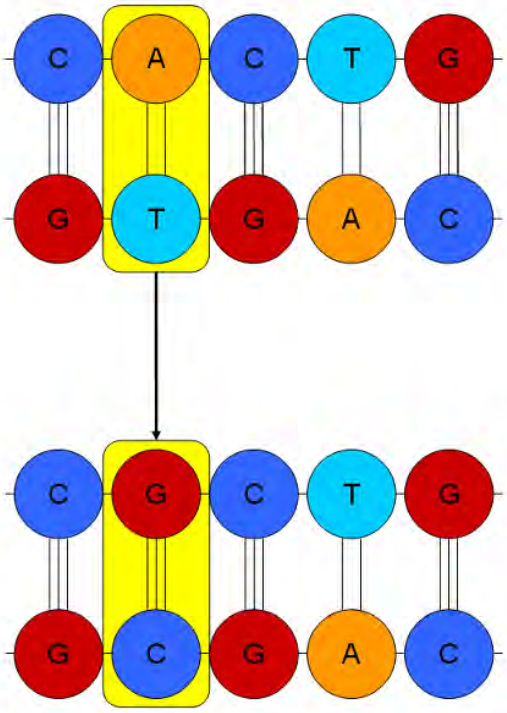
Examensarbete
i ämnet biologi

2011:6

De novo sequencing and SNP discovery in the Scandinavian brown bear (*Ursus arctos*)

Anita J Norman



Cover Figure. Depiction of a SNP in two short segments featuring the same loci of double-stranded DNA.

Sveriges lantbruksuniversitet
Fakulteten för skogsvetenskap
Institutionen för vilt, fisk och miljö
Examensarbete i biologi, 60 hp, E-nivå

Umeå 2011

Examensarbete
i ämnet biologi 2011:6

De novo sequencing and SNP discovery in the Scandinavian brown bear (*Ursus arctos*)

Sekvensering och SNP svep hos brunbjörn

Anita J Norman

Keywords: SNPs, brown bear, Ursus arctos, reduced representation, high-throughput sequencing

Handledare: Göran Spong
Examinator: Carl-Gustaf Thulin

60 hp, E-nivå
Kurskod EX0595

SLU, Sveriges lantbruksuniversitet
Fakulteten för skogsvetenskap
Institutionen för vilt, fisk och miljö

Swedish University of Agricultural Sciences
Faculty of Forestry
Dept. of Wildlife, Fish, and Environmental Studies

Umeå 2011

Abstract
Development of molecular markers that are well-suited for use in studies on wildlife dispersal can greatly enhance conservation efforts of species affected by climate change and human-induced landscape alterations. Single nucleotide polymorphisms (SNPs) are a promising marker for such studies; however, development of SNPs in non-model species with limited or no sequence data can be challenging. This study addresses these challenges in a unique way. Reduced representation libraries (RRL) of ten individual brown bears representing the geographic range in Sweden were high-throughput sequenced on one lane of the Illumina HiSeq2000, yielding reads 100 bases in length. These reads were filtered and processed following two methods: assembly using *ABySS* and alignment using *Stacks*. Putative SNPs were then called. The alignment-only method using *Stacks* was not only simpler and more efficient, it also utilized 4X more reads and yielded 30X more SNPs than the assembly method. The putative SNPs showed a clear haplotype pattern that distinguished the two known lineages of brown bear in Sweden. The outcome of this study is a good start for further work in the development of a SNP-based dispersal model with the brown bear as the model species.

Keywords: SNPs, brown bear, Ursus arctos, reduced representation, high-throughput sequencing

Introduction
Wildlife dispersal plays a central role in mitigating the threats affecting global biodiversity (Trakhtenbrot et al. 2005). As climate change causes a shift in ranges and as populations become increasingly fragmented due to large-scale changes in the landscape, much of which is human-mediated, it becomes more and more critical to understand the processes behind wildlife dispersal. Any hindrance of these processes can have a considerable impact on the viability of a population and can be a major contributing factor to eventual local or global extinction, especially as populations become small and isolated (Macdonald and Johnson 2001). With an understanding of wildlife dispersal, conservation measures can be implemented to minimize or eliminate the effects of human land-use that are known to cause barriers to dispersal.

It thus becomes a priority to develop methods that can quickly and accurately identify dispersal processes and any changes that occur in these processes over time for the conservation of species. While there are various methods of studying dispersal empirically through observation and theoretically through modelling, the use of genetics to assess dispersal can lead to insightful findings that would otherwise be impossible to detect (Peacock and Ray 2001). Additionally, sampling efforts can be done non-invasively, having minimal or no contact with the individuals being sampled, which is especially important for endangered species with small populations (eg. Henry and Russello 2010). For a detailed review of non-invasive genetic sampling methods, see Waits and Paetkau (2005). The use of genetics to infer dispersal patterns of individuals within a population is a two-step process. The first step is to determine the relatedness of individuals in a population, by reconstructing a pedigree. The second step is to associate the relations between individuals with spatial location to determine the distances between closely related individuals. If there is good life history data on the species, it will then be possible to infer the natal area of the individual and the distance from the natal area. This distance is likely the dispersal distance.

The choice of molecular marker to use for assessing relatedness in a population has recently been under scrutiny (Kuhl et al. 2011; Glover et al. 2010; Haasl and Payseur 2010; Liu et al. 2005; Rosenberg et al. 2003) and largely comes down to two types: microsatellites and single nucleotide polymorphisms (SNPs). While one microsatellite marker is roughly 4 to 12 times more informative than one SNP marker allowing it more statistical power (Liu et al. 2005), microsatellites have limitations that make them less attractive when compared to SNPs. These limitations include size homoplasy, complex mutational patterns, propensity for genotyping errors, impeded cross-laboratory reproducibility (Glover et al. 2010), labour intensive development (Haasl and Payseur 2010) and scoring can often be subjective since it is based on a quantitative difference between alleles as opposed to SNPs which are qualitative. One study indicates that microsatellites only have greater statistical power over SNPs for studies looking at recent time scales (Haasl and Payseur 2010). With the advances in sequencing and genotyping technologies and substantially lower costs than a few years ago, the requirement for more SNPs than microsatellites for comparable analyses becomes much less of an issue. More importantly, recent studies have shown that by careful selection from a large pool of SNPs, individual SNPs may be as informative as individual microsatellites, making the need for a ten-fold increase in the number of SNPs over microsatellites unnecessary (Lewis et al. 2011; Glover et al. 2010; Haasl and Payseur 2010; Lao et al. 2006; Liu et al. 2005; Rosenberg et al. 2003).

While SNPs may prove to be the molecular marker of choice for population structure analyses (Lao et al. 2006; Liu et al. 2005), there are still some obstacles to overcome, especially when working with non-model species. Most species have had very limited or no sequence data published, thus necessitating the need for *de novo* sequencing (Ratan et al. 2010; Williams et al. 2010). With the development of high-throughput sequencing, it is possible to detect upwards of tens of thousands of SNPs in a genome depending on its size, SNP density, and method used. However, with such a substantial amount of sequence data and with no reference genome, the sequence assembly can be quite challenging and time-consuming.

The scope of this study is large and involves two phases. This portion of the study represents the first phase. The aim of the first phase involves *de novo* sequencing and discovery of putative SNPs in a non-model species, the brown bear (*Ursus arctos*). The method used is aimed at simplifying both the wet lab (laboratory procedures involving hands on work with samples, reagents and tools) and dry lab (procedures done with the computer) work that is necessary with typical sequence assemblies. The second phase involves the development of a dispersal model based on pedigree reconstruction and detailed spatial data. The brown bear was chosen as the study species because it is a well-studied species that has clear dispersal patterns already known (Støen et al. 2005), thereby providing a measure of accuracy.

Materials and Methods

To ensure that some targeted SNPs would be included, I used both an anonymous high-throughput approach to find random SNPs throughout the genome and a low-throughput approach to find targeted SNPs in the mitochondria DNA and the y-chromosome.

Expected Results from High Throughput Sequencing

Preliminary calculations were carried out to estimate the expected genomic coverage and read depth from high-throughput sequencing based on a reduced representation of the whole genome. Due to a lack of genomic information on the brown bear, the calculations were based on two assumptions: The first assumption is that the average fragment size in the brown bear genome, resulting from the genomic excisions made by the restriction enzyme *BglII*, is equal to the average fragment size of 3100 bases in the human genome (Altshuler et al. 2000). Since mammalian genomes have a large proportion of conserved synteny, which has been shown between the closely related giant panda (*Ailuropoda melanoleura*) and human genomes (95.3% of the panda genome is syntenic with the human genome) (Li et al. 2010), this assumption should hold. The second assumption is that the genome size of the brown bear is 2.4 billion base pairs. The reasoning behind this assumption is two-fold. First, the C-value (a measurement of DNA content which is proportional to the genome size) is 2.75 pg in the brown bear (Vinogradov 1998) slightly less than in the dog (*Canis familiaris*), which is 2.80 pg (Tiersch et al. 1989). The entire genome of the dog has been sequenced and is known to be approximately 2.5 gigabases in size (Lindblad-Toh et al. 2005). In addition, the size of the giant panda genome is estimated to be approximately 2.4 gigabases (Yang 2010) and is the most closely related species to the bear. Using these assumptions, I calculated the hypothetical number of unique fragments of a given size using the formula in Altshuler et al. (2000). Based on the results of these calculations, I decided to sequence all fragments between 100 and 700 bases long resulting in an approximate yield of 132 000 unique fragments at a depth of 40X (eg. 40 of the same reads) per individual sample per paired read and covering 1.05% of genome (Table 1). Figure 1 shows the number of fragments based on100 nucleotide bins for the entire genome.

I aim to develop a 96-well microarray chip containing SNPs from across the genome, including mitochondrial and Y chromosome SNPs. It is therefore necessary to have a large selection of putative SNPs in order to choose the most informative. I calculated that with paired reads that are 100 bases in length and with a minimum flanking region of 35 bases on either side of the SNP that is necessary for assay development, it should still be possible to find many more putative SNPs than what is required. This affords the possibility to treat all reads as independent and non-contiguous (i.e. non-overlapping). This considerably reduces both the wet lab and dry lab work that is necessary for genome assembly and alignment, saving on time and other resources. The wet lab portion is reduced to the use of one restriction enzyme instead of the two or more that would be necessary for contiguity. The dry lab portion is simplified by eliminating the need for assembly entirely and allowing for greater ease in read alignment.

Sample Collection and DNA Extraction

Twenty tissue samples were obtained from a research biobank at the National Veterinary Institute (Statens Veterinärmedicinska Anstalt) (Uppsala, Sweden). The samples are representative of the full geographic range of brown bears in Sweden and were initially collected between 2006 and 2010. DNA was extracted from individuals (10 from liver and 10 from muscle tissue) with the QIAsymphony DNA kit on the QIAsymphony SP (Qiagen) according to the manufacturer's recommendations. The amount and quality of extracted DNA was quantified on a NanoDrop (ThermoScientific) and visualized through gel electrophoresis using the Kodak Electrophoresis Documentation and Analysis System 120.

Table 1. Estimation of sequencing yield of the brown bear genome. The calculations are broken down into 100 base bins based on fragment size and include all fragments between 100 and 700 bases long. The calculations are based on one brown bear genome estimated to be approximately 2.4 Gb in size and fragmentation of the genome by the restriction enzyme *BglII*. Sequencing is paired-end with 100 bases at each end. Fragments longer than 200 bases will not be sequenced beyond the 100 bases at each end and will thus not overlap.

Fragment Size (bases)	# Unique Fragments	# Unique Bases	Read Depth (Individual)	Genome Coverage (%)
100-200	23 795	3 569 305		0,15
200-300	23 040	4 608 005		0,19
300-400	22 309	4 461 731		0,19
400-500	21 601	4 320 101		0,18
500-600	20 915	4 182 967		0,17
600-700	20 251	4 050 186		0,17
Total:	**131 910**	**25 192 296**	**40**	**1,05**

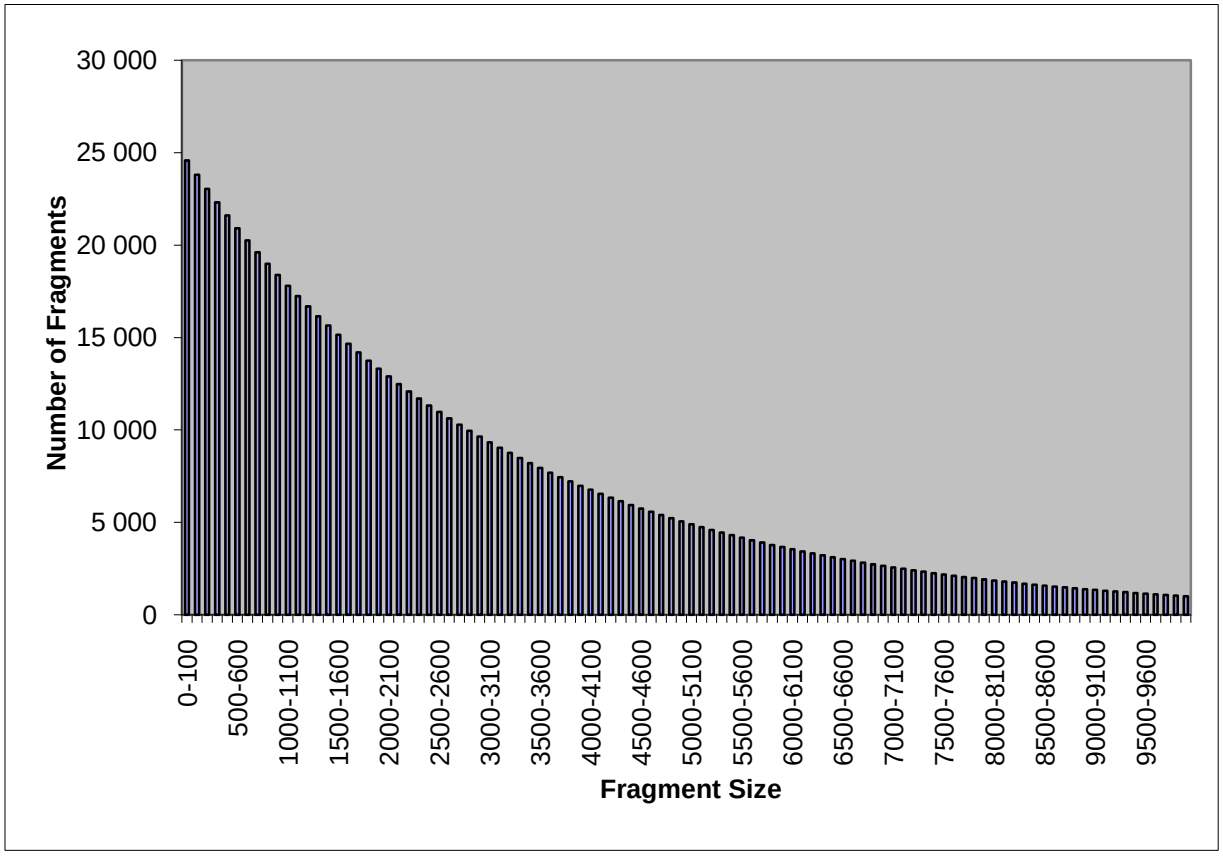


Figure 1. Theoretical whole genome fragmentation by the restriction enzyme *BglII*.

Reduced Representation Library Preparation and High-Throughput Sequencing
The DNA (0.5 µg) of 10 liver samples was digested for 16 hours with the *BglII* (Fermentas) restriction enzyme and then purified using the MinElute Reaction Cleanup kit (Qiagen) following manufacturer's recommendations. The purification was done in two elutions, the second of which was run through gel electrophoresis and the quality of the digestion was then assessed through analysis of the image obtained from the gel. The digested DNA samples were sent to the Science for Life Laboratories (Stockholm, Sweden) for library preparation. Fragments ranging in size from 100 bases to 700 bases were excised and blunt end repaired. Paired-end, multiplexed adaptors were ligated to the fragments. Equimolar concentrations of each sample were measured and run through one lane of the Illumina HiSeq2000 for paired-end sequencing. All paired reads were sequenced to 100 bases on either end of the fragment resulting in an overlap of fragments that are less than 200 bases only.

Targeted Low-Throughput Sequencing
In addition to the high throughput sequencing, I sequenced targeted loci in both the mitochondrial DNA and the Y chromosome.

Mitochondrial DNA. Primers were designed from the published mitochondrial genome (NCBI Accession # EU497665.1) of a European brown bear using Primer3 (Rozen and Skaletsky 2000). Four sets of primers were designed with each fragment approximately 500 bp in length for a total of approximately 2000 bp. The primers were chosen at random and not based on gene locations; however, the fourth primer pair spans the cytochrome-b region of the genome. Table 2 lists the primer sequences, the locations based on the reference sequence, and the product length. The total reaction volume per sample for PCR was 20 µl consisting of 2.5 µl DNA each from muscle tissue of 10 individual bears, 0.5 µl 10 µM each for forward and reverse primer, 12.23 µl distilled water, 0.5 µl 2.5 mM dNTP's, 2.0 µl 10X *Taq* buffer*, 1.6 µl 2.0 mM MgCl2*, and 0.17 µl *Taq* DNA polymerase* (*Fermentas *Taq* DNA Polymerase (native)). The optimized PCR conditions for all primer pairs includes 1 cycle of 94°C for 3 min.; 20 cycles of 94°C for 20s, 60°C less 0.5°C/cycle for 30s, 72°C for 30s; 15 cycles of 94°C for 20s, 50°C for 30s, 72°C for 30s; and 72°C for 5 min. PCR results were visualized using gel electrophoresis on the Kodak Electrophoresis Documentation and Analysis System 120 using 3.8 µl of product and 1.5% agarose gel. The remainder of the PCR product was Sanger sequenced on 3730 xl DNA analyzer (Applied Biosystems).

Table 2. Mitochondrial primers designed from the reference mitochondrial genome (NCBI Accession # EU497665.1)

Forward Sequence (5'-3')	Reverse Sequence (5'-3')	Start Position (based on ref)	Product Length
GCACCTAGCAGGCATCTCTT	CCTGTCGGGATAGCAATGAT	6661	501
GTTCGCTGTAGCCCTCATTC	ACACTCCGGATGCAAGAAGT	9422	499
CGAATCCCCCGTATCATAAA	TCGGATGTTGGTCATTAAGGT	14562	487
CGGACAACTAGCCTCCATTC	GGAGCGAGAAGAGGTACACG	16083	529

Y chromosome. Four Y chromosome primer pairs, DBY3, DBY5, DBY8 and SMCY7 were chosen from Hellborg & Ellegren (2003). The number of nucleotides expected to be sequenced is 1550 bases. The total reaction volume per sample for PCR was 20 µl consisting of 2.5 µl DNA each from muscle or liver tissue of 12 individual male bears, 0.5 µl 10 µM each for

forward and reverse primer, 12.23 µl distilled water for DBY3, DBY5 and 12.63 µl distilled water for DBY8 and SMCY7, 0.5 µl 2.5mM dNTP's, 2.0 µl 10X *Taq* buffer*, 1.6 µl 2.0 mM MgCl2* for DBY3, DBY5 and 1.2 µl 1.5 mM MgCl2* for DBY8 and SMCY7, and 0.17 µl *Taq* DNA polymerase* (*Fermentas *Taq* DNA Polymerase (native)). The optimized PCR conditions for all primer pairs includes 1 cycle of 94°C for 3 min.; 20 cycles of 94°C for 20s, 60°C less 0.5°C/cycle for 30s, 72°C for 30s; 15 cycles of 94°C for 20s, 50°C for 30s, 72°C for 30s; and 72°C for 5 min. PCR results were visualized using gel electrophoresis on the Kodak Electrophoresis Documentation and Analysis System 120 using 3.8 µl of product and 1.5% agarose gel. The remainder of the PCR product was Sanger sequenced on 3730 xl DNA analyzer (Applied Biosystems).

Alignment and SNP Calling

Sequences from the Illumina HiSeq2000 run were first de-multiplexed and placed in separate files, one pair per individual. Sequences were checked for quality with FastQC (v0.9; Babraham Bioinformatics). Adapters were removed from sequences with cutadapt (v0.9.3) and reads containing less than 50 bases were removed. PhiX bacteriophage sequences and reads with a low quality score or with ambiguous nucleotides were filtered out using rNA (v0.9.3; http://iga-rna.sourceforge.net/). The filtered sequences were again checked for quality with FastQC. One of the samples was run through a decontamination filter, DeConSeq, against a human reference and bacterial genomes (http://edwards.sdsu.edu./cgi-bin/deconseq/deconseq.cgi) to ensure the samples were not contaminated.

Two independent alignments of the reads were made for purposes of comparison. First, the overlapping paired reads were joined using a customized perl script and then run through the *ABySS* assembler (v1.2.7; Simpson *et al*. 2009) using a kmer size of 64 and 50X coverage. Reads were indexed and mapped using Bowtie (v0.12.7; Langmead *et. al.* 2010). The resulting sam files were converted to the more compressed bam file and then sorted and indexed using SAMTools (v0.1.13; Li *et. al.* 2009) view tool and sort tool respectively. SAMTools pileup tool was then used to discover variants. The resulting file was then indexed using SAMTools index tool and visualized on IGV (v2.0; Robinson *et al.* 2011).

For the second alignment, further filtering was done to ensure that all reads were the same length. Reads less than 79 bases in length were removed. Reads greater than 79 bases in length were trimmed down to 79 bases. Alignment was carried out using *Stacks* (v 0.984; Catchen *et al*. 2011) with the default parameters.

The mitochondria and y chromosome sequences were aligned using BioEdit (v7.0.9; Tom Hall, Ibis Biosciences, Carlsbad, USA) and manually screened for SNPs.

Results and Discussion

Sequencing

The sequencing run on one lane of the Illumina HiSeq2000 output 20 gigabases of sequencing data with 200 million reads. Adapter removal and quality filtering reduced the amount to 15 gigabases of sequencing data and 170 million reads, which is77% of the original sequenced data. Paired and unpaired reads from the *ABySS* assembly totalled 127 million. Reads used in the *Stacks* program that were filtered for size totalled 136 million. One of the samples did not

yield sequence data and is therefore not accounted for in the results. The one sample that was run through the decontamination filter against human and bacterial genomes resulted in 1.04% contamination using a coverage greater than or equal to 90% and identity greater than or equal to 94%. This result strongly suggests that the DNA sequences were not contaminated.

Alignment. Since the two programs used to align the reads are based on very different algorithms, it is not possible to compare the output directly. *ABySS* is an assembler which uses a de Bruijn algorithm to assemble reads where similar sequences overlap and result in the formation of contigs where two or more reads are assembled together. Since I used a single restriction enzyme, the only reads that are expected to be contiguous are those that are less than 200 bases in length. Reads that do not form contigs are not utilized; therefore, for this study, any reads that originate from fragments that are greater than 200 base pairs in length will not be included in the ABySS assembly. An analysis of the data indicates that approximately 50% of the reads are less than 200 bases in length, which is much higher than the expected number (18% calculated from Table 1). This suggests that the *BglII* restriction site occured more often that what was expected. *Stacks* uses an alignment algorithm and is independent of any contiguity between reads.

With the *ABySS* assembly, the total number of contigs output is 22,025 consisting of 7.2 million reads totalling 5.7% of the total number of paired and unpaired reads. The longest contig is 2368 bases. Since the reads are non-contiguous, the maximum contig length should be no more than 198 bases which is the maximum length of paired reads originating from fragments that were less than 200 bases in length. The number of contigs that are less than or equal to 198 bases is 20,535, totalling 93% of the original contigs. Since most of the contigs fall within the expected range of contig length, it is an indication that the *ABySS* assembly and alignment is fairly accurate. The read depth ranges from 1 to 99,887. Such a large read depth is likely to be the result of repetitive content. The number of contigs with a read depth of less than or equal to 400X (which is the expected read depth for all samples) and greater than 10X is 17,586, totalling 80% of the original contigs.

The *Stacks* alignment output approximately 1 million unique stacks (aligned reads) (Table 3), consisting of 27 million reads, and totalling 20% of the total number of filtered reads. This is four times the amount of reads that *ABySS* utilized. A subset of stacks based on heterozygous diversity and including 133 543 stacks, the read depth ranges from 2 to 5700 within each individual. The number of stacks that have a read depth of a minimum of 10X and a maximum of 100X per individual is 89,528, which is 67% of the total of number of stacks for each individual.

The total number of bases sequenced in the mitochondrial DNA is 2015 and 1489 in the Y-chromosome.

Table 3. Non-parallel comparison of the results of the two programs used in aligning the reads.

	ABySS	*Stacks*
# contigs/stacks	22 025	1 006 948
# reads included	7 243 032	26 990 267
% of total reads	5.7%	20.0%
% contigs < 199 bases	93%	N/A
% contigs/stacks with read depth between:		
10X - 400X (all samples)	80%	
10X -100X (individuals)		67%

Putative SNPs

There are two distinct lineages of brown bears in Sweden separated by a contact zone (Figure 2), one in the north that originates from Eastern Europe and one in the south that originates from Western Europe (Taberlet et al. 1995). Previous studies have revealed very distinct haplotypes in the mitochondria separating the two lineages (Taberlet and Bouvet 1994; Taberlet et al. 1995) (Figure 2). My results show the same pattern in both the mitochondria and the nuclear DNA and correspond correctly based on individual sampling locations. While these haplotype SNPs may be less informative for population structure studies aside from distinguishing between the two populations, it does indicate that the sequencing and alignment are fairly accurate. Further analysis is needed to determine the number of SNPs that are not representative of these two haplotypes. Additionally, two separate *Stacks* runs of the samples split into their respective haplotypes will reveal putative SNPs within the two lineages.

The total number of putative SNPs identified from the *ABySS* alignment after mapping the reads to the contigs is 5974 (Figure 3). The total number of putative SNPs identified from the *Stacks* alignment is 177,176, which is 30 times greater than *ABySS*. However, since half of the reads were not expected to be utilized in ABySS since they would not form contigs, it is possible that the ABySS could output twice the amount. Additionally, more analysis on the *Stacks* output is necessary to ensure that the number of stacks and SNPs are not somehow inflated by repetitive stacks. A manual scan of more than 150 SNPs from the *Stacks* output that were found within the majority of the individuals sampled shows a very strong haplotype pattern that matches the sampling locations of the individual bears separating them into the northern population (Eastern European lineage) or the southern population (Western European lineage) (Figure 4). The total number of putative SNPs identified from the mitochondrial DNA sequences is 57, of which 54 are haplotype SNPs (SNPs that distinguish the two lineages). One individual was found above the contact zone even though it matches the haplotype of the southern lineage, although it was not far from the contact zone (Figure 4). The sequencing of the Y chromosome yielded no SNPs. This lack of variation is supported with findings by (Hellborg and Ellegren 2004) that mammalian Y chromosomes have low levels of nucleotide diversity. It is possible that there are SNPs from the Y chromosome in the sequences from the high-throughput sequencing. A Blast of the SNP-containing reads to the human Y chromosome may reveal these paternally-inherited SNPs. In the absence of Y chromosome SNPs, a Y chromosome marker will be used in the final SNP chip for sex determination.

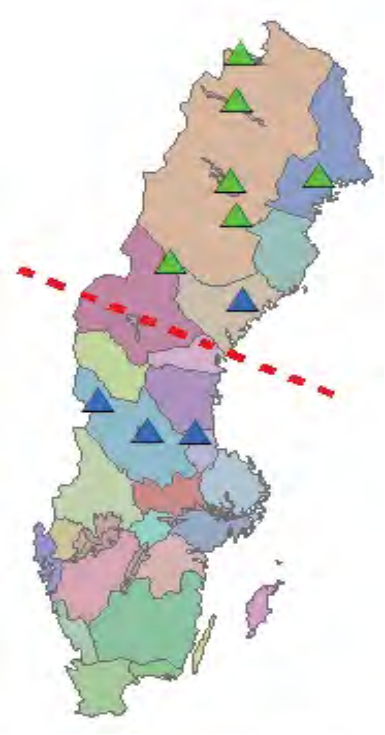


Figure 2. County map of Sweden with sampling locations of 10 bears whose mitochondria were sequenced. The mitochondrial haplotypes distiguish the eastern lineage (green triangles) and the western lineage (blue triangles). The dashed red line represents the contact zone between the two lineages.

The SNPs that are useful for developing assays are the ones that have a flanking region of 35 bases on either side of the SNP. Since the SNPs found in the *Stacks* output have a read length of 79 bases, the SNPs would need to fall within the positions 35 and 44. The number of SNPs that fall within these positions are 22 475, a total of 13% of all putative SNPs found, which remains a good amount for choosing highly informative SNPs.

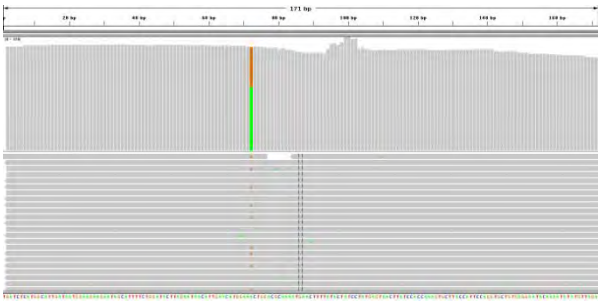


Figure 3. An example of one of the SNPs from the *ABySS* run viewed through IGV.

Figure 4. Three SNPs output from the *Stacks* program showing the two distinct haplotypes separating the northern and southern populations.

Conclusion

More putative SNPs were found than what is needed for development of a 96-well SNP microarray chip allowing for careful selection of the most informative SNPs and increasing the statistical power for pedigree reconstruction. The SNPs found between individuals show a very strong haplotype distinguishing the northern Swedish population (eastern European lineage) and the southern Swedish population (western European lineage). This result was expected and infers that sequencing and SNP-calling were quite successful.

The method chosen to discover putative SNPs across the brown bear genome has proven to be much simpler than the typical assembly methods yet yields equally robust results. When comparing the two programs *Stacks* and *ABySS*, for studies that use one restriction enzyme making assembly no longer necessary, *Stacks* had a considerably better outcome. However, more analysis is necessary to ensure that the output by *Stacks* is not somehow inflated. This study also illustrates that the need for complex analysis and intense computation can be made more efficient by simplifying the process without reducing the quality of the results.

References

Altshuler, D., Pollara, V. J., Cowles, C. R., Van Etten, W. J., Baldwin, J., Linton, L., et al. (2000). An SNP map of the human genome generated by reduced representation shotgun sequencing. Nature, 407(6803), 513-6. doi: 10.1038/35035083.

Catchen, J., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. (2011) Stacks: building and genotyping loci de novo from short-read sequences. G3: Genes, Genomes, Genetics, In Press.

Glover, K.A, Hansen, M.M., Lien, S., Als, T. D., Høyheim, B., & Skaala, O. (2010). A comparison of SNP and STR loci for delineating population structure and performing individual genetic assignment. BMC genetics, 11, 2. doi: 10.1186/1471-2156-11-2.

Haasl, R. J., & Payseur, B. A. (2010). Multi-locus inference of population structure: a comparison between single nucleotide polymorphisms and microsatellites. Heredity, 106(1), 158-171. Nature Publishing Group. doi: 10.1038/hdy.2010.21.

Hellborg, L., & Ellegren, H. (2003). Y chromosome conserved anchored tagged sequences (YCATS) for the analysis of mammalian male-specific DNA. Molecular ecology, 12(1), 283-91. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12492896.

Hellborg, L., & Ellegren, H. (2004). Low levels of nucleotide diversity in mammalian Y chromosomes. Molecular biology and evolution, 21(1), 158-63. doi: 10.1093/molbev/msh008.

Henry, P., & Russello, M. a. (2010). Obtaining high-quality DNA from elusive small mammals using low-tech hair snares. European Journal of Wildlife Research, 57(3), 429-435. doi: 10.1007/s10344-010-0449-y.

Kuhl, H., Tine, M., Hecht, J., Knaust, F., & Reinhardt, R. (2011). Analysis of single nucleotide polymorphisms in three chromosomes of European sea bass Dicentrarchus labrax. Comparative biochemistry and physiology. Part D, Genomics & Proteomics, 6(1), 70-5. Elsevier Inc. doi: 10.1016/j.cbd.2010.04.003.

Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25.

Lao, O., Duijn, K. van, Kersbergen, P., Knijff, P. de, & Kayser, M. (2006). Proportioning whole-genome single-nucleotide-polymorphism diversity for the identification of geographic population structure and genetic ancestry. American journal of human genetics, 78(4), 680-90. doi: 10.1086/501531.

Lewis, J., Abas, Z., Dadousis, C., Lykidis, D., Paschou, P., & Drineas, P. (2011). Tracing Cattle Breeds with Principal Components Analysis Ancestry Informative SNPs. PloS one, 6(4), e18007. doi: 10.1371/journal.pone.0018007.

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R. and 1000 Genome Project Data Processing Subgroup (2009) The Sequence alignment/map (SAM) format and SAMtools. Bioinformatics, 25, 2078-9. [PMID: 19505943]

Li, R., Fan, W., Tian, G., Zhu, H., He, L., Cai, J., et al. (2010). The sequence and de novo assembly of the giant panda genome. Nature, 463(7279), 311-7. doi:10.1038/nature08696.

Lindblad-Toh, K, et al. (2005). Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature 438, 803-819.

Liu, N., Chen, L., Wang, S., Oh, C., & Zhao, H. (2005). Comparison of single-nucleotide polymorphisms and microsatellites in inference of population structure. BMC genetics, 6 Suppl 1, S26. doi: 10.1186/1471-2156-6-S1-S26.

Macdonald , D.W., Johnson, D.D.P. (2001). Dispersal in theory and practice: consequences for conservation biology. In: Clobert J., Danchin E., Dhondt A.A., Nichols J.D. (eds) *Dispersal*. Oxford University Press, Oxford, UK, pp 358–372.

Peacock, M.M., Ray, C. (2001). Dispersal in picas (Onchotona princeps): combining genetic and demographic approaches to reveal spatial and temporal patterns. In: Clobert J., Danchin E., Dhondt A.A., Nichols J.D. (eds) *Dispersal*. Oxford University Press, Oxford, UK, pp 43-56.

Ratan, A., Zhang, Y., Hayes, V. M., Schuster, S. C., & Miller, W. (2010). Calling SNPs without a reference sequence. BMC bioinformatics, 11, 130. doi: 10.1186/1471-2105-11-130.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., Mesirov, J.P. (2011) Integrative Genomics Viewer. Nature Biotechnology 29, 24–26.

Rosenberg, N. a, Li, L. M., Ward, R., & Pritchard, J. K. (2003). Informativeness of genetic markers for inference of ancestry. American journal of human genetics, 73(6), 1402-22. doi: 10.1086/380416.

Rozen, S., Skaletsky, H.J. (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Humana Press, Totowa, NJ, pp 365-386 Source code available at http://fokker.wi.mit.edu/primer3/.

Simpson, J.T., Wong, K., Jackman, S.D., Schein, J.E., Jones, S.J., Birol, I. (2009) ABySS: A parallel assembler for short read sequence data. *Genome Research* (Genome Research, PubMed).

Støen, O.-G., Bellemain, E., Sæbø, S., & Swenson, J. E. (2005). Kin-related spatial structure in brown bears Ursus arctos. *Behavioral Ecology and Sociobiology*, *59*(2), 191-197. doi: 10.1007/s00265-005-0024-9.

Taberlet, P, & Bouvet, J. (1994). Mitochondrial DNA polymorphism, phylogeography, and conservation genetics of the brown bear Ursus arctos in Europe. *Proceedings. Biological sciences / The Royal Society*, *255*(1344), 195-200. doi: 400 10.1098/rspb.1994.0028.

Taberlet, Pierre, Swenson, J. E., Sandegren, F., & Bjarvall, A. (1995). Localization of a Contact Zone between Two Highly Divergent Mitochondrial DNA Lineages of the Brown Bear Ursus arctos in Scandinavia. *Conservation Biology*, *9*(5), 1255-1261.

Tiersch, T.R., Chandler, R.W., Wachtel, S.S., Elias, S. (1989a). Reference standards for flow cytometry and application in comparative studies of nuclear DNA content. *Cytometry* 10: 706-710.

Trakhtenbrot, Ana, Ran Nathan, Gad Perry, and David M. Richardson. 2005. “The importance of long-distance dispersal in biodiversity conservation.” Diversity and Distributions 11:173-181. http://blackwell-synergy.com/doi/abs/10.1111/j.1366-9516.2005.00156.x.

Waits, Lisette P, and David Paetkau. 2005. “Noninvasive genetic sampling tools for wildlife biologists : a review of applications and recommendations for accurate data collection noninvasive genetic sampling tools for wildlife biologists : a review of applications and recommendations for accurate d.” Society.

Williams, Larissa M, Xin Ma, Adam R Boyko, Carlos D Bustamante, and Marjorie F Oleksiak. 2010. “SNP identification, verification, and utility for population genetics in a non-model genus.” BMC genetics 11:32. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2874759&tool=pmcentrez&rendertype=abstract.

Yang, Huanming. 2010. “Sequencing and analysis of the giant panda genome.” Science China. Life sciences 53:1047-8. http://www.ncbi.nlm.nih.gov/pubmed/20821305 (Accessed June 3, 2011).

SENASTE UTGIVNA NUMMER

2010:6 Determinants of winter browsing intensity on young Scots pine (*Pinus sylvestris*) by moose (*Alces alces*) across a bio-geographical gradient in Sweden.
Författare: Lenka Vyšínová

2010:7 Reintroduction of the noble crayfish in the lake Bornsjön.
Författare: Susanna Schröder

2010:8 Human attitudes toward large carnivores bear, wolf, lynx and wolverine. A case study of Västerbotten County.
Författare: Robert Mannelqvist

2010:9 The distribution of Moose (*Alces alces*) during winter in southern Sweden: A respons to food sources?
Författare: Mikael Wallén

2010:10 Training identification tracking dogs (*Canis familiaris*): evaluating the effect of novel trackdown training methods in real life situations.
Författare: Erik Håff

2010:11 Hotade arter i tallmiljöer på Sveaskogs mark i Västerbotten och Norrbotten. Skötselförslag och analys av potentiell habitatutbredning.
Författare: Karin Lundberg

2010:12 Migration losses of Atlantic salmon (*Salmo salar* L.) smolts at a hydropower station area in River Åbyälven, Northern Sweden.
Författare: Stina Gustafsson

2010:13 Do grizzly bears use or avoid wellsites in west-central Alberta, Canada?
Författare: Ellinor Sahlén

2011:1 Pre-spawning habitat selection of subarctic brown trout (*Salmo trutta* L.) in the River Vindelälven, Sweden.
Författare: Erik Spade

2011:2 Vilka faktorer samvarierar med användandet av viltkött, vildfångad fisk, bär och svamp i svenska hushåll? – Stad vs. Landsbygd.
Författare: Jerker Hellstadius

2011:3 Konsekvenser av födoval och minskande sorkstammar för populationer av sorkätande ugglor och rovfåglar.
Författare: Katie Andrle

2011:4 Tjäderns (Tetrao urogallus L.) vinterdiet i norra Sverige: Är gran (Picea abies) viktig i vissa habitat?
Författare: Staffan Öberg

2011:5 Grey-sided vole and bank vole abundance in old-growth forest patches of different size and connectivity
Författare: Niklas Paulsson

Hela förteckningen på utgivna nummer hittar du på www.slu.se/viltfiskmiljo