

RESOURCE ARTICLE

Targeted resequencing of coding DNA sequences for SNP discovery in nonmodel species

Daniel W. Förster¹  | James K. Bull¹ | Dorina Lenz¹ | Marijke Autenrieth^{1,2} | Johanna L. A. Paijmans² | Robert H. S. Kraus^{3,4} | Carsten Nowak⁵ | Helmut Bayerl⁶ | Ralph Kuehn^{6,7}  | Alexander P. Saveljev⁸ | Magda Sindičić⁹ | Michael Hofreiter² | Krzysztof Schmidt¹⁰ | Jörns Fickel^{1,2}

¹Department of Evolutionary Genetics, Leibniz Institute for Zoo and Wildlife Research, Berlin, Germany

²Institute for Biochemistry and Biology, University of Potsdam, Potsdam, Germany

³Department of Biology, University of Konstanz, Konstanz, Germany

⁴Department of Migration and Immuno-Ecology, Max Planck Institute for Ornithology, Radolfzell, Germany

⁵Conservation Genetics Group, Senckenberg Research Institute and Natural History Museum Frankfurt, Gelnhausen, Germany

⁶Unit of Molecular Zoology, Chair of Zoology, Department of Animal Science, Technical University of Munich, Freising, Germany

⁷Department of Fish, Wildlife and Conservation Ecology, New Mexico State University, Las Cruces, New Mexico

⁸Department of Animal Ecology, Russian Research Institute of Game Management and Fur Farming, Kirov, Russia

⁹Faculty of Veterinary Medicine, University of Zagreb, Zagreb, Croatia

¹⁰Mammal Research Institute, Polish Academy of Sciences, Białowieża, Poland

Correspondence

Daniel W. Förster, Department of Evolutionary Genetics, Leibniz Institute for Zoo and Wildlife Research, Alfred-Kowalke-Str. 17, D-10315 Berlin, Germany.
Email: dwgfoerster@gmail.com

Funding information

Leibniz-Gemeinschaft, Grant/Award Number: SAW-2011-SGN-3; Deutscher Akademischer Austauschdienst, Grant/Award Number: A/11/93927; Narodowe Centrum Nauki, Grant/Award Number: 2014/15/B/NZ8/00212

Abstract

Targeted capture coupled with high-throughput sequencing can be used to gain information about nuclear sequence variation at hundreds to thousands of loci. Divergent reference capture makes use of molecular data of one species to enrich target loci in other (related) species. This is particularly valuable for nonmodel organisms, for which often no a priori knowledge exists regarding these loci. Here, we have used targeted capture to obtain data for 809 nuclear coding DNA sequences (CDS) in a nonmodel organism, the Eurasian lynx *Lynx lynx*, using baits designed with the help of the published genome of a related model organism (the domestic cat *Felis catus*). Using this approach, we were able to survey intraspecific variation at hundreds of nuclear loci in *L. lynx* across the species' European range. A large set of biallelic candidate SNPs was then evaluated using a high-throughput SNP genotyping platform (Fluidigm), which we then reduced to a final 96 SNP-panel based on assay performance and reliability; validation was carried out with 100 additional Eurasian lynx samples not included in the SNP discovery phase. The 96 SNP-panel developed from CDS performed very successfully in the identification of individuals and in population genetic structure inference (including the assignment of individuals to their source population). In keeping with recent studies, our results show that genic SNPs can be valuable for genetic monitoring of wildlife species.

KEYWORDS

CDS, conservation genetics, Eurasian lynx, genetic monitoring, hybridization capture, single nucleotide polymorphism

1 | INTRODUCTION

Genetic markers are important tools for assessing and monitoring wildlife species (Frankham, Ballou, & Briscoe, 2010), providing valuable information at both the population (e.g., genetic diversity, census size and migration) and the individual level (e.g., reproductive success, relatedness and home ranges). They serve a particularly important role in the study and conservation of rare and elusive species, as genetic data obtained from noninvasively collected samples (e.g., hair or faeces) provide information that is more difficult to obtain using traditional monitoring (Schwartz, Luikart, & Waples, 2007). In an important way, genetic monitoring can inform researchers and agencies about population parameters that are nearly impossible to retrieve from observation alone.

The past years have seen an ongoing transition from the use of fragment length polymorphisms (e.g., microsatellites) to the use of single nucleotide polymorphisms (SNPs) in molecular studies of wildlife species (Seeb et al., 2011). SNP-based marker panels have several advantages compared with microsatellites (e.g., Brumfield, Beerli, Nickerson, & Edwards, 2003; Garvin, Saitoh, & Gharrett, 2010; Morin, Luikart, & Wayne, 2004), including their genome-wide distribution in coding and noncoding regions, the ease of allele scoring, data transferability due to independence of genotyping technology and the potential for high-throughput screening. Provided that a sufficient number of markers are available, SNPs also compare favourably with microsatellites in applications relevant to conservation, for example individual identification, inference of population structure and admixture, assignment of individuals to populations or parentage and kinship assignment (e.g., Gärke et al., 2012; Glover et al., 2010; Hauser, Baird, Hilborn, Seeb, & Seeb, 2011; Kleinman-Ruiz et al., 2017; Muñoz et al., 2017). Also relevant for the application of SNPs in conservation genetics is that SNP assays typically target shorter sequences (50–60 bp) than microsatellites (80–300 bp) (Morin et al., 2004), which results in lower rates of genotyping errors (e.g., allelic dropout and false alleles) when working with degraded samples characterized by fragmented DNA, such as noninvasively collected samples (Bonin et al., 2004; Broquet, Ménard, & Petit, 2007; Kraus et al., 2015; Norman & Spong, 2015).

Technological advancements such as nanofluidics have substantially reduced the required reaction volumes for SNP genotyping, making it possible to simultaneously type numerous SNP loci even from very little template material. This enables the economical use of noninvasively collected samples, as the often very low amount of template DNA extractable from such samples had previously constrained the number of SNP loci that could be genotyped (e.g., Kraus et al., 2015; and references therein). The low per-sample cost (incl. in comparison with microsatellite-based approaches; Kraus et al., 2015) and the high degree of automation possible with nanofluidics allow for cost-effective and high-throughput screening of samples. Thus, routine use of SNP-panels for genetic monitoring of wildlife species based on noninvasively collected samples is now achievable

and will likely become the marker of choice in noninvasive genetic wildlife studies.

The development of SNP-panels for wildlife species without available genomic resources relies on a preceding phase of SNP discovery. As SNP-panels utilized in genetic monitoring of wildlife species require only a limited number of markers—usually in the order of 24–384 SNPs (e.g., Bayerl et al., 2018; Fitak, Naidu, Thompson, & Culver, 2016; Kleinman-Ruiz et al., 2017; Kraus et al., 2015; Norman, Street, & Spong, 2013; Nussberger, Wandeler, & Camenisch, 2014; von Thaden et al., 2017)—high-throughput sequencing approaches that retrieve data for only a subset of the genome can be used because they yield sufficient genome-wide sequence variants. The most popular and widespread techniques used in population genetics and evolutionary biology are restriction site-associated DNA sequencing (RAD-seq; Miller, Dunham, Amores, Cresko, & Johnson, 2007) and variants thereof (e.g., Elshire et al., 2011; Kraus et al., 2011), whole-transcriptome shotgun sequencing (RNA-seq; Wang, Gerstein, & Snyder, 2009) and targeted capture (Gnirke et al., 2009; Olson, 2007). Targeted capture has several advantages over the alternative techniques (reviewed in Jones & Good, 2016), including greater reproducibility among samples, a higher degree of scalability, more consistent coverage of targeted regions and more accurate variant detection (Gnirke et al., 2009; Harvey, Smith, Glenn, Faircloth, & Brumfield, 2016; Ku et al., 2012; Tewhey et al., 2009). In addition, the utility of targeted capture is far less restricted by sample quality and source (Jones & Good, 2016), permitting a wide range of sample types to be exploited in studies (e.g., noninvasively collected material, Perry, Marioni, Melsted, & Gilad, 2010; archival material, Bi et al., 2014; ancient material, Burbano et al., 2010; formalin-fixed material, Ruane & Austin, 2017).

However, unlike RNA-seq and RAD-seq, targeted capture requires prior sequence knowledge of target loci for the development of baits (DNA/RNA molecules) that are used to enrich complementary genomic regions in genetic libraries prior to sequencing. At present, this sequence knowledge is often lacking for nonmodel species. This can be overcome by either generating required data for study species through prior *de novo* sequence assembly or by utilizing genomic resources from other (related) species (reviewed in Jones & Good, 2016). Studies employing the latter approach can be divided among those targeting genomic regions that are conserved among divergent species (e.g., ultra-conserved element sequencing, Faircloth et al., 2012; anchored hybrid enrichment, Lemmon, Emme, & Lemmon, 2012) and those targeting exons or whole exomes (e.g., Bi et al., 2012; Vallender, 2011). Conserved elements are by definition minimally variable and rely on the retrieval of flanking sequences to identify intraspecific variation (Faircloth et al., 2012; Smith, Harvey, Faircloth, Glenn, & Brumfield, 2014), which appears to become problematic for degraded samples (i.e., samples with high levels of DNA fragmentation; e.g., McCormack, Tsai, & Faircloth, 2016; Ruane & Austin, 2017), and are primarily employed in inter-specific studies to address taxonomic questions (e.g., Faircloth, Sorenson, Santini, & Alfaro, 2013; Faircloth et al., 2012; Hawkins et

al., 2016; Leaché et al., 2015; Manthey, Campillo, Burns, & Moyle, 2016; Smith et al., 2014). Exon/exome capture is not restricted in this manner and is thus more amenable for surveying intraspecific variation (e.g., Bi et al., 2012; Hancock-Hanser et al., 2013; Hodges et al., 2007; Jin et al., 2012; Neves, Davis, Barbazuk, & Kirst, 2013; Powell, Amish, Haynes, Luikart, & Latch, 2016; Roffler et al., 2016; Suren et al., 2016; Yi et al., 2010).

In this study, we demonstrate the application of targeted capture of coding DNA sequences ("exon-capture," Bi et al., 2012) for SNP discovery in a nonmodel species and show how the identified SNPs can be utilized for cost-effective genetic monitoring of an elusive carnivore, the Eurasian lynx *Lynx lynx*. We used baits designed from a fully annotated reference genome of a related species, namely the domestic cat (*Felis catus*), to enrich target loci from our study species (lynx). In order to avoid capturing paralogues, yet have loci evenly distributed throughout the lynx genome, we designed cat baits to target single-copy coding DNA sequences (CDS). To minimize ascertainment bias, lynx samples used for SNP discovery covered the European distribution range of the species, including some reintroduced populations. Following SNP discovery, we developed a SNP-panel that can be routinely used with the nanofluidic Dynamic Array Chip technology implemented in the Fluidigm platform (Fluidigm Corp., San Francisco, CA, USA). This platform has been used successfully to genotype SNPs in a range of noninvasively collected material (e.g., hair and faeces; von Thaden et al., 2017), which represent an important resource for the genetic monitoring of rare and elusive species.

Here, we outline how we (a) used the genomic resources available for model organisms to design baits and then enriched target loci in our study species; (b) filtered the intraspecific variation in our study species for candidate SNP loci; (c) evaluated a large set of candidate SNPs in our chosen genotyping platform; and (d) settled on 96 loci for the final SNP-panel (Figure 1a). We further tested how the newly developed SNP-panel performs in the genetic monitoring of our study species using 100 additional samples not included in the SNP discovery. For this, we specifically focused on the ability of a SNP-panel developed from CDS for (a) individual discrimination and (b) analysis of genetic population structure, including the correct assignment of individuals to their source population.

2 | MATERIALS AND METHODS

2.1 | Biological samples

DNA was extracted from tissue (liver or muscle) or blood using the commercial First-DNA all tissue kit (GEN-IAL GmbH, Troisdorf, Germany). Twenty-six Eurasian lynx from four European populations (Von Arx, Breitenmoser-Würsten, Zimmermann, & Breitenmoser, 2004) were used in the targeted capture: *Baltic*—Estonia ($N = 2$), Latvia ($N = 3$), Poland ($N = 2$) and Russia ($N = 4$); *Nordic*—Finland ($N = 3$) and Norway ($N = 3$); *Carpathian*—Romania ($N = 3$) and Slovakia ($N = 2$); and *Dinaric*—Croatia ($N = 2$) and Slovenia ($N = 2$) (Table 1; Figure 1b).

Applicability of the final 96 SNP-panel was then assessed by genotyping an additional 100 lynx samples originating from the same populations (below, but see also Supporting Information Table S1).

The *Dinaric* population was intentionally sampled as a distinct reintroduced population (originating from the Carpathian population), for the assessment of correct individual assignment (below).

2.2 | Bait design

We compared the annotated genomes of the domestic cat (*Felis catus* v6.2), domestic dog (*Canis lupus familiaris* v3.1), horse (*Equus caballus* v2.0), cow (*Bos taurus* v3.1) and pig (*Sus scrofa* v10.2) using EVOLMARKERS (Li, Riethoven, & Naylor, 2012) to identify single-copy protein-coding genes present in all of these taxa. In order to avoid paralogues, candidate target loci were restricted to CDS with less than 40% similarity to intraspecific sequences using a BLAST approach (Li et al., 2012). This restriction ensured that target loci would be unambiguously identifiable. To increase the chance of SNP discovery and facilitate the development of assays, we excluded short exons and thus chose targets with a minimum length of 400 bp, which yielded 1,357 CDS markers. Potential targets were then further filtered by selecting a single exon per autosomal gene, reducing the set to 809 CDS markers (exons). As capture was to be performed on a felid species, we used *F. catus* CDS to design the baits (Figure 1c, see also information summary in Supporting Information Table S2). The custom tailored MYbaits® target enrichment kit (MYcroarray, Ann Arbor, MI, USA) covering all 809 CDS (having a total length of 618,547 bp) finally consisted of 8,922 biotinylated RNA baits (2× tiling, 120 bp length). The bait design is available as a FASTA file on DRYAD (<https://doi.org/10.5061/dryad.3f4jr01>).

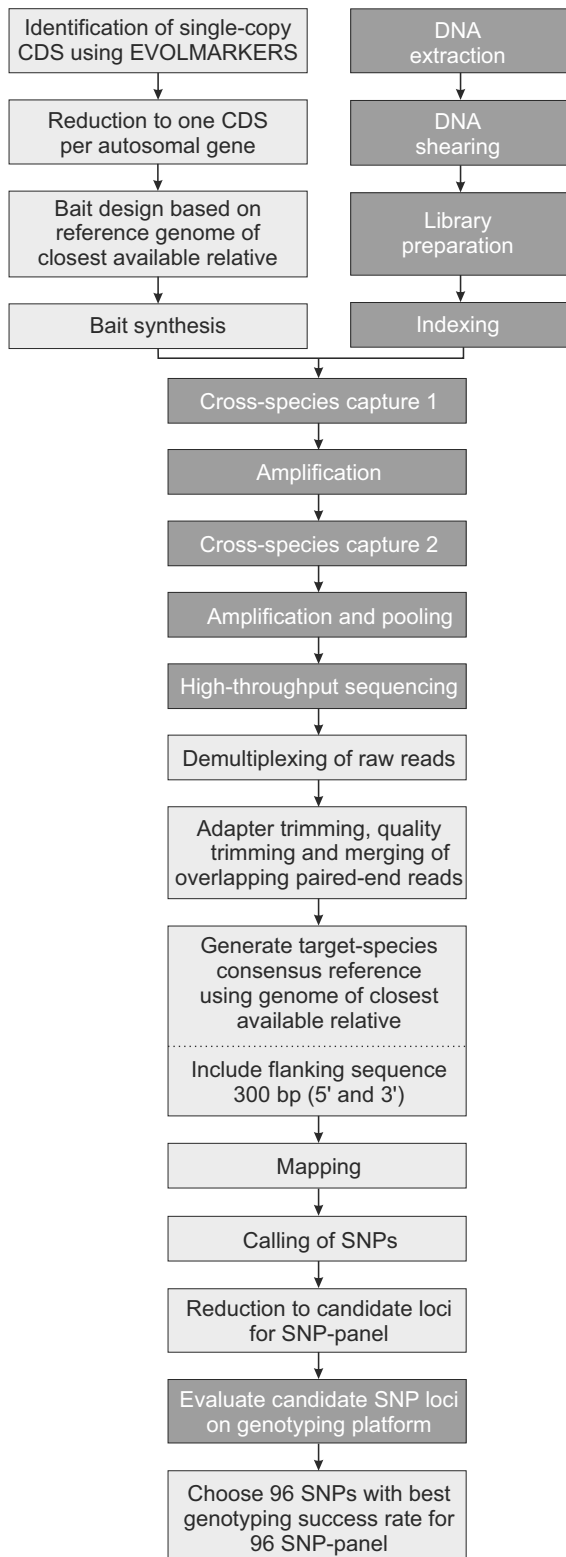
2.3 | Capture and sequencing

Illumina sequencing libraries were built following a published DIY protocol (Meyer & Kircher, 2010) with some modifications reducing both loss of template and costs (Fortes & Pajmans, 2015).

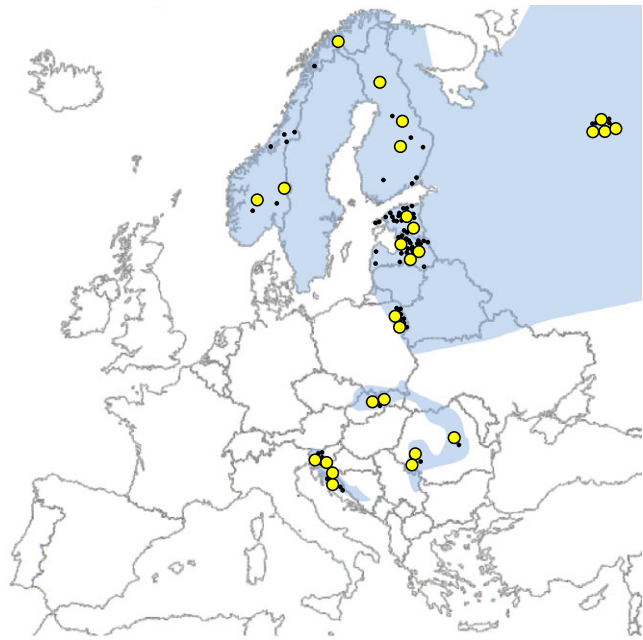
Libraries were captured individually following suggested modifications of the MYbaits protocol (Li, Hofreiter, Straube, Corrigan, & Naylor, 2013). In brief, volumes were reduced, the amount of synthetic RNA baits used per capture was reduced (eightfold), and the standard hybridization temperature of 65°C was replaced by a "touchdown" protocol with hybridization temperature decreasing from 65°C to 50°C in 5°C increments every 11 hr (for details on library building and capture please refer to the Appendix S1).

The library of each individual ($N = 26$) was captured two consecutive times; the eluate from the first capture (i.e., the enriched library) was amplified and used in a second round of capture, as this has been reported to increase the number of on-target reads for both within-species and cross-species capture (Li et al., 2013; Templeton et al., 2013). We confirmed this by sequencing the first five samples processed also after the first capture round, observing a four- to sevenfold increase in the number of unique on-target

(a) Workflow



(b) Sampling



(c) CDS mapped to cat genome

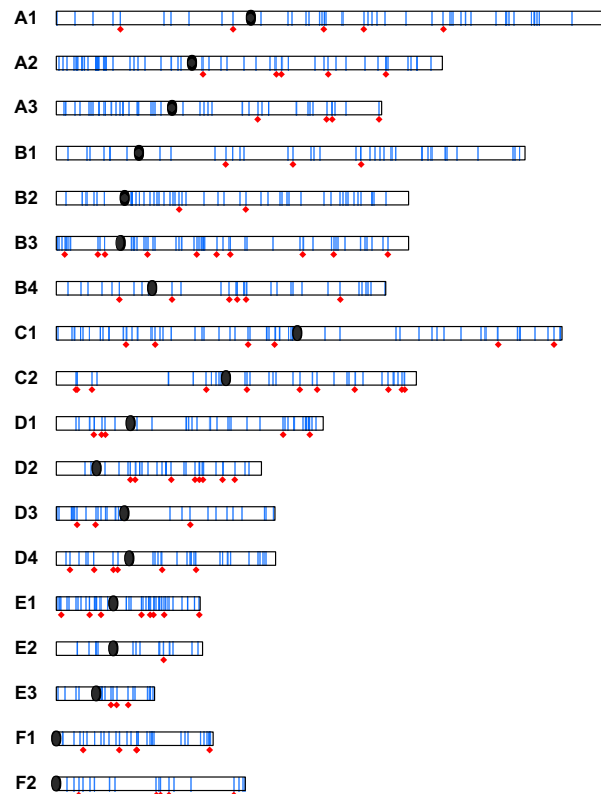


FIGURE 1 Schematic overview of study. (a) Schematic summary of the workflow, presenting both computational steps (light grey boxes) and laboratory steps (dark grey boxes). (b) Sampling localities of Eurasian lynx *Lynx lynx* across Europe. The current distribution of the species is shaded in blue. Large yellow dots represent samples used during the SNP discovery phase; small black dots represent additional samples genotyped using the developed 96 SNP-panel. (c) Schematic representation of the distribution of CDS targeted for enrichment (blue bars), projected onto cat chromosomes. The positions of the 96 SNP loci used for the SNP-panel are indicated by red diamonds. Black ovals show centromeres

TABLE 1 Sampling localities, sequencing results and cross-species capture results of 26 Eurasian lynx *Lynx lynx*

Sample ID	Country	Population ^a	Raw reads	Filtered sequences ^b	On-target reads (%)	% of target covered at 15× ^c
Cro.1 ^d	Croatia	Reintroduced	42,321,520	28,717,105	19.6	79.0
Cro.2	Croatia	Reintroduced	19,467,436	15,178,879	26.2	66.2
Est.1	Estonia	Natural	20,936,570	15,527,441	24.6	80.3
Est.2	Estonia	Natural	25,697,066	17,894,280	25.6	81.9
Fin.1	Finland	Natural	22,477,300	13,883,204	27.6	82.4
Fin.2	Finland	Natural	22,967,626	16,569,199	24.7	82.8
Fin.3	Finland	Natural	22,894,258	16,450,208	24.2	63.4
Latv.1	Latvia	Natural	22,321,270	17,341,768	23.6	77.7
Latv.2	Latvia	Natural	24,394,838	18,562,462	23.7	80.1
Latv.3	Latvia	Natural	21,356,226	16,677,724	20.8	81.6
Nor.1	Norway	Natural	25,157,782	15,829,112	26.5	66.3
Nor.2	Norway	Natural	25,893,728	18,727,706	24.1	82.9
Nor.3 ^d	Norway	Natural	51,194,776	33,117,564	19.9	81.8
Pol.1	Poland	Natural	36,350,048	22,862,227	27.0	83.1
Pol.2	Poland	Natural	23,139,694	15,996,340	26.8	73.0
Rom.1	Romania	Natural	23,078,140	17,044,276	27.4	83.1
Rom.2	Romania	Natural	24,678,618	16,330,922	29.5	83.2
Rom.3 ^d	Romania	Natural	45,479,818	28,824,269	21.3	78.3
Rus.1	Russia	Natural	28,224,284	20,381,841	22.9	80.6
Rus.2 ^d	Russia	Natural	36,384,372	25,422,519	19.9	83.0
Rus.3	Russia	Natural	24,140,510	18,351,015	23.1	81.0
Rus.4	Russia	Natural	23,737,994	17,874,068	22.8	81.6
Svk.1	Slovakia	Natural	21,955,584	16,006,621	29.2	82.8
Svk.2	Slovakia	Natural	25,109,422	18,732,792	26.8	83.7
Svn.1	Slovenia	Reintroduced	24,523,316	18,384,101	24.9	85.1
Svn.2 ^d	Slovenia	Reintroduced	46,579,692	28,828,601	23.0	86.8
		Mean	28,094,688	19,596,779	24.4	76.7

^a“natural” refers to autochthonous populations or populations that have recovered from population bottlenecks through natural means (i.e., without human assistance); “reintroduced” populations are those established through human assistance (translocations, etc.). ^bAdapter and quality-trimmed, merged sequences and sequences that could not be merged. ^cOf the targeted 809 CDS with a total length of 618,547 bp (i.e., not incl. flanking sequences). ^dincl. data from single capture (see Section 2).

sequences after two consecutive captures (Supporting Information Figure S1).

Enriched libraries were paired-end sequenced on two Illumina platforms (Illumina, San Diego, CA, USA): MISEQ using version 2 300-cycle kits and NEXTSEQ using version 1 150-cycle MidOutput kits.

having a minimal overlap of 10 bp were merged using the software FLASH version 1.2.8 (Magoč & Salzberg, 2011). Merged and nonmerged sequences were used as input for mapping using BURROWS-WHEELER ALIGNER version 0.7.10 (BWA; Li & Durbin, 2009) with seeding disabled (Schubert et al., 2012).

2.4 | Data processing

2.4.1 | Preprocessing

Demultiplexing of paired-end reads using BCL2FASTQ version 2.17.1.14 (Illumina, Inc.) was followed by removal of adapter sequences using CUTADAPT version 1.3 (Martin, 2011). Adapter-clipped reads were then quality-trimmed using a sliding window approach in TRIMMOMATIC (Bolger, Lohse, & Usadel, 2014), with the phred quality threshold set at Q = 20 and window length of 10 bp. Overlapping paired-end reads

2.4.2 | Reference processing

Because lynx CDS display little sequence divergence from orthologous cat CDS (~0% to 4% sequence divergence; DWF, unpublished data), we were able to use a simple approach to generate lynx CDS references. First, we retrieved 300 bp of flanking sequences (both 5' and 3') of each cat CDS (cat genome v8.0) to extend the cat reference sequence beyond CDS boundaries—by extending the reference sequences in this manner, we aimed to improve read coverage at the boundaries of target loci. Flanking

sequences were retrieved by querying the CDS bait sequences versus the *F. catus* v8.0 genome using BLASTN (BLAST+ version 2.2.29); the resulting hits were restricted to only one target hit (-max_target_seqs 1) and were further filtered to match only the predicted chromosome. The actual CDS sequence including flanking sequence was retrieved by applying BEDTOOLS GETFASTA (version 2.17.0) using ± 300 bp CDS positions on the cat reference. Then, these extended cat sequences (CDS + flanks) served as reference for mapping all lynx sequences (i.e., from all samples). Aligned sequences were deduplicated using MarkDuplicates from PICARD-TOOLS version 1.106 (<https://github.com/broadinstitute/picard>). Variant calling was carried out using SAMTOOLS version 1.1 (Li et al., 2009) and BCFTOOLS version 1.2 (<http://github.com/samtools/bcftools>) to determine the most common lynx variant at every position; the cat reference was modified accordingly. For variants present at $\geq 3\times$ and alternate base frequency >0.5 , the alternate (lynx) base was used to generate the lynx consensus sequence. In this manner, we converted the cat sequence into a lynx “consensus” sequence for each CDS, which included 300 bp of 5’ and 3’ flanking sequence and henceforth the “lynx CDS reference” (of 1,102,167 bp length). We carried out the same procedure using the recently published Iberian lynx (*Lynx pardinus*) genome as reference (Abascal et al., 2016), representing a close relative of our study species, albeit with a less complete genome assembly. It should be noted, however, that in cases where the target species is highly divergent from the reference species, other approaches are advised to generate target species reference sequences (e.g., Portik, Smith, & Bi, 2016; Yuan et al., 2016).

2.4.3 | Sample processing

We generated a separate consensus sequence for each lynx sample by mapping against the lynx CDS reference. Following a second round of mapping against this sample-specific consensus sequence (to recover as much data as possible), GATK UnifiedGenotyper (v1.6) was then used to identify sequence variants (SNPs and InDels) in the 26 Eurasian lynx samples used for SNP discovery. Of all SNPs identified, only SNPs with a coverage $\geq 15\times$ in at least 20 of the 26 samples were retained as candidates for the SNP-panel. These candidate loci were then reduced to 144 SNPs (representing a 50% surplus over the target number of 96 SNPs for the final panel) based on the following criteria: (a) SNPs had to be biallelic (a requirement of the genotyping platform); (b) minor allele frequency (MAF) should be greater than 10% (to avoid loci with rare alleles and hence potentially low applicability); (c) no variants (SNPs or InDels) within 100 bp of the candidate SNP to avoid interference in the genotyping process; (d) only one SNP allowed per CDS to avoid physical linkage; (e) SNPs should be distributed across the genome as widely as possible (using the cat genome as reference, Figure 1c); and (f) SNPs should not lay in the flanking regions of the CDS, because our goal was to create a SNP-panel from data obtained following cross-species enrichment of CDS (flanking sequences may not be available for all study species).

2.5 | SNP-panel development

We aimed to generate a lynx SNP-panel with 96 SNPs for high-throughput genotyping using Fluidigm's SNPTYPE™ assays (Fluidigm Corp., San Francisco, CA, USA), which has proven effective for SNP genotyping when using DNA extracted from noninvasively collected samples (see Von Thaden et al., 2017). Alternative high-throughput SNP genotyping techniques such as GT-seq (Campbell, Harmon, & Narum, 2015) were not considered; although these perform well when invasive sampling is possible, they have lower success rates when working with low DNA concentrations (e.g., Carroll et al., 2018).

In particular, we intended to use the “96.96 Dynamic Array Chip for Genotyping” that allows the simultaneous genotyping of 96 samples at 96 biallelic SNPs, which is particularly useful when only little sample material is available. The latter was an important criterion for platform selection because noninvasively collected samples represent an important resource for genetic monitoring of wildlife populations, and these often yield very little DNA for analysis, which thus needs to be used very efficiently. The nanofluidic Dynamic Array Chip technology employed in this platform reduces PCR volumes to nanolitres, performing 9,216 (96×96) single-plex reactions in a highly automated fashion. As little as 1.25 μ l of DNA extract (0.5 ng/ μ l) is sufficient to genotype one sample at 96 SNPs. Genotyping itself is accomplished using allele-specific primers labelled with fluorescent dyes. For samples with low amounts of template (e.g., noninvasively collected samples), a preamplification is strongly recommended (Kraus et al., 2015; Nussberger et al., 2014) to reduce the likelihood of incorrect genotypes, which can be caused by insufficient template molecules in the 6.7 nL reaction volumes.

Prior to the selection of 96 SNPs for the final genotyping panel, we preselected 144 candidate SNPs ($96 + 48$) for evaluation (above), of which we randomly chose 15 (10.4%) for verification in five of the 26 Eurasian lynx samples using Sanger sequencing. After their successful verification (100% match between Illumina and Sanger sequencing data), we designed SNPTYPE™ assays for all 144 SNPs.

Using previously established procedures (Kraus et al., 2015), we assessed genotyping errors for all 144 SNPTYPE™ assays. In particular, we genotyped the 26 Eurasian lynx samples that had been sequenced for SNP discovery, two times with template DNA at a “standard concentration” (50 ng/ μ l) and two times with template DNA at a “low concentration” (0.5 ng/ μ l). The latter served to approximate poor DNA-quality samples (e.g., coming from noninvasively collected samples). We used “genotyping treatment c2” (Kraus et al., 2015; less dilution of specific target amplification [STA] products, 42 cycles of amplification) and evaluated (by locus) the following properties: genotype consistency among replicates, the incidences of missing data and genotype consistency with the Illumina sequencing data.

For the assembly of the final 96 SNP-panel, we chose the assays with the most consistent genotyping performance across the 26 lynx samples (i.e., showed lowest rates of missing data and genotyping errors) and for which the three genotypes (AA, AB and BB) could be

unambiguously distinguished in the scatter plots generated as output by the EP1 genotyping software (Fluidigm Corp.). To test the general applicability of the final 96 SNP-panel, we then genotyped a set of 100 Eurasian lynx samples from across Europe (Supporting Information Table S1), none of which had been used in the SNP discovery. For most samples, we performed four replicates: two using template DNA at standard concentration and two using template DNA at low concentration (see above); for some samples, we were only able to perform the two replicates at low concentration. Again, we randomly chose 10 of 96 loci (10.4%) to be verified by Sanger sequencing in five of the 100 Eurasian lynx.

As DNA of prey species may contaminate DNA extracted from noninvasively collected faecal samples, we also performed cross-species testing of the 96 SNP-panel on typical representatives of common prey taxa: roe deer *Capreolus capreolus*, European red deer *Cervus elaphus*, European hare *Lepus europaeus*, house mouse *Mus musculus*, pine marten *Martes martes* and American mink *Neovison vison*.

2.6 | Individual identification

We used two approaches to examine the power of the newly designed 96 SNP-panel to discriminate among individuals. First, we used GIMLET version 1.3.3 (Valière, 2002) to estimate (across loci) both the unbiased “probability of identity” (PID_{unb}) and the more conservative “probability of identity given siblings” (PID_{sib}). These probabilities were estimated separately for each population with at least 10 individuals (Estonia, Latvia, Poland, Finland, Norway, Slovenia), as well as for the genetic clusters identified in the population structure analyses (below).

Second, we directly examined the performance of the SNPs to differentiate individuals in our data set. In particular, we examined how well different subsets of loci performed for identifying individuals. We examined a range of subset sizes (10, 20, 30, 40, 50, 60, 70, 80 and 90 loci) and examined the results for 10,000 permutations per subset size; for this, subsets of loci were randomly drawn from the entire data set without replacement. This analysis was conducted in the statistical programming environment R (<http://www.cran.r-project.org>) using a custom script (deposited on DRYAD under <https://doi.org/10.5061/dryad.3f4jr01>). Only samples without any missing data were used in this analysis, to avoid an additional character state (“missing data”) from influencing results (that is, inflating the ability of n loci to discriminate individuals in the data set).

2.7 | Population structure inferences

For the assessment of SNP performance regarding the detection of genetic substructure and population assignment, we used two common methods: principal component analysis (PCA) and Bayesian population assignment. PCA was carried out using the R package ADEGENET version 2.0.1 (Jombart, 2008). Bayesian assignment to genetic clusters (populations) was carried out using the software STRUCTURE version 2.3 (Pritchard, Stephens, & Donnelly, 2000). In the

latter, we were interested in the number of genetic clusters identified, but also whether the SNPs could be used to assign lynx to their correct source cluster. To examine this, we first conducted the STRUCTURE analysis on samples ($N = 104$) from naturally occurring populations and then tested whether samples from reintroduced populations (Croatia, Slovenia; $N = 21$) would be assigned to the genetic cluster corresponding to their source population, that is the population from which lynx had been translocated to establish the reintroduced populations. First, for all 104 lynx from naturally occurring populations, we ran 10 replicates for values of K (inferred number of genetic clusters) from 1 to 8 for 600,000 iteration steps, the first 150,000 of which were discarded as burn-in, and allowing for correlated allele frequencies in the admixture model (Falush, Stephens, & Pritchard, 2003). The most likely number of genetic clusters (K) was then determined following the ΔK method (Evanno, Regnaut, & Goudet, 2005) implemented on the STRUCTURE HARVESTER website (<http://taylor0.biology.ucla.edu/structureHarvester>) (Earl & vonHoldt, 2012). Using the inferred K , we then re-ran STRUCTURE with all lynx ($N = 125$) to check for correct cluster assignment.

We tested for linkage disequilibrium (LD) and deviation from Hardy–Weinberg equilibrium (HWE) with the R package GENEPOP (version 1.0.5, Rousset, 2008) in sampling localities and inferred genetic clusters; the R package LDHEATMAP (Shin, Blay, McNeney, & Graham, 2006) was used to visualize pairwise r^2 values.

2.8 | Applicability of baits for other taxa

To assess the potential taxonomic breadth for which our cat CDS-derived baits may be applicable, we queried our baits against the published genomes of other carnivore species using BLASTN: Iberian lynx (*Lynx pardinus* v1.0), cheetah (*Acinonyx jubatus* v1.0), leopard (*Panthera pardus* v1.0), tiger (*Panthera tigris altaica* v1.0), dog (*Canis lupus familiaris* v3.1), ferret (*Mustela putorius furo* v1.0), polar bear (*Ursus maritimus* v1.0) and giant panda (*Ailuropoda melanoleuca* v1.0). As sequence divergence between bait and target impacts enrichment (Bi et al., 2012; Bragg, Potter, Bi, & Moritz, 2016; Hedtke, Morgan, Cannatella, & Hillis, 2013; Pajmans, Fickel, Courtiol, Hofreiter, & Förster, 2016; Peñalba et al., 2014; Portik et al., 2016; Vallender, 2011), we determined the sequence similarity between the baits and their corresponding best resulting hit in the queried genomes. While capture performance is dependent on multiple factors, including concentration, length and tiling of baits, diversity and quantity of input library and hybridization temperature, the divergence between bait and target has received the most attention—and it is only for the latter that a predictive framework is available. Portik et al. (2016) investigated the relationship between sequence divergence and several metrics of exon capture performance in a large number of taxa (264 samples across 15 genera of frogs). Provided that the sequence divergence from the target species is known, their regression formulas can be used to make predictions about capture results. Here, we employ the prediction for *missing data* % (i.e., proportion of target not recovered) as a consequence of sequence divergence between bait

and target to address the utility of our cat CDS-derived baits in other carnivorans.

3 | RESULTS

3.1 | Sequence and target enrichment results

The number of raw reads varied from 19,467,436 to 46,579,692 among the 26 lynx samples (mean: 28,094,688), and the percentage of on-target sequences ranged from 19.6% to 29.5% (mean: 24.4%; Table 1). As the target region corresponds to 0.026% of the cat genome (Supporting Information Table S2), and the lynx genome is roughly equivalent in size (Animal Genome Size Database, <http://www.genomesize.com>), the cross-species capture resulted in a greater than 900-fold increase in on-target sequences.

For half of the CDS, the enrichment was very successful: On average, 90% or more of the target region was covered at $\geq 15\times$ depth per sample (Figure 2a). Across all samples, the majority of CDS (632 loci; 78.1%) had a coverage of $\geq 15\times$ for 50% or more of their lengths. Some CDS (61 loci; 7.5%) were poorly enriched in all samples (less than 10% of target region $\geq 15\times$ depth), among them 31 CDS were not enriched in any sample.

As expected, the inclusion of flanking sequences in the lynx CDS references (300 bp, both 5' and 3', see Section 2) increased the number of mapped sequences and improved coverage at CDS boundaries. This yielded between 16,701 and 30,309 more bases of the target with $\geq 15\times$ depth (mean: 20,674 bases; 3.34% of the target region).

3.2 | Variation in CDS

Of the 809 CDS analysed, 61 were excluded due to insufficient data, 207 showed no variation among lynx samples (most probably due to poor coverage of CDS at $\geq 15\times$ depth) and one showed no variation between lynx and cat. Three further CDS were no longer present in the newest build of the cat genome (v8.0). The remaining 537 CDS (66.4% of CDS; length ranging from 961 to 5,113 bp incl. flanking sequence) showed intraspecific variation, which consisted of 1,186 SNPs and 109 InDels (Figure 2b).

Use of the Iberian lynx genome as reference (see Section 2) revealed fewer intraspecific variants among Eurasian lynx samples (993 SNPs and 66 InDels). Most of this difference in the number of variants detected reflects differences in the genome assemblies of the cat and the Iberian lynx (i.e., fewer CDS present in the genome assembly, less flanking sequence retrievable)—in other words, the variation detected using the Iberian lynx genome was mostly a subset of the variation detected using the cat genome. Variants detected only when using the Iberian lynx genome as reference were located in the flanking regions of 14 CDS (18 SNPs, three InDels), which were characterized by low or no coverage when using the cat genome. Thus, although the Iberian lynx is a closer relative of the Eurasian lynx than the domestic cat, the more complete genome assembly of the cat

enabled us to retrieve more data about intraspecific variation in the Eurasian lynx.

3.3 | SNP-panel

Of the 686 candidate SNPs inside CDS (57.8% of SNPs), we selected 144 for further evaluation (see Section 2 for selection criteria). This number was then further reduced to 96 SNPs based on reliability for genotyping on the chosen platform (Fluidigm Corp., “96.96 Dynamic Array Chip for Genotyping”). Based on the distribution of the selected SNPs when projected onto the cat genome (Figure 1c) and assuming a similar distribution in lynx, we estimated the average distance between SNPs on the same chromosome in the final set of 96 SNPs to be 17.55 Mb.

As indicated by the improved genotyping success (Figure 3), the final set of 96 SNPs had less missing data or genotyping inconsistencies in the 26 lynx of the discovery panel (Figure 3b), than did the set of all 144 SNPs (Figure 3a). Some sample replicates—mostly those using low concentrations of template DNA—had a high incidence of missing data (Figure 3a,b). We obtained complete or nearly complete genotypes for the 100 additional lynx samples (Figure 3c); mean genotyping success was 98.8% for replicates using the standard concentration of template DNA (red dots in Figure 3c) and 96.0% for replicates using the low concentration of template DNA (blue dots in Figure 3c). Considering the whole data set ($N = 125$), we found no significant linkage between loci (LD heatmap in Supporting Information Figure S2) and detected no deviation from HWE. In some subsets of the data (sampling localities or genetic clusters, see below), a small number of SNPs were in LD or deviated from HWE (see Supporting Information Table S3).

Cross-species tests resulted in low overall genotyping success, except for the domestic cat (upon which the cross-species capture was based), which generated data at 83 loci, of which only five were heterozygous (6%). The other prey species displayed signals at a far lower rate: roe deer at 13 loci, European red deer at nine loci, European hare at eight loci, house mouse at two loci, pine marten at 27 loci and American mink at 24 loci. Thus, the most likely prey species (roe deer, red deer, hare [Jobin, Molinari, & Breitenmoser, 2000; Belotti, Kreisinger, Romportl, Heurich, & Bufka, 2014]) would not generate false-positive lynx genotypes; for the domestic cat (and presumably also the wildcat *Felis silvestris*), species identification using commonly employed mtDNA markers may be necessary to exclude false-positive lynx genotypes.

3.4 | Identification of individuals

All 96 SNPs of the final panel were polymorphic for the complete lynx data set ($N = 125$), with a MAF between 6% and 50% (mean = 28%). The cumulative PID_{unb} (unbiased probability of identity) within geographic localities ranged from 1.71×10^{-17} to 7.28×10^{-35} (Table 2), and in the genetic clusters identified in the population structure analyses (below) from 1.51×10^{-21} to 1.19×10^{-32} . The more conservative estimate of PID_{sib} (probability

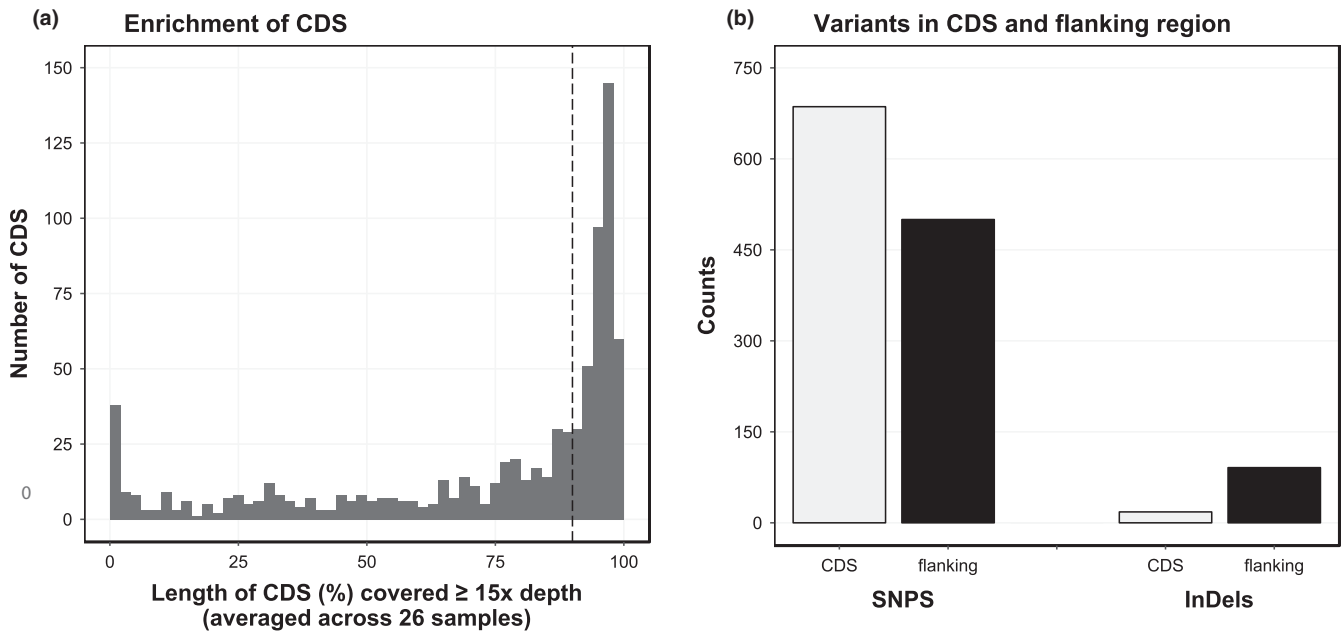
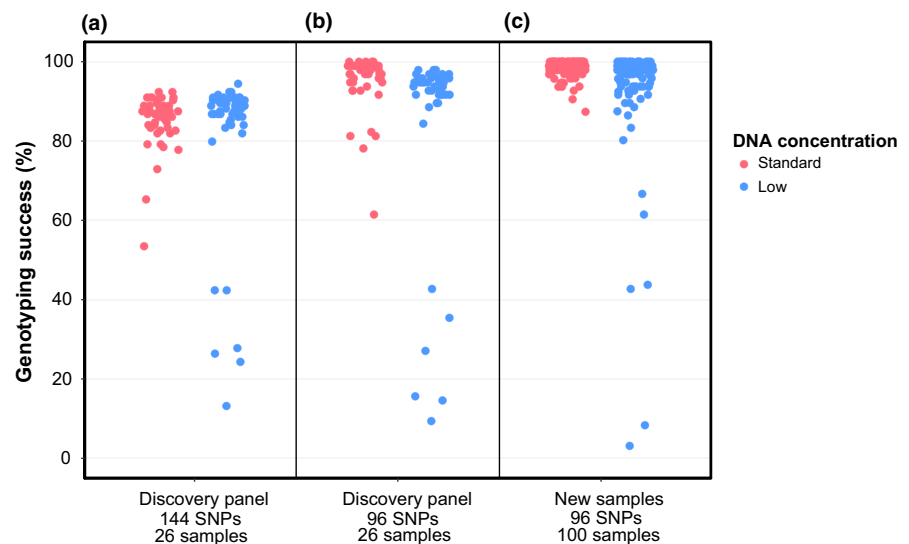


FIGURE 2 (a) Average recovery of the 809 CDS enriched in the 26 Eurasian lynx in the SNP discovery panel. For each CDS, the average percentage of the locus length covered at $\geq 15\times$ depth is plotted (using 2% bins). 50% of CDS have above 90% sequence length coverage at $\geq 15\times$ depth, indicated by the black vertical line. (b) Summary of the distribution of variation in CDS and flanking sequences, shown separately for SNPs and InDels

FIGURE 3 Genotyping success of Eurasian lynx samples at SNP loci: (a) the discovery panel (26 lynx) for the initial set of 144 SNPs tested on the Fluidigm genotyping platform; (b) the final 96 SNP-panel genotyped on the discovery panel (26 lynx); and (c) the final 96 SNP-panel on an additional 100 lynx (not part of discovery panel). Colours are used to differentiate between genotyping success at standard template DNA concentration (red circles, 50 ng/ μ l) and low template DNA concentration (blue circles, 0.5 ng/ μ l). Scatter was added to visualize density



of identity given siblings) ranged from 2.32×10^{-8} to 9.26×10^{-17} within geographic localities and from 7.19×10^{-11} to 7.12×10^{-17} in genetic clusters (Table 2). As little as 19 loci were already sufficient to achieve a $PID_{sib} < 10^{-4}$, regardless of locality or genetic cluster (Table 2). When we examined the performance of SNPs to differentiate all lynx without missing data ($N = 102$) using various subsets of loci (ranging from 10 to 90 loci; Figure 4), we found that 40 SNPs were sufficient to differentiate all individuals in more than 99% of 10,000 random permutations of SNP loci (sampled without replacement); similar results were obtained when examining lynx from the two subspecies separately (see Appendix S1).

Given the genotyping success rate we observed (Figure 3), our final 96 SNP-panel should perform very well for individual identification.

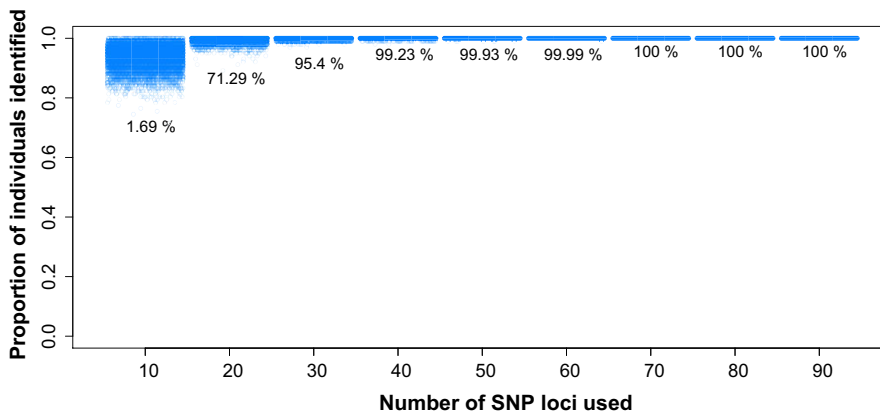
3.5 | Population structure inferences

The PCA showed a clear separation of two distinct clusters along the first principal component axis (Figure 5a), which explained 18.5% of variation. These two clusters corresponded to the Eurasian lynx subspecies *Lynx lynx lynx* (left) and *L. l. carpathicus* (right), which can also be differentiated using microsatellites (e.g., Bull et al., 2016;

TABLE 2 Probability of identity calculated for different subsets of the SNP data

Sample set	N	PID _{unb}	PID _{sib}	Minimum no. of loci needed for PID _{sib} <1 in 10,000 ^a
Geographic localities				
Estonia	27	7.68×10^{-32}	7.81×10^{-16}	18
Latvia	23	1.28×10^{-33}	9.26×10^{-17}	18
Poland	16	8.01×10^{-28}	2.73×10^{-13}	19
Finland	10	7.28×10^{-35}	1.22×10^{-15}	18
Norway	10	3.15×10^{-25}	1.43×10^{-11}	19
Slovenia ^b	13	1.71×10^{-17}	2.32×10^{-8}	19
Genetic clusters (K = 2)				
Cluster 1 Estonia, Latvia, Poland, Russia, Finland, Norway	94	1.19×10^{-32}	7.12×10^{-17}	18
Cluster 2 Romania, Slovakia, Croatia ^b , Slovenia ^b	31	1.51×10^{-21}	7.19×10^{-11}	19
Genetic clusters (K = 5)				
Cluster 1 Estonia, Latvia	50	2.39×10^{-32}	1.62×10^{-16}	18
Cluster 2 Poland	16	8.01×10^{-28}	2.73×10^{-13}	19
Cluster 3 Russia, Finland	18	1.12×10^{-33}	2.18×10^{-16}	18
Cluster 4 Norway	10	3.16×10^{-25}	1.43×10^{-11}	19
Cluster 5 Romania, Slovakia, Croatia ^b , Slovenia ^b	31	1.51×10^{-21}	7.19×10^{-11}	19

Note. PID_{unb} (unbiased probability of identity) and PID_{sib} (probability of identity given siblings) were calculated using GIMLET version 1.3.3 (Valière, 2002).
^aLoci were ranked by informativeness (per-sample set), and then, the number of loci was determined for which PID_{sib} was <10⁻⁴. ^bReintroduced population.

**FIGURE 4** Comparison of the proportion of individuals recovered using different subsets of the SNP data, ranging from 10 to 90 loci. For each subset size, 10,000 permutations (random selection of loci without replacement) were plotted; percentage values indicate the number of permutations in which all individuals in the data set without missing data (N = 102) were identified. Scatter was added on the x-axis for each subset to visualize density

Ratkiewicz et al., 2014). Some substructure within subspecies is also apparent.

In STRUCTURE, the most likely number of genotypic clusters was K = 2 (Figure 5b, upper panel). However, the probability was also high for K = 5 (see Appendix S1). As the Evanno method is designed to detect the highest hierarchical level of genetic structure (Evanno et al., 2005), we also considered K = 5 (Figure 5b lower panel).

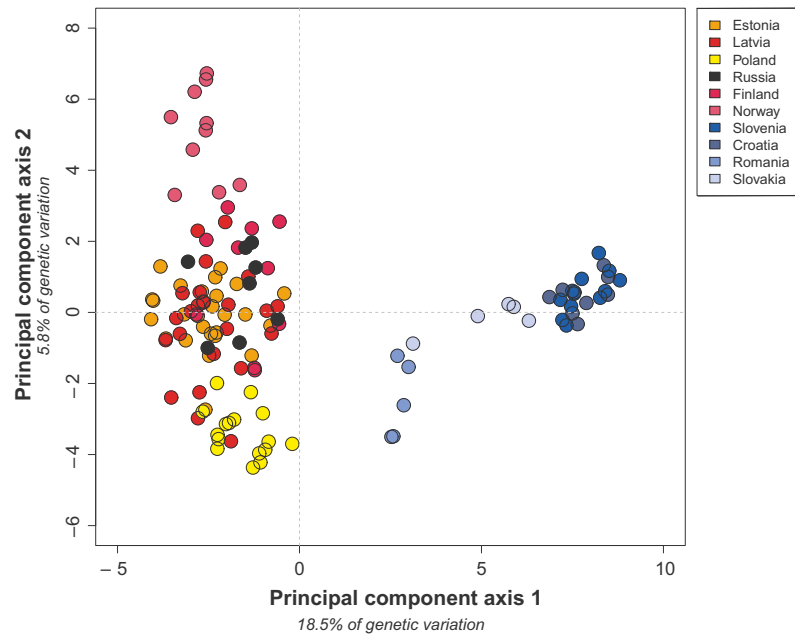
Like in the PCA, the two inferred genotypic clusters (K = 2) corresponded to the two subspecies *L. l. lynx* (Figure 5b, upper panel, yellow) and *L. l. carpathicus* (Figure 5b, upper panel, blue). The lynx from the reintroduced populations in Croatia and Slovenia were assigned with high Q-values (0.75–0.99, mean 0.91) to the cluster of their source population (represented by Romanian and Slovakian samples).

The K = 5 plot (Figure 5b, lower panel) displayed a substructure in *L. l. lynx*: individuals from Estonia + Latvia, Poland, Russia + Finland and Norway now formed their own clusters. There was also admixture between these cluster; for example from Russia + Finland into Estonia + Latvia. This population genetic structure within *L. l. lynx* is similar to the one previously reported using microsatellites (Ratkiewicz et al., 2014). Again, all lynx from reintroduced populations were assigned with high Q-values (0.74–0.98, mean 0.89) to the cluster of their source population.

3.6 | Applicability of baits in other taxa

We found high sequence similarity between the cat CDS-derived baits and their targets in four other felid species, Iberian lynx, cheetah, leopard and tiger (all four: median sequence divergence of 0.8%; Figure 6).

(a) PCA



(b) Bayesian population assignment (STRUCTURE)

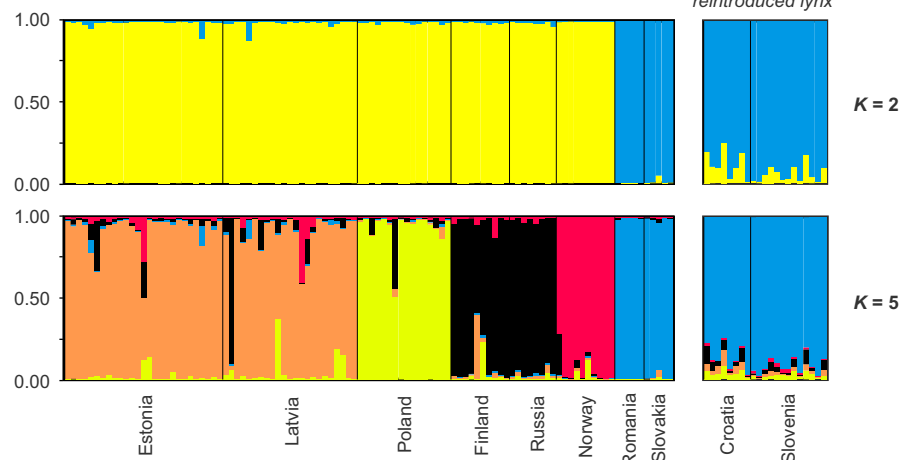


FIGURE 5 Population structure inferences using the 96 SNP-panel. (a) Principal component analysis (PCA); the first two principal component axes are plotted, with the geographic origin of samples indicated by colour. (b) Bayesian population assignment using STRUCTURE; results for $K = 2$ (top) and $K = 5$ (bottom) are displayed. Results for lynx from reintroduced populations (Croatia, Slovenia) are shown on the right

As expected, sequence divergence between cat CDS-derived baits and their targets in other, more distantly related carnivoran species was higher, with a median sequence divergence of 5.8% (Figure 6). Even for the most distantly related taxa considered, the amount of missing data predicted using the regression formula of Portik et al. (2016) was low-to-moderate (up to ~20%; right y-axis in Figure 6) for the majority of targeted regions (i.e., sequences covered by baits).

4 | DISCUSSION

Our results show that targeted capture of CDS can be used for SNP discovery in nonmodel organisms and that a subset of the identified SNPs can be successfully implemented in a high-throughput genotyping platform to accurately identify individuals and to infer population genetic structure of the species of interest.

Using publicly available genomic resources for model organisms, we were able to design baits for 100s of target CDS loci that were then enriched in our nonmodel study species, the Eurasian lynx *Lynx lynx*. We successfully surveyed intraspecific variation in *L. lynx* across its European range and generated a large data set of SNPs inside CDS and their flanking regions. A large proportion of CDS had good or complete coverage of the target region ($\geq 15\times$ depth per sample) and yielded 1,186 SNP loci for downstream applications.

4.1 | Targeted capture for SNP discovery

Targeted capture does not require an exact sequence match between bait and target for successful enrichment. While decreasing sequence similarity between bait and target reduces the efficiency of capture (e.g., Pajmans et al., 2016; Portik et al., 2016), successful enrichment has been reported for species with up to 40% sequence

divergence (Li et al., 2013). This “mismatch tolerance” is utilized in divergent reference capture, where baits designed using molecular data of one species are used to enrich complementary sequences in one or more other species. In interspecific studies, this yields comparable data (100s–1000s loci) across distantly related species (e.g., Bragg et al., 2016; Newman & Austin, 2016; Portik et al., 2016; Yuan et al., 2016).

Rather than generating comparable data across multiple species, we used exon-capture (Bi et al., 2012) to gain comparable data across many samples within the same species. In this manner, we generated high sequence coverage for a small portion of the nuclear genome—the targeted CDS—across all samples. This portion of the genome was then screened for variation in the form of SNPs, a subset of which was then used to design the SNP-panel.

Capture is scalable, with the number of targets determined by bait design. The requirements for the development of a SNP assay—such as technological constraints based on genotyping technology (e.g., no variants flanking SNPs of interest; restriction to biallelic loci), minimum allele frequencies and setting a limit of one SNP per target locus (to avoid physical linkage)—limit the number of identified SNPs that can be utilized in a SNP-panel. To ensure the recovery of sufficient variation for the development of a 96 SNP-panel, we thus chose to target several hundred CDS spread throughout the genome. However, considering that we exceeded the number of SNPs required for the development of a 96 SNP-panel (incl. the mentioned restrictions), a smaller number of targets would have been sufficient. For the purpose of developing a SNP-panel of similar size, we would still recommend a number of target loci exceeding what is practical using DIY protocols (e.g., 51 loci, Peñalba et al., 2014). In particular, as the per-sample costs of using custom baits (e.g., from MYcroarray) can be reduced using smaller reaction volumes in combination with a dilution of synthetic RNA baits (this study; Li et al., 2013; Yuan et al., 2016; Cruz-Dávalos et al., 2017) and by pooling barcoded libraries of multiple samples prior to hybridization (see, for example Portik et al., 2016; for capture

success following different levels of sample pooling). In addition, lowering the sequence depth requirement for SNP calling permits pooling of more samples during sequencing and can further reduce costs (e.g., $\geq 8\times$ depth, Lim & Braun, 2016). Considering our results, and the filters we used to select candidate SNPs for the SNP-panel (see Section 2), we would recommend a minimum of ~250 target loci (400 bp or longer) for SNP discovery using an approach like the one described here.

The increasing availability of genomic resources for nonmodel organisms, particularly annotated genomes and transcriptomes, improves the chances of finding species for bait design that is not too distantly related to target species. While successful enrichment of sequences has been reported for “bait species” with very high divergence times from the target species (up to ~250 million years; Li et al., 2013; Hedtke et al., 2013), several studies have observed a drop in capture efficiency with increasing sequence divergence between bait and target (e.g., Bi et al., 2012; Bragg et al., 2016; Pajmans et al., 2016; Peñalba et al., 2014; Portik et al., 2016; Vallender, 2011). For this reason, we examined the taxonomic breadth at which the baits used here should still retrieve sufficient data for the design of a similarly sized SNP-panel. The relatively limited sequence divergence between our cat CDS-derived baits and their targets in other carnivorans with published genomes (Iberian lynx, cheetah, leopard, tiger, domestic dog, ferret, polar bear and giant panda), suggests a broad utility of these baits. Unsurprisingly, the representatives of the carnivoran family *Felidae* (Iberian lynx, cheetah, leopard and tiger) display very low sequence divergence between the cat CDS-derived baits and the target loci. The baits that we used for exon-capture in lynx should thus perform equally well for other species in the family *Felidae* (consisting of 14 genera and 37 species) that diverged approximately 11 million years ago (Johnson et al., 2006; Li, Davis, Eizirik, & Murphy, 2016). The more distantly related carnivorans (domestic dog, ferret, polar bear and giant panda) display higher sequence divergence between bait and target. Despite a low-to-moderate amount of missing data (up to ~20%) predicted for these

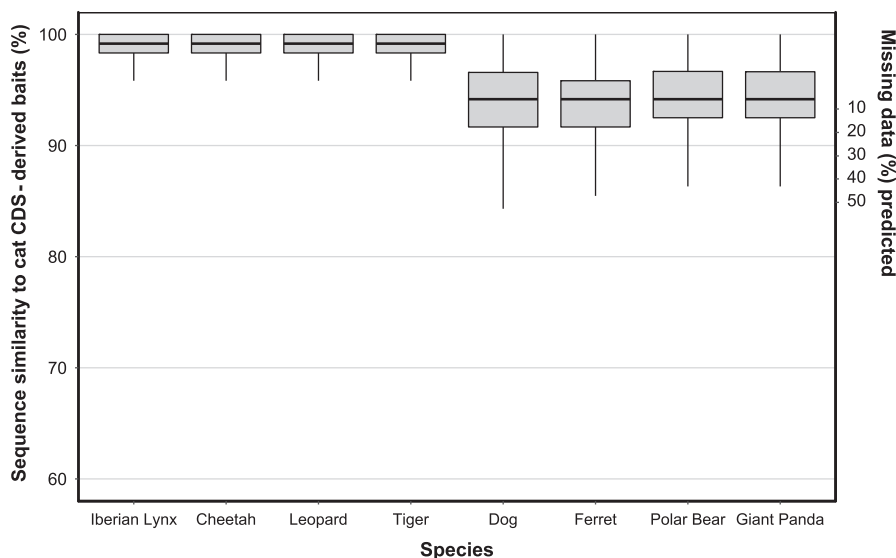


FIGURE 6 Sequence similarity between the domestic cat CDS-derived baits and their targets in other carnivorans with published genomes, including four felids (Iberian lynx, cheetah, leopard and tiger) and four caniform carnivorans (dog, ferret, polar bear and giant panda). The right y-axis displays estimates for *missing data (%)* (proportion of target not recovered) based on sequence similarity between bait and target (Portik et al., 2016)

taxa (Figure 6), sufficient data would be retrieved for the design of a SNP assay. In fact, our results for the Eurasian lynx suggest that even with 50% missing data, enough loci would likely be recovered to exceed the requirements for the design of a 96 SNP-panel (see above). Our cat CDS-derived baits should thus be suitable to examine intraspecific variation in species across the whole carnivoran order (286 species; Wozencraft, 2005) and over substantial divergence times (approx. 59 million years, Eizirik et al., 2010). Exon-capture (Bi et al., 2012) as employed here (targeting single-copy CDS) can thus provide intraspecific sequence variants from a subset of the genome, across a broad range of species and divergence times, supplying the necessary information to develop SNP-panels for non-model species.

Because targeted capture can be used on samples of poor quality (noninvasively collected material, Hernandez-Rodriguez et al., 2018; Perry et al., 2010; archival material, Lim & Braun, 2016; McCormack et al., 2016; ancient material, Carpenter et al., 2013; Enk et al., 2014; formalin-fixed material, Ruane & Austin, 2017), it is also valuable for studies of taxa for which samples are difficult to obtain for genetic analyses (rare and elusive species, those in difficult to reach habitats and degraded samples such as archival material). For example, noninvasively collected material can be incorporated in the SNP discovery process to cover portions of a species' distribution without having fresh tissue samples available, or, archival and ancient samples can be added to existing data sets, providing information about historical (e.g., extinct) populations (Bi et al., 2014; Lim & Braun, 2016).

4.2 | Comparison to other targeted capture approaches

Of late, several approaches have been developed that combine targeted capture with other methods of genome reduction to obtain information about nuclear sequence variation in nonmodel species. Some of these approaches combine targeted capture with a RAD-seq workflow (RADcap, Hoffberg et al., 2016; Rapture, Ali et al., 2016; hyRAD, Suchan et al., 2016). These can overcome some of the challenges or drawbacks of RAD-seq (e.g., sequence polymorphism at a restriction site resulting in null alleles, variable coverage across samples and application to historical samples), and either generate baits directly (hyRAD) or supply the sequence information needed for bait synthesis (RADcap & Rapture). These approaches yield improved matrix occupancy (i.e., more complete data sets) at the cost of expanded laboratory and bioinformatic workflows (modification of RAD-seq and addition of targeted capture). The hyRAD-X approach (Schmid et al., 2017) includes the additional incorporation of an RNA-seq workflow and captures RAD loci located in the transcriptome. In this approach, RAD loci are not anonymous, as they map to an assembled transcriptome. Another approach that combines targeted capture with RNA-seq is expressed exome capture sequencing (EecSeq; Puritz & Lotterhos, 2018), in which mRNA is used to generate baits, after reducing (normalizing) the abundance of highly expressed genes in the cDNA pool.

All of these approaches represent methodological improvements and have great potential for surveying intraspecific variation in non-model species. Moreover, they are cost-effective for population-level studies involving 100s of individuals. This makes them attractive for studies of taxa with little-or-no genomic resources and for which invasive sampling is possible. However, due to their complexity and requirements, they may be less suitable for SNP discovery in rare and protected species that are to be genetically monitored using noninvasively collected samples. In part, the benefits (incl. cost saving) of these approaches are achieved by high sample throughput—something that may not be possible or necessary for SNP discovery (which can be based on 10s of samples, as demonstrated here). Obtaining the requisite mRNA for some of these approaches will be a significant constraint when working with rare and/or elusive species. The investment of time and money for pilot studies required for some of the aforementioned approaches, in addition to the added laboratory and bioinformatic expertise needed, may also limit their practicality.

For most species of conservation concern that can be genetically monitored using noninvasively collected samples (e.g., faeces or hair), the requisite genomic resources for the SNP discovery workflow presented here are available; that is, data from a not too distantly related species (see above). A user-friendly BLAST-based pipeline (Li et al., 2012) can thus be used to identify target loci (single-copy CDS), which are used to design baits for exon-capture (Figure 1a). As these baits can be used for many taxa, spanning broad divergence times, implementation of the workflow for even distantly related species requires no additional investment. The only addition to standard molecular laboratory equipment required to implement this workflow is the magnetic rack needed for targeted capture, and potential cost-saving measures have already been mentioned (dilution of baits, pooling of samples before capture, lower sequence depth to call SNPs). Furthermore, the bioinformatic pipeline postcapture is a standard variant calling procedure, which only involves sequence read processing (adapter and quality trimming), mapping and variant calling. Our workflow thus offers a straightforward means to obtain nuclear sequence variation at 100s of loci—that can be implemented for many taxa without modification—and uses mostly standard molecular laboratory equipment and standard bioinformatic procedures.

4.3 | Genetic monitoring using SNPs in CDS

The ability to accurately identify and differentiate individuals is central to population monitoring (Frankham et al., 2010). Thus, molecular markers used in noninvasive genetic monitoring must have sufficient power to differentiate individuals—even closely related individuals—and overcome the analytical difficulties often associated with noninvasively collected sample material, namely DNA extractions of low volume and low concentration that are characterized by varying levels of DNA degradation.

For our SNP-panel, we adopted a SNP typing platform, Fluidigm's Dynamic Array Chips, which has been successfully used to

genotype SNPs in a range of noninvasively collected material, such as individual hair, faecal samples and urine samples (Kraus et al., 2015; Nussberger et al., 2014). Using this platform, we observed a high genotyping success rate for samples with very low DNA concentrations (0.5 ng/μl), which we used in lieu of noninvasively collected samples. Conducting multiple replicates per sample is unproblematic, as only a limited amount of template is required per sample, and the costs of genotyping samples are relatively low (Kraus et al., 2015).

Regarding the identification of individuals, the 96 SNP format performed very well. This is in line with previous studies, which found that 40–100 SNP loci performed similarly well or better than the typical number (10–20) of microsatellite loci used for the purpose of individual identification and kinship analysis (e.g., Gärke et al., 2012; Hauser et al., 2011; Kaiser et al., 2017; Morin et al., 2012; Tokarska et al., 2009; Weinman, Solomon, & Rubenstein, 2015). Here, we observed that a low number (18 or 19) of informative loci was more than sufficient to distinguish individuals of a given population. Our permutation test showed that even with a substantial number of locus dropouts (up 30%–40%), we were able to distinguish individuals, indicating that this 96 SNP-panel is robust enough for genotyping poor quality (e.g., noninvasively collected) samples.

The ability to make population structure inferences and to assign individuals to populations is an important component of genetic population monitoring, providing information about animal movements and potential gene flow, the impact of habitat fragmentation, degree of inbreeding and other population parameters (Frankham et al., 2010). In past years, there has been increasing evidence for the suitability of genic SNPs to ascertain population membership of individuals (DeFaveri, Viitaniemi, Leder, & Merila, 2013; Elbers, Clostio, & Taylor, 2017; Freamo, O'Reilly, Berg, Lien, & Boulding, 2011; Oliveira et al., 2015; Zhan et al., 2015). This has been examined by itself and also in comparison with both traditionally used microsatellite markers (e.g., DeFaveri et al., 2013; Elbers et al., 2017) and non-genic SNPs (e.g., DeFaveri et al., 2013; Elbers et al., 2017; Zhan et al., 2015); in all cases, genic SNPs performed equally well or better than alternative markers.

Using our CDS-derived SNPs, we were able to unambiguously delineate the two Eurasian lynx subspecies in our sample set (*L. l. lynx* and *L. l. carpathicus*). Within *L. l. lynx*, which dominated our sample set, further levels of population structure could be resolved. This structure was congruent with the one detected in a larger sample set ($N = 298$) that had been genotyped at 13 microsatellite loci (Ratkiewicz et al., 2014). In the same way, the extent and direction of introgression detected using our 96 SNP-panel mirrored that observed using microsatellites in the aforementioned study. Using our 96 SNP-panel, we were also able to accurately assign individuals from reintroduced populations to the genetic cluster of their source population. The 96 SNP-panel presented here thus appears more than suitable for the genetic monitoring of Eurasian lynx across their European range considering the higher potential for automation of SNP genotyping, better collaboration possibilities and cost reduction potential (Kraus et al., 2015).

5 | CONCLUSION

There is growing evidence for the utility of genic SNPs for the genetic monitoring of populations. As demonstrated here, targeted capture coupled with high-throughput sequencing is well suited for acquiring information regarding such intraspecific variation, even in cases where study species lack genomic resources. With the increasing availability of genomic resources for nonmodel species, this kind of approach will become more broadly applicable—even though capture already shows potential to work across substantial divergence times (e.g., Bragg et al., 2016; Hedtke et al., 2013; Li et al., 2013; Portik et al., 2016).

Baits do of course not need to be explicitly designed with the aim of discovering SNPs for genetic monitoring purposes. Thus, baits designed for other purposes (e.g., resolving taxonomic uncertainties, Yuan et al., 2016; identifying regulatory sequences, Yoshihara et al., 2016; identifying adaptive genes, Roffler et al., 2016; investigating loci linked to traits, Springer et al., 2015) can be used to screen samples from related species for intraspecific variation.

ACKNOWLEDGEMENTS

This research was funded by the Leibniz Association (SAW-2011-SGN-3), by the Deutscher Akademischer Austauschdienst (A/11/93927) and National Science Center, Poland (2014/15/B/NZ8/00212). The licence for lynx live-trapping and blood sampling in Poland was obtained from the National Ethics Committee for Animal Experiments (no. DB/KKE/PL—110/2001) and the Local Ethics Committee for Animal Experiments at the Medical University of Białystok, Poland (no. 52/2007). Import of samples from Russia was licensed by CITES permission no 12RU000512. We thank Janis Ozolins, Peep Männil and Ingrid Reinkind for providing samples from Latvia, Estonia and Norway. We would also like to thank three anonymous reviewers for helpful comments and discussion on the manuscript.

DATA ACCESSIBILITY

An Appendix S1, Supporting Information Tables S1, S2 and S3, and Supporting Information Figures S1 and S2 can be found online with the study. Bait sequences, SNP genotypes, the final 96 SNP-panel assay, as well as an R script, are deposited on DRYAD under <https://doi.org/10.5061/dryad.3f4jr01>. Illumina sequences are deposited in the NCBI SRA under Accession no.: SRP116616.

AUTHOR CONTRIBUTIONS

D.W.F., R.H.S.K., H.B., R.K., C.N. and J.F. designed the study; K.S., A.P.S. and M.S. coordinated sample collection; D.W.F., J.B., M.A., J.L.A.P. carried out the experiments; D.W.F. and D.L. analysed and interpreted data; and D.W.F. and J.F. wrote the manuscript. All authors edited and approved the final manuscript.

ORCID

Daniel W. Förster  <http://orcid.org/0000-0002-6934-0404>

Ralph Kuehn  <http://orcid.org/0000-0001-9518-0590>

REFERENCES

- Abascal, F., Corvelo, A., Cruz, F., Villanueva-Canas, J. L., Vlasova, A., Marcet-Houben, M., ... Godoy, J. A. (2016). Extreme genomic erosion after recurrent demographic bottlenecks in the highly endangered Iberian lynx. *Genome Biology*, 17, 251. <https://doi.org/10.1186/s13059-016-1090-1>
- Ali, O. A., O'Rourke, S. M., Amish, S. J., Meek, M. H., Luikart, G., Jeffres, C., & Miller, M. R. (2016). RAD capture (Rapture): Flexible and efficient sequence-based genotyping. *Genetics*, 202, 389–400. <https://doi.org/10.1534/genetics.115.183665>
- Bayerl, H., Kraus, R. H., Nowak, C., Foerster, D. W., Fickel, J., & Kuehn, R. (2018). Fast and cost-effective single nucleotide polymorphism (SNP) detection in the absence of a reference genome using semi-deep next-generation Random Amplicon Sequencing (RAMseq). *Molecular Ecology Resources*, 18, 107–117. <https://doi.org/10.1111/1755-0998.12717>
- Belotti, E., Kreisinger, J., Romportl, D., Heurich, M., & Bufka, L. (2014). Eurasian lynx hunting red deer: Is there an influence of a winter enclosure system? *European Journal of Wildlife Research*, 60, 441–457. <https://doi.org/10.1007/s10344-014-0801-8>
- Bi, K., Linderroth, T., Vanderpool, D., Good, J. M., Nielsen, R., & Moritz, C. (2014). Unlocking the vault: Next generation museum population genomics. *Molecular Ecology*, 22, 6018–6032.
- Bi, K., Vanderpool, D., Singhal, S., Linderroth, T., Moritz, C., & Good, J. M. (2012). Transcriptome-based exon capture enables highly cost-effective comparative genomic data collection at moderate evolutionary scales. *BMC Genomics*, 13, 403. <https://doi.org/10.1186/1471-2164-13-403>
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30, 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Bonin, A., Bellemain, E., Eidesen, P. B., Pompanon, F., Brochmann, C., & Taberlet, P. (2004). How to track and assess genotyping errors in population genetics studies. *Molecular Ecology*, 13, 3261–3273. <https://doi.org/10.1111/j.1365-294X.2004.02346.x>
- Bragg, J. G., Potter, S., Bi, K., & Moritz, C. (2016). Exon capture phylogenomics: Efficacy across scales of divergence. *Molecular Ecology Resources*, 16, 1059–1068. <https://doi.org/10.1111/1755-0998.12449>
- Broquet, T., Ménard, N., & Petit, E. (2007). Noninvasive population genetics: A review of sample source, diet, fragment length and microsatellite motif effects on amplification success and genotyping error rates. *Conservation Genetics*, 8, 249–260.
- Brumfield, R. T., Beerli, P., Nickerson, D. A., & Edwards, S. V. (2003). The utility of single nucleotide polymorphisms in inferences of population history. *Trends in Ecology & Evolution*, 18, 249–256. [https://doi.org/10.1016/S0169-5347\(03\)00018-1](https://doi.org/10.1016/S0169-5347(03)00018-1)
- Bull, J., Heurich, M., Saveljev, A., Schmidt, K., Fickel, J., & Förster, D. (2016). The effect of reintroduction on the genetic variability in Eurasian lynx populations: The cases of Bohemian-Bavarian and Vosges-Palatinian populations. *Conservation Genetics*, 17, 1229–1234. <https://doi.org/10.1007/s10592-016-0839-0>
- Burbano, H. A., Hodges, E., Green, R. E., Briggs, A. W., Krause, J., Meyer, M., ... Pääbo, S. (2010). Targeted investigation of the Neandertal genome by array-based sequence capture. *Science*, 328, 723–725. <https://doi.org/10.1126/science.1188046>
- Campbell, N. R., Harmon, S. A., & Narum, S. R. (2015). Genotyping-in-Thousands by sequencing (GT-seq): A cost effective SNP genotyping method based on custom amplicon sequencing. *Molecular Ecology Resources*, 15, 855–867. <https://doi.org/10.1111/1755-0998.12357>
- Carpenter, M. L., Buenrostro, J. D., Valdiosera, C., Schroeder, H., Allentoft, M. E., Sikora, M., ... Bustamante, C. D. (2013). Pulling out the 1%: Whole-genome capture for the targeted enrichment of ancient DNA sequencing libraries. *The American Journal of Human Genetics*, 93, 852–864. <https://doi.org/10.1016/j.ajhg.2013.10.002>
- Carroll, E. L., Bruford, M., DeWoody, J. A., Leroy, G., Strand, A., Waits, L., & Wang, J. (2018). Genetic and genomic monitoring with minimally invasive sampling methods. *Evolutionary Applications*, <https://doi.org/10.1111/eva.12600>
- Cruz-Dávalos, D. I., Llamas, B., Gaunitz, C., Fages, A., Gamba, C., Soubrier, J., ... Alquraishi, S. A. (2017). Experimental conditions improving in-solution target enrichment for ancient DNA. *Molecular Ecology Resources*, 17, 508–522. <https://doi.org/10.1111/1755-0998.12595>
- DeFaveri, J., Viitaniemi, H., Leder, E., & Merila, J. (2013). Characterizing genic and nongenic molecular markers: Comparison of microsatellites and SNPs. *Molecular Ecology Resources*, 13, 377–392. <https://doi.org/10.1111/1755-0998.12071>
- Earl, D. A., & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 4, 359–361. <https://doi.org/10.1007/s12686-011-9548-7>
- Eizirik, E., Murphy, W. J., Koepfli, K. P., Johnson, W. E., Dragoo, J. W., Wayne, R. K., & O'Brien, S. J. (2010). Pattern and timing of diversification of the mammalian order Carnivora inferred from multiple nuclear gene sequences. *Molecular Phylogenetics and Evolution*, 56, 49–63. <https://doi.org/10.1016/j.ympev.2010.01.033>
- Elbers, J. P., Clostio, R. W., & Taylor, S. S. (2017). Population genetic inferences using immune gene SNPs mirror patterns inferred by microsatellites. *Molecular Ecology Resources*, 17, 481–491. <https://doi.org/10.1111/1755-0998.12591>
- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*, 6, e19379. <https://doi.org/10.1371/journal.pone.0019379>
- Enk, J. M., Devault, A. M., Kuch, M., Murgha, Y. E., Rouillard, J.-M., & Poinar, H. N. (2014). Ancient whole genome enrichment using baits built from modern DNA. *Molecular Biology and Evolution*, 31, 1292–1294. <https://doi.org/10.1093/molbev/msu074>
- Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, 14, 2611–2620. <https://doi.org/10.1111/j.1365-294X.2005.02553.x>
- Faircloth, B. C., McCormack, J. E., Crawford, N. G., Harvey, M. G., Brumfield, R. T., & Glenn, T. C. (2012). Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary time-scales. *Systematic Biology*, 61, 717–726. <https://doi.org/10.1093/sysbio/sys004>
- Faircloth, B. C., Sorenson, L., Santini, F., & Alfaro, M. E. (2013). A phylogenomic perspective on the radiation of ray-finned fishes based upon targeted sequencing of ultraconserved elements (UCEs). *PLoS ONE*, 8, e65923. <https://doi.org/10.1371/journal.pone.0065923>
- Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics*, 164, 1567–1587.
- Fitak, R. R., Naidu, A., Thompson, R. W., & Culver, M. (2016). A new panel of SNP markers for the individual identification of North American pumas. *Journal of Fish and Wildlife Management*, 7, 13–27. <https://doi.org/10.3996/112014-JFWM-080>
- Fortes, G. G., & Pajmans, J. L. A. (2015). Analysis of whole mitogenomes from ancient samples. In T. Kroneis (Ed.), *Whole genome amplification. Methods in molecular biology*. New York, USA: Humana Press.

- Frankham, R., Ballou, J. D., & Briscoe, D. A. (2010). *Introduction to conservation genetics*, 2nd ed. Cambridge, UK: Cambridge University Press. <https://doi.org/10.1017/CBO9780511809002>
- Freemo, H., O'Reilly, P., Berg, P. R., Lien, S., & Boulding, E. G. (2011). Outlier SNPs show more genetic structure between two Bay of Fundy metapopulations of Atlantic salmon than do neutral SNPs. *Molecular Ecology Resources*, 11, 254–267. <https://doi.org/10.1111/j.1755-0998.2010.02952.x>
- Gärke, C., Ytornel, F., Bed'hom, B., Gut, I., Lathrop, M., Weigend, S., & Simianer, H. (2012). Comparison of SNPs and microsatellites for assessing the genetic structure of chicken populations. *Animal Genetics*, 43, 419–428. <https://doi.org/10.1111/j.1365-2052.2011.02284.x>
- Garvin, M. R., Saitoh, K., & Gharrett, A. J. (2010). Application of single nucleotide polymorphisms to non-model species: A technical review. *Molecular Ecology Resources*, 10, 915–934. <https://doi.org/10.1111/j.1755-0998.2010.02891.x>
- Glover, K. A., Hansen, M. M., Lien, S., Als, T. D., Høyheim, B., & Skaala, Ø. (2010). A comparison of SNP and STR loci for delineating population structure and performing individual genetic assignment. *BMC Genetics*, 11, 2. <https://doi.org/10.1186/1471-2156-11-2>
- Gnirke, A., Melnikov, A., Maguire, J., Rogov, P., LeProust, E. M., Brockman, W., ... Nusbaum, C. (2009). Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature Biotechnology*, 27, 182–189. <https://doi.org/10.1038/nbt.1523>
- Hancock-Hanser, B. L., Frey, A., Leslie, M. S., Dutton, P. H., Archer, F. I., & Morin, P. A. (2013). Targeted multiplex next-generation sequencing: Advances in techniques of mitochondrial and nuclear DNA sequencing for population genomics. *Molecular Ecology Resources*, 13, 254–268. <https://doi.org/10.1111/1755-0998.12059>
- Harvey, M. G., Smith, B. T., Glenn, T. C., Faircloth, B. C., & Brumfield, R. T. (2016). Sequence capture versus restriction site associated DNA sequencing for shallow systematics. *Systematic Biology*, 65, 910–924. <https://doi.org/10.1093/sysbio/syw036>
- Hauser, L., Baird, M., Hilborn, R. A. Y., Seeb, L. W., & Seeb, J. E. (2011). An empirical comparison of SNPs and microsatellites for parentage and kinship assignment in a wild sockeye salmon (*Oncorhynchus nerka*) population. *Molecular Ecology Resources*, 11, 150–161. <https://doi.org/10.1111/j.1755-0998.2010.02961.x>
- Hawkins, M. T., Leonard, J. A., Helgen, K. M., McDonough, M. M., Rockwood, L. L., & Maldonado, J. E. (2016). Evolutionary history of endemic Sulawesi squirrels constructed from UCEs and mitogenomes sequenced from museum specimens. *BMC Evolutionary Biology*, 16, 80. <https://doi.org/10.1186/s12862-016-0650-z>
- Hedtke, S. M., Morgan, M. J., Cannatella, D. C., & Hillis, D. M. (2013). Targeted enrichment: Maximizing orthologous gene comparisons across deep evolutionary time. *PLoS ONE*, 8, e67908. <https://doi.org/10.1371/journal.pone.0067908>
- Hernandez-Rodriguez, J., Arandjelovic, M., Lester, J., de Filippo, C., Weihmann, A., Meyer, M., ... Marques-Bonet, T. (2018). The impact of endogenous content, replicates and pooling on genome capture from faecal samples. *Molecular Ecology Resources*, 18, 319–333. <https://doi.org/10.1111/1755-0998.12728>
- Hodges, E., Xuan, Z., Balija, V., Kramer, M., Molla, M. N., Smith, S. W., ... McCombie, W. R. (2007). Genome-wide in situ exon capture for selective resequencing. *Nature Genetics*, 39, 1522–1527. <https://doi.org/10.1038/ng.2007.42>
- Hoffberg, S. L., Kieran, T. J., Catchen, J. M., Devault, A., Faircloth, B. C., Mauricio, R., & Glenn, T. C. (2016). RADcap: Sequence capture of dual-digest RADseq libraries with identifiable duplicates and reduced missing data. *Molecular Ecology Resources*, 16, 1264–1278. <https://doi.org/10.1111/1755-0998.12566>
- Jin, X., He, M., Ferguson, B., Meng, Y., Ouyang, L., Ren, J., ... Zhuo, M. (2012). An effort to use human-based exome capture methods to analyze chimpanzee and macaque exomes. *PLoS ONE*, 7, e40637. <https://doi.org/10.1371/journal.pone.0040637>
- Jobin, A., Molinari, P., & Breitenmoser, U. (2000). Prey spectrum, prey preference and consumption rates of Eurasian lynx in the Swiss Jura Mountains. *Acta Theriologica*, 45, 243–252. <https://doi.org/10.4098/0001-7051>
- Johnson, W. E., Eizirik, E., Pecon-Slattery, J., Murphy, W. J., Antunes, A., Teeling, E., & O'Brien, S. J. (2006). The late Miocene radiation of modern Felidae: A genetic assessment. *Science*, 311, 73–77. <https://doi.org/10.1126/science.1122277>
- Jombart, T. (2008). adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24, 1403–1405. <https://doi.org/10.1093/bioinformatics/btn129>
- Jones, M. R., & Good, J. M. (2016). Targeted capture in evolutionary and ecological genomics. *Molecular Ecology*, 25, 185–202. <https://doi.org/10.1111/mec.13304>
- Kaiser, S. A., Taylor, S. A., Chen, N., Sillet, T. S., Bondra, E. R., & Webster, M. S. (2017). A comparative assessment of SNP and microsatellite markers for assigning parentage in a socially monogamous bird. *Molecular Ecology Resources*, 17, 183–193. <https://doi.org/10.1111/1755-0998.12589>
- Kleinman-Ruiz, D., Martínez-Cruz, B., Soriano, L., Lucena-Perez, M., Cruz, F., Villanueva, B., ... Godoy, J. A. (2017). Novel efficient genome-wide SNP panels for the conservation of the highly endangered Iberian lynx. *BMC Genomics*, 18, 556. <https://doi.org/10.1186/s12864-017-3946-5>
- Kraus, R. H., Kerstens, H. H., Van Hooft, P., Crooijmans, R. P., Van Der Poel, J. J., Elmberg, J., ... Groenen, M. A. (2011). Genome wide SNP discovery, analysis and evaluation in mallard (*Anas platyrhynchos*). *BMC Genomics*, 12, 150. <https://doi.org/10.1186/1471-2164-12-150>
- Kraus, R. H. S., VonHoldt, B., Cocchiari, B., Harms, V., Bayerl, H., Kühn, H., ... Nowak, C. (2015). A single-nucleotide polymorphism-based approach for rapid and cost-effective genetic wolf monitoring in Europe based on non-invasively collected samples. *Molecular Ecology Resources*, 15, 295–305. <https://doi.org/10.1111/1755-0998.12307>
- Ku, C. S., Wu, M., Cooper, D. N., Naidoo, N., Pawitan, Y., Pang, B., ... Soong, R. (2012). Exome versus transcriptome sequencing in identifying coding region variants. *Expert Review of Molecular Diagnostics*, 12, 241–251. <https://doi.org/10.1586/erm.12.10>
- Leaché, A. D., Chavez, A. S., Jones, L. N., Grummer, J. A., Gottscho, A. D., & Linkem, C. W. (2015). Phylogenomics of phrynosomatid lizards: Conflicting signals from sequence capture versus restriction site associated DNA sequencing. *Genome Biology and Evolution*, 7, 706–719. <https://doi.org/10.1093/gbe/evv026>
- Lemmon, A. R., Emme, S. A., & Lemmon, E. M. (2012). Anchored hybrid enrichment for massively high-throughput phylogenomics. *Systematic Biology*, 61, 727–744. <https://doi.org/10.1093/sysbio/sys049>
- Li, G., Davis, B. W., Eizirik, E., & Murphy, W. J. (2016). Phylogenomic evidence for ancient hybridization in the genomes of living cats (Felidae). *Genome Research*, 26, 1–11. <https://doi.org/10.1101/gr.186668.114>
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25, 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25, 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Li, C., Hofreiter, M., Straube, N., Corrigan, S., & Naylor, G. J. (2013). Capturing protein-coding genes across highly divergent species. *BioTechniques*, 54, 321–326.
- Li, C., Riethoven, J. J., & Naylor, G. J. (2012). EvolMarkers: A database for mining exon and intron markers for evolution, ecology and conservation studies. *Molecular Ecology Resources*, 12, 967–971. <https://doi.org/10.1111/j.1755-0998.2012.03167.x>

- Lim, H. C., & Braun, M. J. (2016). High-throughput SNP genotyping of historical and modern samples of five bird species via sequence capture of ultraconserved elements. *Molecular Ecology Resources*, 16, 1204–1223. <https://doi.org/10.1111/1755-0998.12568>
- Magoč, T., & Salzberg, S. L. (2011). FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, 27, 2957–2963.
- Manthey, J. D., Campillo, L. C., Burns, K. J., & Moyle, R. G. (2016). Comparison of target-capture and restriction-site associated DNA sequencing for phylogenomics: A test in cardinalid tanagers (Aves, Genus: Piranga). *Systematic Biology*, 65, 640–650. <https://doi.org/10.1093/sysbio/syw005>
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. Journal*, 17, 10–12. <https://doi.org/10.14806/ej.17.1.200>
- McCormack, J. E., Tsai, W. L. E., & Faircloth, B. C. (2016). Sequence capture of ultraconserved elements from bird museum specimens. *Molecular Ecology Resources*, 16, 1189–1203. <https://doi.org/10.1111/1755-0998.12466>
- Meyer, M., & Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols*, 6, pdb-prot5448. <https://doi.org/10.1101/pdb.prot5448>
- Miller, M. R., Dunham, J. P., Amores, A., Cresko, W. A., & Johnson, E. A. (2007). Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research*, 17, 240–248. <https://doi.org/10.1101/gr.5681207>
- Morin, P. A., Archer, F. I., Pease, V. L., Hancock-Hanser, B. L., Robertson, K. M., Huebinger, R. M., ... Taylor, B. L. (2012). An empirical comparison of SNPs and microsatellites for population structure, assignment, and demographic analyses of bowhead whale populations. *Endangered Species Research*, 19, 129–147. <https://doi.org/10.3354/esr00459>
- Morin, P. A., Luikart, G., & Wayne, R. K. (2004). SNPs in ecology, evolution and conservation. *Trends in Ecology & Evolution*, 19, 208–216. <https://doi.org/10.1016/j.tree.2004.01.009>
- Muñoz, I., Henriques, D., Jara, L., Johnston, J. S., Chávez-Galarza, J., De La Rúa, P., & Pinto, M. A. (2017). SNPs selected by information content outperform randomly selected microsatellite loci for delineating genetic identification and introgression in the endangered dark European honeybee (*Apis mellifera mellifera*). *Molecular Ecology Resources*, 17, 783–795.
- Neves, L. G., Davis, J. M., Barbazuk, W. B., & Kirst, M. (2013). Whole-exome targeted sequencing of the uncharacterized pine genome. *The Plant Journal*, 75, 146–156. <https://doi.org/10.1111/tpj.12193>
- Newman, C. E., & Austin, C. C. (2016). Sequence capture and next-generation sequencing of ultraconserved elements in a large-genome salamander. *Molecular Ecology*, 25, 6162–6174. <https://doi.org/10.1111/mec.13909>
- Norman, A. J., & Spong, G. (2015). Single nucleotide polymorphism-based dispersal estimates using noninvasive sampling. *Ecology and Evolution*, 5, 3056–3065. <https://doi.org/10.1002/ece3.1588>
- Norman, A. J., Street, N. R., & Spong, G. (2013). De novo SNP discovery in the Scandinavian brown bear (*Ursus arctos*). *PLoS ONE*, 8, e81012. <https://doi.org/10.1371/journal.pone.0081012>
- Nussberger, B., Wandeler, P., & Camenisch, C. (2014). A SNP chip to detect introgression in wildcats allows accurate genotyping of low quality samples. *European Journal of Wildlife Research*, 60, 405–410. <https://doi.org/10.1007/s10344-014-0806-3>
- Oliveira, R., Randi, E., Mattucci, F., Kurushima, J. D., Lyons, L. A., & Alves, P. C. (2015). Toward a genome-wide approach for detecting hybrids: Informative SNPs to detect introgression between domestic cats and European wildcats (*Felis silvestris*). *Heredity*, 115, 195–205. <https://doi.org/10.1038/hdy.2015.25>
- Olson, M. (2007). Enrichment of super-sized resequencing targets from the human genome. *Nature Methods*, 1, 891–892. <https://doi.org/10.1038/nmeth1107-891>
- Pajmians, J. L. A., Fickel, J., Courtiol, A., Hofreiter, M., & Förster, D. (2016). Impact of enrichment conditions on cross-species capture of fresh and degraded DNA. *Molecular Ecology Resources*, 16, 42–55. <https://doi.org/10.1111/1755-0998.12420>
- Peñalba, J. V., Smith, L. L., Tonione, M. A., Sass, C., Hykin, S. M., Skipwith, P. L., ... Moritz, C. (2014). Sequence capture using PCR-generated probes: A cost-effective method of targeted high-throughput sequencing for nonmodel organisms. *Molecular Ecology Resources*, 14, 1000–1010.
- Perry, G. H., Marioni, J. C., Melsted, P., & Gilad, Y. (2010). Genomic-scale capture and sequencing of endogenous DNA from feces. *Molecular Ecology*, 19, 5332–5344. <https://doi.org/10.1111/j.1365-294X.2010.04888.x>
- Portik, D. M., Smith, L. L., & Bi, K. (2016). An evaluation of transcriptome-based exon capture for frog phylogenomics across multiple scales of divergence (Class: Amphibia, Order: Anura). *Molecular Ecology Resources*, 16, 1069–1083. <https://doi.org/10.1111/1755-0998.12541>
- Powell, J. H., Amish, S. J., Haynes, G. D., Luikart, G., & Latch, E. K. (2016). Candidate adaptive genes associated with lineage divergence: Identifying SNPs via next-generation targeted resequencing in mule deer (*Odocoileus hemionus*). *Molecular Ecology Resources*, 16, 1165–1172. <https://doi.org/10.1111/1755-0998.12572>
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155, 945–959.
- Puritz, J. B., & Lotterhos, K. E. (2018). Expressed Exome Capture Sequencing (EecSeq): A method for cost-effective exome sequencing for all organisms. *Molecular Ecology Resources*, <https://doi.org/10.1111/1755-0998.12905>
- Ratkiewicz, M., Matusiuk, M., Saveljev, A. P., Sidorovich, V., Ozolins, J., Männil, P., ... Schmidt, K. (2014). Long-range gene flow and the effects of climatic and ecological factors on genetic structuring in a large, solitary carnivore: The Eurasian lynx. *PLoS ONE*, 9, e115160. <https://doi.org/10.1371/journal.pone.0115160>
- Roffler, G. H., Amish, S. J., Smith, S., Cosart, T., Kardos, M., Schwartz, M. K., & Luikart, G. (2016). SNP discovery in candidate adaptive genes using exon capture in a free-ranging alpine ungulate. *Molecular Ecology Resources*, 16, 1147–1164. <https://doi.org/10.1111/1755-0998.12560>
- Rousset, F. (2008). Genepop'007: A complete reimplementations of the Genepop software for Windows and Linux. *Molecular Ecology Resources*, 8, 103–106. <https://doi.org/10.1111/j.1471-8286.2007.01931.x>
- Ruane, S., & Austin, C. C. (2017). Phylogenomics using formalin-fixed and 100+ year-old intractable natural history specimens. *Molecular Ecology Resources*, <https://doi.org/10.1111/1755-0998.12655>
- Schmid, S., Genevest, R., Gobet, E., Suchan, T., Sperisen, C., Tinner, W., & Alvarez, N. (2017). HyRAD-X, a versatile method combining exome capture and RAD sequencing to extract genomic information from ancient DNA. *Methods in Ecology and Evolution*, 8, 1374–1388. <https://doi.org/10.1111/2041-210X.12785>
- Schubert, M., Ginolhac, A., Lindgreen, S., Thompson, J. F., Al-Rasheid, K. A. S., Willerslev, E., ... Orlando, L. (2012). Improving ancient DNA read mapping against modern reference genomes. *BMC Genomics*, 13, 178. <https://doi.org/10.1186/1471-2164-13-178>
- Schwartz, M. K., Luikart, G., & Waples, R. S. (2007). Genetic monitoring as a promising tool for conservation and management. *Trends in Ecology & Evolution*, 22, 25–33. <https://doi.org/10.1016/j.tree.2006.08.009>
- Seeb, J. E., Carvalho, G., Hauser, L., Naish, K., Roberts, S., & Seeb, L. W. (2011). Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms. *Molecular Ecology Resources*, 11, 1–8. <https://doi.org/10.1111/j.1755-0998.2010.02979.x>

- Shin, J. H., Blay, S., McNeney, B., & Graham, J. (2006). LDheatmap: An R function for graphical display of pairwise linkage disequilibria between single nucleotide polymorphisms. *Journal of Statistical Software*, 16, 1–10.
- Smith, B. T., Harvey, M. G., Faircloth, B. C., Glenn, T. C., & Brumfield, R. T. (2014). Target capture and massively parallel sequencing of ultra-conserved elements for comparative studies at shallow evolutionary time scales. *Systematic Biology*, 63, 83–95. <https://doi.org/10.1093/sysbio/syt061>
- Springer, M. S., Signore, A. V., Pajmans, J. L. A., Vélez-Juarbe, J., Domning, D. P., Bauer, C. E., ... Campbell, K. L. (2015). Interordinal gene capture, the phylogenetic position of Steller's sea cow based on molecular and morphological data, and the macroevolutionary history of Sirenia. *Molecular Phylogenetics and Evolution*, 91, 178–193. <https://doi.org/10.1016/j.ympev.2015.05.022>
- Suchan, T., Pitteloud, C., Gerasimova, N. S., Kostikova, A., Schmid, S., Arrigo, N., ... Alvarez, N. (2016). Hybridization capture using RAD probes (hyRAD), a new tool for performing genomic analyses on collection specimens. *PLoS ONE*, 11, e0151651. <https://doi.org/10.1371/journal.pone.0151651>
- Suren, H., Hodgins, K. A., Yeaman, S., Nurkowski, K. A., Smets, P., Rieseberg, L. H., ... Holliday, J. A. (2016). Exome capture from the spruce and pine giga-genomes. *Molecular Ecology Resources*, 16, 1136–1146. <https://doi.org/10.1111/1755-0998.12570>
- Templeton, J. E. L., Brotherton, P. M., Llamas, B., Soubrier, J., Haak, W., Cooper, A., & Austin, J. J. (2013). DNA capture and next-generation sequencing can recover whole mitochondrial genomes from highly degraded samples for human identification. *Investigative Genetics*, 4, 26. <https://doi.org/10.1186/2041-2223-4-26>
- Tewhey, R., Nakano, M., Wang, X., Pabón-Peña, C., Novak, B., Giuffrè, A., ... Topol, E. J. (2009). Enrichment of sequencing targets from the human genome by solution hybridization. *Genome Biology*, 10, R116. <https://doi.org/10.1186/gb-2009-10-10-r116>
- Thaden, V., Cocchiararo, B., Jarausch, A., Jüngling, H., Karamanlis, A. A., Tiesmeyer, A., ... Munoz-Fuentes, V. (2017). Assessing SNP genotyping of noninvasively collected wildlife samples using microfluidic arrays. *Scientific Reports*, 7, 10768. <https://doi.org/10.1038/s41598-017-10647-w>
- Tokarska, M., Marshall, T., Kowalczyk, R., Wójcik, J. M., Pertoldi, C., Kristensen, T. N., ... Bendixen, C. (2009). Effectiveness of microsatellite and SNP markers for parentage and identity analysis in species with low genetic diversity: The case of European bison. *Heredity*, 103, 326–332. <https://doi.org/10.1038/hdy.2009.73>
- Valière, N. (2002). GIMLET: A computer program for analysing genetic individual identification data. *Molecular Ecology Notes*, 2, 377–379.
- Vallender, E. J. (2011). Expanding whole exome resequencing into nonhuman primates. *Genome Biology*, 12, R87. <https://doi.org/10.1186/gb-2011-12-9-r87>
- Von Arx, M., Breitenmoser-Würsten, C., Zimmermann, F., & Breitenmoser, U. (2004). *Status and conservation of the Eurasian lynx (Lynx lynx)* in 2001. Muri: KORA Bericht no 19. 330 pp.
- Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-seq: A revolutionary tool for transcriptomics. *Nature Reviews Genetics*, 10, 57–63. <https://doi.org/10.1038/nrg2484>
- Weinman, L. R., Solomon, J. W., & Rubenstein, D. R. (2015). A comparison of single nucleotide polymorphism and microsatellite markers for analysis of parentage and kinship in a cooperatively breeding bird. *Molecular Ecology Resources*, 15, 502–511. <https://doi.org/10.1111/1755-0998.12330>
- Wozencraft, W. C. (2005). Order Carnivora. In D. E. Wilson, & D. M. Reeder (Eds.), *Mammal species of the world: A taxonomic and geographic reference*, 3rd ed. Baltimore, USA: The Johns Hopkins University Press.
- Yi, X., Liang, Y., Huerta-Sanchez, E., Jin, X., Cuo, Z. X., Pool, J. E., ... Wang, J. (2010). Sequencing of fifty human exomes reveals adaptation to high altitude. *Science*, 329, 75–78. <https://doi.org/10.1126/science.1190371>
- Yoshihara, M., Saito, D., Sato, T., Ohara, O., Kuramoto, T., & Suyama, M. (2016). Design and application of a target capture sequencing of exons and conserved noncoding sequences for the rat. *BMC Genomics*, 17, 593. <https://doi.org/10.1186/s12864-016-2975-9>
- Yuan, H., Jiang, J., Jiménez, F. A., Hoberg, E. P., Cook, J. A., Galbreath, K. E., & Li, C. (2016). Target gene enrichment in the cyclophyllidean cestodes, the most diverse group of tapeworms. *Molecular Ecology Resources*, 16, 1095–1106. <https://doi.org/10.1111/1755-0998.12532>
- Zhan, X., Dixon, A., Batbayar, N., Bragin, E., Ayas, Z., Deutschova, L., ... Bruford, M. W. (2015). Exonic versus intronic SNPs: Contrasting roles in revealing the population genetic differentiation of a widespread bird species. *Heredity*, 114, 1–9. <https://doi.org/10.1038/hdy.2014.59>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Förster DW, Bull JK, Lenz D, et al. Targeted resequencing of coding DNA sequences for SNP discovery in nonmodel species. *Mol Ecol Resour*. 2018;00: 1–18. <https://doi.org/10.1111/1755-0998.12924>