

Article

Not peer-reviewed version

Genome-Wide Identification, Annotation of SNPs, and Selective Sweeps in Indian Horse and Pony Breeds

[Anuradha Bhardwaj](#)^{*,†}, [JayaKumar Sivalingam](#)[†], Yash Pal, [Varij Nayan](#), Mamta Chauhan, Ram Avatar Legha, Sonali Soni, Saroj Panghal, Shiv Kumar Giri, [Sarika Jaiswal](#), Mir Asif Iquebal, Vijay Kumar, Dinesh Kumar, [Tarun Kumar Bhattacharya](#), Bhupendra Nath Tripathi

Posted Date: 30 May 2025

doi: 10.20944/preprints202505.2449.v1

Keywords: genomic diversity; Indian equines; ddRAD; SNPs; selection sweeps



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Genome-Wide Identification, Annotation of SNPs, and Selective Sweeps in Indian Horse and Pony Breeds

Anuradha Bhardwaj ^{1,*†}, Jayakumar Sivalingam ^{2,†}, Yash Pal ^{1,3}, Varij Nayan ^{3,4}, Mamta Chauhan ^{1,5}, Ram Avatar Legha ^{1,6}, Sonali ¹, Saroj Panghal ¹, Shiv Kumar Giri ⁷, Sarika ⁸, Mir Asif Iquebal ⁸, Vijay Kumar ^{1,9}, Dinesh Kumar ⁸, T.K. Bhattacharya ¹ and B. N. Tripathi ^{1,10}

¹ ICAR-National Research Centre on Equines, Sirsa Road, Hisar-125001, Haryana, India

² ICAR-Directorate of Poultry Research, Rajendranagar, Hyderabad-500030, Telangana, India

³ ICAR-Central Institute for Research on Buffaloes, Hisar-125001, Haryana, India

⁴ ICAR-National Dairy Research Institute, Karnal-132001, Haryana, India

⁵ ICAR-National Dairy Research Institute, SRS, Bengaluru -132001, Haryana, India

⁶ ICAR-Central Sheep and Wool Research Institute, Avikanagar-304501, Rajasthan, India

⁷ Maharaja Agrasen University, Baddi, Solan-174103, Himachal Pradesh, India

⁸ ICAR-Indian Agricultural Statistical Research Institute-110012, Himachal Pradesh, India

⁹ ICAR-Central Sheep and Wool Research Institute, Bikaner-334001, Rajasthan, India

¹⁰ Vice-Chancellor, SKUAST-Jammu, India

* Correspondence: dranu.biotech@gmail.com; Tel.: (+91-1662-276748 (Office); +91-9802249964)

† Anuradha Bhardwaj and Jayakumar Sivalingam contributed equally to this work.

Simple Summary: This study explored the genetic differences in Indian horses and ponies to uncover how their unique traits, developed in diverse environments, could boost their health and performance. Using a low-cost genetic method, researchers analyzed over 300,000 DNA variations, identifying key genes linked to critical functions like brain development, muscle strength, and tendon resilience, which enhance athleticism. They also found genes involved in heart function and calcium regulation, potentially helping horses maintain lower heart rates during intense activity, improving stamina. Over 1,600 genetic markers linked to evolutionary adaptation were discovered, with 700 tied to specific genes like SLC4A10 and CADM1. These insights can guide breeding programs to produce healthier, hardier horses suited for work, sports, or farming. By connecting genetic traits to real-world benefits, the study offers practical ways to improve horse welfare and performance globally, supporting communities that depend on these animals for livelihoods and cultural practices.

Abstract: The genomic diversity among Indian horse and pony breeds was investigated to understand how their unique traits, shaped by different environments, might help improve their health and performance. A cost-effective genetic analysis method of the genome-based ddRAD approach was used to identify admixture. The investigators identified 322732 variants, of which 298782 were SNPs and 23950 were Indels across horses and ponies. In the thoroughbred horses, 149134 genetic variants were observed, of which 161002 were SNPs. 1631 positive selection signatures were found common among all horse breeds. 700 selection signatures were enumerated to 306 genes. SLC4A10, CADM1, AUST2, and LOC733058 genes contained more than 10 selection signatures. Many of these selection signatures were linked to important functions, such as brain development (via genes GRM8 and GRIK3), muscle strength (VDR), and tendon durability (TNC)—all of which contribute to horses' athletic abilities. Genes such as ANK2, BCL2, and RYR2 possess positive selection signatures and play central roles in managing calcium movement and heart function. These genes may help horses maintain lower heart rates during intense activity, enhancing their stamina

and performance. This study offered insights to enhance horse welfare and performance globally, benefiting communities reliant on these animals.

Keywords: genomic diversity; Indian equines; ddRAD; SNPs; selection sweeps

1. Introduction

Horses have been used for sporting, hunting, and armed warfare for many generations. Therefore, certain athletic performance-related genes were positively selected in response to the demand created by the environment, resulting in their unique physical morphology and physiological system. Identifying the genetic variants and positive selection regions will help in understanding their role, which may improve selection based on environmental adaptation and performance. India is bestowed with rich equine germplasm. However, these animals have not been studied in detail on the genomics diversity scale. India possesses about 0.54 million equines as per the 2019 livestock census [1]. The major population of equids comprises donkeys, mules, horses, and ponies. These provide livelihood means to the rural and semi-urban societies living in arid, semi-arid, and hilly regions, especially within the foothills of Himalayas, through transport and draught whereas the remaining small population of equines is employed in the army, police, border security force, racing industry and equestrian sports [2,3]. The population of horses and ponies were reduced by 1.81% per year from 1956 to 2019. A negative annual growth rate (-3.28%) of equids for draught work over the years 1956 to 2019 (2.64 to 0.54 million) indicates farmers' preference for a faster mode of mechanised transport [1]. India has three horses (Kathiawari, Kachchhi-Sindhi, and Marwari) and four ponies (Manipuri, Spiti, Zanskari, and Bhutia) inhabiting different agro-climatic regions [4–9]. Over time, these breeds have adapted with certain unique traits like endurance, relative disease tolerance, sturdiness, and sure-footedness to sustain and work in harsh environmental conditions. Among horse breeds, Marwari is the best-adapted breed in hot, dry, and desert conditions of the Marwar area of Rajasthan. The animals of this breed are selected for their conformation, speed, and stamina, endurance capacity with good export potential, while Kathiawari, the native breed of Kathiawar area in Gujarat, is also known for its speed, stamina, endurance, and capability of fast movement. In India, seven breeds of horses have been identified based on their geographical localisation [4–8]. Different attempts have been made for characterization of horses, ponies and donkey breeds at the phenotypic, biochemical and molecular levels [4–14], but still, it is difficult to assign a breed to an unknown sample. In this study we will characterize the genetic diversity of unique and valuable indigenous horses and quantify their relationship with other breeds, aiming to identify breed-specific variants underlying unique phenotypic traits such as inward turning of ears, endurance and racing performance.

Since the domestication of the horse approximately 5,000 years ago, selective breeding has been directed mainly towards the use of the horse in agriculture, transportation, and warfare. Within the past 400 years, the founding of formal breed registries and continued breed specialisation have focused more upon preserving and improving traits related to aesthetics and performance. As a result, most horse breeds today are closed populations with high phenotypic and genetic uniformity of individuals within the breed, but with a great deal of variation among breeds. The use of genetic markers now allows for the identification of genomic variation specific to breeds to be uncovered [15], and to trace genomic relationships and pedigrees [16–18]. Genetic markers (i.e. single-nucleotide polymorphisms; SNPs) are an additional source of information that could be incorporated into the breeding evaluation to enable accurate early selection of horses for coat colour, health, and performance traits. High-throughput, whole-genome SNP arrays are now available for these purposes beyond that possible from conventional microsatellite markers. Other newer approaches include the use of genotyping-by-sequencing (GBS), RAD-seq, and whole-genome sequencing to characterise genomic diversity among individuals and within breeds. Once genomic regions targeted by selection are detected, the variants and processes that have contributed to desired phenotypes

within breeds and across performance groups can be more readily identified. Population-based approaches to identify signals of selection using the loss of heterozygosity and/or other diversity indices have been successful in several domestic species [19,20]. GBS offers great potential for characterising molecular variants and linking these variants with basic phenotypes and other desired characteristics in livestock populations because it makes it possible to cover large fractions of the genome and to vary the sequence read depth per individual. The ddRAD-Seq, a popular tool for MAS in plants, is a simple, reproducible, highly multiplexed approach based on the Illumina sequencing platform [21]. The method has been proven suitable for population studies, germplasm characterisation, genetic improvement, and trait mapping in a variety of diverse organisms. The major advantages over other protocols are both technical simplicity and that informatics pipelines are publicly available. Very few works with MAS have been reported in livestock such as cattle [22], Yak [23], chicken [24] and ducks [25] and horses [26].

2. Materials and Methods

Location of the Study and Source of Samples

The blood samples of adult, healthy and true-to-breed equines of each breed were selected based on their morphological features and breeders' information, from their breeding tracts in different parts of India (Figure 1).

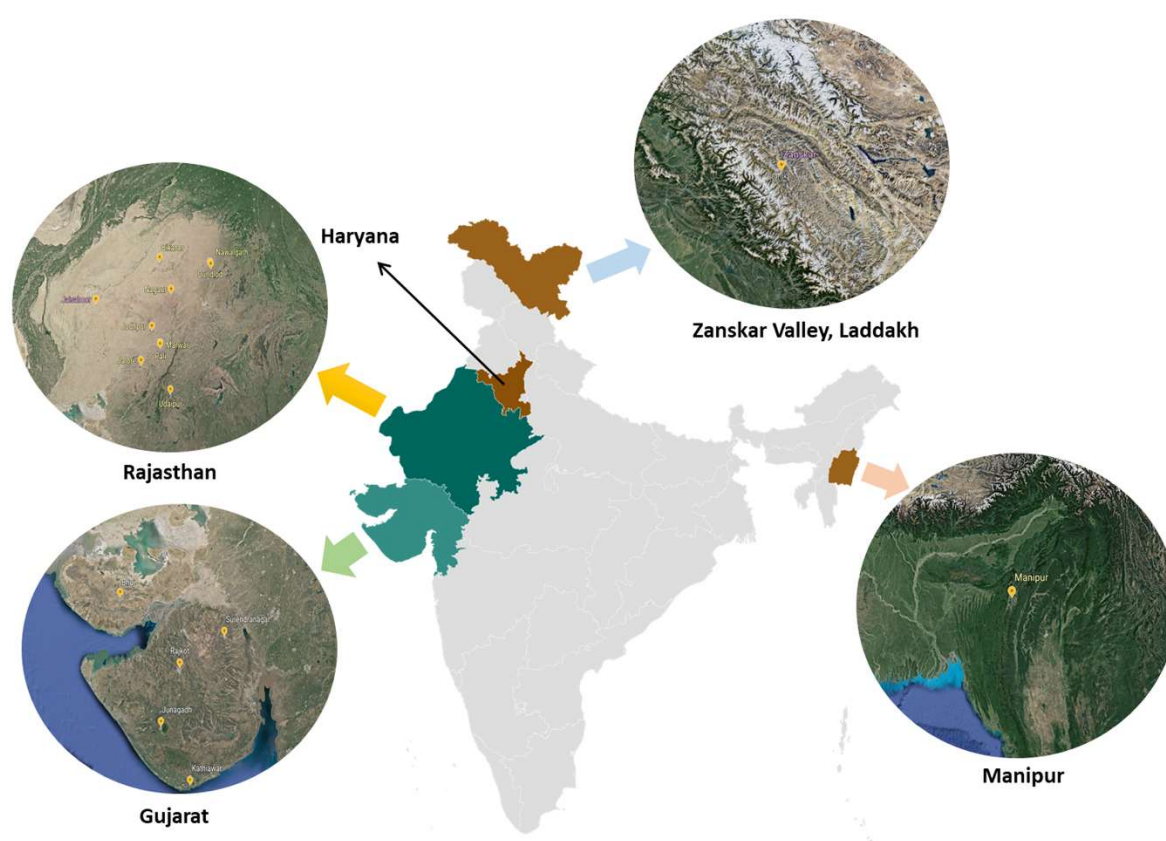


Figure 1. Sample collection from the equine breeding tracts.

These selected animals either belonged to government organizations and/or private small or big equine breeders maintaining these animals for breeding purposes. About 19 Kathiawari horses were sampled from state equine farms at Junagarh and Inaz, Police horses from Junagarh, Rajkot, Surrender

Nagar as well as some private breeders in other parts of Kathiawar (Gujarat), Marwari horses (N=26) from private breeders of Jodhpur, Pali Marwar, Udaipur, Dundlod, Nawalgarh, Jalore, Nagore, Bikaner areas (Rajasthan), the Kachchhi-Sindhi (n=24) were selected based on their morphological features in their home tracts in Bhuj (Gujarat, India) and Jaisalmer (Rajasthan, India). The selected animals belonged to the private horse breeders maintaining the animals for breeding purposes. Efforts were made to select animals from different parts of their home tract to include adequate representatives of the breed. The Manipuri ponies (N=5) and Zanskari ponies (N=5) from Equine Production Centre (EPC), ICAR-NRCE, Bikaner (Rajasthan) (the animals being procured earlier from their native breeding tracts and now maintained at Bikaner). Besides this, Thoroughbred horses (N=16) bred in India and available in different parts of Haryana and Rajasthan states were also included as an outgroup in this study. Efforts were made to cover the entire breeding tract of each horse and pony breed in India. The blood sampling of animals was performed following the relevant guidelines and regulations as approved by the Institutional Animal Ethics Committee (IAEC).

Blood Collection and DNA Isolation

About 8-10 ml of blood from the jugular vein of true-to-breed individuals of the horse/pony breeds were collected aseptically into vacutainers coated with EDTA (0.5 mM, pH 8.0). DNA was extracted from whole blood samples according to the kit protocol (ReliaPrep™ Blood gDNA Miniprep System, Promega).

Quality of Genomic DNA

DNA quality was assessed in 0.8 % Agarose w/v suspension in 0.5X TBE buffer (pH 8) as described by [27]. DNA samples devoid of smearing were considered as good. The Qubit dsDNA HS (high sensitivity) assay kit (Thermo Fisher Scientific) was used to quantify DNA in the Qubit 3.0 fluorometer (Thermo Fisher Scientific) following the manufacturer's instructions.

ddRAD-Based Genotyping Library Preparation

After the initial quality and quantity check of DNA, standard RAD protocol as described by [15] was used for further sample processing. Double digestion of DNA with Sph I and MluCI restriction enzymes was carried out followed by combinatorial barcoding including both Illumina index and an inline barcode for library preparation. Ligation of P1 and P2 adaptors was carried out using T4 DNA ligase, followed by pooling and clean-up of the ligated product. Size selection of the digested DNA product was done after running 2% agarose gel electrophoresis. Then, PCR amplification was done to enrich and assign the Illumina-specific adaptors and flow cell annealing sequences were carried out. Further, all the samples were pooled after adapter ligation and size selection was carried out. Finally, the samples were sequenced on Illumina Hi-seq 2000 and bioinformatic analysis was carried out (Figure 2).

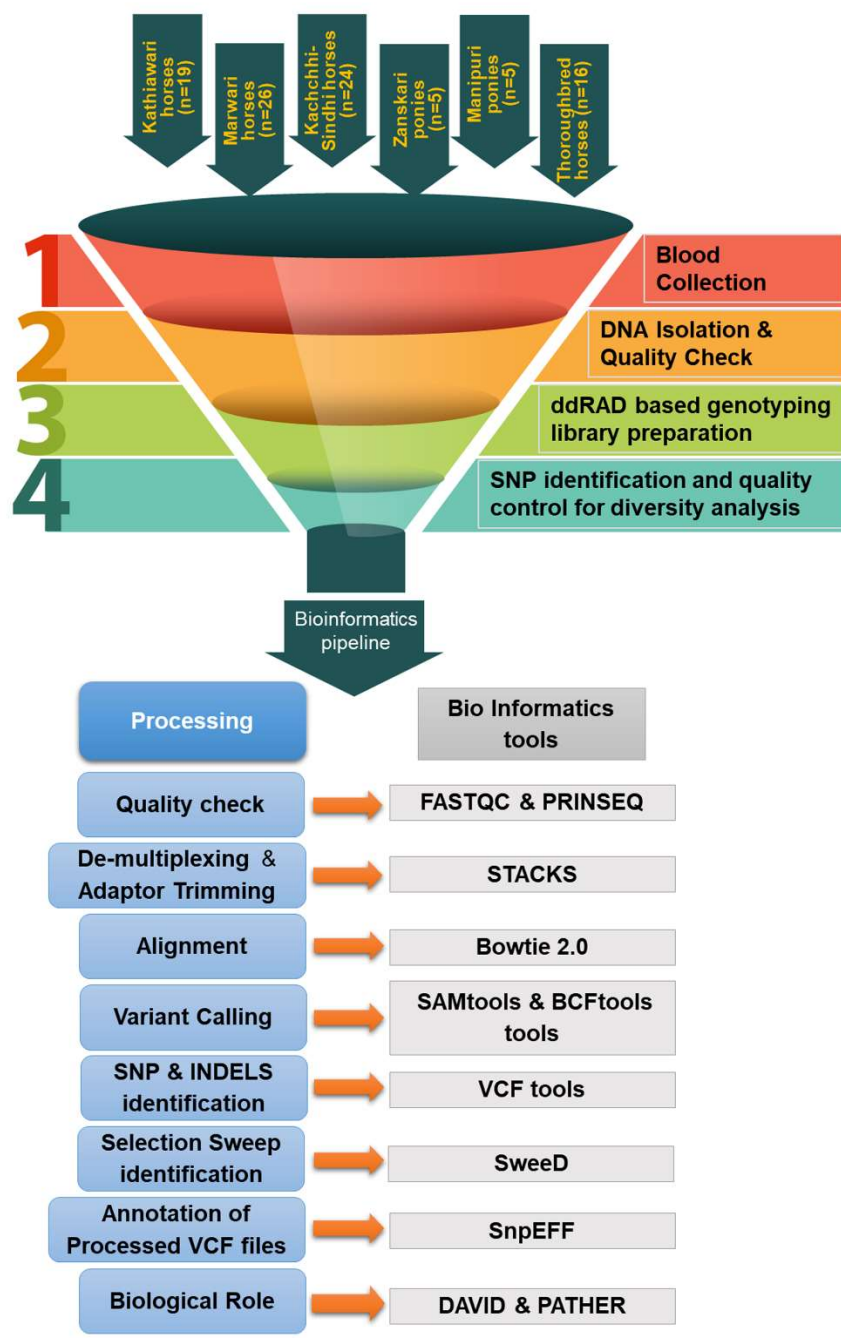


Figure 2. Bioinformatics pipeline.

SNP Identification and Quality Control for Diversity Analysis

Paired-end fastq files were trimmed for barcodes using PRINSEQ [28]. Low-quality reads with an average PHRED score of < 15 per read were discarded using STACKS 2.2 [29]. After the initial quality control, the reads were aligned with the Equus caballus reference genome (*EquCab-3.0*) using Bowtie 2 [30]. SNPs and Indels were filtered using VCFtools [31]. SNPs with good quality (Q = 30) at read depth 10 were retained for further analysis. Also, the SNPs with missing genotypes, minor allele frequency (MAF) <0.05, and those significantly (p <0.0001) deviating from the Hardy Weinberg Equilibrium were further removed by VCFtools.

Selection Sweep Identification

All breeds combined SNPs annotated vcf file at RD10 was used as input in *SweeD* [32] to find common selection signatures among indigenous and thoroughbred horse breeds. Later chromosome-wise selected regions were screened for genes using UCSC Genome Table Browser [33] after figuring out genes present in selected regions were further subjected to functional annotation using Panther GO [34]. Gene to Gene network interaction carried out using Cytoscape [35].

3. Results

Around 252.60 million raw reads from 95 samples of several horse breeds were processed, and of this, 248.08 million reads were of good quality after trimming out adaptors and removal of bad quality sequences. About 122.25 million reads (i.e. 49.20 %) of total processed reads were aligned with the reference genome EquCab-3.0 with an overall alignment rate of 92.32 %. A total of 84.07 million reads were found to be uniquely aligned among the total aligned reads, accounting for 68.11%. SNP and Indels were filtered both as combined and breed-wise horses, respectively, at many read depths (RD @ 2, 5 and 10) (Table 1), and chromosome-wise distribution of genetic variants was found for a combined dataset of Indian horses and ponies (Figure 3).

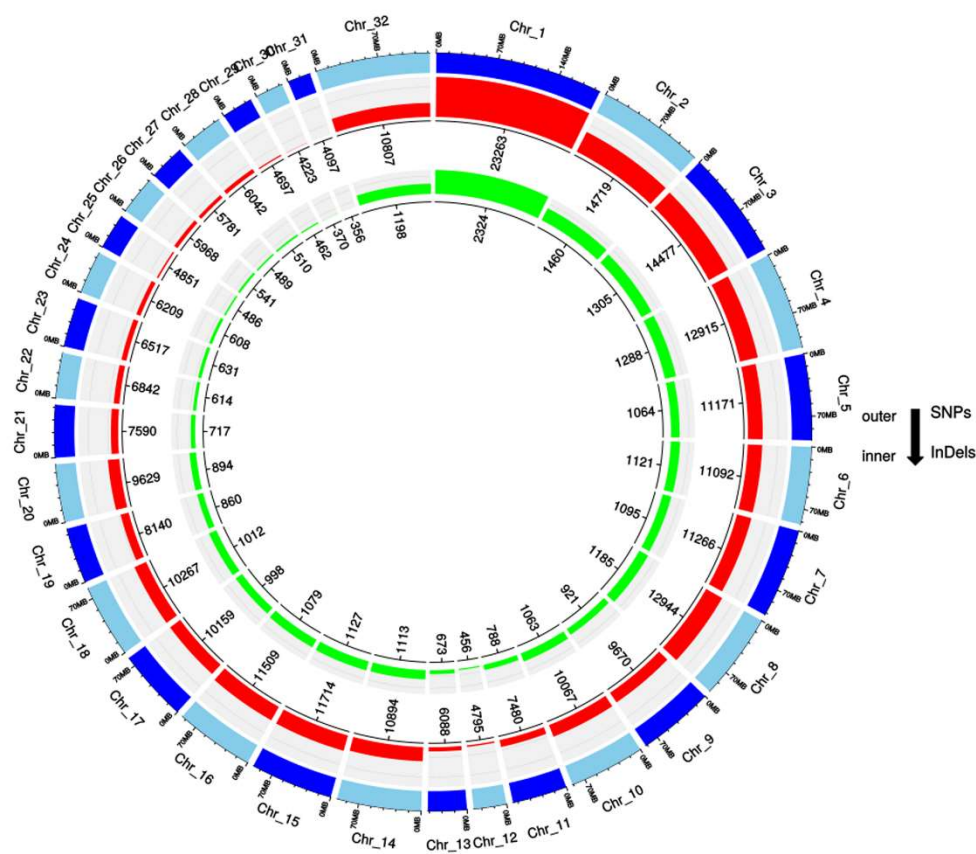


Figure 3. Chromosome-wise SNPs and Indels in Indian equines.

Table 1. Genetic variants identified in several horse breeds.

Breed	Genetic variants	RD02	RD05	RD10
Indian Horses Combined (n=76)	SNPs	328106	311554	298782
	Indels	27027	25108	23950

	Total variants	355133	336662	322732
Kachchhi-Sindhi (n=24)	SNPs	217785	210007	203071
	Indels	17094	16210	15629
	Total variants	234879	226217	218700
Kathiawari (n=17)	SNPs	185631	177773	170395
	Indels	14858	14155	13469
	Total variants	200489	191928	183864
Manipuri (n=5)	SNPs	136117	129558	118801
	Indels	10190	9697	8707
	Total variants	146367	139255	127508
Marwari (n=25)	SNPs	210509	203049	196199
	Indels	16503	15841	15322
	Total variants	220712	218890	211521
Zanskari (n=5)	SNPs	151870	146329	138934
	Indels	11292	10759	10135
	Total variants	163162	157008	149069
Thoroughbred (n=15)	SNPs	148913	144008	138022
	Indels	12089	11688	11112
	Total variants	161002	155696	149134

3.1. Genome-Wide Annotation of SNPs

Filtered SNPs of combined horses and ponies at a read depth of 10 with quality above 30 were subjected to annotation using a database created using the EquCab 3.0 reference genome. About 38.30 % and 37.95 % of total variants account for transcript and intron regions, followed by intergenic regions with 11.28% of variants (Figure 4). The substitution ratio of transitions (Ts) is 68.35%, and transversions (Tv) is 31.75 %, with a ratio of 2.16 transitions per transversion.

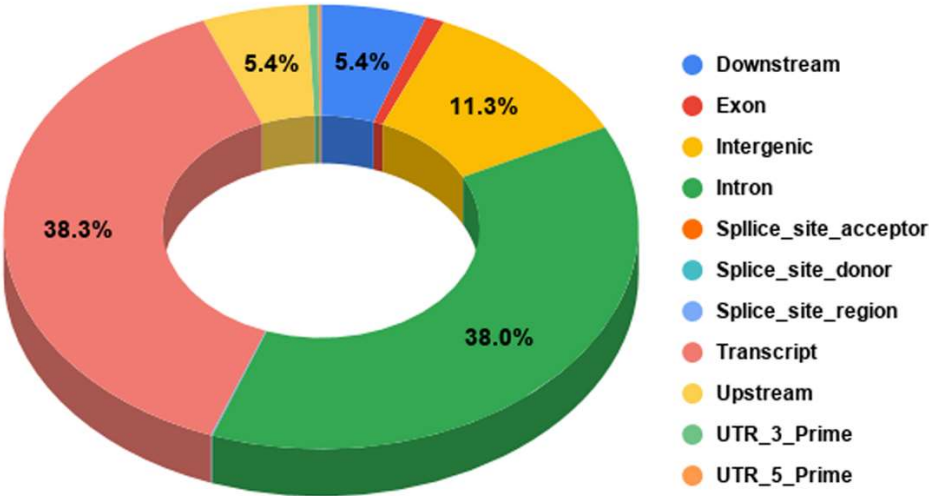


Figure 4. Proportion of the region-wise effect of variants.

3.2. Gene-Wise Mapping of SNPs and Indels

A total of 460 high-quality SNPs and 328 Indels were mapped with 30 already reported candidate genes that have many functions, and their role is discussed in Table 2.

Table 2. List of candidate genes along with their variants and their role in biological function.

Functions	Candidate genes	SNPs	Indels	Role	Reference
Racing Performance	PDZRN3 (PDZ domain-containing RING finger protein 3)	30	14	Differentiation of myoblasts	Ko et al.,2006
Racing Performance	ARL15 (ADP-ribosylation factor-like 15)	16	8	Regulator of myoblast fusion	Bach et al.,2010
Racing Performance	CNTN3 (Contactin 3)	20	4	Muscle maintenance	Jelinsky et al.,2010
Racing Performance	CCT5 (Chaperonin Containing TCP1 Subunit 5)	2	2	Muscle maintenance	Kim et al.,2008
Racing Performance	VARs2 (Valyl-TRNA Synthetase 2, Mitochondrial)	6	-	Muscle maintenance	Shin et al.,2015
Racing Performance	INPP5J (Inositol Polyphosphate-5-Phosphatase J)	2	-	Muscle maintenance	Shin et al.,2015

Racing Performance	SORCS3 (Sortilin Related VPS10 Domain Containing Receptor 3)	42	26	Endurance and speed	Velie et al.,2019;Ricard et al.,2017
Racing Performance	SLC39A2 (Solute Carrier Family 39 Member 2)	10	10	Endurance and speed	Ricard et al.,2017
Racing Performance	CCDC148 (Coiled-Coil Domain Containing 148)	12	16	Speed index	Meira et al.,2014
Racing Performance	TNR (Tenascin R)	24	6	Maintain muscle integrity	Gu et al.,2009
Racing Performance	TNC (tenascin C)	10	4	Maintain muscle integrity	Gu et al.,2009
Racing Performance	SHQ1 (Saccharomyces cerevisiae)	6	10	Differentiation of myoblasts	Shin et al.,2015
Morphology and Skeletal Development	WWOX (WW Domain Containing Oxidoreductase)	74	28	Weight and Rump length	Meira et al.,2014(b)
Morphology and Skeletal Development	Runt-related transcription factor 2 (RUNX2)	14	20	Osteoblastic differentiation and Skeletal morphology	Meira et al.,2014(b)
Morphology and Skeletal Development	Collagen alpha-1(XXVII) chain precursor	12	16	Cartilage calcification	Pereira et al.,2018
Morphology and Skeletal Development	ZFAT (Zinc Finger And AT-Hook Domain Containing)	12	6	Wither height	Signer-Hasler et al.,2012
Morphology and Skeletal Development	LCORL (ligand-dependent nuclear receptor corepressor-like)	6	4	Wither height	Tetens et al.,2013

Morphology and Skeletal Development	HNRNPU (Heterogeneous Nuclear Ribonucleoprotein U)	12	16	Muscle development	Pereira et al.,2019
Neural regulation	GRM8 ((Glutamate receptor, Metabotropic 8)	48	18	Neural transmitter; Behavioural and learning process	Meira et al.,2014 ; Murphy and Arkins.,2007
Neural regulation	GRIK2 (Glutamate Receptor, Ionotropic, Kainate 2)	34	16	Neural transmitter; Behavioural and learning process	Meira et al.,2014 ; Murphy and Arkins.,2007
Neural regulation	RXRA (Retinoid X receptor alpha)	4	14	Ocular and CNS development	Girardi et al.,2019
Neural regulation	RYBP (RING1 and YY1 binding protein)	4	14	Ocular and CNS development	Pirity et al.,2007
Immunity	RC3H2(Roquin 2)	2	14	Regulate inflammatory signals	Schaefer & Klein ,2016
Immunity	STAT3 (Signal transducer and activator of transcription 3)	4	10	Anti-inflammatory response	Leise et al.,2012
Reproduction	STRBP (Spermatid perinuclear RNA-binding protein)	4	10	Spermatid and Spermatogenesis	Meng et al.,2018
Energy metabolism	COX4I1 (Cytochrome C Oxidase Subunit 4I1)	2	4	Oxidative phosphorylation	Ricard et al.,2017
Energy metabolism	CBLB (Cbl Proto-Oncogene B)	4	2	Insulin signalling	Gurgul et al.,2019
Energy metabolism	PPARGC1A(Peroxisome proliferator-activated receptor- γ coactivator 1 α)	4	4	oxidative energy metabolism	Eivers et al.,2012

Cardiac development	ANK1 (ankyrin 1, erythrocytic)	38	18	Cardiac muscle homeostasis; calcium mediation	Meira et al.,2014
Lipid metabolism,	ACACA	2	14	Fatty acid synthesis	Dharuri et al.,2014

3.3. Selection Sweeps

A total of 1631 selection signatures were enumerated at the top 1 percentile with a greater than 1.72 likelihood score from 1,77,726 filtered SNPs. 700 selected regions were found incorporated in 306 genes; of these, 125 genes have greater than one selected region, and the top 10 genes with their corresponding selected regions were listed in Table 3.

Table 3. Top 10 genes with more number of selection regions.

Sl. No	Chromosome	Gene Name	No. of selection signature
1	18	Solute Carrier Family 4 Member 10 (SLC4A10)	15
2	7	Cell adhesion molecule 1 (CADM1)	12
3	13	Autism susceptibility gene 2 (AUTS2)	12
4	13	LOC733058	10
5	7	Zinc Finger protein 94 (ZFP94)	8
6	1	Zinc Finger Protein 658 (ZFP658)	7
7	1	General Transcription Factor IIA Subunit 2(GTF2A2)	6
8	5	Ganglioside-Induced Differentiation Associated Protein 2 (GDAP2)	6
9	11	Galectin 9 (LGALS9)	6
10	14	Erythrocyte Membrane Protein Band 4.1 Like 4a (EPB41L4A)	6

On functional annotation, the majority of genes, about 28 %, accounted for the cellular process (GO: 0009987) followed by biological regulation (GO: 0022610) and signalling (GO: 0023052), with

15.8% & 15.3% respectively of total annotated genes (Figure 5, panel A). At the molecular level, nearly half of the genes aid in binding function (GO: 0005488) and act as function regulators (GO: 0098772) (Figure 5, panel B). Annotated genes in greater proportion assist in integrity, cellular anatomical entity (GO: 0110165), and intracellular (GO: 0005622) cellular components (Figure 5, Panel C).

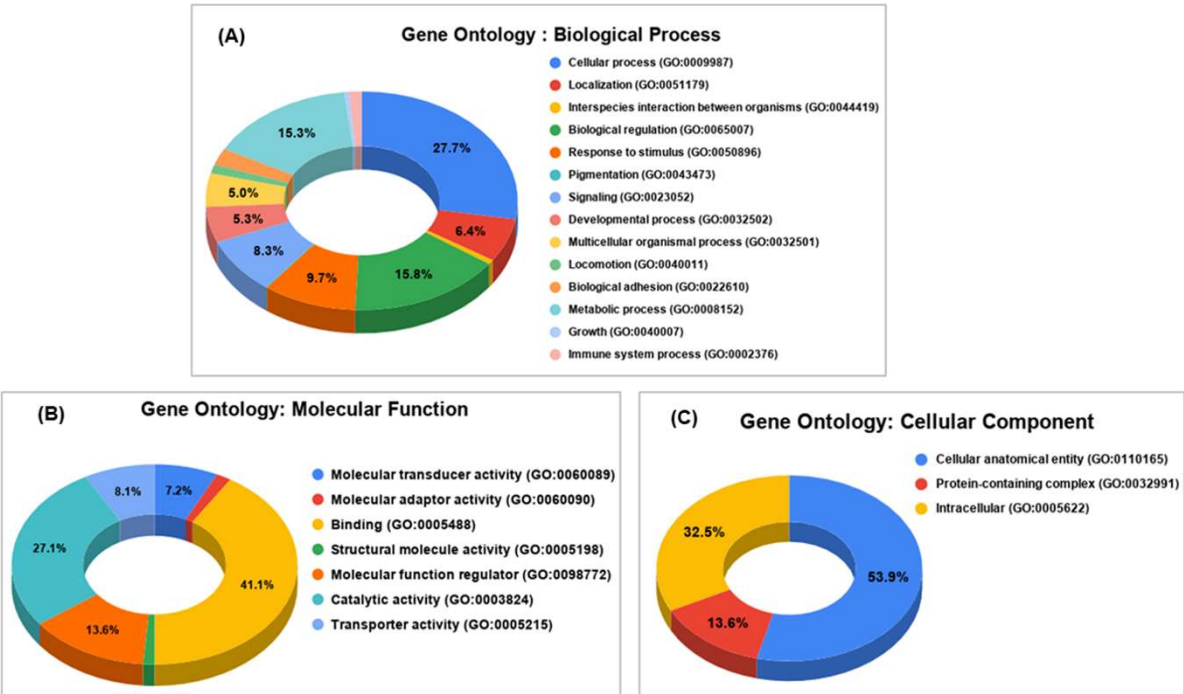


Figure 5. Gene Ontology (GO) analysis for the functional annotation of the genes. (A) GO-biological process (B) GO- Molecular function (C) GO-Cellular component.

Pathway analysis of selected region genes associated in several functional pathways revealed their roles in many vital functions related to growth, regulation of normal cell function, immunity, cardiac conduction and its regulation, neuron development and conduction of signal, hormonal regulation, transportation of mineral ion, and many signalling pathways (Table 4).

Table 4. Top 10 pathways associated with positively selected genes.

Sl. No	Name of Pathway	No of genes involved	Genes list
1	Wnt signalling pathway (P00057)	10	CDH10, PCDH15, FAT1, CDH20, PCDH9, CDH8, FAT2, PRKCQ, ARID1B, CTNNA3
2	Cadherin signalling pathway (P00012)	8	CDH10, PCDH15, FAT1, CDH20, PCDH9, CDH8, FAT2, CTNNA3

3	Metabotropic glutamate receptor group III pathway (P00039)	6	GRM4, GRIK3, GRM8, PRKX, PRKACA, GRM7
4	Heterotrimeric G-protein signalling pathway- Gq alpha and Go alpha mediated pathway (P00027)	6	GRM4, RGS7, GRM8, RGS6, PRKACA, GRM7
5	Heterotrimeric G-protein signalling pathway- Gi alpha and Gs alpha mediated pathway (P00026)	6	GRM4, GRIK3, GRM8, PRKCQ, RGS6, GRM7
6	Beta2 adrenergic receptor signalling pathway (P04378)	5	CACNAID, RYR2, PRKX, CACNB2, PRKACA
7	EGF receptor signalling pathway (P00018)	5	MRPL38, NRG4, PRKCQ, RASAL2, NRG2
8	CCKR signalling map (P06959)	5	TCF4, RYR2, BCL2, PRKCQ, PRKACA
9	Nicotinic acetylcholine receptor signalling pathway (P00044)	4	MYO10, CACNA1D, CACNB2, CHRNA7
10	Angiogenesis (P00005)	3	EPHB2, PRKCQ, EPHB1

3.4. Gene-Gene Interaction Network Analysis

On exploring gene interaction among selected region genes, many genes were found as hub genes involved in multiple biological roles that also may account for the physiological adaptation of the horse towards its high performance. Notably, the ANK2, BCL2, and RYR2 were involved in the transportation of calcium ions and cardiac regulation (Figure 6).

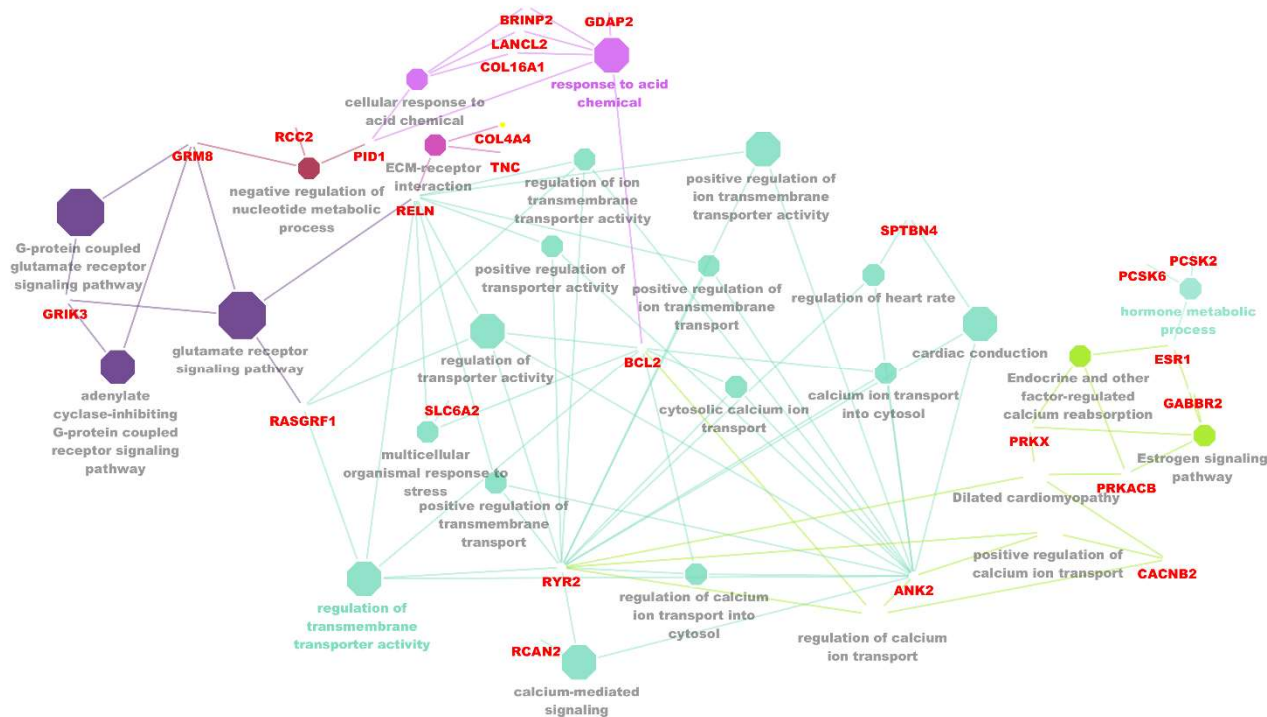


Figure 6. Gene-to-Gene interaction network.

4. Discussion

This study aimed to understand the genetic architecture of Indian horses and ponies through a genome-wide selection signature scan. Earlier, to understand the genetic variants in horses, CNVs from six Indian horse breeds, namely, Manipuri, Zanskari, Bhutia, Spiti, Kathiawari and Marwari, were discovered using a genotyping array [36]. Genome-wide single-nucleotide polymorphism-based genomic diversity and runs of homozygosity (ROH) were also elucidated for selection signatures in Indian equine Breeds [37]. The results of the presented study gave an idea of how the putative positive selected genes involved in the physiological regulation of horses in terms of cardiac, neuronal and hormonal. CADM1 and SLC4A10 genes have a greater number of selected regions. Cell adhesion molecule-1 (CADM1), an immune gene that assists in the regulation of cell-to-cell adhesion in diverse human epithelial cells and consequently provides defence as an anti-malignant agent, which inactivates a variety of cancers such as small cell lung cancer and breast cancer in humans [38,39]. Numerically high selected regions in CADM1 genes gave a glimpse that its selection might be a reason for normal lung function regulation without malignancy in horses, even after their long-term high stress involves performance. The solute carrier 4 family (SLC4A10) gene is involved in the transportation of sodium-coupled bicarbonate ions, mainly in the central nervous system and a minor proportion in the ocular receptors [40]. In a study involving positive selected regions over a variety of horse breeds, this gene is found to be common among all relating to the function of regulation of synapse assembly along with SPARC (Osteonectin), Cwcv, and Kazal Like Domains Proteoglycan 1 (SPOCK1) which resembles our findings and it was hypothesized that these genes might be involved in controlling ocular and hearing processes in horses [41].

Ryanodine Receptor type (RYR2) is a candidate gene for athletic performance, predominantly involved in the Ca⁺⁺ release channel in cardiac muscle [42]. In the mice, the phosphorylated protein encoded by RYR2 increases the inflow of intracellular calcium in myocytes, resulting in cardiac contractility and ventricular arrhythmia during exercise [43]. Researchers reported that the RYR2

gene is closely associated with the QTL region of jumping performance in sport horses [44]. ANK2 (Ankyrin 2) is a vital gene that acts in the regulation of ion channels and transporters to cardiac cell membranes and is also involved in the organisation of cellular receptors such as Na⁺/K⁺ ATPase, the Na⁺/Ca⁺⁺ exchanger, and inositol-1,4,5-trisphosphate receptors [45]. ANK2 causes a longer QT interval in the cardiac cycle via an increase in contractility by the influx of calcium in the sarcoplasmic reticulum, resulting in bradycardia in humans. It was observed that a positive selection signature in this gene might be a reason for the lower heartbeat to compensate for its high performance [46]. B-cell lymphoma 2 (Bcl-2) is an anti-apoptotic survival protein in several cells, including cardiac myocytes, against external stress factors such as heavy exercise, training, and toxic drugs, causing cardiac ischemia resulting in cardiac failure [47,48]. This gene has a closer association in the regulation of integrity and cell survival of myocytes through influencing mitochondrial membrane permeability and also by activation of caspases [49]. GRM8 (glutamate receptor, metabotropic 8) and GRIK3 (glutamate receptor, ionotropic, kainate 2) are both genes encoding for glutamate receptor regulation and activation of L-glutamate, a major excitatory neurotransmitter in the central nervous system [50]. Earlier works [51–53] idealised the glutamate involved in nervous development, synaptic assembly, behavioural and learning processes, which might influence athletic success and the usefulness of the horse. The Vitamin D receptor (VDR) gene is involved in vitamin D3 metabolism and regulation of a variety of other metabolic pathways, such as those involved in the immune response and muscle strength [54]. A study [55] found that a positive association between vitamin D and muscle strength also improves athletic performance. Tenascin C (TNC) is a glycoprotein abundant in developing tendons, bone, and cartilage, and also commands the interaction of cell-matrix [56]. [57] found tenascin coding genes (i.e., TNC, TNN, TNR) have positive-selected regions which are associated with performance and tendon pathology in thoroughbred horses.

The study provided a genome-wide scan of selection signatures in Indian horses and ponies, significantly supplementing our prior work on CNVs and ROH-based diversity in these breeds. Our identification of 1,631 positive selection signatures—700 mapped to 306 genes—reveals a genetic architecture fine-tuned for performance adaptation. Three key functional clusters emerge, with profound evolutionary and translational implications. These pertain to cardiovascular-exercise Adaptations, neural and sensory Specialization and structural integrity.

5. Conclusions

These findings contextualize Indian equine breeds as reservoirs of adaptive genetic variants shaped by diverse environmental pressures. The convergence of signatures in cardiac, neural, and structural genes across breeds suggests universal selection pressures for performance. Beyond breeding applications (e.g., selecting for TNC variants to reduce tendon injuries), our data hold cross-species relevance: CADM1's potential role in cancer resistance warrants exploration in comparative oncology, while ANK2/BCL2 insights could inform human cardiovascular research. Future work should validate these signatures in functional assays (e.g., RYR2 calcium flux studies), assess genotype-phenotype links in athletic cohorts, and expand the sampling to underrepresented breeds to decode local adaptations. Such efforts will maximize the translational potential of this genetic blueprint for enhancing equine welfare and performance globally.

Author Contributions: For research articles with several authors, a short paragraph specifying their individual contributions must be provided. The following statements should be used “Conceptualization, A.B.; methodology, A.B., J.S., V.N., S., S.P. and M.C.; Sampling, Y.P., R.A.L., A.B., V.K. and M.C.; software, A.B., J.S., V.N., S., M.A.I. and D.K.; formal analysis, A.B., J.S., V.N., and S.K.G.; investigation, A.B., J.S., V.N., S. and M.A.I.; resources, A.B.; writing—original draft preparation, A.B., J.S. and S.K.G.; writing—review and editing, A.B., J.S., V.N., D.K., T.K.B. and B.N.T.; visualization, A.B., J.S., V.N. and S.K.G.; supervision, A.B., YP, TKB, BNT; project administration, A.B.; funding acquisition, A.B. All authors have read and agreed to the published version of the manuscript.”

Funding: Please add: This research was funded by ICAR-NRCE Institution Project grant (IXX12220), DST-SERB-ECRA (grant number ECR/2017/000696) and CABIn project grant (grant no. Agril. Edn. 4-1/2013-A&P). The APC will be funded by ICAR-NRCE.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institute Animal Ethics Committee of the ICAR-National Research Centre on Equines, Hisar (Haryana), INDIA. Sampling was performed following the relevant guidelines and regulations.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data supporting this paper were generated by ICAR-National Research Centre on Equines, Hisar, and are available from the corresponding author upon reasonable request.

Acknowledgements: The authors would like to thank the funding agencies for the funds and facilities for research, which were provided through the ICAR-NRCE Institution Project grant (IXX12220), DST-SERB-ECRA grant (ECR/2017/000696) and CABIn project grant (F. no Agril. Edn. 4-1/2013-A&P).

Conflicts of Interest: There are no conflicts to declare. The authors declare that they have no conflict of interest and no competing financial interest.

Abbreviations

The following abbreviations are used in this manuscript:

ddRAD-Seq	Double Digest Restriction Associated DNA (ddRAD) Sequencing
GBS	Genotyping By Sequencing
MAS	Marker-Assisted Selection
SNP	Single Nucleotide Polymorphism

References

1. Livestock Census. Government of India, 20th Livestock Census-2019. Ministry of Fisheries, Animal Husbandry & Dairying; Department of animal husbandry and dairying, Krishi Bhawan, New Delhi, accessible at: <http://dadf.gov.in/sites/default/files/Key%20Results%2BAnnexure%2018.10.2019.pdf> (2019)
2. McSweeney C, Mackie R. Micro-organisms and ruminant digestion: state of knowledge, trends and future prospects. Commission on Genetic Resources for Food and Agriculture. 61 1-62. Accessed on 04 April 2021, available online: <http://www.fao.org/docrep/016/me992e/me992e.pdf> (2012).
3. Pal Y, Sharma P, Bhardwaj A et al. Agri-entrepreneurship development through equines, in: P. Kashyap, A.K. Prusty, A.S. Panwar, S. Kumar, P. Punia, N. Ravisankar & V. Kumar (Eds.). Agri-Entrepreneurship Challenges and Opportunities, Today and tomorrow's printers and publishers, New Delhi, pp. 165-176.(2019).
4. Pal Y, et al. "Phenotypic characterization of Kachchhi-Sindhi horses of India." Indian Journal of Animal Research 55.11 (2021): 1371-1376.
5. Pal, Yash, et al. "Status and conservation of equine biodiversity in India." Indian Journal of Comparative Microbiology, Immunology and Infectious Diseases 41.2 (2020): 174-184.
6. Gupta, AK, et al. "Phenotypic characterization of Indian equine breeds: a comparative study." Animal Genetic Resources/Recursos genéticos animales/Recursos genéticos Animales 50 (2012): 49-58.
7. Gupta AK, Chauhan M, Bhardwaj A, et al. Microsatellite markers based genetic diversity and bottleneck studies in Zanskari pony. Gene. 2012;499(2):357-361.
8. Gupta AK, Chauhan M, Bhardwaj A, et al. Comparative genetic diversity analysis among six Indian breeds and English Thoroughbred horses. Livestock Science. 2014;163:1-11.
9. Gupta AK, Chauhan M, Bhardwaj A. Genetic diversity and bottleneck studies in endangered Bhutia and Manipuri pony breeds. Molecular Biology Reports. 2013;40(12):6935-6943.
10. Gupta AK, Chauhan M, Bhardwaj A, et al. Assessment of demographic bottleneck in Indian horse and endangered pony breeds. Journal of Genetics. 2015;94(2):56-62.

11. Gupta, A. K., Kumar, S., Pal, Y., Bhardwaj, A., Chauhan, M., & Kumar, B. (2018). Genetic diversity and structure analysis of donkey population clusters in different Indian agro-climatic regions. *J Biodivers Endanger Species*, 6(006), 2.
12. Pal, Y., Legha, R. A., Lal, N., Bhardwaj, A., Chauhan, M., Kumar, S., ... & Gupta, A. K. (2013). Management and phenotypic characterization of donkeys of Rajasthan. *Indian Journal of Animal Sciences*, 83(8), 793-797.
13. Gupta, A. K., Kumar, S., Sharma, P., Pal, Y., Dedar, R. K., Singh, J., ... & Kumar, B. (2016). Biochemical profiles of Indian donkey population located in six different agro-climatic zones. *Comparative Clinical Pathology*, 25, 631-637.
14. Gupta, A., Bhardwaj, A., Sharma, P., Pal, Y., & Kumar, M. (2015). Mitochondrial DNA-a tool for phylogenetic and biodiversity search in equines. *J Biodivers Endanger Species*, 1(006).
15. Peterson BK, Weber JN, Kay EH, et al. Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. *PLoS One*. 2012;7(5):e37135.
16. Pérez-Enciso M. Genomic relationships computed from either next-generation sequence or array SNP data. *Journal of Animal Breeding and Genetics*. 2014;131(2):85-96.
17. Pérez-Enciso M, Rincón JC, Legarra A. Sequence- vs. chip-assisted genomic selection: accurate biological information is advised. *Genetics Selection Evolution*. 2015;47(1):43.
18. Pérez-Enciso M, Steibel JP. Phenomes: the current frontier in animal breeding. *Genetics Selection Evolution*. 2021;53(1):22.
19. Stella A, Ajmone-Marsan P, Lazzari B, et al. Identification of Selection Signatures in Cattle Breeds Selected for Dairy Production. *Genetics*. 2010;185(4):1451-1461.
20. Xu L, Bickhart DM, Cole JB, et al. Genomic Signatures Reveal New Evidences for Selection of Important Traits in Domestic Cattle. *Molecular Biology and Evolution*. 2015;32(3):711-725.
21. Gurgul A, Miksza-Cybulska A, Szmatoła T, et al. Genotyping-by-sequencing performance in selected livestock species. *Genomics*. 2019;111(2):186-195.
22. De Donato M, Peters SO, Mitchell SE, et al. Genotyping-by-Sequencing (GBS): A Novel, Efficient and Cost-Effective Genotyping Method for Cattle Using Next-Generation Sequencing. *PLoS One*. 2013;8(5):e62137.
23. Sivalingam J, Vineeth MR, Surya T, et al. Genomic divergence reveals unique populations among Indian Yaks. *Scientific Reports*. 2020;10(1):3636.
24. Pértille F, Guerrero-Bosagna C, Silva VH, et al. High-throughput and Cost-effective Chicken Genotyping Using Next-Generation Sequencing. *Sci Rep*. 2016;6:26929.
25. Zhu Z, Miao Z, Chen H, et al. Ovarian transcriptomic analysis of Shan Ma ducks at peak and late stages of egg production. *Asian-Australas J Anim Sci*. 2017;30(9):1215-1224.
26. Tezuka A, Takasu M, Tozaki T, et al. Genetic analysis of Taishu horses on and off Tsushima Island: Implications for conservation. *Journal of Equine Science*. 2019;30(2):33-40.
27. Sambrook J, Russell D. *Molecular cloning: a laboratory manual*. 2001. (Ed. 3).
28. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics*. 2011;27(6):863-864.
29. Catchen JM, Amores A, Hohenlohe P, et al. Stacks: Building and Genotyping Loci De Novo From Short-Read Sequences. *G3 Genes|Genomes|Genetics*. 2011;1(3):171-182.
30. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nature Methods*. 2012;9(4):357-359.
31. Danecek P, Auton A, Abecasis G, et al. The variant call format and VCFtools. *Bioinformatics*. 2011;27(15):2156-2158.
32. Pavlidis P, Živković D, Stamatakis A, et al. SweeD: Likelihood-Based Detection of Selective Sweeps in Thousands of Genomes. *Molecular Biology and Evolution*. 2013;30(9):2224-2234.
33. Kuhn RM, Haussler D, Kent WJ. The UCSC genome browser and associated tools. *Briefings in Bioinformatics*. 2013;14(2):144-161.
34. Thomas PD, Campbell MJ, Kejariwal A, et al. PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res*. 2003;13(9):2129-41.
35. Otasek D, Morris JH, Bouças J, et al. Cytoscape Automation: empowering workflow-based network analysis. *Genome Biology*. 2019;20(1):185.

36. Sharma NK, Prashant S, Bibek S, et al. Genome wide landscaping of copy number variations for horse inter-breed variability. *Animal Biotechnology*. 2025;36(1):2446251.
37. Bhardwaj A, Tandon G, Pal Y, et al. Genome-Wide Single-Nucleotide Polymorphism-Based Genomic Diversity and Runs of Homozygosity for Selection Signatures in Equine Breeds. *Genes*. 2023;14(8):1623.
38. Saito M, Goto A, Abe N, et al. Decreased expression of CADM1 and CADM4 are associated with advanced stage breast cancer. *Oncol Lett*. 2018;15(2):2401-2406.
39. Wikman H, Westphal L, Schmid F, et al. Loss of CADM1 expression is associated with poor prognosis and brain metastasis in breast cancer patients. *Oncotarget*. 2014;5(10):3076-87.
40. Liu Y, Wang D-K, Jiang D-Z, et al. Cloning and Functional Characterization of Novel Variants and Tissue-Specific Expression of Alternative Amino and Carboxyl Termini of Products of Slc4a10. *PLoS One*. 2013;8(2):e55974.
41. Lee W, Park KD, Taye M, et al. Analysis of cross-population differentiation between Thoroughbred and Jeju horses. *Asian-Australas J Anim Sci*. 2018;31(8):1110-1118.
42. Lanner JT, Georgiou DK, Joshi AD, et al. Ryanodine receptors: structure, expression, molecular details, and function in calcium release. *Cold Spring Harb Perspect Biol*. 2010;2(11):a003996.
43. Wehrens XHT, Lehnart SE, Huang F, et al. FKBP12.6 Deficiency and Defective Calcium Release Channel (Ryanodine Receptor) Function Linked to Exercise-Induced Sudden Cardiac Death. *Cell*. 2003;113(7):829-840.
44. Brard S, Ricard A. Genome-wide association study for jumping performances in French sport horses. *Animal Genetics*. 2015;46(1):78-81.
45. Huq AJ, Pertile MD, Davis AM, et al. A Novel Mechanism for Human Cardiac Ankyrin-B Syndrome due to Reciprocal Chromosomal Translocation. *Heart, Lung and Circulation*. 2017;26(6):612-618.
46. Swayne LA, Murphy NP, Asuri S, et al. Novel Variant in the ANK2 Membrane-Binding Domain Is Associated With Ankyrin-B Syndrome and Structural Heart Disease in a First Nations Population With a High Rate of Long QT Syndrome. *Circ Cardiovasc Genet*. 2017;10(1).
47. Imahashi K, Schneider MD, Steenbergen C, et al. Transgenic expression of Bcl-2 modulates energy metabolism, prevents cytosolic acidification during ischemia, and reduces ischemia/reperfusion injury. *Circ Res*. 2004;95(7):734-41.
48. Kobayashi S, Lackey T, Huang Y, et al. Transcription factor GATA4 regulates cardiac BCL2 gene expression in vitro and in vivo. *The FASEB Journal*. 2006;20(6):800-802.
49. Jafari, Afshar, et al. "Effect of exercise training on Bcl-2 and bax gene expression in the rat heart." *Gene, Cell and Tissue* 2.4 (2015): e32833.
50. Corrêa MJM, da Mota MDS. Genetic evaluation of performance traits in Brazilian Quarter Horse. *Journal of Applied Genetics*. 2007;48(2):145-151.
51. Carobrez, Antonio de Pádua. "Transmissão pelo glutamato como alvo molecular na ansiedade." *Brazilian Journal of Psychiatry* 25 (2003): 52-58.
52. Murphy J, Arkins S. Equine learning behaviour. *Behavioural Processes*. 2007;76(1):1-13.
53. Marinier SL, Alexander AJ. The use of a maze in testing learning and memory in horses. *Appl Anim Behav Sci*. 1994;39(2):177-182.
54. Ceglia L. Vitamin D and skeletal muscle tissue and function. *Molecular Aspects of Medicine*. 2008;29(6):407-414.
55. Hopkinson NS, Li KW, Kehoe A, et al. Vitamin D receptor genotypes influence quadriceps strength in chronic obstructive pulmonary disease2. *The American Journal of Clinical Nutrition*. 2008;87(2):385-390.
56. Schröder W, Klostermann A, Distl O. Candidate genes for physical performance in the horse. *The Veterinary Journal*. 2011;190(1):39-48.
57. Gu J, Orr N, Park SD, et al. A Genome Scan for Positive Selection in Thoroughbred Horses. *PLoS One*. 2009;4(6):e5767.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s)

disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.