Conservation Biology

Contributed Paper

Genome sequencing and conservation genomics in the Scandinavian wolverine population

Robert Ekblom,[1]* Birte Brechlin,[1] Jens Persson,[2] Linnéa Smeds,[1] Malin Johansson,[1] Jessica Magnusson,[1] Øystein Flagstad,[3] and Hans Ellegren[1]

[1]Department of Ecology and Genetics, Uppsala University, Uppsala, Sweden
[2]Grimsö Wildlife Research Station, Department of Ecology, Swedish University of Agricultural Sciences, Riddarhyttan, Sweden
[3]Norwegian Institute for Nature Research, Trondheim, Norway

Abstract: *Genetic approaches have proved valuable to the study and conservation of endangered populations, especially for monitoring programs, and there is potential for further developments in this direction by extending analyses to the genomic level. We assembled the genome of the wolverine (*Gulo gulo*), a mustelid that in Scandinavia has recently recovered from a significant population decline, and obtained a 2.42 Gb draft sequence representing >85% of the genome and including >21,000 protein-coding genes. We then performed whole-genome resequencing of 10 Scandinavian wolverines for population genomic and demographic analyses. Genetic diversity was among the lowest detected in a red-listed population (mean genome-wide nucleotide diversity of 0.05%). Results of the demographic analyses indicated a long-term decline of the effective population size (N_e) from 10,000 well before the last glaciation to <500 after this period. Current N_e appeared even lower. The genome-wide F_{IS} level was 0.089 (possibly signaling inbreeding), but this effect was not observed when analyzing a set of highly variable SNP markers, illustrating that such markers can give a biased picture of the overall character of genetic diversity. We found significant population structure, which has implications for population connectivity and conservation. We used an integrated microfluidic circuit chip technology to develop an SNP-array consisting of 96 highly informative markers that, together with a multiplex pre-amplification step, was successfully applied to low-quality DNA from scat samples. Our findings will inform management, conservation, and genetic monitoring of wolverines and serve as a genomic roadmap that can be applied to other endangered species. The approach used here can be generally utilized in other systems, but we acknowledge the trade-off between investing in genomic resources and direct conservation actions.*

Keywords: genome assembly, non-invasive sampling, population genetics, single nucleotide polymorphisms

Secuenciación de Genomas y Genómica de la Conservación para la Población Escandinava de Glotones

Resumen: *Las estrategias genéticas han mostrado su importancia para el estudio y la conservación de poblaciones en peligro de extinción, especialmente para los programas de monitoreo, y todavía hay potencial para futuros desarrollos en esta dirección si se extienden los análisis hacia el nivel genómico. Ensamblamos el genoma del glotón (Gulo gulo), un mustélido que se ha recuperado recientemente de una declinación poblacional significativa en Escandinavia, y obtuvimos una secuencia inicial de 2.42 Gb que representó >85% del genoma e incluyó >21, 000 genes codificadores de proteínas. Después realizamos una resecuenciación de todo el genoma de diez glotones escandinavos para su análisis demográfico y de genómica poblacional. La diversidad genética estuvo entre las más bajas detectadas para una población en la lista roja (la diversidad promedio de nucleótidos en todo el genoma fue de 0.05%). Los resultados de los análisis demográficos indicaron una declinación a largo plazo del tamaño efectivo de la población (N_e) de 10, 000 individuos previo a la última glaciación a <500 después de este periodo. El N_e actual pareció ser incluso más bajo. El nivel de F_{IS} a lo largo del genoma fue de 0.089 (lo que posiblemente indique endogamia), pero este efecto*

*email robert.ekblom@ebc.uu.se
Article impact statement: *There is a direct link between genomics and management of a red-listed population of wolverines.*
Paper submitted October 31, 2017; revised manuscript accepted June 6, 2018.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

Conservation Biology, Volume 32, No. 6, 1301–1312
© 2018 The Authors. *Conservation Biology* published by Wiley Periodicals, Inc. on behalf of Society for Conservation Biology.
DOI: 10.1111/cobi.13157

no se observó cuando se analizó un conjunto de marcadores SNP altamente variables, ilustrando que dichos marcadores pueden brindar una imagen sesgada del carácter general de la diversidad genética. Encontramos una estructura poblacional significativa, lo que tiene implicaciones para la conectividad y la conservación de la población. Usamos tecnología de chip de circuito microfluido integrado para desarrollar una variedad de SNP que consistió de 96 marcadores altamente informativos que, junto con un paso multiplex previo a la amplificación, se aplicó exitosamente a ADN de baja calidad obtenido de muestras de excretas. Nuestros resultados informarán al manejo, la conservación, y el monitoreo genético de los glotones y funcionará como un mapa genómico que puede aplicarse a otras especies en peligro de extinción. La estrategia usada puede ser aplicada de manera general a otros sistemas, pero reconocemos la compensación existente entre la inversión en los recursos genómicos y las acciones directas de conservación.

Palabras Clave: ensamblado de genomas, genética poblacional, muestro no invasivo, polimorfismo de un solo nucleótido

Introduction

Studies characterizing levels of genetic variation in endangered species have benefitted from technical developments over the last 50 years (e.g., allozymes, restriction fragment length polymorphisms [RFLPs], microsatellite markers, other DNA-based markers) and from recent state-of-the-art whole genome sequencing (Ellegren 2014). Accordingly, there has recently been an increasing interest in using large sets of single nucleotide polymorphisms (SNPs) in conservation genetics (Morin et al. 2004; Garvin et al. 2010; Helyar et al. 2011). If sampled genome-wide, such data offer the possibility to estimate levels and character of genetic variation with high precision (Höglund 2009; Brodersen & Seehausen 2014). Draft genome assemblies, which form the ideal starting point for conservation genomic studies based on SNP markers, have recently been generated for several species of conservation concern (cf. Li et al. 2010; Prufer et al. 2012; Dobrynin et al. 2015).

Long-term population monitoring programs form the basis for conservation efforts in many parts of the world (Barea-Azcón et al. 2007). Including genetic analyses in such monitoring schemes can provide information on, for example, levels of inbreeding, population structure, and migration rates (Schwartz et al. 2007; Frankham 2010). Moreover, genetic monitoring based on noninvasive sampling is a useful means for censuses, identification of individuals, and relatedness estimations (Miller et al. 2012; Stronen et al. 2013; Liu et al. 2014). A typical example is the red-listed wolverine (*Gulo gulo*), an opportunistic predator and scavenger with a circumpolar species distribution across the Northern hemisphere that occurs at low densities in central to northern Scandinavia. The species has been closely monitored since the millennium (Aronsson & Persson 2017), and genetic analyses, mainly performed on noninvasively collected samples, constitutes a central part of this ongoing work (Hedmark & Ellegren 2007; Brøseth et al. 2010).

In large parts of their distribution, the wolverine conflicts with husbandry due to its depredation on domestic sheep (*Ovis aries*) and semidomestic reindeer (*Rangifer tarandus*) (Persson et al. 2015). Consequently, it has been persecuted by humans, which has driven the population close to extinction. In the 1960s, there were probably no more than 100 individuals in Sweden (Haglund 1965), and only small numbers remained in Norway (Landa & Skogland 1995). The species became legally protected in Sweden 1969, southern Norway in 1973, and northern Norway in 1982. Since protection, the Scandinavian wolverine population has recovered and expanded in range and size (Aronsson & Persson 2017). There are now about 850 adult individuals (Eklund et al. 2017). The species is still legally protected, but damage-mitigating lethal control in Sweden is limited, whereas culling is applied extensively to regulate the population in Norway (Gervasi et al. 2015). A more detailed description of wolverine biology is included in the Supporting Information.

Even though wolverines have a large dispersal capacity (Flagstad et al. 2004), molecular studies have raised concerns about the genetic status of the Scandinavian population. Microsatellite variability among Scandinavian wolverines is comparatively low; mean $H_0 < 0.4$ (Walker et al. 2001). Furthermore, limited intron sequencing indicates low levels of nucleotide diversity (Väli et al. 2008), and the population is fixed for a single mtDNA haplotype (Ekblom et al. 2014).

We conducted a genome-wide analysis of wolverine genetic variation and placed our results in the context of genetic monitoring and conservation. We first sequenced and produced a draft assembly of the wolverine genome. We then performed whole-genome resequencing of Scandinavian population samples to estimate levels and character of genome-wide genetic diversity. Finally, we developed an array for SNP genotyping of low-quality samples.

Methods

Additional and more detailed descriptions of methods are provided in Supporting Information.

Sample Collection, Library Preparation, and Sequencing

Tissue samples for genome sequencing (1 female from Jämtland County, Sweden) and whole-genome resequencing (10 males from throughout the Scandinavian distribution range) were provided by the Swedish National Veterinary Institute (Supporting Information). High-quality DNA was extracted using DNeasy Blood & Tissue Kit (Qiagen). Additional tissue and scat samples for SNP genotyping were collected in the field as part of a wolverine monitoring program, and were either frozen or kept in silica bead tubes (details in Brøseth et al. [2010]). The DNA-extraction from scat samples was performed using a Maxwell 16 MDx Instrument (Promega, Madison). Paired-end (library insert size range: 200–500 bp) and mate-pair (3,000–4,500 bp) sequencing for genome assembly was performed on an Illumina HiSeq 2000 instrument in 12 lanes and with read lengths of 100–144 bp.

De Novo Assembly

The genome assembly process is graphically outlined and described in detail in Supporting Information. Briefly, PCR duplicates, Illumina adapter sequences and low-quality regions were trimmed with ConDeTri (Smeds & Künstner 2011) and cutadapt (Martin 2011). The PhiX, mtDNA sequences, and reads shorter than 39 bp after trimming were discarded. We performed de novo assembly with SOAPdenovo (Luo et al. 2012), GapCloser (Li et al. 2010), and SSPACE (Boetzer et al. 2011). We further assembled putative Y-chromosome sequences by utilizing resequencing reads from the 10 males not mapping to the genome assembly (Supporting Information). We used a k-mer count approach to estimate the total genome size.

Quality Control and Annotation

We assessed the completeness of the genome assembly using CEGMA (Parra et al. 2007). Several genome annotation approaches were used in parallel, including synteny mapping to the dog genome, annotation liftovers from the ferret (*Mustela putorius furo*) genome (Peng et al. 2014) using the Kraken package (Zamani et al. 2014), and evidence-based as well as ab-initio gene predictions using Augustus (Stanke et al. 2006). Repetitive elements in the genome assembly were identified and masked using RepeatMasker applying the repeat element library of mustelids.

SNP Identification, Genotyping and Population Genomics

We estimated genome-wide levels of genetic diversity in wolverines by whole-genome resequencing of 10 males from throughout the Scandinavian population, each sequenced to 8–11 times coverage. Reads from each resequenced individual were mapped to the genome assembly using bwa (Li & Durbin 2009). Variable sites (SNPs and InDels) were called using GATK (McKenna et al. 2010). Hard filtering was applied to the raw variant calls according to the GATK guidelines. Nucleotide diversity (π) and Tajima's D were calculated with ANGSD (Korneliussen et al. 2014), and individual level of heterozygosity for each sample was estimated using VCFtools (Danecek et al. 2011). We estimated long-term N_e with the observed nucleotide diversity as a proxy for theta in the formula $\Theta = 4N_e\mu$ and assumed a per-generation mutation rate of 10^{-8}. We used PLINK (Purcell et al. 2007) to calculate linkage disequilibrium (LD) between pairs of SNP markers and NeEstimator (Do et al. 2014; Wang 2016) to produce LD-based estimates of current N_e. The pairwise sequentially Markovian coalescent (PSMC) method (Li & Durbin 2011) was used for estimation of the long term demographic history (temporal variation in N_e). An average generation time of 6 years (Nilsson 2013) was used in these calculations and the per-generation mutation rate was set to 1×10^{-8} based on previously published estimates from related species (Cahill et al. 2013; Dobrynin et al. 2015).

Population Genetics

To validate a fraction of the identified SNPs and conduct population genetic analyses, 384 high-quality and information-rich SNPs (markers with high minor allele frequency [MAF] and a maximum of 1 marker per scaffold) (see Supporting Information for details) were selected for independent genotyping using the Golden Gate assay (Illumina, San Diego, CA). The markers were successfully genotyped in 234 samples originating from throughout the Scandinavian distribution range. After removing noninformative markers and markers suggestive of being sex linked or strongly deviating from Hardy–Weinberg equilibrium, data for 357 SNPs were available. Scat samples for these analyses were collected as part of a standardized genetic monitoring effort (Brøseth et al. 2010), and tissue samples came from either dead animals sent to the National Veterinary Institute (SVA) or from animals handled in the field. The samples were divided into 4 groups according to geographic origin (northern Scandinavia, middle Scandinavia, southern Scandinavia, and southwestern Norway). Population genetic analyses of the resulting SNP data were performed using PLINK (Purcell et al. 2007), GenePop (Raymond & Rousset 1995), STRUCTURE (Falush et al. 2003), and BAPS (Corander et al. 2003).

Development of SNP-Array for Routine Genotyping of Scat Samples

With the aim of developing a SNP genotyping panel applicable to low-quality DNA samples from scats, we evaluated a PCR-based method for high throughput SNP

Table 1. Properties of the assembled draft genome sequence of the wolverine, including summary statistics of annotation based on evidence-based and ab initio gene builds.

Feature		*Quantification*
Number of scaffolds		47,417
Total assembly length		2.4 Gbp (2.2 Gbp excluding Ns)
Average length of scaffolds		51 Kbp (range 0.5 – 1,631)
Scaffold N50		178,272 bp
Contig N50		3,846 bp
GC content		41.4%
Total repeat content		829.3 Mbp
mtDNA sequence		1 contig (16,537 bp)[a]
Putative Y-chromosome sequences:		12 scaffolds (752-9,093 bp) total 44,526 bp
Genome annotation	*Evidence based*	*Ab initio*
Number of genes	26,043	21,856
Number of CDS[b]	452,356	140,589
Number of exons	514,839	176,936
Total length of gene sequences	599,879,240	551,726,382
Fraction of genome covered by genes	24.76%	22.8 %

[a] *GenBank: KF415127.1.*
[b] *Coding DNA sequences.*

genotyping using a 96 samples × 96 markers chip (Fluidigm, San Francisco, CA). We selected 96 SNP markers (out of the 384 verified using Illumina Golden Gate) based on a combination of reliable genotypes from the Golden Gate assay, high FLUIDIGM design scores, high MAFs and low levels of linkage between markers. To increase genotyping success, we applied specific target amplification (STA), a highly multiplexed pre-amplification of the SNP regions (Norman & Spong 2015).

A total of 164 noninvasively collected samples were selected for genotyping. The rate of genotyping error was assessed in 102 samples run in duplicates. Sample dropout was defined as a sample with a marker dropout rate >20 %, or a genotyping error rate >3.5 %. For such low-quality samples, no genotype calls were made. Fourteen of the samples came from individuals where tissue samples had already been genotyped using Illumina Golden Gate, thus enabling us to investigate the distribution of different types of genotyping errors in detail (Supporting Information).

Results

Draft Assembly of the Wolverine Genome

We assembled the wolverine genome based on data from 1 female sequenced to 76 X coverage (Table 1). The total genome size was estimated to 2.7 billion base pairs (Gbp). The draft assembly contained 47,417 sequences (contigs and scaffolds) (Fig. 1) representing 2.423 Gbp (2.202 excluding Ns and gaps) and thus covering >85% of the estimated genome size. Given the high depth of sequence coverage, this is likely to represent the vast majority of the genome that can be sequenced with Illumina technology. Scaffold N50 was 178,272 bp, meaning that 50% of all nucleotides were found in sequence stretches of at least this length. Supporting Information contains a detailed description of the genome assembly.

Genome Annotation

Gene modeling identified 21,856 genes and 176,936 exons, whereas evidence-based prediction gave 26,043 genes and 514,839 exons (Table 1). The number of genes and exons from ab initio modeling were very similar to the number of genes (21,877) and exons (187,198) inferred from a lift over of gene annotations from the genome of the closely related ferret. About half (51%) of 248 ultraconserved genes found in most eukaryotes (CEGs, Parra et al. 2007) were completely covered in the assembly, and 90% of these genes were at least partially sequenced.

A large proportion of vertebrate genomes consist of repeat elements. The repeat content of the wolverine draft genome assembly was 829 Mbp, or 34% of the total assembly length (Table 1). The most common class of repeat element was long interspersed nuclear elements (LINEs), especially LINE1 (432 Mbp or 18% of the assembly). Repeat regions were masked in all subsequent analyses. All but 1 of 19 previously published microsatellite regions from the wolverine or closely related species (see Supporting Information) could be fully or partially recovered in the genome assembly.

Genomic Variability

Whole-genome resequencing of 10 individuals identified 1,473,629 polymorphic sites (1,305,461 SNPs, and

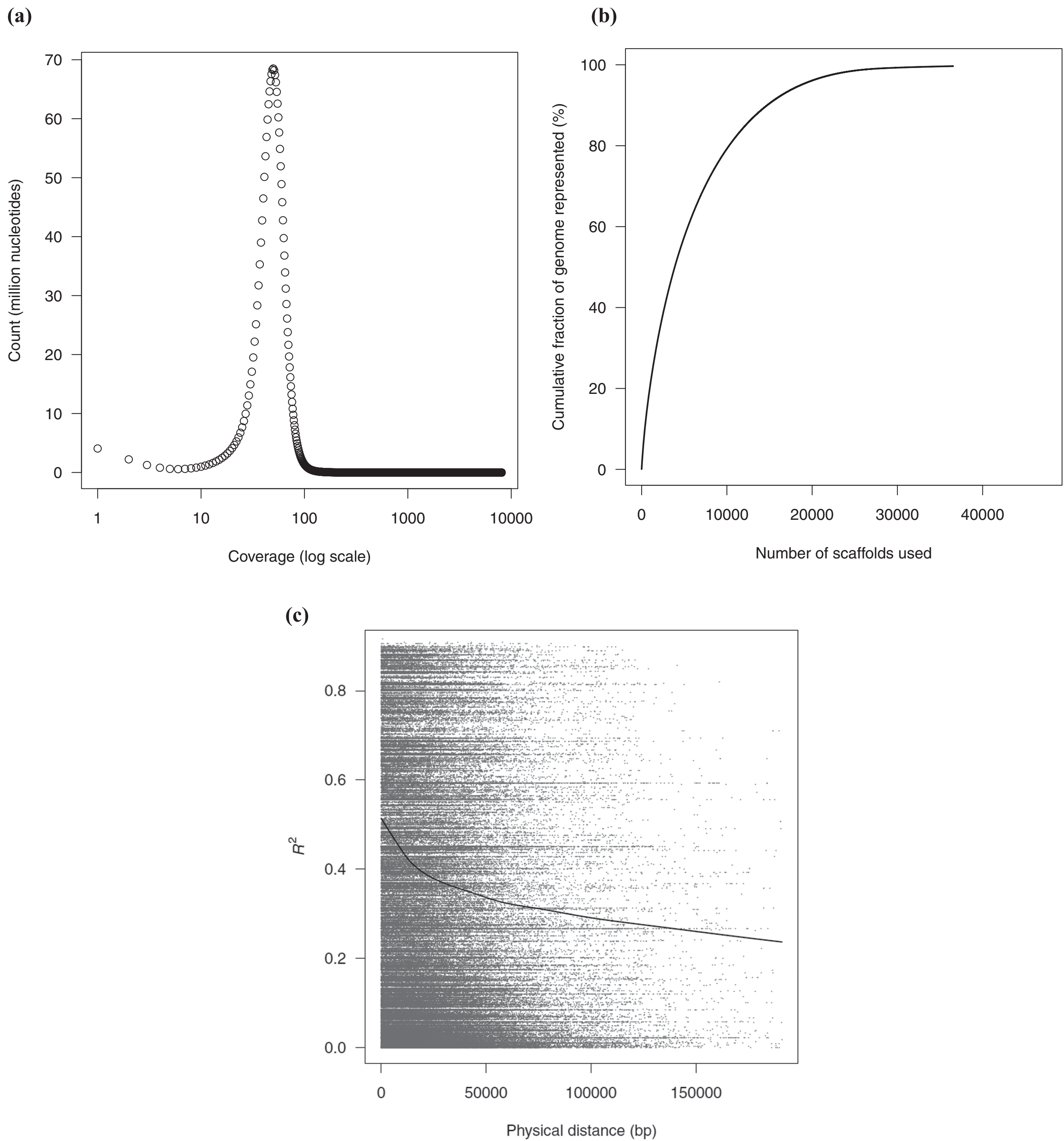

Figure 1. Properties of the wolverine draft genome assembly: (a) distribution of read coverage for all nucleotide positions (distribution peak is at a coverage of 50 reads per site; 98% of the genome [2.2 Gbp] had a read coverage between 20 × and 200 ×), (b) cumulative fraction represented in a given number of genomic scaffolds by size starting with the largest, and (c) pairwise levels of linkage disequilibrium (R^2) between markers relative to physical distance (data from all scaffolds >1 Mbp analyzed in windows with <26 markers; thick black line, nonparametric smoothing curve [lowess]).

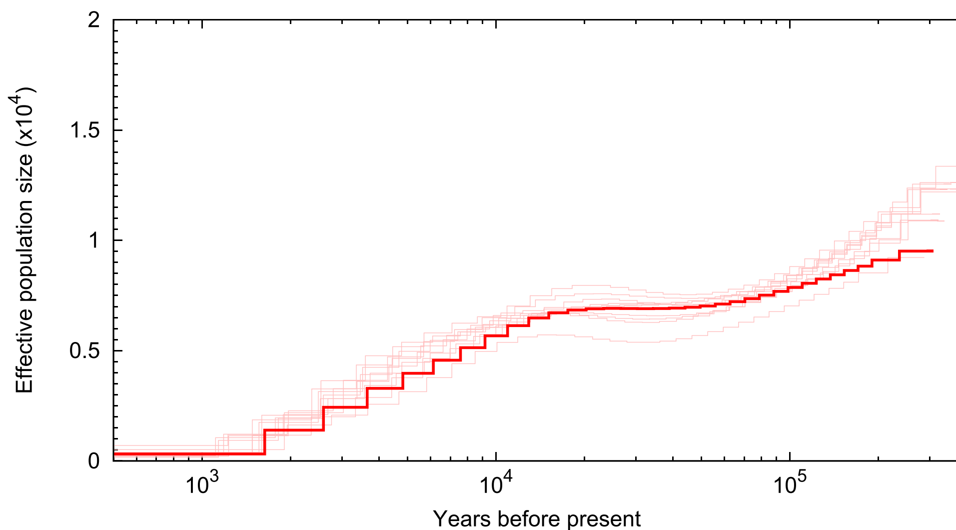

Figure 2. Pairwise sequentially Markovian coalescent plot of long-term demographic history of the Scandinavian wolverine population (bold line, reference individual; thin lines, 10 resequenced individuals). See Supporting Information for results from X-chromosome and autosomal data separated.

168,168 indels). This number of variable sites is very low for a large vertebrate genome, given a sample size of some 10 individuals, and was reflected in an estimated genome-wide nucleotide diversity (π) as low as 0.00051. There was significant variation in individual heterozygosity (h), ranging from 0.00024 to 0.00032 (mean = 0.00028). The discrepancy between the estimates of π and h translates into a strong signal of inbreeding (mean F_{IS} of 0.089). Nucleotide diversity in genic regions was lower than the genome-average, with $\pi = 0.00037$ for coding sequence, 0.00032 for 3′-untranslated regions (UTR) and 0.00031 for 5′UTR (Supporting Information). As expected from a population with a small N_e, the extent of LD was high. The r^2 decreased to 0.3 at a distance of about 50,000 bp and to 0.2 well above 100,000 bp (Fig. 1c).

Although the draft genome assembly did not contain sequences from the Y chromosome given that a female was sequenced, resequencing of males followed by assembly of reads not mapping to the reference genome (Supporting Information) allowed the identification of 12 putative Y-chromosome scaffolds, comprising 44,526 bp (Table 1). No variable positions were identified in these regions among the 10 resequenced males. Neither did we find any sequence variation in mtDNA across the 11 sequenced individuals (Ekblom et al. 2014).

Estimates of the Effective Population Size

Long-term N_e was estimated to 13,000 individuals. In sharp contrast, current N_e was estimated to only 87 individuals based on the extent of LD, an estimate that may be biased due to genetic structure and inbreeding in the population. We observed a long-term decline in N_e over the last 300,000 years, from approximately 10,000 to < 500 individuals (Fig. 2). Genome-wide Tajima's D was close to 0 (mean = 0.037, median = −0.094), which implies no strong genomic signal of a recent population bottleneck.

Population Structure

We found a signal of population structure with a genome-wide mean F_{ST} of 0.04 among the 4 resequenced samples from southwestern Norway and the 7 samples from Sweden and northern Norway. To study genetic diversity and structure in a larger set of samples, we selected 384 presumably unlinked SNPs (see Methods and Supporting Information for details) in the genome for genotyping.

MAFs for the SNP markers ranged between 0.015 and 0.50 with a mean of 0.38. Mean observed heterozygosity (0.44, range 0.03–0.61) was very similar to expected heterozygosity (0.45, range 0.03–0.50), resulting in F_{IS} estimates very close to zero (ranging between −0.015 and 0.016; Supporting Information). There was thus a significant discrepancy between estimates of inbreeding levels based on whole-genome data compared to estimates based on data from a selected set of highly variable genetic markers.

The SNP genotyping data also showed evidence for genetic substructure within the Scandinavian wolverine population with a mean F_{ST} value of 0.047 across 4 arbitrarily defined geographical regions (southwestern Norway, southern, middle, and northern Scandinavia). Samples from southwestern Norway was especially distinct from the other geographical regions (F_{ST} value range 0.049–0.074) (Fig. 3 & Supporting Information), suggesting there were 2 subpopulations within Scandinavia. There were also weak but significant signals of isolation-by-distance between individuals, both in the entire data (Mantel test, $r_S = 0.020$, $p < 0.001$) and within the geographic regions (southwestern Norway, $r_S = 0.032$, $p < 0.001$; all other regions, $r_S = 0.011$, $p < 0.001$). Spatial analyses of genetic clustering using STRUCTURE (Fig. 3c) and BAPS (Supporting Information) further corroborated that the samples clustered into 2 genetically distinct subpopulations, 1 mainly composed of samples from southwestern Norway and 1 by all other samples.

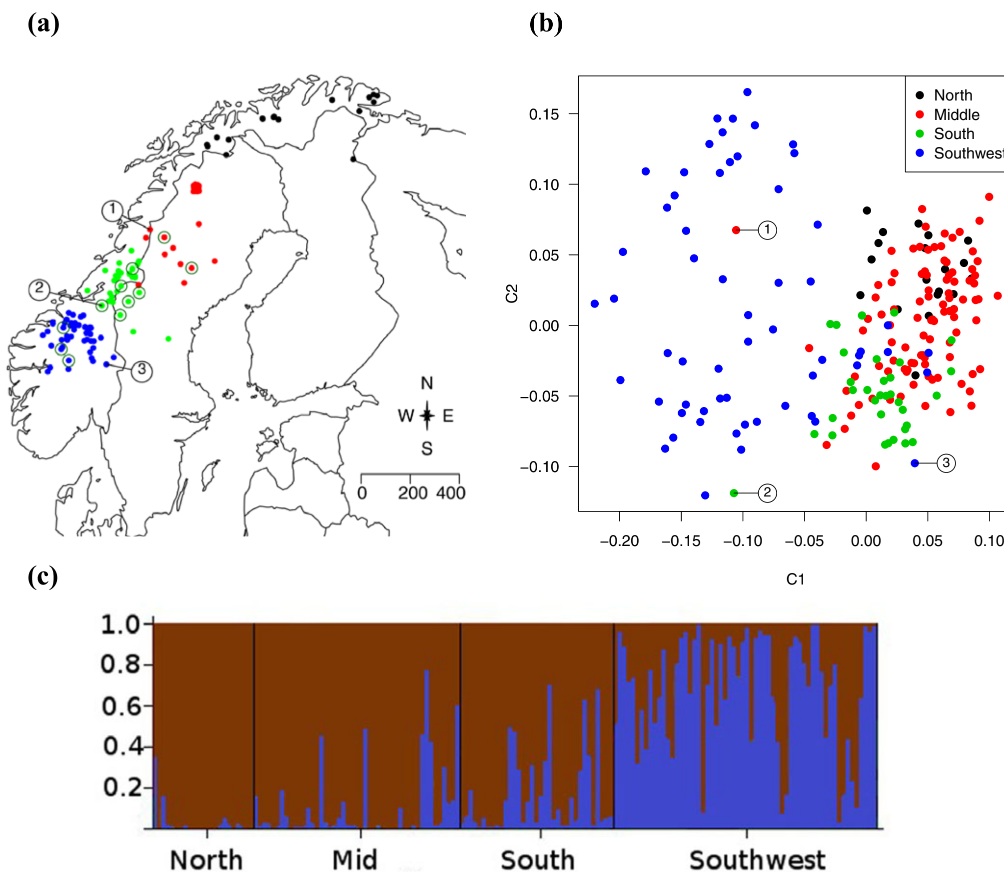

Figure 3. (a) Geographic positions for all wolverine samples included in the population genetic study (n = 234, mainly tissue samples collected from 1993 to 2011) (encircled points, samples used for genome sequencing and resequencing; other points, samples used for single nucleotide polymorphism [SNP] genotyping only; sampling regions: black, north; red, middle; green, southern Scandinavia; blue, southwestern Norway). (b) Multidimensional scaling plot showing genetic similarities among samples (samples 1–3, individuals for which sampling location does not match the genetically defined geographical area to which they cluster). (c) Population structure inferred by Bayesian clustering in STRUCTURE with K of 2 (color of vertical lines, genetic composition of 1 individual; black lines, separate different sampling regions). See Supporting Information for plots with other K levels.

Some samples (denoted 1–3 in Fig. 3) clustered with a different population than their location of sampling. This suggests recent dispersal events, either by the sampled individuals themselves or by their close ancestors. Indeed, a striking case was an individual (1 in Fig. 3) that was born and first sampled as a juvenile in southwestern Norway and subsequently sampled 500 km to the north.

Development of SNP-Array for Routine Genotyping of Scat Samples

The Illumina GoldenGate SNP genotyping assay had very low success rate for DNA prepared from scat samples (91.6% sample dropout rate; 163 out of 178 samples [Supporting Information]). We therefore evaluated a PCR-based genotyping approach (Fluidigm) by selecting 96 unlinked and reliable markers from the original 384 SNP panel (details in Supporting Information). Aided by a multiplexed pre-amplification step (STA), 95 of 96 SNP markers gave reliable genotypes for 124 of the 164 scat samples tested (sample dropout rate = 24.4%). For samples run in duplicate the mean marker dropout rate was 8.6 % and mean genotyping error rate was 0.4%. The error rate was highly skewed as most samples had no genotyping errors while a few had as much as 3.3% errors (Supporting Information). The rate of marker dropout and genotyping error among samples were highly correlated ($r = 0.46$, $df = 60$, $p < 0.001$). In summary, the PCR-based approach performed significantly better than the GoldenGate assay.

All individuals sampled (individual assignment had already been conducted based on microsatellite genotype profiles) had unique multilocus genotypes, with similarities between pairs of samples originating from different individuals having a mean genetic similarity of 0.64 (range 0.48–0.91). Thus, all samples (also those from closely related individuals, see below) could be reliably assigned to

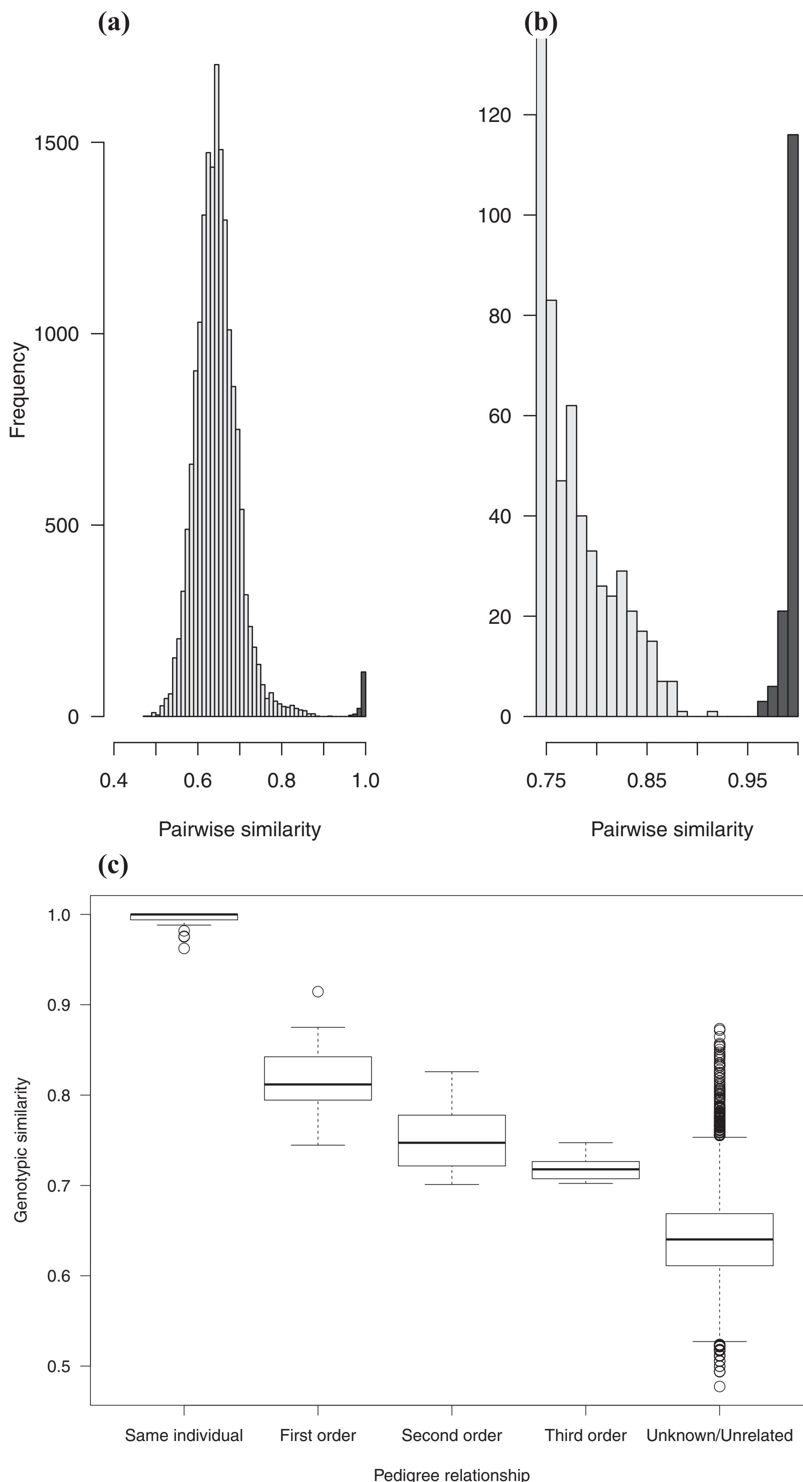

Figure 4. Genotype similarities among wolverine samples obtained from single nucleotide polymorphism (SNP) genotyping with a 96 SNP array: (a) distribution of pairwise multilocus SNP genotype similarity from pairs of samples from different individuals (light grey) and from different samples from the same individual (dark grey), (b) zoom-in of genetic similarities >0.75, (c) levels of pairwise genotype similarity in relation to pedigree relatedness of samples (same individual, multiple samples genotyped from the same individual; first order, parent-offspring and full-sibling relationships; second order, half-sibling and grand-parent-offspring relationships; third order, more distant relationships: whiskers, 95% CI; bars, upper and lower quartiles; horizontal lines within bars, median; circles, extreme values).

an individual based on their SNP genotype, indicating that this marker set would be useful for genetic monitoring. There were some cases of multiple samples collected from the same individual included in the assay. Because of rare genotyping errors, a few such pairs of samples had a genetic similarity of <1, but it was always >0.96 (mean = 0.99) (Fig. 4).

Pedigree relationships had been previously inferred for 19 of the genotyped individuals based on ecological data and multilocus microsatellite genotypes (Hedmark & Ellegren 2007). Mean genotypic similarity between first-order (0.82) and second-order (0.75) relatives were higher than that of unrelated individuals (Fig. 4) and of individuals with unknown relatedness. The 2 most genetically similar

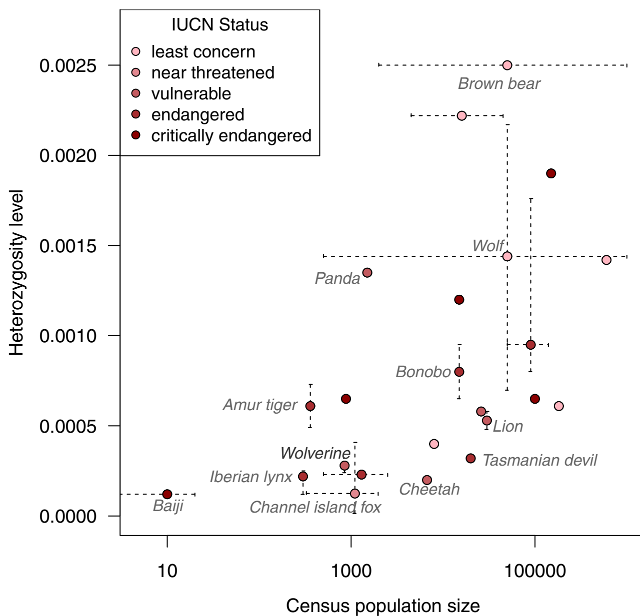

Figure 5. Genome-wide estimates of heterozygosity in relation to population census size and International Union for Conservation of Nature (IUCN) conservation status in a range of mammals (vertical bars, range of individual heterozygosity; horizontal bars, range of census size for sampled populations of the same species or uncertainties in the population size estimates). For clarity some species names are omitted from the plot, these are (from top to bottom) bighorn sheep, western gorilla, bottle nose dolphin, orangutan, chimpanzee, eastern gorilla, minke whale, polar bear, Eurasian lynx, and snow leopard. For details and references see Supporting Information.

individuals (0.91) across the whole SNP data set (and the only pair with genetic similarity >90%) were a known mother-offspring pair.

Discussion

This study provides a draft sequence of the wolverine genome and offers insight on genome-wide levels of genetic diversity from resequencing of a Scandinavian population sample. Nucleotide diversity was as low as 0.00051 (i.e., approximately 1 variable position per 2,000 bp on average). This is considerably lower than that found in other carnivore populations such as wolf, bear, panda, lion, and amur tiger (Cho et al. 2013; Fan et al. 2016; Hailer et al. 2012) (Fig. 5 & Supporting Information). Our genome-wide estimate of genetic diversity is in fact among the lowest recorded for natural mammalian populations and in the same range as some species of severe conservation genetic concern including snow leopard (*Panthera uncia*), Iberian lynx (*Lynx pardinus*), cheetah (*Acinonyx jubatus*), and Tasmanian devil (*Sarcophilus harrisii*) (Cho et al. 2013; Dobrynin et al. 2015; Abascal et al. 2016; Hendricks et al. 2017). However, comparisons of levels of genetic diversity among species have to be treated with some caution because reported estimates of nucleotide diversity are sensitive to the exact bioinformatics pipeline used for variant detection, especially regarding the levels of quality filtering applied. Moreover, as we demonstrate herein, inbreeding and population structure can make estimates of individual heterozygosity (which is used as a measure of nucleotide diversity if only a single individual has been sequenced, as in some of the studies cited above) significantly lower than population-level nucleotide diversity (see below).

Low genetic diversity observed in Scandinavian wolverines based on microsatellite markers may be a consequence of recent bottlenecks and founder events (Flagstad et al. 2004). However, our demographic analyses suggest that N_e has been decreasing since the onset of the last glaciation. The decrease seems to have been particularly pronounced during the last 10,000 years with a drop in N_e from several thousand to a current estimate of <500 individuals. This result highlights the importance of interpreting current levels of genetic variation in a vulnerable species in the context of its long-term population dynamics. However, the results from such demographic analyses need to be interpreted with caution because they are sensitive to underlying assumptions, for example concerning effects of genetic population structure. The decrease in N_e of wolverines must also be put in the perspective of range dynamics. Scandinavia was covered by ice during much of the last glaciation, and refugial wolverine populations in central Europe may have been genetically fragmented with only limited gene flow with the main Eurasian population. Genetic variation may also have been reduced through edge processes during the colonization phase (Zigouris et al. 2013). Future studies need to establish the genetic relationship between contemporary populations throughout the wolverine distribution to understand the effects of such dynamics. This will also be important from a practical conservation perspective, for example if considering the possibility of genetic rescue actions and genetic supplementation (Ingvarsson 2001; Gompert 2012).

Individual heterozygosities (0.00024–0.00032) were considerably lower than nucleotide diversity in the population (0.00051), indicative of inbreeding with a mean inbreeding coefficient of 8.9%. An alternative explanation would be population substructuring (i.e., a Wahlund effect) (Peter et al. 2010). As expected from a small and inbred population, high levels of LD extended over long distances in the wolverine genome. Signatures of inbreeding coupled with the observation of population structure suggest that there are low levels of realized gene flow into Scandinavia from adjacent populations, despite the high dispersal capacity of wolverines (Flagstad et al.

2004). This could have both short- and long-term conservation consequences (Charlesworth & Charlesworth 1987). A long-term decline in N_e may have resulted in an effective purging of deleterious recessives (Fowler & Whitlock 1999), leaving the population less exposed to inbreeding depression. However, loss of genetic diversity may leave the population vulnerable to novel environmental conditions, for example as a result of climate change (England et al. 2003; Brodersen & Seehausen 2014). Facilitation of gene flow from eastern populations may thus be an important goal for long-term conservation of the Scandinavian wolverines.

An interesting methodological, and from a practical perspective important, aspect was that although we found clear signatures of inbreeding when analyzing whole-genome resequencing data, these signals were not present when using data from a subsample of 357 highly variable SNP markers. A likely explanation for this discrepancy is that markers selected for a high MAF will provide a biased picture of the character of individual genetic diversity by overestimating the proportion of heterozygous sites compared to a random draw from sites segregating in the population. It should thus not be the lower number of markers in the genotyping panel per se—which actually could be considered not so low—that generates the discrepancy. A similar bias may apply to the use of microsatellite markers because they are generally hypervariable. An important conclusion from these observations is therefore that inferences made regarding the genetic status of endangered species based only on data from SNP or microsatellite marker sets selected for high variability may need to be treated with caution.

There were close to 1.5 million variable sites in the wolverine genome identified from sequencing of 11 Scandinavian individuals. Several hundred of these variants were verified in a larger set of samples from the same population. A subset of these markers (96) will be applied in a genetic monitoring program aimed to differentiate between individuals based on noninvasive sampling and for estimating local population densities. We found that the panel of SNP markers can be used successfully to discriminate between different individuals based on their multilocus genotype profile. Multiplex pre-amplification together with an integrated microfluidic circuit chip technology led to improved genotyping success of noninvasively collected samples. Compared with a microsatellite genotyping scheme that is currently used (Hedmark et al. 2004), application of SNP methodology will provide improved genotyping quality, a more efficient work flow (both in terms of wet-lab procedures and data handling), and may also enable additional applications such as kinship analyses (Morin et al. 2004; Helyar et al. 2011). A similar approach has recently been successfully used for genotyping of low-quality samples in wildcats (Nussberger et al. 2014), wolves (Kraus et al. 2015), and bears (Norman & Spong 2015).

In summary, we obtained and annotated a draft sequence of the wolverine genome. Based on this genetic resource we determined genome-wide levels of genetic diversity in the vulnerable Scandinavian wolverine population. Diversity was among the lowest recorded for a natural mammal population. Signatures of inbreeding were seen using the full genomic data but not with an SNP marker panel selected based on a high degree of informativeness. However, this panel proved useful for genetic monitoring based on non-invasive samples, illustrating how genomic approaches can be directly applied in conservation of endangered populations, thus bridging the conservation genomics gap (McMahon et al. 2014; Shafer et al. 2015). The cost and effort involved in establishing genomic resources for a nonmodel organism is still considerable. For many systems there will be a trade-off between investing in genomic work versus applied conservation actions. However, the continuous reduction in costs for sequencing and the possibility of using low-budget approaches such as genotyping by sequencing may shift the trade-off toward an increasing use of genomic approaches.

Acknowledgments

We thank D. Scofield, P. Olason, L. Dutoit, K. Nadachowska, and R. Burri for bioinformatic assistance and 3 anonymous reviewers for constructive comments on an earlier version of the manuscript. Annotation was performed by M. Höppner at BILS (Bioinformatics Infrastructure for Life Sciences), a distributed national research infrastructure supported by the Swedish Research Council (VR). Sequencing and SNP genotyping was performed by the SNP&SEQ Technology Platform (www.sequencing.se), Science for Life Laboratory at Uppsala University, a national infrastructure supported by the Swedish Research Council (VR-RFI) and the Knut and Alice Wallenberg Foundation. H. Vretman, at the Clinical Biomarkers Facility on Uppsala University, and L. Lindbom, at Olink Bioscience, performed Fluidigm SNP genotyping of scat samples. Tissue samples for genome sequencing were obtained from SVA (the National Veterinary Institute). High-performance computing was conducted using the Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX). This project was funded by the Swedish Nature Protection Agency (Agreements Dnr 235-12-11 and Dnr 11/110) and the Norwegian Directorate for Nature Management.

Supporting Information

Detailed methods and supplementary results (Appendix S1) and information about the validated SNPs (including flanking sequences) (Appendix S2) are available

online. The authors are solely responsible for the content and functionality of these materials. Queries (other than absence of the material) should be directed to the corresponding author. All raw sequencing reads produced for this project are available at the European Nucleotide Archive under study accession number PRJEB10674. The annotated genome assembly is available at the European Nucleotide Archive under study accession number PRJEB10674 and with scaffold accession numbers CYRY01000001–CYRY01047417. The complete mitochondrial genome sequence is available at NCBI GenBank with accession number KF415127. The assembly of Y-chromosome sequences is available at the European Nucleotide Archive with scaffold accession numbers LN877956–LN877967.

Literature Cited

Abascal F, et al. 2016. Extreme genomic erosion after recurrent demographic bottlenecks in the highly endangered Iberian lynx. Genome Biology **17:**251.

Aronsson M, Persson J. 2017. Mismatch between goals and the scale of actions constrains adaptive carnivore management: the case of the wolverine in Sweden. Animal Conservation **20:**261–269.

Barea-Azcón J, Virgós E, Ballesteros-Duperón E, Moleón M, Chirosa M. 2007. Surveying carnivores at large spatial scales: a comparison of four broad-applied methods. Biodiversity and Conservation **16:**1213–1230.

Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics **27:**578–579.

Brodersen J, Seehausen O. 2014. Why evolutionary biologists should get seriously involved in ecological monitoring and applied biodiversity assessment programs. Evolutionary Applications **7:**968–983.

Brøseth H, Flagstad Ø, Wärdig C, Johansson M, Ellegren H. 2010. Large-scale noninvasive genetic monitoring of wolverines using scats reveals density dependent adult survival. Biological Conservation **143:**113–120.

Cahill JA, et al. 2013. Genomic evidence for island population conversion resolves conflicting theories of polar bear evolution. PLOS Genetics **9:**e1003345.

Charlesworth D, Charlesworth B. 1987. Inbreeding depression and its evolutionary consequences. Annual Review of Ecology and Systematics **18:**237–268.

Cho YS, et al. 2013. The tiger genome and comparative analysis with lion and snow leopard genomes. Nature Communications **4:**2433.

Corander J, Waldmann P, Sillanpää MJ. 2003. Bayesian analysis of genetic differentiation between populations. Genetics **163:**367–374.

Danecek P, et al. 2011. The variant call format and VCFtools. Bioinformatics **27:**2156–2158.

Do C, Waples RS, Peel D, Macbeth GM, Tillett BJ, Ovenden JR. 2014. NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. Molecular Ecology Resources **14:**209–214.

Dobrynin P, et al. 2015. Genomic legacy of the African cheetah, *Acinonyx jubatus*. Genome Biology **16:**277.

Ekblom R, Smeds L, Ellegren H. 2014. Patterns of sequencing coverage bias revealed by ultra-deep sequencing of vertebrate mitochondria. BMC Genomics **15:**467.

Eklund A, Mattisson J, Höglund L, Tovmo M. 2017. Inventering av järv 2017. Page 31. Beståndsstatus för stora rovdjur i Skandinavien. Viltskadecenter, SLU och Rovdata, Grimsö och Trondheim.

Ellegren H. 2014. Genome sequencing and population genomics in non-model organisms. Trends in Ecology & Evolution **29:** 51–63.

England PR, Osler GHR, Woodworth LM, Montgomery ME, Briscoe DA, Frankham R. 2003. Effects of intense versus diffuse population bottlenecks on microsatellite genetic diversity and evolutionary potential. Conservation Genetics **4:**595–604.

Falush D, Stephens M, Pritchard JK. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics **164:**1567–1587.

Fan Z, et al. 2016. Worldwide patterns of genomic variation and admixture in gray wolves. Genome Research **26:**163–173.

Flagstad Ø, Hedmark E, Landa A, Brøseth H, Persson J, Andersen R, Segerström P, Ellegren H. 2004. Colonization history and noninvasive monitoring of a reestablished wolverine population. Conservation Biology **18:**676–688.

Fowler K, Whitlock MC. 1999. The variance in inbreeding depression and the recovery of fitness in bottlenecked populations. Proceedings of the Royal Society B: Biological Sciences **266:** 2061–2066.

Frankham R. 2010. Where are we in conservation genetics and where do we need to go? Conservation Genetics **11:**661–663.

Garvin MR, Saitoh K, Gharrett AJ. 2010. Application of single nucleotide polymorphisms to non-model species: a technical review. Molecular Ecology Resources **10:**915–934.

Gervasi V, Brøseth H, Nilsen EB, Ellegren H, Flagstad Ø, Linnell JDC. 2015. Compensatory immigration counteracts contrasting conservation strategies of wolverines (*Gulo gulo*) within Scandinavia. Biological Conservation **191:**632–639.

Gompert Z. 2012. Population genomics as a new tool for wildlife management. Molecular Ecology **21:**1542–1544.

Haglund B. 1965. Järv och varg. P.A. Norstedt & Söners Förlag, Stockholm.

Hailer F, Kutschera VE, Hallström BM, Klassert D, Fain SR, Leonard JA, Arnason U, Janke A. 2012. Nuclear genomic sequences reveal that polar bears are an old and distinct bear lineage. Science **336:**344–347.

Hedmark E, Ellegren H. 2007. DNA-based monitoring of two newly founded Scandinavian wolverine populations. Conservation Genetics **8:**843–852.

Hedmark E, Flagstad Ø, Segerström P, Persson J, Landa A, Ellegren H. 2004. DNA-based individual and sex identification from wolverine (*Gulo Gulo*) faeces and urine. Conservation Genetics **5:**405–410.

Helyar SJ, et al. 2011. Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. Molecular Ecology Resources **11:**123–136.

Hendricks S, Epstein B, Schönfeld B, Wiench C, Hamede R, Jones M, Storfer A, Hohenlohe P. 2017. Conservation implications of limited genetic diversity and population structure in Tasmanian devils (*Sarcophilus harrisii*). Conservation Genetics 1–6.

Höglund J. 2009. Evolutionary conservation genetics. Oxford University Press, Oxford.

Ingvarsson PK. 2001. Restoration of genetic variation lost – the genetic rescue hypothesis. Trends in Ecology & Evolution **16:**62–63.

Korneliussen T, Albrechtsen A, Nielsen R. 2014. ANGSD: analysis of next generation sequencing data. BMC Bioinformatics **15:** 356.

Kraus RHS, vonHoldt B, Cocchiararo B, Harms V, Bayerl H, Kühn R, Förster DW, Fickel J, Roos C, Nowak C. 2015. A single-nucleotide polymorphism-based approach for rapid and cost-effective genetic wolf monitoring in Europe based on noninvasively collected samples. Molecular Ecology Resources **15:**295–305.

Landa A, Skogland T. 1995. The relationship between population density and body size of wolverines *Gulo gulo* in Scandinavia. Wildlife Biology **1:**165–175.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics **25:**1754–1760.

Li H, Durbin R. 2011. Inference of human population history from individual whole-genome sequences. Nature **475:**493–496.

Li R, et al. 2010. The sequence and de novo assembly of the giant panda genome. Nature **463:**311–317.

Liu S, et al. 2014. Population genomics reveal recent speciation and rapid evolutionary adaptation in polar bears. Cell **157:**785–794.

Luo R, et al. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. GigaScience **1:**18.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet Journal **17**:10–12.

McKenna A, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research **20:**1297–1303.

McMahon BJ, Teeling EC, Höglund J. 2014. How and why should we implement genomics into conservation? Evolutionary Applications **7:**999–1007.

Miller W, et al. 2012. Polar and brown bear genomes reveal ancient admixture and demographic footprints of past climate change. Proceedings of the National Academy of Sciences of the USA **109:**E2382–E2390.

Morin PA, Luikart G, Wayne RK, the SNP workshop group. 2004. SNPs in ecology, evolution and conservation. Trends in Ecology & Evolution **19:**208–216.

Nilsson T. 2013. Population viability analyses of the Scandinavian populations of bear (*Ursus arctos*), lynx (*Lynx lynx*) and wolverine (*Gulo gulo*). Swedish Environmental Protection Agency [Naturvårdsverket].

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

Nussberger B, Wandeler P, Camenisch G. 2014. A SNP chip to detect introgression in wildcats allows accurate genotyping of single hairs. European Journal of Wildlife Research **60:**405–410.

Parra G, Bradnam K, Korf I. 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics **23:**1061–1067.

Peng X, et al. 2014. The draft genome sequence of the ferret (*Mustela putorius furo*) facilitates study of human respiratory disease. Nature Biotechnology **32:**1250–1255.

Persson J, Rauset GR, Chapron G. 2015. Paying for an endangered predator leads to population recovery. Conservation Letters **8:**345–350.

Peter BM, Wegmann D, Excoffier L. 2010. Distinguishing between population bottleneck and population subdivision by a Bayesian model choice procedure. Molecular Ecology **19:**4648–4660.

Prufer K, et al. 2012. The bonobo genome compared with the chimpanzee and human genomes. Nature **486:**527–531.

Purcell S, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. American Journal of Human Genetics **81:**559–575.

Raymond M, Rousset F. 1995. Genepop (Version 3.1d): population genetics software for exact tests and ecumenicism. Journal of Heredity **86:**248–249.

Schwartz MK, Luikart G, Waples RS. 2007. Genetic monitoring as a promising tool for conservation and management. Trends in Ecology & Evolution **22:**25–33.

Shafer ABA, et al. 2015. Genomics and the challenging translation into conservation practice. Trends in Ecology & Evolution **30:** 78–87.

Smeds L, Künstner A. 2011. ConDeTri - a content dependent read trimmer for Illumina Data. PLoS ONE **6**: e26314 https://doi.org/10.1371/journal.pone.0026314.

Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. 2006. AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Research **34:**W435–W439.

Stronen AV, et al. 2013. North-south differentiation and a region of high diversity in European wolves (*Canis lupus*). PLoS ONE **8:** e76454. https://doi.org/10.1371/journal.pone.0076454.

Walker CW, Vilà C, Landa A, Lindén M, Ellegren H. 2001. Genetic variation and population structure in Scandinavian wolverine (*Gulo gulo*) populations. Molecular Ecology **10:**53–63.

Wang, J. 2016. A comparison of single-sample estimators of effective population sizes from genetic marker data. Molecular Ecology **25:**4692–4711.

Väli Ü, Einarsson A, Waits L, Ellegren H. 2008. To what extent do microsatellite markers reflect genome-wide genetic diversity in natural populations? Molecular Ecology **17:**3808–3817.

Zamani N, Sundstrom G, Meadows J, Hoppner M, Dainat J, Lantz H, Haas B, Grabherr M. 2014. A universal genomic coordinate translator for comparative genomics. BMC Bioinformatics **15:**227.

Zigouris J, Schaefer JA, Fortin C, Kyle CJ. 2013. Phylogeography and post-glacial recolonization in wolverines (Gulo gulo) from across their circumpolar distribution. PLoS ONE **8:**e83837. https://doi.org/10.1371/journal.pone.0083837.