MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2011) **11 (Suppl. 1),** 9–19 doi: 10.1111/j.1755-0998.2010.02941.x

SNP DISCOVERY: CHAIN TERMINATION SEQUENCING

A simple route to single-nucleotide polymorphisms in a nonmodel species: identification and characterization of SNPs in the Artic ringed seal (*Pusa hispida hispida*)

MORTEN TANGE OLSEN,* VERONICA HIRSH VOLNY,† MARTINE BÉRUBÉ,*† RUNE DIETZ,‡ CHRISTIAN LYDERSEN,§ KIT M. KOVACS,§ RICHARD S. DODD† and PER J. PALSBØLL*†
Evolutionary Genetics Group, Department of Genetics, Microbiology, and Toxicology, Stockholm University, Svante Arrhenius Väg 20C, S-106 91 Stockholm, Sweden, †Department of Environmental Science, Policy and Management, University of California Berkeley, 321 Mulford Hall, Berkeley, CA 94720, USA, ‡Department of Arctic Environment, National Environmental Research Institute, Aarhus University, Frederiksborgvej 399, DK-4000 Roskilde, Denmark, §Polar Environmental Centre, Norwegian Polar Institute, NO-9296 Tromsø, Norway

Abstract

Although single-nucleotide polymorphisms (SNPs) have become the marker of choice in the field of human genetics, these markers are only slowly emerging in ecological, evolutionary and conservation genetic analyses of nonmodel species. This is partly because of difficulties associated with the discovery and characterization of SNP markers. Herein, we adopted a simple straightforward approach to identifying SNPs, based on screening of a random genomic library. In total, we identified 768 SNPs in the ringed seal, *Pusa hispida hispida*, in samples from Greenland and Svalbard. Using three seal samples, SNPs were discovered at a rate of one SNP per 402 bp, whereas re-sequencing of 96 seals increased the density to one SNP per 29 bp. Although applicable to any species of interest, the approach is especially well suited for SNP discovery in nonmodel organisms and is easily implemented in any standard genetics laboratory, circumventing the need for prior genomic data and use of next-generation sequencing facilities.

Keywords: Arctic, discovery, genomics, natural populations, single-nucleotide polymorphism

Received 17 May 2010; revision received 19 August 2010; accepted 26 August 2010

Introduction

Analysis of single-nucleotide polymorphisms (SNPs) has now become standard in human genetics and genome mapping in model organisms. However, despite multiple and distinct advantages of SNP genotyping, this approach is still rare in nonmodel organisms. The attractiveness of SNPs relative to e.g. microsatellite markers is a result of the abundance of SNPs throughout the genome, the binary nature of the data and the high potential for automation, both in terms of data extraction and exchange (reviewed in Brumfield *et al.* 2003; Morin *et al.* 2004; Syvänen 2001). Moreover, contrary to microsatellites, SNPs are assumed to evolve in a manner more similar to the infinite site mutation model (Kimura 1969), upon which many population genetic approaches are based (Hedrick 1999). Because SNPs usually are biallelic, more SNPs are required comparable to microsatellites [which typically have 5–20 alleles (Anderson & Garza 2006)] to achieve a similar power. However, the high abundance of SNPs and the automation potential make such an increase in experimental throughput a nonissue.

There is, however, a major stumbling block hindering wide adoption of SNPs in nonmodel species–the isolation and characterization of SNPs. This was also the case during early stages in the adoption of microsatellites, but methods of isolation and characterization of microsatellites has since evolved into standard population genetic protocols. Given the potential of SNPs in ecology, evolution and conservation (Belfiore *et al.* 2003; Bensch *et al.* 2002; Primmer *et al.* 2002; Seddon *et al.* 2005; Smith *et al.* 2005), developing simple and efficient protocols to identify and characterize SNPs in nonmodel species that can be achieved with standard laboratory equipment is a high priority.

There are essentially two main venues for SNP discovery in nonmodel species; (i) utilizing DNA sequences from related species; or (ii) *de novo* screening of shot-gun libraries in target species. Applications of the former

Correspondence: Morten Tange Olsen, Fax: (+46) 08 16 43 15; E-mail: mtolsen@gmt.su.se

approach include placing primers in exons to amplify the variable intervening introns (Aitken *et al.* 2004; Cappuccio *et al.* 2006; Primmer *et al.* 2002; Sacks & Louie 2008) or targeting expressed sequence tags (ESTs), by either creating new libraries (Campbell *et al.* 2009) or utilizing existing libraries (Ferber *et al.* 2008). The reliance on existing sequence data limits such targeted gene methods to well-described species complexes (Aitken *et al.* 2004), a limitation that will diminish as more genome sequences become available, but which currently represents a significant barrier.

The latter strategy, sequencing random sections of genomic DNA, has been used in various forms, such as: sequencing AFLP (Amplified Fragment Length Polymorphism) bands (Bensch *et al.* 2002; Karl & Avise 1993; Nicod & Largiader 2003; Roden *et al.* 2009); RAD tags using next-generation sequencing (Baird *et al.* 2008); or sequencing of shot-gun libraries (Cramer *et al.* 2008; Primmer *et al.* 2002). Among all of these approaches, the latter is best suited to nonmodel species. It requires no prior genomic information or access to next-generation sequencing facilities, and it is thus the strategy with the broadest potential applicability. Importantly, it enables a stepwise approach, whereby new SNPs can be discovered as the need arises. Here, we describe how the construction and sequencing of a random genomic library allowed us to identify a total of 768 SNPs in a sample of 96 Arctic ringed seals, *Pusa hispida hispida*.

Material and methods

Tissue collection, storage and DNA extraction

Tissue samples from ringed seals collected in subsistence hunting in Greenland were obtained by biologists from the National Environmental Research Institute, Aarhus University, Denmark, and samples from Svalbard were collected by biologists from the Norwegian Polar Institute. Samples were frozen immediately and stored at −20 °C until analyses were performed. We employed the Qiagen™ DNeasy Blood and Tissue kit to extract genomic DNA from muscle or kidney tissue following the manufacturer's instructions. Sex of sampled animals was determined by visual examination of reproductive organs.

Cloning

Genomic DNA was digested using either *Hin*dIII or *Eco*RI restriction enzymes (New England BioLabs, Inc.). Digested DNA was ligated into pBlueScript II KS(+) vectors (Stratagene, Inc.) and transformed in XL1-Blue cells (Stratagene, Inc.) following the protocol of Hillis *et al.* (1996). The transformed cells were plated onto agar containing ampicillin, X-Gal and IPTG for blue/white selection of positive clones. Sterile toothpicks were used to transfer white colonies to 50 μL ddH_2O. Cells were lysed by incubation at 100 °C for 5 min, and cellular debris was removed by centrifugation.

The supernatant containing plasmid DNA was used to amplify the inserted DNA using standard T3 and T7 oligonucleotide primers. Polymerase chain reactions (PCR) were carried out in a final volume of 10 μL containing 2 μL purified plasmid, 0.4 units of *Taq* DNA polymerase (New England BioLabs, Inc.), 100 μM of each dNTP, 4 mM $MgCl_2$ (Gibco, Inc.), 1 mM Tris–HCl (pH 8.3), 5 mM KCl and 0.6 μM of each oligonucleotide primer. PCR conditions were 95 °C for 5 min followed by three cycles of 95 °C for 1 min, 50 °C for 1 min and 30 s and 72 °C for 2 min. This was followed by 32 cycles of 95 °C for 30 s, 50 °C for 30 s and 72 °C for 2 min. Finally, 8 min of extension was performed at 72 °C completing the reaction. PCR products were separated by electrophoresis through 1.7% agarose gels (FMC, Inc.) in 1× TBE buffer (Sambrook *et al.* 1989).

Clones with insert sizes at ≥500 bp were subsequently sequenced using the T3 oligonucleotide primer and the ABI BigDye™ Terminator v3.1 cycle sequencing kit (Applied Biosystems Inc.) according to the manufacturer's protocol. The order of sequencing fragments was resolved by electrophoresis through a 5% denaturing LongRanger™ (FMC, Inc.) on an Applied Biosystems Inc., DNA Prism377™ automated sequencer. Inserts that were not sequenced in their entirety using the T3 oligonucleotide primer were sequenced in the reverse direction using the T7 oligonucleotide primer following the same protocol.

SNP discovery

The program Primer3 (Rozen & Skaletsky 2000) was employed to design ringed seal-specific forward and reverse oligonucleotide primers to amplify and sequence the largest possible fragment of each cloned insert. Fragments were sequenced in both directions in one individual, randomly chosen among three seals constituting the discovery panel. The three seals originated from three different localities selected to encompass the range of the study population. These included Avanersuaq (Thule), on the Northwest coast of Greenland; Nanortalik, on the southern tip of Greenland; and King Oscars Fjord, on the central East coast of Greenland close to Ittoqqortoormiit (Scoresby Sound). Initially, the discovery panel also included a forth sample from Svalbard, located northeast of Greenland, but this sample was later dropped because of inconsistent PCR success.

Single-nucleotide polymorphisms were identified as double peaks (representing two different bases) at a single

position in a single sequence. If no SNPs were detected, the fragment was sequenced in a seal from one of the other geographical locations, until either a SNP was revealed or samples from all geographical locations had been screened. Sequencing in both directions served to confirm candidate SNPs and reduce any possible false discovery rate. However, in a few cases, an array of tandem repeats close to a candidate SNP precluded sequencing in both directions. In these cases, SNPs were confirmed by performing an additional (independent) amplification of the target region followed by sequencing in the same direction in which the SNP was originally discovered.

PCRs were carried out in a final volume of 20 μL, containing 0.4 units of *Taq* DNA polymerase (New England BioLabs, Inc.), 100 μM of each dNTP, 67 mM Tris–HCl (pH 8.0), 2 mM $MgCl_2$, 16.6 mM $(NH_4)_2SO_4$, 11 mM b-mercaptoethanol, 0.2 mM dNTPs, 0.6 μM of each oligonucleotide primer and approximately 10 ng of genomic DNA. PCR conditions were 95 °C for 5 min, followed by 35 cycles of 95 °C for 30 s, 60 °C for 1 min and 72 °C for 1 min and 30 s. Reactions were completed at 72 °C for 8 min. Amplifications were assessed by electrophoresis, cleaned using the SAP-EXO protocol (Werle *et al.* 1994) and subsequently sequenced by cycle sequencing. Cycle sequencing and electrophoresis conditions were as described earlier.

Re-sequencing

The fragments containing putative SNPs were targeted for re-sequencing in a panel of 96 ringed seal samples originating from three different geographical locations. This included 35 seals from Avanersuaq, 35 from Ittoqqortoormiit, close to King Oscars Fjord, and 26 seals from Svalbard. In addition to the re-sequencing panel, we included eight DNA extraction controls (i.e. samples that were re-extracted and genotyped) and 26 genotyping controls (i.e. samples for which the same extraction was genotyped twice) to evaluate data quality and sources of error. These 34 controls were genotyped in a subset of 22 loci, representing the best (n = 7), middle (n = 8) and worst (n = 7) performing, respectively, of all loci in terms of average quality values (QV). First, each SNP was assigned a QV estimated as 10*log (P_e), where P_e is the probability of error. Second, the per locus average QV was determined for each locus as the average QV of the SNPs within a locus. Finally, loci were ranked according to their average QV and the best, middle and worst performing loci determined. Note that we throughout use the term 'loci' when referring to DNA sequences containing putative SNPs and the term 'SNPs' when referring to single polymorphic sites within such loci.

Re-sequencing was performed by Polymorphic DNA Technologies, Inc., Alameda, CA, USA, according to the following criteria: (i) loci should preferably be between 200 and 300 bp in length; (ii) new primers should be designed to provide reliable bidirectional coverage of the entire locus; (iii) more than 95% of the samples should yield reliable sequences for a given locus; (iv) loci should contain at least one polymorphic site that was not a singleton and the number of SNPs deviating from within population Hardy–Weinberg equilibrium should be at a minimum.

Primers were designed using the company's proprietary primer design tools to create two pairs of nested PCR primers. Amplifications were carried out in nested PCRs. The first reaction was conducted in a total volume of 5 μL, consisting of: 5 ng genomic DNA; 10 picomoles of each primer; and AmpliTaq Gold® PCR Master Mix [250 U (0.05 U/μL) AmpliTaq Gold DNA polymerase, 30 mM Tris/HCl, pH 8.05, 100 mM KCl GeneAmp PCR Gold Buffer, 400 μM of each dNTP, 5 mM $MgCl_2$ and stabilizers] (Applied Biosystems, Inc.). In the nested reaction, 1 μL of boost product was used as a template; except for that, reactions were similar. PCR was conducted using an Eppendorf Mastercycler 384 (Eppendorf, AG.). Cycling conditions were 94 °C for 4 min, followed by 40 cycles of 94 °C for 15 s, 60 °C for 15 s, 72 °C for 1 min and finally one cycle of 72 °C for 7 min. The nested PCR cycling conditions were 94 °C for 4 min followed by 10 cycles of touchdown PCR with 94 °C for 15 s, 70 °C for 15 s with a 1 °C decrease each cycle and 72 °C for 1 min. Hereafter followed 30 cycles of 94 °C for 15 s, 60 °C for 15 s, 72 °C for 1 min and finally one cycle of 72 °C for 7 min. Millipore Montage PCR384 filter plates (Millipore, Inc.) were used for PCR clean-up of the nested reaction prior to cycle sequencing. Cycle sequencing was conducted in a total volume of 5 μL (0.25× reaction) using BigDye® Terminator v3.1 (Applied Biosystems, Inc.) and 1.6 pmol of each primer following the manufacturer's instructions. Millipore SEQ384 filter plates (Millipore, Inc.) where used for dye terminator removal and the length and order of cycle sequencing products were resolved on an ABI 3730xl DNA Analyzer™ (Applied Biosystems, Inc.).

SNP characterization

Genotype data was uploaded to SNPator (Morcillo-Suarez *et al.* 2008) for comparing the genotypes of samples and controls, as well as estimating allele frequencies, observed and expected heterozygosity and deviations from Hardy–Weinberg equilibrium at each SNP. We used GENEPOP v. 4.0 (Rousset 2008) to assess the degree of linkage among loci. However, rather than making pairwise comparisons for all SNPs, we randomly chose one SNP from each locus. We performed two tests: one among the presumably independent loci and one

between pairs of physically linked loci (see Results). Confidence intervals were calculated using the standard error (SE) and corrected for small sample size using the *t*-distribution when necessary (Fowler *et al.* 1998). Analyses were performed on the full set of 96 samples and in some instances also on the three subsets corresponding to geographical location, i.e. Avanersuaq, Ittoqqortoormiit and Svalbard.

Calculating allele frequency spectra

The allele frequency spectrum (AFS) of a population is the distribution of allele frequencies of polymorphic sites (Ewens 1972). This distribution reflects the demographic and molecular processes a population has experienced, but is also sensitive to ascertainment bias. Ascertainment bias is of special concern for biallelic markers, such as SNPs, and may arise when using an unrepresentative discovery panel (Ellegren *et al.* 1995; Nielsen 2000; Wakeley *et al.* 2001). As described earlier, the number of geographical localities represented in our initial discovery panel was reduced from four to three, as the samples obtained from Svalbard had a low PCR success rate. We assessed the effects of excluding Svalbard from the discovery panel, but including the locality in the re-sequencing panel, by calculating AFSs for each of the three localities in the re-sequencing panel (Avernasuaq, Ittoqqortoormiit and Svalbard), including only those SNPs that were initially detected in the discovery panel. The rationale was that if populations are highly structured, the SNPs identified in the discovery panel are less likely to occur in Svalbard, leading to a bias in the distribution of alleles. More specifically, because the discovery panel only consisted of three individuals, the SNPs identified are expected by chance to be common at localities included in the discovery panel (i.e. Avernasuaq and Ittoqqortoormiit), but rare in Svalbard, biasing the AFS of the latter towards rare alleles. Because the ancestral allelic state is unknown, we used the frequency of the minor, less frequent, allele, which results in folded AFSs (Fu 1995). To facilitate comparison among subsamples, we estimated the proportion of SNPs, rather than the absolute number of SNPs, within each minor allele frequency class. Thus, each AFS was obtained by grouping SNPs into minor allele frequency classes and estimating the relative proportions of SNPs in different allele frequency classes. Given the relatively low number of SNPs in the discovery panel, SNPs were divided into minor allele frequency classes of 5% (i.e. 0%, 0–5%, 5–10%, etc.). Next, because unbiased AFSs also provide information about demographic and molecular processes, we estimated the AFSs again, now including all the additional SNPs detected in the re-sequencing panel. We did this for each of the three localities Avanersuaq, Ittoqqortoormiit and Svalbard, as well as for the entire sample of 96 individuals. The latter represents the unbiased AFS of the study population and might serve as a reference for comparison among localities. Calculations were performed as above-mentioned with the exception that minor alleles were divided into frequency classes of 2%.

Results

Cloning and SNP discovery

We designed pairs of oligonucleotide primers for 201 of the loci sequenced from the shot-gun library of ringed seal genomic DNA. Of these, 145 were sequenced and 117 loci were found that contained one or more candidate SNPs. Among these, approximately one-third were discovered in DNA from the ringed seal collected off Avanersuaq, one-third from King Oscars Fjord and one-third from Nanortalik. A total of 261 SNPs were detected yielding an average number of SNPs per locus of 1.8 (95% CI: 1.52–2.08), or 2.2 (95% CI: 1.93–2.53) if we exclude the loci where no SNPs were detected. Overall our sequencing effort covered 86 992 bps, resulting in a frequency of one SNP per 402 bp. Of the 216 SNPs initially discovered, 169 were transitions and 47 were transversions. The false discovery rate of SNPs was assessed by reverse sequencing or performing independent PCRs on 109 of the 117 loci found to contain candidate SNPs. Eighty loci contained at least one SNP that was confirmed by reverse sequencing, 23 loci contained at least one SNP that was confirmed by independent PCR, and six loci contained SNPs that were not confirmed. Thus, 94–95% of the putative SNPs, identified in the discovery panel, were confirmed by reverse sequencing or independent PCR, implying a false SNP discovery rate of 5–6%.

Re-sequencing

Among the 117 loci found to contain candidate SNPs in the discovery panel, we supplied Polymorphic DNA Technologies, Inc. with primer and reference information for 107 loci, varying from a few 100 to more than 1000 bp in length (Fig. 1a). These loci were subsequently split into 279 shorter loci of 200–300 bp in length following Polymorphic's criterion (i) for re-sequencing (Fig. 1b). Of these 279 loci, 134 passed quality control as formulated by criteria (ii) and (iii), and 116 were recommended for genotyping following the re-sequencing criterion iv). Among these, we selected 96 loci for re-sequencing (Fig. 1c, Table S1, Supporting Information, GenBank accession numbers HN268891–HN268986). Note that, as a consequence of the initial splitting up of longer loci into shorter loci, these 96 re-sequencing loci were situated on only 77 of the initial and presumably unlinked loci we

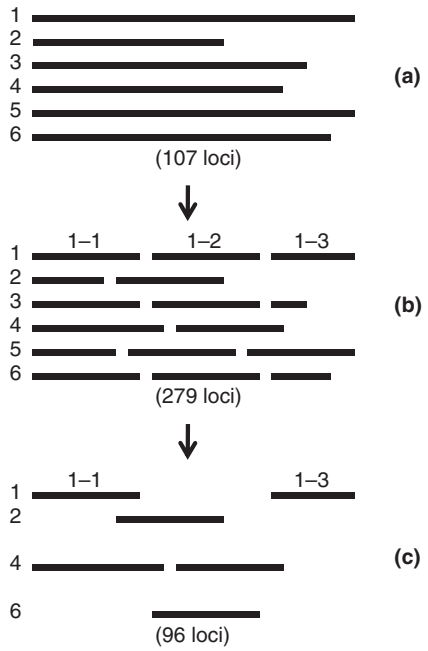

Fig. 1 Flowchart illustrating the development and selection of loci for re-sequencing in ringed seals. Initially, primer and reference information for 107 loci varying in length from a few 100 to more than 1000 bp was supplied by us to polymorphic, as represented by locus 1–6 in (a). These 107 loci were split up in 279 shorter loci of 200–300 bp in length, as illustrated in (b). Following primer design, optimization and quality control, these 279 loci were reduced to 96 loci, as illustrated in (c). Note that among these final 96 loci, some are located on the same initial locus and are thus known to be physically linked, e.g. 1-1 and 1-3 in (c). GenBank accession numbers are HN268891–HN268986.

identified during the discovery process. The test of linkage equilibrium performed among these 77 loci suggested significant deviations at the 5% level between 75 pairs of loci, corresponding to about 2.6% of the total 2926 locus-by-locus tests. In the 19 pairs of linked loci, the maximum distance between two loci was approximately 500 bp, and consequently the maximum distance between two SNPs located on separate but physically linked loci was approximately 1100 bp (given a distance of 500 bp between loci and a length of 200–300 bp per locus). When testing each pair of loci, significant deviations from linkage equilibrium at the 5% level were only observed between nine pairs.

Extraction and genotyping controls

The eight extraction controls did not match their samples in 1.4% of the single-SNP comparisons, with two loci PH192-3 and PH337-1 accounting for 26.1% and 39.1% of the mismatches, respectively. The 26 genotyping controls did not match their samples in 1.0% of the single-SNP comparisons. Here, locus PH151-2 and PH337-1 accounted for 14.3% and 67.9% of all mismatches. All putative mismatches were evaluated by visual inspection of the raw sequence files. In PH192-3, PH151-2 and the other loci containing mismatches, mismatches could be classified as genotyping errors. However, in locus PH337-1, visual inspection of the raw sequence files did not resolve the mismatching issue. Here, one of the eight extraction controls and eight of the 26 genotyping controls did not match the corresponding samples in one to several SNPs. The average number of SNP mismatches per sample in those nine samples was 4.1 (95% CI: 2.79–5.44). This pattern was repeated when evaluating a larger data set than presented here (M. T. Olsen, unpublished results).

SNP characterization

A total of 768 SNPs were detected by re-sequencing the 96 loci in the panel of 96 ringed seals (Table S2, Supporting Information). Of these, 572 were transitions, 184 were transversions, and 12 (or close to 1.6%) were triallelic for which mutation type (i.e. transition or transversion) could not be clearly determined. Triallelic SNPs were all confirmed by visual inspection of aligned sequence files (Table 1). In the majority, the rarest alleles were singletons, but in two SNPs they were doubletons, and one SNP (position 397 in locus PH391-2) contained 29 copies (>15%) of the third and rarest allele.

Re-sequencing covered a total of 21 981 bp, resulting in a frequency of one SNP for every 29 bp. This frequency was reduced to one SNP for every 63 bp if only SNPs with a minor allele frequency above 2% was considered (n = 351). The fraction of missing genotypes was 0.7%.

The number of SNPs per locus varied from 1 to 22, with an average of 8.0 (95% CI: 7.16–8.83). All loci, except

Table 1 Characteristics of the 12 triallelic single-nucleotide polymorphisms identified in ringed seals

		Allele count				
Locus	Position	A	C	G	T	Total
PH017-2	308	0	189	2	1	192
PH045-2	303	156	35	1	0	192
PH045-2	430	0	189	1	2	192
PH133-1	135	0	66	125	1	192
PH192-3	608	1	189	0	2	192
PH197-2	278	101	1	90	0	192
PH201-2	389	2	1	189	0	192
PH246-2	266	36	0	155	1	192
PH345-2	301	2	0	96	94	192
PH374-2	396	0	3	83	106	192
PH391-2	397	48	115	29	0	192
PH391-5	930	1	180	0	7	188

for PH046-2, contained at least one SNP with minor allele frequency above 1%, 85 loci (89%) contained at least one SNP with minor allele frequency above 5%, 73 loci (76%) had a SNP with minor allele frequency above 10%, and 49 loci (51%) contained a SNP with minor allele frequency above 20%.

In the combined sample, average expected and observed heterozygosity was estimated to be H_E = 0.091 (95% CI: 0.0823–0.1000) and H_O = 0.105 (95% CI: 0.0949–0.1145) (Table 2). Overall, 50 SNPs, corresponding to 6.3% (95% CI: 4.61–8.06%), deviated from Hardy–Weinberg expectations, and all of these exhibited an excess of heterozygotes. Approximately 50% of these were located on only three loci (PH064-1, PH337-1 and PH345-1). All deviating SNPs were inspected visually to evaluate the quality of genotypes. SNPs in the three loci PH064-1, PH337-1 and PH345-1 were generally difficult to score from their chromatograms, with most dubious genotype calls resembling heterozygotes. Another locus, PH133-1, containing two SNPs that deviated from Hardy–Weinberg expectations, was interpreted as being X-linked because all males (n = 45) were homozygous. The remaining 23 SNPs were easily genotyped, suggesting that deviations from Hardy–Weinberg proportions in this case could be because of chance.

Inference from allele frequency spectra

The AFSs of the SNPs detected in the discovery panel were similar in the three localities Svalbard, Avernasuaq and Ittoqqortoormiit (Fig. 2a). In a few classes, one or two localities had significantly higher or lower proportions of SNPs relative to the others, but there was no consistent bias towards rare alleles in Svalbard relative to the two other localities, suggesting that the AFS of Svalbard had not been biased by the SNP ascertainment scheme.

The AFSs of the SNPs detected in the re-sequencing panel are depicted in Fig. 2b, while Fig. 2c shows the lowest allele frequency classes including their 95% confidence intervals. The AFSs of the three localities resemble that of the entire sample, suggesting little ascertainment bias and overall similarity in the demographic and molecular processes that have affected the seals at the different sampling localities. However, 54.3% (95% CI: 50.77–57.82%) of the SNPs detected in the entire sample had a minor allele frequency of less than 2%, and a large proportion of these were monomorphic (i.e. had a minor allele frequency of 0%) in one or more of the three separate localities. This proportion was significantly higher in Svalbard at 38.8% (95% CI: 35.37–42.28%) than in Avernasuaq and Ittoqqortoormiit at 30.0% (95% CI: 26.74–33.26%) and 30.6% (95% CI: 27.36–33.90%), respectively. In the 0–2% frequency class, Svalbard at 18.7% (95% CI: 15.93–21.46%) had a significantly lower proportion of SNPs relative to the other two localities, which had approximately 25%–a pattern that could be because of the smaller sample size of Svalbard, or its geographical isolation from the other localities.

Discussion

SNP discovery and characterization

In this study, we readily isolated a large number of SNPs from an organism without any prior genomic data. We found that approximately 55% of the cloned inserts from a random ringed seal genomic DNA library could be amplified and sequenced and that they contained at least one SNP. Combining this with a relatively low false discovery rate of 5–6% suggests that constructing and sequencing random genomic libraries is a powerful alternative to other SNP discovery strategies.

The frequency of SNPs in our small discovery panel of three seals was one SNP per 402 bp. This is similar to rates reported for other mammals, such as one per 400 bp in the domestic dog (*Canis lupus familiaris*) (Brouillette *et al.* 2000), one per 306 bp in wolves (*Canis lupus*) (Seddon *et al.* 2005), one per 262 bp in the red fox (*Vulpes vulpes*) (Sacks & Louie 2008), one per 400 bp in chimpanzees (*Pan troglodytes*) (Aitken *et al.* 2004), one per 613 bp in

Table 2 Characteristics of the three sample localities included in the re-sequencing panel

Sampling locality	N	SNPs per locus	Proportion of polymorphic SNPs (%)	H_E	H_O	Proportion of SNPs deviating from HW (%)
Avanersuaq	35	5.5 (4.87–6.21)	68.9 (65.00–72.86)	0.149 (0.1369–0.1629)	0.137 (0.1248–0.1484)	5.3 (3.36–7.15)
Ittoqqortoormiit	35	5.5 (4.85–6.20)	69.4 (64.97–73.73)	0.147 (0.1347–0.1598)	0.135 (0.1224–0.1469)	4.1 (2.38–5.74)
Svalbard	26	4.9 (4.29–5.46)	60.3 (56.20–64.32)	0.165 (0.1523–0.1783)	0.153 (0.1406–0.1657)	2.6 (1.17–4.06)
Overall	96	8.0 (7.16–8.83)	–	0.091 (0.0823–0.1000)	0.105 (0.0949–0.1145)	6.3 (4.61–8.06)

Values in parentheses are 95% confidence intervals estimated by the standard error of the mean.
N, the number of diploid individuals genotyped; H_E, expected heterozygosity; H_O, observed heterozygosity; HW, Hardy–Weinberg; SNPs, single-nucleotide polymorphisms.

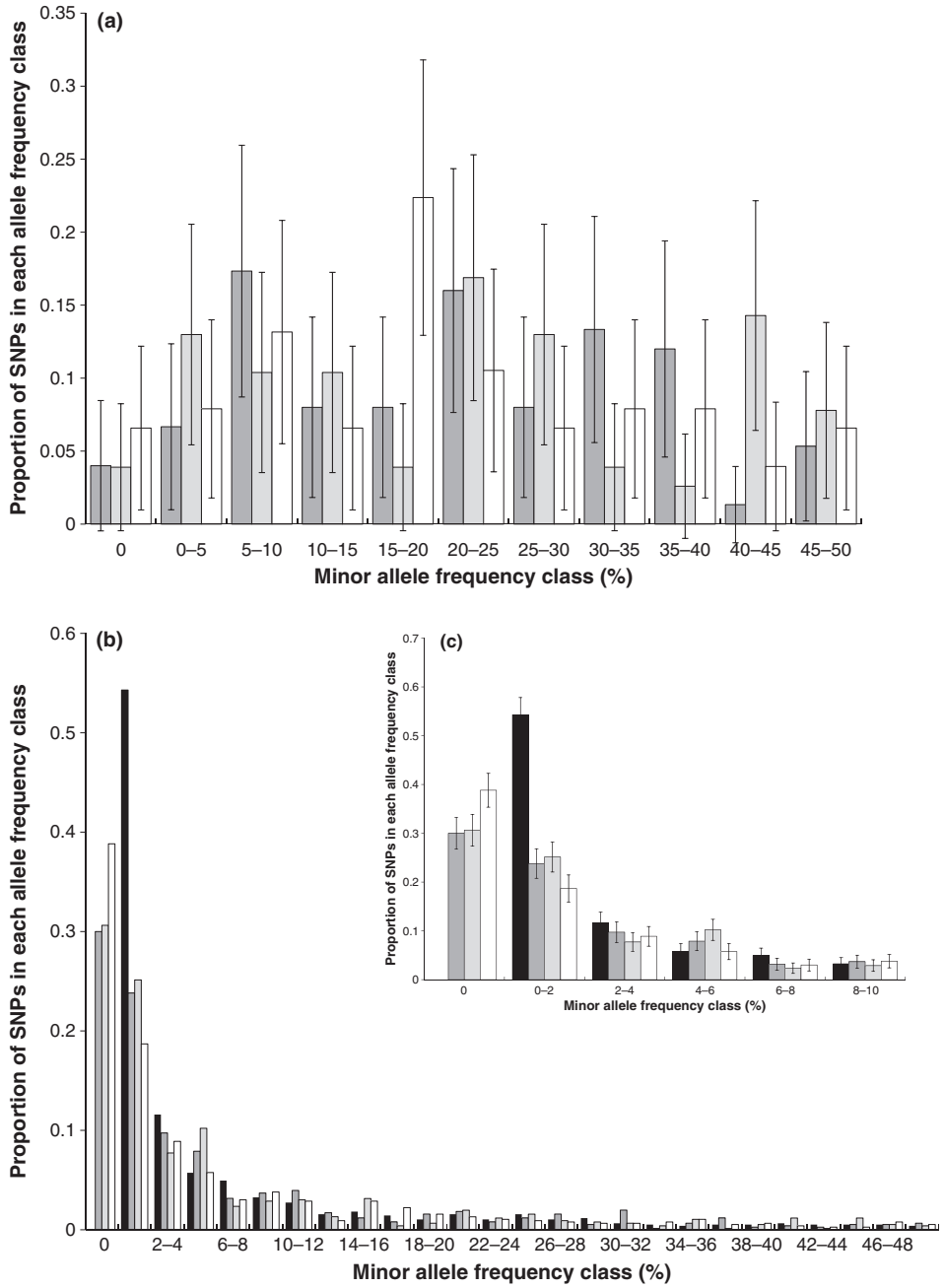

Fig. 2 The folded allele frequency spectrum of the SNPs detected (a) in the discovery panel and (b) by re-sequencing. (c) is a subset of (b), depicting the proportion of SNPs in the three lowest minor allele frequency classes. Error bars are 95% confidence intervals estimated by the standard error of the mean. Black histogram: the entire sample; dark grey: Avanersuaq; light grey: Ittoqqortoormiit; white: Svalbard.

horses (*Equus ferus caballus*) (Shubitowski *et al*. 2001) and approximately one per 1900 bp in humans (Sachidanandam *et al*. 2001).

It is noteworthy that re-sequencing of 96 ringed seals from three distinct geographical locations increased the density of SNPs to a final estimate of one SNP per 29 bp, corresponding to a density at least one order of magnitude higher than those reported for most other species. Part of this difference is probably attributed to a larger sample size in terms of individuals and number of nucleotides sequenced. Also, it is possible that some of the identified SNPs result from sequencing errors or are in fact paralogous sequence variants (PSVs). Sequencing errors could inflate the number of low frequency SNPs

and cannot be disregarded despite our extensive quality control. Similarly, it has been shown that a fraction—up to 10% in coding SNPs—of the human SNPs deposited in the SNP database (dbSNP) at the National Center for Biotechnology Information (NCBI) in reality are PSVs and not true orthologous polymorphims (Estivill *et al.* 2002; Fredman *et al.* 2004; Musumeci *et al.* 2010). If we conservatively assume that all SNPs with a minor allele frequency below 2% are caused by sequencing errors the number of SNPs are reduced to 351. If we further assume that 10% of these SNPs are PSVs, we end up with approximately 315 SNPs, corresponding to one SNP per 67 bp, or approximately half the density observed when considering all putative SNPs. Even this figure is strikingly high compared to most other mammals.

In concert, this indicates that—experimental issues aside—the high observed SNP density reflects a very large effective population size of ringed seals, perhaps coupled with recent population expansions. Such explanation agrees well with preliminary estimates of mtDNA control region haplotype diversity (M. T. Olsen, unpublished results) and is supported by accounts on ringed seal reproductive biology and abundance (Holst *et al.* 1999; Krafft *et al.* 2006; Reeves 1998; Stirling 2005). Also, although there is still uncertainty as to ringed seal population structure, the hypothesized large effective population size of ringed seals is in agreement with the general observation that ringed seals undertake long-distance movements, at least on a seasonal basis (Born *et al.* 2002; Freitas *et al.* 2008; Heide-Jørgensen *et al.* 1992; Lydersen *et al.* 2004; Teilmann *et al.* 1999) and the findings of Davis *et al.* (2008), who reported little genetic differentiation across a vast part of the ringed seals range.

Data quality

Data quality was assessed by including a number of extraction and genotyping controls as well testing all SNPs for deviations from Hardy–Weinberg expectations. In 34 sample–control pairs, 1.2% of the single-SNP comparisons did not match. The vast majority of mismatches were observed in the loci PH151-1, PH192-3 and in particular PH337-1, which alone accounted for approximately 45% of the mismatches (i.e. 0.5% of all single-SNP comparisons). With the exception of locus PH337-1, all mismatches were classified as genotyping errors upon subsequent visual inspection of the raw data. The cause of the high error rate observed at locus PH337-1 is unknown. Sample mix-up during the processes of DNA extraction, PCR amplification or genotyping can be ruled out as it would have affected all control samples at all loci and not only PH337-1. A possible explanation is that locus PH337-1 exists in duplicated segments and the inferred SNPs reflect PSVs or multisite variants (MSVs) (Fredman *et al.* 2004; Gut & Lathrop 2004). This could also account for the heterozygote excess relative to Hardy–Weinberg expectations observed for this and other loci such as PH064-1 and PH345-1. In summary, tests for Hardy–Weinberg equilibrium and comparisons between samples and controls suggests that (i) automatization of the genotyping process in this study is associated with an average error rate of approximately 0.5% and this error rate is likely to differ considerably for different SNPs and loci; (ii) that SNPs in locus PH337-1 cannot be reliably genotyped; and (iii) that deviations from Hardy–Weinberg expectations might be caused by PSVs, MSVs, sequencing errors, or chance.

Ascertainment bias

Ascertainment bias may affect estimates of population size, demographic changes and inferences of population structure, and thus is of immediate concern in terms of the inferences that may be made from SNP genotypes (Kuhner *et al.* 2000; Nielsen 2000; Wakeley *et al.* 2001). We did not select loci for re-sequencing entirely at random, but logically focused on those containing candidate SNPs in the discovery panel. This probably directed SNP discovery towards the more variable regions of the genome, and thus may result in inflated estimates of genetic diversity. However, assuming that the increase applies to all sample populations, it should not influence within-study data analyses and inference.

Deficiency of rare SNPs because of a small discovery panel is another common source of bias for SNP data sets, where discovery is carried out from existing ESTs or sequencing of novel variants, such as has been the case in several human SNP identification projects (Altshuler *et al.* 2000; Buetow *et al.* 1999; Cargill *et al.* 1999; Wang *et al.* 1998). Here, this bias was avoided by re-sequencing the entire locus in which one or more SNPs were initially discovered. This assured the additional detection of rare SNPs, an overall unbiased distribution of SNP frequency classes, as well as a three- to fourfold increase in the total number of SNPs identified.

We attempted to minimize potential bias introduced by a geographically unrepresentative discovery panel by screening for SNPs in individuals from four distinct locations spanning the range of the study population(s). However, one of these localities (Svalbard) was later excluded from the ascertainment panel. We evaluated the degree of bias potentially introduced by this exclusion by means of the AFS and found no indications of a systematic bias in the Svalbard sample. It was recently found that even for highly stationary species, a random ascertainment panel performed as well as a panel intentionally designed to include all sampling localities within

the study population (Rosenblum & Novembre 2007). Thus, although further tests and more careful examination of the ascertainment scheme relating to Svalbard is needed, we currently consider it unlikely that the data from Svalbard should be biased in a manner that significantly influences further analyses.

Finally, despite our attempts to minimize ascertainment bias, it might still have affected our data. If so, our rules for SNP selection and detailed records of the methods employed will allow for bias modelling, which has been shown to correct the process of parameter estimation (Achaz 2009; Keinan *et al.* 2007; Kuhner *et al.* 2000; Marth *et al.* 2004; Nielsen 2000; Nielsen *et al.* 2004; Nielsen & Signorovitch 2003; Wakeley *et al.* 2001).

Applicability of data

The above-mentioned analyses indicated that our data is relatively unbiased in terms of deviations from Hardy–Weinberg expectations and linkage equilibrium, as well as potential bias introduced by the ascertainment scheme. So what can we expect from this data set in terms of the population genetic inference that can be made? Most studies applying SNP markers to the study of nonmodel species have utilized a relatively small set of unlinked SNPs (Campbell *et al.* 2009; Cappuccio *et al.* 2006; Ferber *et al.* 2008; Narum *et al.* 2008; Seddon *et al.* 2005). The present data set consists of several 100 SNPs distributed across 96 loci. Of these, 77 loci are presumed to be unlinked, an assumption that was somewhat supported by our initial tests of linkage equilibrium. In addition, tests between the 19 pairs of loci known to be physically linked revealed that only approximately half of them were in linkage disequilibrium, a pattern potentially caused by high recombination rates. Thus, we expect to have at least 77 unlinked loci of 200–300 bp in length, each containing between one and 22 SNPs.

It has been suggested that if the haplotypic state of such small blocks of linked SNPs can be inferred (Stephens *et al.* 2001; Zhang *et al.* 2005), haplotypes can be treated as separate alleles. Importantly, the number of alleles potentially double for each SNP added to the haplotype (Glaubitz *et al.* 2003). This even goes for SNPs with low minor allele frequencies, which are otherwise typically ignored in population genetic studies. Such haplotype blocks might offer significant advantages in terms of statistical power relative to single-SNP data and microsatellite markers in parentage analysis, individual genetic assignment and inference of population structure (Jones *et al.* 2009; Morin *et al.* 2009; Pool *et al.* 2010). Potential downsides includes a reduced coverage of the genome, and uncertainties as to which extent genotyping and sequencing errors, missing data, recombination and homoplasy will affect haplotype inference and statistical power, although some of these factors might be avoided by only including the best performing SNPs in a given locus (Jones *et al.* 2009; Morin *et al.* 2009).

Recently, development and sequencing of RAD markers using next-generation sequencing resulted in the discovery of thousands of SNPs and population genomic inference in threespine stickleback *Gasterosteus aculeatus* (Baird *et al.* 2008; Hohenlohe *et al.* 2010). Similar to our study, re-sequencing allowed these authors to detect multiple linked SNPs within relatively small fragments of DNA that could be used as haplotype blocks. Clearly, this and future reduced representation shot-gun sequencing (RRSS) techniques offer major advantages in terms of the abundance of SNPs that can be discovered and genotyped in nonmodel organisms, although it might take some time before such techniques become economically realistic for standard population genetic laboratories.

We anticipate that further inference based on the SNP markers presented here will contribute towards an understanding of these and more specific issues relating to the ecology and evolutionary history of Arctic ringed seals. Moreover, because Arctic and North Atlantic seals in general are closely related (Fulton & Strobeck 2010a,b) and microsatellite markers developed for a specific species generally works well in other pinnipeds (Coltman *et al.* 1996; Davis *et al.* 2002), it is likely that the primers and sequence loci described here will be useful for SNP discovery in a broader range of pinnipeds, for example by means of a targeted gene approach. Finally, it is becoming clear that large-scale SNP discovery in nonmodel organisms is indeed feasible and that isolating novel SNPs in species with little genomic data is a straightforward exercise.

Acknowledgements

The project was funded by NSF grants OPP-0240992 and OPP-0520529 to Per J. Palsbøll. The authors thank Greenlandic subsistence hunters and personnel at the National Environmental Research Institute, Aarhus University, Denmark, and the Norwegian Polar Institute, Norway, for aiding in tissue sampling. We also thank James Seeb and two anonymous reviewers for thoughtful comments on the manuscript.

Conflict of interest

The authors have no conflict of interest to declare and note that the sponsors of the issue had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

Achaz G (2009) Frequency spectrum neutrality tests: one for all and all for one. *Genetics*, **183**, 249–258.

Aitken N, Smith S, Schwarz C, Morin PA (2004) Single nucleotide polymorphism (SNP) discovery in mammals: a targeted-gene approach. *Molecular Ecology*, **13**, 1423–1431.

Altshuler D, Pollara VJ, Cowles CR *et al.* (2000) An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature*, **407**, 513–516.

Anderson EC, Garza JC (2006) The power of single-nucleotide polymorphisms for large-scale parentage inference. *Genetics*, **172**, 2567–2582.

Baird NA, Etter PD, Atwood TS *et al.* (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, **3**, 7.

Belfiore NM, Hoffman FG, Baker RJ, DeWoody JA (2003) The use of nuclear and mitochondrial single nucleotide polymorphisms to identify cryptic species. *Molecular Ecology*, **12**, 2011–2017.

Bensch S, Åkesson S, Irwin DE (2002) The use of AFLP to find an informative SNP: genetic differences across a migratory divide in willow warblers. *Molecular Ecology*, **11**, 2359–2366.

Born EW, Teilmann J, Riget F (2002) Haul-out activity of ringed seals (Phoca hispida) determined from satellite telemetry. *Marine Mammal Science*, **18**, 167–181.

Brouillette JA, Andrew JA, Venta PJ (2000) Estimate of nucleotide diversity in dogs with a pool-and-sequence method. *Mammalian Genome*, **11**, 1079–1086.

Brumfield RT, Beerli P, Nickerson DA, Edwards SV (2003) The utility of single nucleotide polymorphisms in inferences of population history. *Trends in Ecology & Evolution*, **18**, 249–256.

Buetow KH, Edmonson MN, Cassidy AB (1999) Reliable identification of large numbers of candidate SNPs from public EST data. *Nature Genetics*, **21**, 323–325.

Campbell NR, Overturf K, Narum SR (2009) Characterization of 22 novel single nucleotide polymorphism markers in steelhead and rainbow trout. *Molecular Ecology Resources*, **9**, 318–322.

Cappuccio I, Pariset L, Ajmone-Marsan P *et al.* (2006) Allele frequencies and diversity parameters of 27 single nucleotide polymorphisms within and across goat breeds. *Molecular Ecology Notes*, **6**, 992–997.

Cargill M, Altshuler D, Ireland J *et al.* (1999) Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nature Genetics*, **22**, 231–238.

Coltman DW, Bowen WD, Wright JM (1996) PCR primers for harbour seal (Phoca vitulina concolour) microsatellites amplify polymorphic loci in other pinniped species. *Molecular Ecology*, **5**, 161–163.

Cramer ERA, Stenzler L, Talaba AL *et al.* (2008) Isolation and characterization of SNP variation at 90 anonymous loci in the banded wren (Thryothorus pleurostictus). *Conservation Genetics*, **9**, 1657–1660.

Davis CS, Gelatt TS, Siniff D, Strobeck A (2002) Dinucleotide microsatellite markers from the Antarctic seals and their use in other Pinnipeds. *Molecular Ecology Notes*, **2**, 203–208.

Davis CS, Stirling I, Strobeck C, Coltman DW (2008) Population structure of ice-breeding seals. *Molecular Ecology*, **17**, 3078–3094.

Ellegren H, Primmer CR, Sheldon BC (1995) Microsatellite evolution—directionality or bias? *Nature Genetics*, **11**, 360–362.

Estivill X, Cheung J, Pujana MA *et al.* (2002) Chromosomal regions containing high-density and ambiguously mapped putative single nucleotide polymorphisms (SNPs) correlate with segmental duplications in the human genome. *Human Molecular Genetics*, **11**, 1987–1995.

Ewens WJ (1972) Sampling theory of selectively neutral alleles. *Theoretical Population Biology*, **3**, 87–112.

Ferber S, Reusch TBH, Stam WT, Olsen JL (2008) Characterization of single nucleotide polymorphism markers for eelgrass (Zostera marina). *Molecular Ecology Resources*, **8**, 1429–1435.

Fowler J, Cohen L, Jarvis P (1998) *Practical Statistics for Field Biology*, 2nd edn. John Wiley & Sons Ltd., West Sussex.

Fredman D, White SJ, Potter S *et al.* (2004) Complex SNP-related sequence variation in segmental genome duplications. *Nature Genetics*, **36**, 861–866.

Freitas C, Kovacs KM, Ims RA, Fedak MA, Lydersen C (2008) Ringed seal post-moulting movement tactics and habitat selection. *Oecologia*, **155**, 193–204.

Fu YX (1995) Statistical properties of segregating sites. *Theoretical Population Biology*, **48**, 172–197.

Fulton TL, Strobeck C (2010a) Multiple fossil calibrations, nuclear loci and mitochondrial genomes provide new insight into biogeography and divergence timing for true seals (Phocidae, Pinnipedia). *Journal of Biogeography*, **37**, 814–829.

Fulton TL, Strobeck C (2010b) Multiple markers and multiple individuals refine true seal phylogeny and bring molecules and morphology back in line. *Proceedings of the Royal Society B: Biological Sciences*, **277**, 1065–1070.

Glaubitz JC, Rhodes OE, Dewoody JA (2003) Prospects for inferring pairwise relationships with single nucleotide polymorphisms. *Molecular Ecology*, **12**, 1039–1047.

Gut IG, Lathrop GM (2004) Duplicating SNPs. *Nature Genetics*, **36**, 789–790.

Hedrick PW (1999) Perspective: highly variable loci and their interpretation in evolution and conservation. *Evolution*, **53**, 313–318.

Heide-Jørgensen MP, Stewart BS, Leatherwood S (1992) Satellite tracking of ringed seals *Phoca hispida* off Northwest Greenland. *Ecography*, **15**, 56–61.

Hillis DM, Moritz C, Mable BK (1996) *Molecular Systematics*. Sinauer Associates, Inc., Sunderland, MA.

Hohenlohe PA, Bassham S, Etter PD *et al.* (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics*, **6**, 23.

Holst M, Stirling I, Calvert W (1999) Age structure and reproductive rates of ringed seals (Phoca hispida) on the northwestern coast of Hudson Bay in 1991 and 1992. *Marine Mammal Science*, **15**, 1357–1364.

Jones B, Walsh D, Werner L, Fiumera A (2009) Using blocks of linked single nucleotide polymorphisms as highly polymorphic genetic markers for parentage analysis. *Molecular Ecology Resources*, **9**, 487–497.

Karl SA, Avise JC (1993) PCR-based assays of Mendelian polymorphisms from anonymous single-copy nuclear DNA: techniques and applications for population genetics. *Molecular Biology and Evolution*, **10**, 342–361.

Keinan A, Mullikin JC, Patterson N, Reich D (2007) Measurement of the human allele frequency spectrum demonstrates greater genetic drift in East Asians than in Europeans. *Nature Genetics*, **39**, 1251–1255.

Kimura M (1969) Number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. *Genetics*, **61**, 893–903.

Krafft BA, Kovacs KM, Frie AK, Haug T, Lydersen C (2006) Growth and population parameters of ringed seals (Pusa hispida) from Svalbard, Norway, 2002-2004. *Ices Journal of Marine Science*, **63**, 1136–1144.

Kuhner MK, Beerli P, Yamato J, Felsenstein J (2000) Usefulness of single nucleotide polymorphism data for estimating population parameters. *Genetics*, **156**, 439–447.

Lydersen C, Nost OA, Kovacs KM, Fedak MA (2004) Temperature data from Norwegian and Russian waters of the northern Barents Sea collected by free-living ringed seals. *Journal of Marine Systems*, **46**, 99–108.

Marth GT, Czabarka E, Murvai J, Sherry ST (2004) The allele frequency spectrum in genome-wide human variation data reveals signals of differential demographic history in three large world populations. *Genetics*, **166**, 351–372.

Morcillo-Suarez C, Alegre J, Sangros R *et al.* (2008) SNP analysis to results (SNPator): a web-based environment oriented to statistical genomics analyses upon SNP data. *Bioinformatics*, **24**, 1643–1644.

Morin PA, Luikart G, Wayne RK (2004) SNPs in ecology, evolution and conservation. *Trends in Ecology & Evolution*, **19**, 208–216.

Morin PA, Martien KK, Taylor BL (2009) Assessing statistical power of SNPs for population structure and conservation studies. *Molecular Ecology Resources*, **9**, 66–73.

Musumeci L, Arthur JW, Cheung FSG *et al.* (2010) Single nucleotide differences (SNDs) in the dbSNP database may lead to errors in genotyping and haplotyping studies. *Human Mutation*, **31**, 67–73.

Narum SR, Banks M, Beacham TD *et al.* (2008) Differentiating salmon populations at broad and fine geographical scales with microsatellites

and single nucleotide polymorphisms. *Molecular Ecology*, **17**, 3464–3477.

Nicod J-C, Largiader CR (2003) SNPs by AFLP (SBA): a rapid SNP isolation strategy for non-model organisms. *Nucleic Acids Research*, **31**, e19.

Nielsen R (2000) Estimation of population parameters and recombination rates from single nucleotide polymorphisms. *Genetics*, **154**, 931–942.

Nielsen R, Signorovitch J (2003) Correcting for ascertainment biases when analyzing SNP data: applications to the estimation of linkage disequilibrium. *Theoretical Population Biology*, **63**, 245–255.

Nielsen R, Hubisz MJ, Clark AG (2004) Reconstituting the frequency spectrum of ascertained single-nucleotide polymorphism data. *Genetics*, **168**, 2373–2382.

Pool JE, Hellmann I, Jensen JD, Nielsen R (2010) Population genetic inference from genomic sequence variation. *Genome Research*, **20**, 291–300.

Primmer CR, Borge T, Lindell J, Sætre G-P (2002) Single-nucleotide polymorphism characterization in species with limited available sequence information: high nucleotide diversity revealed in the avian genome. *Molecular Ecology*, **11**, 603–612.

Reeves RR (1998) Distribution, abundance and biology of ringed seals (Phoca hispida): an overview. In: *Ringed seals in the North Atlantic* (eds Heide-Jørgensen MP, Lydersen C), pp. 9–45. The North Atlantic Marine Mammal Commission, Tromsø, Norway.

Roden SE, Dutton PH, Morin PA (2009) AFLP fragment isolation technique as a method to produce random sequences for single nucleotide polymorphism discovery in the green turtle, Chelonia mydas. *Journal of Heredity*, **100**, 390–393.

Rosenblum EB, Novembre J (2007) Ascertainment bias in spatially structured populations: a case study in the eastern fence lizard. *Journal of Heredity*, **98**, 331–336.

Rousset F (2008) genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103–106.

Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In: *Bioinformatics Methods and Protocols: Methods in Molecular Biology* (eds Krawetz S, Misener S), pp. 365–386. Humana Press, Totowa, NJ.

Sachidanandam R, Weissman D, Schmidt SC, Kakol JM (2001) A map of the human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*, **409**, 928–933.

Sacks BN, Louie S (2008) Using the dog genome to find single nucleotide polymorphisms in red foxes and other distantly related members of the Canidae. *Molecular Ecology Resources*, **8**, 35–49.

Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning. A Laboratory Manual*, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

Seddon JM, Parker HG, Ostrander EA, Ellegren H (2005) SNPs in ecological and conservation studies: a test in the Scandinavian wolf population. *Molecular Ecology*, **14**, 503–511.

Shubitowski DM, Venta PJ, Douglas CL, Zhou R-X, Ewart SL (2001) Polymorphism identification within 50 equine gene-specific sequence tagged sites. *Animal Genetics*, **32**, 78–88.

Smith CT, Seeb JE, Schwenke P, Seeb LW (2005) Use of the 5'-nuclease reaction for single nucleotide polymorphism genotyping in chinook salmon. *Transactions of the American Fisheries Society*, **134**, 207–217.

Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics*, **68**, 978–989.

Stirling I (2005) Reproductive rates of ringed seals and survival of pups in Northwestern Hudson Bay, Canada, 1991–2000. *Polar Biology*, **28**, 381–387.

Syvänen A-C (2001) Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nature Reviews Genetics*, **2**, 930–942.

Teilmann J, Born EW, Acquarone M (1999) Behaviour of ringed seals tagged with satellite transmitters in the North Water polynya during fast-ice formation. *Canadian Journal of Zoology*, **77**, 1934–1946.

Wakeley J, Nielsen R, Liu-Cordero SN, Ardlie K (2001) The discovery of single-nucleotide polymorphisms—and inferences about human demographic history. *American Journal of Human Genetics*, **69**, 1332–1347.

Wang DG, Fan JB, Siao CJ *et al*. (1998) Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science*, **280**, 1077–1082.

Werle E, Schneider C, Renner M, Volker M, Fiehn W (1994) Convenient single-step, one tube purification of PCR products for direct sequencing. *Nucleic Acids Research*, **22**, 4354–4355.

Zhang K, Sun FZ, Zhao HY (2005) HAPLORE: a program for haplotype reconstruction in general pedigrees without recombination. *Bioinformatics*, **21**, 90–103.

Supporting Information

Additional supporting information may be found in the online version of this article.

Table S1 Characteristics of the 96 loci for which SNPs were genotyped in the re-sequencing panel

Table S2 Characteristics of each of the 768 SNPs identified in the re-sequencing panel

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.