Deciphering local adaptation of native Indian cattle (*Bos indicus*) breeds using landscape genomics and in-silico prediction of deleterious SNP effects on protein structure and function

Shivam Bhardwaj1 · Sanjeev Singh2 · Indrajit Ganguly2 · Avnish Kumar Bhatia2 · S. P. Dixit2

Received: 7 September 2022 / Accepted: 25 January 2023 / Published online: 15 February 2023
© King Abdulaziz City for Science and Technology 2023

Abstract
India has 50 registered breeds of native cattle (*Bos indicus*) which are locally adapted to diverse environmental conditions. This study aimed to investigate the genomic basis of adaptation of native Indian cattle and to predict the impact of key SNPs on the amino acid changes that affect protein function. The Illumina 777 K BovineHD BeadChip was used to genotype 178 native cattle belonging to contrasting landscapes and agro-climatic conditions. The genotype-environment association was investigated with R. SamBada, using 5,74,382 QC passed SNPs and 11 predictor variables (10 multi-collinearity controlled environmental variables and 1 variable as “score of PCA” on ancestry coefficients of individuals). In total, 1,12,780 models were selected as significant \((q < 0.05)\) based on \(G\) score. The pathway ontology of the annotated genes revealed many important pathways and genes having a direct and indirect role in cold and hot adaptation. Only ten SNP variants had a SIFT score of < 0.05 (deleterious), and only two of them, each lying in the genes CRYBA1 and USP18, were predicted to be deleterious with high confidence. RaptorX predicted the tertiary structures of proteins encoded by wild and mutant variants of these genes. The quality of the models was determined using Ramachandran plots and RaptorX parameters, indicating that they are accurate. RaptorX and I-Mutant 2.0 softwares revealed significant differences among wild and mutant proteins. Adaptive alleles identified in the present investigation might be responsible for the local adaptation of these cattle breeds.

Keywords Adaptation · *Bos indicus* · Deleterious mutation · Ecological genetics · Environmental stress · Landscape genomics · Protein structure · SNP array

Introduction
India has a variable climate that comprises of a wide range of weather conditions across a vast geographic scale and topography and is distributed over 15 agro-climatic regions. It has resulted in availability of different types of feed and fodder resources. These selective forces have contributed to the development of diverse locally adapted breeds/populations of livestock species for a wide range of characteristics/traits over time. There are 50 registered cattle breeds in India (https://nbagr.icar.gov.in/en/registered-cattle/) that are well-tailored to a variety of climatic conditions, endowed with varying qualities, and mainly categorized into three utility groups: dairy, draft, and dual. These functional categorizations of Indian zebu breeds have also been endorsed by paternal lineages (Ganguly et al. 2020) and genome-wide Runs of Homozygosity (Dixit et al. 2020).

Higher resolution of livestock genome and the discovery of loci with potential ecological importance have been made possible by advances in genomics and bioinformatics. Landscape genomics examines the geographical and environmental factors that form the genetic structure of breeds or populations by integrating population genetics, spatial statistics, and landscape ecology (Manel et al. 2010). Landscape genomics-based environmental association studies in livestock species are important for understanding the genetic basis (Joost et al. 2007) and exploring the genomic regions associated with local adaptation, and are becoming increasingly popular with the availability of whole genome sequencing and/or SNP chip genotyping data.
In India, the majority of cattle adaptation research has been focused on the polymorphism of major heat shock protein genes (Sodhi et al. 2013) and their comparative expression (Kumar et al. 2015; Deb et al. 2014; Maibam et al. 2017). Impact of heat stress on cellular and transcriptional adaptation of mammary epithelial cells in riverine buffalo has also been investigated (Kapila et al. 2016). The genomic variants associated with local adaptation have been identified in Holstein Friesian, Ugandan, Russian and Sibe- rian cattle (Hayes et al. 2009; Stucki et al. 2017; Yurchenko et al. 2018; Igoshin et al. 2019), as well as South African (Mdladla et al. 2018) and worldwide goat populations (Bertolini et al. 2018). Similar work has also been reported in honeybee populations in the Iberian Peninsula and Africa (Henriques et al. 2018; Eimanifar et al. 2018). Under climate change scenarios, these identified markers may be useful in marker-assisted animal selection for better performance (Hayes et al. 2009). However, no such systematic genome-wide association analysis has been conducted to date to investigate the genomic basis of local adaptation of Indian cattle breeds. Hence, the aim of this research was to investigate the genomic variants linked to local adaptation in Indian cattle breeds living in contrasting landscapes and predict the resulting change in protein function, which could aid in adaptation and boost productivity and health.

Materials and methods

Collection of animal samples and SNP genotyping

The blood samples were collected randomly in compliance with the guidelines and regulations of the Institute Animal Ethics Committee (IAEC), National Bureau of Animal Genetics Resources (ICAR-NBAGR 2023), Karnal under the supervision of Veterinarians. All the experiments and protocols were duly approved by the Institutional Research Committee of ICAR-NBAGR (project code 7.70). To investigate the genetic basis of local adaptation, a total of 178 samples belonging to breeds from different agroclimatic conditions were genotyped using 777 K BovineHD BeadChip (Illumina) from eleven native cattle breeds. The genomic DNA was extracted from the whole blood using HiPurA™ SPP Blood DNA isolation kit (Himedia, Lab. Pvt. Ltd.). The quality of DNA samples was tested in 1% agarose gel electrophoresis and the concentrations were determined using a Quantus™ fluorometer (Promega Coorporation, USA).

Preparation of the SNP genotype data file

The quality control (QC) check on the SNP data was carried out using PLINK v 1.9 software (Purcell et al. 2007; Chang et al. 2015). The unmapped SNPs along with those located on X, Y chromosomes, and mitochondrial DNA were filtered out. SNPs with call rate <10%, minor allele frequency <0.05, and with significant deviation from Hardy–Weinberg equilibrium ($p < 10^{-3}$) were removed. All individuals with >10% missing SNPs were also excluded. The quality control (QC) measures led to a final data on 158 cattle samples belonging to Sahiwal (13), Tharparkar (17), Gir (15), Ongole (17), Hariana (18), Kangayam (16), Vechur (16), Siri (9), Ladakhi (11), Kankrej (14) and Hallikar (12) breeds.

Linkage disequilibrium (LD) and pruning of the data

The $r^2$ values were calculated using PLINK v 1.9 keeping the window size limit of 500 kb between pair-wise SNPs. The autosomal SNPs were further pruned out with an $r^2$ value >0.5 with sliding window of 100 Kb and step-size of 50 SNPs (Malomane et al. 2018) using PLINK v 1.9 so as to reduce the ascertainment biasness and overestimation of the diversity parameters.

Diversity and population structure analysis

The $r^2$ values were calculated using PLINK v 1.9 keeping the window size limit of 500 kb between pair-wise SNPs. The autosomal SNPs were further pruned out with an $r^2$ value >0.5 with sliding window of 100 Kb and step-size of 50 SNPs (Malomane et al. 2018) using PLINK v 1.9 so as to reduce the ascertainment biasness and overestimation of the diversity parameters.

Analysis of the population structure was carried out by STRUCTURE software (Pritchard et al. 2000) so as to extract the ‘population structure’ variable which was subsequently used along with the other variables in the phenotypic file (environmental file). This new variable ‘population structure’ was defined by performing a principal component analysis (PCA) on the coefficients of ancestry and used to represent the population structure in Samβada bivariate analyses so as to reduce the spurious associations between the genotypes and environmental variables.

Out of 178 samples, 158 individuals of 11 breeds remained after quality checking. Structure analysis clustered them into 5 sub-populations. The Principal Component Analysis (PCA) on the co-ancestry coefficient of these individuals at $K=5$ revealed that 29.5% variation of the data set was contributed by 1st Principal Axis (PC-1). When the number of animals was reduced to 69 by selecting less admixed individuals from structure plot of 158 individuals, these 69 animals were structured into 4 sub-populations. The Principal Component Analysis of the co-ancestry coefficient of these 69 individual at $K=4$ revealed 37.19% variation by 1st Principal Axis (PC-1). Animals of different breeds were grouped on the basis of the diverse agroclimatic conditions in which they inhabited i.e. hot versus cold (group1), cold versus humid (group2), and hot versus humid (group3). The breeds taken for hot-arid adaptation were Gir, Tharparkar, Kankrej; Ladakhi & Siri were taken for cold-high altitude adaptation; whereas Kangayam & Hallikar were considered for hot-humid. PCA of co-ancestry coefficients explained 100% variance in 1st PC itself for all these three groups.
Therefore in final data analysis, 46 animals belonging to the contrasting landscape (hot versus cold) conditions were used for further analysis so as to reduce the spurious associations between the genotypes and environmental variables while the admixture effect was taken care by PC-1. Pairwise $F_{ST}$ value was estimated for each locus by Arlequin 3.5 (Excoffier and Lischer 2010) and the most significant 3565 SNPs ($F_{ST}$ $p$-value <0.01) were further used in STRUCTURE software (Pritchard et al. 2000) to decipher the presence of population structure and account for individual ancestries. Unsupervised clustering analysis was performed with 100,000 MCMC and 100,000 Burnin (Porras-Hurtado et al. 2013) for $K$ = 1 to 5, each with 7 iterations. The STRUCTURE results were plotted by ‘pophelper’ package (Francis 2017) in R software.

### Determination of the environmental variables and preparation of the environmental (phenotypic) file

The 100 years historic climatic data was obtained from the India Meteorological Department (IMD) which consisted of average monthly maximum and minimum temperature and average rainfall. This data from the weather stations of the sampling location of animals was used to calculate 55 biopredictor variables (O'Donnell and Ignizio 2012) Samβada package in R was used for the association analysis. Location coordinates were also obtained for the sampling locations by setLocation() function of R.Samβada package. The monthly average UV index of sampling locations was obtained from Weather Atlas (2023) (https://www.weather-atlas.com). Finally, 59 environmental variables were obtained for each sample. To lower the dependency between models and spare computation time, the correlation coefficient between pair of variables was used to control the multicollinearity. The bivariate analysis was carried out by R.Samβada package in which population structure was also taken into account. PCA was performed using princomp function of R package. The maximum coefficient of correlation was set to 0.8 between pairs of variables using prepareEnv() function of R.Samβada package which reduced the data set to 10 variables (Fig. 2). Finally, 11 predictor variables (including the ‘population structure’ variable) were considered for Samβada bivariate analysis.

### Genotype-environmental association study

The .ped file of QC data was converted to .gds format by R package ‘SNPRelate’ (Zheng et al. 2012) for analysis by R.SamBada. Samβada was launched via R.Samβada’s SamβadaParallel() function that implements supervision by default, as it relies on the ‘doparallel’ R package. To detect selection signatures, Samβada processes each genotype independently. A locus is defined as ‘detected’ by Samβada if at least one of its three genotypes showed a significant association with an environmental variable. Samβada uses logistic regressions model to know the probability of observing a particular genotype of a polymorphic marker under the given environmental conditions at the sampling locations (Joost et al. 2007). These logistic models are calibrated using a maximum likelihood procedure. In bivariate models, the selection procedure accessed whether the environmental variable is associated with the genotype while taking into account the possible effect of admixture (population variable).

False positives were controlled by taking $q$-value for significance testing (Storey and Tibshirani 2003). The models having $q$-value below the threshold levels are selected as significant and the SNPs which were significantly associated with different environmental variables were obtained from the models showing $q$-value for the $G$-score below the threshold level (0.05 and 0.01).

### Gene annotation of the significantly associated SNPs

rsIDs of all the significantly associated SNPs were submitted to the Ensembl Variant Effect Predictor (VEP) tool to assess the effect of SNP variants on genes, transcripts, and protein sequence (McLaren et al. 2016) and a list of defined genes was prepared. These are the genes that are influenced by environmental variables. The functional classification of the genes obtained from VEP was performed by PANTHER ver.15 (Thomas 2003).

### Detection of deleterious non-synonymous SNPs

VEP gives the location of the variants in the genes. The genes showing variants in the coding region were identified. The genes showing the consequence of variants as non-synonymous, were extracted. VEP gives the SIFT scores for these missense variants. A SIFT score predicts whether an amino acid substitution affects protein function (Ng and Henikoff 2003). Variants with scores in the range of 0.0 to < 0.05 are considered deleterious while variants with score $\geq$ 0.05 were considered as tolerant.

### In silico analysis of the protein structure

The genes showing deleterious SNP mutations based on SIFT scores were selected for comparative structural analysis of the tertiary structure of the proteins encoded by the wild and mutant form of the genes. Both the sequences of the proteins (wild and mutant) were submitted to the RaptorX online server (http://raptorx.uchicago.edu/) which performs Domain parsing for protein prediction (Kallberg et al.
The 3D structure predicted by the RaptorX server was visualized in JMOL viewer directly on the RaptorX server. The “.pdb” files generated were also visualized by Pymol software (https://pymol.org).

**Assessment of model quality**

RaptorX gives various measures to determine the consistency of a projected 3D structure model, including \( P \)-value, Score, uGDT, and GDT. Further, PROCHECK v.3.5 (Laskowski et al. 1993) (https://servicesn.mbi.ucla.edu/PROCHECK) was used to generate Ramachandran plots for predicted protein structure of wild and mutant genes. Other than the General Ramachandran, separate plots were generated for Glycine and Proline residues, as the favorable and unfavorable regions of these residue types differ greatly from those of the other residues. A good quality protein structure prediction contains all the set of torsional angles in the allowed regions whereas, in a bad quality or low-resolution protein structure, large number of amino acid residues lies in the forbidden region.

**Binding site and ligand prediction**

RaptorX binding online server was used to determine the binding sites on the predicted protein structures of wild and mutant proteins encoded by genes showing deleterious mutations due to non-synonymous SNPs. The predicted pockets are listed in order of their likelihood of being a binding site. Pocket multiplicity is given to indicate the quality of the predicted pockets, which represents the frequency with which the selected pocket was found in the template structures. When the pocket multiplicity is above 40, there is a good chance that the predicted pocket is true. List of ligands which may bind with the predicted pocket was also obtained.

**Protein stability and structure alignment**

The protein variants were submitted to I-Mutant2.0 online web server to predict protein stability change due to single-site mutation in the protein sequence (Capriotti et al. 2005). DDG value was calculated as the unfolding Gibbs free energy value of the mutated protein minus the unfolding Gibbs free energy value of wild type (Kcal/mol). Under ternary classification DDG value < −0.5 indicates large decrease in stability of mutant protein; whereas > −0.5 indicates large increase in stability. Further, the protein sequences were submitted to RaptorX structure alignment server to get RMSD value (Å) and template modeling score (TM-score). If TM-score is > 0.6 two proteins share a similar fold (90% chances); whereas < 0.4 indicates that the two proteins have different folds (Kallberg et al. 2012). Usually protein pairs with a TM-score > 0.5 are mostly in the same fold while those with a TM-score < 0.5 are mainly not in the same fold (Xu and Zhang 2010).

**Results**

**Quality checking of the samples for filtration of genotypic data**

Initially (before QC), the number of loci for all the 46 animals were 7,77,962 (~ 777 k). Firstly, SNPs lying in autosomes were extracted which accounted for a total of 7,38,042 loci. Various QC parameters applied by PLINK finally reduced the effective number of loci to 5,74,382 (Table 1). 3,83,031 SNP variants out of 5,74,382 variants were further removed by LD based pruning.

**Generating predictor variables**

**Population variable**

Locus wise \( F_{ST} \) estimates obtained from Arlequin are presented in Manhattan graph (Fig. S1). A total of 3565 significant SNPs (\( P < 0.01 \)) were used to assess the structuring of the breeds under study. The delta \( K \) value for STRUCTURE was maximum when \( K = 2 \) (Fig. S2), suggesting two distinct clusters of these cattle which was in agreement with another model-based clustering approach, ADMIXTURE, where the lowest CV error was found at \( K = 2 \) (Fig. S3). PCA was performed on the ancestry coefficients obtained from STRUCTURE at \( K = 2 \) and the first PC itself explained 100% of the variability (Fig. 1). Table S1 shows ancestry coefficients of each individual for both the clusters along with the estimated PC score.

| QC parameter                                    | SNPs removed | SNPs remaining |
|------------------------------------------------|--------------|----------------|
| 777 K genotyping                                | 0            | 7,77,962       |
| Unmapped/sex chromosomes/mitochondrial DNA     | 39,920       | 7,38,042       |
| Hardy–Weinberg equilibrium (\( p < 0.001 \))   | 15,883       | 7,22,159       |
| Missing genotype data (\( > 10\% \) missing SNPs) | 11,722       | 7,10,437       |
| Minor allele threshold (\( < 0.05 \))           | 1,36,055     | 5,74,382       |
Environmental variables

The location coordinates were obtained from “setLocation” function of R.SamBada for four breeds of cattle viz. Siri, Hallikar, Kankrej and Ladakhi from Gangtok, Bangalore, Jodhpur and Leh; respectively. The monthly average climatic data obtained from India Meteorological Department for four sampling locations, which consisted of three parameters i.e. $T_{\text{MAX}}$—Maximum Mean Temperature (°C), $T_{\text{MIN}}$—Minimum Mean Temperature (°C) and MR—Mean Rainfall in mm (Table S2). Further, using this data other bio-predictor variables were estimated (Table S3). Finally, environmental file was generated which consisted 59 environmental variables (bio-predictors (19), altitude (1), UV index (1), monthly $T_{\text{MAX}}$ (12), $T_{\text{MIN}}$ (12) and MR (12), as well as two location coordinates) (Table S4). The correlation plot for these 59 variables is shown in (Fig. S4), here degree of blueness and redness represents amount of positive and negative correlation; respectively.

For the “maxcorr” argument of “prepareEnv” function under R.SamBada package, a threshold value of 0.8 was set which reduced the dataset to 10 variables out of 59 (Fig. 2). Thus, the variables left for further analysis were “longitude, latitude, Bio1, Bio2, Bio3, Bio4a, Bio12, tmax7, prec10, prec12”. A new variable “pop1” was added which was created as the “score of PCA” on ancestry coefficients of individuals.

---

**Fig. 1** Principal component analysis performed on the ancestry coefficients obtained from STRUCTURE at $K=2$ (two distinct clusters in the population) and the first principal component explained 100% of the variability.

**Fig. 2** To control multi-collinearity between environmental variables, a threshold value of 0.8 was set for maximum correlation which reduced the dataset to 10 variables from 59.
Bivariate analysis

A locus-specific landscape genomics analysis was performed with 5,74,382 SNPs to discover the SNPs substantially associated with environmental factors. SamBada was used to analyze all 46 samples using the R package “R.SamBada” which was run on 12 cores of High performing computation facility available at ICAR-NBAGR, simultaneously using the “doParallel” R package. It resulted in a total of 14,096,124 bivariate models, with environmental as the first variable and "pop1" as the second. A total of 1,12,780 (0.8%) estimated models had a \( q \)-value of \( G \)-score < 0.05 and were considered significant models. The number of significant models dropped dramatically to 10,155 (0.072%) when the threshold was raised to a \( q \)-value of \( G \)-score < 0.01. Manhattan plots (Fig. S5) were generated to depict the \( q \)-values for each environmental variable used in the model. The 30,350 distinct SNPs (Table S5) were extracted for further analysis from the models with \( q \)-value < 0.05, which accounted for 5.28% of the total SNPs (5,74,382) used in the association study. The number of genotypes associated with a variable was highest for “tmax7” (14,157) and lowest for “Bio1” (7,251). It was observed that all the variables had more than 10,000 associations, except "Bio1". 3,169 distinct SNPs associated with environmental variables were observed at \( q \)-value < 0.01. When individual variables were screened for associations they showed similar trend as observed in previous threshold of \( q \)-value < 0.05 (Fig. 3) i.e. “tmax7” showed highest (1,381) and “Bio1” lowest (503) number of associations.

Annotation of significant SNPs

The reference SNP ID numbers (rsIDs) of 30,350 significantly associated SNPs \( q < 0.05 \) were submitted to Ensembl Variant Effect Predictor (VEP) tool which processed 30,121 SNPs. A total of 4,435 gene annotations including 95 miRNA binding sites (Table S6) were obtained.

Many genes related to cold adaptation like \textit{EPAS1}, \textit{EGLN2}, \textit{EGLN3} etc. and chaperones involved in hot adaptation were identified. Majority of variants were lying in non-coding region, maximum of which were intronic variants (56%), followed by intergenic variants (32%). Other consequences are depicted in Fig. 4. Variants lying in coding regions were merely about 1% (70% synonymous and 30% missense variants) of the total variants (Fig. 5).

SIFT score classification for all the 104 unique missense variants is depicted in Table 2. The list of non-synonymous SNPs (72) and the genes (66) harbouring them showing tolerant mutations with high confidence is given in Table S7. Only 10 variants showed SIFT score < 0.05 (deleterious) out of which, 6 variants were lying in annotated genes. Only 2 variants, i.e. “rs110287779”
and “rs136891050” lying in genes CRYBA1 and USP18, respectively were predicted to be deleterious with high confidence (Table 3). The former was found to be associated with Bio3 and Latitude and later with Bio12 and longitude (q < 0.05). Table S8a and Table S8b show the gene and genotypic frequencies of SNP loci with rsID rs110287779 and rs136891050 in the CRYBA1 and USP18 genes, respectively. For rs110287779, the ‘A’ allele was found to be fixed in the Kankrej breed, while the deleterious mutant allele ‘G’ accumulated in the other three breeds. The minor allele ‘G’ frequency was almost equal (0.22) for high-altitude breeds (Siri and Ladakhi). However, for rs136891050, the major allele ‘G’ was found to be almost fixed in Ladakhi (0.955), whereas the ‘A’ allele was found to be more prevalent in Siri.

### Pathway and gene ontology analysis by PANTHER software ver.15

A pathway component represents a group of homologous proteins across various organisms that participate in the same specific biochemical reactions within the pathway. Since, evolution acts on genes involved in similar functional pathways within a network rather than on single genes (Wu et al. 2020), functional classification of the annotated genes was performed using PANTHER ver.15 (Thomas 2003). Pathway ontology determined a total of 141 pathways (Table S9). Some important interconnected pathways were also observed (Table S10).

### Protein structure prediction and visualization

The protein sequences were queried in RaptorX which returned PDB files holding the coordinates of the structures
of interest. These files were visualized by Pymol software, and the models for wild and mutant proteins were coloured according to the secondary structures as green for loop, red for helix and yellow for sheet. For CRYBA1 protein, both the sequences were predicted in a single domain. Best template was 3LWK from PDB archive, which encodes for crystal structure of human Beta-crystallin A4 (CRYBA4). USP18 structure was predicted in 2 domains. Domain1 of wild form consisted residues 1–106 and domain2 consisted of sequences from 167–406 amino acid positions. Model excluded 60 amino acid residues ranging from 107–166. In mutant form of the predicted protein structure, the two domains were, domain1 (1–108) and domain2 (164–406), and the model excluded 55 residues ranging from 109–163 in the amino acid sequence. Best template from PDB archive was 5CHT which encodes for crystal structure of USP18 of mouse. The predicted structures for wild and mutant forms of CRYBA1 and USP18 containing the loop, sheet and helix structures are depicted in Figs. 6 and 7, respectively.

The effect of mutations in the secondary structure distribution (Helices, Sheets and Loops) and the solvent accessibility of the predicted protein models for CRYBA1 and USP18 are given in Table S11. The solvent accessibility is divided into three states namely, ‘Buried’ for less than 10%, ‘Exposed’ for larger than 42% and ‘Medium’ for between 10 and 42%. It was observed that most of the secondary structures represented loops for both the genes variants with CRYBA1 protein having higher values (65–67% in CRYBA1 and 49–50% in USP18).

Model quality assessment

Quality parameters given by RaptorX

All the residues (215) were included in the predicted models of wild and mutant proteins of CRYBA1, whereas the predicted model consisted of 346 (85%) and 351 (86%) residues of wild and mutant protein of USP18, respectively. The relative quality of models based on p-value is excellent for both the forms of CRYBA1 and domain1 of USP18 (P-value < 10⁻⁴), but there may be some uncertainty in domain2 of USP18 as both the predicted structures were showing P-value > 10⁻⁴. The unnormalized Global Distance Test (uGDT) score is seen for models having more than 100 residues and for smaller proteins GDT is considered. All the predicted structures have a uGDT score > 50 (Table 4) which is an indicator of good model.

Ramachandran plot

Three Ramachandran plots were generated for each form of both the proteins, i.e. for general, glycine and proline.

CRYBA1

The general Ramachandran plot for wild form was generated for 186 residues, which excludes 17 glycine (shown as triangles), 10 proline and 2 residues from all the 215 residues. Here 91.40% residues (170) were falling in most favoured region, whereas only 2 residues (1.1%) i.e. Tyrosine-105 and...
Aspargine-155 were in disallowed region. In comparison to general Ramachandran plot for wild CRYBA1 protein, a little reduction to 89.80% (167 residues) in residues lying in most favoured region was seen in mutant type. Only one residue, Alanine-202 was observed to lie in disallowed region, which is an add-on to the predicted mutant protein (Fig. 8 and Table 5). Ramachandran plots for Glycine and Proline residues were generated separately for each of the protein variant. In wild form of Glycine Ramachandran plot one residue at position 199 was observed in unfavourable region while rest of the residues in Glycine and Proline Ramachandran plot for wild CRYBA1 protein were in favourable regions (Fig. S6a). The Glycine and Proline Ramachandran plots for mutant CRYBA1 are given in Fig. S6b, and none of the residues were observed in disallowed region for both of the plots.

**USP18**

The general Ramachandran plot for wild form was generated for 330 residues, which excludes glycine (shown as triangles), proline and end residues. 92.1% residues (304) were falling in most favoured region, 20 residues in additional allowed and 4 in generously allowed regions, whereas only 2 residues (0.6%) i.e. Valine-167 and Threonine-192 were in disallowed region. When compared to general Ramachandran plot for mutant USP18 protein, a little reduction in residues lying in most favoured region was seen, which decreased to 91% (304 residues). Number of residues lying in additional allowed and generously allowed regions were 21 and 7, respectively. Serine-371 and Isoleucine-295 were observed in disallowed region. (Fig. 9 and Table 5). Ramachandran plots for Glycine and Proline residues revealed Glycine-10 and Proline-69 in disallowed region for both wild and mutant proteins. Additionally, Proline-191 was also observed in disallowed region in wild variant of predicted USP18 (Fig. S7).

### Protein stability and structure alignment

For the variations N59S in CRYBA1 and V125M in USP18 protein, we obtained DDG value of -1.73 and -1.19, with a reliability index of 9 and 8, respectively at temperature 25 °C and neutral pH, which indicates the decrease in stability of both
the proteins due to mutation. The root-mean-square deviation (RMSD) of atomic positions is the measure of the average distance between the backbone atoms of superimposed proteins. RaptorX structure alignment server gave RMSD value (Å) of 0.69 and 2.17 and template modeling score (TM-score) of 0.88 and 0.625 for the variations N59S in CRYBA1 and V125M in USP18, respectively. Thus, different variants of both the proteins share similar kind of folds (TM-score > 0.6) but wild types are relatively more stable.

### Table 5 Distribution of amino-acid residues in different regions of general Ramachandran plots for CRYBA1 and USP18 protein

| Item | CRYBA1 | USP18 |
|------|--------|-------|
| Region of Ramachandran plots | Wild variant | Mutant variant | Wild variant | Mutant variant |
| Residues in most favoured regions A,B,L | 170 (91.4%) | 167 (89.80%) | 304 (92.1%) | 304 (91%) |
| Residues in additional allowed regions a,b,l,p | 11 (5.9%) | 16 (8.60%) | 20 (6.1%) | 21 (6.3%) |
| Residues in generously allowed regions ~ a,~ b,~ l,~ p | 3 (1.6%) | 2 (1.1%) | 4 (1.2%) | 7 (2.1%) |
| Residues in disallowed regions | 2 (1.1%) | 1 (0.5%) | 2 (0.6%) | 2 (0.6%) |
| Total Number of non-glycine and non-proline residues | 186 (100%) | 186 (100%) | 330 (100%) | 334 (100%) |

### Binding site and ligand prediction in wild and mutant variants

#### CRYBA1

RaptorX Binding predicted 5 pockets for each variant of CRYBA1 protein. The first and the second predicted pockets were same for both the variants. The ranking of third pocket of wild variant reduced to fourth in mutant variant.
Fig. 9 The general Ramachandran plot for wild form of USP18 was generated for 330 residues, which excludes glycine (shown as triangles), proline and end residues. 92.1% residues (304) were falling in most favoured region, 20 residues in additional allowed and 4 in generously allowed regions, whereas only 2 residues (0.6%) i.e. Valine-167 and Threonine-192 were in disallowed region. When compared to general Ramachandran plot for mutant USP18 protein, a little reduction in residues lying in most favoured region was seen, which decreased to 91% (304 residues). Number of residues lying in additional allowed and generously allowed regions were 21 and 7, respectively. Serine-371 and Isoleucine-295 were observed in disallowed region.

Table 6 Different pockets predicted on CRYBA1 and USP18 variants by RaptorX Binding

| Pocket | Multiplicity | Top ligand | Binding residues | Pocket | Multiplicity | Top ligand | Binding residues |
|--------|--------------|------------|------------------|--------|--------------|------------|------------------|
| CRYBA1 |
| DOM1   |
| 1      | 4            | ACT        | D37 K44 D61      | 1      | 4            | ACT        | D37 K44 D61      |
| 2      | 4            | ACT        | P30 W31 G71      | 2      | 4            | ACT        | P30 W31 G71      |
| 3      | 3            | SO4        | Q164 C165 G166 C185 | 3      | 4            | ACT        | R114 I182 E184 Y192 |
| 4      | 2            | SO4        | R114             | 4      | 3            | SO4        | Q164 C165 G166 C185 |
| 5      | 2            | ACT        | I182 E184 H187 Y192 | 5      | 2            | SO4        | H78 T79 M111     |
| USP18  |
| DOM1   |
| 1      | 16           | ZN         | C212 C215 C264 C267 | 1      | 16           | ZN         | C212 C215 C264 C267 |
| 2      | 6            | NEH        | A170 G351 H352   | 2      | 6            | NEH        | P169 A170 G351 H352 |
| DOM2   |
| 1      | 6            | NEH        | Q11 T12 C13 Q86  | 1      | 6            | NEH        | Q11 T12 C13 Q86  |
| 2      | 2            | SO4        | G27 K30          | 2      | 2            | SO4        | G27 K30          |
| 3      | 2            | SO4        | H7 K65 A66       | –      | –            | –          | –               |

ZN Zinc (+2) cation, NEH Ethanolamine, SO4 Sulphate ion, ACT Acetate ion
This vacancy of the mutant variant was occupied by a new pocket, comprised of binding residues observed in 4th and 5th pocket of wild variant. The 5th pocket of mutant variant was completely new having binding residues not observed in any of the pocket of wild variant (Table 6). The predicted pockets can be visualized in Fig. S8.

**USP18**

Two identical pockets were predicted for domain1 in both the variants. However, the only difference was an additional binding residue, “P169” in the 2nd pocket of mutant variant. For domain 2, two pockets for wild and mutant variants were identical and third pocket was predicted only for wild variant (Table 6). The positions of various pockets on the predicted protein structures can be visualized in Fig. S9.

**Discussion**

The present study investigated the genome-wide association of environmental variables with SNP markers, as well as their impact on protein structure and stability. The admixture analysis revealed the differentiation of temperate cattle (Ladakhi and Siri) from the tropical cattle (Kankrej and Hallikar). Earlier, different livestock species including sheep (Yang et al. 2016), goat (Song et al. 2016) and cattle (Edea et al. 2013; Stucki et al. 2017) had been grouped into different clusters indicating their distinct altitude or local adaptation. SamBada’s bivariate analysis led to a considerably lower FPR than the other methods (ARLEQUIN, LFMM, BAYENV) (Stucki et al. 2017). This indicates that including population structure as a set of covariates improves SamBada’s ability to distinguish between signals of selection and differences in allelic frequencies due to isolation by distance/local adaptation.

Monthly bioclimatic variables were used in our analysis because several management and production systems like calving interval /season of calving are based on monthly variations. The association was carried out using 11 independent predictor variables ($r^2$ < 0.80). Various workers have used different number of environmental variables (2 to 118) in diverse livestock and arthropod species (Stucki et al. 2017; Joost et al. 2007; Duruz et al. 2019; Vajana et al. 2018; Eimanifar et al. 2018) for the identification of signatures of local adaptation using SamBada software. Based on G-score’s q-value, 1,12,780 models were identified as significant, with 30,350 unique SNPs or 5.28% of the total (5,74,382 SNPs) SNPs used in the analysis. This is in consonance with the number of significant SNPs (5.9%) obtained earlier (Stucki et al. 2017) in landscape genomic study on Uganda cattle. In another study on Iberian honeybees, SamBada processed a total of 38,683,470 univariate models, of which 1305 SNPs were found to be significant (FDR < 0.05).

The present study revealed a number of important pathways including Hypoxia response via HIF activation. Several studies have identified the HIF pathway as a natural selection target in Tibetan adaptation to high altitude (Cheviron and Brumfield 2012). Angiogenesis is another promising candidate, as cold exposure increases capillary density (Egginton 2002) and angiogenic factors such as VEGFA, VEGFR2, HIF-1a, PAI1, PEDF (Luo et al. 2017). VEGF signaling pathway is a key regulator of angiogenesis in response to tissue hypoxia and plays an important role in vascular vasodilation (Li et al. 2013). Angiotensin II-stimulated signaling through G proteins and beta- arrestin, as well as the endothelin signaling pathway, are linked to cutaneous vasoconstriction, which is crucial for reducing convective heat loss and preserving body temperature during cold exposure. The pathway for cholesterol biosynthesis was also identified. Short term cold exposure (Leppaluoto et al. 1988), cold seasons (Robinson et al. 1992), and long-term migration to cold regions (Gichev 1990) have been linked to an increase in serum cholesterol levels. Genetic selection resulting in increased serum cholesterol level has been reported in Finland inhabitants (Kontula 1990). In addition to light exposure, the daily rise and fall in ambient temperature may be an important input to the circadian clock system (Van Someren 2003). Thermal environment is a key determinant of sleep (Gilbert et al. 2004). Further, it has been demonstrated (Okamoto-Mizuno and Mizuno 2012) that the impact of cold exposure on circadian rhythm may be greater than that of heat. Antioxidant upregulation is crucial for responding to hypoxia-mediated oxidative stress at high altitudes (Vij et al. 2005). Thus temperate type breeds (Siri and Ladakhi) must have been thriving under these cold and high altitude environments due to the upregulation of the relevant genes/pathways under such conditions.

The present research identified 14 important genes in the TLR signaling pathway, which is involved in the heat stress response (Eicher et al. 2004; Zhou et al. 2005). Insulin/IGF pathway-protein kinase B signaling cascade and Insulin/IGF pathway-mitogen activated protein kinase/MAP kinase cascade have also been identified. Heat stress or hyperthermia causes hyperinsulinemia, resulting in a metabolic profile close to that of an immune-stimulated system in heat stressed bovine. This hyperinsulinemia is associated with lipopolysaccharide (LPS) and Hsp60 (Asea 2008). The p38 MAPK pathway identified in this study is triggered in response to several environmental stressors and promotes inhibition of cell growth, apoptosis (Wada and Penninger 2004) and heat-induced sperm damage (Rahman et al. 2014) is also crucial for the tropical type breeds (Hallikar and Kankrej) thriving under stressful conditions.
Present investigation also identified some important hypoxia related genes that play a direct role in cold adaptation. Two members of EGLN family (EGLN2 and EGLN3) have been identified, each of which catalyses oxygen-dependent hydroxylation of HIF-1α (Bruick 2001). EPAS1 (HIF2A) identified in this study is probably the most prominent hypoxia-related gene that shows signatures of adaptation in multiple species (Friedrich and Wiener 2020). Regulatory genes are ‘hot spots’ of convergent selection since they coordinate the expression of downstream target genes (Martin and Orgogozo 2013). As EPAS1 is a key transcription factor in the HIF pathway, it may be more likely to be targeted by selection. Gene HSPBAP1 (HSPB1 Associated Protein 1) identified in this study has been shown to express more in goat peripheral blood mononuclear cells in response to short-term cold stress (Mohanarao et al. 2014). The present investigation also identified SLC2A1 gene. The expression of this gene was reported to be higher in peripheral blood mononuclear cells of Ladakhi cattle, as compared to low-altitude cattle (Verma et al. 2018). Genes involved in metabolism (CA2, MIO188B) and heat stress (CDK1) were found to be common in a genetic comparison of low and high-altitude Ethiopian cattle populations (Edea et al. 2013). ICAM1, which is related to the cardiovascular system, was also identified in the goat breeds from the Tibetan highlands (Song et al. 2016). A nutrition related pathway gene CAMK2 was found to be common with selection signals in yak, indicating the importance of nutrition assimilation in adaptation to high altitude (Qiu et al. 2012). Several candidate genes for environmental adaptation and acclimation were identified in the present study. Genes identified for cold adaptation (AGTRAP, COBL, KCNMA1, PLA2G4, SLC8A1, PKLR, TCF7L2), coat colour (KIT, KITLG, EDN3), milk & growth (GHR, NCAPG, LCORL, LAP3, ABCG2), light stress (CERKL), acclimation (AQP5, RGS7), reproduction (ANXA10, BCL2), adipose tissue (TNKS, ARRDC3), growth (HMGA2, XKR4), milk production (KLHL1, PCCA), meat quality (IGFBP3, NRAP, PC, KAT2B, SLC8A1), disease resistance (PFKM and SIRPA) are in agreement with studies on Russian cattle (Yurchenko et al. 2018; Weldenegoudguad et al. 2019). Genes AMPD2, PLPP3, SP4, RFX4, LEF1, SLIT2, IGFBP7, STK32B, ADRA1D, UBE2E3, CLPB, ADAMTS16 and LAMA1 identified by R.SamBada were also showing strong signatures for environmental adaptation in Chinese cattle, with one group consisting of cold adapted Tibetan cattle (Gao et al. 2017). In the tropical composition breed Brangus, RFX4 has been found to influence heifer fertility (Fortes et al. 2012). Coat colour gene LEF1 also overlaps with pigmentation QTL regions underlying UV-protection which is more intense in high altitude and snow covered terrains as compared to lowlands. SOD1, a candidate gene for thermoregulation (B. indicus and African cattle), KIT, MITF, PDGFRA genes for coat color (Ankole) and EPB42, an Anaemia related gene (N’Dama) were reported (Kim et al. 2017). Many genes related to meat and growth traits (NCAPG, IGFBP2, IGFBP5, MYH9 and R3HDM1), and coat colour (MAP2K1) were found to be common with study conducted on Chinese cattle (Mei et al. 2018). GRIA4, GRM7, LMCD1 and AASDHPPT have been found under selection in Siberian cattle (Igoshin et al. 2019). GRIA4 encodes for the glutamate ionotropic receptor AMPA type subunit 4, which mediates excitatory synaptic transmission (Platt 2007). Activation of AMPA receptors in the medial preoptic region of the hypothalamus leads to a rise in body temperature in rats (Sengupta et al. 2016), indicating that GRIA4 expression could be involved in the thermoregulation response to acute cold stress in cattle as well. RHOD (Ras Homolog Family Member A) gene which encodes for a GTPase protein, is responsible for the Ca2+ sensitization of the contractile proteins that underlies the tonic component of vascular smooth muscle contraction and plays an important role in vasoconstriction mediated hypertension (Carbone et al. 2015). Interestingly, 112 genes found to be involved in cold tolerance in this study were previously described in mammals from the Arctic or Antarctic (Yudin et al. 2017) (Table S12).

The present study identified many chaperones, including HSPA4 (Heat Shock Protein Family A (Hsp70) Member 4), TP53INP1 etc. Heat shock factors (HSFs) and factors involved in protein folding are activated as a result of heat stress. Heat is a proteotoxic stress that causes denatured proteins to aggregate and become cytotoxic (Fink 1999; Liu et al. 2013). Many HSFs serve as molecular chaperones and assist in protein folding thus prevent protein aggregation during the cellular response to heat stress (Lindquist and Craig 1988; Hightower 1991; Moseley 1997), resulting in protein homeostasis.

Only draft breeds (Siri, Ladakhi, and Hallikar) had ‘Allele G’ for rs110287779 (found in CRYBA1), and it was highest in the Hallikar, which is considered the best draft breed in South India. While in Kankrej, a well-known dual-purpose breed, the minor allele was completely absent. As a consequence, the allelic richness of the ‘G allele’ might be related to draughtability. Table S13 shows the monthly average UV index/altitude for four places of sampling location, with the highest UV index in Bangalore (Source: IMD; obtained from www.weather-ind.com). Leh and Gangtok are at a high elevation. Many environmental intricacies are observed in Leh due to its extreme high altitude, such as low oxygen and humidity, as well as more than three times the exposure to ultraviolet (UV) light as compared to plains (Singh et al. 2013). A thinner atmosphere filters less UV radiation at higher altitudes, and UV levels rise 10% to 12% for every 1000 m gained in altitude. Sikkim and Leh experience heavy winter snowfall, which can reflect as much as 80% of UV radiation (Singh 2004). UV is a major contributing factor...
in causing human environmental dermatosis, which directs towards the harmful effects of UV in Ladakh region (Singh et al. 2013). It is thought that in cases of ocular exposure to radiation that causes keratitis and cataracts, the eyes are oriented towards the surface and that the ocular exposure is due to reflected radiation (Ambach et al. 1993). Several mutations in crystallin genes (including CRYBA1) have been linked to different types of pediatric cataract in humans (Devi et al. 2008). The SNP mutation in the CRYBA1 gene has been found in animals from areas with high biologically active UV radiation. This gene has also been linked to cataract in humans and is responsible for preserving the transparency of the eye lens. Therefore, the accumulation of this deleterious SNP mutation (G allele at higher frequency of 27.1%) in Siri, Ladakhi and Hallikar breeds might be due to more biologically active UV radiations in the regions harbouring these breeds.

Ubiquitin-specific peptidase 18 (USP18) is a negative regulator of type I and type III interferon signalling (François-Newton et al. 2011). Multiple immunological and biological functions of USP18 in cell and organ development, infection, autoimmunity and tumour immunology have been observed. In mammalian cells, many proteins are modified by ubiquitination, which at the immune level, is essential for antigen presentation. The USP18 gene is rapidly and strongly upregulated after viral infection or by type I and type III IFNs, lipopolysaccharide (LPS) (MacParland et al. 2016), tumour necrosis factor alpha (TNF-α) (MacParland et al. 2002), or genotoxic stress (Yang et al. 2015). The higher expression of this gene reduces the responsiveness to IFN-I, allowing virus replication to be restricted locally. This replication is necessary for the adaptive immune system to be triggered (Honke et al. 2016), eliciting a strong immune response to prevent the fatal outcome of infection. Down-regulation of USP18 expression can boost antiviral signaling of IFN-I. Several studies have shown that a drop in ambient temperature is linked to the incidence of influenza infection in humans and livestock species (Gunning et al. 1999; Su et al. 2017). The absence of USP18 is associated with prolonged activation of JAK-STAT signaling. This further activates the transcription of antiviral genes (Malakhov et al. 2002). The present analysis also revealed the JAK/STAT signaling pathway (P00038) (Table S9). Various mutations have been identified (Honke et al. 2016) which abolishes the isopeptidase activity of this protein. The more stable wild form of USP18 protein was almost fixed in Ladakhi cattle (G allele frequency 95.5%), thriving in high-altitude cold and hypoxic environment. The deleterious mutation was observed to be most widespread and prevalent in Siri cattle (A allele frequency 72.2%). The structure analysis also revealed that the animals of Siri breed are more admixed in comparison to other breeds. This might have impacted the Minor allele frequency of USP18 gene in this breed.

The accumulation of deleterious mutation of USP18 gene in this breed might have influenced the adaptive/innate immunity. Moreover, many common viral infections of cattle have adversely affected the fertility of dairy cow including the rise in age at first calving (Wathes et al. 2020).

In conclusion, temperate cattle (Ladakhi and Siri) are a better fit to utilize 777 K genotyping array as they show high polymorphism for HD chip. PCA-based clustering and model-based structuring have placed the four breeds into two distinct clusters i.e. Siri-Ladakhi (cold adapted) and Hallikar–Kankrej (hot adapted). The detection of selection signatures in indigenous cattle for adaptation to high altitude and hot environment has provided useful insights into the specific and general adaptation mechanisms. Many important pathways and genes have been identified which are directly or indirectly involved in cold and hot adaptation of the animals. The deleterious mutations identified in our study resulted in major structural differences between wild and mutant CRYBA1 and USB18 protein variants, with the wild type being more stable than the mutant for both proteins. To better understand the biological basis of high-altitude adaptation, the candidate gene variants found in the present study can be tested in in-vivo and in-vitro experiments. Furthermore, when developing breeding plans, the adaptive alleles discovered in native breeds can be combined with the high productivity of developed commercial breeds.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s13205-023-03493-3.

Acknowledgements The Authors are thankful to the Director, ICAR-NBAGR, Karnal, Incharges of various livestock farms, Veterinary Officers of State Animal Husbandry Department and Dr. Vijay K. Bharti (DRDO-Leh) for providing the necessary facilities/blood sample collection for the research work. The authors acknowledge the funding support received from the ICAR-National Bureau of Animal Genetic Resources, Karnal for the institute project (project code 7.70).

Author contributions SB: investigation, methodology, software, SS: investigation, supervision, conceptualization, data curation, draft preparation, resources. IG: visualization, investigation, supervision, editing, AKB: script writing, software and SPD: conceptualization, reviewing and editing.

Data availability The data used in the study have been uploaded on ICAR-Krishi portal and is in public domain with the URL http://krishi.icar.gov.in/jspsui/handle/123456789/31167 and http://krishi.icar.gov.in/jspsui/handle/123456789/69832.

Declarations Conflict of interest The authors declare no conflict of interest for the research paper entitled “Deciphering local adaptation of native Indian cattle (Bos indicus) breeds using landscape genomics and in-silico prediction of deleterious SNP effects on protein structure and function”.

Ethics statement The animal study was reviewed and approved by ICAR-NBAGR, Institute Animal Ethical Committee.
ICAR (2023) National Bureau of Animal Genetic Resources. Registered breeds of cattle. https://nbagr.icar.gov.in/en/registered-cattle/

Igoshin AV, Yurchenko AA, Belonogova NM, Petrovsky DV, Aitnazarov RB, Soloshenko VA, Yudin NS, Larkin DM (2019) Genome-wide association study and scan for signatures of selection point to candidate genes for body temperature maintenance under the cold stress in Siberian cattle populations. BMC Genet 20(51):26. https://doi.org/10.1186/s12863-019-0725-0

Joost S, Bonin A, Bruford MW, Després 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(18):3955–3969. https://doi.org/10.1111/j.1365-294X.2007.03442.x

Kallberg M, Wang H, Wang S, Peng J, Wang Z, Lu H, Xu J (2012) Template-based protein structure modeling using the RaptorX web server. Nat Protoc 7(8):1511–1522. https://doi.org/10.1038/nprot.2012.085

Kapila N, Sharma A, Kishore A, Sodhi M, Tripathi PK, Mohanty AK, Mukesh M (2016) Impact of heat stress on cellular and transcriptional adaptation of mammary epithelial cells in riverine buffalo (Bubalus Bubalis). PLoS ONE 11(9):e0157237. https://doi.org/10.1371/journal.pone.0157237

Kim J, Hanotte O, Mwai OA, Dessie T, Bashir S, Diallo B, Agaba Kontula K (1990) Molecular basis of hyperlipidemias: lessons from the Li H, Guo S, Ren Y, Wang D, Yu H, Li W, Zhao X, Chang Z (2013) Lindquist S, Craig EA (1988) The heat-shock proteins. Annu Rev 33:157–177. https://doi.org/10.1146/annurev.cellbio.22.101808.003215

Luo X, Jia R, Luo X-q, Wang G, Zhang Q-l, Qiao H, Wang N, Yan J-q (2017) Cold exposure differentially stimulates angiogenesis in BAT and WAT of mice: implication in adrenergic activation. Cell Physiol Biochem 42(3):974–986. https://doi.org/10.1159/000478680

MacParland SA, Ma X-Z, Chen L, Khattar R, Cherepanov V, Selzner M, Feld JJ, Selzner N, McGilvray ID (2016) Lipopolysaccharide and tumor necrosis factor alpha inhibit interferon signaling in hepatocytes by increasing ubiquitin-like protease 18 (USP18) expression. J Virol 90(12):5549–5560. https://doi.org/10.1128/JVI.02557-15

Maibam U, Hoddia OK, Sharma PS, Singh SV, Upadhyay RC (2017) Seasonal change in oxidative stress markers in blood plasma of Thar parkar (Bos indicus) and Karan fries (Bos indicus x Bos taurus) Cattle under tropical climate. Int J Curr Microbiol App Sci 6(2):1720–1730. https://doi.org/10.20546/ijcmas.2017.602.192

Malakov MP, Malakhova OA, Kim KI, Ritchie KJ, Zhang D-E (2002) UBP43 (USP18) specifically removes ISG15 from conjugated proteins. J Biol Chem 277(12):9976–9981. https://doi.org/10.1074/jbc.M109078200

Malomane DK, Reimer C, Weigend S, Weigend A, Sharifi AR, Simiender H (2012) Efficiency of different strategies to mitigate ascertainment bias when using SNP panels in diversity studies. BMC Genomics 19(1):22. https://doi.org/10.1186/s12864-017-4416-9

Mandel S, Joost S, Epperson BK, Holderegger R, Storfer A, Rosenberg MS, Scribner KT, Bonin A, Fortin M-J (2010) Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field: adaptive genetic variation and landscape genetics. Mol Ecol 19(17):3760–3772. https://doi.org/10.1111/j.1365-294X.2010.04717.x

Martin A, Orgogozo V (2013) The loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation: perspective. Evolution. https://doi.org/10.1111/evo.12081

McLaren W, Gill L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, Flicek P, Cunningham F (2010) The ensemble variant effect predictor. Genome Biol 11(7):122. https://doi.org/10.1186/s13059-016-0974-4

Mladilka K, Dzombe EF, Muchadeyi FC (2018) Landscape genomics and pathway analysis to understand genetic adaptation of South African indigenous goat populations. Heredity 120(4):369–378. https://doi.org/10.1038/s41437-017-0044-z

Mei C, Wang H, Liao Q, Wang L, Cheng G, Wang H, Zhao C, Zhao S, Song J, Guan X, Liu GE, Li A, Wu X, Wang C, Fang X, Zhao X, Smith SB, Yang W, Tian W, Gui L, Zhang Y, Hill RA, Jiang Z, Xiong Y, Jia C, Sun X, Wang S, Yang H, Wang J, Zhu W, Zan L (2018) Genetic architecture and selection of Chinese cattle revealed by whole genome resequencing. Mol Biol Evol 35(3):688–699. https://doi.org/10.1093/molbev/msx322

Mohana Rao GJ, Mukherjee A, Banerjee D, Gohain M, Dass G, Brahma B, Datta TK, Upadhyay RC, De S (2014) HSP70 family genes and HSP27 expression in response to heat and cold stress in vitro in peripheral blood mononuclear cells of goat (Capra hircus). Small Rumin Res 116(2–3):94–99. https://doi.org/10.1016/j.smallrumres.2013.10.014

Moseley PL (1997) Heat shock proteins and heat adaptation of the whole organism. J Appl Physiol 83(5):1413–1417. https://doi.org/10.1152/jappl.1997.83.5.1413

Ng PC, Henikoff S (2003) SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Res 31(13):3812–3814. https://doi.org/10.1093/nar/gkg509

O’Donnell MS, Ignizio DA (2012) Bioclimatic predictors for supporting ecological applications in the conterminous United States. US Geological Survey Data Series 691:10

Okamoto-Mizuno K, Mizuno K (2012) Effects of thermal environment on sleep and circadian rhythm. J Physiol Anthropol 31(1):14. https://doi.org/10.1186/1880-6805-31-14

Platt SR (2007) The role of glutamate in central nervous system health and disease—A review. Vet J 173(2):278–286. https://doi.org/10.1016/j.tvjl.2005.11.007

Porras-Hurtado L, Ruiz Y, Santos C, Phillips C, Carracedo Â, Lareu MV (2013) An overview of STRUCTURE: applications, parameter settings, and supporting software. Front Genet. https://doi.org/10.1226/vj.1389/fjge.2013.00098
Singh S (2004) High-altitude clear-sky direct solar ultraviolet irradiation.

Thomas PD (2003) PANTHER: a library of protein families and subfamilies indexed by function. Genome Res 13(9):2129–2141. https://doi.org/10.1010/3717243

Rahman MB, Vandaele L, Rijsselaere T, El-Deen MS, Maes D, Shamsuddin M, Van Soom A (2014) Bovine spermatooza react to in vitro heat stress by activating the mitogen-activated protein kinase 14 signalling pathway. Reprod Fertil Dev 26(2):245. https://doi.org/10.1071/RD12198

Sengupta T, Jaryal AK, Mallick HN (2016) Effects of NMDA and non-NMDA ionotropic glutamate receptors in the medial preoptic area on body temperature in awake rats. J Therm Biol 61:1–7. https://doi.org/10.1016/j.jtherbio.2016.07.020

Singh S (2004) High-altitude clear-sky direct solar ultraviolet irradiance at Leh and Hanle in the western Himalayas: observations and model calculations. J Geophys Res 109(D19):D19201.

Singh G, Chatterjee M, Grewal R, Verma R (2012) The yak genome and adaptation to life at high altitude. Nat Genet 44(8):946–949. https://doi.org/10.1038/ng.2343

Stucki S, Orozco-terWengel P, Forster BR, Duruz S, Colli L, Masembe C, Negriini R, The NextGen C, Joost S, Ajmone-Marsan P (2018) Combining landscape genomics and ecological modelling to investigate local adaptation of indigenous Ugandan cattle to east coast fever. Front Genet 9:385. https://doi.org/10.3389/genet.2018.00385

Van Soeren EJW (2003) Thermosensitivity of the circadian timing system. Sleep Biol Rhythms 1(1):55–64. https://doi.org/10.1046/j.1446-9235.2003.00002.x

Verma P, Sharma A, Sudhi M, Thakur K, Kataria RS, Niranjan SK, Bharti VK, Kumar P, Giri A, Kalia S, Mukesh M (2018) Transcriptome analysis of circulating PBMCs to understand mechanism of high altitude adaptation in native cattle of Ladakh region. Sci Rep 8(1):7681. https://doi.org/10.1038/s41598-018-25736-7

Vij AG, Dutta R, Satrija NK (2005) Acclimatization to oxidative stress at high altitude. High Alt Med Biol 6(4):301–310. https://doi.org/10.1089/ham.2005.6.301

Wada T, Penninger JM (2004) Mitogen-activated protein kinases in apoptosis regulation. Oncogene 23(16):2838–2849. https://doi.org/10.1038/sj.ong.1027556

Wathes DC, Ougejofoor CF, Thomas C, Cheng Z (2020) Importance of viral disease in dairy cow fertility. Engineering 6(1):26–33. https://doi.org/10.1016/j.eng.2019.07.020

Weather Atlas. https://www.weather-atlas.com.

Weldenegogaud M, Popov R, Pohkarel K, Ammosov I, Ming Y, Ivanova Z, Kantenan J (2019) Whole-genome sequencing of three native cattle breeds originating from the northernmost cattle farming regions. Front Genet 9:728. https://doi.org/10.3389/genet.2018.00728

Wu D-D, Yang C-P, Wang M-S, Dong K-Z, Yan D-W, Hao Z-Q, Fan Y-L, Chu S-Z, Shen Q-S, Jiang L-P, Li Y, Zeng L, Liu H-Q, Xie H-B, Ma Y-F, Kong X-Y, Yang S-L, Dong X-X, Esmailihadeh A, Irwin DM, Xiao X, Li M, Dong Y, Wang W, Shi P, Li H-P, Ma Y-H, Gou X, Chen Y-B, Zhang Y-P (2020) Convergent genomic signatures of high-altitude adaptation among domestic mammals. Natl Sci Rev 7(6):952–963. https://doi.org/10.1093/nsr/nzw213

Xu J, Zhang Y (2010) How significant is a protein structure similarity with TM-score = 0.5? Bioinformatics 26(7):889–895

Yang Z, Xian H, Hu J, Tian S, Qin Y, Wang R-F, Cui J (2015) USP18 negatively regulates NF-κB signaling by targeting TAK1 and NEMO for deubiquitination through distinct mechanisms. Sci Rep 5(1):12738. https://doi.org/10.1038/srep12738

Yan J, Li W-R, Lv F-H, He S-G, Tian S-L, Peng W-F, Sun Y-W, Zhao Y-X, Tu X-L, Zhang M, Xie X-L, Wang Y-T, Li J-Q, Liu Y-G, Sheng Z-Q, Wang F, Liu G-J, Lu H-F, Kantenan J, Han J-L, Li M-H, Liu M-J (2016) Whole-genome sequencing of native sheep provides insights into rapid adaptations to extreme environments. Mol Biol Evol 33(10):2576–2592. https://doi.org/10.1093/molbev/msw129

Yudin NS, Larkin DM, Ignatieva EV (2017) A compendium and functional characterization of mammalian genes involved in adaptation to Arctic or Antarctic environments. BMC Genet 18(1):111. https://doi.org/10.1186/s12863-017-0580-9

Yurchenko AA, Daetwyler HD, Yudin N, Schnabel RD, Vander Jagt DM (2018) A high-performance computing toolset for relatedness analysis at scale. Genome Res 28(12):1975–1988. https://doi.org/10.1101/sgr.2018.00924

Yudin N, Schnabel RD, Vander Jagt DM, Oudega BM, Weigel KA, Evers GW, Waldis F, Barendse W, Wahlund S, Estefan A, Visscher PM, Zehetmeier S, Grosskopf S, Weigel K, Friesen H, van den Berg M, Zandstra B, Meijer J, Elzinga L, van der Burg R, van Ooijen JW, Wathes DC, Oguejiofor CF, Thomas C, Cheng Z (2020) Importance of viral disease in dairy cow fertility. Engineering 6(1):26–33. https://doi.org/10.1016/j.eng.2019.07.020

Weather Atlas. https://www.weather-atlas.com.

Weldenegogaud M, Popov R, Pohkarel K, Ammosov I, Ming Y, Ivanova Z, Kantenan J (2019) Whole-genome sequencing of three native cattle breeds originating from the northernmost cattle farming regions. Front Genet 9:728. https://doi.org/10.3389/genet.2018.00728

Wu D-D, Yang C-P, Wang M-S, Dong K-Z, Yan D-W, Hao Z-Q, Fan Y-L, Chu S-Z, Shen Q-S, Jiang L-P, Li Y, Zeng L, Liu H-Q, Xie H-B, Ma Y-F, Kong X-Y, Yang S-L, Dong X-X, Esmailhaedh A, Irwin DM, Xiao X, Li M, Dong Y, Wang W, Shi P, Li H-P, Ma Y-H, Gou X, Chen Y-B, Zhang Y-P (2020) Convergent genomic signatures of high-altitude adaptation among domestic mammals. Natl Sci Rev 7(6):952–963. https://doi.org/10.1093/nsr/nzw213

Xu J, Zhang Y (2010) How significant is a protein structure similarity with TM-score = 0.5? Bioinformatics 26(7):889–895

Yang Z, Xian H, Hu J, Tian S, Qin Y, Wang R-F, Cui J (2015) USP18 negatively regulates NF-κB signaling by targeting TAK1 and NEMO for deubiquitination through distinct mechanisms. Sci Rep 5(1):12738. https://doi.org/10.1038/srep12738

Yan J, Li W-R, Lv F-H, He S-G, Tian S-L, Peng W-F, Sun Y-W, Zhao Y-X, Tu X-L, Zhang M, Xie X-L, Wang Y-T, Li J-Q, Liu Y-G, Sheng Z-Q, Wang F, Liu G-J, Lu H-F, Kantenan J, Han J-L, Li M-H, Liu M-J (2016) Whole-genome sequencing of native sheep provides insights into rapid adaptations to extreme environments. Mol Biol Evol 33(10):2576–2592. https://doi.org/10.1093/molbev/msw129

Yudin NS, Larkin DM, Ignatieva EV (2017) A compendium and functional characterization of mammalian genes involved in adaptation to Arctic or Antarctic environments. BMC Genet 18(1):111. https://doi.org/10.1186/s12863-017-0580-9

Yurchenko AA, Daetwyler HD, Yudin N, Schnabel RD, Vander Jagt DM, Oudega BM, Weigel KA, Evers GW, Waldis F, Barendse W, Wahlund S, Estefan A, Visscher PM, Zehetmeier S, Grosskopf S, Weigel K, Friesen H, van den Berg M, Zandstra B, Meijer J, van Ooijen JW, Wathes DC, Oguejiofor CF, Thomas C, Cheng Z (2020) Importance of viral disease in dairy cow fertility. Engineering 6(1):26–33. https://doi.org/10.1016/j.eng.2019.07.020

Weather Atlas. https://www.weather-atlas.com.
Zhou J, An H, Xu H, Liu S, Cao X (2005) Heat shock up-regulates expression of Toll-like receptor-2 and Toll-like receptor-4 in human monocytes via p38 kinase signal pathway. Immunology 114(4):522–530. https://doi.org/10.1111/j.1365-2567.2004.02112.x