

Adaptations to Climate-Mediated Selective Pressures in Sheep

Feng-Hua Lv,¹ Saif Agha,^{2,3} Juha Kantanen,^{4,5} Licia Colli,^{6,7} Sylvie Stucki,² James W. Kijas,⁸ Stéphane Joost,² Meng-Hua Li,^{*1} and Paolo Ajmone Marsan^{6,7}

¹CAS Key Laboratory of Animal Ecology and Conservation Biology, Institute of Zoology, Chinese Academy of Sciences (CAS), Beijing, China

²Laboratory of Geographic Information Systems (LASIG), School of Architecture, Civil and Environmental Engineering (ENAC), Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

³Department of Animal Science, Faculty of Agriculture, Ain Shams University, Cairo, Egypt,

⁴Biotechnology and Food Research, MTT Agrifood Research Finland, Jokioinen, Finland

⁵Department of Biology, University of Eastern Finland, Kuopio, Finland

⁶Istituto di Zootecnica, Facoltà di Agraria, Università Cattolica del Sacro Cuore, Piacenza, Italy

⁷Biodiversity and Ancient DNA Research Center—BioDNA, Università Cattolica del Sacro Cuore, Piacenza, Italy

⁸CSIRO Livestock Industries, St Lucia, Brisbane, Qld, Australia

*Corresponding author: E-mail: menghua.li@ioz.ac.cn.

Associate editor: Yuseob Kim

Abstract

Following domestication, sheep (*Ovis aries*) have become essential farmed animals across the world through adaptation to a diverse range of environments and varied production systems. Climate-mediated selective pressure has shaped phenotypic variation and has left genetic “footprints” in the genome of breeds raised in different agroecological zones. Unlike numerous studies that have searched for evidence of selection using only population genetics data, here, we conducted an integrated coanalysis of environmental data with single nucleotide polymorphism (SNP) variation. By examining 49,034 SNPs from 32 old, autochthonous sheep breeds that are adapted to a spectrum of different regional climates, we identified 230 SNPs with evidence for selection that is likely due to climate-mediated pressure. Among them, 189 (82%) showed significant correlation ($P \leq 0.05$) between allele frequency and climatic variables in a larger set of native populations from a worldwide range of geographic areas and climates. Gene ontology analysis of genes collocated with significant SNPs identified 17 candidates related to GTPase regulator and peptide receptor activities in the biological processes of energy metabolism and endocrine and autoimmune regulation. We also observed high linkage disequilibrium and significant extended haplotype homozygosity for the core haplotype *TBC1D12-CH1* of *TBC1D12*. The global frequency distribution of the core haplotype and allele OAR22_18929579-A showed an apparent geographic pattern and significant ($P \leq 0.05$) correlations with climatic variation. Our results imply that adaptations to local climates have shaped the spatial distribution of some variants that are candidates to underpin adaptive variation in sheep.

Key words: adaptation, climate-mediated selection, genome-wide scans, GTPase regulator, peptide receptor, *TBC1D12*, sheep.

Introduction

Environmental heterogeneity and differences in climatic factors (e.g., temperature and precipitation) influence the spatial distribution of phenotypic and genetic variation across populations of a variety of organisms including loblolly pine, *Arabidopsis*, *Drosophila*, goat, and human (Hancock et al. 2008; Pariset et al. 2009; Eckert et al. 2010; González et al. 2010; Hancock, Brachi, et al. 2011; Hancock, Witonsky, et al. 2011). The detection of climate-mediated selective signatures is thus one of the central research themes in evolutionary biology, with the potential to shed light on the genetic basis of local adaptation and speciation in response to changing climates (MacCallum and Hill 2006; Joost et al. 2007). The identification of adaptive variation

also holds the promise of providing insight into functionally important variants (see the reviews in Bamshad and Wooding 2003; Nielsen et al. 2007).

To date, the identification of environment-driven selection has been largely restricted to species with finished genome sequences or model species, such as *Drosophila* and humans (see the review in Oleksyk et al. 2010). These successful studies have revealed examples of climatic adaptation in traits including pigmentation (Hancock, Witonsky, et al. 2011), body size (Gardner et al. 2011), and thermal response (Karell et al. 2011). Few published studies have been conducted in livestock, despite the fact that many farmyard animal species have a global range and exhibit phenotypic diversity and adaptation to disparate environments. One recent exception is

© The Author 2014. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Open Access

a study of domesticated yaks (*Bos grunniens*; Qiu et al. 2012); however, yaks have a restricted range compared with other livestock species. Thus, the recent availability of genome-wide single nucleotide polymorphism (SNP) sets in livestock would allow the identification of environment-associated selection.

Livestock have a population history characterized by domestication and subsequent human-mediated selection for favorable production traits. The comparison of genomic patterns of SNP variability, often between divergent breeds, has successfully identified many genomic regions and genes that have undergone selection sweeps (Gu et al. 2009; Qanbari et al. 2010; Stella et al. 2010; Amaral et al. 2011). Most studies have employed analyses of allele frequency differences, with measures such as F_{ST} -based outliers or long-range haplotype (LRH) tests. Importantly, these analyses have proceeded without the integration of genomic and environmental data (Kijas et al. 2012; Ai et al. 2013; Ramey et al. 2013). As a result, it is still impossible to link the signatures of selection to specific spatially varying selective pressures (e.g., a specific environmental variable). In recent years, several approaches have been developed in landscape genomics to detect adaptation to different climate pressures by examining correlations or the association between SNP alleles and climate variables (e.g., BayEnv in Coop et al. 2010; latent factor mixed model (LFMM) in Frichot et al. 2013; see also the review in Joost et al. 2013). These approaches have strengths and weaknesses due to different implicit assumptions in the models. By applying these approaches, several studies have succeeded in searching for evidence of genetic adaptation to different climatic pressures by scanning the genome for environmental correlations in a variety of organisms (Coop et al. 2010; Hancock et al. 2010; Meier et al. 2011; Shimada et al. 2011).

Ruminants such as sheep can be directly affected by climate effects on thermoregulation, pasture quality, and biomass (Nielsen et al. 2013). Although there is extensive evidence for phenotypic variation due to genetic adaptation and/or nongenetic acclimatization to different climates in sheep (Nielsen et al. 2012, 2013), the extent to which this variation is the result of genetic adaptation at the whole genome-wide level is still unclear. In the face of globally changing climates that may favor more woody vegetation (browse) at the expense of grasses (Gordon and Prins 2008), these questions are highly relevant for sheep genetics and breeding (e.g., marker-assisted selection and breeding) to identify breeds of sheep or produce better derived breeds that are more robustly suited to future climates, that is, have increased feed efficiency on novel vegetation communities (Gordon and Prins 2008; Franks and Hoffmann 2012; Nielsen et al. 2012, 2013).

To the best of our knowledge, this is the first high density SNP genome scan for climate-induced selection in livestock that combines molecular and environmental data. The aim of this study is to characterize the genetic legacy that centuries of climate-induced adaptations have imparted to the sheep genome by identifying genome-wide signatures of selection. From a data set of genome-wide (~50 K) SNPs in 74 sheep populations/breeds that were sampled and genotyped within the sheep HapMap project (<http://www.sheephapmap.org/>

[hapmap.php](#), last accessed June 3, 2014), we selected genotypes of 32 old, autochthonous sheep breeds (see fig. 1). We performed a variety of selection tests using approaches based on different assumptions (e.g., genetic differentiation of SNPs, haplotype structure, and genetic–environmental correlations) and different data sets (genomic data alone and the combination of genomic and environmental data). We identified a set of candidate SNPs, genes, and core haplotypes under climate-driven adaptation that were enriched in two clusters of gene ontology (GO) terms related to the biological processes of energy metabolism and endocrine and autoimmune regulation. These results will advance our understanding of the genetic architecture of climate-driven adaptive evolution and are of significance for their potential applications in functional genomics and selective breeding (Joost et al. 2007), as well as in the creation of conservation management programs (Luikart et al. 2003) to cope with rapid global climate change in sheep and other livestock.

Results

Relationships between Breeds Based on Climate Variables and Genomic Data

Principal component analysis (PCA) on the basis of climatic variables (fig. 2A and B) and the analysis of genetic relationships between breeds was performed to identify a set of distantly related breeds adapted to extreme environments. In this subset, signatures for climatic adaptation are expected to be stronger and easier to detect while spurious signals due to common origins between breeds will be reduced.

PCA clustered 32 native sheep breeds according to the environment they are adapted to inhabit (see fig. 3). The first two principal components (PC1 and PC2) explain more than 69% of the total variance (PC1 accounts for 50.78% and PC2 for 18.31%). PC1 divides breeds as a result of the contributions of multiple environmental climate variables. This component represents a synthetic parameter that principally summarizes the information of three climatic variables (supplementary table S1, Supplementary Material online): The number of days with >0.1 mm of rain per month (RDO; 18.49%), the percent maximum possible sunshine (SUN; 17.89%), and the mean diurnal temperature range in °C (DTR; 15.91%). These are critical factors for vegetation growth and terrestrial primary production (Nemani et al. 2003). PC2 and PC3 do not reveal a clear geographic divergence associated with environmental variables (see fig. 3A). The plot revealed that eight breeds have positive extreme PC1 values: Four breeds from the United Kingdom (Border Leicester, Boreray, Scottish Blackface, and Soay Sheep), two from Switzerland (Swiss Mirror Sheep, Valais Blacknose Sheep), one from Norway (Spael-colored Sheep), and one from Finland (Finnsheep). These extreme PC1 values likely arise due to high values of precipitation and days of rainfall (PR and RDO). Six breeds have PC1 values at the negative extreme mainly because of the high values of temperature, sunshine, and distribution of precipitation (TMP, DTR, SUN, and PRCV): One each from Iran (Afshar), Turkey (Karakas), Cyprus (Cyprus Fat-Tail), India (Indian Garole),

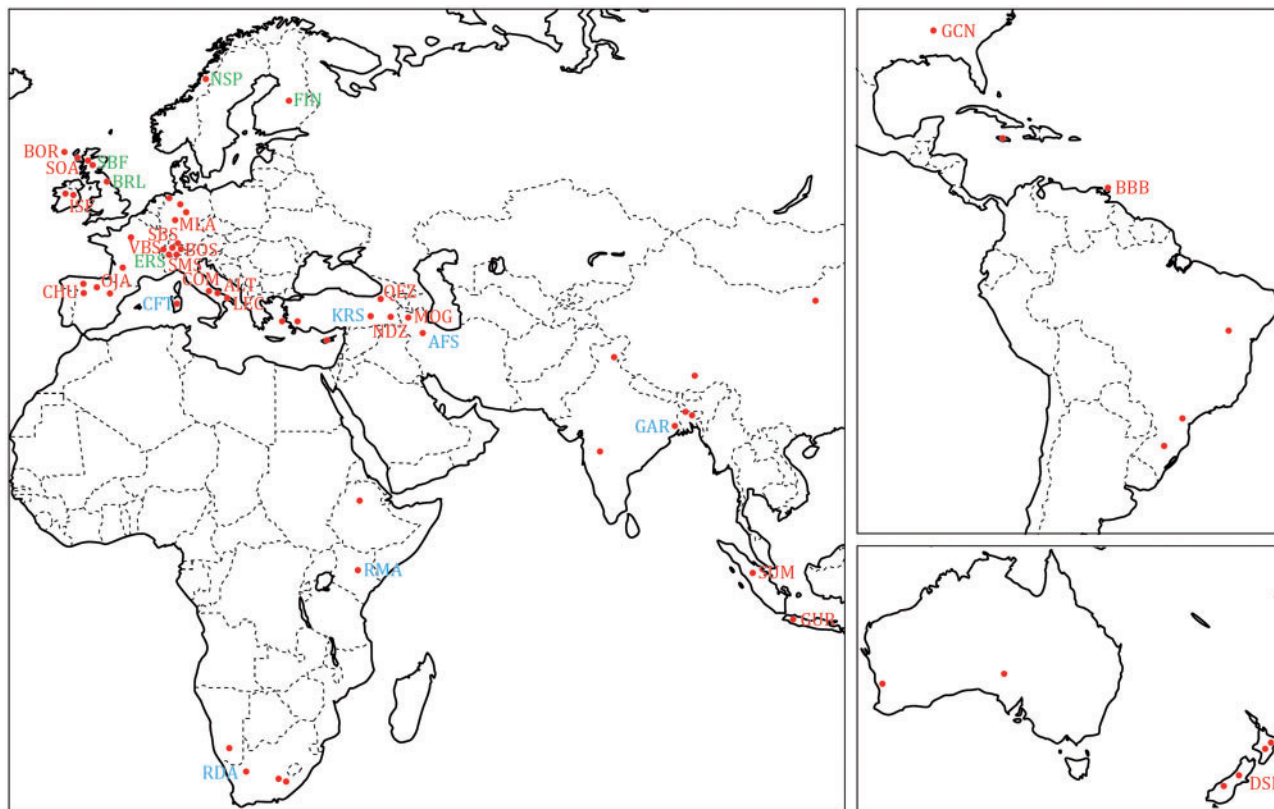


Fig. 1. Geographic origins of the world's sheep breeds. Sampling locations of the 74 breeds/populations (for details of the breeds/populations, see supplementary table S1, Supplementary Material online, in Kijas et al. [2012] or at <http://www.sheephapmap.org/hapmap.php>, last accessed June 3, 2014) are indicated by red dots. Only the 32 old, autochthonous breeds selected in the analyses are represented by the codes. Codes in green (NSP, FIN, SBF, BRL, ERS) and in blue (CFT, KRS, AFS, GAR, RMA, RDA) represent the 2 groups of 11 breeds selected for the selection tests. Breed names as represented by the codes are detailed in supplementary table S8, Supplementary Material online.

South Africa (Ronderib Afrikaner), and Kenya (Red Maasai) (fig. 3B).

Genetic Relationship between Breeds

A neighbor-net graph of D_R was constructed to explore the genomic relationship between breeds (fig. 4). The breeds grouped into two main clusters. One main cluster (cluster I) included breeds from South Asia, the Middle East, Africa and South America, whereas the other cluster (cluster II) was composed of breeds from Europe, New Zealand and the United States. This grouping is consistent with previous findings concerning the phylogeography of the examined breeds (Kijas et al. 2009, 2012). All the breeds in cluster I showed negative PC1 values in the PCA plot based on environmental variables. Conversely, not all the breeds in cluster II showed positive PC1 values (see figs. 3 and 4). Further, all the breeds with positive PC1 values were classified into cluster II, but all the breeds with negative PC1 values did not belong to cluster I (see figs. 3 and 4). Among the 15 breeds located at the extremes of PC1 in the PCA plot, 11 were selected for further analyses. Four breeds were excluded due to their shared ancestry with other breeds (e.g., Moghani is closely related to Afshari; Swiss Mirror Sheep is closely related to Engadine Red Sheep) or high levels of inbreeding (Soay, $F_{IS} = 0.33$; Boreray, $F_{IS} = 0.28$; see supplementary table S3, Supplementary Material online, in Kijas et al. 2012). Therefore, by combining

the results of the interpopulation genetic relationship analysis and the PCAs, 11 sheep breeds (Afshari [AFS, Iran], Ronder Afrikaner [RDA, South Africa], Indian Garole [GAR, India], Karakas [KRS, Turkey], Red Maasai [RMA, Kenya], Cyprus Fat-Tail [CFT, Cyprus], Border Leicester [BRL, United Kingdom], Spael-white [NSP, Norway], Finnsheep [FIN, Finland], Engadine Red Sheep [ERS, Switzerland], and Scottish Blackface Sheep [SBF, United Kingdom]) were chosen for genome-wide selection tests.

Detection of Selective Sweeps

In the first analysis for selection, we searched for values of F_{ST} that were either higher or lower than expected after controlling for the expected genetic heterozygosity (H_E). This approach was applied to the 11 breeds in the two clusters presented in figure 5A. The summary statistical method based on simulated and observed pairwise F_{ST} values identified a total of 2,353 SNPs beyond the 95th percentile of the empirical distribution (see Materials and Methods) as outliers across the genome (fig. 5A and supplementary table S2, Supplementary Material online). As this figure is approximately the expected number of false positives, only regions carrying five consecutive significant (window-averaged P values ≤ 0.05) SNPs, an event that is highly unlikely to occur by chance ($P < 10^{-8}$), were considered for further analyses. Using this strategy, we identified 29 sweep regions that

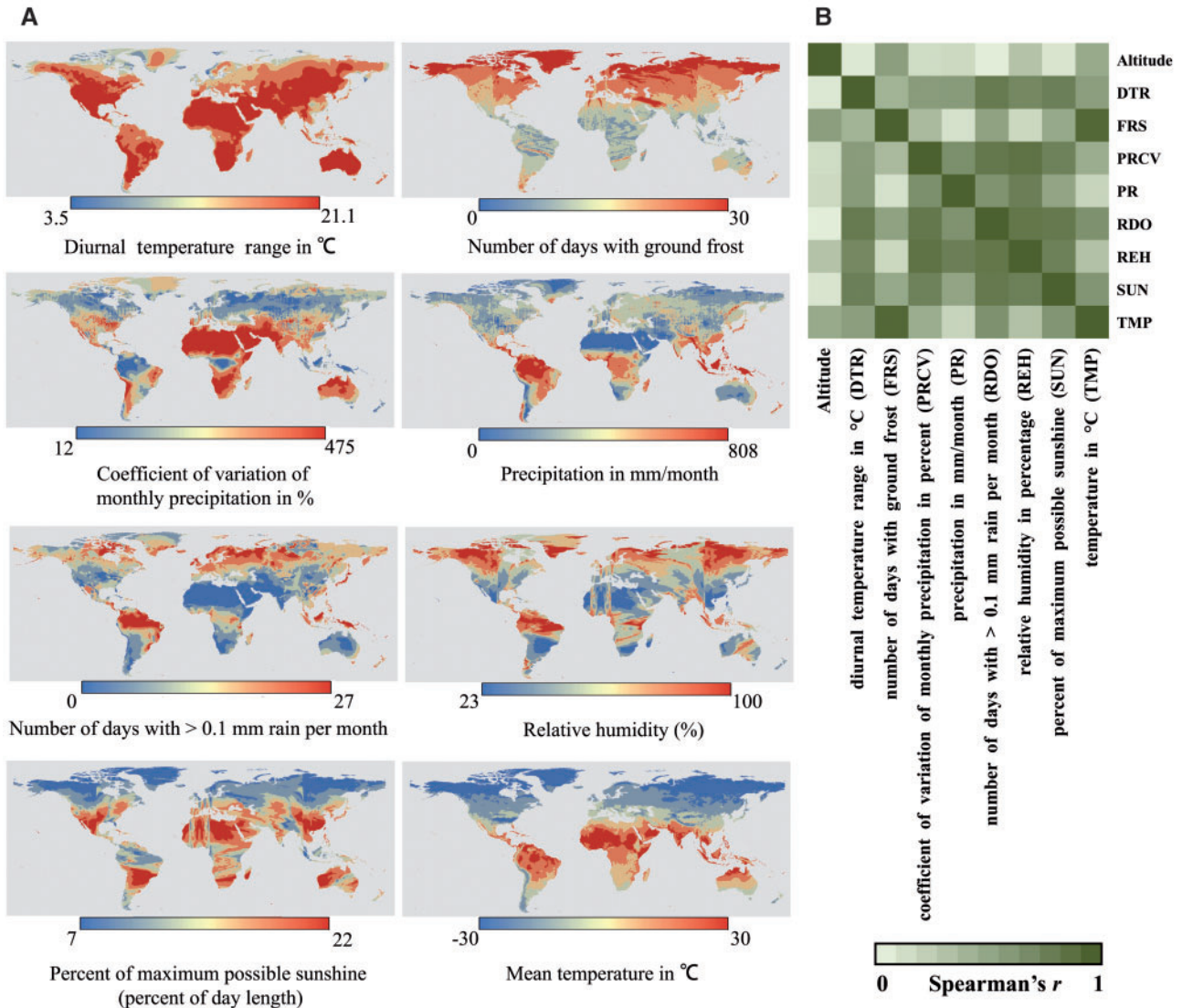


Fig. 2. Climatic variable used in the analysis. (A) Maps show the geographic patterns of the yearly mean values of the eight climate variables: Diurnal temperature range (DTR), number of days with ground frost (FRS), coefficient of variation of monthly precipitation (PRCV), precipitation in mm/month (PR), number of days with >0.1 mm of rain per month (RDO), relative humidity (REH), percent maximum possible sunshine (SUN), and mean temperature (TMP); (B) a heatmap shows the absolute values of Spearman's rank correlation coefficients for the altitude and annual mean values of the eight climatic variables.

showed significant genetic differentiation between the two population groups. We further identified two regions at the more stringent cutoff level of window-averaged P values ≤ 0.01 (OAR1:119.38–119.80 Mb, OAR4:48.50–48.80 Mb; table 1). The 29 regions were distributed across chromosomes OAR1, OAR2, and OAR3. Most regions (27/29; average values of $r^2 < 0.5$; table 1) showed generally low levels of linkage disequilibrium (LD) between SNPs within a given region. A fragment on OAR2 was the largest, extending over 0.35 Mb (table 1). A total of 91 genes were located in or close to these regions by aligning the “target region” to Ovine (Texel) version 3.1 Genome Assembly (see Materials and Methods).

Of the 29 candidate sweep regions (table 1), three overlapped with those previously identified in the analysis that excluded the consideration of environmental parameters (~10%; OAR2: 51.72–51.95 Mb, OAR6:36.61–36.87 Mb, and OAR10:30.49–30.70 Mb; see table 1 in Kijas et al. 2012).

These regions spanned six genes (*MELK*, *GNE*, *SPP1*, *IBSP*, *MEPE*, and *HMGB1*), which were not biologically and functionally relevant candidate genes for selection in either this study or the earlier study. The majority of strong signals was specific to the subset analyzed here and did not overlap with those found in earlier worldwide analyses, suggesting that true selection pressures are more localized rather than distributed across populations. The difference could also be due to the relative effects of natural and artificial selections on the sheep genome, which were targeted by this and the earlier study, respectively.

To search for selection observed across multiple breeds, pairwise F_{ST} outlier tests were implemented between breeds belonging to each of the two groups. SNP outliers were distributed across the entire genome, but none was detected to be under divergent selection ($P \leq 0.05$) across all the 30 (5×6) pairwise tests (data not shown). The number of

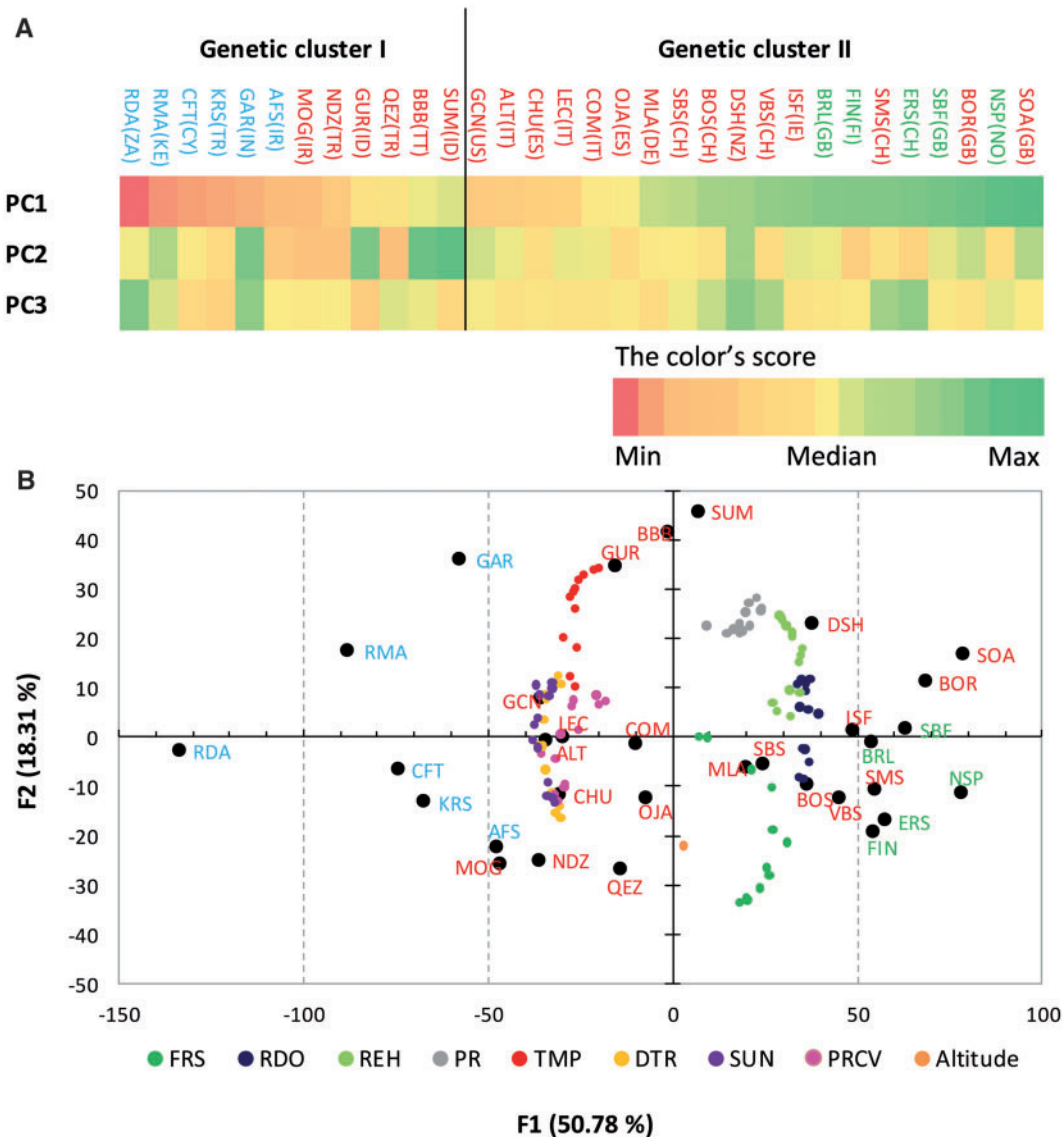


Fig. 3. PCA of environmental variables and 32 old, autochthonous sheep breeds. (A) Heat strips for each of the first three PCs are shown for the 32 sheep breeds assigned to the two genetic clusters (I and II; see fig. 4); the two letters before the breed codes indicate the country of origin: CH, Switzerland; CY, Cyprus; DE, Germany; ES, Spain; FI, Finland; IE, Ireland; ID, Indonesia; IN, India; IR, Iran; IT, Italy; KE, Kenya; NO, Norway; NZ, New Zealand; TR, Turkey; TT, Trinidad and Tobago; UK, United Kingdom; US, United States; and ZA, South Africa; (B) the score plots of PC1 versus PC2 for the 32 old, autochthonous sheep breeds and the environmental variables of their geographic origins. The breeds in contrasting environments selected for the selection tests are indicated in blue and green, respectively. The nine environmental variables are mean diurnal temperature range in °C, DTR; number of days with ground frost, FRS; precipitation in mm/month, PR; the coefficient of variation of monthly precipitation in percent, PRCV; relative humidity in percentage, REH; percent of maximum possible sunshine, SUN; mean temperature in °C, TMP; and number of days with > 0.1 mm rain per month, RDO; and altitude. The small colored circles represent the altitude or the yearly mean and monthly parameters of one of the eight climate variables across all the 32 breeds.

pairwise populations that displayed significant divergence with P values ≤ 0.05 was plotted across the genome (supplementary fig. S1, Supplementary Material online). These data revealed peaks and troughs where selection was shared across breeds and absent or unique to only a small number of breeds, respectively. In total, 101 SNPs (supplementary fig. S1, Supplementary Material online) were detected with divergent selection ($P \leq 0.05$) shared by half (15/30) of the comparisons. The observation that signals were not detected across all the comparisons is expected. Adaptation is due to the interaction of a number of complex traits, and many

genes are likely involved in the control of each trait. This indicates that a high level of genetic heterogeneity is expected and that breeds may have adapted to a similar environment using different “genomic strategies.”

Signatures of Genomic Adaptation to Local Environments

We processed 15,445,710 univariate models (147,102 genotypes \times 105 environmental parameters). Due to the limitation of the software in which the P values provided by

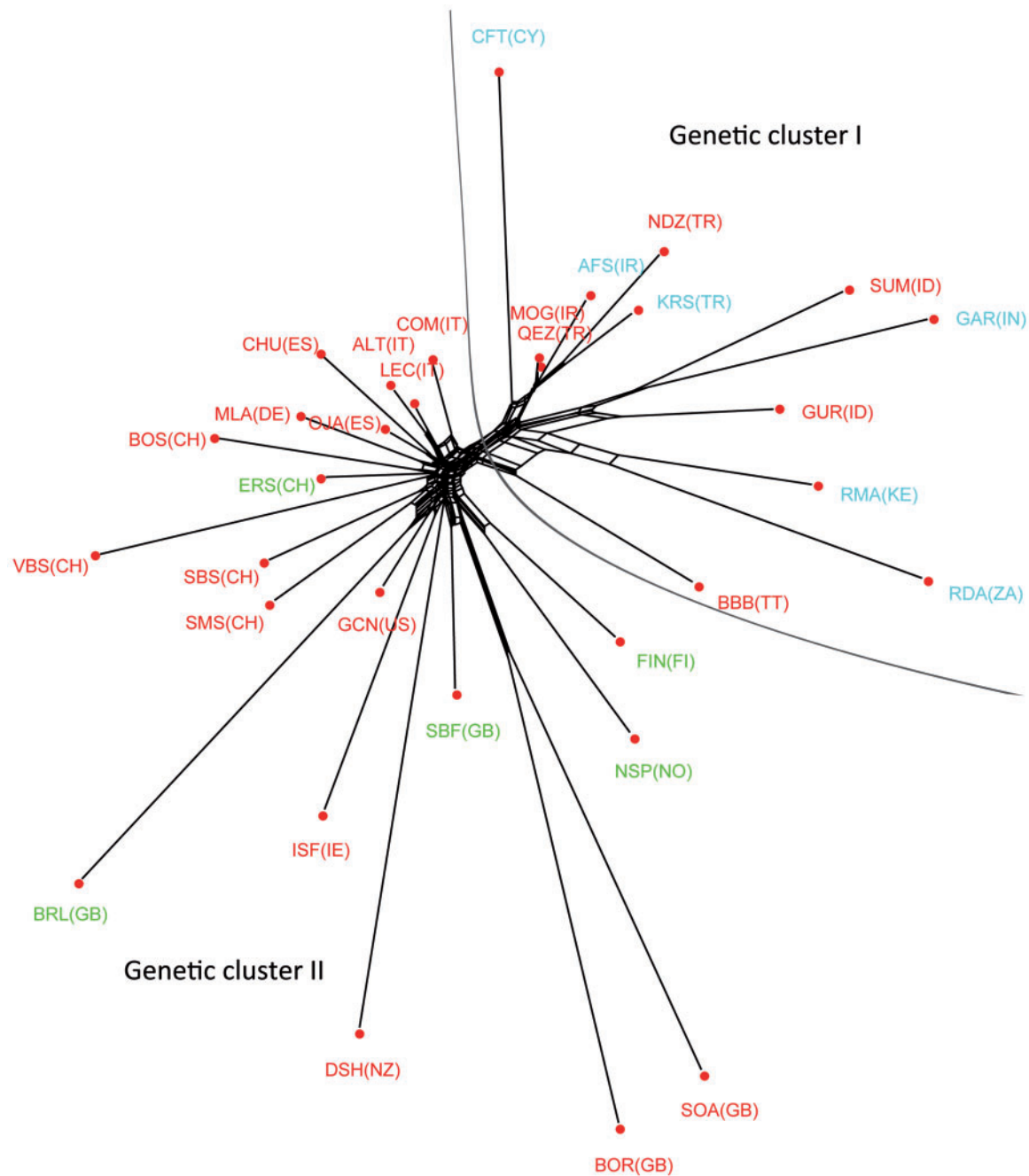


Fig. 4. Genetic relationship between the 32 old, autochthonous sheep breeds based on Reynolds' genetic distance. A metric of Reynolds' genetic distance was used to construct a neighbor-net graph relating the breeds. The 11 sheep breeds selected for selection tests from the two genetic clusters (I and II) are indicated in green and blue, respectively. The codes in the parentheses indicate their geographic origins of development: CH, Switzerland; CY, Cyprus; DE, Germany; ES, Spain; FI, Finland; IE, Ireland; ID, Indonesia; IN, India; IR, Iran; IT, Italy; KE, Kenya; NO, Norway; NZ, New Zealand; TR, Turkey; TT, Trinidad and Tobago; UK, United Kingdom; US, United States; and ZA, South Africa; the breed names, as represented by the codes, are detailed in [supplementary table S8, Supplementary Material online](#).

MatSAM are limited to $1E-20$, we did not set a standard confidence level before correcting for multiple comparisons. Instead, we sorted out the models according to Wald statistics and selected the first approximately 1,000 showing the highest values ($\sim 0.06\%$ of the total models processed). The Wald statistic of the selected models ranged between 14.1 and 21.7 ([supplementary table S3, Supplementary Material online](#)). Of the nine environmental factors, PR (34.19%, 330/965), SUN (26.73%, 258/965), and RDO (24.97%, 241/965) are the

predominant variables involved in the best models ([supplementary table S3, Supplementary Material online](#)). In particular, of a total of 32 selective SNPs located near or within the 17 strong candidate genes (see next section), 16 (50%, 16/32) SNPs are associated with SUN ([supplementary table S3, Supplementary Material online](#)). The LFMM approach identified a total of 3,756 SNPs showing $|z|$ scores greater than 5 (two-sided test; [fig. 5B](#) and [supplementary table S4, Supplementary Material online](#)). The cutoff $|z|$ score > 5

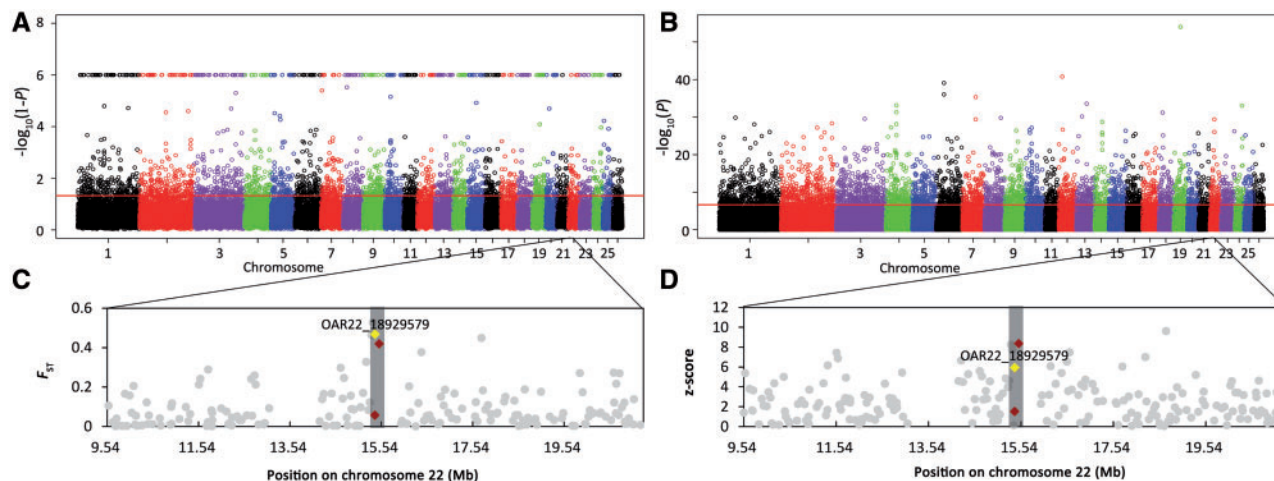


Fig. 5. F_{ST} -based outlier and LFMM selection tests. (A) Genome-wide distribution of $[-\log_{10}(1 - P)]$ values for genetic differentiation (F_{ST}) between the two groups of 11 breeds (P value is the probability of simulated $F_{ST} < \text{sample } F_{ST}$; Antao et al. 2008; [supplementary table S2, Supplementary Material online](#); see also Materials and Methods) by the LOSITAN program (Antao et al. 2008); the red line indicates the outlier SNPs beyond the upper 95th percentile, which were assumed to be under selection; the ceiling of maximum values at 6 is due to a limitation of the calculation of P values, and only six digits after the decimal point can be calculated; (B) genome-wide distribution of significance values $[-\log_{10}(P)]$ for the correlation between the frequencies of SNPs and the environmental variables in the LFMM test; the red line indicates the significance level of $P < 10^{-7}$; (C) regional distribution of F_{ST} for the SNPs within and around the strong candidate gene *TCB1D12*; and (D) regional distribution of $|z|$ scores for the SNPs within and around the strong candidate gene *TCB1D12*; SNP OAR22_18929579 is in yellow, and the two other SNPs (OAR22_18924267 and OAR22_19020858) in the core haplotype are in red.

indicated significant SNP effects at the level of $P \leq 10^{-7}$ after applying a standard Bonferroni correction for $\alpha = 0.01$ and $L = 10^5$ (α , type I error; L , number of loci). Among these, 382 SNPs were also detected by MatSAM ([fig. 6A](#) and [supplementary table S4, Supplementary Material online](#)).

GO Enrichments and Core Haplotypes

We identified 230 overlapping SNPs ([fig. 6A](#)) among the selection tests using approaches based on different models and assumptions (F_{ST} outlier, MatSAM, and LFMM). Compared with the overlap expected by chance, there was a significant excess of overlapping selective signals that were shared between pairs of approaches or among all three approaches (observed SNPs $n = 230$, SNPs expected by chance $n = 4$; $P < 0.001$; [fig. 6A](#)). We also detected a significant (observed SNPs $n = 79$, SNPs expected by chance $n = 13.3$; $P < 0.001$; [fig. 6B](#)) enrichment of overlapping signals between the SNPs with $|z| > 5$ and the SNPs in the 29 candidate sweep regions ([table 1](#)). The 230 overlapping SNPs were located across the entire genome with a large number of SNPs located on OAR1, OAR2, OAR3, and OAR4 ([supplementary table S5, Supplementary Material online](#)). We found a total of 175 candidate genes ([supplementary table S5, Supplementary Material online](#)). Two genes (*MAPK14* and *ANTXR2*; Hancock et al. 2008; Fumagalli et al. 2011) were previously detected to be associated with regional variations in climate and pathogens in humans ([supplementary table S6, Supplementary Material online](#)). In addition, only six genes were associated with growth and production (*POL* and *LRR2*; Zhang et al. 2013), reproduction (*FSHR* and *PRL*; Rannikki et al. 1995; Chu et al. 2007, 2009, 2012), coat color

(*EDNRB*; Metallinos et al. 1998; Wilkinson et al. 2013), and fat deposits (*ACSS2*; Bichi et al. 2013), as identified by earlier genetic studies in sheep ([supplementary tables S5 and S7, Supplementary Material online](#)) and other livestock.

We compared the distribution of gene sizes in the candidate gene set against the background set, and the Kolmogorov–Smirnov test of significance indicated that the distribution of gene size for the candidate gene set is indeed significantly larger ($P < 0.01$; see [supplementary fig. S2, Supplementary Material online](#)). After further filtering out 28 SNPs that showed strong LD ($r^2 > 0.5$) with one or more nearby loci in the same genomic regions (or genes), we obtained 202 SNPs and 175 relevant candidate genes (see [supplementary table S3, Supplementary Material online](#)). GO enrichment among the 175 environmentally associated candidate genes was evaluated for evidence of functional enrichment in specific categories of biological processes, molecular functions, and KEGG pathways. Of a total of 39 GO categories, we revealed 13 GO terms enriched with 17 genes using a threshold of $P < 0.05$ ([tables 2 and 3](#)). These GO terms grouped into two clusters; one was mainly related to GTPase activity (cluster A, enrichment score = 2.18), and the other was mainly related to peptide and chemokine receptor activity (cluster B, enrichment score = 1.95), which are mainly involved in the biological processes of energy metabolism and endocrine and autoimmune regulation (e.g., insulin, gonadotropin-releasing hormone, and formyl peptide receptors). Among the 17 genes implicated in climate-driven selection, nine were directly related to energy sources and transfers. These genes encode signaling molecules involved in GTPase activator and regulator activities, regulation of Ras protein signal transduction, such as *ARHGEF18*,

Table 1. Regions under Climate-Associated Selection in the Sheep Genome.

Region	Chr.	Position (Mb) (version 3.1)	Position (Mb) (version 1.0)	r^2	Peak SNP (F_{ST})	Top SNPs	No. of Genes	Candidate Genes ^a
1*	1	69.00–69.19	73.58–73.78	0.68	OAR1_73583793 (0.54)	6	2	<i>EVIS</i>
2*	1	105.35–105.70	113.14–113.48	0.20	s47856 (0.43)	9	8	
3**	1	119.38–119.80	129.37–129.96	0.41	s70345 (0.58)	5	1	
4* ^b	2	51.72–51.95	55.31–55.54	0.15	OAR2_55416697 (0.51)	5	2	
5*	2	164.90–165.03	174.54–174.66	0.32	OAR2_174558603 (0.61)	5	1	
6*	2	184.00–184.30	195.09–195.39	0.39	OAR2_195251529 (0.42)	6	1	
7*	2	232.47–232.59	245.55–245.68	0.42	s13779 (0.44)	5	4	<i>NMUR1</i>
8*	2	234.33–234.61	247.47–247.76	0.28	s09024 (0.38)	7	11	
9*	3	11.66–11.89	12.24–12.47	0.32	OAR3_12358231 (0.66)	5	2	
10*	3	44.11–44.30	47.15–47.35	0.35	OAR3_47149563 (0.46)	5	2	
11*	3	106.05–106.36	112.82–113.18	0.28	OAR3_112822823 (0.48)	6	4	
12*	3	144.25–144.55	154.11–154.43	0.33	OAR3_154209677 (0.32)	5	3	
13*	3	186.59–186.74	200.81–200.98	0.28	OAR3_200805613 (0.54)	5	1	
14**	4	48.50–48.80	51.32–51.63	0.37	OAR4_51489408 (0.73)	7	6	
15* ^b	6	36.63–36.87	40.83–41.04	0.10	OAR6_40955920 (0.33)	5	3	
16*	6	115.21–115.56	116.66–117.05	0.25	s65350 (0.44)	7	4	
17*	7	6.55–6.87	6.45–6.74	0.15	OAR7_6587255 (0.60)	6	3	
18*	7	55.98–56.18	61.90–62.17	0.40	OAR7_61967937 (0.65)	5	2	
19*	10	28.71–29.00	28.73–29.04	0.24	OAR10_28772065 (0.47)	6	5	
20* ^b	10	30.49–30.70	30.54–30.75	0.20	OAR10_30746533 (0.70)	6	1	
21*	13	27.24–27.43	30.18–30.37	0.61	s41783 (0.45)	5	2	
22*	13	53.41–53.63	58.10–58.35	0.48	s61722 (0.59)	6	10	
23*	14	35.21–35.50	36.64–36.94	0.21	s18388 (0.56)	6	4	
24*	16	3.36–3.52	3.43–3.63	0.24	s25980 (0.45)	5		
25*	17	4.78–5.17	5.30–5.78	0.14	OAR17_5388531 (0.39)	6	2	
26*	21	17.85–18.08	20.16–20.40	0.29	OAR21_20371526 (0.39)	8	1	
27*	21	45.17–45.37	50.17–50.38	0.25	s48673 (0.43)	5	2	
28*	23	26.31–26.64	27.44–27.77	0.22	s11742 (0.37)	5	2	
29*	24	33.80–33.96	36.90–37.06	0.24	s43015 (0.64)	5	2	

NOTE.— r^2 is the average LD value for all the pairwise SNPs within the regions.

^aOverlap of the 17 strong candidate genes enriched in the GO terms.

^bOverlapping regions with those previously identified in Kijas et al. (2012).

Average P value of five consecutive SNPs *less than 0.05 and **less than 0.01.

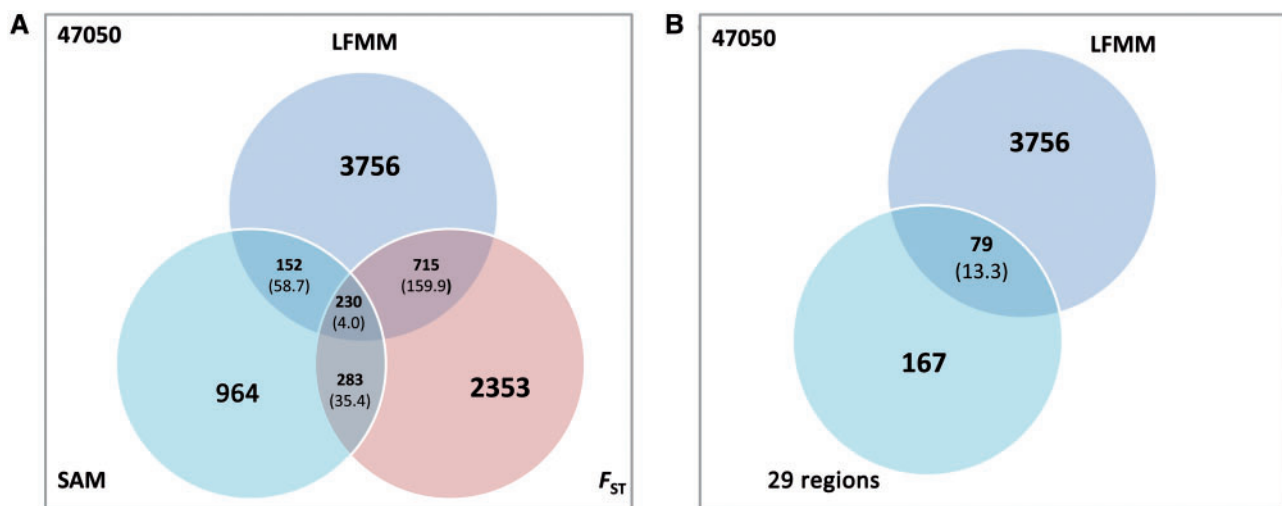


Fig. 6. The overlapping SNPs under different selection tests. (A) Numbers in the intersection regions are the observed overlapping SNPs between two methods or among all three methods; (B) numbers in the intersection region are overlapping SNPs between the 29 candidate sweep regions (the window-averaged P values ≤ 0.05 in the F_{ST} -outlier test) and LFMM analysis ($|z|$ score > 5 in LFMM). Numbers in parentheses show the overlapping SNPs expected by chance if signals were independent across populations. The total number of SNPs is reported in the upper left corner.

Table 2. Overrepresented GO Terms among the 175 Candidate Genes Identified to Be under Environmentally Associated Selection in Sheep.

Cluster (enrichment score)	Term	Description	Genes	P value	
Cluster A (2.18)	GO:0030695	Molecular function	GTPase regulator activity	ARAP1, EVIS, ARHGEF18, PLCE1, CHN1, THY1, TBC1D12, FBXO8	1.9E-03
	GO:0060589	Molecular function	Nucleoside-triphosphatase regulator activity	ARAP1, EVIS, ARHGEF18, PLCE1, CHN1, THY1, TBC1D12, FBXO8	2.3E-03
	GO:0008047	Molecular function	Enzyme activator activity	ARAP1, EVIS, CHN1, ALOX5AP, THY1, TBC1D12,	4.9E-03
	GO:0005096	Molecular function	GTPase activator activity	ARAP1, EVIS, CHN1, THY1, TBC1D12	7.7E-03
	GO:0005083	Molecular function	Small GTPase regulator activity	ARAP1, EVIS, ARHGEF18, THY1, TBC1D12, FBXO8	8.0E-02
	GO:0043087	Biological process	Regulation of GTPase activity	ARAP1, EVIS, THY1, TBC1D12	1.7E-02
	GO:0046578	Biological process	Regulation of Ras protein signal transduction	ARAP1, EVIS, ARHGEF18, TBC1D12, FBXO8	2.3E-02
	Cluster B (1.95)	GO:0001653	Molecular function	Peptide receptor activity	XCR1, CCR9, CXCR6, EDNRB, NMUR1
GO:0008528		Molecular function	Peptide receptor activity, G-protein coupled	XCR1, CCR9, CXCR6, EDNRB, NMUR1	5.0E-03
GO:0042277		Molecular function	Peptide binding	XCR1, CCR9, CXCR6, EDNRB, NMUR1	1.0E-02
GO:0004950		Molecular function	Chemokine receptor activity	XCR1, CCR9, CXCR6	1.6E-02
GO:0019956		Molecular function	Chemokine binding	XCR1, CCR9, CXCR6	1.9E-02
bta04060		KEGG pathway	Cytokine–cytokine receptor interaction	XCR1, CCR9, CXCR6, PRL, IL12RB1, ACVR2A	2.8E-02

PLCE1, *TBC1D12* and *FBXO8*, and enzyme activators, such as *ARAP1*, *EVIS*, *CHN1*, *ALOX5AP* and *THY1* (tables 2 and 3). The other eight genes encode signaling molecules and cell surface proteins implicated in peptide and chemokine receptor activities, such as *XCR1*, *CCR9*, *CXCR6*, *EDNRB* and *NMUR1*, and cytokine–cytokine receptor interaction pathways, such as the peptide and chemokine receptors *PRL*, *IL12RB1*, and *ACVR2A*, which participate in endocrine and autoimmune regulatory processes (tables 2 and 3).

On the basis of LRH test, the strongest signal among those detected in the 17 strong candidate genes implicated in response to the local climatic variation (tables 2 and 3) resides within a 92-kb region that lies entirely within the gene *TBC1D12*. Only in this gene core haplotypes show significantly ($P < 0.01$; table 4) high extended haplotype homozygosity (EHH)/relative EHH (REHH) values. In *TBC1D12*, three SNPs define four core haplotypes (denoted *TBC1D12-CH1* to *CH4*; table 4). The OAR22_18929579-A allele is carried only on the core haplotype *TBC1D12-CH1* (table 4), which is common (61%) in populations under low temperature and high precipitation climate (with positive PC1 values in fig. 3B). *TBC1D12-CH1* demonstrates clear long-range, high LD, as shown in the haplotype bifurcation diagrams (fig. 7A) and has correspondingly high EHH values at long distances ($EHH \geq 0.8$ at the 1-Mb distance tested; fig. 7B). We further compared the REHH values of core haplotypes and found that *TBC1D12-CH1* is a clear outlier (fig. 7C) with a statistically significant ($P < 0.01$; table 4) higher REHH value than those of the other haplotypes of comparable frequencies.

Testing Candidate SNPs and Genes under Divergent Selection

We did not observe large allele frequency variation in 2,000 randomly selected SNPs among the 32 old, autochthonous

populations (supplementary fig. S3A, Supplementary Material online), which have a worldwide range of geographic origin and climate adaptation. In contrast, allele frequencies of the 230 overlapping candidate SNPs showed a large range of variation among these populations (supplementary fig. S3A, Supplementary Material online). We also observed significant differences in the distribution of Spearman's rank correlation coefficients between the 230 candidate SNPs and the 2,000 randomly selected loci (supplementary fig. S3B, Supplementary Material online): The distribution appears to be normal for the 2,000 random SNPs, whereas the values are mostly at the positive and negative extremes for the 230 candidate SNPs. A large majority (82%, 189/230; data not shown) of the 230 candidate SNPs showed significant ($P < 0.05$) correlations between their frequencies and the PC1 values. In particular, the frequencies of the OAR22_18929579-A allele and core haplotype *TBC1D12-CH1* (table 4) in *TBC1D12* showed significant ($\rho = 0.636$, $P < 0.001$; $\rho = 0.756$, $P < 0.001$) correlation with the variation of the PC1 value in the populations ($n = 32$). We further examined the global distribution of the allele OAR22_18929579-A (fig. 8A) and of the haplotype *TBC1D12-CH1* (fig. 8B) and found that both are at high frequency in cold and humid regions, as northern Europe and United Kingdom, and at low frequency in high temperature and dry regions, as the Near East, South Asia, and Africa. Our results further confirmed that the allele OAR22_18929579-A and the core haplotype *TBC1D12-CH1* in the strong candidate gene *TBC1D12* appear to be under strong selection in response to environmental stress.

Discussion

By combining the patterns in SNP variation with environmental variables, we present the results of a genome scan that was used to detect the signatures of natural selection in response to climatic variation. In the genomic regions significantly

Table 3. The 17 Strong Candidate Genes and Associated SNPs within or Neighboring the Genes under Environmental Adaptive Selection Based on Three Different Approaches.

Chr.	Gene	Position (bp)	Gene Size (kb)	SNP	F_{ST} ^a	z Score ^b	Max Wald ^c
1	<i>EVIS</i>	68957918–69191464	233.50	OAR1_73673800 ^d	0.55**	5.35	15.81
2	<i>FBXO8</i>	105386295–105423924	37.63	OAR2_113355547	0.37*	9.74	17.59
2	<i>CHN1</i>	134023000–134022999	70.92	OAR2_142314137	0.29**	5.42	14.37
				OAR2_143195752	0.46**	5.48	15.62
2	<i>ACVR2A</i>	160457581–160549550	91.97	OAR2_170004218	0.35**	6.99	15.06
2	<i>NMUR1</i>	232366028–232483197	117.20	s74305	0.51**	11.20	16.41
				OAR2_245260141	0.32*	8.71	17.33
				OAR2_245638268	0.39*	8.35	14.93
				OAR2_245680195	0.42*	10.37	14.57
5	<i>IL12RB1</i>	4854306–4869950	15.64	s33778 ^d	0.35**	5.40	15.87
5	<i>ARHGEF18</i>	13935201–14021434	86.23	s48780 ^d	0.38**	8.02	14.88
10	<i>ALOX5AP</i>	30365435–30388526	23.09	s18834	0.84**	8.79	17.66
				s68983	0.67**	8.61	18.85
				s12004	0.54**	7.06	14.67
				OAR10_30746533	0.70**	7.24	17.37
10	<i>EDNRB</i>	53508345–53534498	26.15	OAR10_53694894	0.23*	6.83	19.21
				OAR10_54710595	0.33**	5.46	17.06
15	<i>THY1</i>	29450455–29452173	1.72	s13895	0.36*	6.74	16.08
				s02747	0.38*	6.38	15.70
15	<i>ARAP1</i>	50442784–50529042	86.26	OAR15_55184101 ^d	0.63**	7.54	15.70
19	<i>XCR1</i>	53235736–53236748	1.01	s18532	0.61**	8.60	18.16
				OAR19_56419475	0.27*	5.25	16.21
19	<i>CXCR6</i>	53290059–53291081	1.02	s18532	0.61**	8.60	18.16
				OAR19_56419475	0.27*	5.25	16.21
19	<i>CCR9</i>	53333188–53340782	7.60	s18532	0.61**	8.60	18.16
				OAR19_56419475	0.27*	5.25	16.21
20	<i>PRL</i>	34258080–34266415	8.34	OAR20_37437726	0.57**	9.54	19.18
				s39515	0.62**	9.90	19.15
22	<i>PLCE1</i>	15009437–15337740	328.30	OAR22_18841208_X ^d	0.46**	8.24	14.55
				OAR22_18876523	0.53**	7.75	16.38
				OAR22_18929579	0.47**	5.90	15.00
				OAR22_19052408	0.44**	5.89	16.91
22	<i>TBC1D12</i>	15398171–15490853	92.68	OAR22_18841208_X	0.46**	8.24	14.55
				OAR22_18876523	0.53**	7.75	16.38
				OAR22_18929579 ^d	0.47**	5.90	15.00
				OAR22_19052408	0.44**	5.89	16.91

^aBased on the F_{ST} -based selection test between the 2 groups of 11 populations (see Materials and Methods).

^bBased on the LFMM test (Frichot et al. 2013).

^cBased on the spatial analysis method (Joost et al. 2007).

^dSNPs within genes.

Significant at the level of * $P < 0.05$ and ** $P < 0.01$.

Table 4. Core Haplotypes of Gene *TBC1D12* in the 2 Groups of 11 Old, Autochthonous Breeds.

Core Haplotype	Core SNP Alleles (distance to the peak SNP)			Core Haplotype Frequencies			EHH ^b	REHH ^b	REHH ^b P Value
	OAR22_18924267 (–5 kb)	OAR22_18929579 (0 kb)	OAR22_19020858 (87 kb)	Total	Group 1 ^a	Group 2 ^a			
CH1	A	A	A	0.28	0.03	0.61	0.02/0.02	1.95/1.09	0.02*/0.14
CH2	A	G	A	0.20	0.20	0.17	0.01/0.02	0.50/1.14	0.31/0.12
CH3	C	G	A	0.08	0.07	0.11	0.01/0.01	0.43/0.62	0.33/0.27
CH4	C	G	G	0.38	0.62	0.10	0.01/0.01	0.72/0.68	0.23/0.25

^aThe selected 2 groups of 11 old, autochthonous breeds are detailed in Materials and Methods. Group 1 includes the six breeds of Indian Garole, Red Maasai, Ronderib Afrikaner, Cyprus Fat-Tail, Karakas and Afshari, and group 2 includes the five breeds of Scottish Blackface, Border Leicester, Spael-white, Engadine Red Sheep and Finnsheep.

^bEHH, REHH, and the P value of REHH are shown for the upstream and downstream sides from each core in group 2 (see fig. 7).

*Significant at the level of $P < 0.05$.

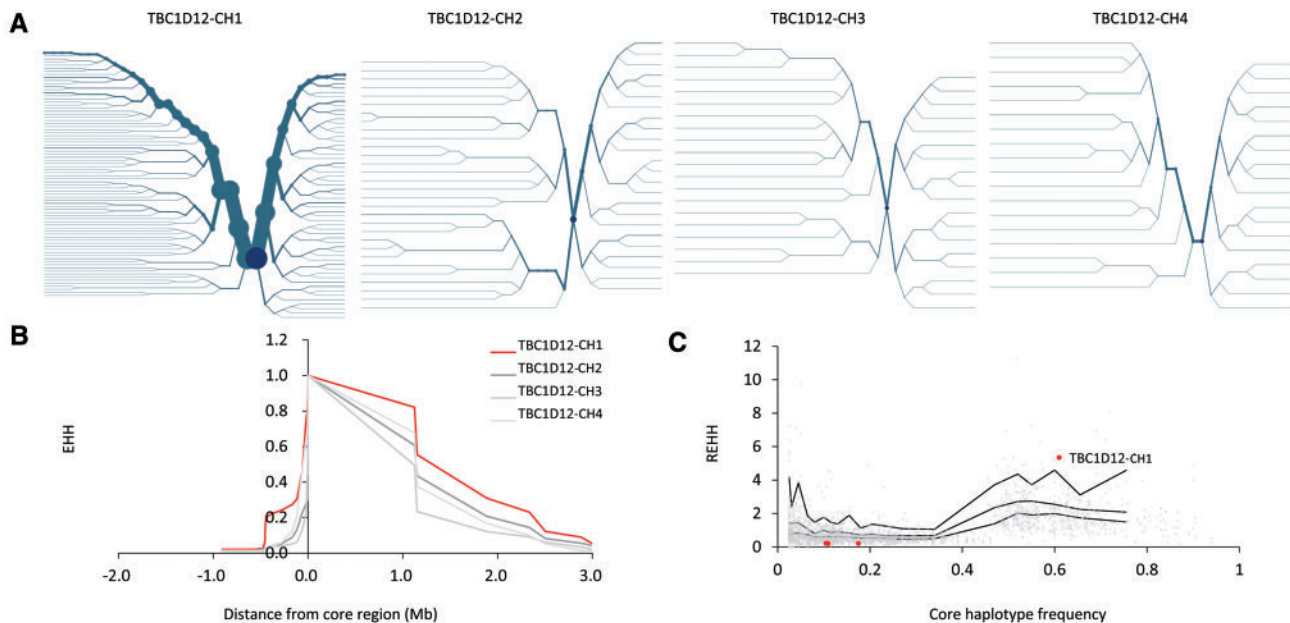


Fig. 7. Haplotype bifurcation diagrams, EHH and REHH of the gene *TBC1D12*. They are indicated in group 2, including five breeds of Scottish Blackface, Border Leicester, Spael-white, Engadine Red Sheep, and Finnsheep (see Materials and Methods). (A) Bifurcation diagrams (see Materials and Methods) for the four core haplotypes (CH1–CH4), respectively; (B) EHH values at varying distances from the core region on each of the four core haplotypes at *TBC1D12*; (C) distribution of REHHs versus core haplotype frequency. The REHHs were compared with the distribution of all of the other possible core haplotypes, as represented by gray dots; the 95th, 75th, and 50th percentiles are shown. The four core haplotypes at *TBC1D12* are marked in red, and *TBC1D12-CH1* is beyond the 95th percentile.

associated with climate-driven genetic adaptation, we found 17 strong candidate genes (see table 3). These results, together with knowledge regarding the molecular function of the candidate genes, provide new information on the genetic mechanisms likely underlying environmental adaptation in domestic animals.

Impact of Climate on Energy Metabolism and Endocrine and Autoimmune Regulation

Our results suggest that the process of autochthonous sheep breeds adaptation to extreme climates is principally mediated by complex, integrated energy metabolic responses, as observed in rodents (Rezende et al. 2004). Climate is known to have an important impact on animal physiology and fitness, particularly those of ruminants (Hofmann 1989; Bradshaw and Holzapfel 2010). Factors such as the temperature, solar radiation, UV radiation, precipitation and humidity all have direct effects, whereas factors such as the digestibility, quality and quantity of forage, which are themselves directly influenced by climate variables such as sunlight, precipitation and temperature, all have indirect strong effects on sheep (McManus et al. 2011). These three climatic variables can affect sheep directly, through effects on thermoregulation (Parker and Robbins 1985), but the stronger effect is expected to operate indirectly, through the metabolic regulation of plant quality and biomass (e.g., Mysterud et al. 2001; Mysterud and Austrheim 2008; Nielsen et al. 2013). For example, when facing thermal stress (i.e., a deviation from the relatively narrow range of body temperature that is optimal

for the coordination of molecular and cellular processes), animals minimize adverse effects by adjusting feed intake (variety and amount) and therefore energy metabolic processes (Hahn 1999; Mader 2003; Hancock, Brachi, et al. 2011; McManus et al. 2011). Consequently, long-term thermal stress can result in energy metabolic adaptation, as well as heat and cold tolerance, in particular breeds. Meanwhile, variation in animal morphology, including body size (large vs. small) and shape (fat-tailed vs. thin-/short-tailed), also follows basic thermoregulatory principles to dissipate or conserve body energy in different climates. In addition, several breeds (e.g., Norwegian White Sheep) grow particularly large in cold regions that are dry/warm in early spring and warm/wet in late summer, where conditions favor continued grass growth (Nielsen et al. 2013); however, several breeds grow small in the tropics and subtropics, where the tropical grasses (C4) have a lower nutritional value and higher fiber content (McManus et al. 2011). Thus, climates have a great impact on energy metabolic adaptations across native sheep breeds.

We also demonstrate evidence for the selection of candidate genes involved in endocrine regulation, which is consistent with the impact of the day length (e.g., photoperiod: the duration of light in a light/dark cycle), refractory period (animals do not respond to day length), and seasonal timing on animal physiology and evolutionary fitness (Bradshaw and Holzapfel 2010). The sunlight (or daytime) length, which is necessary for organisms to program seasonal changes in their lifecycles, changes as climate changes. For example, climates in the north have shorter growing seasons, and winter comes earlier in the year than in the south. Typically, the sunlight

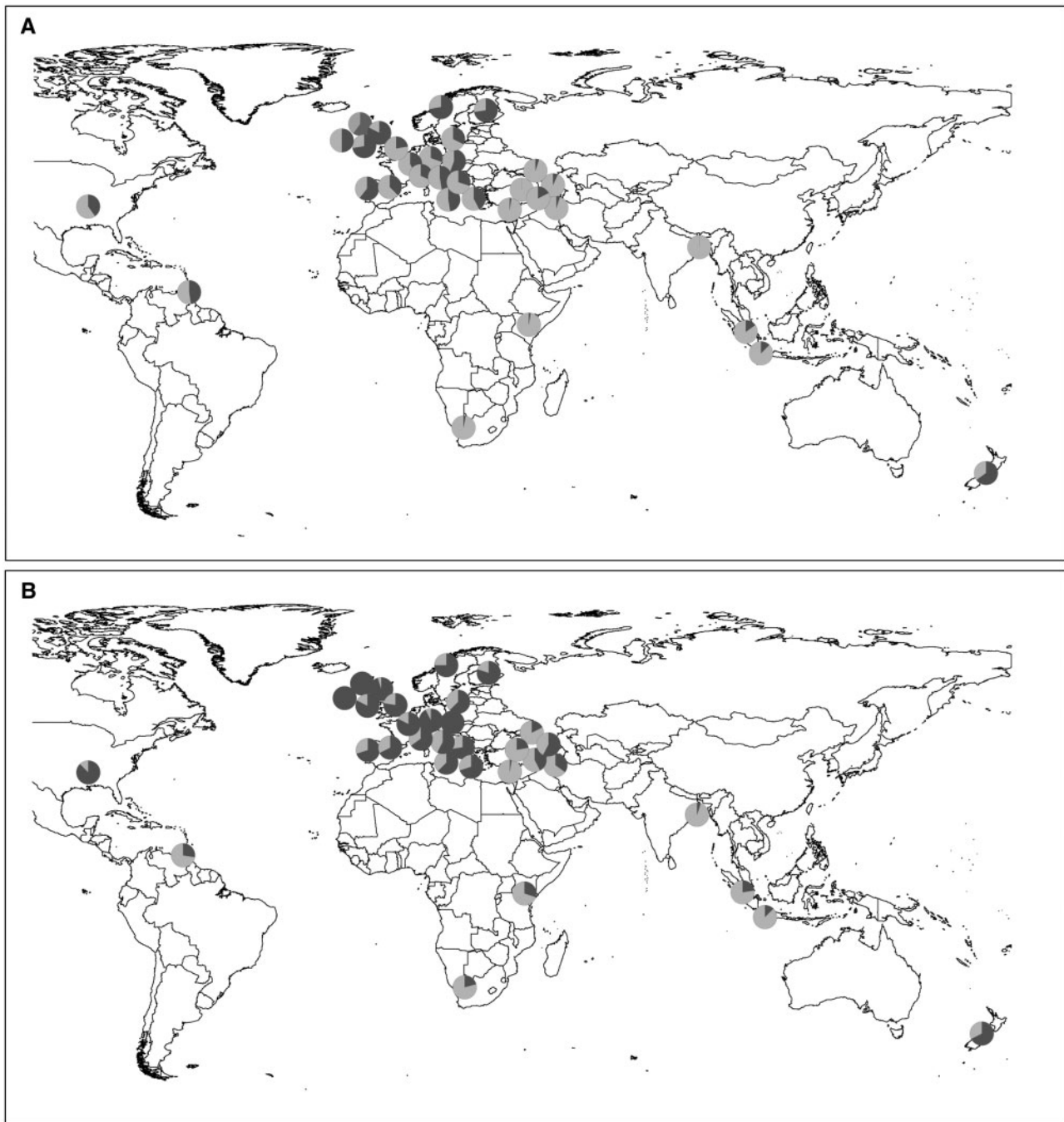


FIG. 8. Worldwide frequency distribution of the SNP allele and core haplotype. (A) The allele OAR22_18929579-A and (B) the core haplotype *TBC1D12-CH1*. The frequency is indicated in dark gray.

length influences physiological activities by mediating specific endocrine hormones that ultimately determine, for example, the timing of reproduction in mammals. In sheep, most breeds are seasonal breeders. Similar to seasonal reproduction in other mammals, the priming of the neuroendocrine axis of sheep begins with the input of sunlight and terminates with the secretion of a gonadotropin. Sunlight thus plays an important role in initiating a cascade of physiological events and serves as an anticipatory cue for the timing of reproduction events in sheep (Gomez-Brunet et al. 2008; Chemineau et al. 2010). As recent rapid climatic changes have occurred, one of

the major biotic responses in sheep has been the altered timing of seasonal events such as reproduction (see the review in Rosa and Bryant 2003), and the response has imposed selection on the interpretation of sunlight and its hormonal integration (Bradshaw and Holzapfel 2010).

Moreover, our findings indicate that climate can also affect autoimmune regulation in animals indirectly, through its effects on the environments that sheep inhabit. Among these the shaping of pathogens and patterns has evolutionary consequences (Chessa et al. 2009; Kenyon et al. 2009), including both the emergence of new disease syndrome and the change

in the prevalence of existing vectors (Summers 2009; Tabachnick 2010; McManus et al. 2011). For example, field studies on *P. odocoilei* have shown a phylogenetically distinct protostrongylid of Dall's sheep in a subarctic alpine habitat, and the results indicated that the possible effects of climate change include parasite range expansion and amplification in endemic regions (Kutz et al. 2005; Jenkins et al. 2006).

Putative Candidate Genes for Environmental Genetic Adaptation

Analysis using the EHH/REHH and correlation analyses provided independent evidence supporting the proposed signature of environmental genetic adaptation in *TBC1D12*. The *TBC1D12* gene is a GTPase activator for Rab family protein(s) and plays a key role in increasing the GTPase activity (Ishibashi et al. 2009). GTPases are enzymes that catalyze the hydrolysis of GTP, which serves as an energy source or activator of substrates in metabolic reactions within the cell. Thus, the signatures of local genetic adaptation in *TBC1D12* are most likely caused by direct environment stressors that mainly stem from temperature and sunlight.

More specifically, the frequencies of the allele OAR22_18929579-A and the core haplotype *TBC1D12-CH1* have an apparent pattern of global distribution (fig. 8). This observation is consistent with the conclusion that variants that are deleterious in hot equatorial climates become advantageous (rather than simply neutral) in colder climates (Hancock et al. 2008). Therefore, the selected SNP OAR22_18929579 and haplotype (*TBC1D12-CH1*) in gene *TBC1D12* identified by our scan can be considered a most probable "selection candidate" for complex adaptive traits. The results suggest that during the process of breed development, sheep with varying performances in energy metabolism were strongly selected due to their capability to adapt to the local environment.

Apart from the *TBC1D12* gene, several SNPs were also found in other candidate genes with enriched functions (table 2). None of the 17 major candidate genes has shown strong evidence of environmental genetic adaptation in previous human studies (supplementary table S6, Supplementary Material online). One of the main reasons is most likely the diverse sets of selective pressures acting on humans and sheep. In humans, environmental pressure is primarily driven by factors such as pathogens (Fumagalli et al. 2011), disease (Sabeti et al. 2007), diet (Hancock et al. 2010), altitude (Huerta-Sánchez et al. 2013), and climate (Frichot et al. 2013). As such, previous studies have identified a number of candidate genes related to skin pigmentation, metabolism, and immune function. However, among a remarkably diverse range of environments, climatic factors such as temperature, precipitation and sunlight (Nielsen et al. 2012, 2013) have imposed major environmental stress on sheep indirectly through forage, and thus, the selection of genetic variants in genes associated with development and energy metabolism is likely to have occurred. Although metabolism seems to apply to both humans and sheep, indirect climatic effects through the diet should be much larger

on sheep than on human due to 1) forage is the main diet source for sheep; and 2) ruminant has a particular plant digestion and metabolism system. Of the association models identified by the MatSAM analysis (supplementary table S3, Supplementary Material online), the variable SUN—which is the percentage of sunlight available—is one of the variables predominantly involved in the best models, together with precipitation variables (e.g., OAR1_88143×SUN-March, s73618×SUN-May, s53592×SUN-June, s71447×SUN-May, s74273×SUN-July, s04088×PR-October, OAR1_8932175×PR-October, and OAR1_16697706×PR-October; supplementary table S2, Supplementary Material online). These climatic parameters are closely associated with the energy metabolism of sheep through direct and indirect effects (e.g., through the feed resources for sheep; Nemani et al. 2003; Hancock et al. 2010).

Methodological Considerations and Caveats

Discovery of the same set of candidate SNPs, genes, and genomic regions under selection using alternative approaches can provide sound evidence for selective signatures (Oleksyk et al. 2010). The strategy used in this study of combining complementary statistical approaches has the power to identify alleles that experience small shifts in frequency by selection through decreasing the number of false-positive associations (Frichot et al. 2013); thus, this strategy allows us to detect novel loci in which SNPs show subtle but consistent patterns across populations. Despite the occurrence of false positives, we believe that they would be minimal in the significant overlapping SNPs and in the genomic regions containing five consecutive significant SNPs identified by the approaches. Our results differ from the results of previous analyses that were based merely on broad-scale population differentiation (Kijas et al. 2012).

Earlier simulations showed that the LFMM approach is robust to population structure and demographic history to some extent in distinguishing signals of environmental correlation due to selection from correlations due to population structure (Frichot et al. 2013). Although we selected only some of the breeds in the selection tests, the confounding effect between interpopulation genetic relationships and environmental gradients still exists due to the isolation-by-distance (IBD) patterns of the genotyped breeds (fig. 4; see also Kijas et al. 2012). Thus, by introducing population structure as hidden variables, the LFMM approach can efficiently estimate the background residual levels of population structure and control for random effects due to population history and IBD (Frichot et al. 2013). In addition, the MatSAM approach makes the implicit assumption that each data point is equally informative and the errors are not correlated (e.g. due to population structure; Joost et al. 2008, 2013). Given the differences among approaches, there still seem to be some interesting signals in the tails of distributions identified by the individual approaches (supplementary table S3; Supplementary Material online). In particular, given that the LFMM approach is the only one that accounts for population structure, more focus should be given to the selective

SNPs/genes (supplementary table S4, Supplementary Material online) identified using it in future studies.

There are several caveats to our study. Because several other environmental variables are partially correlated with climatic factors (e.g., the strong correlation between temperature and the number of days with ground frost; fig. 2B), the signals of natural selection detected in this study are likely to be the result of both the direct and the indirect influences of climate on the sheep genome. It is difficult to distinguish the causal selective pressures from the other pressures by correlation because many environmental and ecological variables covary (Coop et al. 2010). In addition, we note that the environmental variable for each breed might vary within the habitat region; further, our SNP data represent only a fraction of the coding portion of the sheep genome. Thus, fine-scale environmental data throughout the habitat region for each breed and higher density SNPs would be more informative. In addition, the geographic structure that correlates with PC1 may have biased the detection of selective signals. By examining the distributions of allele frequencies (supplementary fig. S3A, Supplementary Material online) and their Spearman's rho with PC1 (supplementary fig. S3B and C, Supplementary Material online) for the 230 candidate SNPs and 2,000 randomly selected SNPs, we found that the effect of geographic structure should be very small in this case. Several recent papers noted that signals of population differentiation (e.g., F_{ST} and climate correlations) did not necessarily contain signals of hard selective sweeps (see the reviews in Pritchard et al. 2010; Frichot et al. 2013; Duforet-Frebourg et al. 2014; Ferrer-Admetlla et al. 2014). Thus, if the selection on standing variation underlies adaptation, signals of hard sweeps are not necessarily expected. Moreover, hard selective sweeps may be infrequent in species with moderate to small population sizes and may only play a more prominent role in species with large population sizes (see the review in Joost et al. 2013). Overall, we view this method as a powerful strategy to highlight interesting loci and correlations that will be further explored in future studies.

Conclusions

In conclusion, our study investigated the genetic architecture of climate-mediated adaptations in farm animals using a genome-wide scanning approach. We identified several candidate genomic regions and candidate genes underlying local adaptation in sheep. The 17 strong candidate genes (see table 3), in particular, the target gene *TBC1D12*, which is involved in controlling the activities of GTPase activators and regulators, can now serve as starting points for investigating biological processes and possible underlying mechanisms of environmental genetic adaptation. Our results suggest that the core haplotype *TBC1D12-CH1* and allele OAR22_18929579-A were most likely involved in the adaptation to local climates during the development of native sheep breeds. These results not only increase our understanding of the genetic landscape of climate-induced adaptation across the sheep genome but also have a more practical value in marker-assisted breeding in sheep. Future studies are needed to further confirm (e.g., experimentally measure the phenotypic difference between

sheep with and without the mutations) and refine our results by integrating additional genomic data (e.g., candidate gene sequencing, mRNA and microRNA expression profiling, and DNA methylation) with a more comprehensive array of environmental variables.

Materials and Methods

Sheep Breeds and Samples

An initial set of 32 old, autochthonous breeds (1,224 individuals; supplementary table S8, Supplementary Material online) were selected out of 74 domestic sheep breeds/populations (2,819 individuals) that had been sampled and genotyped within the Sheep HapMap project (for information on the populations and their geographic origins, see fig. 1 and Kijas et al. 2012). Within each breed, pairwise interindividual kinship coefficients (Φ) were calculated using genome-wide SNPs (~50 K SNPs) according to the methods detailed in Li et al. (2011). To reduce the effect of sample size differences on the following estimates (Kijas et al. 2012), the 20 individuals showing the least pairwise relatedness were selected in the analysis for the breeds with more than 20 samples, whereas all the individuals were included in the analysis for the three breeds (Boreray, Karakas, and Ronderib Afrikaner) with less than 20 samples (supplementary table S8, Supplementary Material online). Thus, a total of 632 individuals were selected, and most of the pairwise kinship coefficients (5,441/5,935, 91.68%) between individuals were less than that between half-sibs ($\Phi = 0.125$). Sampling locations of the sheep breeds were first obtained from the International Sheep Genomics Consortium (ISGC) contributors who had collected the samples; if no information was received from the contributors, the sampling location was assigned as the geographic coordinates of the centroid of the breed's traditional rearing area. The number of samples and geographic origins of the breeds, as well as the sampling locations (fig. 1) and coordinates, are provided in supplementary table S8, Supplementary Material online.

PCA Based on Climatic Variables

A total of 105 environmental parameters describing the sampling locations were used in PCA to distinguish breeds on the basis of their adaptation to different agroclimatic zones. Environmental variables including climatic and altitude data (supplementary table S8, Supplementary Material online) were obtained for each breed based on the coordinates of the sampling locations. The altitude was estimated with the help of the digital elevation model SRTM30 (Shuttle Radar Topography Mission; <http://www2.jpl.nasa.gov/srtm/>, last accessed June 3, 2014) developed by the National Aeronautics and Space Administration (NASA), which has 30 arc seconds of spatial resolution. Climatic data were composed of latitude/longitude grids with a resolution of 10 min (approximately 12 km at the latitude of Switzerland), containing yearly mean and monthly values of the eight variables over global land areas (see also Joost et al. 2007; Pariset et al. 2009). In this study, monthly parameters of the variables were also considered in order to take account of seasonality of, for

example, vegetation growth and lambing (Joost et al. 2007; Pariset et al. 2009). Climatic data were available in the global climate data set (<http://www.cru.uea.ac.uk/data>, last accessed June 3, 2014) of the Climatic Research Unit, Norwich and covered 40 years from 1961 to 2001 (New et al. 2002). The following climate variables were included: Mean diurnal temperature range in °C (DTR), number of days with ground frost (FRS), precipitation in mm/month (PR), the coefficient of variation of monthly precipitation in percent (PRCV), relative humidity in percentage (REH), percent of maximum possible sunshine (percent of day-length, SUN), mean temperature in °C (TMP), and number of days with >0.1 mm rain per month (RDO). High-resolution global distributions of the climate variables (yearly mean values) and their correlations are shown in figure 2. The correlations among the 105 climatic parameters (yearly mean and monthly values) are also shown in supplementary table S9, Supplementary Material online.

Ovine SNP50K BeadChip Genotyping and Quality Control

The ovine SNP50K BeadChip was developed by the ISGC (<http://www.sheepmap.org>, last accessed June 3, 2014; see also in Kijas et al. 2009). Details on the SNP discovery, design of the ovine array, and genotyping procedures can be found at the following address: <http://www.sheepmap.org/hapmap.php> (last accessed June 3, 2014). SNP quality control has been detailed elsewhere in Kijas et al. (2009, 2012) and Miller et al. (2011). The HapMap clean data set has 49,034 SNPs (the CSIRO Data Access Portal: <https://data.csiro.au>, last accessed May 9, 2014), which passed both the manufacturing process and rigorous quality controls. In brief, the quality control steps included the exclusion of markers based on assay abnormalities annotated by Illumina, Mendelian inconsistencies documented by the ISGC, discordant genotypes between experiments, minor allele frequency = 0, call rate <0.99, atypical X-clustering, evidence for a nearby polymorphism, compression, intensity values only, or evidence of a deletion. Genotypes in the PLINK format (Purcell et al. 2007) are available from the ISGC website. Furthermore, we excluded SNPs without chromosomal or physical locations and those on X- and Y-chromosomes, based on the information detailed in the Sheep Genome Browser Oar v3.1 (<http://www.livestockgenomics.csiro.au/cgi-bin/gbrowse/oarv3.1/>, last accessed June 20, 2014). Our working data set contained 47,050 SNPs (supplementary table S10, Supplementary Material online).

Analysis of Genetic Relationship among Populations

We evaluated the genetic relationships among the 32 sheep breeds previously selected to exclude closely related breeds in the analysis and to minimize the confounding signals of selection due to population structure. The pairwise Reynolds' genetic distances (D_R ; Reynolds et al. 1983) between populations were computed using the Arlequin v3.11 software package (Excoffier and Lischer 2010). Because the removal of SNPs in high LD has been shown to at least partially counter the effect of ascertainment bias on the estimates of genetic

relationships between populations (Herráez et al. 2009), the pairwise matrix of Reynolds' genetic distance was calculated based on 22,861 SNPs identified by the application of the LD pruning algorithm implemented in the PLINK indep-pairwise (50 5 0.05). This procedure calculates the LD between SNPs in windows containing 50 markers and removes one SNP from each pair when the r^2 LD index exceeds 0.05. Reynolds' genetic distances were then used to construct a neighbor-net network of genetic relationships among breeds (Bryant and Moulton 2004), employing the SplitsTree package v4.12 (Huson and Bryant 2006).

Screening for SNPs and Genomic Regions under Selection

The first method we used to detect selection signatures was the Bayesian test that was developed by Beaumont and Balding (2004). This test was applied to a subset of 11 breeds (Scottish Blackface, Border Leicester, Spael-white, Engadine Red Sheep, Finnsheep, Indian Garole, Red Maasai, Ronderib Afrikaner, Cyprus Fat-Tail, Karakas, and Afshari) selected from the 32 breeds as belonging to two well-separated clusters defined by the PCA of environmental data and being loosely related on the basis of Reynold's genetic distance. The Beaumont and Balding test, as implemented in the LOSITAN program (Antao et al. 2008; available from <http://popgen.eu/soft/lositan/>, last accessed April 29, 2014), was performed on 1) two groups of animals: One comprised individuals belonging to the six breeds with negative extreme PCA1 coordinates (group 1: Indian Garole, Red Maasai, Ronderib Afrikaner, Cyprus Fat Tail, Karakas, and Afshari) and the second group of individuals was from the five breeds with positive extreme PC1 coordinates (group 2: Scottish Blackface, Border Leicester, Spael-white, Engadine Red Sheep, and Finnsheep); and 2) the 30 (6 × 5) possible pairwise combinations of breeds belonging to the two opposite PCA clusters.

The program LOSITAN was used (<http://popgen.eu/soft/lositan/>, last accessed April 29, 2014) to generate 100,000 simulated loci. It produced an expected neutral distribution of F_{ST} values and a P value estimate (the probability of simulated $F_{ST} < \text{sample } F_{ST}$) for each SNP. Each simulation included 100 individuals per population, 2 populations, 47,050 loci, an expected F_{ST} value of 0.070, and a confidence interval (CI) = 0.95 under the infinite allele mutation model. This method identifies SNPs under selection based on the distributions of F_{ST} that were higher or lower than expected, controlling for the expected genetic heterozygosity (H_E). Two options—"neutral mean F_{ST} " and "force mean F_{ST} "—were adopted, as recommended in Antao et al. (2008). F_{ST} -outlier SNPs were extracted based on the 95th percentiles (Li et al. 2014). To putatively identify specific genomic regions under strong selective sweeps, we focused on "windows" of consecutive SNPs. We used a standard sliding window size of five SNPs across each chromosome to calculate the average P values. An average P value > 0.95 for at least five consecutive SNPs indicated a genomic region under divergent selection. Finally, we integrated the results of all of the pairwise

comparisons by plotting the genome against the number of times they show P value > 0.95 in the 30 contrasts.

Testing for Signatures of Local Adaptation

We used MatSAM v1.0, which was developed by Joost et al. (2008), to detect the markers associated with environmental variables. Rather than relying on population genetics theoretical models, this spatial analysis approach uses spatial coincidence (Goodchild 1996) to relate the genetic profile of study sheep to the environmental parameters measured at the geographic coordinates of their sampling sites. The data used for the analysis are in the form of a matrix, in which each row corresponds to an individual and to the geographic coordinates where it was sampled; the columns contain 1) binary information (1 or 0) for the presence or absence of a given SNP allele; and 2) values of environmental parameters at the sampling location. The approach is performed at the individual genotype level, and multiple univariate logistic regression analyses are calculated to determine the degree of association between the frequencies of each allele and the values of the environmental parameters. The significance of the associations is determined with a log-likelihood (G) test and a Wald test (Joost et al. 2007). Bonferroni correction is applied to correct for multiple comparisons (Joost et al. 2008). By calculating the significance of the models generated by all possible pairwise combinations (allele vs. environmental parameter), the markers implicated in the models that emerge as statistically significant can be detected, and these loci are likely to be under the selective sweeps of environmental adaptations.

We further calculated the correlations between SNPs and climate variables using new algorithms implemented in the computer program LFMM (Frichot et al. 2013; available from <http://membres-timc.imag.fr/Eric.Frichot/lfmm/index.htm>, last accessed January 1, 2014). The new algorithms, which are based on population genomics, ecological modeling and statistical learning techniques, have proven to be efficient in screening genomes for signatures of local adaptation by decreasing the number of false-positive associations due to, for example, population structure and random effects (Frichot et al. 2013). Because the environmental variables were mainly related to temperature, sunlight and precipitation variables and PC1 explained most of the total variance (50.78%), which greatly exceeded that explained by PC2 (18.31%), we summarized the variables by using the first axis of the PCA (see above) for all of the environmental variables. We applied the LFMM algorithm and calculated the $|z|$ scores for all of the SNPs using 100 sweeps for burn-in and 1,000 additional sweeps. We used $K = 3$ latent factors based on population structure analyses using the program SmartPCA from the EIGENSOFT v5.0 package (<http://www.hsph.harvard.edu/alkes-price/software/>, last accessed January 1, 2014) and the Bayesian clustering program STRUCTURE v2.3.4 (Pritchard et al. 2000; for the results, see [supplementary material](#) and [fig. S4](#), [Supplementary Material](#) online).

Candidate SNP Annotation, GO, and EHH Analysis

We annotated the overlapping candidate SNPs detected by the selection tests to identify candidate functional genes under selection. We inferred the gene annotation of the mapped interval from the Ovine (Texel) v3.1 Genome Assembly (<http://www.livestockgenomics.csiro.au/cgi-bin/gbrowse/oarv3.1/>, last accessed June 20, 2014). In this analysis, we defined the “target region” as approximately 5,000 bp upstream and downstream of the significant SNPs (the genomic position of significant SNP $\pm 5,000$ bp; see also Li et al. 2013). Target genes were also searched for between two significant SNPs ≤ 1 Mb and identified as candidate genes under selection.

GO analyses were further performed using the function annotation clustering tool from DAVID Bioinformatics Resources 6.7 (available from <http://david.abcc.ncifcrf.gov>, last accessed July 2, 2014; e.g., Huang et al. 2008, 2009). When assessing evidence for enrichment from SNP-based analyses using this approach, a bias for larger genes is likely due to the increased probability of finding a signal by chance. Of the 230 candidate SNPs identified above, we further removed the SNPs that showed significant ($r^2 > 0.5$) LDs with one or multiple other loci and only considered a single signal from the SNPs showing significant LDs between each other. The DAVID enrichment analysis was then based on the SNPs and candidate genes after the filtering. We intended to determine which GO categories are statistically overrepresented in a set of genes compared with that expected in random positions (Fumagalli et al. 2011). We chose bovine (*Bos taurus*) known genes as the background set of genes, because a large portion of the candidate genes were mapped on a bovine rather than ovine genome (see Results). All the GO term accession numbers for each gene from the bovine genome were downloaded from Ensembl (<http://www.ensembl.org>, last accessed July 2, 2014, release 74); these data included the entire annotation and were considered as the reference set. Enriched terms were retrieved with a significantly higher than expected number of associated genes using Fisher's exact test. Any predominant functional theme of interesting gene sets in the GO hierarchy was shown if the P value was less than 0.05.

Candidate genes under recent strong selection should demonstrate EHH (Sabeti et al. 2002; Huerta-Sánchez et al. 2013) because of the low number of recombinations occurring in only a few generations (Sabeti et al. 2002). We further performed LRH tests in the two groups of populations (i.e., 11 populations) to identify the core haplotypes using the SWEEP software v1.1 (available from: <http://www.broadinstitute.org/mpg/sweep/>, last accessed July 10, 2014). We examined the EHH and REHH values in the 300 kb upstream and downstream of each core region within the 17 strong candidate genes (see Results) involved in the GO terms (Qanbari et al. 2010; Li et al. 2013). A pair of SNPs with the upper 95% CI of $D' = 0.7-0.98$ was defined to be in strong LD (Gabriel et al. 2002; Qanbari et al. 2010). Core haplotypes were set to include ≥ 3 SNPs. All of the haplotypes in the major candidate genes were divided into 20 bins based on their frequencies. We

compared the frequencies and REHHs for the core haplotypes in a candidate gene with those across the genome. *P* values were obtained by log-transforming the REHH values in the bin to achieve normality. Core haplotypes with extreme REHHs (beyond the 95th percentiles) in the empirical distribution were considered significant (Sabeti et al. 2002; Qanbari et al. 2010).

Validating the Evidence for Candidate SNPs and Genes under Selection

As true causal variants should display most signatures of positive selection (e.g., high-derived allele frequencies and long extended haplotype; e.g., Andersen et al. 2012; Li et al. 2013), we further validated the evidence for candidate SNPs and major genes under selection as identified above. Because favorable alleles of candidate genes under divergent selection tend to have greater frequencies in populations with higher relevant trait values (Orr and Kim 1998; Pritchard and Di Rienzo 2010; Turchin et al. 2012), we tested the correlation between the allele frequencies of the 230 candidate SNPs and PC1 values of the environmental variables, which explained most (50.78%) of the total variance of the environmental data, by examining the Spearman's rank correlation coefficient (ρ) and its statistical significance in the 21 breeds (by excluding the 11 breeds included in the F_{ST} -outlier selection tests from the 32 initially selected worldwide old, autochthonous sheep breeds) with a worldwide range of geographic origins and climates. We computed SNP allele frequencies in the population using the PLINK program (Purcell et al. 2007). The same statistical tests were also applied between frequencies of the core haplotypes for the 17 strong candidate genes and the PC1 values in the 21 populations. We further examined the distributions of allele frequencies, as well as their correlation coefficients with PC1, for the 230 candidate SNPs and 2,000 randomly selected loci from the total SNP set in the 32 populations.

Supplementary Material

Supplementary material, figures S1–S4, and tables S1–S10 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

The authors acknowledge the European Science Foundation (ESF)—Genomic Resources Programme and MTT Agrifood Research Finland international mobility grant for supporting Meng-Hua Li's visit to Università Cattolica del Sacro Cuore, Piacenza, Italy. This work was supported by the Chinese Academy of Sciences (the 100-talent Program of the Chinese Academy of Sciences), the National Natural Science Foundation of China (grant No. 31272413), and the Academy of Finland (grant Nos. 250633 and 256077) to M.-H.L. The International Sheep Genomics Consortium (<http://www.sheephapmap.org>) members who contributed to this work and/or coauthors of Kijas et al. (2012) include Juan-José Arranz, Universidad de León; Georgios Banos, Aristotle University of Thessaloniki; William Barendse, CSIRO,

Australia; Ahmed El Beltagy, Animal Production Research Institute; Jörn Bennewitz, University of Hohenheim; Simon Boitard, INRA, France; Steve Bishop, The Roslin Institute; Lutz Bunger, Scottish Agricultural College; Jorge H Calvo, CITA; Antonello Carta, AGRIS SARDEGNA; Ibrahim Cemal, Adnan Menderes University; Elena Ciani, University of Bari, Italy; Noelle Cockett, University of Utah; Brian Dalrymple, CSIRO, Australia; David Coltman, University of Alberta; Magali San Cristobal, INRA, France; Maria Silvia D'Andrea, Università degli Studi del Molise; Ottmar Distl, University of Veterinary Medicine Hannover; Cord Drögemüller, Institute of Genetics, University of Bern; Georg Erhardt, Institut für Tierzucht und Haustiergenetik Justus-Liebig-Universität Giessen; Emma Eythorsdottir, Agricultural University of Iceland; Kimberly Gietzen, Illumina Inc., United States of America; Elisha Gootwine, The Volcani Center; Vidja S. Gupta, National Chemical Laboratory; Olivier Hanotte, University of Nottingham; Ben Hayes, Department of Primary Industries Victoria, Australia; Mike Heaton, USDA; Stefan Hiendleder, University of Adelaide; Han Jianlin, ILRI and CAAS; Matthew Kent, CiGene; Johannes A. Lenstra, Utrecht University, the Netherlands; Terry Longhurst, Meat and Livestock Australia, Runlin Ma, Chinese Academy of Sciences; David MacHugh, University College Dublin, Jill Maddox, University of Melbourne; Massoud Malek, IAEA; Russell McCulloch, CSIRO, Australia; Md. Omar Faruque, Bangladesh Agriculture University; John McEwan, AgResearch, Invermay Agricultural Center, New Zealand; Despoina Miltiadou, Cyprus University of Technology; Carole Moreno, INRA; V Hutton Oddy, University of New England; Samuel Paiva, Genetic Resources and Biotechnology, Embrapa, Brasília, Brazil; Josephine Pemberton, University of Edinburgh; Fabio Pilla, Università degli Studi del Molise; Laercio R. Porto Neto, CSIRO, Queensland, Australia; Herman Raadsma, University of Sydney, Australia; Cyril Roberts, Caribbean Agricultural Research and Development Institute; Tiziana Sechi, AGRIS SARDEGNA; Bertrand Servin, INRA, France; Paul Scheet, University of Texas MD Anderson Cancer Center; Pradeepa Silva, University of Peradeniya; Henner Simianer, Universität Göttingen; Jon Slate, University of Sheffield; Miika Tapio, MTT Agrifood Research Finland; Selina Vattathil, University of Texas MD Anderson Cancer Center; and Vicki Whan, CSIRO, Australia.

References

- Ai HS, Huang LS, Ren J. 2013. Genetic diversity, linkage disequilibrium and selection signatures in Chinese and western pigs revealed by genome-wide SNP markers. *PLoS One* 8:e56001.
- Amaral AJ, Ferretti L, Megens HJ, Crooijmans RPMA, Nie HS, Ramos-Onsins SE, Perez-Enciso M, Schook LB, Groenen MAM. 2011. Genome-wide footprints of pig domestication and selection revealed through massive parallel sequencing of pooled DNA. *PLoS One* 6:e14782.
- Andersen KG, Shylakhter I, Tabrizi S, Grossman SR, Happi CT, Sabeti PC. 2012. Genome-wide scans provide evidence for positive selection of genes implicated in Lassa fever. *Philos Trans R Soc Lond B Biol Sci* 367:868–877.

- Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G. 2008. LOSITAN: a workbench to detect molecular adaptation based on a F_{ST} -outlier method. *BMC Bioinformatics* 9:323.
- Bamshad M, Wooding SP. 2003. Signatures of natural selection in the human genome. *Nat Rev Genet.* 4:99–111.
- Beaumont MA, Balding DJ. 2004. Identifying adaptive genetic divergence among populations from genome scans. *Mol Ecol.* 13:969–980.
- Bichi E, Frutos P, Toral PG, Keisler D, Hervás G, Lóor JJ. 2013. Dietary marine algae and its influence on tissue gene network expression during milk fat depression in dairy ewes. *Anim Feed Sci Technol.* 186: 36–44.
- Bradshaw WE, Holzapfel CM. 2010. Light, time, and the physiology of biotic response to rapid climate change in animals. *Annu Rev Physiol.* 72:147–166.
- Bryant D, Moulton V. 2004. Neighbor-Net: an agglomerative method for the construction of phylogenetic networks. *Mol Biol Evol.* 21: 255–265.
- Chemineau P, Malpoux B, Brillard J-P, Fostier A. 2010. Photoperiodic treatments and reproduction in farm animals. *Bull Acad Vet Fr.* 163: 19–26.
- Chessa B, Pereira F, Arnaud F, Amorim A, Goyache F, Mainland I, Kao RR, Pemberton JM, Beraldi D, Stear MJ, et al. 2009. Revealing the history of sheep domestication using retrovirus integrations. *Science* 324:532–536.
- Chu MX, Guo XH, Feng CJ, Li Y, Huang DW, Feng T, Cao GL, Fang L, Di R, Tang QQ, et al. 2012. Polymorphism of 5' regulatory region of ovine FSHR gene and its association with litter size in Small Tail Han sheep. *Mol Biol Rep.* 39:3721–3725.
- Chu MX, Mu YL, Fang L, Ye SC, Sun SH. 2007. Prolactin receptor as a candidate gene for prolificacy of Small Tail Han sheep. *Anim Biotechnol.* 18:65–73.
- Chu MX, Wang XC, Jin M, Di R, Chen HQ, Zhu GQ, Fang L, Ma YH, Li K. 2009. DNA polymorphism of 5' flanking region of prolactin gene and its association with litter size in sheep. *J Anim Breed Genet.* 126: 63–68.
- Coop G, Witonsky D, Di Rienzo A, Pritchard JK. 2010. Using environmental correlations to identify loci underlying local adaptation. *Genetics* 185:1411–1423.
- Duforet-Frebourg N, Bazin E, Blum MGB. 2014. Genome scans for detecting footprints of local adaptation using a Bayesian factor model. *Mol Biol Evol.* 31:2483–2495.
- Eckert AJ, van Heerwaarden J, Wegrzyn JL, Nelson CD, Ross-Ibarra J, González-Martínez SC, Neale DB. 2010. Patterns of population structure and environmental associations to aridity across the range of Loblolly Pine (*Pinus taeda* L., Pinaceae). *Genetics* 185:969–982.
- Excoffier L, Lischer HEL. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour.* 10:564–567.
- Ferrer-Admetlla A, Liang M, Korneliusen T, Nielsen R. 2014. On detecting incomplete soft or hard selective sweeps using haplotype structure. *Mol Biol Evol.* 31:1275–1291.
- Franks SJ, Hoffmann AA. 2012. Genetics of climate change adaptation. *Annu Rev Genet.* 46:185–208.
- Frichot E, Schoville SD, Bouchard G, François O. 2013. Testing for associations between loci and environmental gradients using latent factor mixed models. *Mol Biol Evol.* 30:1687–1699.
- Fumagalli M, Sironi M, Pozzoli U, Ferrer-Admetlla A, Pattini L, Nielsen R. 2011. Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. *PLoS Genet.* 7:e1002355.
- Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, et al. 2002. The structure of haplotype blocks in the human genome. *Science* 296:2225–2229.
- Gardner JL, Peters A, Kearney MR, Joseph L, Heinsohn R. 2011. Declining body size: a third universal response to warming? *Trends Ecol Evol.* 26:285–291.
- Gomez-Brunet A, Santiago-Moreno J, del Campo A, Malpoux B, Chemineau P, Tortonese DJ, Gonzalez-Bulnes A, López-Sebastián A. 2008. Endogenous circannual cycles of ovarian activity and changes in prolactin and melatonin secretion in wild and domestic female sheep maintained under a long-day photoperiod. *Biol Reprod.* 78:552–562.
- González J, Karasov TL, Messer PW, Petrov DA. 2010. Genome-wide patterns of adaptation to temperate environments associated with transposable elements in *Drosophila*. *PLoS Genet.* 6:e1000905.
- Goodchild MF. 1996. Geographic information systems and spatial analysis in the social sciences. In: Aldenderfer M, Maschner HDG, editors. *Anthropology, space, and geographic information systems*. New York: Oxford University Press. p. 214–250.
- Gordon IJ, Prins HHT. 2008. Grazers and browsers in a changing world: conclusions. In: Gordon IJ, Prins HHT, editors. *The ecology of browsing and grazing*. Berlin (Germany): Springer-Verlag. p. 309–321.
- Gu JJ, Orr N, Park SD, Katz LM, Sulimova G, MacHugh DE, Hill EW. 2009. A genome scan for positive selection in thoroughbred horses. *PLoS One* 4:e5767.
- Hahn GL. 1999. Dynamic responses of cattle to thermal heat loads. *J Anim Sci.* 77:10–20.
- Hancock AM, Alkorta-Aranburu G, Witonsky DB, Di Rienzo A. 2010. Adaptations to new environments in humans: the role of subtle allele frequency shifts. *Philos Trans R Soc Lond B Biol Sci.* 365: 2459–2468.
- Hancock AM, Brachi B, Faure N, Horton MW, Jarymowycz LB, Sperone FG, Toomajian C, Roux F, Bergelson J. 2011. Adaptation to climate across the *Arabidopsis thaliana* genome. *Science* 334:83–86.
- Hancock AM, Witonsky DB, Alkorta-Aranburu G, Beall CM, Gebremedhin A, Sukernik R, Utermann G, Pritchard JK, Coop G, Di Rienzo A. 2011. Adaptations to climate-mediated selective pressures in humans. *PLoS Genet.* 7:e1001375.
- Hancock AM, Witonsky DB, Gordon AS, Eshel G, Pritchard JK, Coop G, Di Rienzo A. 2008. Adaptations to climate in candidate genes for common metabolic disorders. *PLoS Genet.* 4:e32.
- Herráez DL, Bauchet M, Tang K, Theunert C, Pugach I, Li J, Nandineni MR, Gross A, Scholz M, Stoneking M. 2009. Genetic variation and recent positive selection in worldwide human populations: evidence from nearly 1 million SNPs. *PLoS One* 4:e7888.
- Hofmann RR. 1989. Evolutionary steps of ecophysiological adaptation and diversification of ruminants: a comparative view of their digestive system. *Oecologia* 78:443–457.
- Huang DW, Sherman BT, Lempicki RA. 2008. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* 4:44–57.
- Huang DW, Sherman BT, Lempicki RA. 2009. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37:1–13.
- Huerta-Sánchez E, DeGiorgio M, Pagani L, Tarekegn A, Ekong R, Antao T, Cardona A, Montgomery HE, Cavalleri GL, Robbins PA, et al. 2013. Genetic signatures reveal high-altitude adaptation in a set of Ethiopian populations. *Mol Biol Evol.* 30:1877–1888.
- Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol.* 23:254–267.
- Ishibashi K, Kanno E, Itoh T, Fukuda M. 2009. Identification and characterization of a novel Tre-2/Bub2/Cdc16 (TBC) protein that possesses Rab3A-GAP activity. *Genes Cells* 14:41–52.
- 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.
- Joost S, Bonin A, Bruford MW, Despres L, Conord C, Erhardt G, Taberlet P. 2007. A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. *Mol Ecol.* 16:3955–3969.
- Joost S, Kalbermatten M, Bonin A. 2008. Spatial analysis method (SAM): a software tool combining molecular and environmental data to identify candidate loci for selection. *Mol Ecol Resour.* 8: 957–960.
- Joost S, Vuilleumier S, Jensen JD, Schoville S, Leempoel K, Stucki S, Widmer I, Melodelima C, Rolland J, Manel S. 2013. Uncovering the genetic basis of adaptive change: on the intersection of

- landscape genomics and theoretical population genetics. *Mol Ecol*. 22:3659–3665.
- Karell P, Ahola K, Karstinen T, Valkama J, Brommer JE. 2011. Climate change drives microevolution in a wild bird. *Nat Commun*. 2:208.
- Kenyon F, Sargison ND, Skuce PJ, Jackson F. 2009. Sheep helminth parasitic disease in south eastern Scotland arising as a possible consequence of climate change. *Vet Parasitol*. 163:293–297.
- Kijas JW, Lenstra JA, Hayes B, Boitard S, Neto LRP, San Cristobal M, Servin B, McCulloch R, Whan V, Gietzen K, et al. 2012. Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS Biol*. 10: e1001258.
- Kijas JW, Townley D, Dalrymple BP, Heaton MP, Maddox JF, McGrath A, Wilson P, Ingersoll RG, McCulloch R, McWilliam S, et al. 2009. A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. *PLoS One* 4:e4668.
- Kutz SJ, Hoberg EP, Polley L, Jenkins EJ. 2005. Global warming is changing the dynamics of arctic host–parasite systems. *Proc Biol Sci*. 272: 2571–2576.
- Li MH, Strandén I, Tiirikka T, Sevón-Aimonen ML, Kantanen J. 2011. A comparison of approaches to estimate the inbreeding coefficient and pairwise relatedness using genomic and pedigree data in a sheep population. *PLoS One* 6:e26256.
- Li MH, Tiirikka T, Kantanen J. 2014. A genome-wide scan study identifies a single nucleotide substitution in *ASIP* associated with white versus non-white coat-colour variation in sheep (*Ovis aries*). *Heredity* 112: 122–131.
- Li Y, Reynolds A, Boyko AR, Wayne RK, Wu D-D, Zhang Y-P. 2013. Artificial selection on brain-expressed genes during the domestication of dog. *Mol Biol Evol*. 30:1867–1876.
- Luikart G, England PR, Tallon D, Jordan S, Taberlet P. 2003. The power and promise of population genomics: from genotyping to genome typing. *Nat Rev Genet*. 4:981–994.
- MacCallum C, Hill E. 2006. Being positive about selection. *PLoS Biol*. 4: 293–295.
- Mader TL. 2003. Environmental stress in confined beef cattle. *J Anim Sci*. 81:110–119.
- McManus C, Louvandini H, Gugel R, Sasaki L, Bianchini E, Bernal F, Paiva S, Paim T. 2011. Skin and coat traits in sheep in Brazil and their relation with heat tolerance. *Trop Anim Health Prod*. 43: 121–126.
- Meier K, Hansen MM, Bekkevold D, Skaala O, Mensberg KLD. 2011. An assessment of the spatial scale of local adaptation in brown trout (*Salmo trutta* L.): footprints of selection at microsatellite DNA loci. *Heredity* 106:488–499.
- Metallinos DL, Bowling AT, Rine J. 1998. A missense mutation in the endothelin-B receptor gene is associated with Lethal White Foal Syndrome: an equine version of Hirschsprung Disease. *Mamm Genome*. 9:426–431.
- Miller JM, Poissant J, Kijas JW, Coltman DW, Consortium ISG. 2011. A genome-wide set of SNPs detects population substructure and long range linkage disequilibrium in wild sheep. *Mol Ecol Resour*. 11: 314–322.
- Mysterud A, Austrheim G. 2008. The effect of domestic sheep on forage plants of wild reindeer; a landscape scale experiment. *Eur J Wildl Res*. 54:461–468.
- Mysterud A, Stenseth NC, Yoccoz NG, Langvatn R, Steinheim G. 2001. Nonlinear effects of large-scale climatic variability on wild and domestic herbivores. *Nature* 410:1096–1099.
- Nemani RR, Keeling CD, Hashimoto H, Jolly WM, Piper SC, Tucker CJ, Myrneni RB, Running SW. 2003. Climate-driven increases in global terrestrial net primary production from 1982 to 1999. *Science* 300: 1560–1563.
- New M, Lister D, Hulme M, Makin I. 2002. A high-resolution data set of surface climate over global land areas. *Clim Res*. 21:1–25.
- Nielsen A, Steinheim G, Mysterud A. 2013. Do different sheep breeds show equal responses to climate fluctuations? *Basic Appl Ecol*. 14: 137–145.
- Nielsen A, Yoccoz NG, Steinheim G, Storvik GO, Rekdal Y, Angeloff M, Pettorelli N, Holand Ø, Mysterud A. 2012. Are responses of herbivores to environmental variability spatially consistent in alpine ecosystems? *Glob Chang Biol*. 18:3050–3062.
- Nielsen R, Hellmann I, Hubisz M, Bustamante C, Clark AG. 2007. Recent and ongoing selection in the human genome. *Nat Rev Genet*. 8: 857–868.
- Oleksyk TK, Smith MW, O'Brien SJ. 2010. Genome-wide scans for footprints of natural selection. *Philos Trans R Soc Lond B Biol Sci*. 365: 185–205.
- Orr HA, Kim Y. 1998. An adaptive hypothesis for the evolution of the Y chromosome. *Genetics* 150:1693–1698.
- Pariset L, Joost S, Marsan PA, Valentini A, Econogene Consortium. 2009. Landscape genomics and biased F_{ST} approaches reveal single nucleotide polymorphisms under selection in goat breeds of North-East Mediterranean. *BMC Genet*. 10:7.
- Parker KL, Robbins CT. 1985. Bioenergetics of wild herbivores. In: Hudson RJ, White RG, editors. Thermoregulation in ungulates. Boca Raton (FL): CRC Press. p. 161–182.
- Pritchard JK, Di Rienzo A. 2010. Adaptation—not by sweeps alone. *Nat Rev Genet*. 11:665–667.
- Pritchard JK, Pickrell JK, Coop G. 2010. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Curr Biol*. 20:208–215.
- Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945–959.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 81:559–575.
- Qanbari S, Pimentel ECG, Tetens J, Thaller G, Lichtner P, Sharifi AR, Simianer H. 2010. A genome-wide scan for signatures of recent selection in Holstein cattle. *Anim Genet*. 41:377–389.
- Qiu Q, Zhang G, Ma T, Qian W, Wang J, Ye Z, Cao C, Hu Q, Kim J, Larkin DM, et al. 2012. The yak genome and adaptation to life at high altitude. *Nat Genet*. 44:946–949.
- Ramey H, Decker J, McKay S, Rolf M, Schnabel R, Taylor J. 2013. Detection of selective sweeps in cattle using genome-wide SNP data. *BMC Genomics* 14:382.
- Rannikki AS, Zhang F-P, Huhtaniemi IT. 1995. Ontogeny of follicle-stimulating hormone receptor gene expression in the rat testis and ovary. *Mol Cell Endocrinol*. 107:199–208.
- Reynolds J, Weir BS, Cockerham CC. 1983. Estimation of the coancestry coefficient basis for a short-term genetic distance. *Genetics* 105: 767–779.
- Rezende EL, Bozinovic F, Garland T Jr. 2004. Climatic adaptation and the evolution of basal and maximum rates of metabolism in rodents. *Evolution* 58:1361–1374.
- Rosa HJD, Bryant MJ. 2003. Seasonality of reproduction in sheep. *Small Rumin Res*. 48:155–171.
- Sabeti PC, Reich DE, Higgins JM, Levine HZP, Richter DJ, Schaffner SF, Gabriel SB, Platko JV, Patterson NJ, McDonald CJ, et al. 2002. Detecting recent positive selection in the human genome from haplotype structure. *Nature* 419:832–837.
- Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, Xie XH, Byrne EH, McCarroll SA, Gaudet R, et al. 2007. Genome-wide detection and characterization of positive selection in human populations. *Nature* 449:913–918.
- Shimada Y, Shikano T, Merilä J. 2011. A high incidence of selection on physiologically important genes in the three-spined stickleback, *Gasterosteus aculeatus*. *Mol Biol Evol*. 28:181–193.
- Stella A, Ajmone-Marsan P, Lazzari B, Boettcher P. 2010. Identification of selection signatures in cattle breeds selected for dairy production. *Genetics* 185:1451–1498.
- Summers BA. 2009. Climate change and animal disease. *Vet Pathol*. 46: 1185–1186.
- Tabachnick WJ. 2010. Challenges in predicting climate and environmental effects on vector-borne disease epizootics in a changing world. *J Exp Biol*. 213:946–954.

Turchin MC, Chiang CWK, Palmer CD, Sankararaman S, Reich D, Hirschhorn JN, ANthropometric GI. 2012. Evidence of widespread selection on standing variation in Europe at height-associated SNPs. *Nat Genet.* 44:1015–1019.

Wilkinson S, Lu ZH, Megens H-J, Archibald AL, Haley C, Jackson JJ, Groenen MAM, Crooijmans RPMA, Ogden R, Wiener P. 2013.

Signatures of diversifying selection in European pig breeds. *PLoS Genet.* 9:e1003453.

Zhang L, Liu J, Zhao F, Ren H, Xu L, Lu J, Zhang S, Zhang X, Wei C, Lu G, et al. 2013. Genome-wide association studies for growth and meat production traits in sheep. *PLoS One* 8: e66569.