

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 Bhardwaj

ICAR-National Dairy Research Institute, Karnal, Haryana-132001

Sanjeev Singh (✉ sssanjeev197@gmail.com)

ICAR-National Bureau of Animal Genetic Resources, Karnal, Haryana-132001

Indrajit Ganguly

ICAR-National Bureau of Animal Genetic Resources, Karnal, Haryana-132001

Avnish Kumar Bhatia

ICAR-National Bureau of Animal Genetic Resources, Karnal, Haryana-132001

S. P. Dixit

ICAR-National Bureau of Animal Genetic Resources, Karnal, Haryana-132001

Research Article

Keywords: Landscape genomics, Bos indicus, Adaptation, SNP Chip

Posted Date: June 9th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-533895/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Present study aimed to explore genomic basis of adaptation of Indian native cattle and to predict the impact of key SNPs on amino acid changes that affect protein function. Four native cattle breeds belonging to contrasting landscape and climatic conditions were genotyped using Illumina 777 K BovineHD BeadChip: Siri & Ladakhi from cold hilly areas, and Kankrej and Hallikar from hot arid and semi-arid regions, respectively. The R.SamBada package in R was used to perform the genotype-environment association analysis. A total of 1,12,780 significant ($q < 0.05$), models with 30,350 unique SNPs were obtained. Significantly associated SNPs had impact on 4,435 genes and 141 pathways. 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. Identified adaptive alleles might be responsible for the local adaptation of these cattle breeds.

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. Despite the fact that many of these breeds do not outperform internationally recognized breeds for economically important traits, they are well suited to local environmental conditions. 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 lineages¹ and genome-wide Runs of Homozygosity².

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 ecology³. Landscape genomics-based environmental association studies in livestock species are important for understanding the genetic basis⁴ 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 genes⁵ and their comparative expression^{6,7,8}. Impact of heat stress on cellular and transcriptional adaptation of mammary epithelial cells in riverine buffalo has also been investigated⁹. The genomic variants associated with local adaptation have been identified in Holstein Friesian, Ugandan, Russian and Siberian cattle^{10,11,12,13}, as well as South African¹⁴ and worldwide goat populations¹⁵. Similar work has also been reported in honeybee populations in the Iberian Peninsula and Africa^{16,17}. Under climate change scenarios, these identified markers may be useful in marker-assisted animal selection for better performance¹⁰. 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

Animal samples and SNP genotyping.

The blood samples were collected randomly in compliance with the guidelines and regulations of the Institutional Animal Ethics Committee (IAEC), National Bureau of Animal Genetics Resources (ICAR-NBAGR), Karnal. To investigate the genetic basis of local adaptation, a total of 46 samples were genotyped using 777 K BovineHD BeadChip (Illumina) from four native cattle breeds belonging to contrasting landscape and climatic conditions: Siri (9) & Ladakhi (11) from cold hilly; Kankrej (14) and Hallikar (12) from hot arid and semi-arid regions, respectively.

Preparing SNP genotype data.

The quality control (QC) was carried out by using PLINK software. The unmapped SNPs along with those located on X, Y chromosomes, and mitochondrial DNA was filtered out. SNPs with call rate $< 10\%$, minor allele frequency < 0.05 , and with significant deviation from Hardy-Weinberg equilibrium ($p < 10^{-3}$) were also removed. All individuals with $> 10\%$ missing SNPs were also excluded. The QC data-set was pruned for Linkage Disequilibrium (LD) by using PLINK software keeping $r^2 < 0.5$, scanning window of 100 Kb, and step-size of 50 SNP. The .ped file of QC data was converted to .gds format by R package 'SNPRelate' for analysis by R.SamBada.

STRUCTURE Analysis.

Pairwise F_{ST} value was estimated for each locus by Arlequin 3.5¹⁸ and the most significant 3565 SNPs (F_{ST} p-value < 0.01) were further used in STRUCTURE software¹⁹ to decipher the presence of population structure and account for individual ancestries. Unsupervised clustering analysis was performed with 100,000 MCMC and 100,000 Burnin²⁰ for $K = 1$ to 5, each with 7 iterations. The STRUCTURE results were plotted by 'pophelper' package in R.

Environmental Variables.

The 100 years 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 individuals was used to estimate 55 bio-predictor variables. Along with these the 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 <https://www.weather-ind.com> and Altitude from en.wikipedia.org. Finally, the values of the 59 environmental variables were extracted for each sampling locality.

To lower the dependency between models and spare computation time, the correlation coefficient between pair of variables was used to control for multicollinearity. The multivariate analysis with Samβada consisted of bivariate models along with their corresponding univariate and constant models. In this study, the multivariate models were used to take population structure into account. A 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 analyses so as to reduce the spurious associations. PCA was performed by using princomp function of R package. The maximum coefficient of correlation was set to 0.8 between pairs of variables by using prepareEnv() function of R.Samβada package which reduced the data set to 10 variables (Fig. 2). On this basis, 11 predictor variables (including the 'population structure' variable) were considered for Samβada multivariate analysis.

Genotype-environmental association.

Samβada was launched via R.Samβada's SambadaParallel() 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 to model the probability of observing a particular genotype of a polymorphic marker under the given environmental conditions at the sampling locations⁴. These logistic models are calibrated using a maximum likelihood procedure. In our 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).

Samβada estimated G-scores and Wald scores for the models. The multiple testing problem occurs when one considers a set of statistical inferences simultaneously which causes an increased in false positives. This was controlled by taking q-value for significance testing²¹. 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).

SNP annotation.

rsIDs of all the significantly associated SNPs were submitted to the Ensembl Variant Effect Predictor tool to assess the effect of SNP variants on genes, transcripts, and protein sequence²². The results were downloaded 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.

Detection of deleterious non-synonymous SNPs.

VEP gives the location of the variants in the genes. We identified the genes showing variants in the coding region. The genes showing the consequence of variants as non-synonymous, *i.e.* causing a change in the resulting amino acid, were extracted. VEP gives the SIFT scores for these missense variants. A SIFT score predicts whether an amino acid substitution affects protein function²³. The SIFT score ranges from 0.0 (deleterious) to 1.0 (tolerated). Variants with scores in the range of 0.0 to 0.05 are considered deleterious.

Structural Analysis of Protein.

The genes showing deleterious SNP mutations based on SIFT scores were selected for comparative structural analysis of the tertiary structure of the protein in the wild and mutant form of the genes. Both the sequences of the protein (wild and mutant) were submitted to the RaptorX online server (<http://raptorx.uchicago.edu/>) which does Domain parsing for protein prediction²⁴. The 3D structure predicted by the RaptorX server can be visualized in JMOL viewer directly on the RaptorX server, or the ".pdb" file can be downloaded and visualized by Pymol software.

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 (<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 predict 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. The server also gives a list of ligands which may bind with the predicted pocket.

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 sequence²⁵. 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). There are 90% chances that if the TM-score is >0.6 the two proteins share a similar fold whereas for <0.4, the two proteins have different folds.

Results

Quality checking of the samples for filtration of genotypic data.

Initial number of loci before QC for all the 46 animals were 7,77,962 (~ 777k). 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.

Table 1
Number of SNPs removed by various Quality check (QC) parameters.

QC parameter	SNPs removed	SNPs remaining
777K 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

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.

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_{MAX} -Maximum Mean Temperature (°C), T_{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_{MAX} (12), T_{MIN} (12) and MR (12), as well as two location coordinates) (Table S4). The correlation plot for 59 variables is shown in (Fig. S4) where degree of blueness and redness represents amount 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.

Bivariate Analysis.

With 5,74,382 SNPs passing the quality filter, a locus-specific landscape genomics analysis was performed 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 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%) of the 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) have been used to illustrate the q-values for each environmental variable, with the red line representing the threshold level of 0.05 and the blue line depicting the threshold level of 0.01. The 30,350 distinct SNPs (Table S5) were extracted for further analysis from the major models with $q < 0.05$, accounting for 5.28 percent of the total SNPs (5,74,382) used in the study. The number of genotypes associated with a variable was highest for "tmax7" (14,157) and lowest for "Bio1" (7,251). All other variables had more than 10K associations, with the exception of "Bio1". We found 3,169 distinct SNPs associated with environmental variables when we used the $q < 0.01$ threshold. When individual variables were screened for associations they showed similar trend as observed in previous threshold of $q < 0.05$ (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 in Ensembl Variant Effect Predictor (VEP) of which 30,121 SNPs were processed by VEP tool. The processed variants were lying in 5,481 overlapped genes out of which 1,046 genes had unannotated ensembl gene IDs. Thus, a total of 4,435 annotations including 95 miRNA binding sites (Table S6) were obtained. Many genes related to cold adaptation like *EPAS1*, *EGLN2*, *EGLN3*, etc. and chaperones involved in hot adaptation were identified. Majority of variants were lying in non-coding region, maximum of which were intron variants (56%), followed by intergenic variants (32%). Other consequences are depicted in Fig. 4. Variants lying in coding region were merely ~ 1% of the variants of which 70% variants were synonymous and 30% were missense 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 most common in Ladakhi (0.955), but the 'A' allele was found to be more prevalent in Siri.

Table 2
Numbers of unique SNPs underlying various SIFT prediction classes.

Feature	Deleterious	Deleterious	Tolerated	Tolerated	Total
	(High confidence)	(Low confidence)	(High confidence)	(Low confidence)	
Unique variants in characterized proteins	2	4	72	14	92
Unique variants in uncharacterized proteins	3	1	6	2	12
Total no. of unique variants	5	5	78	16	104

Table 3
List of missense deleterious variants (SIFT score < 0.05)

S. No.	Uploaded variant	Location	Gene	Symbol	Nucleotide variation	Amino-acid position	Amino-acid variation	SIFT score and Prediction
1	rs109247682	18:62924074	ENSBTAG000000050536	-	Gaa/Aaa	281	E/K	Deleterious_low_confidence (0.01)
2	rs110142915	21:34506379	ENSBTAG00000007348	STRA6	aCc/aAc	671	T/N	Deleterious_low_confidence (0.01)
3	rs136762779	3:559222	ENSBTAG00000000616	GPR161	Act/Cct	445	T/P	Deleterious_low_confidence (0)
4	rs136295412	4:31921506	ENSBTAG00000019403	MALSU1	aGt/aTt	41	S/I	Deleterious_low_confidence (0)
5	rs137354560	8:59322576	ENSBTAG00000015000	VCP	cGt/cAt	768	R/H	Deleterious_low_confidence (0)
6	rs133239723	15:49435044	ENSBTAG000000039735	-	cGc/cAc	265	R/H	Deleterious (0.03)
7	rs110320975	6:84511587	ENSBTAG000000049291	-	tTg/tCg	249	L/S	Deleterious (0.03)
8	rs43513552	7:41322999	ENSBTAG000000053872	-	Ctg/Atg	211	L/M	Deleterious (0.03)
9	rs110287779	19:20587324	ENSBTAG00000005352	CRYBA1	aGt/aAt	59	S/N	Deleterious (0.01)
10	rs136891050	5:76009296	ENSBTAG00000016661	USP18	Gtg/Atg	125	V/M	Deleterious (0)

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 genes²⁶, functional classification of the 4,435 annotations was performed by using PANTHER ver.15²⁷. 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. The 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 sequences 167–406. 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 Fig. 6 and Fig. 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 variants of both the genes 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.

Table 4
Model quality parameters for CRYBA1 and USP18 variants obtained by RaptorX software.

Model quality assessment parameter	Wild form of CRYBA1	Mutant form of CRYBA1	Wild form of USP18		Mutant form of USP18	
	Domain1	Domain1	Domain1	Domain2	Domain1	Domain2
Score	147	145	175	116	172	117
p-value	2.83E-10	3.89E-10	1.56E-13	2.26E-04	1.19E-13	1.82E-04
uGDT (GDT)	166 (77)	167 (78)	176 (73)	102 (96)	177 (73)	105 (97)
Residues modelled	215(100%)	215(100%)	346 (85%)		351 (86%)	

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 Asparagine-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.

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

Item	CRYBA1		USP18	
	Wild variant	Mutant variant	Wild variant	Mutant variant
Region of Ramachandran plots				
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%)

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.

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. 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.

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

CRYBA1							
WILD				MUTANT			
DOM1				DOM1			
Pocket	Multiplicity	Top ligand	Binding residues	Pocket	Multiplicity	Top ligand	Binding residues
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				DOM1			
Pocket	Multiplicity	Top Ligand	Binding residues	Pocket	Multiplicity	Top ligand	Binding residues
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				DOM2			
Pocket	Multiplicity	Top Ligand	Binding residues	Pocket	Multiplicity	Top ligand	Binding residues
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	-	-	-	-
Here, ZN: Zinc (+ 2) cation; NEH: Ethanamine; SO4: Sulphate ion and ACT: Acetate ion							

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 current study looked at the genome-wide association of environmental variables with SNP markers, as well as their impact on protein structure and stability. Genotype-Environment Association (GEA) was performed by R.SamBada package, using 5,74,382 QC passed SNPs and 11 predictor variables, one of which was included specifically to take care of the population structure to rule out false positive associations.

The BovineHD BeadChip covers SNPs validated from mostly *Bos taurus taurus* breeds (21), a few *Bos taurus indicus* (4), and hybrids *Bti* × *Btt* breeds (5) (<https://www.illumina.com>). The samples taken in this study are purely from *Bos indicus* cattle. This could explain why a large number of SNP loci were found to be monomorphic (MAF = 0) or had rare minor alleles (MAF ≤ 0.05). As a result, about 60–70% of SNPs were eliminated from each chromosome (Fig. S10).

The admixture analysis revealed the distinction of temperate cattle (Ladakhi and Siri) from the tropical cattle (Kankrej and Hallikar). Individuals from Tibetan sheep²⁸, Tibetan cashmere goats²⁹, and Ethiopian cattle³⁰ have also been grouped into lowland and highland populations using genetic information, indicating distinct altitude adaptation. Siri cattle had relatively more admixed than other breeds; however, the degree of admixture varied from individual to individual. Two clusters were achieved in Ugandan cattle¹¹ and Lida cattle³¹, respectively, on the basis of which a population variable was estimated, which is similar to our work. SamBada's bivariate analysis led to a considerably lower FPR than the other methods (ARLEQUIN, LFMM, BAYENV) (2–4% vs. 39–45%)¹¹. 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/population structuring, reducing the non-spurious association between genotype and environmental variable.

Monthly 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 ($r^2 < 80\%$). After removing variables with a r^2 greater than 90%, a final data collection of 16 environmental variables for Moroccan sheep (13 Bioclim variables, 2 raw WorldClim variables, and altitude) and 15 variables for Lida cattle (10 Bioclim and 5 raw WorldClim variables) were retained³¹. Likewise, a total of 15 predictor variables were

considered for SamBada's multivariate analysis¹¹. Various workers have used different number of environmental variables (2 to 118) in diverse livestock and arthropod species like Cattle, Sheep, Pine Weevil, and Honeybee^{11,4,31,32,17} for the identification of signatures of local adaptation by using SamBada software. Based on G-score's q-value, we identified 1, 12,780 models 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 results (5.9%) obtained earlier¹¹ in a similar study on Uganda cattle. The q-values based on G-scores using a significance threshold of 0.05 has also been used in a study on Moroccan sheep³¹. In another study on Iberian honeybees, SamBada processed a total of 38,683,470 univariate models, of which 1,305 SNPs were found to be important (FDR < 0.05).

The current study uncovered a number of important pathways, one of which is the Hypoxia response via HIF activation (P00030). A number of studies were reviewed to pinpoint the HIF pathway as a natural selection target in Tibetan adaptation to high altitude³³. Angiogenesis (P00005) is another promising candidate, as cold exposure increases capillary density³⁴ and angiogenic factors such as *VEGFA*, *VEGFR2*, *HIF-1 α* , *PAI1*, *PEDF*³⁵. VEGF signaling pathway (P00056) is a key regulator of angiogenesis in response to tissue hypoxia and plays an important role in vascular vasodilation³⁶. Angiotensin II-stimulated signaling through G proteins and beta-arrestin (P05911), as well as the endothelin signaling pathway (P00019), 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 (P00014) has also been identified. Short term cold exposure³⁷, cold seasons³⁸, and long-term migration to cold regions³⁹ have all been linked to an increase in serum cholesterol levels. Genetic selection resulting in increased serum cholesterol level has been reported in Finland inhabitants⁴⁰. In addition to light exposure, the daily rise and fall in ambient temperature may be an important input to the circadian clock system (P00015)⁴¹. Since thermoregulation is closely related to the mechanism that controls sleep, the thermal environment is a key determinant of sleep⁴². Further, it has been demonstrated⁴³ that the impact of cold exposure on circadian rhythm may be greater than that of heat. The oxidative stress response (P00046) is a major cold adaptive pathway because the amount of ROS produced during a limited supply of oxygen (at high altitude) exceeds the amount of oxygen available, resulting in oxidative stress. As a consequence, antioxidant upregulation is crucial for responding to hypoxia-mediated oxidative stress at high altitudes⁴⁴.

Our research identified 14 important genes in the TLR signaling pathway (P00054), which is involved in the heat stress response^{45,46}. Insulin/IGF pathway-protein kinase B signaling cascade (P00033) and Insulin/IGF pathway-mitogen activated protein kinase kinase/MAP kinase cascade (P00032) 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 believed to be associated with lipopolysaccharide (LPS) and Hsp60⁴⁷. The p38 MAPK pathway (P05918) is another important candidate for hot adaptation as MAPK signal transduction is triggered in response to several environmental stressors and promotes inhibition of cell growth and apoptosis⁴⁸. Furthermore, the MAPK signaling pathway has been linked to heat-induced sperm damage⁴⁹.

We 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-1 α ⁵⁰. *EPAS1* (HIF2A) identified in our study is probably the most prominent hypoxia-related gene that shows signatures of adaptation in multiple species⁵¹. Regulatory genes are 'hot spots' of convergent selection since they coordinate the expression of downstream target genes⁵². As *EPAS1* is a key transcription factor in the HIF pathway, it may be more likely to be targeted by selection. We discovered the gene *HSPBAP1* (HSPB1 Associated Protein 1), which has been shown to express more in goat peripheral blood mononuclear cells in response to short-term cold stress⁵³. We also identified *SLC2A1* gene in our study. The expression of this gene reported to be higher in peripheral blood mononuclear cells of Ladakhi cattle, as compared to low-altitude cattle⁵⁴. Genes involved in metabolism (*CA2*, *MYO18B*) and heat stress (*CDK1*) were found to be common in a genetic comparison of low and high altitude Ethiopian cattle populations³⁰. *ICAM1*, which is related to the cardiovascular system, was also identified in the goat breeds from the Tibetan highlands²⁹. 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 altitude⁵⁵. Several candidate genes for environmental adaptation and acclimation were identified in our 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 (*HMG2A*, *XKR4*), milk production (*KLHL1*, *PCCA*), meat quality (*IGFBP5*, *NRAP*, *PC*, *KAT2B*, *SLC8A1*), disease resistance (*PFKM* and *SIRPA*) are in agreement with studies on Russian cattle^{12,56}. Genes like *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 cattle⁵⁷. In the tropical composition breed Brangus, *RFX4* has been found to influence heifer fertility⁵⁸. 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 for thermoregulation (*B. indicus* and African cattle), *KIT*, *MITF*, *PDGFRA* genes for coat color (Ankole) and *EPB42*, an Anaemia related gene (N'Dama) were reported⁵⁹. 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 cattle⁶⁰. *GRIA4*, *GRM7*, *LMCD1* and *AASDHPT* have been found under selection in Siberian cattle¹³. *GRIA4* encodes for the glutamate ionotropic receptor AMPA type subunit 4, which mediates excitatory synaptic transmission⁶¹. Activation of AMPA receptors in the medial preoptic region of the hypothalamus leads to a rise in body temperature in rats⁶², indicating that *GRIA4* expression could be involved in the thermoregulation response to acute cold stress in cattle as well. RHOA (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 hypertension⁶³. Interestingly, 112 genes found to be involved in cold tolerance in this study were previously described in mammals from the Arctic or Antarctic⁶⁴ (Table S12).

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^{65,66}. Many HSFs serve as chaperones, assisting in protein folding and thus preventing protein aggregation during the cellular response to heat stress^{67,68,69}, resulting in protein homeostasis. We obtained many chaperones, including *HSPA4* (Heat Shock Protein Family A (Hsp70) Member 4), *TP53INP1* (Tumor Protein P53 Inducible Nuclear Protein 1), *AHSA2* (Activator of 90 kDa heat shock protein ATPase homolog 2), *TBCD* (Tubulin-specific chaperone D), *CDC37L1* (Hsp90 co-chaperone Cdc37-like 1) and *TBCC* (Tubulin-specific chaperone C).

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 draftability. The UV Index is a global standard for assessing the strength of ultraviolet (UV) radiation at a given location and time. Table S13 shows the monthly average UV index for four places, with the highest UV index in Bangalore (Source: IMD; obtained from www.weather-ind.com). The altitude of the sampling locations are Jodhpur-231m, Bangalore-920m, Gangtok-1650m and Leh-3524m (Wikipedia). We can see that 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 plains⁷⁰. A thinner atmosphere filters less UV radiation at higher altitudes, and UV levels rise 10–12% for every 1000 meters gained in altitude (WHO). Sikkim and Leh experience heavy winter snowfall, which can reflect as much as 80% of UV radiation (WHO). A rise in altitude of 1076 m from Leh to Hanle (A village in Ladakh UT) resulted in a 20–35% increase in irradiance⁷¹. The UV region covers the wavelength range 100–400 nm and is divided into three bands: UVA (315–400 nm); UVB (280–315 nm); UVC (100–280 nm). The UV radiation reaching the Earth's surface is largely composed of UVA with a small UVB component (WHO). The maximum value obtained at local noon at Hanle for UVB dose rate is 4.62 W/m², whereas it is 4.25 W/m² at Leh⁷¹. A study concluded that in Ladakh, UV is a major contributing factor in causing human environmental dermatosis, which directs towards the harmful effects of UV in Ladakh region⁷⁰. Other UV-related diseases are likely to be seen in animals living in high-altitude areas where UV radiation directly affects the individuals. Because of the low air pressure, less scattering, low aerosol content, and higher albedo at high altitudes, UV exposure is high, resulting in a substantial spike in biologically effective radiation. 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 radiation⁷². Several mutations in crystallin genes (including *CRYBA1*) have been linked to different types of pediatric cataract in humans⁷³. 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*), also known as ISG15 isopeptidase, is a negative regulator of type I and type III interferon signalling⁷⁴. Multiple immunological and biological functions of *USP18* in cell and organ development, infection, autoimmunity and tumour immunology have been observed. The upregulation and downregulation of this gene contributes in adaptive and innate immune response against viral infection, respectively. 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)⁷⁵, tumour necrosis factor alpha (TNF- α)⁷⁵, or genotoxic stress⁷⁶. The innate immune system reduces viral replication via IFN-I and is essential to prevent the virus from spreading to other organs. Simultaneously, it decreases viral antigen presentation to cells of the adaptive immune response. 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 triggered⁷⁷, eliciting a strong immune response to prevent the fatal outcome of infection. Downregulation of *USP18* expression in antigen-presenting cells 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^{78,79}. The absence of *USP18* enhances IFN-I and IFN-III signaling and is associated with prolonged Janus-activated kinase/signal transducer and activator of transcription (JAK-STAT) signaling. This further activates the transcription of antiviral genes⁸⁰. Our analysis also uncovered the JAK/STAT signaling pathway (P00038) (Table S9). Various mutations have been identified⁷⁷ within the Cys box (C61A), His box (298–315), mutation L361F and the Asn at position 314, 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 observed to be most widespread and prevalent in Siri cattle (A allele frequency 72.2%). The accumulation of deleterious mutation of *USP18* gene in this breed might have influenced the adaptive/innate immunity. The age at first calving, which is an indication of animal's health, is higher (5-5.5 years) in Siri cattle⁸¹ than other temperate type cattle (Ladakhi breed) under study. Moreover calving interval was also higher in Siri than Ladakhi (16.7 vs 11.8 months) which also indicates the higher reproductive efficiency of Ladakhi over Siri cattle. The structure analysis 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. It has also hypothesized that the artificial selection increases the frequency of deleterious mutation in the absence of effective recombination during domestication and reduce the reproductive fitness⁸². This is in consonance with our findings in Siri cattle having small population size of about 12,000. Recently, it has extensively been reviewed⁸³ that many common viral infections of cattle have

adversely affected the fertility of dairy cow including the rise in age at first calving. The authors concluded that viral infection interfere with the ISG (Interferon stimulating genes) regulatory IRF-STAT1 and STAT2 pathways to inhibit IFNT-induced ISG expression including *ISG15*, *HERC5*, *USP18*, *DDX58*, *IFIH1* and *SAMD9* (immune regulators with antiviral activity).

Conclusion

Temperate cattle (Ladakhi and Siri) are a better fit to utilize 777k 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. In order 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.

Declarations

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).

Conflict of interest

The authors declare that they do not have any conflict of interest for this research paper.

Competing interests

The author(s) declare no competing interests.

Author contributions

S.B.: Investigation, Methodology, Software, S.S. : Investigation, Supervision, Conceptualization, Data curation, draft preparation, Resources. I.G.: Visualization, Investigation. Supervision, editing, A.K.B.: Script writing, Software and S.P.D. : Conceptualization, Reviewing and Editing

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Additional information

Supplementary information is available as separate files.

References

1. Ganguly, I. *et al.* Y-chromosome genetic diversity of *Bos indicus* cattle in close proximity to the centre of domestication. *Sci. Rep.* **10**, 1–9 <https://doi.org/10.1038/s41598-020-66133-3> (2020).
2. Dixit, S. P. *et al.* Genome-Wide Runs of Homozygosity Revealed Selection Signatures in *Bos indicus*. *Front. Genet.* **11**, 92 <https://doi.org/10.3389/fgene.2020.00092> (2020).
3. Manel, S. *et al.* Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field. *Mol. Ecol.* **19**, 3760–3772 <https://doi.org/10.1111/j.1365-294X.2010.04717.x> (2010).
4. Joost, S. *et al.* A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. *Mol. Ecol.* **16**, 3955–3969 <https://doi.org/10.1111/j.1365-294X.2007.03442.x> (2007).
5. Sodhi, M. *et al.* Novel polymorphisms in UTR and coding region of inducible heat shock protein 70.1 gene in tropically adapted Indian zebu cattle (*Bos indicus*) and riverine buffalo (*Bubalus bubalis*). *Gene*. **527**, 606–615 <https://doi.org/10.1016/j.gene.2013.05.078> (2013).
6. Kumar, M. *et al.* Dietary Inorganic Chromium in Summer-Exposed Buffalo Calves (*Bubalus bubalis*): Effects on Biomarkers of Heat Stress, Immune Status, and Endocrine Variables. *Biol. Trace Elem. Res.* **167**, 18–27 <https://doi.org/10.1007/s12011-015-0272-0> (2015).

7. Deb, R. *et al.* Effect of heat stress on the expression profile of Hsp90 among Sahiwal (*Bos indicus*) and Frieswal (*Bos indicus*×*Bos taurus*) breed of cattle: A comparative study. *Gene*. **536**, 435–440 <https://doi.org/10.1016/j.gene.2013.11.086> (2014).
8. Maibam, U., Hooda, O. K., Sharma, P. S., Singh, S. V. & Upadhyay, R. C. Seasonal Change in Oxidative Stress Markers in Blood Plasma of Tharparkar (*Bos indicus*) and Karan Fries (*Bos indicus* x *Bos taurus*) Cattle under Tropical Climate. *Int. J. Curr. Microbiol. Appl. Sci.* **6**, 1720–1730 <https://doi.org/10.20546/ijcmas.2017.602.192> (2017).
9. Kapila, N. *et al.* Impact of Heat Stress on Cellular and Transcriptional Adaptation of Mammary Epithelial Cells in Riverine Buffalo (*Bubalus Bubalis*). *PLoS One*. **11**, 157237 <https://doi.org/10.1371/journal.pone.0157237> (2016).
10. Hayes, B. J. *et al.* A Validated Genome Wide Association Study to Breed Cattle Adapted to an Environment Altered by Climate Change. *PLoS One*. **4**, 6676 <https://doi.org/10.1371/journal.pone.0006676> (2009).
11. Stucki, S. *et al.* High performance computation of landscape genomic models including local indicators of spatial association. *Mol. Ecol. Resour.* **17**, 1072–1089 <https://doi.org/10.1111/1755-0998.12629> (2017).
12. Yurchenko, A. A. *et al.* Scans for signatures of selection in Russian cattle breed genomes reveal new candidate genes for environmental adaptation and acclimation. *Sci. Rep.* **8**, 12984 <https://doi.org/10.1038/s41598-018-31304-w> (2018).
13. Igoshin, A. V. *et al.* 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**, 26 <https://doi.org/10.1186/s12863-019-0725-0> (2019).
14. Mdladla, K., Dzomba, E. F. & Muchadeyi, F. C. Landscape genomics and pathway analysis to understand genetic adaptation of South African indigenous goat populations. *Heredity (Edinb.)*. **120**, 369–378 <https://doi.org/10.1038/s41437-017-0044-z> (2018).
15. Bertolini, F. *et al.* Signatures of selection and environmental adaptation across the goat genome post-domestication. *Genet. Sel. Evol.* **50**, 57 <https://doi.org/10.1186/s12711-018-0421-y> (2018).
16. Henriques, D. *et al.* Whole genome SNP-associated signatures of local adaptation in honeybees of the Iberian Peninsula. *Sci. Rep.* **8**, 11145 <https://doi.org/10.1038/s41598-018-29469-5> (2018).
17. Eimanifar, A., Brooks, S. A., Bustamante, T. J. & Ellis D. Population genomics and morphometric assignment of western honey bees (*Apis mellifera* L.) in the Republic of South Africa. *BMC Genomics*. **19**, 615 <https://doi.org/10.1186/s12864-018-4998-x> (2018).
18. Excoffier, L. & Lischer, H. E. L. 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 <https://doi.org/10.1111/j.1755-0998.2010.02847.x> (2010).
19. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics*. **155**, 945–959 (2000).
20. Porras-Hurtado, L. *et al.* An overview of STRUCTURE: applications, parameter settings, and supporting software. *Front. Genet.* **4**, <https://doi.org/10.3389/fgene.2013.00098> (2013).
21. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci.* **100**, 9440–9445 (2003). <https://doi.org/10.1073/pnas.1530509100>.
22. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 <https://doi.org/10.1186/s13059-016-0974-4> (2016).
23. Ng, P. C. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **31**, 3812–3814 <https://doi.org/10.1093/nar/gkg509> (2003).
24. Källberg, M. *et al.* Template-based protein structure modeling using the RaptorX web server. *Nat. Protoc.* **7**, 1511–1522 <https://doi.org/10.1038/nprot.2012.085> (2012).
25. Capriotti, E., Fariselli, P. & Casadio, R. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res.* **33**, 306–310 <https://doi.org/10.1093/nar/gki375> (2005).
26. Wu, D. D. *et al.* Convergent genomic signatures of high-altitude adaptation among domestic mammals. *Natl. Sci. Rev.* **7**, 952–963 <https://doi.org/10.1093/nsr/nwz213> (2020).
27. Thomas, P. D. & PANTHER A Library of Protein Families and Subfamilies Indexed by Function. *Genome Res.* **13**, 2129–2141 <https://doi.org/10.1101/gr.772403> (2003).
28. Yang, J. *et al.* Whole-Genome Sequencing of Native Sheep Provides Insights into Rapid Adaptations to Extreme Environments. *Mol. Biol. Evol.* **33**, 2576–2592 <https://doi.org/10.1093/molbev/msw129> (2016).
29. Song, S. *et al.* Exome sequencing reveals genetic differentiation due to high-altitude adaptation in the Tibetan cashmere goat (*Capra hircus*). *BMC Genomics*. **17**, 122 <https://doi.org/10.1186/s12864-016-2449-0> (2016).
30. Edea, Z. *et al.* Genetic diversity, population structure and relationships in indigenous cattle populations of Ethiopia and Korean Hanwoo breeds using SNP markers. *Front. Genet.* **4**, <https://doi.org/10.3389/fgene.2013.00035> (2013).
31. Duruz, S. *et al.* Rapid identification and interpretation of gene–environment associations using the new R.SamBada landscape genomics pipeline. *Mol. Ecol. Resour.* **19**, 1355–1365 <https://doi.org/10.1111/1755-0998.13044> (2019).
32. Vajana, E. *et al.* Ajmone-Marsan, 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/fgene.2018.00385> (2018).

33. Cheviron, Z. A. & Brumfield, R. T. Genomic insights into adaptation to high-altitude environments. *Heredity (Edinb)*. **108**, 354–361 <https://doi.org/10.1038/hdy.2011.85> (2012).
34. Egginton, S. Temperature and angiogenesis: the possible role of mechanical factors in capillary growth. *Comp. Biochem. Physiol. Part A Mol. Integr. Physiol.* **132**, 773–787 [https://doi.org/10.1016/S1095-6433\(02\)00047-8](https://doi.org/10.1016/S1095-6433(02)00047-8) (2002).
35. Luo, X. *et al.* Cold Exposure Differentially Stimulates Angiogenesis in BAT and WAT of Mice: Implication in Adrenergic Activation. *Cell. Physiol. Biochem.* **42**, 974–986 <https://doi.org/10.1159/000478680> (2017).
36. Li, H. *et al.* VEGF 189 Expression is Highly Related to Adaptation of the Plateau Pika (*Ochotona curzoniae*) Inhabiting High Altitudes. *High Alt. Med. Biol.* **14**, 395–404 <https://doi.org/10.1089/ham.2013.1013> (2013).
37. Leppäluoto, J., Korhonen, I., Huttunen, P. & Hassi, J. Serum levels of thyroid and adrenal hormones, testosterone, TSH, LH, GH and prolactin in men after a 2-h stay in a cold room. *Acta Physiol.Scand.* **132**, 543–548 <https://doi.org/10.1111/j.1748-1716.1988.tb08363.x> (1988).
38. Robinson, D., Bevan, E. A., Hinohara, S. & Takahashi, T. Seasonal variation in serum cholesterol levels – evidence from the UK and Japan. *Atherosclerosis.* **95**, 15–24 [https://doi.org/10.1016/0021-9150\(92\)90171-C](https://doi.org/10.1016/0021-9150(92)90171-C) (1992).
39. Gichev, J. P. Change of liver function during adaptation of man to northern conditions. *Arct. Med Res.* **49**, 68–73 (1990).
40. Kontula, K. Molecular Basis of Hyperlipidemias: Lessons from the Finnish Gene Inheritance. *Ann. Med.* **22**, 303–305 <https://doi.org/10.3109/07853899009147911> (1990).
41. Van Someren, E. J. Thermosensitivity of the circadian timing system. *Sleep Biol. Rhythms.* **1**, 55–64 <https://doi.org/10.1046/j.1446-9235.2003.00002.x> (2003).
42. Gilbert, S. S., van den Heuvel, C. J., Ferguson, S. A. & Dawson, D. Thermoregulation as a sleep signalling system. *Sleep Med. Rev.* **8**, 81–93 [https://doi.org/10.1016/S1087-0792\(03\)00023-6](https://doi.org/10.1016/S1087-0792(03)00023-6) (2004).
43. Okamoto-Mizuno, K. & Mizuno, K. Effects of thermal environment on sleep and circadian rhythm. *J. Physiol. Anthropol.* **31**, 14 <https://doi.org/10.1186/1880-6805-31-14> (2012).
44. Vij, A., Dutta, G. R. & Satija, N. K. Acclimatization to Oxidative Stress at High Altitude. *High Alt. Med. Biol.* **6**, 301–310 <https://doi.org/10.1089/ham.2005.6.301> (2005).
45. Eicher, S. D., McMunn, K. A., Hammon, H. M. & Donkin, S. S. Toll-like receptors 2 and 4, and acute phase cytokine gene expression in dexamethasone and growth hormone treated dairy calves. *Vet. Immunol. Immunopathol.* **98**, 115–125 <https://doi.org/10.1016/j.vetimm.2003.10.009> (2004).
46. Zhou, J., An, H., Xu, H., Liu, S. & Cao, X. 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**, 522–530 <https://doi.org/10.1111/j.1365-2567.2004.02112.x> (2005).
47. Asea, A. Heat Shock Proteins and Toll-Like Receptors. in 2008, 111–127. https://doi.org/10.1007/978-3-540-72167-3_6.
48. Wada, T. & Penninger, J. M. Mitogen-activated protein kinases in apoptosis regulation. *Oncogene.* **23**, 2838–2849 <https://doi.org/10.1038/sj.onc.1207556> (2004).
49. Rahman, M. B. *et al.* Bovine spermatozoa react to in vitro heat stress by activating the mitogen-activated protein kinase 14 signalling pathway. *Reprod. Fertil. Dev.* **26**, 245 <https://doi.org/10.1071/RD12198> (2014).
50. Bruick, R. K. A Conserved Family of Prolyl-4-Hydroxylases That Modify HIF. *Science.* **294**, 55451337–55451340 <https://doi.org/10.1126/science.1066373> (2001).
51. Friedrich, J. & Wiener, P. Selection signatures for high-altitude adaptation in ruminants. *Anim. Genet.* **51**, 157–165 <https://doi.org/10.1111/age.12900> (2020).
52. Martin, A. & Orgogozo, V. The loci of repeated evolution: A catalog of genetic hotspots of phenotypic variation. *Evolution (N. Y)*. <https://doi.org/10.1111/evo.12081> (2013).
53. Mohanarao, G. J. *et al.* 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**, 94–99 <https://doi.org/10.1016/j.smallrumres.2013.10.014> (2014).
54. Verma, P. *et al.* Transcriptome Analysis of Circulating PBMCs to Understand Mechanism of High Altitude Adaptation in Native Cattle of Ladakh Region. *Sci. Rep.* **8**, 7681 <https://doi.org/10.1038/s41598-018-25736-7> (2018).
55. Qiu, Q. *et al.* The yak genome and adaptation to life at high altitude. *Nat. Genet.* **44**, 946–949 <https://doi.org/10.1038/ng.2343> (2012).
56. Weldenogdguad, M. *et al.* Whole-Genome Sequencing of Three Native Cattle Breeds Originating From the Northernmost Cattle Farming Regions. *Front. Genet.* **9**, 728 <https://doi.org/10.3389/fgene.2018.00728> (2019).
57. Gao, Y. *et al.* Species composition and environmental adaptation of indigenous Chinese cattle. *Sci. Rep.* **7**, 16196 <https://doi.org/10.1038/s41598-017-16438-7> (2017).
58. Fortes, M. R. S. *et al.* Gene network analyses of first service conception in Brangus heifers: Use of genome and trait associations, hypothalamic-transcriptome information, and transcription factors1. *J. Anim. Sci.* **90**, 2894–2906 <https://doi.org/10.2527/jas.2011-4601> (2012).
59. Kim, J. *et al.* The genome landscape of indigenous African cattle. *Genome Biol.* **18**, 34 <https://doi.org/10.1186/s13059-017-1153-y> (2017).

60. Mei, C. *et al.* Genetic Architecture and Selection of Chinese Cattle Revealed by Whole Genome Resequencing. *Mol. Biol. Evol.* **35**, 688–699 <https://doi.org/10.1093/molbev/msx322> (2018).
61. Platt, S. R. The role of glutamate in central nervous system health and disease – A review. *Vet. J.* **173**, 278–286 <https://doi.org/10.1016/j.tvjl.2005.11.007> (2007).
62. Sengupta, T., Jaryal, A. K. & Mallick, H. N. 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> (2016).
63. Carbone, M. L. *et al.* Angiotensin II Activates the RhoA Exchange Factor Arhgef1 in Humans. *Hypertension.* **65**, 1273–1278 <https://doi.org/10.1161/HYPERTENSIONAHA.114.05065> (2015).
64. Yudin, N. S., Larkin, D. M. & Ignatieva, E. V. A compendium and functional characterization of mammalian genes involved in adaptation to Arctic or Antarctic environments. *BMC Genet.* **18**, 111 <https://doi.org/10.1186/s12863-017-0580-9> (2017).
65. Fink, A. L., Chaperone-Mediated, P. & Folding *Physiol. Rev.* **79**, 425–449 <https://doi.org/10.1152/physrev.1999.79.2.425> (1999).
66. Liu, J. *et al.* An Autoregulatory Loop Controlling Arabidopsis HsfA2 Expression: Role of Heat Shock-Induced Alternative Splicing. *Plant Physiol.* **162**, 512–521 <https://doi.org/10.1104/pp.112.205864> (2013).
67. Lindquist, S. & Craig, E. A. The Heat-Shock Proteins. *Annu. Rev. Genet.* **22**, 631–677 <https://doi.org/10.1146/annurev.ge.22.120188.003215> (1988).
68. Hightower, L. E. Heat shock, stress proteins, chaperones, and proteotoxicity. *Cell.* **66**, 191–197 [https://doi.org/10.1016/0092-8674\(91\)90611-2](https://doi.org/10.1016/0092-8674(91)90611-2) (1991).
69. Moseley, P. L. Heat shock proteins and heat adaptation of the whole organism. *J. Appl. Physiol.* **83**, 1413–1417 <https://doi.org/10.1152/jappl.1997.83.5.1413> (1997).
70. Singh, G., Chatterjee, M., Grewal, R. & Verma, R. Incidence and care of environmental dermatoses in the high-altitude region of Ladakh, India. *Indian J. Dermatol.* **58**, 107 <https://doi.org/10.4103/0019-5154.108038> (2013).
71. Singh, S. High-altitude clear-sky direct solar ultraviolet irradiance at Leh and Hanle in the western Himalayas: Observations and model calculations. *J. Geophys. Res.* **109**, 19201 <https://doi.org/10.1029/2004JD004854> (2004).
72. Ambach, W., Blumthaler, M. & Schöpf, T. Increase of biologically effective ultraviolet radiation with altitude. *J. Wilderness Med.* **4**, 189–197 <https://doi.org/10.1580/0953-9859-4.2.189> (1993).
73. Devi, R. R. *et al.* Crystallin gene mutations in Indian families with inherited pediatric cataract. *Mol. Vis.* **14**, 1157–1179 (2008).
74. François-Newton, V. *et al.* USP18-Based Negative Feedback Control Is Induced by Type I and Type III Interferons and Specifically Inactivates Interferon α Response. *PLoS One.* **6**, 22200 <https://doi.org/10.1371/journal.pone.0022200> (2011).
75. MacParland, S. A. *et al.* Lipopolysaccharide and Tumor Necrosis Factor Alpha Inhibit Interferon Signaling in Hepatocytes by Increasing Ubiquitin-Like Protease 18 (USP18) Expression. *J. Virol.* **90**, 5549–5560 <https://doi.org/10.1128/JVI.02557-15> (2016).
76. Yang, Z. *et al.* USP18 negatively regulates NF- κ B signaling by targeting TAK1 and NEMO for deubiquitination through distinct mechanisms. *Sci. Rep.* **5**, 12738 <https://doi.org/10.1038/srep12738> (2015).
77. Honke, N., Shaabani, N., Zhang, D. E., Hardt, C. & Lang, K. S. Multiple functions of USP18. *Cell Death Dis.* **7**, 2444 <https://doi.org/10.1038/cddis.2016.326> (2016).
78. Gunning, R., Brown, I. H. & Crawshaw, T. R. Evidence of influenza A virus infection in dairy cows with sporadic milk drop syndrome. *Vet. Rec.* **145**, 556–557 <https://doi.org/10.1136/vr.145.19.556> (1999).
79. Su, S., Fu, X., Li, G., Kerlin, F. & Veit, M. Novel Influenza D virus: Epidemiology, pathology, evolution and biological characteristics. *Virulence.* **8**, 1580–1591 <https://doi.org/10.1080/21505594.2017.1365216> (2017).
80. Malakhov, M. P., Malakhova, O. A., Il Kim, K., Ritchie, K. J. & Zhang, D. E. UBP43 (USP18) Specifically Removes ISG15 from Conjugated Proteins. *J. Biol. Chem.* **277**, 9976–9981 <https://doi.org/10.1074/jbc.M109078200> (2002).
81. Tantia, M. S. *et al.* The cattle of eastern himalayas. *Anim. Genet. Resour. Inf.* **19**, S191–S192 <https://doi.org/10.1017/S1014233900000778> (1996).
82. Lu, J. *et al.* The accumulation of deleterious mutations in rice genomes: a hypothesis on the cost of domestication. *Trends Genet.* **22**, 126–131 <https://doi.org/10.1016/j.tig.2006.01.004> (2006).
83. Wathes, D. C., Oguejiofor, C. F., Thomas, C. & Cheng, Z. Importance of Viral Disease in Dairy Cow Fertility. *Engineering.* **6**, 26–33 <https://doi.org/10.1016/j.eng.2019.07.020> (2020).

Figures

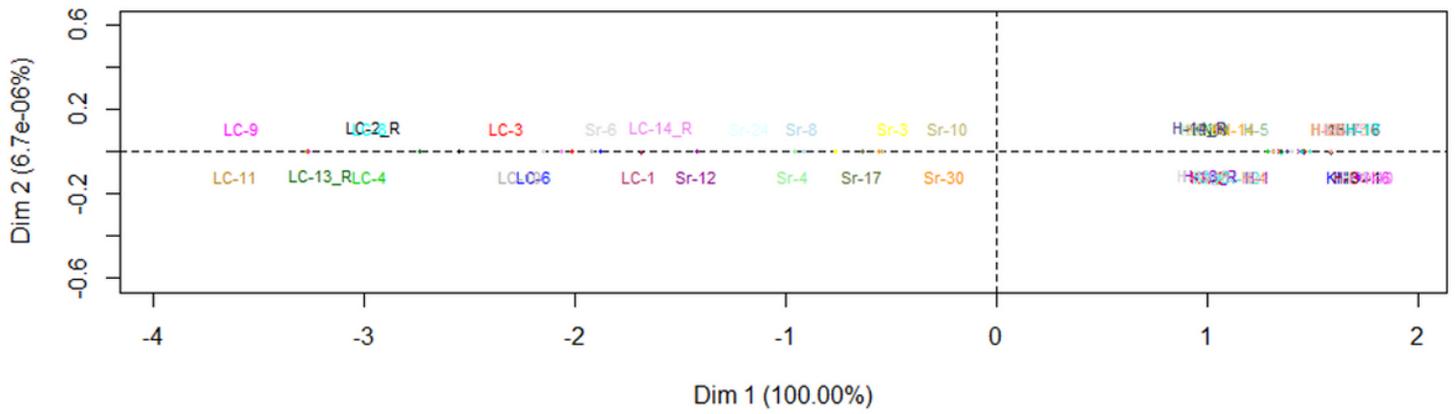


Figure 1
 Principal Component Analysis performed on Ancestry coefficients of all the 46 individuals of four breeds of cattle (Siri & Ladakhi from cold hilly on left side; Kankrej and Hallikar from semi-arid/hot humid on right side)

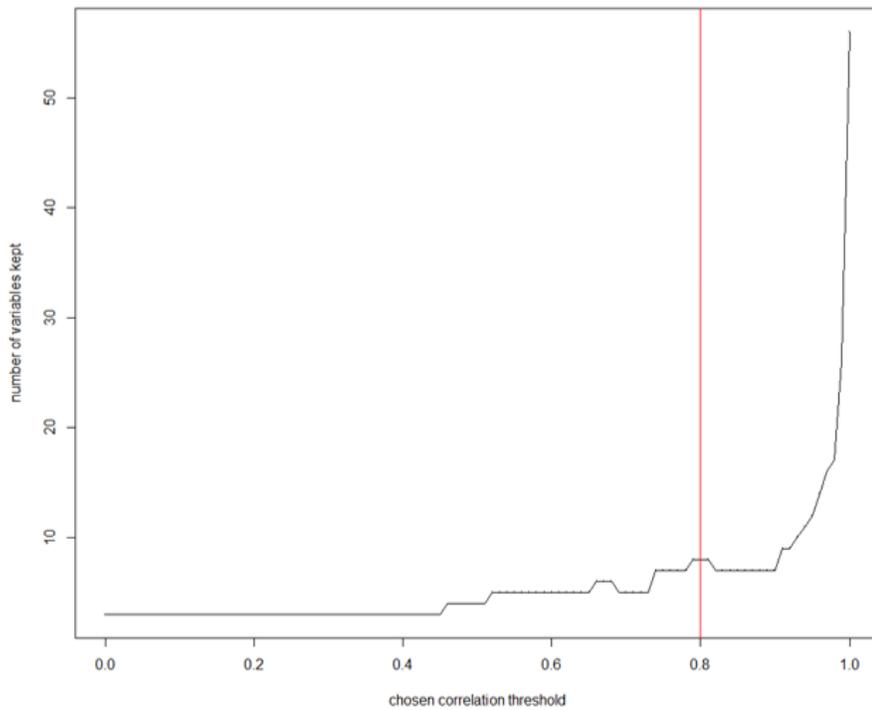


Figure 2
 Number of variables kept according to correlation threshold of 0.8.

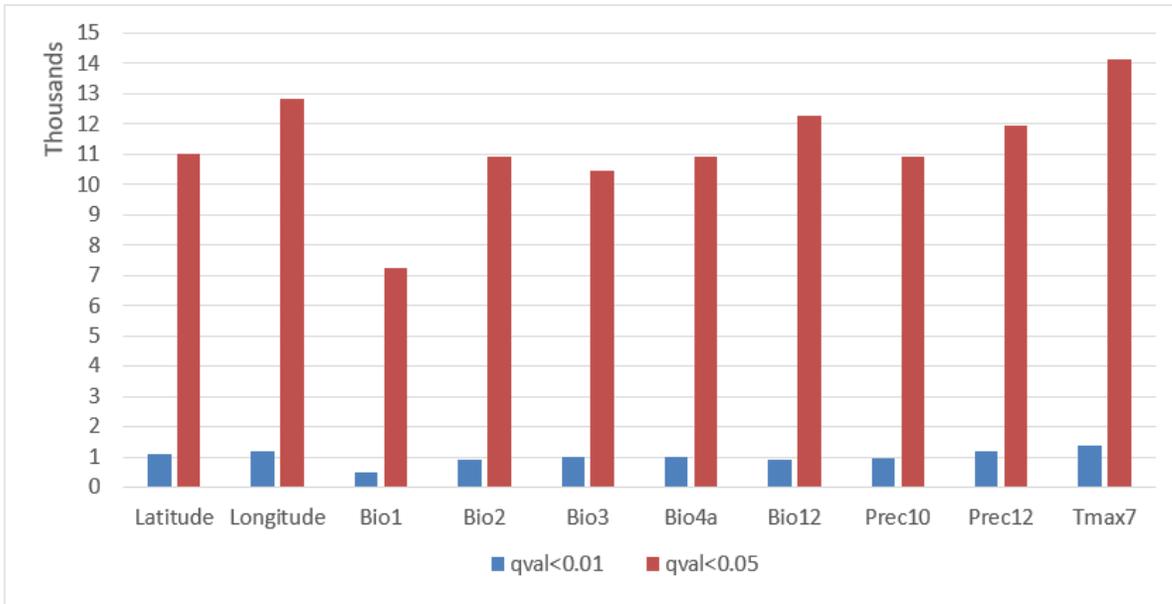


Figure 3

Number of significant models associated with each environmental variable at q-value < 0.05 and q-value < 0.01

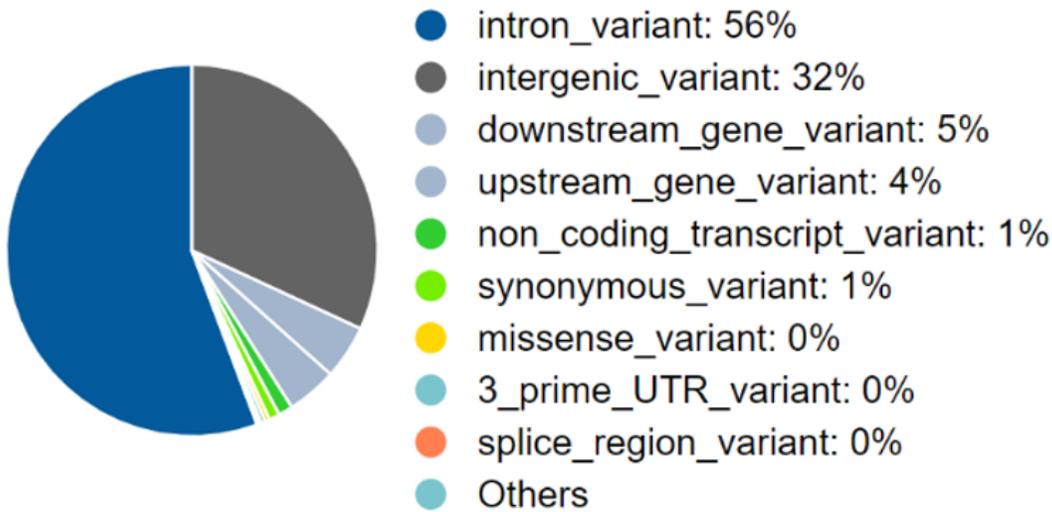


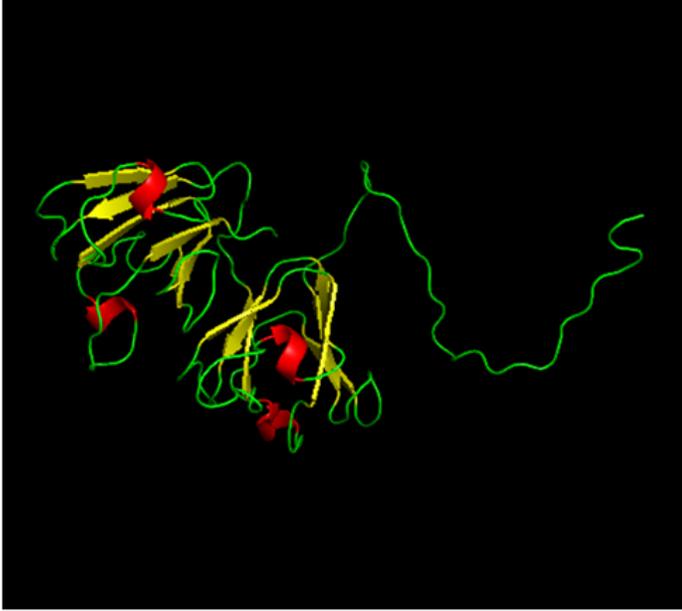
Figure 4

Consequences of the SNP variants obtained from Variant Effect Prediction tool.

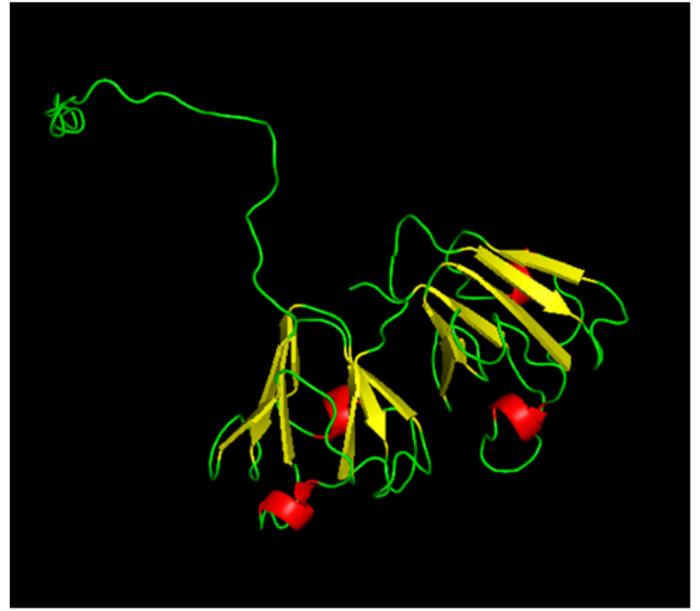


Figure 5

Consequences of variants lying in coding region.



a) Wild form of CRYBA1 protein



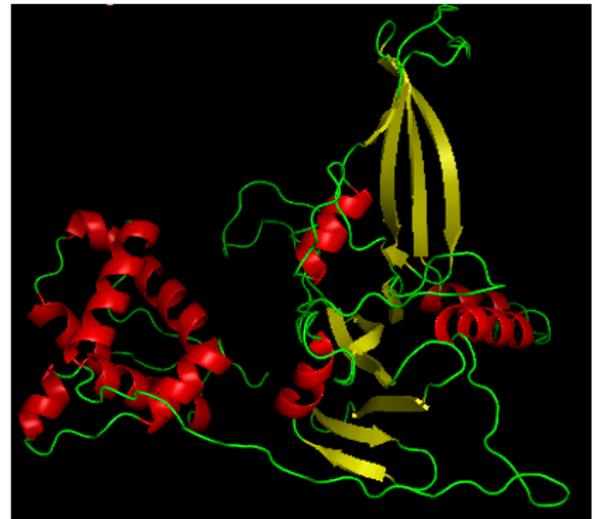
b) Mutant form of CRYBA1 protein

Figure 6

(a&b). Tertiary structure of wild and mutant form of CRYBA1 protein visualized by Pymol software (green-loop, yellow-sheet, red-helix)



a) Wild form of USP18



b) Mutant for of USP18

Figure 7

(a&b). Tertiary structure of wild and mutant form of USP18 protein visualized by Pymol software (green-loop, yellow-sheet, red-helix).

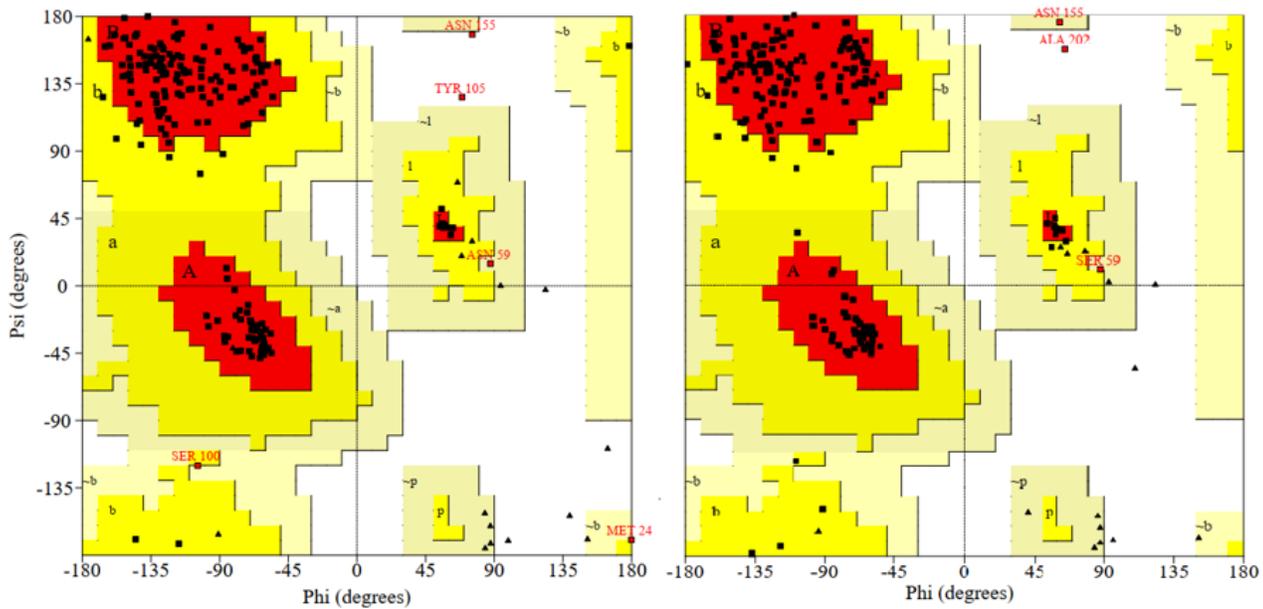


Figure 8

General Ramachandran plots for wild (left) and mutant (right) variant of CRYBA1 protein

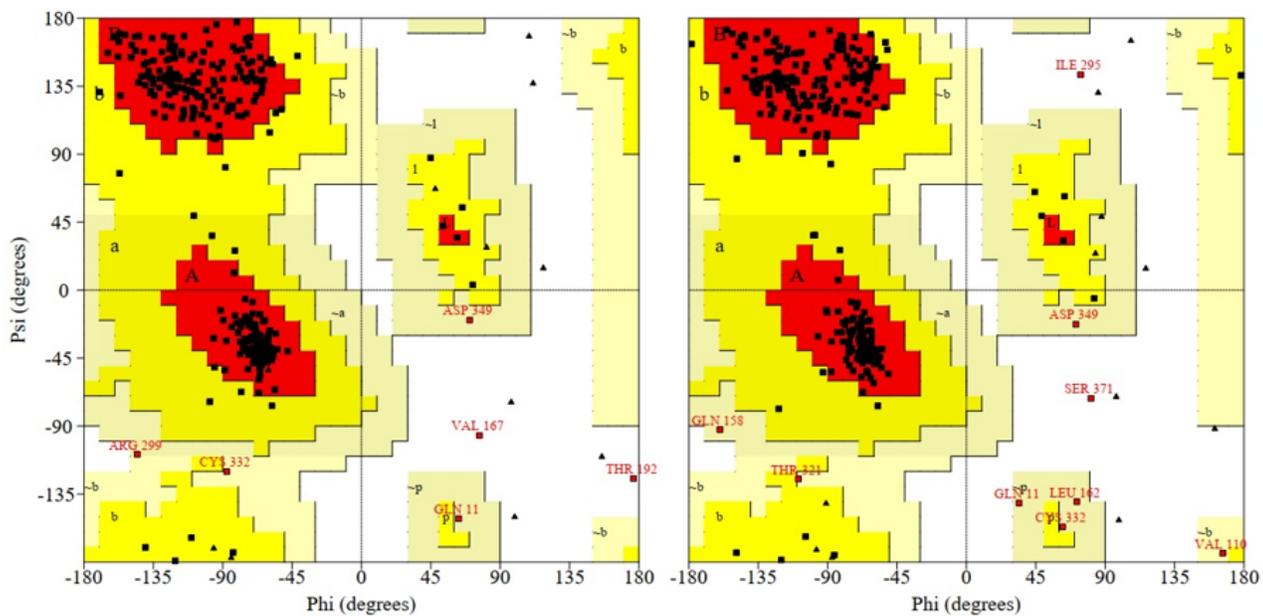


Figure 9

General Ramachandran plots for wild (left) and mutant (right) variant of USP18 protein.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Fig.S1.docx](#)
- [Fig.S2.docx](#)
- [Fig.S3.docx](#)
- [Fig.S4.docx](#)
- [Fig.S5.docx](#)
- [Fig.S6.docx](#)

- Fig.S7.docx
- Fig.S8.docx
- Fig.S9.docx
- Fig.S10.docx
- ST1.docx
- ST2.docx
- ST3.xlsx
- ST4.xlsx
- ST5.xlsx
- ST6.xlsx
- ST7.docx
- ST8.docx
- ST9.docx
- ST10.docx
- ST11.docx
- ST12.docx
- CaptionsoftheSupplementarytablesandfigures.docx
- ST13.docx