (Holand et al. 2013).

Procedure

Interpolations for mean relatedness were calculated based on one focal individual at a time. The geographic location of each non-focal individual was represented by the pairwise r-value with the focal individual, whereas the focal individual was not represented by any value. Interpolations were repeated in the same manner for each individual (N=412) so that every individual was included in an interpolation N-1 times. The results of all interpolations were overlaid together with the overall result being the sum of the individual-based mean relatedness across the landscape. To determine if there existed any sex-specific patterns of relatedness, the same process was repeated using only males or only females as focal individuals. All individuals regardless of sex were still included as non-focal individuals.

Statistical significance was calculated using overall interpolated values at each grid point by dividing mean relatedness with root-mean-square of the standard deviation. We used an alpha level of 0.05 and, being a two-tailed test, normalized values greater than 0.975 or less than 0.025 were considered significant.

Areas of significance were in turn analysed following the same procedure as for the entire study area (global area), but only including the geographic area that contained areas of significance (local area). Here, only individuals located within defined boundaries were included in the analysis; however, the same pairwise r-values based on the allele frequencies of the entire sample set were used. As with the global area, for each local area we first analyzed all individuals and subsequently partitioned males and females. In addition to focusing on areas deemed as significant, we included an area with almost no global significance as a control.

INLA

Prior to running the interpolations, relatedness coefficients were first checked to ensure a Gaussian distribution. Interpolations were performed following the guidelines of Lindgren & Rue (2015) using the spatial SPDE-model (Lindgren et al. 2011). A two-dimensional mesh was created for each area being analysed. INLA was run with family set to Gaussian and using the following model:

\[ y \sim N(1 + \text{Intercept} + f(\text{field}, \text{model=}\text{spde})) \]
where \( y \) is the pairwise \( r \)-value. We ran other models with different families and including Euclidean distance as a coefficient and found that the model above performed best as it resulted in the lowest deviance information criterion (DIC).

Mean relatedness and standard deviations were interpolated at every grid point on grids of 150 x 150 for the global area and 100 x 100 for the local areas. Maps were created using the R package lattice v. 0.20-30 (Sarkar 2008).

**Inbreeding**

Areas that showed statistically significant relatedness patterns in the interpolations (i.e. the local areas) were further investigated for evidence of inbreeding. For each area, we extracted all pairwise relatedness values equal to or exceeding 0.40, to capture all first-order relatives (i.e. parent-offspring and full-siblings). We then ran a Wilcoxon rank sum test between each pairwise area to detect if there was a statistical significance between areas based on an alpha level of 0.05.

**Results**

**Sample Collection and Genotyping**

We obtained DNA extracts and GPS coordinates for 412 individuals: 180 males and 232 females (Figure 2). All individuals were genotyped successfully at all 96 SNP loci. Mean minor allele frequency of autosomal loci was 0.37 (Range: 0.13-0.50). The call rate excluding Y-chromosome loci was 0.997 and error rate based on heterozygous loci was 3.8x10^{-4}.

**Population Structure with PCA**

The PCA resulted in one loose cloud with no apparent substructuring throughout all sampled individuals (Figure 3a). The qqplot indicates a normal distribution with perhaps some outliers at the ends, which could be explained as migrants (Figure 3b). The best interpretation of these results is thus that this group of individuals comprises one population with a few distantly related individuals and, therefore, that no substructuring exists.

**Relatedness Estimates**

Lynch-Ritland \( r \)-values (Lynch and Ritland 1999) were calculated for each possible pairwise comparison between the 412 individuals resulting in 84,666 \( r \)-values. R-values ranged from -0.54 to 0.75 with a mean of 0.00 (SD 0.13).

**Interpolations**

**Interpretation of Maps**

Maps showing interpolations were created as heat maps. Redder areas have high degrees of relatedness to all individuals in the study area (\( r \)-value > 0). Green areas indicate a level of unrelatedness that would be expected in panmictic populations (\( r \)-value ~ 0). Bluer areas indicate increasing degrees of unrelatedness that would be reflective of non-panmictic conditions (\( r \)-value < 0), which can be caused either by structure in populations or isolation-by-distance.
**Global Scale**

Figure 4a shows the overall interpolation for the entire study area. Most of the interpolation is blue, thereby indicating that the population deviates from expectations of panmixia, which is likely attributed to effects of isolation-by-distance. Figure 4b shows areas that are statistically significant (i.e. more unrelated to the population as a whole than expected by chance). The northern parts of the study area show large areas of significance. Partitioning males and females results in differing patterns (see Figure 4c-f). Males show a significant degree of unrelatedness in northern Dalarna and northern Gävleborg relative to the population as a whole, whereas females show a significantly high degree of unrelatedness in northern Gävleborg only. This suggests that individuals in these areas are segregated from the rest of the population, but it does not reveal how individuals in these areas are related to each other. To explore this, we conducted the same analysis but with a focus specifically within these areas (the local scale) (Figure 5).

**Local Scale**

*Control (CA).* First we analysed an area with only one small patch of significance at the global scale as a control (CA). The mean predicted relatedness across CA shows patterns of high relatedness as depicted in orange shades (Figure 6a). This contrasts from results at the global scale and is likely explained by smaller scope and lack of isolation-by-distance. We checked for significance for each sex. Males showed no significant areas (Figure 6b), whereas females showed large patches of significant relatedness (Figure 6c). Since the brown bear exhibits male-biased dispersal and female philopatry, these results are consistent with our expectations: panmictic conditions for males and core areas with highly related females.

*Northern Dalarna (ND).* As for CA, ND shows patterns of relatedness across the landscape (Figure 6d). When testing for statistical significance, almost the entire area is significant. This holds true for males (Figure 6e). Females show significance for a large proportion of the area (Figure 6f). At the global scale, we found that males were significantly unrelated to the population as a whole (Figure 5d). At the local scale, we found that they were significantly more related to each other than expected by chance. This indicates that the individuals in ND are segregated from the rest of the population in the study area.

*Northern Gävleborg (NG).* The predictions in NG follow the same pattern of relatedness across the landscape as other local areas (Figure 6g). Like CA and unlike in ND, there are some areas of high relatedness that are statistically significant next to areas that are not significant. However, unlike CA, the significance exists for both sexes (Figure 6h & i) suggesting that males are not panmictic and hence that there is population structuring occurring.

**Inbreeding**

ND had significantly higher relatedness values (mean = 0.54) than both CA (mean = 0.50) and NG (mean = 0.51) areas (W = 1420, p-value = 0.016; W = 2875, p-value = 0.0079 respectively). The mean relatedness between CA and NG do not differ significantly (W = 3009, p-value = 0.8116).
Discussion
In this study, we developed a new spatially explicit method for detecting population structure in a continuous population that we call landscape relatedness (LandRel). Assuming random mating, a homogenous landscape and non-sex-biased dispersal, populations should be distributed with relatively equal levels of relatedness throughout the landscape. Where levels of relatedness are higher or lower than expected by chance, it can be inferred that one or more of the above assumptions above have been violated and that population structure exists. This is what LandRel seeks to discover. Where \textit{a priori} knowledge of landscape structure, dispersal characteristics or mating behaviours exist and differ from the above assumptions, adjustments of expectations of relatedness across the landscape can be made accordingly.

We empirically tested LandRel in the south-central population of the Scandinavian brown bear, which exhibits male-biased dispersal and female philopatry (Støen et al. 2006; Zedrosser et al. 2007). Our expectations were therefore that males would display equal levels of relatedness throughout the study area, while females would display areas of high relatedness, also known as core areas. Isolation-by-distance was previously tested (Mantel Test) using the same data and resulted in significance for females (p-value < 0.001), but not for males (p-value = 0.080) (see Norman & Spong, 2015), thus further supporting the assumption that males are panmictic throughout the study area.

In contrast to our expectations, our LandRel results showed evidence of population structure in the brown bear in two areas in the northern part of the study area. This was first identified at the global scale, where individuals in these northern areas were more unrelated to the population as a whole than expected by chance. As we focused our analyses on these areas, we determined that these individuals were also significantly more related to each other than expected by chance. Since females are philopatric and are expected to form areas of high relatedness (Støen et al. 2005), it is difficult to draw conclusions regarding population structure of females. However, in areas where males appear to be structured, females show similar patterns thereby further suggesting a barrier to gene flow, and thus population structuring for all individuals.

Since the population is continuous beyond the study area, particularly to the north and northwest, processes occurring beyond the study boundary may influence these northern areas identified as segregated. The influence may be in the form of introgression from the distinct subpopulation that exists north of the study area. Still, this does little to explain the striking differences in population-based relatedness between neighbouring individuals in northern Gävleborg. Additionally, we lack information about individuals from the gap between Dalarna and Gävleborg in the north making it difficult to know if the segregated individuals are a part of one continuous population, or if they are indeed two differentiated areas. At this time, it is unclear how biologically arbitrary sampling boundaries affect relatedness in the landscape when the population is continuous beyond the boundary.

**Implications for Brown Bear Conservation and Management**
If the identified structure is a result of physical or behavioural isolation from the remaining population, there is a risk of inbreeding, especially if there are few individuals in the isolated area. We tested both areas that appear segregated for
signs of inbreeding and found that northern Dalarna had significantly higher relatedness values than the control, which is suggestive that inbreeding has occurred. Tallmon et al. (2004) expressed their concern for possible inbreeding in this subpopulation after detecting a low effective population size and low immigration rate. As inbreeding can have severe consequences on population fitness (Newman and Tallmon 2006), this evidence for inbreeding in northern Dalarna is of conservation concern. Though further investigation is needed to determine what is causing the structuring, Nelleman et al. (2007) conducted a habitat suitability model for brown bear in an area just to the south of our local northern Dalarna area and suggested that due to the close proximity of human settlements to the west, south and east, the population may become confined. The segregated population in NW Gävleborg did not show a significant difference in relatedness values when compared with the control. However, in our analysis NW Gävleborg contains individuals that are not segregated from the main population. We therefore cannot make any conclusive statement that inbreeding has not occurred in NW Gävleborg. Our results pinpoint parts of the population that are at greater risk for reduced fitness due to isolation. Further investigation is needed to determine the cause of isolation, whether it is a barrier in the landscape or deviations from known mating or dispersal patterns.

**Landscape Relatedness (LandRel) Method**

Besides detecting population structuring, the landscape relatedness (LandRel) method can be informative for other factors affecting populations. In addition to detecting areas with inbreeding, it can provide an indication of where barriers exist in the landscape, whether it is a natural barrier or a recently developed human-induced alteration in the landscape. Additionally, if used in a monitoring program, LandRel will be sensitive to changes within the population with no lag effect, making it an ideal monitoring tool.

One key advantage of LandRel is that it provides insights into contemporary population processes. Some of the most common methods used to analyse population structure are based on genetic differentiation between areas. This gives rise to historically derived population structure going back several to hundreds of generations. While this may be highly informative for understanding the demographic history of a species and how it has evolved, it falls short if more recent population processes need to be investigated. LandRel thus provides more immediate feedback from effects of recent anthropogenic alterations to the landscape for example.

Taking genomic data one step further from allele-based to relatedness-based analysis minimizes some complications that arise with other assignment-based methods. While the r-value itself is defined using estimates of allele frequencies in the population, the resulting comparisons between individuals should not be affected due to its relative nature. This is also true for loci that are not in Hardy-Weinberg equilibrium. Additionally, r-values are not sensitive to loci under selection. Unlike most assignment methods, LandRel will indicate if there is no structure in the population. While many kinship-based methods share these advantages, LandRel is useful for continuous populations and requires no a priori knowledge on number of source populations unlike kinship-based methods (Iacchei et al. 2013). Furthermore, LandRel enabled us to detect fine-scale structuring that a commonly used method, PCA, was not able to detect.
There are a few limitations with the LandRel method that should be considered. First, if there exists two panmictic subpopulations of equal size that are isolated from each other, the initial results from LandRel will not be apparent as there would appear to be an even spread of relatedness throughout. Even so, the overall interpolated mean relatedness values will end up being lower than if the entire area was panmictic due to the pairwise comparisons between individuals residing in opposite groups. However, the same result would occur if there exists isolation-by-distance. This can be investigated further by testing for isolation-by-distance and focusing the interpolation on smaller areas, as we did in this study, since any structure between areas will become more apparent. Sampling characteristics pose further limitations to LandRel. Sampling needs to be conducted at a scale large enough to encompass most natal dispersal distances in order for LandRel to be effective. For this, a priori knowledge of dispersal would be beneficial and can be accomplished using the same data (see Norman & Spong 2015). While sampling intensity is not as important as scale per se, the higher the sampling intensity, the more accurate the interpolations will be. Finally, more research is needed to bring LandRel to its full potential. For example, testing how it works with various degrees of structuring and simulating different sampling intensities would identify important features of the method.

Collecting samples noninvasively is one approach that enables data collection on elusive and rare species from individuals that are living without ever needing to disturb them (Taberlet et al. 1999; Waits and Paetkau 2005; Smith and Wang 2014). Since noninvasive samples collected from the environment are often degraded due to exposure to UV rays and other chemical processes, use of high quality molecular markers that are insensitive to fragmented DNA are recommended. SNPs are particularly useful for degraded DNA since they require only small segments (65 to 100 base pairs). Their qualitative nature also makes them less sensitive to false alleles. Additionally, it is possible to genotype using single copy detection thereby dramatically reducing the occurrence of allelic dropouts and making it possible to accurately genotype samples with very low DNA quantities. In this study, we used SNPs with a call rate of almost 1.00 and a genotyping error rate of less than 0.0004 enabling us to maximize the amount of information extracted from the samples.

Conclusion

In this study, we show that the landscape relatedness (LandRel) method can be used for identifying fine-scale population structure within a continuous population of the southernmost Scandinavian brown bear population. It also helped identify areas with inbreeding, which can have negative consequences for a population. LandRel is a complementary method to the many other methods used to identify population structure with a focus on fine-scale spatial structure within a continuously distributed population. The primary advantage of LandRel is that it provides insight into contemporary processes, something that is highly sought after in ecological and conservation-oriented studies.
Acknowledgements
We are grateful to ESF ConGenOmics for providing the funds for a research visit where the idea for this method was first conceived. Thanks to Torsten Nygård Kristensen and Cino Pertoldi for hosting the visit. Thanks to Håvard Rue and his group for hosting another research visit. We are grateful to NIBIO (Bioforsk) and Michael Schneider for providing us with DNA extracts and spatial data. Thanks to all the citizen volunteers and the county administration boards of Dalarna and Gävleborg for organizing the sample collection. Many thanks to Helena Königsson for her tireless work in the lab. This is paper # 201X-XX in the Scandinavian Brown Bear Research Project.
Figure Captions

Figure 1. Map of Sweden with the counties of Dalarna and Gävleborg, comprising the study area, highlighted in green. (p9, line 202)

Figure 2. Map of study area showing sample locations for a) males (n=180) and b) females (n=232). (p14, line 309)

Figure 3. Results from a principal component analysis used to identify genetic differentiation among sampled individuals. Scatterplot (a) and colours indicate genetic distance between individuals. The values from PC1 were plotted using a quantile to quantile plot (b) to visually determine if they are normally distributed. The empirically derived points are the black circles and a normal distribution is expected to follow the blue line. (p14, line 315)

Figure 4. The first column shows interpolations of 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. (p15, line 333)

Figure 5. Global significance areas map from figure 4 with orange as the control area (CA), blue as northern Dalarna (ND), and green as northern Gävleborg (NG) with males as triangles and females as circles. (p16, line 346)

Figure 6. The first column shows the overall interpolation for the three local areas: a) control area (CA); d) northern Dalarna (ND); and g) northern Gävleborg (NG). The second and third column shows significant areas of relatedness for males and females respectively. (p16, line 350)
References


21
skattningar och trender.


