RESEARCH ARTICLE Indian Journal of Animal Research, Volume 59 Issue 8: 1267-1276 (August 2025)

Tracing India's Canine Heritage Through SNP-based Haplotype Identification

Dapinder Singh[1], Shashi Kant Mahajan[2], Neeraj Kashyap[3], Chandra Sekhar Mukhopadhyay[1]

10.18805/IJAR.B-5480

ABSTRACT

Background: The majority of dog breeds and germplasm in India remain largely unexplored, with no previous studies conducted on the population structure of owned dogs. This study aimed to fill this gap by identifying haplotypes and exploring population structure among different breeds of dogs using genome-wide distributed SNPs.

Methods: The research utilized custom double-digest restriction site-associated DNA sequencing genotyping-by-sequencing (ddRAD-GBS) with Illumina 150 bp paired-end sequencing of 50 dog samples, resulting in 2,18,433 high-quality SNPs meeting the screening criteria. Results were obtained on 46 dog samples which were analyzed for population structure and haplotype identification using bash and R-environment. Three haplotypes (located on *AFAP1*, *CELSR1* and *GBGT1* genes) were selected based on SNP density and haplotype length for validation *via* PCR, followed by paired-end Sanger sequencing in seven different dog breeds (n= 21).

Result: The results revealed significant connections between dog breeds from Punjab and Haryana, with less pronounced affiliation with Karnataka. Sequencing results showed that *CELSR1* and *GBGT1* genes contained SNPs, while the *AFAP1* gene did not. These findings provided insights into the molecular-level population structure, SNPs and haplotypes of diverse dog breeds in India. The SNP variation identified could be utilized for the molecular characterization of indigenous dogs.

Key words: Dogs, Haplotype, Population structure, R-coding, SNPs.

INTRODUCTION

Dog (*Canis lupus familiaris*) holds a unique position as one of the earliest animals to be tamed, evolving alongside humans over millennia to become one of the most beloved companions in history. With a domestication history spanning over 14,000 years, dogs have been selectively bred for various roles, including as pets, guardians and hunting partners. This deliberate breeding has resulted in the emergence of a diverse array of dog breeds, with 199 recognized internationally by the American Kennel Club (AKC) out of an estimated 340 breeds worldwide, each exhibiting distinct morpho-physiological and behavioral traits (https://www.akc.org/dog-breeds/). The astonishing variety of dog breeds emerging relatively recently can be attributed to factors such as single nucleotide polymorphisms (SNPs) (Ali *et al*., 2020).

Single nucleotide polymorphisms (SNPs) represent variations at a single base pair in the DNA sequence, constituting the most prevalent form of genetic variation within animal genomes. These variations play a crucial role in understanding the genetic diversity of species and are frequently utilized to assess population structures across different organisms (Liu *et al*., 2022). In this study, a comprehensive collection of 2,18,433 genome-wide SNPs obtained through the ddRAD-GBS approach serves as a valuable resource for deciphering the genetic makeup and differentiation among dog breeds in India.

Haplotypes, a group of linked SNPs inherited together on a chromosome, offer additional insight into genetic diversity (Cheng-Li *et al*., 2023). Understanding the distribution and diversity of haplotypes is integral to unravelling the genetic architecture and population structure of different dog breeds (Vychodilova *et al*., 2018). On the other hand, population structure, defined as a systematic differentiation in allele frequencies among sub-populations due to non-random mating, is fundamental in tracing the evolutionary history of dog breeds.

[1]Department of Bioinformatics, College of Animal Biotechnology, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana-141 012, Punjab, India.
[2]Department of Veterinary Surgery and Radiology, College of Veterinary Sciences, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana-141 012, Punjab, India.
[3]Department of Animal Genetics and Breeding, College of Veterinary Science, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana-141 012, Punjab, India.

Corresponding Author: Chandra Sekhar Mukhopadhyay, Department of Bioinformatics, College of Animal Biotechnology, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana-141 004, Punjab, India. Email: csmukhopadhyay@gadvasu.in
ORCIDs: 0009-0000-1739-8383, 0000-0002-2948-1202, 0000-0003-0618-8422, 0000-0002-4545-4204

How to cite this article: Singh, D., Mahajan, S.K., Kashyap, N. and Mukhopadhyay, C.S. (2025). Tracing India's Canine Heritage Through SNP-Based Haplotype Identification. *Indian Journal of Animal Research*. **59(8):** 1267-1276. doi: 10.18805/IJAR.B-5480.
Submitted: 26-08-2024 **Accepted:** 22-01-2025 **Online:** 18-02-2025

To overcome the challenges posed by the high cost and time-consuming nature of high-throughput sequencing, this study adopts the genotyping-by-sequencing (GBS) approach. Genome-wide SNP-based diversity and population structure analysis using ddRAD have been carried out in different livestock species like buffalo, yak, horse and camel (Van Dijk *et al.,* 2014; Tezuka *et al.,* 2018; Kumar *et al.,* 2020; Sivalingam *et al.,* 2020). Notably, the ddRAD-GBS technique, proposed by Poland *et al.* (2012), utilizes two restriction enzymes (*PstI* and *MspI*), effectively reducing genomic complexity and enhancing sequencing library uniformity.

The Indian dog breeds have received little research attention. Research reports on the molecular characterization of indigenous dogs are very limited. The number of indigenous canine SNPs is still unknown and nearly all of them are still unidentified. Identification of haplotypes, along with analysis of population structure, remains unknown in the owned dog breeds reared in India. Therefore, this study seeks to fill this gap by using advanced genomic approaches to understand breed relationships and to look for marker haplotypes for key genes linked with these differentiated characteristics. The results are useful in elucidating the molecular description of the native dogs and provide valuable information about the molecular genetic distinctiveness of these dogs, which will be of great value in breed conservation and the improvement of the breeding program.

MATERIALS AND METHODS

Sample collection and DNA extraction for ddRAD-GBS

The genotyping-by-sequencing (GBS) data, previously generated in the Animal Genomics Laboratory, College of Animal Biotechnology, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana, Punjab, as part of the DBT-funded project entitled "Parentage Determination and Cytogenetic Profiling in Dogs (DBT-19I)," were available. Table 1 presents a brief description of the experimental samples.

The experiment was conducted from February 2023 to August 2023 at the College of Animal Biotechnology, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana, Punjab. Genomic DNA was extracted from the blood samples using the phenol-chloroform-isoamyl alcohol technique (Green and Sambrook, 2018). DNA purity was assessed using the Thermo Scientific™ NanoDrop 2000 spectrophotometer and molecular quality was confirmed via agarose gel electrophoresis. All samples were initially processed for genotyping-by-sequencing (GBS), with an initial QC report received for 50 samples (results of 46 samples were finally received).

ddRAD-genotyping by sequencing library preparation and sequencing

The samples were subsequently forwarded to Novogene Lifesciences Pvt. Ltd., Mumbai (https://en.novogene.com/), where the GBS analysis (Peterson *et al.,* 2012) was carried out along with mapping the SNPs to the reference dog genome (https://www.ncbi.nlm.nih.gov/assembly/GCF_014441545.1/).

ddRAD-GBS is a genomic approach that combines restriction enzyme digestion and next-generation sequencing to identify genetic markers. The GBS DNA library was prepared using the ddRAD approach for each sample by digesting them with two restriction enzymes (*MseI* and *HaeIII_MspI*), to reduce genomic complexity and enhance sequencing library uniformity. The resulting fragments were joined by two barcoded adapters that either had a compatible sticky end with universal Illumina P5 or P7 sequences or with the primary digestion enzyme. This process allows for the multiplexing of samples and enables their identification during sequencing, ensuring accurate data collection and analysis on the Illumina platform.

After the sequencing assessment report, the raw data for 46 samples were sequenced. The effective sequencing data was aligned with the reference sequence using BWA software (Li and Durbin, 2009). The mapping rate and coverage were then calculated based on the alignment results. SAMtools (Li *et al.,* 2009) was used to handle the BAM files.

Filtration of raw reads and SNP identification

SNP-calling is the process of identifying and cataloging single nucleotide polymorphisms (SNPs) in a genome. It was performed using SAMtools (Li *et al.,* 2009) to identify individual SNP variations. To minimize the error rate in SNP identification, the following criteria were utilized: (a) There should be more than 4 support reads for each SNP and (b) Each SNP should have a mapping quality (MQ) greater than 20.

The raw data is further processed by validating accession numbers with chromosomal details and discarding items missing this information. The samples were then separated into 13 trios using R-coding for Mendelian error evaluation. Errors like tri-allelic genotypes separated by a tab were eliminated. Trio's paternal genotypes are tested for Mendelian errors. Non-Mendelian SNPs are filtered based on segregation distortion, departing from Mendelian inheritance. The SNP file is then transformed into genotype files and subsequently used for filtering based on minor allele frequency (MAF<0.05).

Table 1: Number of samples by breed used for ddRAD-GBS data generation.

Dog breeds	Number of samples
German shepherd	5
Labrador retriever	17
Pug	6
Mudhol hound	4
Belgian malinois	6
Beagle	2
Pointer	3
Cane corso	3
Total	46

Table 2: Primer details of target gene amplification.

Gene name	Primer	5′-Sequence-3′	Length	Product Size	Tm (°C)
GBGT1	F	TTGCTGATTGGAGAGATGGGG	21	361	59.79
	R	ACCCGACTTCACACTCATAGGTA	23		60.56
CELSR1	F	TCATGCAACGCGATGTTGTG	20	568	60.11
	R	CACTGAAGGAGCTCAGACCG	20		60.11
AFAP1	F	GAGAAATTCAGGCCACACATGTG	23	524	60.37
	R	CAGGAAAGGAAATTCTTTTGG	21		53.23
AFAP1	F	GCTCCACTGAAGAGGGATATAC	22	501	57.47
	R	AAGGCACAGGCTTTCACTCG	20		60.89

Population structure analysis using R-programming

Filtered VCF files were examined in R v4.3.1 using the vcfR, poppr, ape and RColorBrewer packages. The "dist()" function generates a pairwise genetic distance matrix, reflecting genetic distances between all individuals. Genetic linkages were analysed using a UPGMA tree based on data from Punjab, Haryana and Karnataka (46 samples, threshold 50) (Tomar *et al.,* 2022). For population-wide linkages and multi-loci genotypes (MLG), a minimum spanning network (MSN) was developed utilizing a 'genlight' object and distance matrix. Allelic differences were estimated *via* bitwise.dist() from Poppr.

Principle component analysis (PCA) was performed using the glPCA() method on the genlight object and visualization was performed with ggplot2. Discriminant analysis of principal components (DAPC) also employs a 'genlight' object with predefined populations, utilising PCA findings for sample assignment. Here, the purpose is to allocate each sample to a population based on the findings of the PCA. Therefore, the same parameters used in the PCA (n.pca = 3 and n.da = 2) were used to rebuild the DAPC. ggplot2 and 'tidyr' transform data were used for scatter plot display.

For more straightforward comprehension, the graphs were recreated using ggplot2 and categorized samples by their positions. DAPC-calculated population assignments were extracted and a new data frame was built using original population labels and sample names. Using pivot_longer from 'tidyr', the data frame was rebuilt to match ggplot2's structural requirements and the columns were renamed for familiarity. The reconstructed dapc.results data frame was displayed using ggplot2. Samples are on the X-axis, membership probabilities are on the Y-axis and fill color represents original populations. Each facet in the bar plot reflects the initial population assignment, offering an ordered picture comparing membership probability with the original population.

Identification and validation of haplotypes

Haplotype-based analysis, compared to individual SNPs, may significantly lessen the frequency of false discoveries owing to the reduced number of association tests run (Hamblin and Jannink, 2011). The haplotype identification from the SNP genotyping data was conducted using an R-programming environment in the Ubuntu Linux terminal. The following criteria were followed for the identification of the haplotypes: (a) distance between SNPs: less than 80kb; (b) haplotype length: up to 150kb; and (c) at least 4 SNPs should be included in one haplotype.

Table 3: Blood samples collected from seven different dog breeds for validation studies.

Sample no.	Breed
Sample_1	French mastiff
Sample_2	Pomeranian
Sample_3	German shepherd
Sample_4	Beagle
Sample_5	Labrador retriever
Sample_6	Labrador retriever
Sample_7	Labrador retriever
Sample_8	German shepherd
Sample_9	Labrador retriever
Sample_10	Labrador retriever
Sample_11	Labrador retriever
Sample_12	Labrador retriever
Sample_13	German shepherd
Sample_14	Siberian husky
Sample_15	Labrador retriever
Sample_16	German shepherd
Sample_17	Pomeranian
Sample_18	American pitbull
Sample_19	German shepherd
Sample_20	Labrador retriever
Sample_21	German shepherd

Three haplotypes were selected for the validation trials, considering SNP density and MAF values of 0.05 or 5%. NCBI (https://www.ncbi.nlm.nih.gov/) was searched for the relevant RefSeq IDs to find related genes with the chosen haplotypes. Three genes, namely *AFAP1*, *GBGT1* and *CELSR1*, were identified and the primers for these genes were self-designed using Primer-BLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast/). The details of the primers are shown in Table 2.

Sample collection for haplotype validation

The Institutional Academic Ethical Committee (IAEC) allowed blood sample collection from dogs, as per Memo

No.: IAEC/2023/134-55. Dated 22/05/2023 and Memo No. IAEC/2023/176-186, Dated 14/09/2023. Samples were obtained from Guru Angad Dev Veterinary and Animal Sciences University (GADVASU) multispecialty clinics. Samples from seven distinct breeds were collected using a 21G needle (DISPO VANTM) from the cephalic vein (Table 3). Healthy canine samples were collected, independent of age and sex, into 15-ml sterile tubes containing anticoagulant (EDTA). Samples were kept at -20°C until analysis.

Genomic DNA was obtained from the blood samples using the phenol-chloroform-isoamyl alcohol method (Green and Sambrook, 2018). The purity of the DNA was evaluated using the Thermo Scientific™ NanoDrop 2000 spectrophotometer and the molecular quality was verified through agarose gel electrophoresis.

Sequencing of PCR results for the detection of SNPs

PCR reactions were performed in a 30 µl reaction volume using primers for *AFAP1*, *CELSR1 and GBGT1* genes using GoTaq® Colorless Master Mix (Promega), which includes Taq DNA polymerase, dNTPs, $MgCl_2$, reaction buffers and nuclease-free water. The purified PCR products of the *GBGT1*, *CELSR1* and *AFAP1* (AF3A and AF3B-fragments) genes underwent bidirectional sequencing (forward and reverse) at GeneSpec Pvt. Ltd., Kerala. Sequencing data were then visualized using FinchTV software (https://digitalworldbiology.com/FinchTV). The obtained FASTA sequences were aligned to reference chromosomal sequences of relevant genes using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/). All chromatograms were manually inspected for SNP genotyping at each discovered SNP site.

RESULTS AND DISCUSSION

Population Structure Analysis Using PCA

The GBS data files were submitted to NCBI SRA (BioProject: PRJNA843534; SRA accession number SRR26636791 to SRR26636836). The hierarchical distance tree constructed for the forty-six canine samples revealed distinct clustering patterns that align with geographic origins. Branch lengths and confidence values (100%) indicated genetic distances and sample similarities. Breeds from Punjab and Haryana exhibited closer genetic relationships, whereas Karnataka breeds demonstrated less resemblance, underscoring regional differentiation (Fig 1).

The Minimum Spanning Network (MSN), developed from reference-based SNP data, elucidated breed-specific genetic variations. Node Id 76, representing the Labrador retriever from Punjab, occupied a central position, signifying ancestral connections with shared SNPs across breeds. Branches radiating from this node highlighted regional diversity, with elongated branches corresponding to specific breeds such as Labrador retriever (21 and 79), Pug (44) and Belgian Malinois (90), affirming the genetic resemblance between Punjab, Haryana and Karnataka breeds (Fig 2).

Principle component analysis (PCA) further delineated genetic diversity, with PC2 effectively separating Punjab and Karnataka samples from a cluster predominantly including Punjab and Haryana breeds. The Karnataka samples formed a distinct orange ellipse, while the data of Punjab and Haryana exhibited overlapping clusters (Fig 3). Moreover, DAPC revealed consistent clustering with PCA results, assigning breeds to geographic states (Fig 4). The identification of genetic differences through discriminant analysis emphasized the contribution of breed-specific SNPs to regional clustering.

Membership probability analysis, depicted through colored vertical lines (Fig 5), quantified the likelihood of each sample belonging to a specific region. Visualizing posterior probabilities enhanced the interpretation of genetic affiliations, confirming shared ancestry between

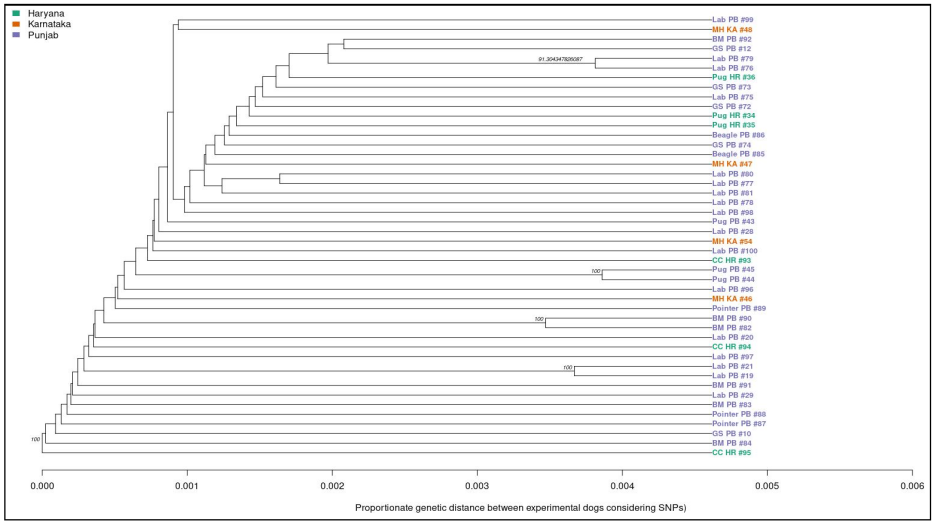


Fig 1: Genetic distance tree cluster samples with respect to their location (PB, HR and KA), highlighting region-specific genetic similarities.

breeds from Punjab, Haryana and Karnataka. This shared lineage was supported by SNP genotype similarity, as reflected in the plot of genetic diversity (Fig 6).

Haplotype identification

A systematic approach was employed to identify haplotypes using stringent criteria: SNP distance thresholds of 80 kb or cumulative differences exceeding 150 kb, with a minimum of four SNPs constituting a haplotype. This methodology uncovered 15,552 haplotypes from 2,18,429 SNP locations.

Isolation and quality check of genomic DNA

High-quality genomic DNA was extracted from 21 canine blood (Table 3) samples using the conventional Phenol: Chloroform: Isoamyl Alcohol (25:24:1) technique (Green and Sambrook, 2018) and verified through spectrophotometric analysis (OD 260/280 nm values: 1.6-1.8). The integrity of DNA was confirmed *via* gel electrophoresis on a 0.8% Agarose gel (70V for 45-50 min.), displaying unambiguous high molecular weight bands under GelDoc system, suitable for downstream applications.

Functional analysis of haplotypes

Three of the identified haplotypes were prioritized based on SNP frequency and minor allele frequency (MAF > 5%).

- **Haplotype 1:** Located on chromosome 3, spanning 187 bp with four SNPs associated with the *AFAP1* gene.
- **Haplotype 2:** Spanning 111 bp on chromosome 9 with five SNPs linked to the *GBGT1* gene.

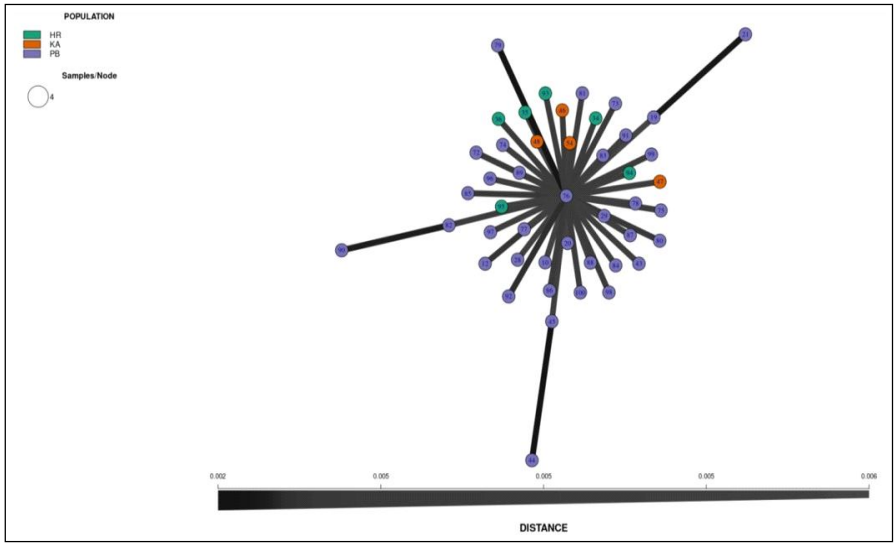


Fig 2: Minimum spanning network (MSN) shows genetic relationships among eight dog breeds, with central nodes indicating similarity and distant nodes reflecting evolutionary divergence.

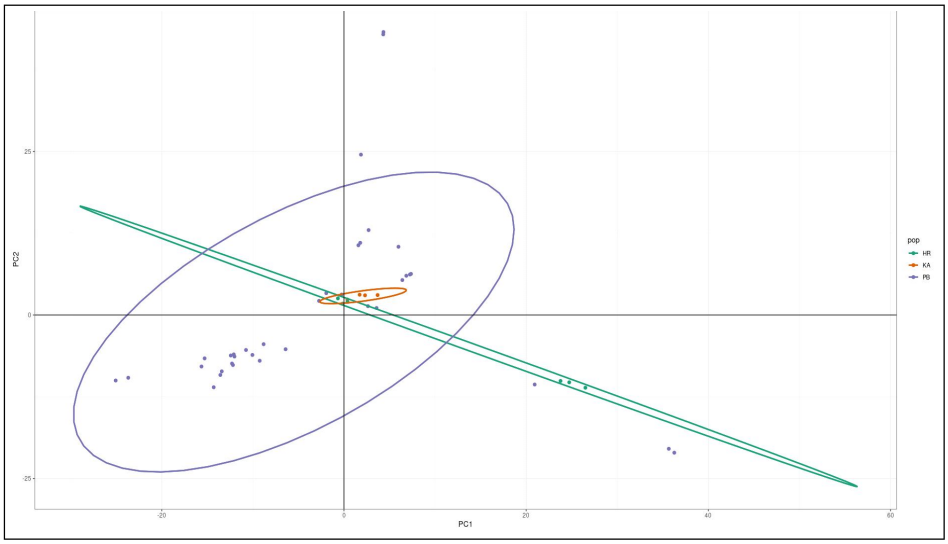


Fig 3: Principal components analysis (PCA) reveals a tight genetic cluster of breeds, with four outliers representing individuals with unique ancestral backgrounds.

- **Haplotype 3:** Covering 117 bp on chromosome 10 with six SNPs connected to the *CELSR1* gene.

Identification of polymorphisms

To identify SNPs, the reference sequence of the corresponding chromosomes of the *GBGT1*, *CELSR1* and *AFAP1* genes was aligned with the FASTA sequences acquired after sequencing using Clustal Omega. Eleven SNPs were discovered for *GBGT1* and *CELSR1*, whereas no SNP was detected in *AFAP1*. In the *GBGT1* gene, SNPs such as 503 g>A and 510 g>A were prevalent across multiple samples, whereas 516 a>G and 426 g>C were breed-specific (Table 4). Whereas, for the *CELSR1* Gene, Six SNPs were identified, with specific variants (*e.g.*, 414 g<A, 469 g>A) displaying consistent presence in specific samples (Table 5).

Dogs (*Canis lupus familiaris*) have been integral companions to humans for thousands of years, fulfilling roles ranging from hunting and protection to emotional and therapeutic support. Domesticated from wolves over 14,000 years ago, dogs are considered the first domesticated animals and have become indispensable members of society (Morell, 1997). Despite their historical and cultural significance, genetic studies on dog breeds reared in India remain limited, focusing primarily on breeds from Western and European origins. This research aims

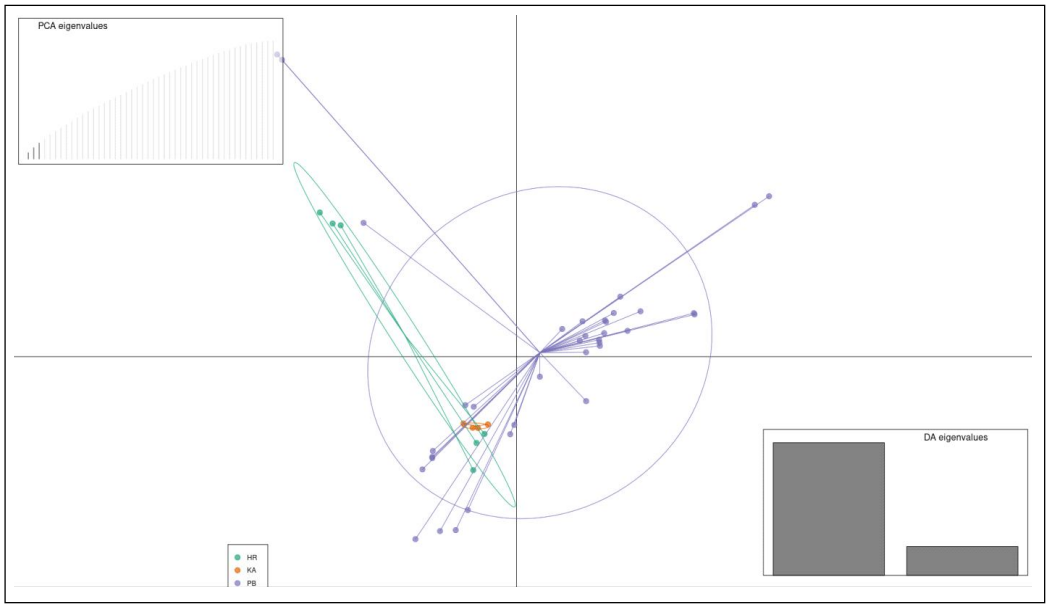


Fig 4: Discriminant analysis of principle components (DAPC) groups samples by genetic similarity, assigning each to its most likely population and highlighting distinct genetic clusters.

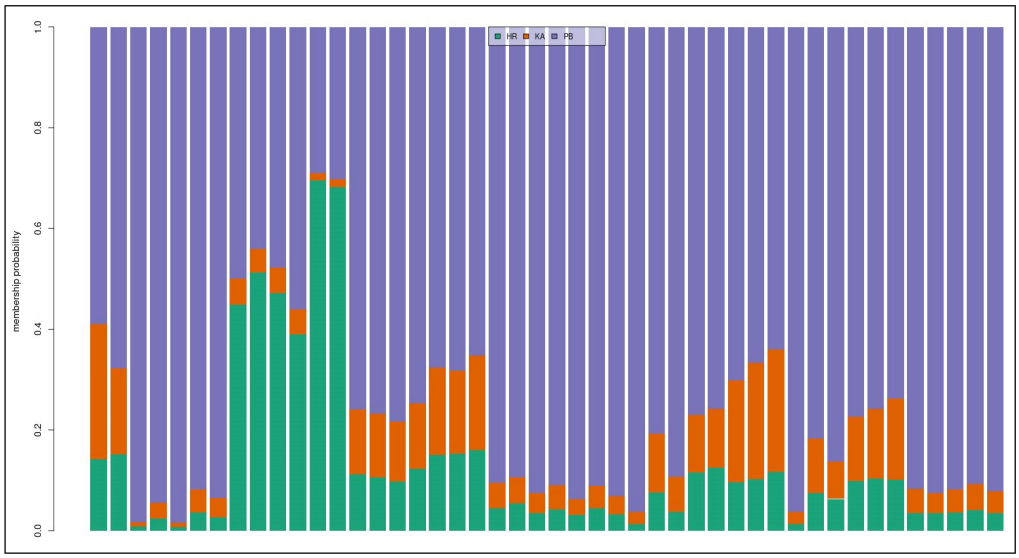


Fig 5: Membership probability of divergent samples, where each bar represents an individual dog, grouped based on their geographical origin.

to bridge this gap by analyzing the genetic structure and diversity of significant dog breeds from North and South India, utilizing high-quality SNP data derived from ddRAD-GBS technology.

Genotyping-by-sequencing (GBS) has become a powerful tool for obtaining genome-wide SNPs, facilitating genetic diversity and population structure studies. The study identified 8,13,580 SNPs, of which 2,18,433 high-quality SNPs were used to analyze the population structure of divergent dog breeds. Previous studies, such as Kaur *et al.* (2023), demonstrated the utility of GBS in analyzing genome-wide SNPs in indigenous Indian breeds like Gaddi dogs, identifying over 75,000 high-quality SNPs. This reinforces the role of GBS as a robust tool for population genetic studies and association mapping.

The population structure analysis revealed tight ancestral relationships among Indian dog breeds, with geographical proximity often correlating with genetic similarity. These findings align with Zhang *et al.* (2018), who highlighted that regional factor strongly influence

Table 4: Summary of SNPs detected in the target genomic region of canine *GBGT1* gene.

Sample no.	Position	Reference sequence	Forward	Reverse	Mutation type
Sample_1	503	G	A	G	Transition
	510	G	A	G	Transition
	516	A	A	G	Transition
Sample_2	426	G	C	G	Transversion
	503	G	A	G	Transition
	510	G	A	G	Transition
Sample_4	503	G	A	G	Transition
	510	G	A	G	Transition
	512	A	T	A	Transversion
Sample_5	503	G	A	G	Transition
	510	G	A	G	Transition
	512	A	A	G	Transition
Sample_6	503	G	A	G	Transition
	510	G	A	G	Transition
Sample_7	426	G	C	G	Transversion
	510	G	A	G	Transition
Sample_8	503	G	A	G	Transition
	510	G	A	G	Transition
	512	A	T	A	Transversion
Sample_9	503	G	A	G	Transition
	510	G	A	G	Transition
Sample_10	503	G	A	G	Transition
	510	G	A	G	Transition
Sample_12	503	G	A	G	Transition
	510	G	A	G	Transition
Sample_13	503	G	A	G	Transition
	510	G	A	G	Transition
Sample_14	503	G	A	G	Transition
	510	G	A	G	Transition
Sample_15	503	G	A	G	Transition
	510	G	A	G	Transition
Sample_16	510	G	A	G	Transition
Sample_17	357	C	C	T	Transition
	510	G	A	G	Transition
Sample_18	503	G	A	G	Transition
	510	G	A	G	Transition
	512	A	T	A	Transversion
Sample_19	510	G	A	G	Transition
	516	A	A	G	Transition
Sample_21	426	G	C	G	Transversion
	510	G	A	G	Transition

genetic structure in domestic animals. In line with this, the phylogenetic analysis confirmed breed-specific genetic differences shaped by environmental adaptations and historical breeding practices. Similar observations were reported by Freedman *et al.* (2014), who documented how selection pressures and geography impact breed evolution in dogs globally.

A significant outcome of this study was the haplotype analysis, which revealed 15,552 haplotypes. These genome regions are critical for understanding genetic architecture, as they represent blocks of conserved alleles inherited together due to linkage disequilibrium. Three haplotypes associated with AFAP1, GBGT1 and CELSR1 genes were particularly interesting. The *AFAP1* gene on chromosome 3 plays a pivotal role in cellular motility and has been linked to physiological adaptations that might benefit specific dog populations (Cunnick *et al.,* 2015).

Similarly, *GBGT1* and *CELSR1*, located on chromosomes 9 and 10, are implicated in glycosylation (Indellicato and Trinchera, 2021) and planar cell polarity pathways (Chen *et al.,* 2022), respectively. These pathways are essential for neurological development and cellular signaling, potentially impacting behavioral traits and environmental adaptability. The functional implications of these genes underscore their importance in understanding breed-specific characteristics and genetic diversity.

These findings not only provide insights into the genetic structure of Indian dog breeds but also contribute to broader efforts in breed conservation and functional genomics. Identifying specific haplotypes linked to traits such as disease resistance or behavioural attributes could guide breeding programs and improve breed welfare (Gutierrez-Reinoso *et al.,* 2021). Furthermore, this study complements global efforts to map dog genetic diversity,

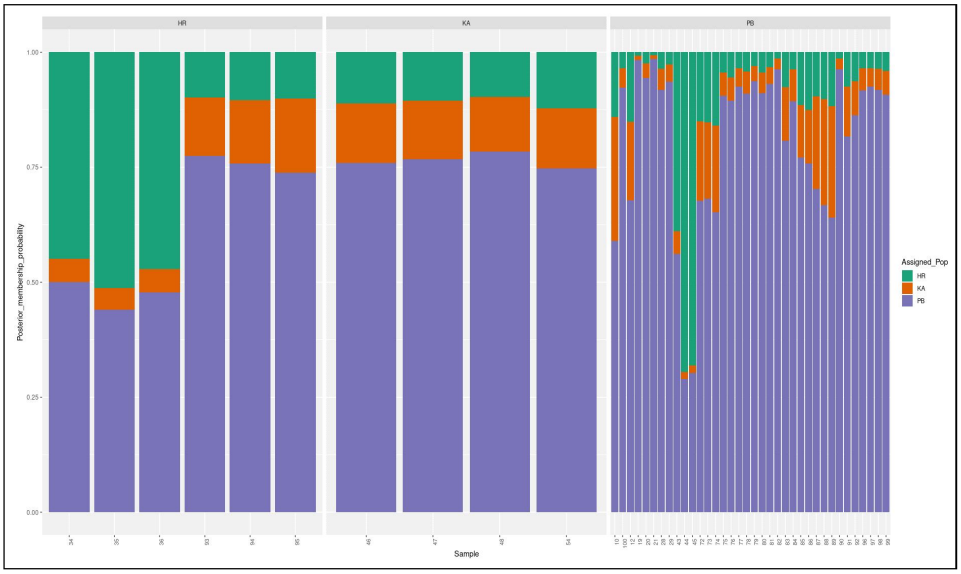


Fig 6: Posterior membership probability bar plot organizes samples by population, contrasting assignment probabilities with their original populations for clarity.

Table 5: Summary of SNPs detected in the target genomic region of canine *CELSR1* gene.

Sample no.	Position	Reference sequence	Forward	Reverse	Mutation type
Sample_4	414	G	A	G	Transition
	439	G	A	G	Transition
	445	G	A	G	Transition
	469	G	A	G	Transition
	488	G	A	G	Transition
Sample_7	414	G	A	G	Transition
	439	G	A	G	Transition
	445	G	A	G	Transition
	469	G	A	G	Transition
	488	G	A	G	Transition
	667	G	A	G	Transition
Sample_8	469	G	A	G	Transition
	488	G	A	G	Transition
Sample_13	469	G	A	G	Transition

as highlighted in studies of conserved haplotypes in humans and animals (Guryev *et al.,* 2006).

CONCLUSION

The application of ddRAD-GBS technology has significantly advanced our understanding of the genetic variation within Indian dog breeds, revealing unique haplotypes associated with critical physiological functions. This study provides valuable insights into the genetic composition, specific genetic variants (SNPs) and distinctive DNA sequences (haplotypes) in prominent canine breeds in North and South India. The extensive SNP variants identified enrich our knowledge of genetic diversity in these breeds and serve as a foundational resource for future research. The findings hold practical implications for breeding programs, enabling breeders to make informed decisions to preserve and enhance indigenous dog populations' genetic health and diversity. Future research should focus on expanding sampling and incorporating advanced genomic tools, such as transcriptomics, epigenomics and functional genomics, to explore the genetic underpinnings of breed-specific traits further. These efforts will contribute to the conservation and sustainable management of India's diverse and culturally significant canine breeds, bridging the gap in genetic studies on indigenous populations.

ACKNOWLEDGEMENT

The authors thankfully acknowledge the funding the Department of Biotechnology, Government of India, provided through the collaborative research project "Parentage Determination and Cytogenetic Profiling in Dogs (DBT-19I)".

Authorship

1) Dapinder Singh: Data analysis, blood sample collection, wet lab analysis for validation studies and manuscript writing.
2) Shashi Kant Mahajan: Provided with the blood samples for validation studies and checked the manuscript.
3) Neeraj Kashyap: Manuscript editing.
4) Chandra Sekhar Mukhopadhyay: Designing, supervising and editing the manuscript.

All authors read and approved the manuscript. All authors contributed to the manuscript revision and read and approved the submitted version.

Conflict of interest

The authors declare no conflict of interest.

REFERENCES

Ali, M.B., Evans, J.M., Parker, H.G., Kim, J., Pearce-Kelling, S., Whitaker, D.T. and Ostrander, E.A. (2020). Genetic analysis of the modern Australian labradoodle dog breed reveals an excess of the poodle genome. *PLoS Genetics*. **16(9):** e100-8956.

Chen, Z., Luo, S., Liu, Z.G., Deng, Y.C., He, S.L., Liu, X.R. and China Epilepsy Gene 1.0 Project. (2022). CELSR1 variants are associated with partial epilepsy of childhood. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*. **189(7-8):** 247-256.

Cheng-Li, L., Ri-Su, N., Wei-Wei, N., Guang-Xin, E., Yan-Guo, H., Yan, Z. and Yong-Fu, H. (2023). Genetic diversity identification and haplotype distribution of Myostatin gene in goats. *Indian Journal of Animal Research*. **57(3):** 273-281. doi: 10.18805/IJAR.B-1305.

Cunnick, J.M., Kim, S., Hadsell, J., Collins, S., Cerra, C., Reiser, P. and Cho, Y. (2015). Actin filament-associated protein 1 is required for cSrc activity and secretory activation in the lactating mammary gland. *Oncogene*. **34(20):** 2640-2649.

Freedman, A.H., Gronau, I., Schweizer, R.M., Ortega-Del Vecchyo, D., Han, E., Silva, P.M. and Novembre, J. (2014). Genome sequencing highlights the dynamic early history of dogs. *PLoS Genetics*. **10(1):** e100-4016.

Green, M. R. and Sambrook, J. (2018). Isolation and quantification of DNA. *Cold Spring Harbor Protocols*. **6:** 093-336.

Guryev, V., Smits, B.M.G., de Belt, J.V., Verheul, M., Hubner, N. and Cuppen, E. (2006). Haplotype block structure is conserved across mammals. *PLoS Genetics*. **2(7):** e121.

Gutierrez-Reinoso, M.A., Aponte, P.M. and Garcia-Herreros, M. (2021). Genomic analysis, progress and future perspectives in dairy cattle selection: A review. *Animals*. **11(3):** 599.

Hamblin, M.T. and Jannink, J.L. (2011). Factors affecting the power of haplotype markers in association studies. *The Plant Genome*. **4(2):** 145-153.

Indellicato, R. and Trinchera, M. (2021). Epigenetic regulation of glycosylation in cancer and other diseases. *International Journal of Molecular Sciences*. **22(6):** 29-80.

Kaur, B., Mahajan, S.K. and Mukhopadhyay, C.S. (2023). Tag-SNPs identification in Gaddi dogs in the light of divergent canine breeds. Manuscript in Pipeline.

Kumar, D.R., Devadasan, M.J., Surya, T., Vineeth, M.R., Choudhary, A., Sivalingam, J. and Verma, A. (2020). Genomic diversity and selection sweeps identified in Indian swamp buffaloes reveals it's uniqueness with riverine buffaloes. *Genomics*. **112(3):** 2385-2392.

Li, H. and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. **25(14):** 1754-1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N. and 1000 Genome Project Data Processing Subgroup. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*. **25(16):** 2078-2079.

Liu, B., Tao, W., Feng, D., Wang, Y., Heizatuola, N., Ahemetbai, T. and Wu, W. (2022). Revealing genetic diversity and population structure of endangered Altay white-headed cattle population using 100 k SNP markers. *Animals*. **12 (22):** 3214.

Morell, V. (1997). The origin of dogs: Running with the wolves. *Science*. **276(5319):** 1647-1648.

Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S. and Hoekstra, H.E. (2012). Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PloS One*. **7(5):** e37-135.

Poland, J.A., Brown, P.J., Sorrells, M.E. and Jannink, J.L. (2012). Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PloS One*. **7(2):** e32-253.

Sivalingam, J., Vineeth, M. R., Surya, T., Singh, K., Dixit, S.P., Niranjan, S.K. and Ravikumar, D. (2020). Genomic divergence reveals unique populations among Indian Yaks. *Scientific Reports*. **10(1).** doi: 10.1038/s41598-020-59887-3.

Tezuka, A., Takasu, M., Tozaki, T. and Nagano, A.J. (2018). The ability of ddRAD-Seq to estimate genetic diversity and genetic introgression in endangered native livestock. *BioRxiv*. 454108.

Tomar, Y.S., Tiwari, S., Tripathi, M.K., Singh, S. and Gupta, N. (2022). Genetic diversity, population structure and biochemical parameters estimations driving variations in groundnut germplasm. *Legume Research-An International Journal*. **47(2):** 190-195. doi: 10.18805/LR-4965.

Van Dijk, E.L., Auger, H., Jaszczyszyn, Y. and Thermes, C. (2014). Ten years of next-generation sequencing technology. *Trends in Genetics*. **30(9):** 418-426.

Vychodilova, L., Necesankova, M., Albrechtova, K., Hlavac, J., Modry, D., Janova, E. and Horin, P. (2018). Genetic diversity and population structure of African village dogs based on microsatellite and immunity-related molecular markers. *PLoS One*. **13(6):** e019-9506.

Zhang, M., Peng, W. F., Hu, X. J., Zhao, Y. X., Lv, F. H. and Yang, J. (2018). Global genomic diversity and conservation priorities for domestic animals are associated with the economies of their regions of origin. *Scientific Reports*. **8(1):** 116-177.