MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2016) **16**, 1147–1164 doi: 10.1111/1755-0998.12560

SPECIAL ISSUE: SEQUENCE CAPTURE

SNP discovery in candidate adaptive genes using exon capture in a free-ranging alpine ungulate

GRETCHEN H. ROFFLER,*† STEPHEN J. AMISH,‡ SETH SMITH,‡ TED COSART,‡ MARTY KARDOS,‡§ MICHAEL K. SCHWARTZ§¶ and GORDON LUIKART‡**

*Alaska Science Center, U.S. Geological Survey, 4210 University Drive, Anchorage, AK 99508, USA, †Wildlife Biology Program, Department of Ecosystem Sciences and Conservation, College of Forestry and Conservation, University of Montana, Missoula, MT 59812, USA, ‡Fish and Wildlife Genomics Group, Division of Biological Sciences, University of Montana, Missoula, MT 59812, USA, §Evolutionary Biology Centre, Uppsala University, SE-75236 Uppsala, Sweden, ¶US Forest Service Rocky Mountain Research Station, National Genomics Center for Wildlife and Fish Conservation, 800 E. Beckwith Ave., Missoula, MT 59801, USA, **Flathead Lake Biological Station, University of Montana, Polson, MT 59860, USA*

Abstract

Identification of genes underlying genomic signatures of natural selection is key to understanding adaptation to local conditions. We used targeted resequencing to identify SNP markers in 5321 candidate adaptive genes associated with known immunological, metabolic and growth functions in ovids and other ungulates. We selectively targeted 8161 exons in protein-coding and nearby 5′ and 3′ untranslated regions of chosen candidate genes. Targeted sequences were taken from bighorn sheep (*Ovis canadensis*) exon capture data and directly from the domestic sheep genome (*Ovis aries* v. 3; oviAri3). The bighorn sheep sequences used in the Dall's sheep (*Ovis dalli dalli*) exon capture aligned to 2350 genes on the oviAri3 genome with an average of 2 exons each. We developed a microfluidic qPCR-based SNP chip to genotype 476 Dall's sheep from locations across their range and test for patterns of selection. Using multiple corroborating approaches (LOSITAN and BAYESCAN), we detected 28 SNP loci potentially under selection. We additionally identified candidate loci significantly associated with latitude, longitude, precipitation and temperature, suggesting local environmental adaptation. The three methods demonstrated consistent support for natural selection on nine genes with immune and disease-regulating functions (e.g. Ovar-DRA, APC, BATF2, MAGEB18), cell regulation signalling pathways (e.g. KRIT1, PI3K, ORRC3), and respiratory health (CYSLTR1). Characterizing adaptive allele distributions from novel genetic techniques will facilitate investigation of the influence of environmental variation on local adaptation of a northern alpine ungulate throughout its range. This research demonstrated the utility of exon capture for gene-targeted SNP discovery and subsequent SNP chip genotyping using low-quality samples in a nonmodel species.

Keywords: candidate genes, exon capture, next-generation sequencing, *Ovis dalli dalli*, population genomics, SNP chip

Received 16 November 2015; revision received 4 June 2016; accepted 7 June 2016

Introduction

Identifying patterns of adaptive variation in nonmodel species in natural environments has been challenging due to the difficulties of conducting experiments (e.g. controlled breeding) and obtaining direct measurements (e.g. common garden), especially for species with long generation intervals (Holderegger & Wagner 2008; Schoville *et al.* 2012). The increased accessibility of high-throughput sequencing and genotyping methods has facilitated discovery and genotyping of SNPs in nonmodel organisms (Brumfield *et al.* 2003; Slate *et al.* 2009; Stapley *et al.* 2010), including SNPs in functional or adaptive genes (Cosart *et al.* 2011). These genomic tools facilitate the identification of genes with an adaptive function with higher certainty than with traditional genetic or observation-based methods (Nielsen *et al.* 2009; Eckert *et al.* 2010; Hohenlohe *et al.* 2011), including candidate genes associated with fitness traits. Candidate genes have a known function in a particular process, a metabolic pathway, or are related to a phenotype, and have been found to be under selection in previous studies. Therefore, candidate genes have a higher likelihood of being under selection than other genes, especially if

Correspondence: Gretchen H. Roffler, Alaska Department of Fish and Game, Division of Wildlife Conservation, 802 3rd Street, Douglas, AK 99824, USA, Fax: 907-465-4272; E-mail: gretchen.roffler@alaska.gov

their function relates to selection pressures that vary across the study area (Luikart *et al.* 2003; Nielsen 2005). These recent genomic developments are making broad-scale investigations of adaptive divergence in free-ranging species more possible than in the past (Allendorf *et al.* 2010; Ekblom & Wolf 2014; Ellegren 2014).

Exon capture (Hodges *et al.* 2007) is an approach for discovering SNPs in wild species that are closely related to domestic species for which a sequenced genome is available and has been used to identify candidate genes associated with disease and other fitness traits in wild ungulates (Cosart *et al.* 2011). Exons are expressed portions of the genome that code for proteins and are therefore generally well conserved. Thus, it is possible to use the fully sequenced genomes of model species (e.g. domestic sheep; Dalrymple *et al.* 2007; Kijas *et al.* 2009; Miller *et al.* 2011) as a reference to identify candidate SNPs in a related nonmodel species. This information in combination with a growing body of animal science studies on domestic sheep breeds (Ibeagha-Awemu *et al.* 2008; Lühken 2012; Heaton *et al.* 2014) can advance research on local adaptation by targeting candidate genes. Candidate genes have been identified in wild sheep and other vertebrates related to immune function and parasite resistance and applied to studies of free-ranging sheep populations to test for areas of adaptive differentiation across a broad geographical scale (Worley *et al.* 2006; Johnston *et al.* 2011; Luikart *et al.* 2011). Furthermore, application of genomic data sets and landscape genomic approaches has recently identified climate-mediated selection in sheep species (Joost *et al.* 2007; Pariset *et al.* 2009; Lv *et al.* 2014) and provides evidence of the value of these methods to understand adaptation to their environments.

Dall's sheep have evolved in Arctic and sub-Arctic environments for >1 million years (Kurten & Anderson 1980; Bunch *et al.* 2006) and are therefore expected to express specific adaptations to variability in environmental factors that affect individual fitness. Environmental extremes are typical in northern latitudes, thus it would be expected that selection on genes associated with specific traits related to physiological limitations and phenological timing would result. For example, exposure to cold temperatures and long winters would select genes associated with metabolic, thermogenic and circadian rhythm functions. The period of lambing in mountain sheep is correlated with plant phenology because it greatly affects the nutritional requirements of lactation which influences lamb survival (Bunnell 1982; Festa-Bianchet 1988; Rachlow & Bowyer 1991). Therefore, the relatively short summers in northern latitudes provide an abbreviated window of time during which ungulates must maximize their body condition and rear offspring. These conditions would select genes associated with signal pathways such as growth factors and lactation. Finally, pathogen exposure levels and parasite abundance in northern ungulates tend to decrease with latitude (Kutz *et al.* 2005; Jenkins *et al.* 2006; Mainguy *et al.* 2007). Therefore, we would expect greater variation of genes associated with immune function at the southern end of the latitude gradient within the Dall's sheep subspecies range.

Our overarching objective was to develop a set of SNPs in candidate genes in a nonmodel organism and to extend this molecular approach into natural landscapes at a broad spatial scale to detect patterns of local adaptation. Based on ecological knowledge of Dall's sheep, we developed hypotheses of factors limiting their fitness. We then selected genes with known functions related to these factors (immunological, metabolic, growth and environmental signalling functions) in ovids, domestic cattle or other well-studied taxa. We used exon capture to selectively target these genomic regions and then genotyped sheep at the range-wide scale. Finally, we identified loci putatively under selection using two F_{ST} outlier approaches and correlation of candidate gene allele frequencies with environmental variables.

Materials and methods

Sample selection and extraction

We collected samples throughout the Dall's sheep subspecies' range in Alaska and the Yukon Territory (Fig. 1). There are ~90 000 Dall's sheep in this region (Festa-Bianchet 2008) distributed across a latitudinal and ecological gradient from the southern, coastal influenced regions of Alaska, to the arid Brooks Range in the Arctic representing the geographical limit of the subspecies. We attempted to include samples representing the major mountain ranges inhabited by Dall's sheep, as genetic structure has been demonstrated to coincide with these landscape features (Worley *et al.* 2004; Roffler *et al.* 2014). We obtained 476 muscle or blood samples collected during 1998–2014 from hunter-killed rams or previous research efforts that had been preserved frozen or stored in tissue preservation buffer or 100% ethanol at room temperature (Table 1).

Genomic DNA was extracted using Gentra Puregene Blood and Tissue kits following the recommended protocols (Qiagen N.V., Hilden, Germany). The quality of the sample DNA was evaluated at several points during the project. Due to the age, collection method and/or preservation conditions, some samples produced low DNA yields or showed high levels of inorganic contaminants. Several different extraction protocols were tested, compared and optimized in an effort to get the highest quality and concentration of DNA possible (Appendix S1, Supporting information).

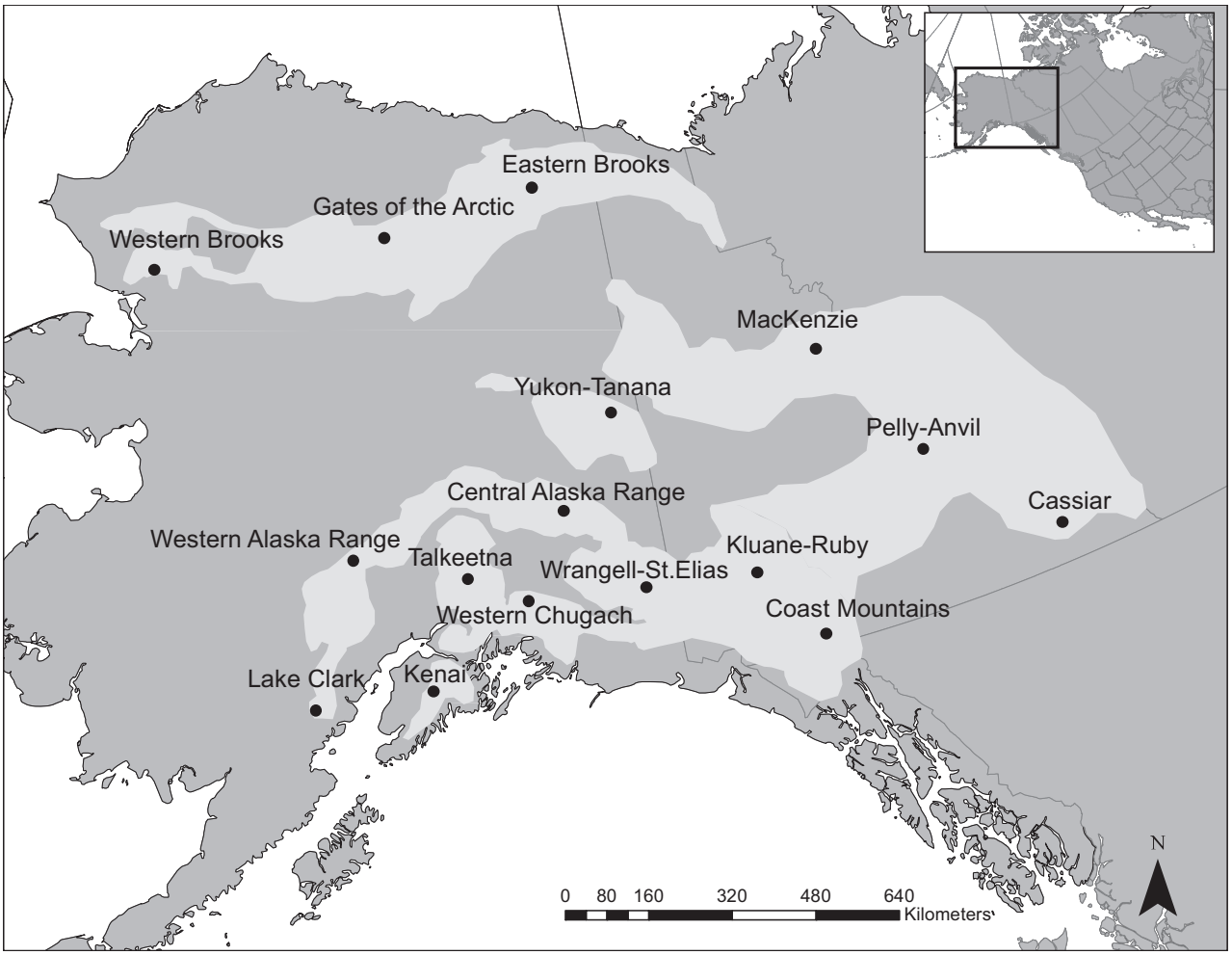

Fig. 1 Map of sample locations and subspecies range (light grey) of Dall's sheep.

Exon target sequences

Sequences for in-solution targeted re-sequencing and downstream SNP discovery and assay design were produced from several sources. Forty-eight per cent of the Dall's sheep targeted exon sequences came from sequence reads from two bighorn sheep (*Ovis canadensis*) from a previous study. The sequence reads were aligned to the domestic sheep (*Ovis aries*) genome v1 (oviAri1; Archibald *et al.* 2010) with BWA (v5.9-r16) and genotyped using SAMTOOLS v0.1.18 (Li *et al.* 2009). The genotyped positions for one of the bighorn individuals were used to create sequences based on exon intervals in the oviAri1 genome, as annotated by the collection of multiple species gene alignments to the oviAri1 genome (obtained at the UCSC Genome Browser server at http://hgdownload.cse.ucsc.edu/goldenPath/oviAri1/database/xenoRefGene.txt.gz). The final set of bighorn sheep sequences consisted of those that aligned to complete exons not only in the sheep genome, but also, in a second check for complete exon coverage, in the alignment of the same reads to the well-annotated cow genome BTAU v4.0 (Elsik *et al.* 2009; NCBI Assembly Accession no. GCF_000003205.2). The bighorn consensus sequences are available on request. The remaining 52% of targets were added from oviAri2 genome (www.sheephapmap.org/isgc_genseq.htm), including additional candidate gene exons and exons in chromosomal regions not enriched in the two bighorn sheep (Fig. 2). Overall, a total of 10 032 genomic regions including 4029 annotated genes were targeted in Dall's sheep samples using 14 205 RNA baits (sequences available upon request).

Exon capture

One and a half micrograms of genomic DNA in 50 μL TE-buffer was fragmented to a median size of 150 bp using the Covaris-S210 instrument following the

Table 1 Summary information for sampled populations of Dall's sheep in Alaska and the Yukon Territory, 2000–2010 (n = 476). Shown for each sample location are the coordinates of the sample area centroid, the number of samples included in the analyses, the expected (H_e) and observed (H_o) heterozygosities, the proportion of polymorphic loci (P_P) and inbreeding coefficient (F_{IS})

Mountain range	Sample location	n	Latitude	Longitude	H_e	H_o	P_P	F_{IS}
Alaska	Central Alaska Range	32	63.2179	−144.1943	0.378	0.321	0.97	0.150
Alaska	Lake Clark NPP	12	60.0805	−154.1653	0.248	0.262	0.73	−0.057
Alaska	Western Alaska Range	21	68.5360	−144.5570	0.339	0.294	0.96	0.134
Brooks	Eastern Brooks Range	33	67.6730	−161.8960	0.341	0.285	0.98	0.165
Brooks	Western Brooks Range	32	67.8380	−160.8240	0.339	0.294	0.96	0.149
Brooks	Gates of the Arctic NPP	25	68.0976	−151.9260	0.288	0.272	0.81	0.056
Kenai	Kenai	10	60.2016	−150.1430	0.259	0.278	0.70	−0.073
Talkeetna	Talkeetna	10	62.2870	−148.4510	0.352	0.332	0.96	0.056
Chugach	Western Chugach	71	61.6007	−145.7410	0.393	0.322	0.99	0.181
Wrangell	Wrangell-St. Elias NPP	122	61.9392	−123.8470	0.385	0.312	0.99	0.188
St. Elias	Kluane-Ruby Ranges	13	61.9402	−140.6860	0.255	0.261	0.70	−0.025
Yukon	Cassiar	2	60.1340	−131.4740	0.265	0.346	0.62	−0.306
Yukon	Coast Mountains	19	60.6760	−136.0780	0.293	0.258	0.87	0.120
Yukon	Pelly-Anvil Ranges	16	62.3590	−133.8608	0.320	0.244	0.92	0.238
MacKenzie	MacKenzie	8	64.7875	−133.4980	0.225	0.249	0.64	−0.106
Yukon-Tanana Uplands	Yukon-Charley River NP	50	64.3919	−142.5850	0.308	0.286	0.93	0.073

recommended settings. Sheared samples were cleaned with Agencourt AMPure XP Beads (Beckman Coulter, Fullerton, CA, USA) using 120 μL of resuspended beads and 40 μL of DNA. DNA fragmentation was evaluated based on the distribution of fragment sizes using capillary electrophoresis on Agilent DNA1000 chips, and the concentration of the DNA was estimated by PicoGreen assay (Invitrogen). Individual sequencing libraries were prepared with 1 μg of fragmented DNA using Bioo Scientific (Austin, TX, USA) NEXTflex DNA kits and barcodes following the standard protocol. Six in-solution hybridization captures of target DNA using biotinylated RNA baits from MYCROARRAY v1.3.5 were carried out following the recommended protocol, with three samples multiplexed in each reaction.

Eighteen individual samples were sequenced in a single HiSeq 2000 (Illumina Inc., San Diego, CA, USA) lane producing approximately 358 million 100 bp single-end reads. The ILLUMINACLIP function with the TruSeq2-SE adapter sequences in TRIMMOMATIC v0.27 (Lohse *et al.* 2012) was used with the recommended settings to filter and trim the raw reads. The number of reads after trimming, the per-base and the per-sequence quality of the trimmed reads, were checked using FASTQC v0.10.1 (www.bioinformatics.babraham.ac.uk/projects/fastqc). The quality filtered and trimmed sequence reads were aligned using BWA mapper software v0.6.2 (Li & Durbin 2009) to the domestic sheep genome oviAri3. CoverageBED in BEDTOOLS v2.24.0 (Quinlan & Hall 2010) was used to calculate the number of reads mapped to each targeted exon nucleotide.

SNP discovery

Samples were genotyped following the Broad Institute's guidelines for nonmodel species (https://www.broadinstitute.org/) and a conservative set of hard filters to remove potentially false SNPs. Using GATK's recommended tools for genotyping (https://www.broadinstitute.org/gatk/guide), and GATK v3.3 (McKenna *et al.* 2009), each sample was genotyped separately with the HaplotypeCaller, followed by a joint genotyping by GATK's GentotypeGVCFs. GATK's VariantFiltration tools, combined with an in-house script, were used to remove potentially false-positive SNP calls. A SNP genotyped locus was excluded if it showed one or more of the following: (i) an allele was observed <8 times; (ii) the inbreeding coefficient at the locus was under −0.8824, the upper bound of the lowest 5% of the inbreeding coefficients calculated for all loci under consideration (GATK's InbreedingCoeff annotator: https://www.broadinstitute.org/gatk/gatkdocs/org_broadinstitute_gatk_tools_walkers_annotator_InbreedingCoeff.php); (iii) more than two individual genotype calls had genotype qualities <30; (iv) more than two individual genotype calls had low coverage depth such that a homozygous alternate genotype call had coverage depth under five reads, or a heterozygous genotype call had coverage depth under eight reads; (v) more than two individual genotypes had coverage depth exceeding 300 reads; (vi) more than 11 individuals genotype calls were missing (i.e. the HaplotypeCaller did not find sufficient read data to call a genotype); (vii) more than 12 of the 18 individuals were genotyped as heterozygotes

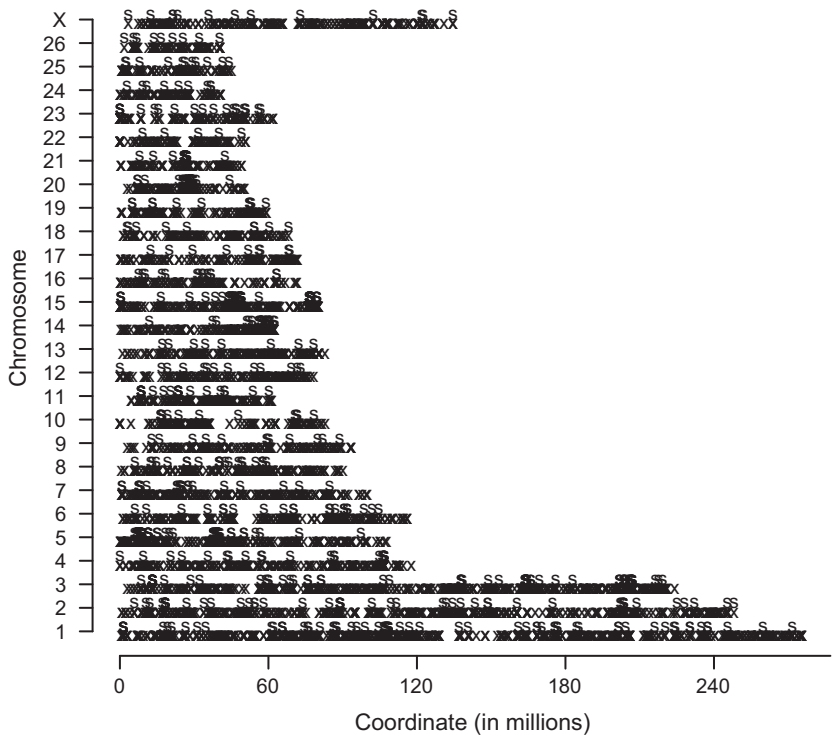

Fig. 2 On the *Ovis aries* genome (Oar_v3.1, GenBank Assession no. 000298735.1), the base pair coordinates of the targeted exon regions are marked with an 'x'. Coordinates of the 964 SNPs that passed our filtering criteria are marked above their corresponding target with an 'S'.

(erroneous read mapping is suspected when substantially more than half of the individuals are genotyped as heterozygous); (viii) a SNP locus was part of a cluster of at least three SNP loci inside a 10 bp window; or (ix) a SNP locus was within 10 bp of an insertion/deletion call.

Consensus sequence generation and genotyping assay design

We used our list of candidate genes to identify exons that could be used to test our hypotheses and contained variable SNPs for consensus sequence generation and later genotyping assay development. For each filtered SNP locus, one sequence was generated for each individual, consisting of the filter-passed SNP and its flanking positions, 101 base pairs upstream, 100 base pairs downstream. Each position in the sequence had either a diploid genotype or an 'N' inserted at positions for which the individual had low coverage ($\leq 3\times$) or low genotype quality (phred score ≤ 30). These individual sequences were then aligned using MAFFT, v7.215 (Katoh & Standley 2013). At each position, a nucleotide was included in the consensus genotype if it was observed at least eight times. A position was designated as having an unknown genotype ('N') if among the individuals, no genotype was observed in 20 or more of the 36 possible observations, more than 16 of the 36 possible observations were gapped by the alignment, or if the sum of nonobservations plus gaps was 20 or more.

Consensus sequences were also designed for five SNPs within and near RXFP2 (Appendix S1, Supporting information), a gene that has been inferred to be under strong selection in both wild bighorn (Kardos *et al*. 2015) and domestic (Kijas *et al*. 2012; Johnston *et al*. 2013) sheep species, and putatively associated with variation in horn size in bighorn sheep (Poissant *et al*. 2012). All consensus sequences were then aligned to the NCBI nucleotide collection (nt) database using BLASTN in BLASTN v2.2.25 (Altschul *et al*. 1990). If a position was not already annotated, the top BLASTN hit was used to annotate the consensus sequence if the E-value for the alignment was $<1 * 10^{-20}$. Consensus sequence quality and suitability for assay design were measured by calculating the number of Ns within 10 bp and 50 bp of all target SNPs. Allele frequencies, observed heterozygosity, and expected heterozygosity were then calculated at the population level for each locus. SNPs were also identified as either transitions or transversions. SNPs were preferentially targeted for assay design if they were in genes of interest (see above); had a consensus sequence with no Ns within 10 bp and fewer than 5Ns within 50 bp of the target SNP; were transversions; or showed evidence of having allele frequencies that were highly variable across the range. Specifically, if the frequency of the minor allele was <0.1 in the eight samples that were sequenced

from one source population, but >0.25 in at least one of the other populations with exon capture genotypes, then the SNP was preferentially targeted for assay design. Additionally, loci that were fixed for different alleles in any combination of two populations were retained. Consensus sequences for SNPs matching these criteria were submitted to Fluidigm for SNP Type assay design using D3 assay design software (Fluidigm Inc., San Francisco, CA, USA). This proprietary software takes into account primer and amplicon specificity and seeks to avoid placing primers over SNPs, highly structured GC rich regions, and repeats (https://www.fluidigm.com/faq/aa-31). Because many of our samples were low quality, we relied on Fluidigm SNP Type assays as this technology has proved to be the most sensitive and reliable for genotyping low DNA concentration samples (Campell & Narum 2009). Primers and sequence probes for SNP assays tested are available upon request.

Genotyping assay evaluation

All assays were run on the Fluidigm EP1 genotyping system using 96.96 Dynamic Array IFCs under the manufacturer recommended reaction conditions. Prior to genotyping, all samples were quantitated using a Nanodrop 2000 Spectrophotometer (Thermo Scientific, Waltham, MA, USA). Samples with concentrations below 30 ng/μL were preamplified using the standard Fluidigm Specific Target Amplification (STA) protocol. STA involves 15 cycles of multiplex PCR using up to 96 pairs of locus-specific (LSP) and STA primers provided with SNP type assays. Amplicons were then diluted 100× and genotyped using the EP1 system.

To test assays, 95 samples from across the study area were each genotyped twice, including a subset of eight control samples that were used in the targeted resequencing experiment. Prior to calculating call rate and concordance statistics, seven samples with abnormally low call rates (<80%) were removed. This step was meant to help avoid a scenario where few, low-quality samples, controlled the majority of the variation in assay quality. Several criteria were used to evaluate the performance of each SNP assay for genotyping individual Dall's sheep samples. First, the plot of the SNP alleles' fluorescent intensity was checked for separation and clustering of each genotype class (heterozygotes and each homozygote) and assays with diffuse or unclear clusters were marked as potentially unusable. We then calculated the call rate for each assay across all replicates. Each sample was run at least twice, so genotypes from each independent PCR reaction were compared for agreement. In addition, genotypes between the original exon sequencing data and the SNP assays were checked for concordance. Finally, five populations (N from 9 to 26) were used to calculate Hardy–Weinberg proportions (HWP) and gametic disequilibrium (GD) between loci that passed initial genotype and clustering quality controls.

Descriptive statistics

We tested for deviation from HWP and for GD using GENEPOP (Raymond & Rousset 1995) across the sampling locations and SNP loci using the Markov Chain Monte Carlo (MCMC) approximation of Fisher's exact test and a simulated exact test respectively. We ran 10 000 dememorizations, 100 batches and 5000 iterations, and applied a Bonferroni-corrected alpha level of 0.05. For each sampling location, we calculated observed and expected heterozygosities (H_o and H_e) using GENEPOP (Table 1). We calculated F_{IS} metrics for each genotyped individual within each population. F_{IS} values range from 1.0 to −1.0, with values less than −0.6 indicating genotyping error (e.g. 2 loci), or selection against homozygotes (Waples 2014).

Identifying outlier loci

We differentiated between neutral SNP loci and those putatively under selection or linked to genes under selection using two approaches. We first tested for departures from neutrality using an F_{ST} outlier approach based on simulation methods (Beaumont & Nichols 1996) implemented in LOSITAN (Antão *et al.* 2008). We used 500 000 replicates assuming an infinite alleles mutation model, with forced mean F_{ST} calculated for each pairwise comparison, and tested for outliers at the 95 and 99% confidence interval levels and used a false discovery rate (FDR) of (0.01 and) 0.05. We also used BAYESCAN ver. 2.1 (Foll & Gaggiotti 2008), a Bayesian approach that directly estimates the posterior probability of a locus being under selection by comparing global and population-specific allele frequencies derived from F_{ST} coefficients. To identify loci under selection using BAYESCAN, we first performed 20 pilot runs of 50 000 iterations, followed by 100 000 iterations on a sample size of 5000 and thinning interval of 10. We used a prior odds value of 10 and identified outlier loci by comparing posterior probabilities and threshold values obtained from the FDR. We identified outlier loci as those with FDR q values <0.05 and 0.01, which approximately correspond to α values of 0.05 and 0.01. LOSITAN and BAYESCAN have different assumptions and algorithms which may lead to differences in the number of outlier loci detected. LOSITAN assumes an island model (with equal population sizes, F_{ST} variances and migration rates), whereas BAYESCAN does not make assumptions about F_{ST} variance equivalence and allows population and locus-specific effects in models. Simulation studies have demonstrated a higher

type I error of LOSITAN vs. BAYESCAN due to the failure to accommodate unequal F_{ST} variances (Peréz-Figueroa *et al.* 2010; Narum & Hess 2011). Type I error for balancing selection was particularly high in this scenario (Narum & Hess 2011; Lotterhos & Whitlock 2014).

Environmental correlations

We tested the hypothesis that loci under selection are correlated with particular environmental variables by implementing a sampling design distributed along environmental gradients (Manel *et al.* 2010). We selected environmental and landscape variables (Appendix S2, Table S1, Supporting information) to use in association tests with SNP loci based on a priori hypotheses about ecological factors that affect Dall's sheep fitness. To test for associations between environmental characteristics and allele frequencies, we used an individual-based spatial analysis method (Sam*β*ada; Joost *et al.* 2007). Sam*β*ada uses logistic regressions to test for associations between all possible pairs of environmental variables and allele frequencies at spatially referenced sample locations and thus estimates the probability of the presence of a specific genetic marker given the sampling site environmental characteristics. Models including and excluding the environmental variables are compared through examination of the significance of regression coefficients. Model significance is then evaluated with likelihood ratio (*G*) and Wald tests; if both tests indicate an environmental variable is more informative than a model with a constant only, the model is considered significant (Joost *et al.* 2007). We recoded each allele from the 188 SNP loci as present ('1') or absent ('0'), producing 564 SNP genotypes (i.e. 00, 01, 11 for each of the 188 SNPs). These genotypes were tested for associations with 25 environmental variables (Table S1, Supporting information), resulting in 14 100 tests. We applied a Bonferroni-corrected threshold corresponding to $\alpha = 0.05$ and 0.01 to significant tests.

Results

Exon capture

Approximately 358 million 100 bp single-end reads were generated in the single HiSeq 2000 (Illumina Inc.) sequencer lane. We obtained an average of approximately 22 million high quality sequence reads from each of 16 individuals distributed across the study area. Two individuals showed low read numbers (<500 000) and failed several FASTQC per-base and per-sequence quality tests and were removed from the data set. After duplicate removal and realignment, individual bam files were used to calculate coverage for targeted exons using CoverageBed. Across these 14 samples, a read depth of 5× or greater was observed for 85% of nucleotides in targeted exons.

Genotyping assay design and evaluation

The initial genotyping yielded 23 757 potential SNPs. After applying our conservative SNP filtration criteria necessary for consensus sequence generation, 1311 SNPs, each of which was the centre of a 202 bp sequence, comprised the initial set of potential SNP assay sequences.

Of the 1311 sequences submitted to Fluidigm, 987 were suitable for SNP type assay design. Two hundred and sixty-one assays were selected from genes of interest and from across the genome and tested on a Fluidigm EP1. Call rates varied from 0% to 100% across the tested assays; however, 234 assay produced call rates >95% (Table 2). Duplicate genotype concordance varied from 64.7% to 100%, with 249 assays having genotypes that were concordant in >95% of comparisons. Concordance between exon capture and SNP type genotypes ranged from 0% to 100%, with 175 assays producing genotypes that matched those from exon capture in >95% of cases.

One hundred and eighty-eight assays were selected to be included in the final SNP panel used to genotype all 476 individuals. The majority of assays were chosen because they had high quality genotype clustering, high call rates and high levels of concordance. However, some loci were selected as a priori targets because they are located in a gene of interest.

SNP chip descriptive statistics

After Bonferroni correction, 230 of 3008 tests for HWP were significant. Seven loci deviated from HWP in >3 populations, thus were removed from further analyses. Gametic disequilibrium was detected in 140 of the 281 248 tests after Bonferroni correction, although GD was not observed for any SNP pair in more than three populations. Negative *F* is values (less than −0.6) were observed for 12 loci across seven populations; no more than two observations per population and no locus was homozygote deficient. The H_o and H_e ranged between 0.244–0.346 and 0.225–0.393, respectively (Table 1).

Outlier tests

We detected a total of 56 and 31 F_{ST} outlier loci potentially under selection using LOSITAN at the 0.05 and 0.01 thresholds, respectively (Table 3). Of the outlier loci, 15 were putatively under directional and 11 under balancing selection at the more conservative 0.01 threshold (Table 3). Using BAYESCAN, we identified a total of 28 and 21 outlier loci at the 0.05 and 0.01 thresholds, respectively, among which 12 were possibly under directional

Table 2 Number of SNP type assays tested by chromosome, mean call rate, duplicate genotype concordance and concordance of SNP type assay genotype with exon capture genotypes

Chromosome	*N*	Call rate	Duplicate genotype concordance	SNP assay and exon capture genotype concordance
1	33	95.00	99.24	92.51
2	21	96.99	99.12	86.03
3	14	97.84	98.41	80.63
4	8	98.71	98.85	100.00
5	16	88.91	96.94	72.71
6	10	96.61	99.08	95.88
7	10	99.64	99.56	98.33
8	5	99.08	98.43	90.00
9	8	99.86	99.15	96.88
10	13	94.40	99.42	76.34
11	6	100.00	99.45	60.42
12	5	98.75	98.25	88.50
13	3	99.81	99.62	100.00
14	2	100.00	98.32	100.00
15	8	99.64	99.43	100.00
16	6	99.62	99.07	100.00
17	3	99.04	100.00	100.00
18	5	99.66	98.60	100.00
19	8	87.50	99.47	91.07
20	32	91.01	99.20	76.83
21	4	97.99	99.43	100.00
22	1	99.43	100.00	100.00
23	6	99.23	93.33	89.29
24	5	98.85	99.30	100.00
25	1	99.43	98.84	100.00
26	3	97.70	98.94	81.25
M	7	99.67	97.21	93.88
X	18	99.77	99.88	83.99
Total	261	96.17	98.87	87.85

(upper right corner Fig. 3a), and nine under balancing selection (lower right corner Fig. 3a). More outliers were detected with LOSITAN than BAYESCAN, and the majority (25 of 29) loci detected by LOSITAN but not by BAYESCAN were for balancing selection. The most robust support (demonstrated by coinciding results of the two methods) was for a subset of 28, and 15 markers were identified as outliers (at the 0.05 and 0.01 thresholds, respectively), including nine for directional and six for balancing selection at the most conservative threshold.

The nine loci putatively under directional selection included loci in immune functions genes including Ovar-DRA (Major Histocompatibility Complex, Class II, DR Alpha), Basic Leucine Zipper Transcription Factor (BATF2), Melanoma Antigen Family B18 (MAGEB18) and Adenomatous Polyposis Coli (APC). Other directional selection candidate loci included SYNJ1 (Synaptojanin 1; synaptic transmission) and CYSLTR1 (Cysteinyl Leukotriene Receptor 1; pulmonary inflammation mediator).

Environmental correlations

Of 14 100 models Sam*β*ada computed (25 environmental parameters × 564 SNP genotypes) at the most conservative 0.01 Bonferroni-corrected confidence level, 781 significant associations (5.5% of the total) for both the likelihood ratio (*G*) or Wald tests were identified in 120 loci. Of these, 28 were significant for at least one F_{ST} outlier test. Applying the most conservative criteria (requiring both F_{ST} outlier tests to be significant at the 0.01 threshold), resulted in a subset of 10 candidate loci with significant environmental associations (Table 3). All nine directional selection candidate loci had statistically significant environmental correlations, while the six balancing selection candidate loci had only one. The majority of the environmental variables with significant associations were related to precipitation, temperature, position (latitude and longitude) and elevation (Table 3).

Discussion

To facilitate studies of local adaptation, we identified 5321 SNPs in candidate adaptive genes using an exon capture approach. We detected adaptive variation of Dall's sheep and uncovered putatively adaptive loci correlations with environmental variables at a broad spatial scale. Our results suggest that these genomics approaches can be used to identify adaptive variation across populations of Dall's sheep which could be useful to address wild sheep conservation by identifying adaptation to local conditions. Moreover, this work provides valuable insights into the general understanding of the genetic basis of adaptation in a free-ranging species.

Utility of a targeted approach

Our approach involved targeting a priori identified candidate genes. Research on adaptive differentiation of free-ranging species is greatly benefited using available information from closely related reference genomes. Coding regions of the genome are generally well conserved, thus knowledge of genes related to fitness in model organisms may be translated into studies of multiple species in natural environments. This 'bottom-up' approach (Sork *et al.* 2013; Harrisson *et al.* 2014) is valuable for gaining insight into an organism's response to specific selective pressures.

Exon capture facilitated identification of putatively adaptive candidate genes (with known fitness trait associations in other species) at the range-wide scale in Dall's sheep. Exons code for proteins that are involved in the

Table 3 SNP loci identified as potentially under selection for Dall's sheep using F_{ST} outlier and environmental correlation tests. For the F_{ST} outlier tests, loci with a 95% (*) and 99% (**) probability thresholds for balancing (B) and directional (D) selection are shown. SNP that are significant for at least one outlier test and also have a genotype significantly correlated with environmental covariates for both the likelihood ratio (*G*) and Wald tests at the Bonferroni-corrected 0.001 alpha levels are shown. For environmental variables abbreviations, see Table S1 (Supporting information)

	F_{ST} outlier detection		Environmental correlation		
Selected loci	LOSITAN	BAYESCAN	Sam*β*ada	Gene	Function
chr1 _42593194	B**			IL12RB2	Interleukin 12 receptor, beta 2; promotes the proliferation of T cells and NK cells
chr1_104552702	D*		lat[2], precip_wquart[2], anmeantemp[2], meantemp_cquart[2]	RXFP4	Relaxin/insulin-like family peptide receptor 4; associated with horn growth in sheep
chr1_121168066	D*	D**	lat[4], lon[6], mintemp_month[5], antemprange[5], meantemp_cquart[5]	SYNJ1	Synaptojanin 1; synaptic transmission and membrane trafficking
chr1_168582008	B*		precip_wquart[1,2], annualprecip[1,2]	ALCAM	Activated leucocyte cell adhesion molecule; interactions between nervous system cells, binding of T and B cells
chr1_176713422	B**	B*		KIAA2018	DNA and calcium ion binding
chr1_223923669	B*			OTOL1	Otolin 1
chr1_259772518	B**			TMPRSS2	Transmembrane protease serine 2
chr1_61277814	B*			LPAR3	Lysophosphatidic acid receptor 3; receptor for lysophosphatidic acid (LPA)
chr1_95724103	D**	D**	min_temp[1], lat[2], precip_wetq[2], anmeantemp[2], anualprecip[2], meantemp_cquart[2], mintemp_month[2]	WARS2	Tryptophanyl-tRNA synthetase 2; encodes mitochondrial tryptophanyl-tRNA synthetase
chr2_160493069	B*		lon[2]	ACVR2A	Activin A receptor, type IIA; susceptibility to pregnancy-related disease
chr2_203655388	B**			BMPR2	Bone morphogenetic protein receptor, type II; endochondral bone formation and embryogenesis
chr2_248078917	D*	D**	lat[1,2,3], precip_wquart[1,2], meantemp_cquart[1,2], annualprecip[1,2], mintemp_month[1,2], antemprange[1,2], anmeantemp[1], precip_season[2]	RGS2	Regulator Of G-protein signalling 2; GTPase activating proteins
chr2_48929874	B*	B**		GABBR2	Gamma-aminobutyric acid (GABA) B receptor, 2; synaptic inhibition, slow wave sleep, muscle relaxation, and antinociception
chr3_164545868	B**	B**		OR10A7	Olfactory receptor, family 10, Subfamily A, Member 7
chr3_204456312	B**	B**		OLR1	Oxidized low-density lipoprotein (lectin-like) receptor 1; recognition, internalization and degradation of oxLDL by vascular endothelial cells
chr4_85338163	B*	B**		TSPAN12	Tetraspanin 12; regulates cell-surface receptor signal transduction

Table 3 (Continued)

Selected loci	F_{ST} outlier detection		Environmental correlation	Gene	Function
	LOSITAN	BAYESCAN	Sam*β*ada		
chr4_9440922	D**	D**	lat[4,5], anmeantemp[4,5], meantemp_cquart[4,5], precip_wquart[4,5], lon[5], maxtemp_month[5], meantemp_wquart[5], mintemp_month[5], annualprecip[5], elv[5]	KRIT1	*Ovis aries* KRIT1, ankyrin repeat containing, transcript
chr5_39660956	B**			OR2G6	Olfactory receptor, family 2, subfamily G, member 6
chr5_54818699	D**	D**	lat[4,5], precip_wquart[4,5], meantemp_cquart[4,5], annualprecip[4,5], mintemp_month[4,5], precip_season[4,5], anmeantemp[4,5], antemprange[4], precip_season[5], lon[5], elv[5]	LARS	leucyl-tRNA synthetase; infantile liver failure
chr5_5961073	B*		lon[1,2], lat[2]	SIN2-family	GTPase that may function in mitochondrial ribosome assembly, promotes growth and cell division
chr6_70228621	B**			KIT	Type 3 transmembrane receptor protein for MGF mast cell growth factor
chr6_90551951	B*		lat[1,2], lon[1,2], precip_season[1], a nmeantemp[1]	IP-10	Chemokine (C-X-C Motif) ligand 10; antimicrobial gene encodes a chemokine of the CXC subfamily
chr7_23042786	B**		annualprecip[2], precip_wquart[2], precip_season[2]	RPGRIP1	Retinitis pigmentosa GTPase regulator interacting protein 1; encodes a photoreceptor gene
chr7_42256561	B*		lon[5]	PTGDR	Prostaglandin D2 receptor (DP); respond to extracellular cues and activate intracellular signal transduction pathways
chr7_84766674	B**	B**		ESRRB	Oestrogen-related receptor beta; placental development
chr7_892425	D**	D**	lat[1,2], lon[1,2], elv[1,2], maxtemp_month[1,2], precip_wquart[2]	APC	Adenomatous polyposis coli; tumour suppressor
chr8_49459850	D**	D**	lat[4,5], precip_wquart[4,5], annualprecip[4,5], lon[4], maxtemp_month[4], elv[4], antemprange[6], meantemp_cquart[5], anmeantemp[5], mintemp_month[5]	ORC3	Origin recognition complex, subunit 3; DNA replication in eukaryotic cell
chr9_12833308	B*	B*		PHF3	PHD finger protein 3; transcription factor
chr9_29404853	B**		lon[2]	ZHX1	Zinc fingers and homeoboxes 1; transcriptional repressor
chr9_59631558	D*	D**	mintemp_month[1,2], meantemp_cquart[1,2], anmeantemp[1,2,3], antemprange[1], lat[2,3], meantemp_wquart[2]	MED30	Mediator complex subunit 30; regulated transcription of nearly all RNA polymerase II-dependent genes
chr9_79888536	B*		elv[1]	CPQ	Carboxypeptidase Q; hydrolysis of circulating peptide

Table 3 (Continued)

Selected loci	F_{ST} outlier detection		Environmental correlation	Gene	Function
	LOSITAN	BAYESCAN	Sam*β*ada		
chr10_29494765	B*			RXFP2-Intronic Variant	Relaxin/insulin-like family peptide receptor 4; associated with horn growth in sheep
chr10_29498259	B*			RXFP2-Intronic Variant	Relaxin/insulin-like family peptide receptor 4; associated with horn growth in sheep
chr11_20513034	B*			SSH2	Slingshot protein phosphatase 2; regulates actin filament dynamics
chr12_43721662	D*			PER3	Period circadian clock 3; encode components of the circadian rhythms of locomotor activity, metabolism and behaviour
chr12_71647443	B*			IRF6	Interferon regulatory factor 6; regulates epithelial cell proliferation
chr14_42498364	B**			TDRD12	Tudor domain containing 12; spermatogenesis
chr15_28166746		D*	precip_wquart[1], annualprecip[1], elv[1], lat[1], onsetgreen[1], endgreen[1], elv[3], precip_wquar[t3], precip_wquart[2]	IL10RA	Interleukin-10 receptor subunit alpha; codes for proteins that inhibit the synthesis of proinflammatory cytokines
chr15_34621586	B**		lat[2], lon[2], precip_season[2]	ABCC8	ATP-Binding Cassette, subfamily C (CFTR/MRP), Member 8; transports molecules across extra- and intracellular membranes
chr15_48961906	B*			OR52M1	Olfactory receptor 52M1-like
chr16_9268098	D**		lat[1], lon[1], anmeantemp[1], meantemp_cquart[1], lat[2], lon[2], anmeantemp[2], meantemp_cquart[2], precip_season[2], mintemp_month[2]	PTCD2	Pentatricopeptide repeat domain 2; involved in mitochondrial RNA maturation and respiratory chain function
chr18_18652916	B**	B*		MRPL46	Mitochondrial ribosomal protein L46; protein synthesis within the mitochondrion
chr18_2872329	B**	B**		GABRB3	Gamma-aminobutyric acid (GABA) A receptor, beta 3; major inhibitory neurotransmitter, histamine receptor
chr18_3374453	B*		maxtemp_month[2]	GABRA5	Gamma-aminobutyric acid (GABA) A receptor, alpha 5; mediates neuronal inhibition
chr18_60363352	B*			VRK1	Vaccinia-related kinase 1; regulate cell proliferation
chr20_10135660	B**	B*		MAPK14	Mitogen-activated protein kinase 14; cellular function
chr20_25774224	D**	D**	lat[1,2], precip_wquart[1,2], meantemp_cquart[1,2], annualprecip[1,2], mintemp_month[1,2], antemprange[1,2], anmeantemp[1,2], precip_season[1,2]	Ovar-DRA	Major histocompatibility complex, Class II, DR alpha; immune function, presents peptides from extracellular proteins
chr20_25774236	D*	D**	lat[1,2], precip_wquart[1,2], meantemp_cquart[1,2], annualprecip[1,2], mintemp_month[1,2], antemprange[1,2], anmeantemp[1,2], precip_season[1,2]	Ovar-DRA	Major histocompatibility complex, Class II, DR alpha; immune function, presents peptides from extracellular proteins

Table 3 (Continued)

Selected loci	F_{ST} outlier detection		Environmental correlation	Gene	Function
	LOSITAN	BAYESCAN	Sam*β*ada		
chr20_26942052	B**			IFITM	Interferon-induced transmembrane protein; inhibits the entry of viruses to the host cell cytoplasm
chr20_26945193	B**	B**		BOLA class I histocompatibility	Class I histocompatibility antigen; immune function
chr21_42526714	D**	D**	mintemp_month[1,2], meantemp_cquart[1,2], antemprange[1,2], anmeantemp[1,2], precip_wquart[1,2], mintemp_month[2], annualprecip[2], precip_season2, lat[2,3]	BATF2	Basic leucine zipper transcription factor, ATF-Like 2; differentiation of lineage-specific cells in the immune system
chr23_14061659	B**	B**	lon[2]	PI3K Family	Phosphatidylinositol 3-kinase regulator activity; positive regulation of development of blood vessels, T cell differentiation
chr24_18992466	B*	B*		ZP2	Zona pellucida glycoprotein 2 (sperm receptor); fertilization and preimplantation development
chr25_25352273	B**	B*		SUPV3L1	suppressor of var1, 3-like; mitochondrial RNA metabolism
chrX_122553725	D**	D**	lat[1,2], anmeantemp[1,2], meantemp_cquart[1,2], meantemp_wquart[1,2], mintemp_month[1,2], maxtemp_month[1,2]	CYSLTR1	Cysteinyl leukotriene receptor 1; contraction and proliferation of smooth muscle, oedema, eosinophil migration and damage to the mucus layer in the lung
chrX_23330270	D**	D**	lat[1,2], lon[1,2], anmeantemp[1,2], precip_season[1,2]	MAGEB18	Melanoma Antigen Family B18, may regulate various malignant phenotypes of cancer cells
chrX_53002036	D**		lon[5,4], lat[4], elv[4]	GATA1	GATA binding protein 1 (globin transcription factor 1); erythroid development; regulates switch of foetal to adult haemoglobin
Total (0.01)	31	21	28		
Total (0.05)	56	28	28		

[1]AA genotype, [2]TT genotype, [3]TA genotype, [4]CC genotype, [5]GG genotype, [6]CG genotype.

production of phenotypes, and phenotype traits may influence the organisms' fitness and therefore be affected by natural selection. Consequently, we expected this approach might have a higher success rate at identifying adaptive variants than a broad-scale genomewide scan of genomic variation. In support of this exon-targeted approach, previous research using quantitative trait loci to identify relationships between fitness traits and candidate genes in domestic cattle, goats and sheep has revealed most genetic polymorphisms associated with production (fatness, growth, milk production) are found in exons (Ibeagha-Awemu *et al.* 2008).

Indeed, our gene-targeted approach produced somewhat higher proportions of candidate adaptive to putatively neutral genes than other next-generation genomics techniques. For example, Lv *et al.* (2014) detected 230 SNPs of 49 034 from the OvineSNP50BeadChip with evidence of selection (0.04%) across 32 domestic sheep breeds. Of these, 17 were strong candidate genes for climate-mediated selection (0.03%). Using 744 alleles from microsatellite markers, Joost *et al.* (2007) found five alleles from four loci had significant correlations with environmental variables (13% of loci, 0.7% of alleles) across 57 breeds of domestic sheep distributed across Europe and the Middle East. The five alleles most likely to be under selection were all associated with precipitation and the number of wet days in a given month (Joost *et al.* 2007), and two of the five had been previously associated with disease resistance. Using genome scan approaches, Limborg *et al.* (2012) identified 6% and Narum *et al.*

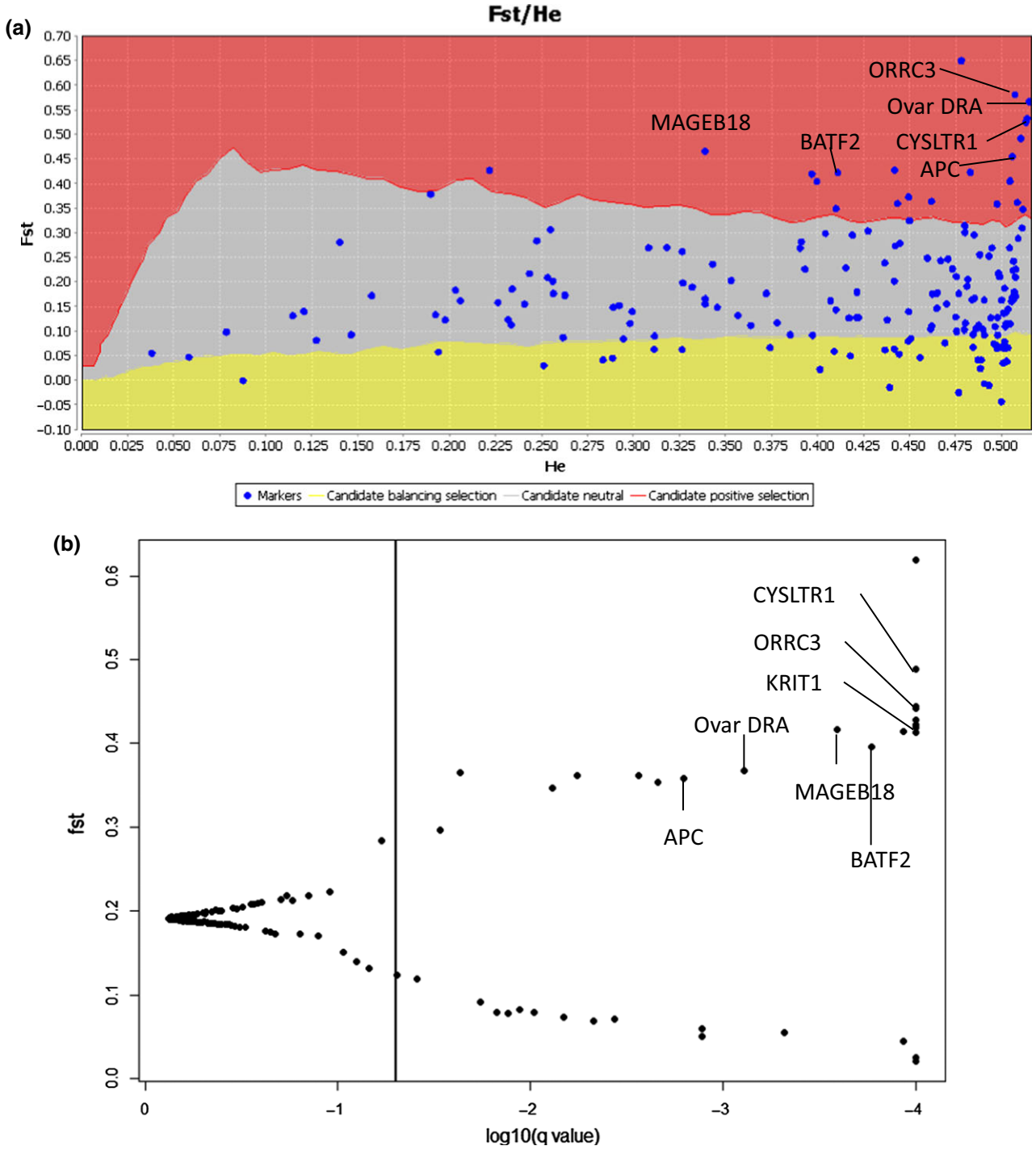

Fig. 3 (a) Outlier loci detected at the 5% level with LOSITAN software. The confidence area for candidate loci under positive selection is shown in red and for candidate loci potentially under balancing selection in yellow. (b) Outlier loci identified by BAYESCAN. *X* axis shows the posterior odds, and the *Y* axis shows the F_{ST} index values.

(2010) identified 8% of SNPs genotyped as putatively under selection.

Here, we identified 57 loci of 181 (31%) as putatively under selection; of these 28 were detected as outliers with both F_{ST}-based approaches (15%). The relatively higher proportion of outliers detected in this study is potentially due to using a focused approach, specifically targeting candidate genes for which we had prior

information (e.g. associated with specific fitness-related traits) and hypothesized would have a greater likelihood of being under selection. In contrast, the other studies referenced above used genome scan approaches to identify outlier loci or alleles significantly correlated with environmental variables. A benefit to using genome scans is that it is not necessary to have a priori knowledge of loci associated with specific traits that are under selection and therefore may detect adaptive alleles without first identifying them (Storz 2005). However, in comparison with our targeted approach, which we were able to carry out due to an existing body of knowledge of domestic animal traits and their genetic associations, it may be necessary to assess a larger number of markers to detect putatively adaptive loci. Although we began with a relatively small number of SNP loci in contrast to other studies, we identified a comparable number of putatively adaptive loci. We expect that expanding our SNP panel would also provide a higher number of candidate outlier loci, loci with significant environmental correlations and a similar proportion of candidate adaptive vs. neutral loci.

Exon capture in nonmodel species

Bighorn sheep sequence reads from a previous exon capture study proved to be an effective source of exon sequences to be used as baits in the Dall sheep exon capture. Existing next-generation sequencing (NGS) data from a nonmodel species, along with their alignment to genome sequences from related model or domesticated and better-studied species like *Ovis aries* and *Bos taurus*, can be used to obtain gene sequence information on yet another nonmodel species. Thus, we could leverage the closer relationship between the bighorn and the Dall's sheep, vs. *Ovis aries* and the Dall's sheep, in the design of an exon capture on which to base the construction of our SNP array. Bighorn and the Dall's sheep are thought to have diverged 1.41 million years ago and are therefore expected to be more genetically similar than Dall's sheep and domestic sheep, who shared a most recent common ancestor 3.12 million years ago (Bunch *et al.* 2006). This likely improved the overall success rate for our assay genotyping by increasing average read depth per target and allowing us to use higher NGS data quality thresholds for both the targeted resequencing genotyping, as well as the consensus sequence construction (Cosart *et al.* 2011). Our work joins a growing body of research that has capitalized on accumulated knowledge of domesticated species to select genes for use in studies of natural selection in wild populations (Kardos *et al.* 2015; Miller *et al.* 2015).

The Fluidigm platform is less commonly used in genomic research than other more high-throughput or massively parallel sequencing technologies that have become available (e.g. genotyping-by-sequencing, etc.). However, there are advantages to using this system including rapid turnaround time, straightforward data analysis, and a vetted methodology for dealing with low quality and low concentration samples. We found it to be an appropriate method for our research primarily because many of our samples were degraded and fragmented (e.g. old or poorly preserved blood samples), and because using the Fluidigm EP1, only a small quantity of DNA is required to obtain accurate and reliable genotypes (e.g. Campbell & Narum 2009; Beja-Pereira *et al.* 2009), particularly if samples are preamped prior to genotyping (Kraus *et al.* 2015). Specific target amplification primer pairs may be multiplexed at low DNA concentration (50 nM per pair) and individual samples may be amplified with few cycles (~15). An additional future goal of our research is to apply the SNP chip to genotyping of Dall's sheep samples collected noninvasively, specifically faecal pellets. Faecal pellets may be collected broadly over large landscapes given access to Dall's sheep bedding sites (Roffler *et al.* 2014). As is common with noninvasive samples, DNA concentration is low, but this obstacle may be overcome using the Fludigm SNP chip developed in this research.

An additional benefit of the Fluidigm platform for our study was that it performed relatively efficiently and economically. A single laboratory technician could feasibly genotype 96 samples at 96 SNPs in 1 day (9216 genotypes total), considering time estimates of approximately 4 h of laboratory preparation, ~2 h of PCR and an hour of data analysis. To the best of our knowledge, exon capture has only been successfully performed once on low-quality noninvasive samples (Perry *et al.* 2010), and rarely on museum samples (e.g. Bi *et al.* 2014; McCormack *et al.* 2015; Lim & Braun 2016), and to apply this technology to our samples would require a large investment in both time and finances. As these technologies are becoming increasingly available and applied to answer questions in molecular ecology, it is possible that in the near future exon capture could be effective with noninvasive samples from a broader array of taxa (Beja-Pereira *et al.* 2009).

Signatures of natural selection

Our results revealed signals of natural selection in Dall's sheep populations throughout their range. Focusing on the results with the strongest support evidenced by multiple significant outlier tests, we found candidate genes for immune function, signalling pathways for regulating cellular processes, tumour suppression, respiratory health, reproduction and olfactory receptors. We also demonstrate the relevance of environmental

heterogeneity in locally adaptive genetic variation in candidate genes. For all loci putatively under directional selection (identified by both outlier methods), temperature variables were significantly associated with SNP variants, with annual mean temperature (anmeantemp) most highly correlated, and also identified by principal component analyses as one of the environmental factors that explained most of the variability among regions (Appendix S2, Supporting information). The most highly correlated precipitation variable was precipitation of the wettest quarter (precip_wquart), significantly associated with 10 of the loci under putative directional selection. Latitude was also identified as a potentially important driver for local adaptation, as it was also significantly associated with all candidate loci for directional selection, whereas longitude and elevation were associated with 4.

Two of the candidate adaptive loci were in the Ovar-DRA gene located in the major histocompatibility complex (MHC) region. The MHC is made up of multiple genes and is generally highly polymorphic (at class I) across taxa. Candidate loci within MHC have a demonstrated role in parasite resistance in sheep (Buitkamp *et al.* 1996; Paterson *et al*. 1998). The Ovar-DRA is a class II MHC gene and plays a central role in the initiation of protective immune function in vertebrates. Specifically, it encodes cell-surface proteins to bind and present peptides from extracellular proteins to T cells and invoke an immune response. This gene is well conserved and believed to have limited polymorphisms, although interestingly, comparatively high levels of diversity were described in domestic sheep breeds, potentially due to an independent evolutionary history (Ballingall *et al.* 2010). In Dall's sheep, we found strong evidence of directional selection in Ovar-DRA, and strong balancing selection in the BOLA Class I histocompatibility antigen. Balancing selection at MHC loci acting over long timescales and across multiple populations can occur simultaneously with directional selection at local populations. This pattern could be a result of different pathogens acting on different alleles at the MHC locus, or spatial inconsistency of pathogen exposure across a metapopulation.

The candidate gene CYSLTR1 is a receptor for cysteinyl leukotrienes which play a role in respiratory health by mediating inflammation, and eosinophil migration to the lung. In humans, variants of this gene are associated with occurrence of asthma and susceptibility to infection (Sokolowska *et al.* 2009). We found strong evidence for directional selection in this gene and environmental correlations with all tested temperature variables and latitude. These results fit our hypothesis that genes related to immune function would vary across a latitudinal gradient, as pathogen exposure and abundance decline with latitude and decreasing temperatures. Wild sheep are believed to be immunologically naïve and therefore more susceptible to infectious diseases, especially in the case of respiratory diseases which have had catastrophic effects on bighorn sheep populations in the U.S. Rockies (Foreyt *et al.* 1996; Garde *et al.* 2005; Jenkins *et al.* 2007; George *et al.* 2008). Greater understanding of adaptive differentiation in disease resistance genes can aid in predictions of how Dall's sheep may respond to the spread of respiratory disease, which is expected to present an increasing threat to northern ungulates as a result of changing climates (Kutz *et al.* 2005; Jenkins *et al.* 2006). Two other candidate genes related to immune functions with strong evidence for directional selection were MAGEB18 and APC. APC has a signalling pathway function which regulates stem cells and cancer tumour suppression, and MAGEB18 is expressed in tumours and may play a role in regulating malignant cancer cells. Both demonstrated homozygous variants associated with latitude, longitude, temperature and precipitation variables, and APC was also significantly correlated with elevation.

The Relaxin/Insulin-Like Family Peptide Receptor 4 (RXFP2) was a LOSITAN outlier for directional selection but not BAYESCAN, providing weaker support of adaptive variation in this gene. However, the locus was significantly correlated with latitude, two precipitation variables and annual mean temperature. Two intronic variants of the RXFP2 gene were also identified as outliers with LOSITAN for balancing selection. This gene is associated with horn growth in sheep and has been demonstrated to be under strong selection in sheep species (Kijas *et al.* 2012; Johnston *et al.* 2013; Kardos *et al.* 2015). In domestic sheep, selective breeding to remove horns resulted in a strong signal of selection (Kijas *et al.* 2012). In free-ranging bighorn sheep, whole genome sequencing revealed a selective sweep at this gene, likely related to selection favouring rams with larger horns (Kardos *et al.* 2015). Our results provide additional evidence for selection on RXFP2 in a free-ranging and widely distributed sheep population.

Conclusions

Using a novel exon capture approach for gene-targeted SNP discovery followed by qPCR-based SNP chip genotyping, we were able to extend population genomic approaches into natural landscapes and discover genes potentially under selection at the range-wide scale. We targeted genes with known functions, many of them in domestic sheep or cattle, and others from well-studied taxa. However, many traits are affected by multiple genes and interactions with the environment, and thus, we cannot account for unmeasured genetic or

environmental effects. We recognize that while these results appear to be promising for future research, the possibility exists that the outlier loci detected may be simply linked to genomic regions experiencing natural selection and not under selection themselves. Additionally, correlation of alleles with environmental variables does not imply causation. Despite these limitations, our approach was valuable because it was founded on ecological knowledge of the organism, applied in a real life setting and can provide a step towards understanding the genetic basis of natural selection. Further insights and increased certainty of selection could be gained by spatial and temporal replication of sample collections across environmental gradients in independent populations.

Detecting adaptive loci and natural selection in free-ranging populations is widely recognized as challenging, yet a critical endeavour for gaining insights into adaptive responses of species to changing environments (Shafer *et al*. 2015). We provide an example of using targeted resequencing and SNP genotyping in a nonmodel species to detect molecular patterns of natural selection and correlation of candidate adaptive genes with environmental variables. Continued development and application of these methods will advance the accumulation of evidence for the genetic basis of adaptation.

Acknowledgements

This work was funded by the National Park Service (NPS) and the U.S. Geological Survey (USGS), and U.S. National Science Foundation grant 1407300. The NPS (J. Lawler, B. Mangipane, K. Rattenbury), the Alaska Department of Fish and Game (N. Cassara, T. Kavalok, T. Lohuis, T. McDonough, T. Peltier, R. Schwanke) and the Yukon Department of Environment (T. Hegel) provided samples and valuable insights. Laboratory assistance was provided by Y. Horeaux and J. Shabacker. L. Adams, S. Dobrowski and M. Hebblewhite provided useful comments on early drafts of the manuscript. This work was supported by the U.S. National Science Foundation grant 1407300. G.R. was supported by the USGS and 3 University of Montana scholarships (Bertha Morton, Les Pengelly, and George and Mildred Cirica). G.L. was supported by NSF grants DEB-1258203 and DEB-1067613. M.K. was supported by the National Science Foundation via the Montana Institute on Ecosystems (EPS-1101342) and the Montana Ecology of Infectious Diseases IGERT Program (DGE-0504628). Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

References

Allendorf FW, Hohenlohe PA, Luikart G (2010) Genomics and the future of conservation genetics. *Nature Reviews Genetics*, **11**, 697–709.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403–410.

Antão T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G (2008) LOSITAN: a workbench to detect molecular adaptation based on a Fst-outlier method. *BMC Bioinformatics*, **9**, 323.

Archibald AL, Cockett NE, Dalrymple BP *et al*. (2010) The sheep genome reference sequence: a work in progress. *Animal Genetics*, **41**, 449–453.

Ballingall KT, Rocchi MS, McKeever DJ, Wright F (2010) Trans-species polymorphism and selection in the MHC class II DRA genes of domestic sheep. *PLoS One*, **5**, e11402.

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society B: Biological Sciences*, **263**, 1619–1626.

Beja-Pereira A, Oliveira R, Alves PC, Schwartz MK, Luikart G (2009) Advancing ecological understandings through technological transformations in noninvasive genetics. *Molecular Ecology Resources*, **9**, 1279–1301.

Bi K, Linderoth T, Vanderpool D *et al*. (2014) Unlocking the vault: next generation museum population genomics. *Molecular Ecology*, **22**, 6018–6032.

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

Buitkamp J, Stear MJ, Epplen JT (1996) Class I and class II major histocompatibility complex alleles are associated with faecal egg counts following natural, predominantly *Ostertagia circumcincta* infections. *Parasitology Research*, **82**, 693–696.

Bunch TD, Wu C, Zhang Y-P, Wang S (2006) Phylogenetic analysis of snow sheep (*Ovis nivicola*) and closely related taxa. *The Journal of Heredity*, **97**, 21–30.

Bunnell FL (1982) The lambing period of mountain sheep: synthesis, hypotheses, and tests. *Canadian Journal of Zoology*, **60**, 1–14.

Campbell NR, Narum SR (2009) Quantitative PCR assessment of microsatellite and SNP genotyping with variable quality DNA extracts. *Conservation Genetics*, **10**, 779–784.

Cosart T, Beja-Pereira A, Chen S *et al*. (2011) Exome-wide DNA capture and next generation sequencing in domestic and wild species. *BMC Genomics*, **12**, 347.

Dalrymple BP, Kirkness EF, Nefedov M *et al*. (2007) Using comparative genomics to reorder the human genome sequence into a virtual sheep genome. *Genome Biology*, **8**, R152.

Eckert AJ, Bower AD, González-Martínez SC *et al*. (2010) Back to nature: ecological genomics of loblolly pine (*Pinus taeda, Pinaceae*). *Molecular Ecology*, **19**, 3789–3805.

Ekblom R, Wolf JBW (2014) A field guide to whole-genome sequencing, assembly and annotation. *Evolutionary Applications*, **7**, 1026–1042.

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

Elsik CG, Tellam RL, Worley KC (2009) The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science*, **324**, 522–528.

Festa-Bianchet M (1988) Seasonal range selection in bighorn sheep: conflicts between forage quality, forage quantity, and predator avoidance. *Oecologia*, **75**, 580–586.

Festa-Bianchet M (2008) *Ovis dalli*. In: *IUCN 2011. IUCN Red List of Threatened Species. Version 2011.2*. Available from www.iucnredlist.org. Accessed 25 March 2010.

Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, **180**, 977–993.

Foreyt J, Silflow M, Lagerquist E (1996) Susceptibility of Dall sheep (*Ovis dalli dalli*) to pneumonia caused by *Pasteurella haemolytica*. *Journal of Wildlife Diseases*, **32**, 586–593.

Garde E, Kutz S, Schwantje H, Veitch A, Jenkins E, Elkin B (2005) Examining the risk of disease transmission between wild Dall's Sheep and mountain goats, and introduced domestic sheep, goats, and llamas in the Northwest Territories. In: *Wildlife Disease and Zoonotics*. University of Nebraska, Lincoln.

George JL, Martin DJ, Lukacs PM, Miller MW (2008) Epidemic pasteurellosis in a bighorn sheep population coinciding with the appearance of a domestic sheep. *Journal of Wildlife Diseases*, **44**, 388–403.

Harrisson KA, Pavlova A, Telonis-Scott M, Sunnucks P (2014) Using genomics to characterize evolutionary potential for conservation of wild populations. *Evolutionary Applications*, **7**, 1008–1025.

Heaton MP, Leymaster KA, Kalbfleisch TS *et al*. (2014) SNPs for parentage testing and traceability in globally diverse breeds of sheep. *PLoS One*, **9**, 1–10.

Hodges E, Xuan Z, Balija V *et al*. (2007) Genome-wide in situ exon capture for selective resequencing. *Nature Genetics*, **39**, 1522–1527.

Hohenlohe PA, Amish SJ, Catchen JM, Allendorf FW, Luikart G (2011) Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. *Molecular Ecology Resources*, **11**(Suppl 1), 117–122.

Holderegger R, Wagner HH (2008) Landscape genetics. *BioScience*, **58**, 199–207.

Ibeagha-Awemu EM, Kgwatalala P, Zhao X (2008) A critical analysis of production-associated DNA polymorphisms in the genes of cattle, goat, sheep, and pig. *Mammalian Genome*, **19**, 591–617.

Jenkins EJ, Veitch AM, Kutz SJ, Hoberg EP, Polley L (2006) Climate change and the epidemiology of protostrongylid nematodes in northern ecosystems: *Parelaphostrongylus odocoilei* and *Protostrongylus stilesi* in Dall's sheep (*Ovis d. dalli*). *Parasitology*, **132**, 387–401.

Jenkins EJ, Veitch AM, Kutz SJ *et al*. (2007) Protostrongylid parasites and pneumonia in captive and wild thinhorn sheep (*Ovis dalli*). *Journal of Wildlife Diseases*, **43**, 189–205.

Johnston SE, McEwan JC, Pickering NK *et al*. (2011) Genome-wide association mapping identifies the genetic basis of discrete and quantitative variation in sexual weaponry in a wild sheep population. *Molecular Ecology*, **20**, 2555–2566.

Johnston SE, Gratten J, Berenos C *et al*. (2013) Life history trade-offs at a single locus maintain sexually selected genetic variation. *Nature*, **502**, 93–95.

Joost S, Bonin A, Bruford MW *et al*. (2007) A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. *Molecular Ecology*, **16**, 3955–3969.

Kardos M, Luikart G, Bunch R *et al*. (2015) Whole-genome resequencing uncovers molecular signatures of natural and sexual selection in wild bighorn sheep. *Molecular Ecology*, **24**, 5616–5632.

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, **30**, 772–780.

Kijas JW, Townley D, Dalrymple BP *et al*. (2009) A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. *PLoS One*, **4**, e4668.

Kijas JW, Lenstra JA, Hayes B *et al*. (2012) Genome-wide analysis of the World's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS Biology*, **10**, e1001258.

Kraus RHS, VonHoldt B, Cocchiararo B *et al*. (2015) A SNP-based approach for rapid and cost-effective genetic wolf monitoring in Europe based on non-invasively collected samples. *Molecular ecology resources*, **15**, 295–305.

Kurten B, Anderson E (1980) *Pleistocene Mammals of North America*. Columbia University Press, New York.

Kutz SJ, Hoberg EP, Polley L, Jenkins EJ (2005) Global warming is changing the dynamics of Arctic host-parasite systems. *Proceedings of the Royal Society of London B: Biological Sciences*, **272**, 2571–2576.

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

Li H, Handsaker B, Wysoker A *et al*. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

Lim HC, Braun MJ (2016) High-throughput SNP genotyping of historical and modern samples of five bird species via sequence capture of ultraconserved elements. *Molecular Ecology Resources*.

Limborg MT, Helyar SJ, De Bruyn M *et al*. (2012) Environmental selection on transcriptome-derived SNPs in a high gene flow marine fish, the Atlantic herring (*Clupea harengus*). *Molecular Ecology*, **21**, 3686–3703.

Lohse M, Bolger AM, Nagel A *et al*. (2012) RobiNA: A user-friendly, integrated software solution for RNA-Seq-based transcriptomics. *Nucleic Acids Research*, **40**, 622–627.

Lotterhos KE, Whitlock MC (2014) Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. *Molecular Ecology*, **23**, 2178–2192.

Lühken G (2012) Genetic testing for phenotype-causing variants in sheep and goats. *Molecular and Cellular Probes*, **26**, 231–237.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews. Genetics*, **4**, 981–994.

Luikart G, Amish SJ, Winnie J *et al*. (2011) High connectivity among argali sheep from Afghanistan and adjacent countries: inferences from neutral and candidate gene microsatellites. *Conservation Genetics*, **12**, 921–931.

Lv F-H, Agha S, Kantanen J *et al*. (2014) Adaptations to Climate-Mediated Selective Pressures in Sheep. *Molecular Biology and Evolution*, **31**, 3324–3343.

Mainguy J, Worley K, Côté SD, Coltman DW (2007) Low MHC DRB class II diversity in the mountain goat: past bottlenecks and possible role of pathogens and parasites. *Conservation Genetics*, **8**, 885–891.

Manel S, Joost S, Epperson BK *et al*. (2010) Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field. *Molecular ecology*, **19**, 3760–3772.

McCormack JE, Tsai WLE, Faircloth BC (2015) Sequence capture of ultraconserved elements from bird museum specimens. *Molecular Ecology Resources*, doi: 10.1111/1755-0998.12466.

McKenna A, Hanna M, Banks E *et al*. (2009) The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, **20**, 254–260.

Miller JM, Poissant J, Kijas JW, Coltman DW (2011) A genome-wide set of SNPs detects population substructure and long range linkage disequilibrium in wild sheep. *Molecular Ecology Resources*, **11**, 314–322.

Miller JM, Moore SS, Stothard P, Liao X, Coltman DW (2015) Harnessing cross-species alignment to discover SNPs and generate a draft genome sequence of a bighorn sheep (*Ovis canadensis*). *BMC Genomics*, **16**, 397.

Narum SR, Hess JE (2011) Comparison of FST outlier tests for SNP loci under selection. *Molecular Ecology Resources*, **11**, 184–194.

Narum SR, Campbell NR, Kozfkay CC, Meyer KA (2010) Adaptation of redband trout in desert and montane environments. *Molecular Ecology*, **19**, 4622–4637.

Nielsen R (2005) Molecular signatures of natural selection. *Annual Review of Genetics*, **39**, 197–218.

Nielsen EE, Hemmer-Hansen J, Larsen PF, Bekkevold D (2009) Population genomics of marine fishes: identifying adaptive variation in space and time. *Molecular Ecology*, **18**, 3128–3150.

Pariset L, Joost S, Marsan PA, Valentini A (2009) Landscape genomics and biased FST approaches reveal single nucleotide polymorphisms under selection in goat breeds of North-East Mediterranean. *BMC Genetics*, **10**, 7.

Paterson S, Wilson K, Pemberton JM (1998) Major histocompatibility complex variation associated with juvenile survival and parasite resistance in a large unmanaged ungulate population (*Ovis aries* L.). *Proceedings of the National Academy of Sciences*, **95**, 3714–3719.

Peréz-Figueroa A, Garcéa-Pereira MJ, Saura M, Rolén-Alvarez E, Caballero A (2010) Comparing three different methods to detect selective loci using dominant markers. *Journal of Evolutionary Biology*, **23**, 2267–2276.

Perry GH, Marioni JC, Melsted P, Gilad Y (2010) Genomic-scale capture and sequencing of endogenous DNA from feces. *Molecular Ecology*, **19**, 5332–5344.

Poissant J, Davis C, Malenfant R, Hogg JT, Coltman D (2012) QTL mapping for sexually dimorphic fitness-related traits in wild bighorn sheep. *Heredity*, **108**, 256–263.

Quinlan AR, Hall IM (2010) BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.

Rachlow JL, Bowyer RT (1991) Interannual Variation in Timing and Synchrony of Parturition in Dall's Sheep. *Journal of Mammalogy*, **72**, 487–492.

Raymond M, Rousset F (1995) GENEPOP: Population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.

Roffler GH, Talbot SL, Luikart G *et al*. (2014) Lack of sex-biased dispersal promotes fine-scale genetic structure in alpine ungulates. *Conservation Genetic*, **15**, 837–851.

Schoville SD, Bonin A, François O *et al*. (2012) Adaptive genetic variation on the landscape: methods and cases. *Annual Review of Ecology, Evolution, and Systematics*, **43**, 23–43.

Shafer ABA, Wolf JBW, Alves PC *et al*. (2015) Genomics and the challenging translation into conservation practice. *Trends in Ecology & Evolution*, **30**, 78–87.

Slate J, Gratten J, Beraldi D *et al*. (2009) Gene mapping in the wild with SNPs: Guidelines and future directions. *Genetica*, **136**, 97–107.

Sokolowska M, Wodz-Naskiewicz K, Cieslak M, Seta K, Bednarek AK (2009) Pawliczak R (2009) Variable expression of cysteinyl leukotriene type I receptor splice variants in asthmatic females with different promoter haplotypes. *BMC Immunology*, **10**, 63.

Sork VL, Aitken SN, Dyer RJ, Eckert AJ, Legendre P, Neale DB (2013) Putting the landscape into the genomics of trees: approaches for understanding local adaptation and population responses to changing climate. *Tree Genetics & Genomes*, **9**, 901–911.

Stapley J, Reger J, Feulner PGD *et al*. (2010) Adaptation genomics: the next generation. *Trends in Ecology & Evolution*, **25**, 705–712.

Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology*, **14**, 671–688.

Waples RS (2014) Testing for Hardy-Weinberg Proportions: Have We Lost the Plot? *Journal of Heredity*, **106**, 1–19.

Worley K, Strobeck C, Arthur S *et al*. (2004) Population genetic structure of North American thinhorn sheep (Ovis dalli). *Molecular Ecology*, **13**, 2545–2556.

Worley K, Carey J, Veitch A, Coltman DW (2006) Detecting the signature of selection on immune genes in highly structured populations of wild sheep (*Ovis dalli*). *Molecular Ecology*, **15**, 623–637.

G.H.R., G.L. and M.K.S. designed the study. S.A. coordinated the exon capture, SNP genotyping and provided analytical input. T.C. conducted bioinformatics analyses. S.S. conducted laboratory analyses. M.K. identified additional sheep genome targets for the capture. G.H.R. performed analyses, coordinated sample collection and led the writing of the manuscript with input from all co-authors.

Data accessibility

SNP data are available in Dryad DOI: doi: 10.5061/dryad.kk466.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1 DNA extraction, consensus sequence generation, and genotyping assay design.

Appendix S2 Environmental and landscape association tests with SNP loci.

Table S1 SNP-type assays and primer sequences tested including alleles targeted by allele specific primer 1 (ASP1) and 2 (ASP2), sequences for allele specific primer 1 (ASP1 sequence) and 2 (ASP2 sequence), the sequences for locus specific primers (LSP sequence), and the sequences for specific target amplification primers (STA sequences)

Table S2 Environmental and landscape variables tested for associations with SNP loci

Fig. S1 Principal component analysis (PCA) plot of environmental variables demonstrating the percentage of variation explained by each axis and correlations among variables.