RESOURCE ARTICLE

A robust sequencing assay of a thousand amplicons for the high-throughput population monitoring of Alpine ibex immunogenetics

Camille Kessler[1] | Alice Brambilla[2,3] | Dominique Waldvogel[2] | Glauco Camenisch[2] | Iris Biebach[2] | Deborah M. Leigh[2,4] [ID] | Christine Grossen[2] | Daniel Croll[1] [ID]

[1]Laboratory of Evolutionary Genetics, Institute of Biology, University of Neuchâtel, Neuchâtel, Switzerland

[2]Department of Evolutionary Biology and Environmental Studies, University of Zürich, Zürich, Switzerland

[3]Alpine Wildlife Research Center, Gran Paradiso National Park, Italy

[4]WSL Swiss Federal Research Institute, Birmensdorf, Switzerland

Correspondence
Christine Grossen, Department of Evolutionary Biology and Environmental Studies, University of Zürich, Zürich, Switzerland.
Email: christine.grossen@uzh.ch

Daniel Croll, Laboratory of Evolutionary Genetics, Institute of Biology, University of Neuchâtel, Neuchâtel, Switzerland.
Email: daniel.croll@unine.ch

Funding information
Interreg Alcotra, Grant/Award Number: LEMED-IBEX; Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung, Grant/Award Number: 198147

Abstract

Polymorphism for immune functions can explain significant variation in health and reproductive success within species. Drastic loss in genetic diversity at such loci constitutes an extinction risk and should be monitored in species of conservation concern. However, effective implementations of genome-wide immune polymorphism sets into high-throughput genotyping assays are scarce. Here, we report the design and validation of a microfluidics-based amplicon sequencing assay to comprehensively capture genetic variation in Alpine ibex (*Capra ibex*). This species represents one of the most successful large mammal restorations recovering from a severely depressed census size and a massive loss in diversity at the major histocompatibility complex (MHC). We analysed 65 whole-genome sequencing sets of the Alpine ibex and related species to select the most representative markers and to prevent primer binding failures. In total, we designed ~1,000 amplicons densely covering the MHC, further immunity-related genes as well as randomly selected genome-wide markers for the assessment of neutral population structure. Our analysis of 158 individuals shows that the genome-wide markers perform equally well at resolving population structure as RAD-sequencing or low-coverage genome sequencing data sets. Immunity-related loci show unexpectedly high degrees of genetic differentiation within the species. Such information can now be used to define highly targeted individual translocations. Our design strategy can be realistically implemented into genetic surveys of a large range of species. In conclusion, leveraging whole-genome sequencing data sets to design targeted amplicon assays allows the simultaneous monitoring of multiple genetic risk factors and can be translated into species conservation recommendations.

KEYWORDS
adaptation, conservation genetics, genomics/proteomics, population genetics - empirical, wildlife management

Christine Grossen and Daniel Croll contributed equally to this work.

1 | INTRODUCTION

Biodiversity is currently undergoing a dramatic decline caused by ecological and anthropological pressures (Barnosky et al., 2011; Ceballos et al., 2015, 2017; WWF, 2018). Endangered species are particularly prone to genetic risks due to population bottlenecks and habitat fragmentation that lead to: genetic diversity loss, inbreeding, introgression from related species and accumulation of deleterious mutations (Allendorf et al., 2010; Frankham, 2005). Genetic diversity is generally thought to be declining in wild populations (Leigh et al., 2019). Low genetic diversity can affect fitness and survival (Reed & Frankham, 2003), as shown for instance in cheetah (*Acinonyx jubatus*, O'Brien et al., 1983; O'Brien et al., 1985), Florida panthers (*Puma concolor coryi*, Pimm et al., 2006; Roelke et al., 1993) and Alpine ibex (*Capra ibex*, Brambilla et al., 2015). Loss of genetic diversity at adaptive immune loci is particularly problematic because populations will lack genetic variants conferring disease resistance, and this has been shown to be problematic in amphibians (Kosch et al., 2019; Savage et al., 2011), Tasmanian devils (*Sarcophilus harrisii*, Siddle et al., 2007), giant panda (*Ailuropoda melanoleuca*, Zhu et al., 2020) and Alpine ibex (Brambilla et al., 2018). However, despite the importance for conservation management, the genetic underpinnings of disease susceptibility are largely unknown in most nonmodel species (Schoville et al., 2012). The major obstacles in investigating wildlife immunogenetics are the challenge of collecting quantitatively and qualitatively adequate phenotypic data on diseases in wild populations, as well as the lack of genetic tools suitable in the conservation framework (Holderegger et al., 2019). Hence, there is a pressing need to establish effective genetic monitoring tools for many wild species (Acevedo-Whitehouse & Cunningham, 2006; Allendorf et al., 2010).

The major histocompatibility complex (MHC), a highly polymorphic region involved in foreign antigen recognition, is an important player for disease susceptibility in vertebrates. Co-evolution with pathogens often causes balancing selection on MHC polymorphism, maintaining high genetic diversity including deeply divergent alleles. Genetic variants under selection often encode the ability to present a wider range of antigens to T-cells and, thus, to recognize a greater variety of parasites. Consequently, heterozygotes often have elevated resistance and are favoured by selection (Bernatchez & Landry, 2003). Species that have undergone a strong bottleneck such as Alpine and Iberian ibex (*Capra pyrenaica*), Tasmanian devils, cheetahs or Galapagos penguins (*Spheniscus mendiculus*) show strongly reduced genetic diversity at the MHC compared to related species (Angelone et al., 2018; Bollmer et al., 2007; Brambilla et al., 2018). Low genetic diversity in cheetahs is thought to have contributed to high mortality rates (O'Brien et al., 1983, 2017). High prevalence of a fatal facial cancer in Tasmanian devil is thought to, at least, partially stem from low MHC diversity (Siddle et al., 2007). Low levels of MHC diversity are consequently a major threat to endangered species. Alongside with the adaptive immune system, the innate immune system plays an important role in the defence against a wide range of pathogens. Genetic variation within some genes, such as Toll-like receptors (TLRs), has been shown to be linked to variation in immune competence (Ammerdorffer et al., 2014; Tschirren et al., 2013). Yet the field of wildlife immunogenetics is only just emerging and immune-related genes outside of the MHC remain heavily understudied in wild species. The major challenge for monitoring nonmodel wild species is that inferring causal disease susceptibility from other species is only rarely possible and associations need to be established for each species–pathogen interaction (Acevedo-Whitehouse & Cunningham, 2006).

To make informed conservation management decisions, the genetic health of populations should be taken in account (Allendorf et al., 2010; Biebach et al., 2016). Yet integrating information about immune-related polymorphisms and genetic diversity across the genome remains a major challenge. Conservation genetics studies have traditionally relied on microsatellite markers (Allendorf et al., 2010; Brambilla et al., 2015; Ouborg et al., 2010; Witzenberger & Hochkirch, 2014), which are inexpensive and relatively easy to use. However, only in exceptional cases do microsatellites reveal adaptive genetic variation and are thus not applicable to most relevant immune-related polymorphism (Allendorf et al., 2010; Steiner et al., 2013). Similarly, conservation genomics methods based on restriction site-associated DNA sequencing (RAD; e.g., ddRAD, GBS, Davey et al., 2011) intrinsically lack the ability to target particular polymorphisms of interest (Garvin et al., 2010). Untargeted single nucleotide polymorphism (SNP) genotyping can also be achieved through the capture of RAD tags as in RAPTURE (Ali et al., 2016; Komoroske et al., 2019; Sard et al., 2020) and RADCAP (Hoffberg et al., 2016). While the development and sample preparation time to capture RAD tags is short, the possibility to target specific loci (e.g., putative adaptive variation) is highly limited (reviewed in Meek & Larson, 2019). Whole genome sequencing resolves the issue of marker selection but is expensive, requires high-quality input material to work reliably and imposes significant computational challenges due to the very large amounts of data (Davey et al., 2011). This is because most genome sequencing techniques are still relatively costly, and require bioinformatic expertise and powerful computing systems. These are major impediments for the successful implementation in conservation efforts (Shafer et al., 2015; Taylor et al., 2017).

Targeted SNP-based genotyping overcomes many of the constraints of the above-mentioned methodologies, because the approach typically tolerates a wide range of input material and is cost-effective (Eriksson et al., 2020; Kleinman-Ruiz et al., 2017; Natesh et al., 2019; Wright et al., 2015). Early applications of targeted SNP genotyping methods relied on hybridization arrays and fluorescent probes. Applications such as TaqMan, Sequenom and SNaPshot arrays have enabled the analyses of tens to hundreds of SNPs in hundreds of samples (reviewed in Syvänen, 2001). The targeted sequencing of amplicons evolved from array-based systems and takes advantage of the large amounts of sequencing data produced by next-generation sequencers. Nonetheless, two important limitations apply to existing targeted sequencing approaches. First, unknown polymorphisms in flanking sequences of the targeted SNPs can lead to amplification dropouts. This shortcoming can be remedied by integrating whole-genome sequencing information

from a limited number of individuals. Second, the parallel amplification of a large number of loci requires optimization and normalization steps to ensure balanced amounts of PCR product for each locus-by-individual combination (Meek & Larson, 2019).

Amplicon sequencing by genotyping-in-thousands (GT-seq) offers a highly flexible and broad approach to analyse hundreds of markers (typically 200–500; Campbell et al., 2015; reviewed in Meek & Larson, 2019). GT-seq saw broad applications across systems including in large-scale salmonid surveys (Matala et al., 2016; May et al., 2020). However, the development of a new GT-seq assay requires significant efforts to optimize multiplex amplifications. A number of similar methods for amplicon sequencing have also been developed (e.g., MTA-seq, three-round multiplex, Onda et al. 2018; Chen et al. 2016). Some alternatives to GT-seq rely partially on commercial solutions, including AmpliSeq, TruSeq Custom Amplicon and the microfluidics-based Juno system. Microfluidics technology is notable because it allows for parallel multiplex amplifications and automated sample–primer pool pairing in miniaturized wells. This lowers overall reaction volumes, ensures accurate mixing and reaction conditions, and reduces primer incompatibility issues. These factors stabilize amplification yields across amplicons and, as a consequence, increase the total number of usable amplicons per assay (Chiu et al., 2017; Hess et al., 2020).

Here, we develop and validate a high-throughput amplicon sequencing assay based on a microfluidics platform, which enables the simultaneous monitoring of population structure, immunogenetics and introgression from a domestic relative in Alpine ibex. The species is an excellent model to assess the broad applicability of novel genotyping assays for conservation. Because of a dramatic recent bottleneck caused by a near extinction event in the 19th century (Grodinsky & Stüwe, 1987), the census size was reduced to fewer than 100 individuals limited to one single population in northern Italy (Gran Paradiso). A captive breeding programme restored the species to a current census size of over 50,000 individuals spread across the European Alps (Brambilla et al., 2020; Toïgo et al., 2020). However, the historical bottleneck left substantial genome-wide signatures of low heterozygosity, in particular at MHC loci, which may threaten long-term species viability (Grossen et al., 2014, 2018, 2020). Hybridization events with domestic goat (*Capra hircus*) produced introgression tracts at the MHC re-establishing some genetic variation, which may have been lost due to the species bottleneck (Grossen et al., 2014). Recently, concerns were raised over population declines as direct and indirect consequences of epizootic disease outbreaks (e.g., sarcoptic mange, respiratory diseases, infectious keratoconjunctivitis, brucellosis). A notable brucellosis outbreak was recently observed in the area of Bargy in France. Cases of brucellosis reported in cattle and humans in the region were allegedly traced back to the ibex population, so that French authorities undertook a massive eradication programme leading to the culling of more than 250 individuals (44% of the estimated population) within 2 years following the outbreak (Mick et al., 2014; Quéméré et al., 2020). In summary, Alpine ibex represent a model species where high-throughput monitoring of inbreeding levels, potential hybridization events and immunogenetic diversity can substantially improve genetic analyses and conservation management efforts.

We established a high-throughput assay of nearly 1,000 loci covering largely genome-wide (putatively neutral) polymorphisms, variants found at the MHC and other immune-related loci, as well as diagnostic variants useful to detect recent hybridization events with domestic goat. We used 51 whole-genome sequences of Alpine ibex and domestic goat to identify relevant SNPs. Furthermore, we used whole genome sequences to mask polymorphism at primer sites to maximize amplification success. Based on highly parallel Illumina amplicon sequencing of 172 Alpine ibex, Iberian ibex and domestic goat samples, we assessed the accuracy and robustness of the assay across variable input sample quality. Finally, we compared the high-throughput assay with RAD-seq and low-coverage whole-genome sequencing data sets on the same populations.

2 | MATERIALS AND METHODS

2.1 | Collection of samples and DNA extraction

Genotyping was performed on 158 Alpine ibex samples, representing eight populations, each with 19–20 sampled individuals. The samples also included an individual suspected to be an albino. Sample material consisted either of tissue or blood. Samples were collected by biopsy darting or during captures carried out in the framework of the EU Interreg-Alcotra 1664 LEMED-ibex project. Further samples were previously collected during captures or legal hunting (Biebach & Keller, 2009; Brambilla et al., 2015; Willisch et al., 2012). In addition to the 158 samples described above, we evaluated the performance of four DNA samples collected from faeces of Alpine ibex in the Gran Paradiso National Park (Italy). The faecal samples originated from two mother–kid pairs sampled on the same day. Visual identification and microsatellite analyses helped confirm that the samples were from different individuals (Willisch et al., 2012). No replicates were performed for these samples in our study. We added six further tissue samples from suspected hybrids (based on field observations), five tissue samples from domestic goats and three from Iberian ibex. Detailed information about the origin, collection method and sampling year of each sample is provided in Table S1. Tissue samples were stored in 95%–100% ethanol at room temperature. Blood samples were collected in EDTA Vacutainer tubes and stored at −20℃. Faecal samples were collected opportunistically immediately after defecation, frozen within 3–4 hr and stored at −20℃ until analysis. DNA extractions were carried out using the DNeasy Blood & Tissue and the QIAmp DNA Stool kits (Qiagen) following the manufacturer's protocol.

DNA concentrations of all samples were assessed using a Qubit fluorometer assay (ThermoFisher). The manufacturer's protocol recommends DNA concentrations of 100–200 ng µl^{-1} for downstream applications. Because these concentrations could not be reached for most of the samples, we chose an *ad hoc* concentration of 50 ng µl^{-1} for normalization of all samples. Samples with initial concentrations

below 50 ng μl^{-1} were purified with AMPure magnetic beads (Beckman Coulter) following the manufacturer's protocol and eluted in smaller volumes to achieve the required final concentration. Samples, for which the standard concentration could not be reached despite AMPure concentration, were also processed but excluded for assay validation steps (see Table S1). To explore effects of low DNA input, we diluted three of the Alpine ibex samples in a dilution series starting from 100 ng μl^{-1} down to 20 ng μl^{-1} (5-fold), 4 ng μl^{-1} (25-fold), 0.8 ng μl^{-1} (125-fold) and 0.16 ng μl^{-1} (625-fold). We selected samples with a concentration of at least 100 ng μl^{-1}, among which we included one corresponding to an individual for which whole-genome sequencing data are available and two were randomly chosen.

2.2 | SNP discovery based on whole-genome sequenced individuals across species

To identify SNPs segregating within the species, we used 29 Alpine ibex whole genome sequences (representing seven different populations). Twenty-two additional genome sequences (16 domestic goats and six bezoar, *Capra aegagrus*) were used to identify species diagnostic markers for introgression analyses. Furthermore, 14 additional genome sequences (four Iberian ibex; two Siberian ibex, *C. sibirica*; two Nubian ibex, *C. nubiana*; one Markhor, *C. falconeri* and five sheep, *Ovis* sp., see Grossen et al., 2020 for details) were included in the analyses to mask polymorphic positions in the domestic goat reference genome (see below). Data from domestic goat, bezoar and sheep were produced by the NextGen Consortium focusing on cattle, sheep and goat genome sequencing (https://www.epfl.ch/labs/nextgen). Raw reads were trimmed using TRIMMOMATIC version 0.36 (Bolger et al., 2014). Trimmed reads were mapped with default settings using BWA-MEM (Li et al., 2009) to the domestic goat reference genome (version ARS1, Bickhart et al., 2017) and duplicates were marked using MARKDUPLICATES from PICARD (http://broadinstitute.github.io/picard, version 1.130). After genotype calling using HAPLOTYPECALLER and GENOTYPEGVCF (McKenna et al., 2010; GATK, version 4.0.8.0, Van der Auwera et al., 2013), SNPs were removed using VariantFiltration of GATK if: QD < 2.0, FS > 40.0, SOR > 5.0, MQ < 20.0, −3.0 > MQRankSum > 3.0, −3.0 > ReadPosRankSum > 3.0 and AN < 46 (80% of all Alpine ibex individuals). Genome-wide polymorphism was used in three ways to design amplicons for the high-throughput sequencing assay: (a) to discover SNPs variable in Alpine ibex; (b) to discover SNPs with fixed alleles between Alpine ibex and domestic goats (i.e., diagnostic markers); and (c) to mask polymorphic sites near targeted SNPs to prevent designing primers spanning polymorphic sites potentially causing amplification drop-outs.

2.3 | SNP effect prediction and selection

We annotated SNPs using SNPEFF version 4.3t 2017-11-24 (Cingolani et al., 2012) with gene annotations produced for the domestic goat genome ARS1 as reference (Bickhart et al., 2017). SNPs located in repeat-masked regions of the reference genome were excluded to avoid designing amplicons in repetitive regions. We designed three main categories of marker sets:

2.3.1 | Marker type 1—Genome-wide, putatively neutral markers

For the design of genome-wide markers, we focused on SNPs segregating only among Alpine ibex individuals from the Gran Paradiso population, the source of all current Alpine ibex populations (four million SNPs). Our aim was to most accurately reflect segregating neutral polymorphism in the species. Hence, including nonsource individuals could lead to ascertainment bias. Loci were selected with a genotyping rate of ≥70%, a minimal genotype quality of 20 and at a minimal distance of 2.3 Mb between adjacent SNPs (in vcftools, Danecek et al., 2011).

2.3.2 | Marker type 2—Diagnostic markers to detect hybridization

To detect recent hybridization events, we identified loci with fixed alleles distinguishing domestic goats and Alpine ibex. Specifically, we chose loci with fixed allele frequency differences between all Alpine ibex and all goats (both domestic goat and bezoar), with a minimal genotyping rate of 80%, a minimal genotype quality of 20 and a minimal distance between adjacent SNPs of 13 Mb (in vcftools, Danecek et al., 2011).

2.3.3 | Marker type 3—Immunogenetic markers

Markers in genes relevant to the immune system were selected for two subgroups (here called MHC region and Immune, outside MHC) with a minimal allele count of 1 among Alpine ibex.

2.3.4 | 3a Immunogenetic markers in the MHC region

Because of the importance of the MHC region for immune functions and evidence for introgression from the domestic goat, we covered the entire MHC region on chromosome 23 (positions 20,892,916–23,588,623 bp). The chromosomal location of the MHC was identified using BLASTN version 2.7.1+ (Altschul et al., 1990) based on gene sequences reported as belonging to the MHC in a previous goat genome assembly version (CHIR1, Dong et al., 2013; Table S17). Additionally, we used gene ontology (GO) and gene homology to search for all MHC-related genes on chromosome 23. Matching gene sequences within 250 kb of the homologous MHC region of the CHIR1 assembly were considered as part of the MHC region

for further analyses. We split the MHC region into 270 windows of 10 kb using BEDTOOLS version 2.27.1 (Quinlan & Hall, 2010) to identify SNPs for the amplicon assay. Evidence for introgression from domestic goat is particularly strong at the *DRB* exon 2 (Grossen et al., 2014). We hence designed the amplicon assay with more dense SNPs in the *DRB* region (positions 23,411,211–23,511,211). The *DRB* region was localized between positions 23,451,944–23,470,477 using BLASTN version 2.7.1+ (Altschul et al., 1990) and the *DRB* sequence provided by Dong et al. (2013). A buffer zone of 40,733 bp before and after the gene was added thereby also including the gene ENSCHIG00000008942, which encodes immune-related functions. We divided the *DRB* region into 20 windows of 5 kb. Using R, we randomly selected a single SNP from every window in the defined MHC and *DRB* by prioritizing MODERATE impact mutations based on SNPEFF and SNPs in coding regions. We manually selected a marker in the *DRB* exon 2 at position 23,460,796 bp.

2.3.5 | 3b Other immunogenetic markers outside MHC region

We targeted immune-related genes, outside of the MHC region on chromosome 23, based on a candidate gene approach using GO, gene homology and literature reports (Acevedo-Whitehouse & Cunningham, 2006; Turner et al., 2011). We searched for all MHC-related GO terms on Ensembl Biomart in the goat genome. Then, we performed INTERPROSCAN version 5.31–70.0 analyses and searched for protein domain annotations matching the term "histocompatibility." For the lists by Acevedo-Whitehouse and Cunningham (2006) and Turner et al. (2011), we focused on 211 genes encoding important immune functions for mammals including interferons, interleukins, TLRs and MHC-related genes outside of the MHC region on chromosome 23. We prioritized SNPs in coding sequences with a moderate impact prediction based on SNPEFF and retained only a single SNP per gene (Table S2).

2.4 | Amplicon design

We extracted a 1,001-bp sequence from the domestic goat reference genome centred around each target SNP using BCFTOOLS version 1.9 (https://www.htslib.org/). The extracted sequences had masked positions for all repetitive regions (masked reference genome) and positions which were polymorphic in Alpine ibex and/or with a minor allele count of three among all other *Capra* species (SAMTOOLS version 1.9 with option -m; Li et al., 2009). Sequences with more than 50% masked bases or sequences overlapping between different amplicon sets were excluded. We obtained a set of 1,589 sequences for assay primer design by Fluidigm. About 25% of the sequences were rejected by Fluidigm due to the lack of primer options matching the design criteria, leaving 1,265 sequences for oligonucleotide primer synthesis. The targeted amplicon length was 200 bp.

2.5 | Targeted DNA sequencing library preparation and SNP calling

Libraries were prepared following the manufacturer's protocol for the Juno LP 192.24 integrated fluidic circuits plate (IFC). After loading all reagents on the IFC, target amplicons were generated for each sample through PCR amplification on a specialized thermocycler (Juno system; Fluidigm). A total of 1,265 primer pairs subdivided into 24 multiplex pools were combined in the IFC with each reagent mix consisting of a DNA sample and an individual barcode. After amplification, samples were pooled into a single tube and purified thrice using AMPure XP magnetic beads (Beckman Coulter) to ensure removal of excess primers before adapter ligation. Finally, sequencing adapters were added by PCR to the purified library followed by a final round of purification according to the manufacturer's protocol. The quantity and quality of the library were assessed using a Qubit fluorometer assay (ThermoFisher) and a 4200 TapeStation electrophoresis instrument (Agilent). The final library was sequenced on a single lane of a NextSeq 500 system (Illumina) in midoutput mode adding ~30% PhiX to avoid potential problems due to low sequence complexity. We demultiplexed raw read data using BCL2FASTQ version 2.19.0.316 and used TRIMMOMATIC version 0.38 (Bolger et al., 2014) for quality trimming. Forward and reverse reads were merged using FLASH version 1.2.11 (Magoč & Salzberg, 2011) and aligned to the goat reference genome ARS1 using BOWTIE2 version 2.3.5 (Langmead & Salzberg, 2012). Read depths for each step were estimated with MULTIQC version 1.7 (Ewels et al., 2016). We called SNPs using HAP-LOTYPECALLER, COMBINEGVCFS and GENOTYPEGVCFS from GATK version 4.0.1 (McKenna et al., 2010; Van der Auwera et al., 2013). Variant sites were excluded if matching any of the following conditions: QD < 5, MQ < 20, −2 > ReadPosRankSum > 2, −2 > MQRankSum > 2, −2 > BaseQRankSum > 2.

2.6 | Marker system performance in Alpine ibex populations

To evaluate key performance metrics of the assay, we compared the outcome against two major classes of current population genomics sequencing approaches: RAD-seq and low-coverage whole genome sequencing. We used data sets reporting analyses on Alpine ibex populations (Grossen et al., 2018, 2020; Leigh et al., 2018) including the founder population of Gran Paradiso and the three Swiss populations Pleureur, Brienzer Rothorn and Albris. Because high-coverage whole genome sequencing is not feasible for large-scale applications, we generated realistic, low-coverage data sets at approximately 1× coverage. For this, we downsampled 15 whole genome Illumina sequencing data sets (Grossen et al., 2020). We used SAMBAMBA version 0.6.6 (Tarasov et al., 2015) to downsample individual bamfiles to a fraction of 0.05 (producing a final coverage of ~1×). We used the software ANGSD (Korneliussen et al., 2014) to calculate genotype likelihoods with the following options: -doGlf 2, -doMajorMinor 1, -doMaf 1, -minMaf 0.05, -SNP_pval 1e-6, -minMapQ 20, -minQ 20,

-skipTriallelic 1, -uniqueOnly 1, -remove_bads 1, -only_proper_pairs 1. The resulting likelihoods were used to run PCANGSD (Meisner & Albrechtsen, 2018) and NGSADMIX (Skotte et al., 2013). For the RAD-seq data set, we first trimmed reads using TRIMMOMATIC version 3.6 (Bolger et al., 2014) and performed read mapping using HISAT2 version 2.1 (Kim et al., 2019) on the ARS1 reference genome. We deduplicated bam files using MARKDUPLICATES from PICARD version 2.5 (http://broad institute.github.io/picard/) and called SNPs on all autosomes (1–29) using the GATK version 4.1 pipeline with HAPLOTYPECALLER, GENOMICSD-BIMPORT and GENOTYPEGVCFS (McKenna et al., 2010). SNPs were flagged using the GATK VariantFiltration tool and removed if any of the conditions were matched: QD < 2.0, FS > 60.0, SOR > 3.0, MQ < 30.0, −12.5 > MQRankSum > 12.5 and −8.0 > ReadPosRankSum. Next, we filtered for SNPs falling within 100 bp of an *Sbf*I restriction cut site identified by *in silico* analyses with the ENSEMBL tool *restrict* (Yates et al., 2019). To compare genotyping performance between RAD-seq and the amplicon sequencing data sets, we filtered for a minimal individual genotyping rate of 0.5 and kept polymorphic sites only. For all further analysis (population differentiation), we kept only bi-allelic SNPs with a minimal genotyping rate of 0.9 and a minor allele frequency of 0.01. Furthermore, we removed SNPs with a heterozygosity above 0.8 from the RAD-seq data set, and for the amplicon sequencing, we only retained SNPs from the genome-wide (neutral) set and removed sites on the Xa/b sex chromosome.

2.7 | Population genetic data analyses

Genetic data analyses were done using R 4.0.2 (R Core Team 2018). The R package {BioCircos} was used to generate the circular plot. Principal component analyses (PCAs) were performed using the *glPCA* function from the R package {adegenet}. The R package {hierfstat} was used for F_{ST} calculations. The SNP intersection matrix was visualized with {UpSetR}. Genotype assignment plots were generated using sparse non-negative matrix factorization algorithms as implemented in the R package {LEA}. For each marker set, we ran 100 repetitions per *K* (*K* = 1–10) with entropy = TRUE to find the most likely number of clusters (i.e., *K* with the lowest entropy). Tajima's *D* estimates in coding sequences (i.e., all immune-related genes targeted by the amplicon) were calculated using the 29 Alpine ibex whole-genome sequencing data sets and the R package {PopGenome}.

3 | RESULTS

3.1 | Assessment of locus quality across the targeted sequencing assay

We analysed whole genome sequencing data sets comprising a total of 138 million SNPs segregating among all samples (65 whole genome sequences, *Capra* and *Ovis*), 5.3 million SNPs segregating among the 29 sequenced Alpine ibex, and 4 million SNPs segregating within the

Alpine ibex founder population Gran Paradiso (Italy). Based on this SNP set, we designed 1,256 amplicons and performed targeted amplicon sequencing covering genome-wide polymorphisms, markers to detect recent hybrids and variants related to immune functions (Figure 1b,c). In contrast to the GT-seq approach, amplifications were performed in 24 pools of ~52 primer pairs using a microfluidics platform (Figure 1a). We analysed a total of 187 samples (representing 172 individuals) in a single run including 158 Alpine ibex, five domestic goats, three Iberian ibex and six suspected hybrids between Alpine ibex and domestic goats (Table S1). Illumina NextSeq 550 sequencing generated a total of 108 million read pairs after removal of PhiX spike-ins (32.4 Gb total sequence length). Of the 1,265 target SNPs, 1,189 SNPs passed hard quality filtering in GATK. To assess genotype quality across loci, we first focused only on samples that satisfied our threshold DNA concentration of ≥50 ng μl⁻¹ (see Methods). Furthermore, we required that each analysed individual produced at least 100,000 mapped reads across all loci (Figure 2a). The 65 retained individuals represented seven different Alpine ibex populations, two domestic goat breeds as well as three Iberian ibex. The number of mapped reads ranged from 230,801 to 1,576,615 reads (Figure 2a; Table S3). Based on these high-quality samples, we found that the median read depth per locus was high with nearly all loci having >20 reads (median across loci =389 reads, Figure 2b). We found seven loci with a median read depth of zero. The highest median read depth was 1,931 for a marker designed in the MHC. A total of 172 loci were invariant among the 65 high-quality samples and were hence removed (retaining 1,017 SNPs). Next, we analysed the genotyping rate across loci and found that 940 loci were genotyped in all high-quality samples and 989 loci were genotyped in more than 75% of samples (Figure 2c). We discarded a further 28 of the 1,017 remaining SNPs due to a genotyping rate (among individuals) below 75% (including the seven loci with median read depth of zero) to prioritize loci providing the highest information content across individuals. We retained a total of 989 high-quality loci for further analysis (Figure 2d).

### 3.2 | Assessment of genotype quality, rate and accuracy

The mean genotype quality (GQ, i.e., the Phred-scaled confidence that the genotype assignment is correct) across the 65 high-quality samples was on average 96 (Figure 3a). Taking advantage of four Alpine ibex individuals, which were both whole genome sequenced and genotyped using our assay, we analysed the overall accuracy of genotypes. Among the four individuals, a total of 62 loci were not assigned a genotype based on whole-genome sequencing. On average, 901 loci (range: 883–910, Figure 3b) showed perfectly matching genotypes and 20 loci (range: 17–27, Figure 3b) showed one mismatching allele (i.e., heterozygote vs. homozygote mismatch). The average genotype quality of the mismatched genotypes was 63 while it was 96 for matching genotypes. We found no complete allelic mismatch in any of the four individuals (i.e., homozygous calls

(a) GT-seq

Multiple rounds of primer testing (~4 months)

Microfluidics-based (*e.g.* Juno system)

Primer design outsourced (1-2 weeks)

Up to 96 samples (regular PCR plates)
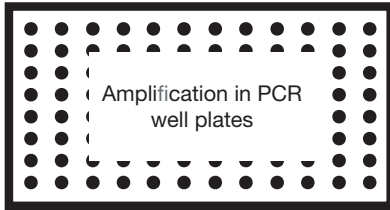
Amplification in PCR well plates

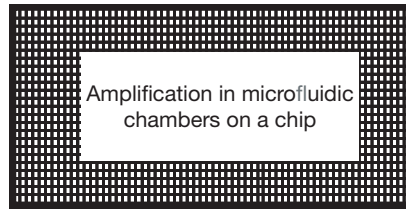PCR 1: up to ~500 pooled primer pairs

PCR 2: TruSeq adapter ligation

Collection, Quality Control

Illumina sequencing

Up to 192 samples

Amplification in microfluidic chambers on a chip

24 pools of up to 100 primer pairs per pool possible

(our study: 24 pools of ~52 primer pairs)

Collected Adapter Ligation & Quality Control

Illumina sequencing

(b)

Marker type
- Genome-wide (Population structure, paternity analyses)
- Immune, outside MHC (immunogenetics)
- MHC region (immunogenetics)
- Diagnostic (hybrid detection)

(c) MHC region

MHC target loci

Genes

Chromosome 23 [position in Mb]

DRB target loci

Genes

DRB

[position in Mb]
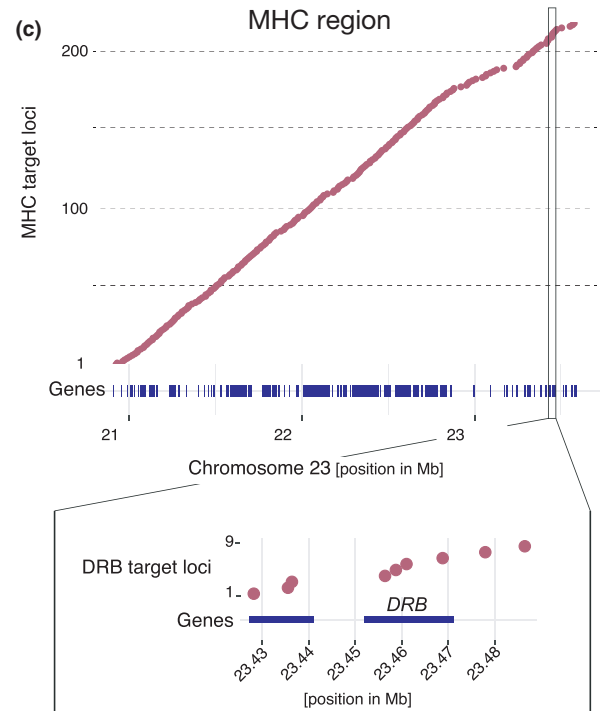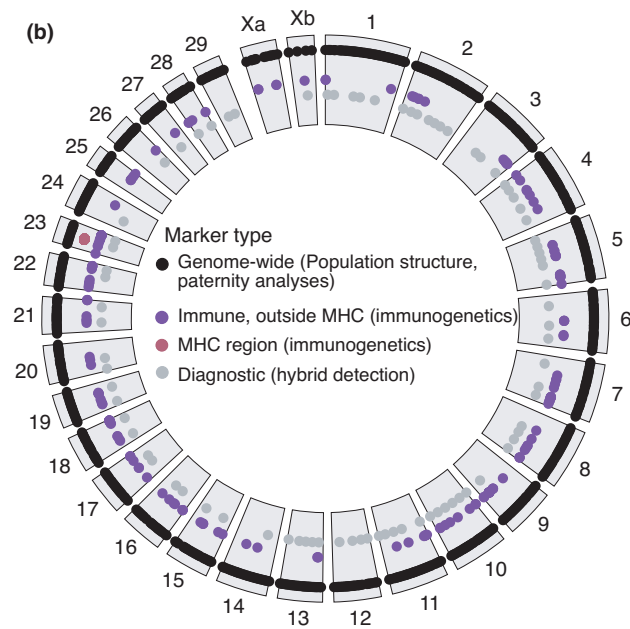
FIGURE 1 Design and distribution of targeted amplicon sequencing loci. (a) Design principles of the genotyping-in-thousands (GT-seq) and the microfluidics-based Juno system approaches. GT-seq relies on the optimization of primer mixes for the simultaneous amplification of hundreds of loci (Campbell et al., 2015; reviewed in Meek & Larson, 2019). By design, microfluidics separates primer pairs into pools to reduce the number of amplifications per microfluidic well (Chiu et al., 2017). (b) Distribution along chromosomes for each marker category (genome-wide, immune loci, MHC region, diagnostic). (c) Regularly spaced loci across the MHC region on chromosome 23 (20.89–23.59 Mb). Blue segments identify coding sequences in the MHC. The *DRB* gene region of the MHC (zoom view) was targeted by a denser array of loci

for distinct alleles). We found that genotyping rates >90% across loci are retained by diluting samples 25-fold from 100 ng μl^{-1} down to 4 ng μl^{-1} (Figure 3c). At 0.8 ng μl^{-1}, the genotyping rate was >75% for two out of three samples. The genotyping rate was >23% for

0.16 ng μl^{-1} of DNA (625-fold dilution). Furthermore, we analysed four faecal samples (without replication). Three of the four faecal samples had a 25% genotyping rate (i.e., 248–270 genotyped loci). The faecal sample with the lowest quality had a 10.3% genotyping
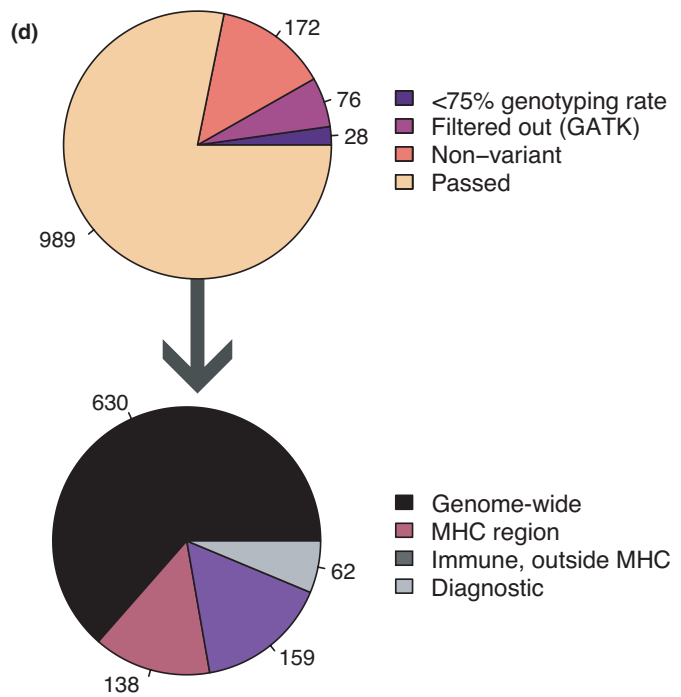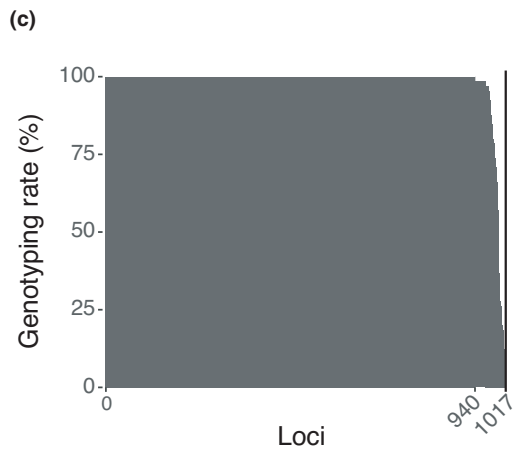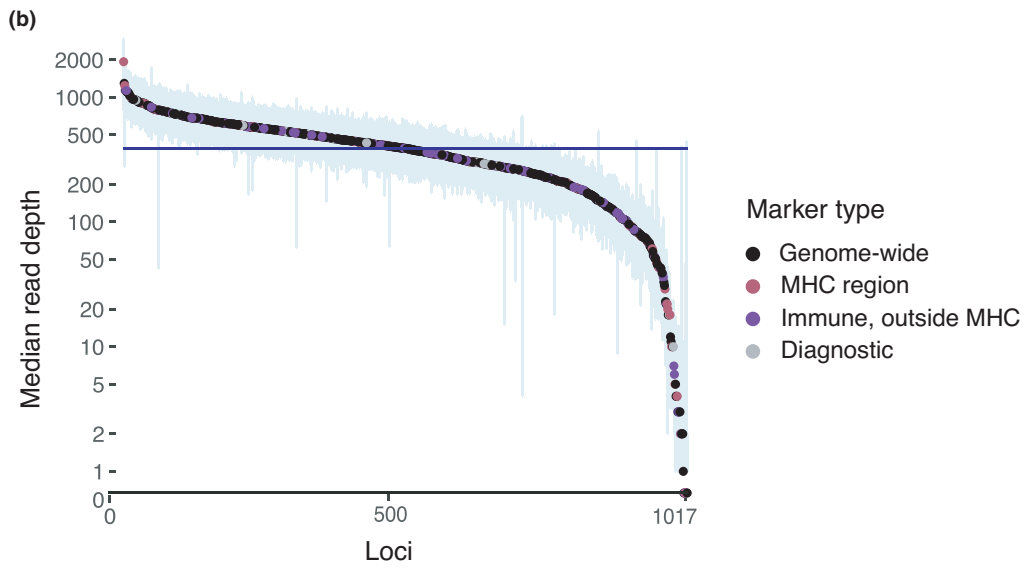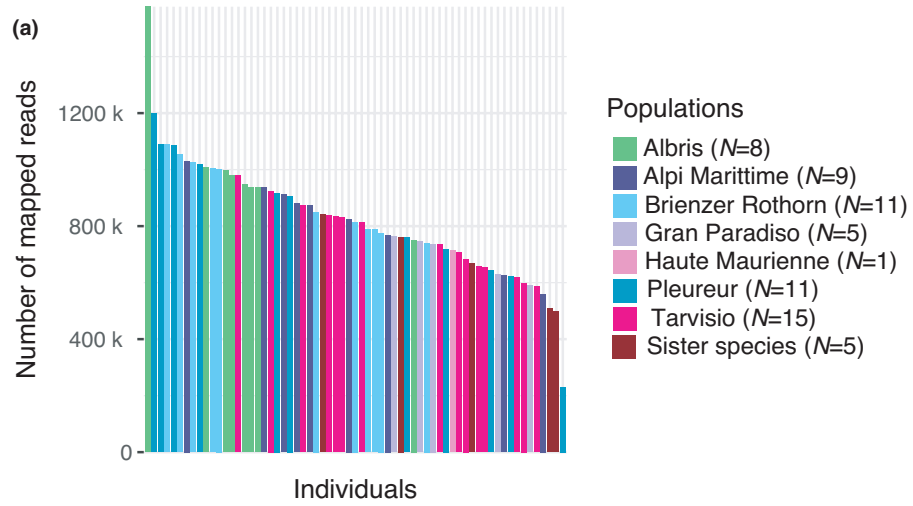
FIGURE 2 Assessment of amplification consistency and genotyping rates. (a) Number of mapped reads per individual sample, colored by population (N = 65). Seven Alpine ibex populations are shown in addition to domestic goats (N = 2) and Iberian ibex (N = 3). Low-quality input DNA samples (faecal samples etc.) were excluded. (b) Median read depth per locus across 1,017 amplicon loci. Markers are coloured according to their category. The blue line represents the median (389) and the blue area represents the 95% confidence interval for each marker. (c) Variation in genotyping rates across the 1,017 loci. (d) Outcomes of the different filtering stages: <75% genotyping rate: loci removed due to low genotyping rate among the 65 high-quality individuals (28 loci); Filtered out (GATK): removed based on SNP hard filtering in GATK (76 loci); Non-variant: monomorphic among the 65 high-quality samples (172 loci); Passed: a total of 989 loci were retained for further analyses. The marker category of the retained markers is shown below. Marker categories are detailed in Figure 1

rate corresponding to 102 genotyped loci. Samples with low genotyping rates typically show inconsistently genotyped loci across samples. We found that this was indeed the case with 41–133 of the loci being genotyped in only one out of four faecal samples. A total of 41 loci were genotyped in at least three fecal samples (Figure 3d).

3.3 | High-resolution population structure

A major reason for genotyping species of conservation concern is the identification of population subdivisions and admixture events. We expanded our genotyping assay to the full set of Alpine ibex samples (n = 158) spanning the extant distribution range across the Alps (Figure 4a). For this, we focused on the specifically designed genome-wide markers excluding immune-related and diagnostic markers as well as sites on the Xa/b sex chromosome (total 617 SNPs). Based on a PCA, we identified three major genotype clusters. The largest cluster was composed of the Gran Paradiso source population and populations reintroduced directly from Gran Paradiso to Italy or Switzerland (Figure 4b). A second cluster grouped the two French populations Haute Maurienne and Champsaur (the latter founded with individuals coming from the former). The third cluster was composed of the isolated Alpi Marittime population, which is thought to have only six effective founder individuals (Terrier & Rossi, 1994). The identified population structure was also supported by individual ancestry coefficients using a sparse non-negative matrix factorization algorithm (K = 3, Figure S1; Frichot et al. 2014). At the K with the lowest entropy (K = 5, Figure S2), the analyses revealed a fine-scale population structure: all populations except for the two French populations Haute Maurienne and Champsaur were clearly distinct (Figure 4d). The genotyping assay performed also well for population-level assignments of the low-input/quality samples. All four Gran Paradiso faecal samples had similar principal component values (Figure 4b) and population assignments (Figure 4d) as other, high-quality samples from the same population. We also analysed population differentiation of Alpine ibex using pairwise F_{ST} (Figure 4c). We found a consistent pattern separating Alpi Marittime from all other populations. Furthermore, the two French populations showed relatively low differentiation consistent with their foundation history: the Champsaur population was founded only 25 years ago (fewer than four ibex generations) with 31 individuals coming from Haute Maurienne. The three Swiss populations (Albris, Brienzer Rothorn and Pleureur) were only weakly differentiated from the Gran Paradiso source population as expected from former analysis based on microsatellites (Biebach & Keller, 2009).

3.4 | Performance contrasts among next-generation sequencing methods

To objectively assess the performance of the newly developed assay, we analysed genotyping outcomes of four core Alpine ibex populations (Gran Paradiso, Pleureur, Albris and Brienzer Rothorn; Figure 5; Figure S2). After quality filtering of each data set to ensure objective comparisons (see Methods), we retained 892 loci from targeted amplicon sequencing (n = 75 individuals), 26,547 RAD-seq loci (n = 82 individuals) and 3 million low-coverage whole genome sequencing loci (n = 15 individuals). Overall, 97% of the individuals had a per-individual genotyping rate of ≥90% for targeted amplicon sequencing loci contrasting with 23% of the individuals genotyped at ≥90% for RAD-seq loci (Figure 5a). Enforcing a per-locus genotyping rate of ≥90% over all individuals, 96% of targeted amplicon sequencing loci but only 39% of RAD-seq loci were retained (Figure 5b). Locus and individual-level genotyping rates cannot meaningfully be retrieved from genotype likelihood-based analyses (low-coverage whole genome sequencing data set). We performed comparative population differentiation analyses and found that the global F_{ST} ranged between 0.071 (RAD-seq, 8,316 SNPs) and 0.077 (targeted amplicon sequencing, 588 SNPs). Pairwise F_{ST} estimates were also similar among marker systems (Figure 5c). PCAs constructed from targeted amplicon sequencing and RAD-seq markers clearly resolved the four populations (Figure 5d,e). However, the low-coverage whole genome sequencing did not resolve Albris and Brienzer Rothorn populations (Figure 5f). The first and second principal component axes explained 6.2% and 4.9% for the targeted amplicon sequencing (Figure 5d), 5.2% and 4.2% for the RAD-seq (Figure 5e) and 23.8%–6.1% for the low-coverage whole genome sequencing (Figure 5f), respectively. Genotype assignments to clusters showed clear population differentiation for the targeted amplicon sequencing markers (Figure 5g) and slightly weaker resolution for RAD-seq markers (Figure 5h). Low-coverage whole genome sequencing genotyping clearly separated Gran Paradiso and Pleureur populations but again failed to resolve Albris and Brienzer Rothorn populations (Figure 5i).

3.5 | Detection of recent hybrids and introgression tracts

We analysed the performance of the SNP markers specifically designed to detect introgression from domestic goats into Alpine ibex. To assess the power to discriminate genotypes suspected to be from hybrid individuals, we performed a PCA including all 158

FIGURE 3 Genotyping accuracy and performance. (a) Mean genotype quality (GQ, i.e., the Phred-scaled confidence that the genotype assignment is correct) for each high-quality locus across the 65 high-quality DNA samples. (b) Genotype accuracy assessed by matching recovered alleles from the targeted amplicon sequencing and whole-genome sequencing of the same four individuals. (c) Assessment of genotyping rates for serial dilutions of three individuals and four faecal samples. (d) Analysis of genotyped loci shared among the four faecal samples. The overlaps show recovered genotypes out of a total of 989 loci
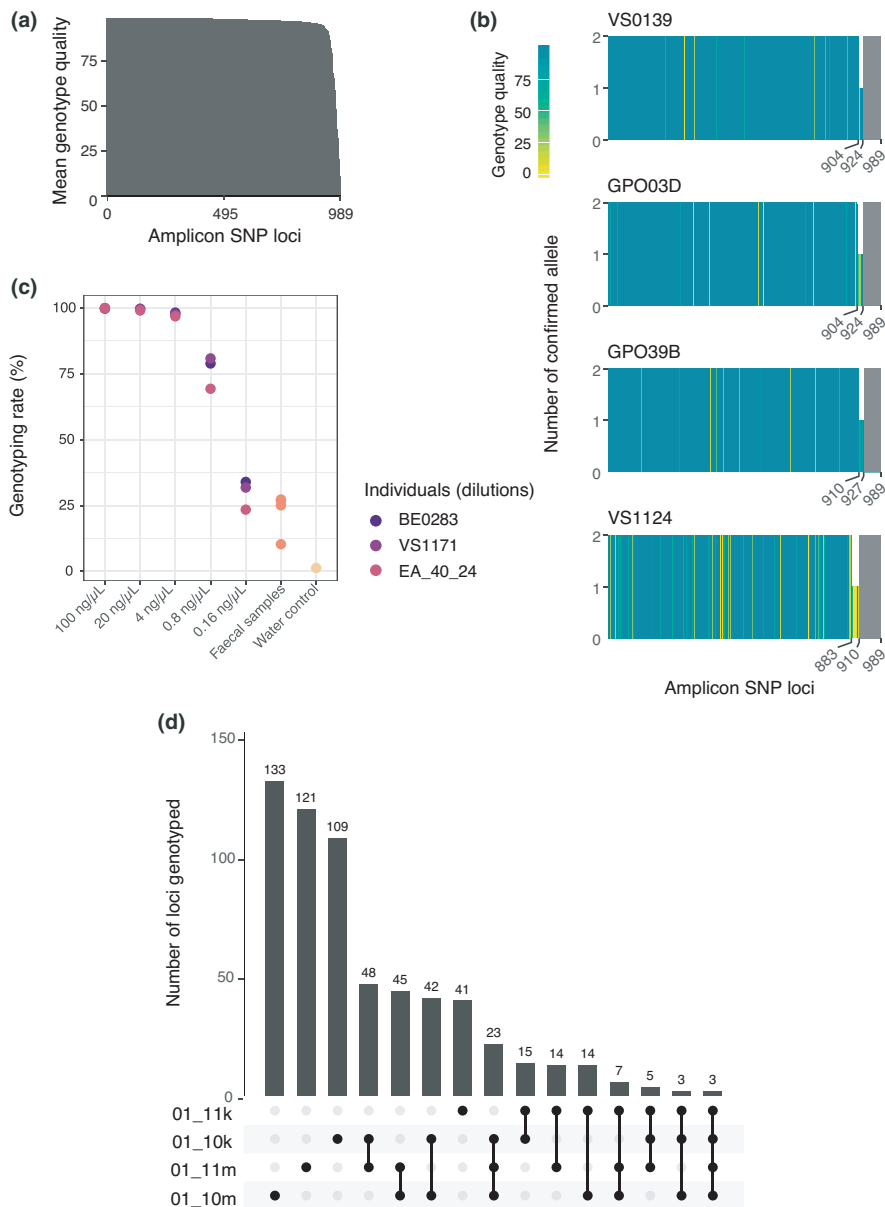

Alpine ibex individuals, suspected hybrids based on field reports of unusual phenotypes ($n = 6$), domestic goats ($n = 5$) and Iberian ibex ($n = 3$; Figure 6a). As expected, the first PC clearly separated domestic goats, Iberian ibex as well as Alpine ibex. The second PC differentiated the Alpi Marittime population from all other Alpine ibex. The suspected hybrid individual from Alpi Marittime (AM-H) clearly clustered with domestic goats. Two additional suspected hybrids (GR-ib1, GR-ib2), as well as a potential albino individual (FR-blanc has largely whitish fur, but no red eyes) clustered with Alpine ibex. Three

suspected hybrids (TI-ib, GPHB1 and GP-ib-V02-17) were located near the midpoint between domestic goats and Alpine ibex matching expectations for recent (F_1 or backcross) hybrid genotypes.

Using the goat–Alpine ibex diagnostic marker set, we analysed the recency of the hybridization event by identifying contributions from each parental species (Figure 6b). Of the six suspected hybrids and one potential albino, two individuals were confirmed to be Alpine ibex (FR-blanc, GR-ib1) and one a domestic goat (AM-H). Our results confirm that the unusual phenotype of FR-blanc reported

FIGURE 4 Population genetic analyses of Alpine ibex. (a) Map showing the extant Alpine ibex distribution (grey) and sampled populations (red). (b) Principal component analysis of 158 Alpine ibex individuals based on 617 genome-wide SNPs designed for capturing genome-wide population differentiation. (c) Pairwise F_{ST} matrix of all Alpine ibex populations based on genome-wide SNPs. Gran Paradiso was the only population surviving the near extinction and hence is the source population of all existing Alpine ibex populations. The Swiss populations Albris, Pleureur and Brienzer Rothorn and the Italian population Alpi Marittime were founded in early 1900 from Gran Paradiso individuals. The populations Tarvisio, Champsaur and Haute Maurienne were later founded independently from Gran Paradiso individuals. (c) Structure-like analysis (based on sparse non-negative matrix factorization algorithms) of all Alpine ibex with $K = 5$. *Faecal samples
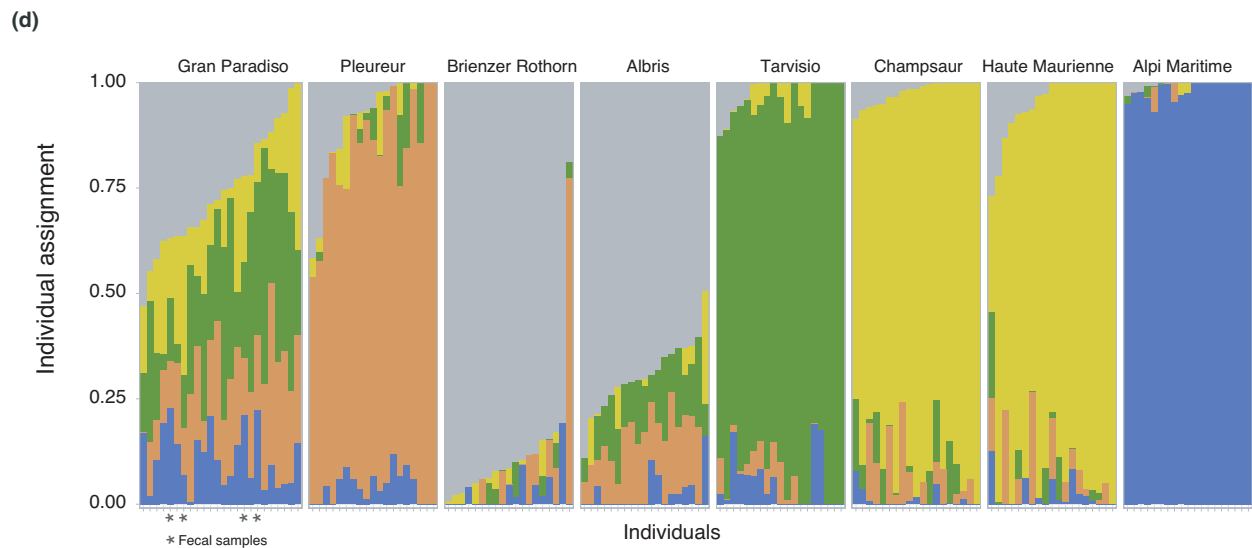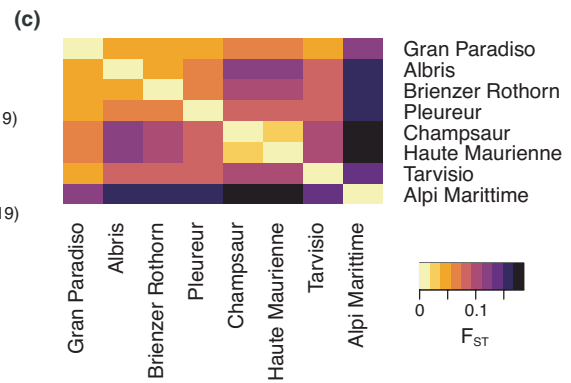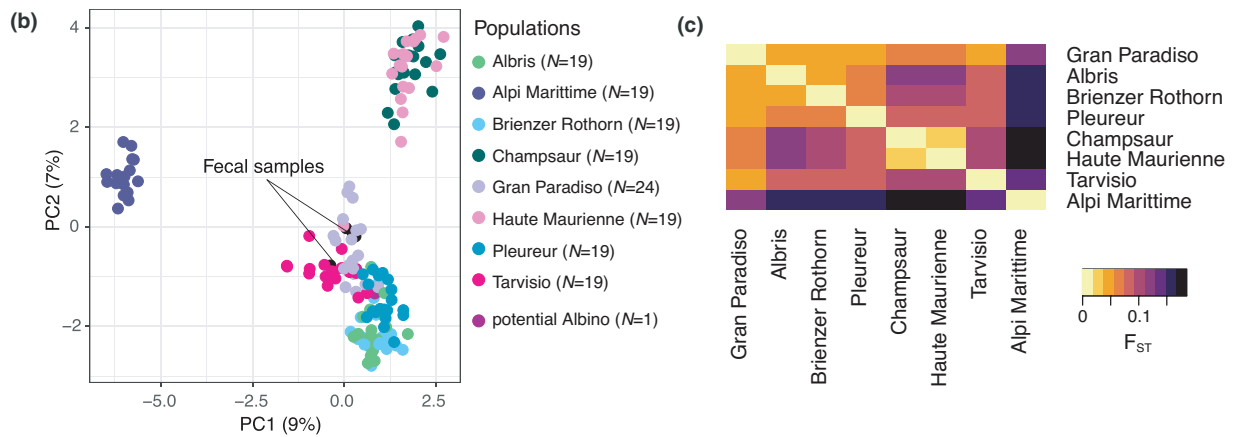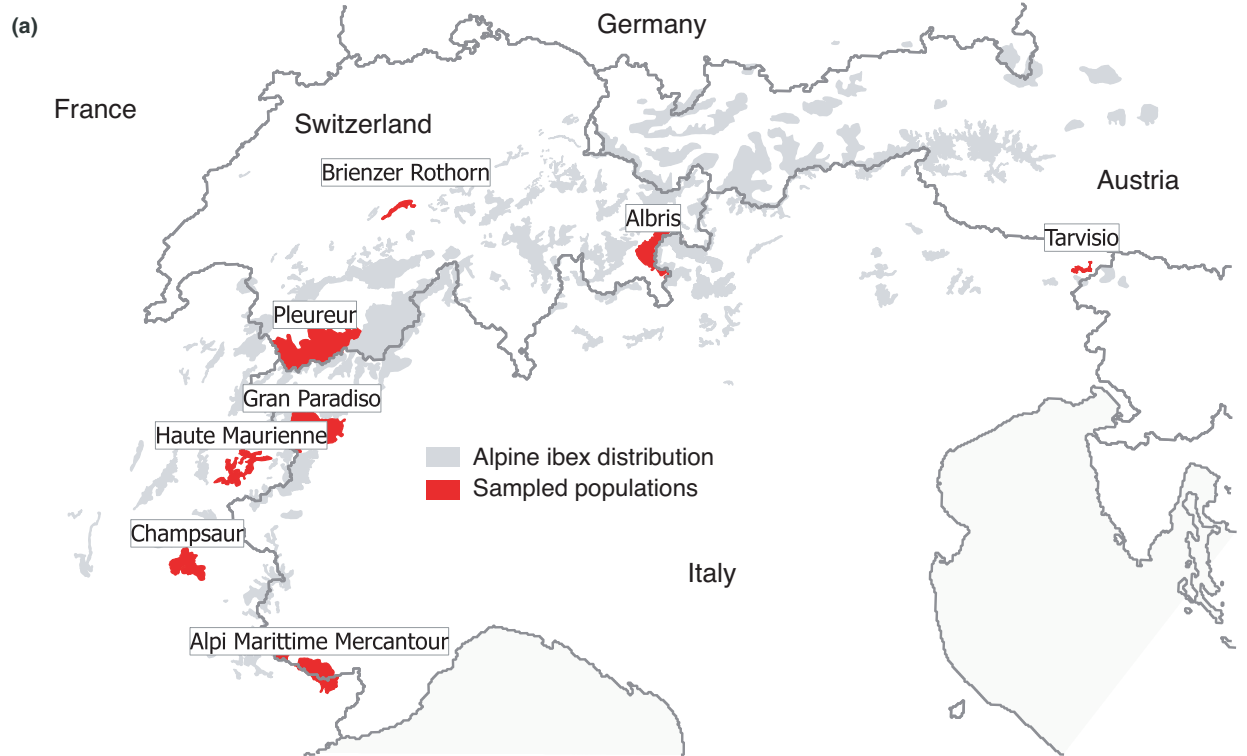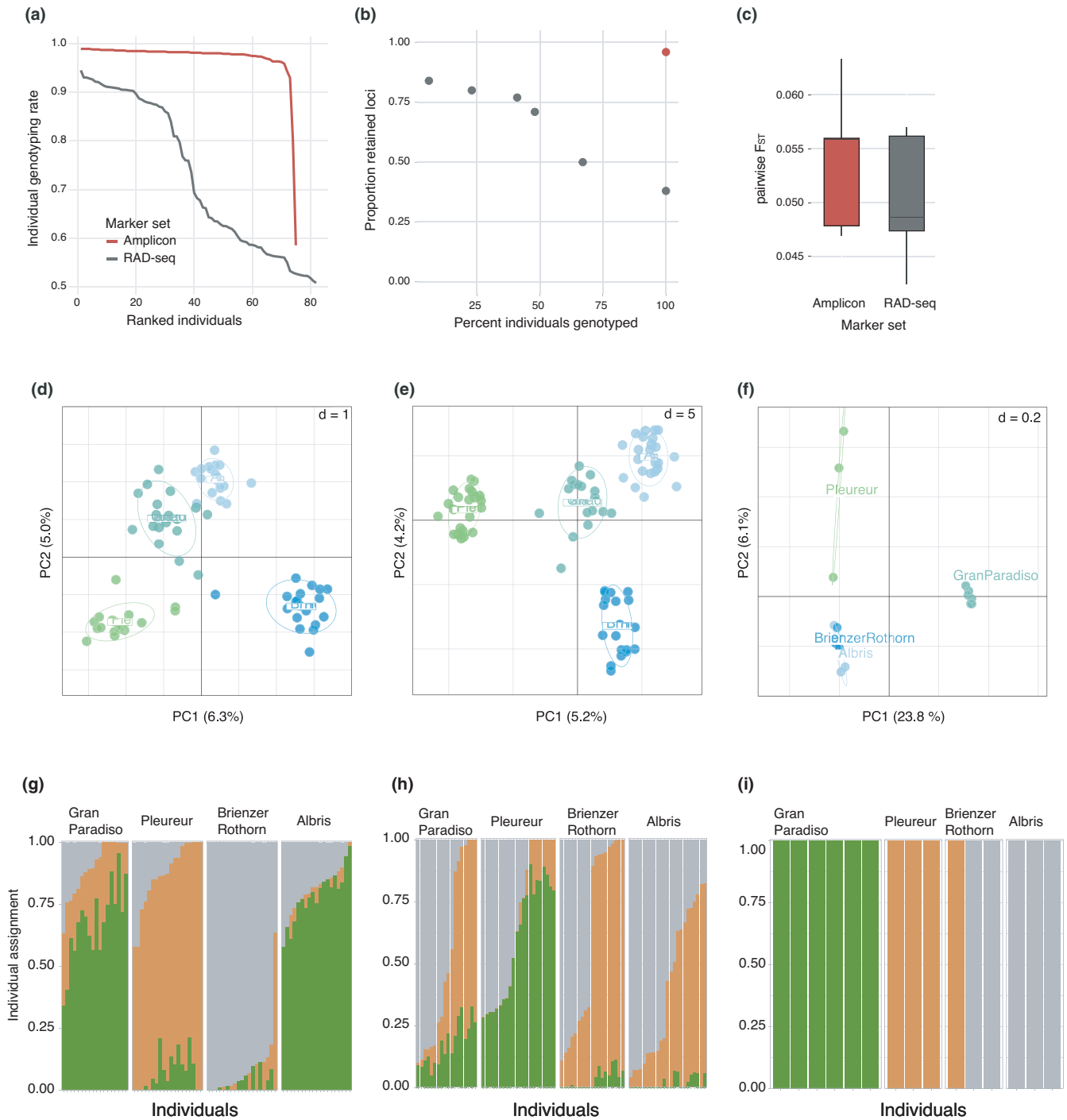
FIGURE 5 Performance comparisons between targeted amplicon sequencing, RAD-seq and low-coverage whole genome sequencing on four Alpine ibex populations. (a) Individual genotyping rate ranked by individual for amplicon sequencing and RAD-seq. (b) Proportion of SNPs retained at a genotyping rate of 90% as a function of the percentage of individuals included. The comparison is only meaningful for amplicon sequencing and RAD-seq. (c) Boxplot of all pairwise F_{ST} estimates among populations for the amplicon sequencing and RAD-seq. (d–f) Principal component analysis of population differentiation for (d) the amplicon sequencing (plotted are −PC1 against −PC2), (e) RAD-seq and (f) low-coverage whole-genome sequencing. Structure-like analysis (based on sparse non-negative matrix factorization algorithms) of (g) the amplicon sequencing, (h) RAD-seq and (i) low-coverage whole-genome sequencing data sets; (i) is based on an NGSADMIX analysis (Skotte et al., 2013)

from the field was not caused by recent domestic goat introgression. Individual GR-ib1 was suspected to be a hybrid because it was behaving in very unusual ways, seeking proximity to buildings.

Individual AM-H was reported to resemble domestic goat but was living among Alpine ibex. Another suspected hybrid (GR-ib2) with white hoofs showed a weak sign of domestic goat introgression (only
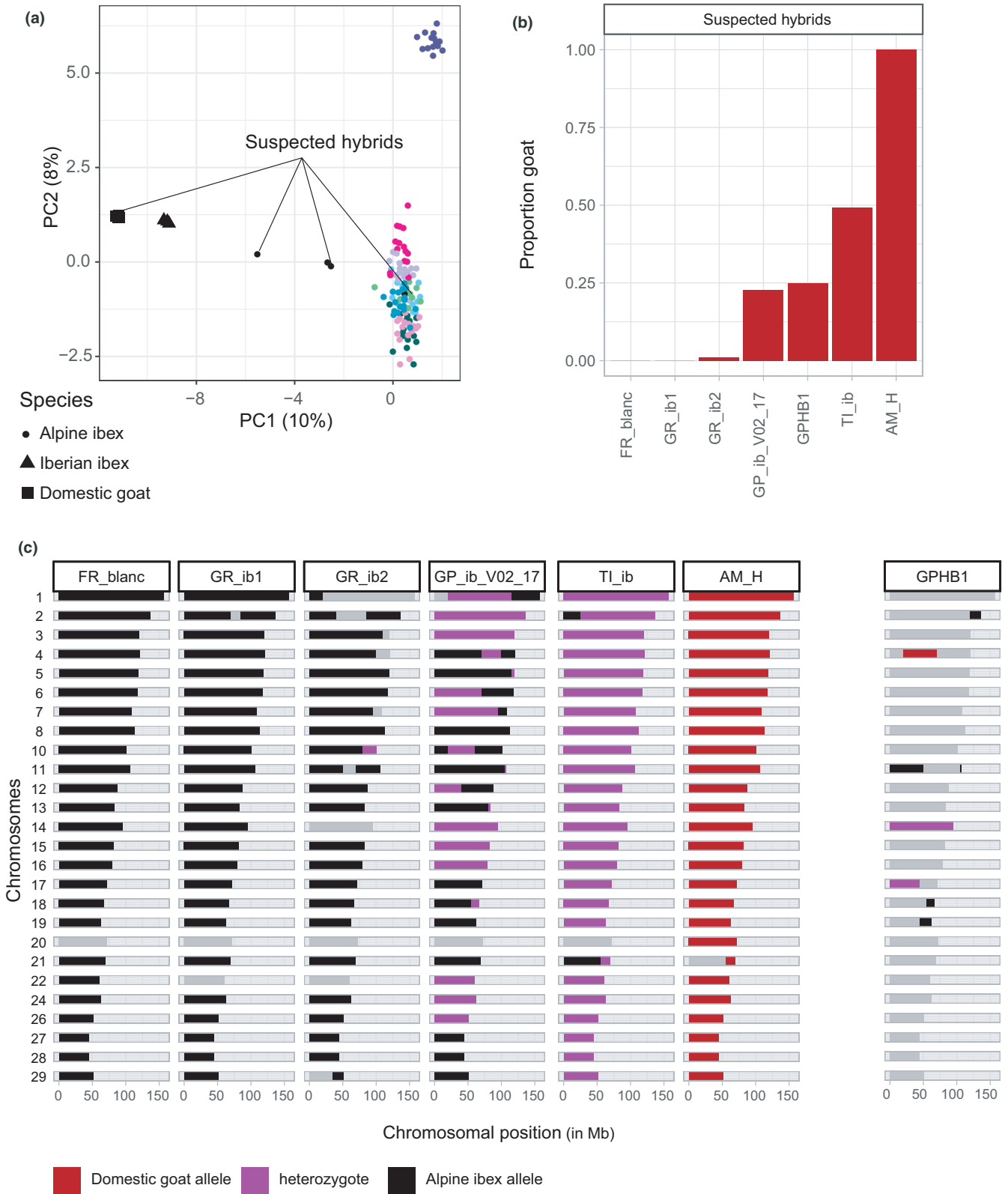
FIGURE 6 Performance of targeted amplicon sequencing to detect recent hybridization events. (a) Principal component analysis of 168 individuals (excluding faecal samples) including domestic goat and Iberian ibex ($N = 5$ and 3) and suspected hybrids based on phenotypic observations ($N = 6$). The PCA was performed based on 617 genome-wide SNPs designed for capturing putatively neutral population differentiation. (b) Analysis of diagnostic markers for the detection of hybrids for six suspected hybrids and one potential albino. The proportion of diagnostic markers assigned to domestic goat are shown, where the proportion is calculated as (1 × homozygous goat genotype + ½ × heterozygous genotype)/(total genotypes at diagnostic markers). (c) Chromosome painting using diagnostic markers along the 29 autosomes. Colours identify Alpine ibex (black), heterozygous (purple) and domestic goat (red) genotypes. Sample GPHB1 was of low DNA quality and quantity (drop of blood conserved on FTA filter paper; see Methods)

one marker on chromosome 10 was heterozygous) suggesting a potential backcross. Note that GR_ib2 showed no clear differentiation from Alpine ibex based on the genome-wide marker PCA underlining the usefulness of specifically designed diagnostic loci. Two individuals (GP_ib_V02_17 and GPHB) showed signs of ~25% domestic goat introgression (i.e., probable F_2 backcrosses). GP_ib_V02_17 had darker fur than commonly seen in Alpine ibex and an unusual horn shape (no nodes and a triangular transverse section). Chromosome painting showed one individual was being heterozygous for all diagnostic markers (TI_ib, Figure 6c), hence representing very probably an F_1 hybrid. This individual was observed going into a stable following goats and had an unusual horn shape. The DNA from GPHB was from a small amount of blood stored on FTA paper explaining the poor genotype quality. Field reports suspected an F_1 hybrid but both the contribution plot (Figure 6b) and the chromosomal painting (Figure 6c) suggest an F_2 hybrid. Our analyses show the power of highly discriminatory markers to detect recent hybridization and introgression events in a pool of individuals with field-reported, suspected admixture.

3.6 | Immunogenetics of Alpine ibex populations

The targeted amplicon sequencing specifically focused on 297 polymorphisms in immune-related genes within and outside the MHC. Using whole-genome sequencing data for Alpine ibex (n = 29), we found that Tajima's D in Alpine ibex ranged from −2.4 to 4.5 across all genes encoded in the MHC region and from −1.4 to 2.5 for all immune-related genes outside of the MHC region targeted by the amplicon sequencing. The median Tajima's D was lower in the MHC (0.45) compared to immune loci outside of the MHC (1.3, Figure 7a,b). The high Tajima's D values in some immune-related loci suggest long-term maintenance of alleles through balancing selection (Figure 7a). Comparing different sets of amplicon targets in our assay, we found that the genome-wide markers aimed at resolving population structure showed overall the highest average heterozygosity (Figure 7c). The lowest genome-wide heterozygosity was found in Alpi Marittime consistent with the severe founding bottleneck. Immune-related and MHC loci showed consistently lower levels of heterozygosity compared to genome-wide loci. A notable exception was the Alpi Marittime population where the MHC showed higher levels of heterozygosity (0.25) than other immune-related and genome-wide markers (Figure 7c). The MHC of the Albris population showed a surprisingly low average heterozygosity (0.11) compared to genome-wide markers. On a PCA, the MHC in domestic goat and Iberian ibex showed a low degree of differentiation (Figure 7d). This is probably explained by the focus on segregating polymorphism in Alpine ibex only. MHC genotypes showed tight clusters among Alpine ibex individuals but only weak population signatures, which is in marked contrast to genome-wide markers (Figures 4b and 7d). Clusters of nearly identical Alpine ibex genotypes were generally composed of genotypes from multiple populations. The two French populations Haute Maurienne and Champsaur shared most MHC genotypes. Individuals from the Alpi Marittime were largely distinct from all other populations with the exception of a shared genotype with the Italian population Tarvisio. Some individuals from Alpi Marittime were translocated to Tarvisio in 1993, which may explain our finding.

4 | DISCUSSION

Genetic monitoring is central to many population surveys and conservation efforts. Here, we developed an accurate and versatile microfluidics-based tool for the monitoring of Alpine ibex population health. The set of nearly 1,000 loci enables the concurrent assessment of population structure, the detection of recent hybridization events and immune function-related genotypes. We show that the assay performs well in comparison to restriction-site based DNA sequencing and low-coverage whole genome sequencing with a probable wider range of acceptable input material. Assessments of immunity-related genotypes can be integrated in individual translocation recommendations for conservation management of the species.

Effective high-throughput genotyping relies on the consistent amplification of a large number of loci, robustness to variation in input DNA quality and affordable costs. We show that the microfluidics-based targeted sequencing approach produces largely uniform coverage both across 989 loci and individuals. Integrating whole-genome sequencing derived SNPs from different species into the design of amplicons allowed us to amplify across Alpine ibex, Iberian ibex and the domestic goat. This extends the usability of the assay with the caveat that the potential ascertainment bias in other species should be considered because SNPs were selected based on polymorphism in Alpine ibex only. Furthermore, the identified genotypes were well validated by cross-referencing with the whole-genome sequencing data sets. For Alpine ibex, uniform, high-accuracy genotyping with low input DNA has previously only been achieved through microsatellite marker analyses (Biebach & Keller, 2009) or SNP chips designed on related model species (Grossen et al., 2014). An SNP chip genotyping study based on a 52 K Illumina Goat SNP Chip recovered 677 polymorphic markers out of a total of ~52,000 markers known to be polymorphic among domestic goats (Grossen et al., 2014; Tosser-Klopp et al., 2015). This shows how even SNP chips designed for closely related species can be largely unsuitable and may suffer from substantial ascertainment bias. High-throughput methods such as RAD-seq and GBS sequencing are widely used next-generation sequencing approaches for nonmodel species often producing at least tens of thousands of SNPs. However, such restriction-based reduced representation approaches do not allow the targeting of specific loci (e.g., immune-relevant loci) and often produce highly uneven read depth across loci (Andrews et al., 2016; Jiang et al., 2016). Furthermore, we show that replicating RAD-seq and low-coverage whole genome sequencing on the same set of populations analysed in our targeted amplicon sequencing approach produces a highly similar resolution of the genetic structure. The comparatively low number of SNPs assayed in the targeted amplicon sequencing is probably compensated for by the highly consistent genotyping rates across loci and individuals.
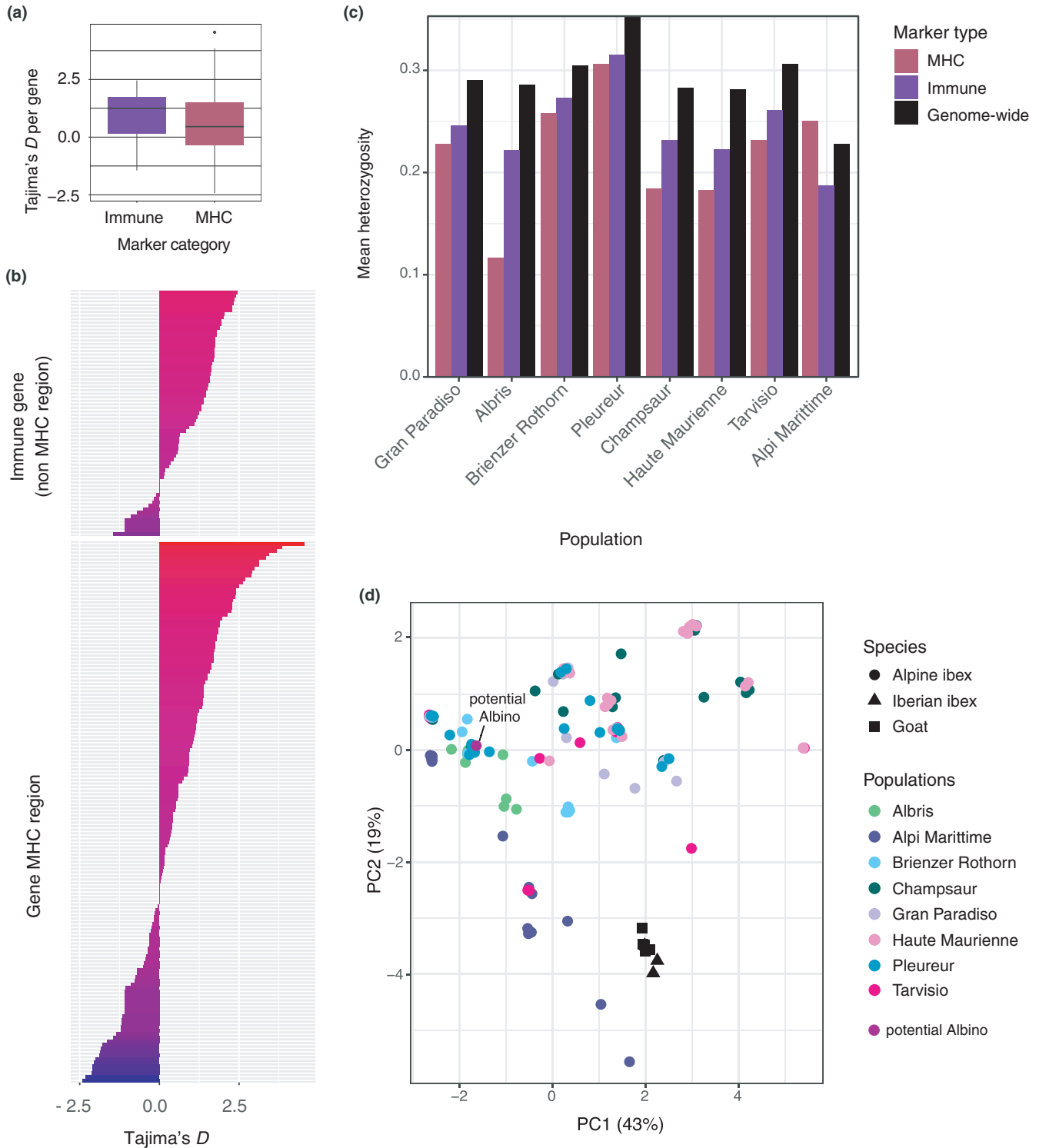
FIGURE 7 Genome-wide immunogenetics survey of Alpine ibex populations. (a) Boxplot of Tajima's D estimates based on whole-genome sequencing of 29 Alpine ibex of immune-related loci covered by the targeted amplicon sequencing assay. The MHC region is shown separately from other immune-related loci. (b) Tajima's D estimates shown per each gene represented by immune-related loci covered by the targeted amplicon sequencing assay. (c) Targeted amplicon sequencing analyses of heterozygosity across Alpine ibex populations ($N = 154$, excluding faecal samples). Heterozygosity is shown separately for genome-wide SNPs, the MHC region and other immune-related SNP loci. (d) Principal component analysis of Alpine ibex and sister species based on 138 SNPs genotyped in the MHC region using targeted amplicon sequencing

We show that the microfluidics approach for amplicon sequencing produces consistent amplification across nearly all loci if minimum DNA requirements are fulfilled. We have set a minimum threshold of 50 ng μl^{-1} (or a total of 100 ng), which is usually achievable for well-preserved tissue samples. We have also investigated the potential to recover genotypes from degraded and low-concentration

samples such as faecal material. Faecal sampling is a widely used noninvasive sampling method and sometimes the only option for elusive species (Beja-Pereira et al., 2009). Faecal samples produce low-quality (i.e., degraded) and low-quantity DNA. In addition, faecal DNA samples can be heavily contaminated with bacteria, plant or prey DNA. Hence, the actual endogenous DNA is often much below the measured total DNA concentration. We analysed faecal DNA quantities from 4–23 ng μl^{-1} but found that the proportion of successfully amplified loci was comparable to diluted nonfaecal DNA at around 0.1 ng μl^{-1}. Our preliminary analyses using four faecal samples suggest that the new amplicon assay produces sufficient genotyping data to assign the origin of faecal samples to individual populations. To assess the resolution of the genome-wide set of amplicon markers, we have genotyped a collection of Alpine ibex samples representing all major reintroduction events. Alpine ibex were limited to the Gran Paradiso National Park in Italy in the 19th century. Starting in the early 20th century, populations were introduced independently from Gran Paradiso to Switzerland, France and Italy and from there further populations were founded. The population structure of extant Alpine ibex populations is dominated by signals of population reintroductions and translocations (Biebach & Keller, 2009; Grossen et al., 2018). The Alpine ibex genotypes assessed by our targeted sequencing approach confirmed all major aspects of the reintroduction history including placing Gran Paradiso at the centre of the extant genetic diversity. Populations reintroduced directly from Gran Paradiso were at the closest genetic distances to the source population yet showed distinct clustering as expected from the strong bottlenecks imposed by the translocation of few individuals. A major concern for wild species co-occurring with closely related domestic animals is the potential for hybridization and introgression. Alpine ibex populations across Europe are monitored for the presence of individuals with atypical phenotypes (e.g., Giacometti et al., 2004; Steyer et al., 2016; Todesco et al., 2016). We assessed genotypes of seven suspected hybrids (including one potential albino) and were able to show that only three individuals are clearly identifiable as recent hybrids. One individual showed possible signs of past introgression and three suspected hybrids clustered either with domestic goat or Alpine ibex genotypes, highlighting the importance of genetic monitoring of suspected hybridization events.

A major genetic factor for the long-term survival of endangered species is genetic diversity at immune loci, in particular the MHC. We successfully amplified hundreds of loci involved in key immune functions focusing on nonsynonymous polymorphisms. In parallel, we amplified a dense array of loci spanning the MHC. This 2-Mb locus on chromosome 23 recently received genetic material from domestic goats, regenerating heterozygosity at the DRB locus (Grossen et al., 2014). Genotypes across the MHC clustered tightly among populations, revealing that Alpine ibex were genetically impoverished across the entire MHC. Comparative analyses with domestic goat breeds will enable a refinement of our understanding of historical and potentially ongoing introgression into the Alpine ibex gene pool. Our findings of low diversity at the MHC are also in accordance with previous studies based on microsatellite and RAD

genotyping (e.g., Alasaad et al., 2012; Brambilla et al., 2018; Grossen et al., 2014). Underlining the relevance of performing large-scale targeted amplicon sequencing of immune loci are recent findings that high heterozygosity at the MHC was correlated with higher resistance to infectious keratoconjunctivitis, a major disease factor in some Alpine ibex populations (Brambilla et al., 2018). Surveillance of disease susceptibility through genetic analyses of immune loci will also significantly improve conservation strategies by informing the choice of founder individuals prior to reintroductions.

Our microfluidics-based amplicon sequencing approach joins GT-seq, the currently most widely used method for amplicon sequencing (Campbell et al., 2015; Meek & Larson, 2019), for the survey of wild populations. The major advantage of GT-seq relative to our approach is the possibility to multiplex thousands of samples (relative to the hundreds currently possible with microfluidics). In particular for conservation genetics applications, the cost per sample is an important factor to consider. We refer here to Campbell et al. (2015) to compare costs with the GT-seq approach. To provide useful comparisons, we assume identical costs for DNA extraction, library purification, quality checks and sequencing, because we expect these steps to be identical between the procedures. For 2,000 samples, GT-seq costs were estimated to be 1.23 USD per sample for points 1 and 6–8 (Appendix S1; Campbell et al., 2015). Our protocol was based on 384 single-index sequencing barcodes, but this could easily be modified to reach higher plexity. The remaining costs for oligo synthesis and amplification were estimated to be 2.75 USD per sample (Campbell et al., 2015; Table S4). In direct comparison, our microfluidics approach costs 14.40 USD per sample, including the cost of outsourced primer design, Juno IFC chips and kits (Table S4). Note that our cost estimates are for a Swiss research facility in 2020 compared to Campbell et al. (2015). The Juno microfluidic system from Fluidigm used in this study is also not standard molecular laboratory equipment. In summary, GT-seq is more cost-effective per sample and is independent from commercial instruments. Note though that GT-seq may imply more development time (~4 months according to Meek & Larson, 2019) to ensure sufficiently balanced amplicons in a single pool (or multiple pools). To our knowledge, the sensitivity to input DNA quantity and quality of GT-seq and the microfluidics approach has not been systematically compared. We expect both methods to perform similarly well, however. Recent targeted sequencing applications focused on amplifying fragments of the MHC, sex-specific markers and marker sets to resolve parentage in large sample pools (Bootsma et al., 2020, Fuselli et al., 2018, Razali et al., 2017, Yu et al., 2020, reviewed in Meek & Larson, 2019). Our approach integrating whole-genome resequencing data sets both to identify the most valuable polymorphisms to target and to mask polymorphisms interfering with amplification should be universally beneficial. Provided that at least a draft genome is available from a closely related species, even low-coverage data sets will provide important benefits to the assay development and validation.

In conclusion, our microfluidics-based targeted amplicon assay allows the simultaneous monitoring of multiple genetic risk factors and demonstrates how next-generation sequencing techniques can

be adapted for the needs of population genetic surveys and conservation management. The whole genome sequencing-informed approach enabled us to select a highly specific set of loci to simultaneously address questions of population structure, recent hybridization events and how polymorphism is shaped across major components of the immune system. Efficient and precise characterization of individual genotypes can be translated into recommendations on how to prioritize translocation events and replenish genetic diversity at immune loci. The versatility to amplify the same loci across related species also enables powerful screens for recent introgression events. Our study shows the relevance of bridging population genomic investigations with assays that can be realistically implemented into population genetic surveys and decision-making for conservation management.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

AUTHOR CONTRIBUTIONS

C.K., C.G. and D.C. conceived the study, C.K., D.W., G.C., I.B. and C.G. performed analyses, A.B. contributed samples, D.L. contributed data sets, C.K., C.G. and D.C. wrote the manuscript. All authors commented on the text.

DATA AVAILABILITY STATEMENT

The raw amplicon sequencing data produced for this project have been deposited at the NSBI Sequence Read Archive under BioProject Accession no. PRJNA669599. The raw whole-genome sequencing data produced for this project have been deposited at the NCBI Sequence Read Archive under Accession nos. SAMN10736122–SAMN10736160 (BioProject PRJNA514886 [https://www.ncbi.nlm.nih.gov/sra/PRJNA514886]). The whole-genome data produced by the NexGen Consortium (*Capra hircus* accessions: ERR470105, ERR470101, ERR313212, ERR313211, ERR313204, ERR297229, ERR313206, ERR405774, ERR405778, ERR315778, ERR318768, ERR246140, ERR340429, ERR246152, ERR345976, *Capra aegagrus* accessions: ERR340334, ERR340340, ERR340333, ERR340331,

ERR340335, ERR340348; *Ovis aries* accessions: ERR157945, ERR299288; *Ovis orientalis*: ERR157938; *Ovis vignei*: ERR454948; *Ovis canadensis*: SRR501898) were downloaded from [ftp://ftp.sra.ebi.ac.uk/vol1/fastq].

ORCID

Deborah M. Leigh https://orcid.org/0000-0003-3902-2568
Daniel Croll https://orcid.org/0000-0002-2072-380X

REFERENCES

Acevedo-Whitehouse, K., & Cunningham, A. A. (2006). Is MHC enough for understanding wildlife immunogenetics? *Trends in Ecology & Evolution*, 21(8), 433–438. https://doi.org/10.1016/j.tree.2006.05.010

Alasaad, S., Biebach, I., Grossen, C., Soriguer, R. C., Pérez, J. M., & Keller, L. F. (2012). Microsatellite-based genotyping of MHC class II DRB1 gene in Iberian and Alpine ibex. *European Journal of Wildlife Research*, 58(4), 743–748. https://doi.org/10.1007/s10344-011-0592-0

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(2), 389–400. https://doi.org/10.1534/genetics.115.183665

Allendorf, F. W., Hohenlohe, P. A., & Luikart, G. (2010). Genomics and the future of conservation genetics. *Nature Reviews. Genetics*, 11(10), 697–709. https://doi.org/10.1038/nrg2844

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Ammerdorffer, A., Roest, H.-I.-J., Dinkla, A., Post, J., Schoffelen, T., van Deuren, M., Sprong, T., & Rebel, J. M. (2014). The effect of *C. burnetii* infection on the cytokine response of PBMCs from pregnant goats. *PLoS One*, 9(10), e109283.

Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, 17, 81–92. https://doi.org/10.1038/nrg.2015.28

Angelone, S., Jowers, M. J., Molinar Min, A. R., Fandos, P., Prieto, P., Pasquetti, M., Cano-Manuel, F. J., Mentaberre, G., Olvera, J. R. L., Ráez-Bravo, A., Espinosa, J., Pérez, J. M., Soriguer, R. C., Rossi, L., & Granados, J. E. (2018). Hidden MHC genetic diversity in the Iberian ibex (*Capra pyrenaica*). *BMC Genetics*, 19(1), 28. https://doi.org/10.1186/s12863-018-0616-9

Barnosky, A. D., Matzke, N., Tomiya, S., Wogan, G. O. U., Swartz, B., Quental, T. B., Marshall, C., McGuire, J. L., Lindsey, E. L., Maguire, K. C., Mersey, B., & Ferrer, E. A. (2011). Has the Earth's sixth mass extinction already arrived? *Nature*, 471(7336), 51–57.

Beja-Pereira, A., Oliveira, R., Alves, P. C., Schwartz, M. K., & Luikart, G. (2009). Advancing ecological understandings through technological transformations in noninvasive genetics. *Molecular Ecology Resources*, 9(5), 1279–1301. https://doi.org/10.1111/j.1755-0998.2009.02699.x

Bernatchez, L., & Landry, C. (2003). MHC studies in nonmodel vertebrates: What have we learned about natural selection in 15 years? *Journal of Evolutionary Biology*, 16(3), 363–377.

Biebach, I., Leigh, D. M., Sluzek, K., & Keller, L. F.(2016). 8. Genetic Issues in Reintroduction. In S. J. David, J. M. Joshua, L. A. Paul, & S. Rob (Eds.), *Reintroduction of Fish and Wildlife Populations*. (149–184). Berkeley: University of California Press. https://doi.org/10.1525/9780520960381-010

Bickhart, D. M., Rosen, B. D., Koren, S., Sayre, B. L., Hastie, A. R., Chan, S., Lee, J., Lam, E. T., Liachko, I., Sullivan, S. T., Burton, J. N., Huson, H. J., Nystrom, J. C., Kelley, C. M., Hutchison, J. L., Zhou, Y., Sun,

J., Crisà, A., Ponce de León, F. A., ... Smith, T. P. L. (2017). Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. *Nature Genetics*, 49(4), 643–650. https://doi.org/10.1038/ng.3802

Biebach, I., & Keller, L. F. (2009). A strong genetic footprint of the re-introduction history of Alpine ibex (*Capra ibex ibex*). *Molecular Ecology*, 18(24), 5046–5058. https://doi.org/10.1111/j.1365-294X.2009.04420.x

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bollmer, J. L., Vargas, F. H., & Parker, P. G. (2007). Low MHC variation in the endangered Galápagos penguin (*Spheniscus mendiculus*). *Immunogenetics*, 59(7), 593–602. https://doi.org/10.1007/s00251-007-0221-y

Bootsma, M. L., Gruenthal, K. M., McKinney, G. J., Simmons, L., Miller, L., Sass, G. G., & Larson, W. A. (2020). A GT-seq panel for walleye (*Sander vitreus*) provides important insights for efficient development and implementation of amplicon panels in non-model organisms. *Molecular Ecology Resources*, 20(6), 1706–1722.

Brambilla, A., Biebach, I., Bassano, B., Bogliani, G., & von Hardenberg, A. (2015). Direct and indirect causal effects of heterozygosity on fitness-related traits in Alpine ibex. *Proceedings. Biological Sciences / the Royal Society*, 282(1798), 20141873. https://doi.org/10.1098/rspb.2014.1873

Brambilla, A., Keller, L., Bassano, B., & Grossen, C. (2018). Heterozygosity-fitness correlation at the major histocompatibility complex despite low variation in Alpine ibex (*Capra ibex*). *Evolutionary Applications*, 11(5), 631–644.

Brambilla, A., Von Hardenberg, A., Nelli, L., & Bassano, B. (2020). Distribution, status, and recent population dynamics of Alpine ibex *Capra ibex* in Europe. *Mammal Review*, 50(3), 267–277.

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(4), 855–867. https://doi.org/10.1111/1755-0998.12357

Ceballos, G., Ehrlich, P. R., Barnosky, A. D., Garcia, A., Pringle, R. M., & Palmer, T. M. (2015). Accelerated modern human-induced species losses: Entering the sixth mass extinction. *Science Advances*, 1(5), e1400253. https://doi.org/10.1126/sciadv.1400253

Ceballos, G., Ehrlich, P. R., & Dirzo, R. (2017). Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. *Proceedings of the National Academy of Sciences*, 114(30), E6089–E6096.

Chen, K., Zhou, Y. X., Li, K., Qi, L. X., Zhang, Q. F., Wang, M. C., & Xiao, J. H. (2016). A novel three-round multiplex PCR for SNP genotyping with next generation sequencing. *Analytical and bioanalytical chemistry*, 408(16), 4371–4377.

Chiu, D. T., deMello, A. J., Di Carlo, D., Doyle, P. S., Hansen, C., Maceiczyk, R. M., & Wootton, R. C. R. (2017). Small but perfectly formed? Successes, challenges, and opportunities for microfluidics in the chemical and biological sciences. *Chem*, 2(2), 201–223. https://doi.org/10.1016/j.chempr.2017.01.009

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., & Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly*, 6(2), 80–92. https://doi.org/10.4161/fly.19695

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. & 1000 Genomes Project Analysis Group (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., & Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, 12(7), 499–510. https://doi.org/10.1038/nrg3012

Dong, Y., Xie, M., Jiang, Y. U., Xiao, N., Du, X., Zhang, W., Tosser-Klopp, G., Wang, J., Yang, S., Liang, J., Chen, W., Chen, J., Zeng, P., Hou, Y., Bian, C., Pan, S., Li, Y., Liu, X., Wang, W., ... Wang, W. (2013). Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (*Capra hircus*). *Nature Biotechnology*, 31(2), 135–141. https://doi.org/10.1038/nbt.2478

Eriksson, C. E., Ruprecht, J., & Levi, T. (2020). More affordable and effective noninvasive single nucleotide polymorphism genotyping using high-throughput amplicon sequencing. *Molecular Ecology Resources*, 20(6), 1505–1516. https://doi.org/10.1111/1755-0998.13208

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, 32(19), 3047–3048. https://doi.org/10.1093/bioinformatics/btw354

Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., & François, O. (2014). Fast and efficient estimation of individual ancestry coefficients. *Genetics*, 196(4), 973–983. https://doi.org/10.1093/bioinformatics/btw354

Frankham, R. (2005). Genetics and extinction. *Biological Conservation*, 126(2), 131–140. https://doi.org/10.1016/j.biocon.2005.05.002

Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., & François, O. (2014). Fast and efficient estimation of individual ancestry co-efficients. *Genetics*, 196(4), 973–983. https://doi.org/10.1016/j.biocon.2005.05.002

Fuselli, S., Baptista, R. P., Panziera, A., Magi, A., Guglielmi, S., Tonin, R., Benazzo, A., Bauzer, L. G., Mazzoni, C. J., & Bertorelle, G. (2018). A new hybrid approach for MHC genotyping: high-throughput NGS and long read MinION nanopore sequencing, with application to the non-model vertebrate Alpine chamois (*Rupicapra rupicapra*). *Heredity*, 121(4), 293–303. https://doi.org/10.1038/s41437-018-0070-5

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(6), 915–934. https://doi.org/10.1111/j.1755-0998.2010.02891.x

Giacometti, M., Roganti, R., Tann, D. D., Stahlberger-Saitbekova, N., & Obexer-Ruff, G. (2004). Alpine ibex *Capra ibex ibex* x domestic goat *C. aegagrus domestica* hybrids in a restricted area of southern Switzerland. *Wildlife Biology*, 10(1), 137–143.

Grodinsky, C., & Stüwe, M. (1987). With lots of help alpine ibex return to their mountains. *Smithsonian*, 18(9), 68–77.

Grossen, C., Biebach, I., Angelone-Alasaad, S., Keller, L. F., & Croll, D. (2018). Population genomics analyses of European ibex species show lower diversity and higher inbreeding in reintroduced populations. *Evolutionary Applications*, 11(2), 123–139. https://doi.org/10.1111/eva.12490

Grossen, C., Guillaume, F., Keller, L. F., & Croll, D. (2020). Purging of highly deleterious mutations through severe bottlenecks in Alpine ibex. *Nature Communications*. Retrieved from https://www.nature.com/articles/s41467-020-14803-1

Grossen, C., Keller, L. F., & Biebach, I., International Goat Genome Consortium, & Croll, D. (2014). Introgression from domestic goat generated variation at the major histocompatibility complex of Alpine ibex. *PLoS Genetics*, 10(6), e1004438. https://doi.org/10.1371/journal.pgen.1004438

Hess, J. F., Kohl, T. A., Kotrová, M., Rönsch, K., Paprotka, T., Mohr, V., Hutzenlaub, T., Brüggemann, M., Zengerle, R., Niemann, S., & Paust, N. (2020). Library preparation for next generation sequencing: A review of automation strategies. *Biotechnology Advances*, 41, 107537. https://doi.org/10.1016/j.biotechadv.2020.107537

Hoffberg, S. L., Kieran, T. J., Catchen, J. M., Devault, A., Faircloth, B. C., Mauricio, R., & Glenn, T. C. (2016). RAD cap: Sequence capture of dual-digest RAD seq libraries with identifiable duplicates and reduced missing data. *Molecular Ecology Resources*, 16(5), 1264–1278.

Holderegger, R., Balkenhol, N., Bolliger, J., Engler, J. O., Gugerli, F., Hochkirch, A., Nowak, C., Segelbacher, G., Widmer, A., & Zachos, F. E. (2019). Conservation genetics: Linking science with practice. *Molecular Ecology*, *28*(17), 3848–3856. https://doi.org/10.1111/mec.15202

Jiang, Z., Wang, H., Michal, J. J., Zhou, X., Liu, B., Woods, L. C. S., & Fuchs, R. A. (2016). Genome wide sampling sequencing for SNP genotyping: Methods, challenges and future development. *International Journal of Biological Sciences*, *12*(1), 100–108. https://doi.org/10.7150/ijbs.13498

Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, *37*(8), 907–915. https://doi.org/10.1038/s41587-019-0201-4

Kleinman-Ruiz, D., Martínez-Cruz, B., Soriano, L., Lucena-Perez, M., Cruz, F., Villanueva, B., Fernández, J., & Godoy, J. A. (2017). Novel efficient genome-wide SNP panels for the conservation of the highly endangered Iberian lynx. *BMC Genomics*, *18*(1), 1–12. https://doi.org/10.1186/s12864-017-3946-5

Komoroske, L. M., Miller, M. R., O'Rourke, S. M., Stewart, K. R., Jensen, M. P., & Dutton, P. H. (2019). A versatile Rapture (RAD-Capture) platform for genotyping marine turtles. *Molecular Ecology Resources*, *19*(2), 497–511. https://doi.org/10.1111/1755-0998.12980

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of next generation sequencing data. *BMC Bioinformatics*, *15*(1), 356. https://doi.org/10.1186/s12859-014-0356-4

Kosch, T. A., Silva, C. N. S., Brannelly, L. A., Roberts, A. A., Lau, Q., Marantelli, G., Berger, L., & Skerratt, L. F. (2019). Genetic potential for disease resistance in critically endangered amphibians decimated by chytridiomycosis. *Animal: an International Journal of Animal Bioscience*. Retrieved from https://zslpublications.onlinelibrary.wiley.com/doi/abs/10.1111/acv.12459

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Leigh, D. M., Hendry, A. P., Vázquez-Domínguez, E., & Friesen, V. L. (2019). Estimated six per cent loss of genetic variation in wild populations since the industrial revolution. *Evolutionary Applications*, *12*, 1505–1512. https://doi.org/10.1111/eva.12810

Leigh, D. M., Lischer, H. E. L., Grossen, C., & Keller, L. F. (2018). Batch effects in a multiyear sequencing study: False biological trends due to changes in read lengths. *Molecular Ecology Resources*, *18*(4), 778–788. https://doi.org/10.1111/1755-0998.12779

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. & 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Magoč, T., & Salzberg, S. L. (2011). FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, *27*(21), 2957–2963. https://doi.org/10.1093/bioinformatics/btr507

Matala, A. P., Hatch, D. R., Everett, S., Ackerman, M. W., Bowersox, B., Campbell, M., & Narum, S. (2016). What goes up does not come down: The stock composition and demographic characteristics of upstream migrating steelhead differ from post-spawn emigrating kelts. *ICES Journal of Marine Science*, *73*(10), 2595–2605. https://doi.org/10.1093/icesjms/fsw109

May, S. A., McKinney, G. J., Hilborn, R., Hauser, L., & Naish, K. A. (2020). Power of a dual-use SNP panel for pedigree reconstruction and population assignment. *Ecology and Evolution*, *10*(17), 9522–9531. https://doi.org/10.1002/ece3.6645

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. https://doi.org/10.1101/gr.107524.110

Meek, M. H., & Larson, W. A. (2019). The future is now: Amplicon sequencing and sequence capture usher in the conservation genomics era. *Molecular Ecology Resources*, *19*(4), 795–803. https://doi.org/10.1111/1755-0998.12998

Meisner, J., & Albrechtsen, A. (2018). Inferring population structure and admixture proportions in low-depth NGS. *Genetics*, *210*, 719–731. https://doi.org/10.1534/genetics.118.301336

Mick, V., Le Carrou, G., Corde, Y., Game, Y., Jay, M., & Garin-Bastuji, B. (2014). Brucella melitensis in France: Persistence in wildlife and probable spillover from Alpine ibex to domestic animals. *PLoS One*, *9*(4), e94168. https://doi.org/10.1371/journal.pone.0094168

Natesh, M., Taylor, R. W., Truelove, N. K., Hadly, E. A., Palumbi, S. R., Petrov, D. A., & Ramakrishnan, U. (2019). Empowering conservation practice with efficient and economical genotyping from poor quality samples. *Methods in Ecology and Evolution*, *10*(6), 853–859. https://doi.org/10.1111/2041-210X.13173

O'Brien, S. J., Johnson, W. E., Driscoll, C. A., Dobrynin, P., & Marker, L. (2017). Conservation genetics of the cheetah: Lessons learned and new opportunities. *The Journal of Heredity*, *108*(6), 671–677. https://doi.org/10.1093/jhered/esx047

O'Brien, S. J., Roelke, M., Marker, L., Newman, A., Winkler, C., Meltzer, D., Colly, L., Evermann, J., Bush, M., & Wildt, D. (1985). Genetic basis for species vulnerability in the cheetah. *Science*, *227*(4693), 1428–1434.

O'Brien, S. J., Wildt, D. E., Goldman, D., Merril, C. R., & Bush, M. (1983). The cheetah is depauperate in genetic variation. *Science*, *221*(4609), 459–462.

Ouborg, N. J., Pertoldi, C., Loeschcke, V., Bijlsma, R. K., & Hedrick, P. W. (2010). Conservation genetics in transition to conservation genomics. *Trends in Genetics: TIG*, *26*(4), 177–187. https://doi.org/10.1016/j.tig.2010.01.001

Pimm, S. L., Dollar, L., & Bass, O. L. (2006). The genetic rescue of the Florida panther. *Animal Conservation*, *9*(2), 115–122. https://doi.org/10.1111/j.1469-1795.2005.00010.x

Quéméré, E., Rossi, S., Petit, E., Marchand, P., Merlet, J., Game, Y., Galan, M., & Gilot-Fromont, E. (2020). Genetic epidemiology of the Alpine ibex reservoir of persistent and virulent brucellosis outbreak. *Scientific Reports*, *10*(1), 4400. https://doi.org/10.1038/s41598-020-61299-2

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

Razali, H., O'Connor, E., Drews, A., Burke, T., & Westerdahl, H. (2017). A quantitative and qualitative comparison of illumina MiSeq and 454 amplicon sequencing for genotyping the highly polymorphic major histocompatibility complex (MHC) in a non-model species. *BMC Research Notes*, *10*(1), 1–10. https://doi.org/10.1186/s13104-017-2654-1

Reed, D. H., & Frankham, R. (2003). Correlation between fitness and genetic diversity. *Conservation Biology: The Journal of the Society for Conservation Biology*, *17*(1), 230–237. https://doi.org/10.1046/j.1523-1739.2003.01236.x

Roelke, M. E., Martenson, J. S., & O'Brien, S. J. (1993). The consequences of demographic reduction and genetic depletion in the endangered Florida panther. *Current Biology: CB*, *3*(6), 340–350. https://doi.org/10.1016/0960-9822(93)90197-V

R Core Team (2018). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. *Vienna*, https://www.R-project.org

Sard, N. M., Smith, S. R., Homola, J. J., Kanefsky, J., Bravener, G., Adams, J. V., Holbrook, C. M., Hrodey, P. J., Tallon, K., & Scribner, K. T. (2020). RAPTURE (RAD capture) panel facilitates analyses characterizing sea lamprey reproductive ecology and movement dynamics.

Ecology and Evolution, 10(3), 1469–1488. https://doi.org/10.1002/ece3.6001

Savage, A. E., Sredl, M. J., & Zamudio, K. R. (2011). Disease dynamics vary spatially and temporally in a North American amphibian. *Biological Conservation*, 144(6), 1910–1915. https://doi.org/10.1016/j.biocon.2011.03.018

Schoville, S. D., Bonin, A., François, O., Lobreaux, S., Melodelima, C., & Manel, S. (2012). Adaptive genetic variation on the landscape: Methods and cases. *Annual Review of Ecology, Evolution, and Systematics*, 43(1), 23–43. https://doi.org/10.1146/annurev-ecolsys-110411-160248

Shafer, A. B. A., Wolf, J. B. W., Alves, P. C., Bergström, L., Bruford, M. W., Brännström, I., Colling, G., Dalén, L., De Meester, L., Ekblom, R., Fawcett, K. D., Fior, S., Hajibabaei, M., Hill, J. A., Hoezel, A. R., Höglund, J., Jensen, E. L., Krause, J., Kristensen, T. N., ... Zieliński, P. (2015). Genomics and the challenging translation into conservation practice. *Trends in Ecology & Evolution*, 30(2), 78–87. https://doi.org/10.1016/j.tree.2014.11.009

Siddle, H. V., Kreiss, A., Eldridge, M. D. B., Noonan, E., Clarke, C. J., Pyecroft, S., Woods, G. M., & Belov, K. (2007). Transmission of a fatal clonal tumor by biting occurs due to depleted MHC diversity in a threatened carnivorous marsupial. *Proceedings of the National Academy of Sciences of the United States of America*, 104(41), 16221–16226. https://doi.org/10.1073/pnas.0704580104

Skotte, L., Korneliussen, T. S., & Albrechtsen, A. (2013). Estimating individual admixture proportions from next generation sequencing data. *Genetics*, 195(3), 693–702. https://doi.org/10.1534/genetics.113.154138

Steiner, C. C., Putnam, A. S., Hoeck, P. E. A., & Ryder, O. A. (2013). Conservation genomics of threatened animal species. *Annual Review of Animal Biosciences*, 1, 261–281. https://doi.org/10.1146/annurev-animal-031412-103636

Steyer, K., Kraus, R. H. S., Mölich, T., Anders, O., Cocchiararo, B., Frosch, C., Geib, A., Götz, M., Herrmann, M., Hupe, K., Kohnen, A., Krüger, M., Müller, F., Pir, J. B., Reiners, T. E., Roch, S., Schade, U., Schiefenhövel, P., Siemund, M., ... Nowak, C. (2016). Large-scale genetic census of an elusive carnivore, the European wildcat (*Felis s. silvestris*). *Conservation Genetics*, 17(5), 1183–1199. https://doi.org/10.1007/s10592-016-0853-2

Syvänen, A. C. (2001). Accessing genetic variation: Genotyping single nucleotide polymorphisms. *Nature Reviews Genetics*, 2(12), 930–942. https://doi.org/10.1038/35103535

Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J., & Prins, P. (2015). Sambamba: Fast processing of NGS alignment formats. *Bioinformatics*, 31(12), 2032–2034. https://doi.org/10.1093/bioinformatics/btv098

Taylor, H. R., Dussex, N., & van Heezik, Y. (2017). De-extinction needs consultation. *Nature Ecology & Evolution*, 1(1), 1–2. https://doi.org/10.1038/s41559-017-0198

Terrier, G., & Rossi, P. (1994). Le bouquetin (*Capra ibex ibex*) dans les Alpes Maritimes Franco-Italiennes: Occupation de l'espace, colonisation et régulation naturelles. *Travaux Scientifiques Du Parc National De La Vanoise*, 18, 271–287.

Todesco, M., Pascual, M. A., Owens, G. L., Ostevik, K. L., Moyers, B. T., Hübner, S., Heredia, S. M., Hahn, M. A., Caseys, C., Bock, D. G., & Rieseberg, L. H. (2016). Hybridization and extinction. *Evolutionary Applications*, 9(7), 892–908. https://doi.org/10.1111/eva.12367

Toïgo, C., Brambilla, A., Grignolio, S., & Pedrotti, L. (2020). *Capra ibex. The IUCN Red List of Threatened Species*. 2020: e.T42397A161916377

Tosser-Klopp, G., Rothschild, M. F., Huson, H. J., Nicolazzi, E. L., Sonstegard, T. S., Amills, M., Riggs, P., Van Tassell, C. P., Marsan, P. A., Stella, A., Sayre, B., & Rupp, R., & et al. (2015). Update on the international goat genome consortium projects. *Plant and Animal Genome*. Retrieved from https://hal.inrae.fr/hal-02800945

Tschirren, B., Andersson, M., Scherman, K., Westerdahl, H., Mittl, P. R. E., & Råberg, L. (2013). Polymorphisms at the innate immune receptor TLR2 are associated with Borrelia infection in a wild rodent population. *Proceedings. Biological Sciences/The Royal Society*, 280(1759), 20130364.

Turner, A. K., Begon, M., Jackson, J. A., Bradley, J. E., & Paterson, S. (2011). Genetic diversity in cytokines associated with immune variation and resistance to multiple pathogens in a natural rodent population. *PLoS Genetics*, 7(10), e1002343. https://doi.org/10.1371/journal.pgen.1002343

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K. V., Altshuler, D., Gabriel, S., & DePristo, M. A. (2013). From FastQ data to high confidence variant calls: The Genome Analysis Toolkit best practices pipeline. *Current Protocols in*, 43, 11.10.1–11.10.33.

Willisch, C. S., Biebach, I., Koller, U., Bucher, T., Marreros, N., Ryser-Degiorgis, M.-P., Keller, L. F., & Neuhaus, P. (2012). Male reproductive pattern in a polygynous ungulate with a slow life-history: The role of age, social status and alternative mating tactics. *Evolutionary Ecology*, 26(1), 187–206. https://doi.org/10.1007/s10682-011-9486-6

Witzenberger, K. A., & Hochkirch, A. (2014). The genetic integrity of the ex situ population of the European wildcat (*Felis silvestris silvestris*) is seriously threatened by introgression from domestic cats (*Felis silvestris catus*). *PLoS One*, 9(8), e106083. https://doi.org/10.1371/journal.pone.0106083

Wright, B., Morris, K., Grueber, C. E., Willet, C. E., Gooley, R., Hogg, C. J., O'Meally, D., Hamede, R., Jones, M., Wade, C., & Belov, K. (2015). Development of a SNP-based assay for measuring genetic diversity in the Tasmanian devil insurance population. *BMC Genomics*, 16(1), 1–11. https://doi.org/10.1186/s12864-015-2020-4

WWF (2018). In M. Grooten, & R. E. A. Almond (Eds.), *Living Planet Report - 2018: Aiming Higher*. WWF.

Yates, A. D., Achuthan, P., Akanni, W., Allen, J., Allen, J., Alvarez-Jarreta, J., Amode, M. R., Armean, I. M., Azov, A. G., Bennett, R., Bhai, J., Billis, K., Boddu, S., Marugán, J. C., Cummins, C., Davidson, C., Dodiya, K., Fatima, R., Gall, A., ... Flicek, P. (2019). Ensembl 2020. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkz966

Yu, Y., Luo, Z., Wang, Q., Zhang, Q., Zhang, X., Xiang, J., & Li, F. (2020). Development of high throughput SNP genotyping approach using target sequencing in Pacific white shrimp and its application for genetic study. *Aquaculture*, 528, 735549. https://doi.org/10.1016/j.aquaculture.2020.735549

Zhu, Y., Grueber, C., Li, Y., He, M., Hu, L., He, K., Liu, H., Zhang, H., & Wu, H. (2020). MHC-associated *Baylisascaris schroederi* load informs the giant panda reintroduction program. *International Journal for Parasitology: Parasites and Wildlife*, 12, 113–120. https://doi.org/10.1016/j.ijppaw.2020.05.010

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.