Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Genome-Wide Analyses of Breed Differentiation and Candidate Functional Genomic Differentiation in Domestic Dogs

Rongxing Wei
Guangzhou Key Laboratory of Forensic Multi-Omics for Precision Identification
jiaqian Le
Guangzhou Key Laboratory of Forensic Multi-Omics for Precision Identification
Linying Ye
Guangzhou Key Laboratory of Forensic Multi-Omics for Precision Identification
Linyuan Guo
Gene Editing Technology Center of Guangdong Province
Chang Su
Guangzhou Key Laboratory of Forensic Multi-Omics for Precision Identification
Mingyue Zhao
Guangzhou Key Laboratory of Forensic Multi-Omics for Precision Identification
Jiarong Chang
Guangzhou Key Laboratory of Forensic Multi-Omics for Precision Identification
Weibin Wu
Shaoguan Forensic Science Institute
Chao Liu
State Key Laboratory of Bioactive Molecules and Druggability Assessment, Guangdong Basic Research Center of Excellence for Natural Bioactive Molecules and Discovery of Innovative Drugs
Ling Chen
Guangzhou Key Laboratory of Forensic Multi-Omics for Precision Identification
Weian Du
duwa168@163.com

Department of Forensic Medicine, School of Basic Medical Sciences and Public Health, Jinan University

Article

Keywords: Domestic dogs, Population structure, Genomic homozygosity, Functional differentiation, Ancestry-informative markers

Posted Date: November 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11356138/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Genome-Wide Analyses of Breed Differentiation and Candidate Functional Genomic Differentiation in Domestic Dogs

Rongxing Wei[a,b†], Jiaqian Le [a†], Linying Ye [a], Linyuan Guo[c] , Chang Su [a], Mingyue Zhao[a], Jiarong Chang[a], Weibin Wu[d], Chao Liu [e], Ling Chen[a*], Weian Du[f*]

[a] *Guangzhou Key Laboratory of Forensic Multi-Omics for Precision Identification, School of Forensic Medicine, Southern Medical University, Guangzhou, Guangdong 510515, PR China.*

[b] *Nanchang Police Dog Base of the Ministry of Public Security, Nanchang, Jiangxi 330103, PR China.*

[c] *Gene Editing Technology Center of Guangdong Province, School of Medicine, Foshan University, Foshan, Guangdong 528225, PR China.*

[d] *Shaoguan Forensic Science Institute, Shaoguan, Guangdong, 512029, PR China.*

[e] *State Key Laboratory of Bioactive Molecules and Druggability Assessment, Guangdong Basic Research Center of Excellence for Natural Bioactive Molecules and Discovery of Innovative Drugs, Jinan University, Guangzhou 510632, PR China; Guangdong Province Key Laboratory of Pharmacodynamic Constituents of TCM and New Drugs Research, College of Pharmacy, Jinan University, Guangzhou 510632, PR China.*

[f] *Department of Forensic Medicine, School of Basic Medical Sciences and Public Health, Jinan University，Guangzhou, Guangdong 510632, China.*

*Corresponding author：Ling Chen, E-mail: lingpzy@163.com;

Weian Du, E-mail:duwa168@163.com

† Co-first Author：Rongxing Wei and Jiaqian Le contributed equally to this work.

Abstract

Domestic dog breeds show extensive phenotypic diversity shaped by recent breed formation, closed breeding, founder effects, genetic drift, and artificial selection. However, how these processes have contributed to genomic differentiation among modern breeds remains incompletely characterized. Here, we integrated genome wide

SNP genotype data from more than 5,000 dogs as a global reference dataset and newly genotyped approximately 300 dogs from eight focal breeds to investigate population structure, genomic inbreeding, and breed group differentiation. Genome wide dimensionality reduction and ADMIXTURE analyses revealed clear genetic structuring among modern breeds, with newly genotyped populations clustering with their corresponding public reference breeds. Genomic inbreeding analyses showed that modern purebred dogs generally exhibited elevated FROH compared with village dogs and wolves, and population level FROH was positively correlated with mean pairwise FST. Window based FST scans across selected FCI breed group comparisons identified highly differentiated genomic regions. Functional annotation of candidate genes revealed comparison specific enrichment patterns, including muscle and cytoskeletal terms in some French Bulldog related comparisons, but no uniform functional profile across all breed group contrasts. Finally, we developed compact ancestry informative SNP panels for genetic breed assignment among the eight focal breeds, with the 200 SNP panel providing clear resolution of breed structure in the present dataset. Together, these findings provide a genome wide overview of differentiation among selected dog breeds and identify candidate regions and marker sets for future studies of canine breed history, breed associated morphology, and genotype-based breed assignment.

Keywords: Domestic dogs; Population structure; Genomic homozygosity; Functional differentiation; Ancestry-informative markers;

Introduction

Human-dog coexistence can be traced back approximately 14,000-32,000 years, representing one of the earliest and longest-standing relationships between humans and animals [1-4]. During the domestication process, dogs (*Canis lupus familiaris*) experienced extensive diversification under long-term human-mediated selection. Today, more than 450 breeds are officially recognized by major kennel organizations such as the American Kennel Club and the Fédération Cynologique Internationale [5].

This remarkable phenotypic diversity has made dogs an important model for comparative genomics and studies of phenotypic evolution. Genome-wide association studies have identified the genetic basis of numerous canine traits, including body size, coat color, limb length, skull morphology, and lifespan, highlighting the power of the dog model for dissecting genotype-phenotype relationships [6-9]. Additionally, the dog model has been utilized to identify genes with translational potential for human health and biology, including both rare and common human disorders, such as autoimmune disease, neuromuscular disorders, and cancer [10-15].

Recent ancient genomic studies further suggest that dogs and humans dispersed together across Eurasia, indicating a long history of co-dispersal between the two species [16-18]. In some cases, dogs possessing valuable working or physical traits were exchanged among human populations, facilitating the spread of functionally specialized lineages. However, most modern dog breeds emerged relatively recently, primarily during the mid-19th century, when deliberate reproductive isolation and intensive artificial selection led to the establishment of standardized breeds [19-22]. These processes, often accompanied by strong population bottlenecks and closed breeding systems, generated pronounced population structure and genetic differentiation among breeds [23,24].

Selective breeding has contributed substantially to the remarkable phenotypic and genetic diversity observed among modern dog breeds [8]. Different breeds have historically been developed for a wide range of purposes, including hunting, guarding, herding, retrieving, scent detection, and companionship [22,25,26]. These breeding practices have resulted in distinct breed lineages that differ in morphology, behavior, and genetic composition. However, despite extensive research on canine population structure and breed history, the genomic basis of breed differentiation among many modern dog breeds remains incompletely understood. In particular, widely used working associated breeds, including German Shepherd Dog (GS), Belgian Malinois (BM), Kunming Dog (KM), Bloodhound (BH), Labrador Retriever (LR), Rottweiler (RT), and English Springer Spaniel (SS), together with the French Bulldog (FB),

provide an opportunity to investigate patterns of genomic differentiation among representative dog breeds.

In this study, we analyzed genome wide SNP genotypes from a large canine dataset to characterize population structure, genetic relationships, and genomic differentiation among dog breeds. We identified highly differentiated genomic regions through genome wide FST analyses and explored the biological functions of genes located within these regions. In addition, we developed compact ancestry informative SNP panels for breed assignment among eight focal breeds. These analyses provide insights into patterns of breed differentiation in domestic dogs and contribute genomic resources that may support breed assignment, population genetic studies, and breeding management.

Material and methods

Sample Collection

Newly collected blood samples from the eight focal breeds were used for SNP array genotyping, including German Shepherd Dog (GS), Belgian Malinois Shepherd Dog (BM), Kunming Dog (KM), Bloodhound (BH), Labrador Retriever (LR), Rottweiler (RT), English Springer Spaniel (SS), and French Bulldog (FB). According to the Fédération Cynologique Internationale (FCI) classification, GS and BM belong to Group 1 (Sheepdogs and Cattle Dogs), RT to Group 2 (Pinscher and Schnauzer - Molossoid and Swiss Mountain and Cattledogs), BH to Group 6 (Scent hounds and related breeds), LR and SS to Group 8 (Retrievers - Flushing Dogs - Water Dogs), and FB to Group 9 (Companion and Toy Dogs). KM is a Chinese native breed that is not recognized by the FCI. Samples were obtained from the Nanjing Police Dog Research Institute, Kunming Police Dog Base, and Nanchang Police Dog Base. A total of 312 individuals were included: 36 BM, 20 BH, 71 GS, 46 FB, 27 KM, 25 SS, 42 LR, and 45 RT (Supplementary Table S1). All sampled dogs were unrelated for at least two generations.

Blood samples were collected by routine venipuncture performed by trained personnel under veterinary supervision. No animals were anesthetized, euthanized, or killed for the purposes of this study. All methods were carried out in compliance with institutional and national animal welfare regulation and guidelines.

DNA Extraction

Genomic DNA was extracted from peripheral blood samples using the DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. Briefly, three 3×3 mm blood spot punches were incubated with lysis buffer and Proteinase K until fully solubilized. DNA was then bound to the column in the presence of ethanol, followed by sequential washes to remove impurities and residual salts. Columns were air-dried to eliminate residual ethanol, and DNA was finally eluted in 20-50 μL elution buffer for downstream analyses.

SNP Genotyping and Quality Control

Genomic DNA was genotyped using the Illumina Canine 230K SNP array following the manufacturer's protocol. DNA concentrations were measured and normalized to 50 ng/μL, with 200 ng used for the Infinium HD Assay. As part of the standard Illumina Infinium HD Assay workflow, genomic DNA was denatured and subjected to the manufacturer specified amplification step before enzymatic fragmentation and array hybridization. Amplified DNA was enzymatically fragmented in a controlled end-point reaction, precipitated with isopropanol, and resuspended in hybridization buffer. DNA samples were hybridized to the bead-based array overnight, allowing 50-base site-specific probes on each bead to anneal to the complementary DNA fragments. Non-specifically bound DNA was removed, followed by single-base extension and fluorescent labeling. Arrays were scanned using an Illumina scanner, and fluorescence intensities were converted to SNP genotypes using BeadStudio software. Genotyping quality was high, with a mean call rate of 96.57%. Ten individuals with call rates below 90% were excluded from further analyses. After applying additional filtering, a total of

290 individuals (34 BM, 15 BH, 68 GS, 39 FB, 23 KM, 25 SS, 42 LR, and 44 RT) and 196,125 markers were retained for downstream analyses.

Integration of Modern Genomic Data

We further integrated publicly available genotype data from a global canine dataset [27]. For population structure analyses (Dimensionality Reduction, and ADMIXTURE Analysis), only breeds represented by more than 10 individuals were retained (Table S2, total 95 breeds, 5,145 dogs). For FST, FROH, and TreeMix analyses, a representative subset of breeds was selected based on both geographic origin and FCI classification. Specifically, breeds from Africa, America, the Middle East, Australia, Central Asia, East Asia, Europe, and the Arctic were included, together with representatives from multiple FCI groups (28 breeds). This sampling strategy was designed to maximize geographic and breed diversity while maintaining balanced population representation. Details of the breeds included in each analysis are provided in Supplementary Table S2. Only SNPs genotyped on both arrays and passing quality control were included (99,018 SNPs).

Population genetic analysis

Dimensionality Reduction

Principal component analysis (PCA) of unrelated modern dogs was performed using the smartpca program implemented in EIGENSOFT v3.0 [28]. The two-dimensional PHATE embedding was generated using the first 20 PCs as input with phate (pca_matrix, ndim=2, knn=80, gamma=0, decay=100) [29]. Slingshot v1.4.0 was used for identifying lineages using PHATE coordinates as input and clusters defined using kmeans (centers=40) [30].

ADMIXTURE Analysis

To estimate the ancestry of dogs, SNPs were pruned for linkage disequilibrium (LD) using PLINK v1.9 (--indep-pairwise 50 5 2) [31]. ADMIXTURE v1.3.0 was run on the pruned dataset for K = 2-20 with 10-fold cross-validation and 100 bootstrap replicates

[32]. Cross-validation (CV) errors were calculated for each K to determine the optimal model, and the best-supported K was identified based on the lowest CV error. Ancestry proportions were visualized using AncestryPainterV2 [33].

Runs of Homozygosity

We used PLINK v1.9 to identify individual-level ROHs and calculated the population average level among geographically different dogs [31]. Following the parameter settings described by Sams et al. [34], ROHs were identified using a minimum ROH length of 500 kb (--homozyg-kb 500), a maximum SNP density of 5000 kb per SNP (--homozyg-density 5000), a maximum gap of 1000 kb between consecutive homozygous SNPs (--homozyg-gap 1000), a sliding window of 41 SNPs (--homozyg-window-snp 41), and no heterozygous sites allowed per window (--homozyg-window-het 0).For each detected ROH segment, physical length (in kilobases) was converted to megabases (Mb). For each individual, the total length of ROH segments was calculated by summing all ROH segments across autosomes. Genomic inbreeding coefficients based on ROH (FROH) were calculated as:

$$F_{ROH}=\frac{\text{Total ROH length}}{L}$$

where L=2203.76Mb corresponds to the total length of the CanFam3.1 autosomal genome.

Pairwise Fixation Index (FST) Estimation and Treemix Analysis

We calculated pairwise fixation indexes (FST) among representative dog populations selected from the public dataset using PLINK v1.9 with the --fst flag [31,35]. These populations were selected to represent broad geographic regions, including Africa, America, the Middle East, Australia, Central Asia, East Asia, Europe, and the Arctic, and to include multiple FCI groups while avoiding overrepresentation of closely related breeds. Population trees and historical gene flow were inferred using Treemix v1.13 [36]. Population allele frequencies were then estimated using PLINK and converted into TreeMix input format. TreeMix was performed as an unrooted analysis with the number of migration edges ranging from 0 to 7. For each migration model, five independent

replicates were conducted using 500-SNP blocks (-k 500), bootstrap resampling (-bootstrap), standard errors of migration edges (-se), and global optimization (-global). The optimal number of migration edges was determined using OptM.

Identification and functional characterization of highly differentiated loci

Eight focal breeds (BM, BH, GS, FB, KM, SS, LR, and RT) were assigned to breed groups according to the FCI classification system (Table S1). GS, BM, and KM were assigned to FCI "Sheepdogs Group 1". RT was assigned to FCI "Molossoid dogs Group 2". BH was assigned to FCI "Scent Hounds Group 6". LR and SS were assigned to FCI "Retrievers Group 8". FB was assigned to FCI "Companion and toy dogs Group 9". Because of its close genetic relationship with GS and its similar historical working roles, KM was grouped with the Group 1 breeds in subsequent analyses.

To investigate genomic differentiation among the selected FCI breed groups, pairwise FST analyses were performed among the five breed groups using PLINK v1.9. A total of ten pairwise comparisons were conducted: Group 1 versus Group 2, Group 1 versus Group 6, Group 1 versus Group 8, Group 1 versus Group 9, Group 2 versus Group 6, Group 2 versus Group 8, Group 2 versus Group 9, Group 6 versus Group 8, Group 6 versus Group 9, and Group 8 versus Group 9. Genome wide FST values were summarized using a sliding window approach with a window size of 200 kb and a step size of 100 kb. Only windows containing at least one SNP were retained for downstream analyses. For each sliding window, the mean FST was calculated as the arithmetic average of all SNP specific FST values located within the window.

$$F_{ST}^{\omega} = \frac{1}{n_{\omega}} \sum_{i=1}^{n_{\omega}} F_{ST,i}$$

n_{ω} represents the number of SNPs within window ω.

Genomic regions within the top 1% of the empirical FST distribution were defined as highly differentiated loci. Genes located within these candidate regions were annotated based on the Ensembl gene annotation database (CanFam3.1, release 104). Functional enrichment analyses, including Gene Ontology (GO) analyses, were conducted using

the R package clusterProfiler (v4.0) [37]. Gene annotation and identifier conversion were performed using the org.Cf.eg.db database. Significantly enriched GO terms were identified using a Benjamini-Hochberg adjusted P value threshold of 0.05. In addition, gene-concept network visualization was performed using the cnetplot function to illustrate the relationships between enriched functional categories and the associated genes. Finally, to identify recurrently differentiated candidate genes associated with each FCI breed group, candidate gene sets from all pairwise comparisons involving the same group were intersected. For example, Group 1 related recurrent candidate genes were defined as genes shared across the Group 1 versus Group 2, Group 1 versus Group 6, Group 1 versus Group 8, and Group 1 versus Group 9 comparisons. The same strategy was applied to Groups 2, 6, 8, and 9. UpSet plots were generated using the R package UpSetR to visualize [38].

Identification of ancestry-informative SNP panels and breed classification using population structure and machine learning analyses

Allele frequencies and pairwise linkage disequilibrium for each SNP were calculated using PLINK v1.9. Before marker selection, the dataset was stratified by breed and randomly divided into a training set comprising 70% of the individuals and an independent test set comprising the remaining 30%. All marker selection and model fitting procedures were conducted exclusively using the training set. AIM-generator[39] was applied to the training data to identify ancestry-informative markers capable of discriminating among the eight breeds, and panels containing 10, 100, and 200 SNPs were selected. For each panel, PCA was fitted using the training individuals, and the test individuals were projected onto the corresponding principal component space. ADMIXTURE was fitted to the training data assuming K = 8 ancestral populations, and ancestry proportions for the test individuals were estimated using the allele-frequency parameters inferred from the training set. Random Forest and XGBoost classifiers were then constructed using the PCA coordinates and ADMIXTURE ancestry proportions derived from the training data. Model hyperparameters were optimized by grid search

with cross-validation conducted within the training set. The final models were applied once to the independent test set, and classification performance was evaluated using overall accuracy, breed-specific classification metrics, and confusion matrices.

Results

Genome-wide Dimensionality Reduction Identifies Dog Lineages Shaped by Divergent Working Roles

All breed dogs in this study were categorized according to the Fédération Cynologique Internationale (FCI) classification system, with additional stratification based on their documented geographic origins to facilitate biologically meaningful comparisons.

To investigate genetic relationships among individuals, we performed PCA using genome-wide SNP data. The first two principal components revealed clear genetic structuring among dog breeds (Fig. 1A). A star-like configuration was observed in the two-dimensional projection, with gray wolves positioned near the center of the plot and multiple breed groups distributed along distinct trajectories. Most European breeds were distributed along three well-defined clines extending outward from the wolf cluster, indicating directional divergence along partially independent genetic axes. Among the eight focal dog breeds, GS, BM, KM, LR, RT, and FB were distinctly separated from gray wolves along the primary axis of variation, forming a tightly grouped cluster. The KM occupied an intermediate position along one of the European-associated clines, consistent with its breeding history, which involved admixture between German Shepherd Dogs imported from the Soviet Union and indigenous village dogs from Kunming in the 1950s [40]. In contrast, BH and SS were positioned closer to the wolf cluster, reflecting reduced genetic differentiation relative to other breeds.

To capture global relationships beyond discrete breed clusters, we applied potential of heat diffusion for affinity-based transition embedding (PHATE). PHATE preserves both local and global structure in high-dimensional data and models pairwise

relationships among samples as diffusion probabilities. The PHATE embedding also displayed a star-like configuration, with gray wolves located near the center and multiple breed groups distributed along distinct trajectories.

Mapping breeds according to the ten FCI groups revealed that breeds from different functional or geographic categories generally followed distinct trajectories within the PHATE manifold (Supplementary Fig. S1). For example, Sighthounds Group breeds aligned along one trajectory, Spitz and primitive types Group along another, and freely breeding village dogs clustered near the center, consistent with lower differentiation. To formally characterize these patterns, we applied pseudotemporal reconstruction, which orders clusters of individuals along branching trajectories in high-dimensional space. Clusters were defined using k-means clustering, with gray wolves serving as a central reference. In this unsupervised framework, nine major lineages were identified, whereas breeds in Dachshunds Group and Scent hounds Group remained near the central region, reflecting their relatively low genetic differentiation compared to other modern breeds (Fig. 1B).

To further investigate genetic composition, we performed ADMIXTURE analysis at K = 15, which minimized the cross-validation error. The results showed that GS, RT, LR, and FB were each characterized by a predominant inferred ancestry component (Fig. 1C). In contrast, KM, BH, BM, and SS exhibited more admixed genetic profiles, with contributions from multiple inferred ancestry components. Notably, the Kunming Dog genome contained a substantial proportion of the ancestry component that was also predominant in German Shepherd Dogs. Similarly, Bloodhounds, Belgian Malinois, and English Springer Spaniels shared varying proportions of the ancestry component that was most common in English Springer Spaniels. These patterns are consistent with the genetic relationships observed in the PCA analyses.

Genomic Inbreeding and Genetic Drift Reinforce Breed Differentiation

To investigate the demographic and evolutionary processes underlying the lineage structure identified above, we examined patterns of runs of homozygosity (ROH)

within breeds, genome-wide genetic differentiation among breeds, and historical gene flow using TreeMix. Modern purebred dogs exhibited markedly elevated genomic inbreeding coefficients (FROH) compared with village dogs and wolves (Fig. 2A). Breeds such as English Setter, BH, Irish Wolfhound, and GS showed the highest mean FROH values, frequently exceeding 0.30-0.40. In contrast, village dogs and several indigenous or indigenous populations displayed substantially lower genomic inbreeding coefficients, often below 0.10. The distribution of FROH within breeds was generally tight, suggesting breed-specific demographic constraints and closed breeding structures. Pairwise FST analysis revealed pronounced genetic differentiation among modern dog breeds (Fig. 2B). Most breed pairs exhibited moderate to high levels of differentiation (FST = 0.20-0.30), with several comparisons exceeding 0.30, indicative of strong population structure. FST values within FCI-defined groups are consistently lower than those observed between groups. For instance, Maltese and Havanese, both classified in “Companion and Toy dogs Group”, display closer genetic affinity despite originating from distinct geographic regions. Importantly, breeds characterized by elevated genomic inbreeding (high FROH) tended to exhibit greater average pairwise FST values. Pearson correlation analysis confirmed a strong positive association between population-level FROH and mean pairwise FST ($r = 0.762$, $p = 9.28 \times 10^{-21}$), suggesting that increased genomic homozygosity is tightly coupled with enhanced between-breed genetic divergence. To further investigate the evolutionary processes underlying breed differentiation, we inferred population splits and migration events using TreeMix (Fig. 2C). The optimal number of migration edges was determined to be $m = 1$ based on the Δm statistic. The resulting tree topology revealed pronounced branch elongation among modern purebred dogs, consistent with substantial genetic drift following breed formation. Most breeds belonging to the same FCI group clustered together within the tree. Notably, newly genotyped populations clustered with their corresponding public reference breeds, including LR with Labrador Retriever, GS with German Shepherd Dog, and SS with English Springer Spaniel. Together with the strong positive association between FROH and mean pairwise FST, the TreeMix analysis

supports a model in which intensive artificial breeding has simultaneously increased within-breed homozygosity and amplified between-breed genetic divergence.

Identification of Highly Differentiated Genomic Regions Across Breed Comparisons

To characterize genomic differentiation among the selected FCI breed groups, we performed pairwise window based FST scans among five breed groups, including Group 1, Sheepdogs and Cattle Dogs, Group 2, Pinscher and Schnauzer, Molossoid and Swiss Mountain and Cattle Dogs, Group 6, Scent Hounds and Related Breeds, Group 8, Retrievers, Flushing Dogs and Water Dogs, and Group 9, Companion and Toy Dogs. A total of ten pairwise comparisons were conducted (Supplementary Tabel S3-S12). For each comparison, genomic windows within the top 1% of the empirical FST distribution were defined as highly differentiated regions, and genes overlapping these regions were considered candidate genes (Supplementary Fig. S2 and S3). The top 1% FST thresholds varied across comparisons, ranging from 0.3777 in the Group 1 versus Group 8 comparison to 0.7217 in the Group 2 versus Group 6 comparison. This variation indicates different levels of genomic differentiation among the selected breed groups. The number of candidate genes identified in each comparison ranged from 345 to 455. Comparisons involving French Bulldogs showed different enrichment patterns. The Group 1 versus Group 9 comparison identified 385 candidate genes but showed no significant Gene Ontology enrichment after multiple testing correction. In contrast, the Group 2 versus Group 9 comparison identified 364 candidate genes and showed significant enrichment of muscle and contractile structure related categories, including myofibril, myosin filament, contractile muscle fiber, myosin II complex, myosin complex, and ESCRT complex (Fig. 3A). The Group 6 versus Group 9 comparison identified 455 candidate genes and was enriched mainly for dioxygenase activity, oxidoreductase activity, purinergic nucleotide receptor activity, and nucleotide receptor activity (Fig. 3B). The Group 8 versus Group 9 comparison identified 389 candidate genes and showed enrichment of myosin filament, myosin II complex, myosin complex, centrosome, myofibril, contractile muscle fiber, dioxygenase activity, and

363 microfilament motor activity (Fig. 3C). Pairwise comparisons among non Group 9
364 breed groups further revealed heterogeneous enrichment patterns. The Group 1 versus
365 Group 2 comparison identified 382 candidate genes but showed no significant GO
366 enrichment. The Group 1 versus Group 6 comparison identified 379 candidate genes
367 and showed enrichment of serine type peptidase activity, serine hydrolase activity, and
368 serine type endopeptidase activity (Fig. 3D). The Group 1 versus Group 8 comparison
369 identified 397 candidate genes and was enriched for chemokine activity (Fig. 3E). The
370 Group 2 versus Group 6 comparison identified 364 candidate genes and was enriched
371 for calmodulin binding (Fig. 3F), whereas the Group 2 versus Group 8 and Group 6
372 versus Group 8 comparisons identified 345 and 369 candidate genes, respectively, with
373 no significant GO enrichment. To further illustrate the relationships between enriched
374 functional categories and their associated genes, we constructed a gene–concept
375 network using cnetplot (Fig. 3G). The network revealed that multiple candidate genes
376 were shared among several muscle-related functional terms, forming a highly
377 interconnected module centered on myofibril and contractile muscle fiber. Notably,
378 several members of the myosin gene family, including *MYH1*, *MYH2*, *MYH3*, *MYH4*,
379 *MYH8*, and *MYL1*, were linked to multiple enriched categories, highlighting their
380 central roles in muscle structure and contraction. Additional genes such as *TNNI1* and
381 *LMOD1* were also connected to key muscle-related terms, suggesting coordinated
382 genetic differentiation in pathways related to skeletal muscle organization and
383 contractile function. These results indicate that functional annotations of highly
384 differentiated regions varied substantially among breed group comparisons.
385 To further identify candidate genes repeatedly detected in comparisons involving the
386 same FCI group, we intersected candidate gene sets across all pairwise comparisons
387 involving each group (Fig. 3H, Supplementary Table S13). Five genes, *B3GALNT2*,
388 *GNG4*, *LEMD3*, *TBCE*, and *WIF1*, were repeatedly detected in all Group 1 related
389 comparisons. Among them, *WIF1* has been reported to be related to the ear morphology
390 of canines [41]. Two genes, *BAZ2B* and *NOL4L*, were repeatedly detected in all Group 2
391 related comparisons and were located on chromosomes 36 and 24, respectively. Seven
392 genes, *CAMSAP2*, *CDK8*, *CETN3*, *GPR25*, *INAVA*, *RNF6*, and *TFEC*, were repeatedly

detected in all Group 6 related comparisons. The largest recurrent gene set was observed for Group 9 related comparisons. A total of 66 candidate genes were shared, with most of these genes located on chromosomes 5 and 26. Chr5 region may reflect breed specific morphological differentiation, particularly craniofacial morphology associated with French Bulldog brachycephaly [42]. Among these, 10 genes were annotated in the GO database as being associated with ATP related pathways (GO:0016887 and GO:0005524).

Development of a genetic breed assignment SNP panel for dog breeds

To develop a compact SNP panel for breed assignment and genetic management of police dogs, the dataset was first stratified by breed and divided into a training set comprising 70% of the individuals and an independent test set comprising the remaining 30%. All marker selection and model development procedures were restricted to the training set. Allele frequencies and linkage disequilibrium patterns were calculated in the training individuals using PLINK, and AIM-generator was then used to rank ancestry-informative SNPs according to their discriminatory ability among the eight breeds. Based on these rankings, panels containing 10, 100, and 200 SNPs were constructed. PCA fitted to the training data showed that the 10-SNP panel provided limited separation among several breeds, whereas breed clustering became progressively clearer with the 100-SNP and 200-SNP panels. The 200-SNP panel provided the clearest separation of the training individuals along the first two principal components (Fig. 4A; Supplementary Fig. S4). Test individuals were subsequently projected onto the corresponding principal component spaces. ADMIXTURE analysis was fitted to the training data at K = 8, and ancestry proportions for the independent test individuals were estimated using the parameters inferred from the training set (Fig. 4B). Random Forest and XGBoost models were trained using PCA coordinates and ADMIXTURE ancestry proportions derived from the training data, with hyperparameters optimized by cross-validation within the training set. When applied to the independent test set, all three panels achieved 100% classification accuracy with

both Random Forest and XGBoost, with no misclassified individuals in the corresponding confusion matrices (Fig. 4C and 4D). These results demonstrate that the selected markers retained high discriminatory performance in individuals that were not used for marker selection or model development. Although the three panels showed identical classification accuracy in the present test set, the 200-SNP panel provided clearer separation in PCA and more distinct ancestry profiles at K = 8. It therefore offered a more complete representation of genetic differentiation among the eight focal breeds.

Discussion

In this study, we provide a comprehensive genomic characterization of eight target dog breeds, integrating genome-wide SNP data, demographic inference, and functional analyses. We show that modern dog breeds form distinct lineages radiating from an ancestral gray wolf-like gene pool, with clear patterns of genomic differentiation reinforced by intensive artificial breeding. Our analyses demonstrate that FCI-defined breed groups largely correspond to genomic trajectories, while breeds with complex admixture histories occupy intermediate positions. Pairwise FST scans further showed that the functional annotations of candidate genes varied across comparisons, indicating that genomic differentiation among the focal breed groups is heterogeneous rather than explained by a single shared pattern. In addition, the ancestry informative SNP panels showed strong breed assignment performance within the eight focal breeds, providing a potential tool for genetic breed assignment when phenotypic information is unavailable.

Population structure and genomic differentiation among dog breeds

Breed classification systems, such as those used by kennel clubs, are largely based on the historical working functions of breeds [43], providing an ecologically meaningful framework for investigating behavioural and genetic differentiation among dog breeds. The star-like pattern observed in PCA and PHATE embeddings reflects rapid divergence of modern breeds from a common ancestral pool. Modern standardized

breeds generally occupied distinct regions of the PCA and PHATE embeddings, consistent with the effects of recent breed formation, closed breeding, founder events, and artificial selection. In contrast, village dogs and certain indigenous populations remain near the central cluster, consistent with limited human intervention and reduced divergence [44,45]. FROH in modern breeds, combined with high pairwise FST values, indicates that closed breeding structures amplify both within-breed homozygosity and between-breed divergence [46]. The strong positive correlation between FROH and FST underscores that inbreeding and genetic drift are closely linked processes driving the reinforcement of breed boundaries. TreeMix analysis further revealed that most gene flow occurs within closely related breeds or within functional FCI groups, rather than across major breed categories, consistent with deliberate human management of mating practices to preserve breed identity. These results indicate that breed membership and recent breeding history are the major determinants of genome wide similarity, whereas broad functional categories alone do not fully explain the observed genetic relationships among breeds [47,48].

Highly differentiated regions across FCI breed group comparisons

Window based FST scans were used to evaluate regional patterns of genomic differentiation among the selected FCI breed groups. This approach reduces the influence of stochastic variation at individual SNPs and summarizes differentiation across neighbouring markers, which are often correlated through linkage disequilibrium. The pairwise scans revealed marked heterogeneity in the magnitude of differentiation among the selected FCI breed groups. The top 1% FST thresholds varied substantially among comparisons, ranging from 0.3777 for Sheepdogs and Cattle Dogs versus Retrievers and Flushing Dogs to 0.7217 for Molossoid dogs versus Scent Hounds. This variation indicates that genomic differentiation was not uniform across breed group pairs. Such heterogeneity is expected in domestic dogs, because modern breeds were formed through recent breed creation, founder effects, closed breeding, and selection for breed defining traits [49-51]. Previous canine genomic studies have shown that

demographic history and breed formation can generate signals that resemble selection, especially in small or subdivided populations [49,52-55]. Accordingly, the highly differentiated regions identified here provide a framework for comparing how genomic differentiation varies across FCI breed group contrasts. The functional enrichment results showed that the biological annotations of candidate regions differed among comparisons. In comparisons involving French Bulldogs, the Molossoid dogs versus French Bulldog and Retrievers versus French Bulldog comparisons were enriched for muscle and cytoskeletal terms, including myofibril, myosin filament, myosin complex, myosin II complex, contractile muscle fiber, and microfilament motor activity [56,57]. These terms were connected to a cluster of muscle associated genes, including *MYH1, MYH2, MYH3, MYH4, MYH8, MYL1, TNNI1*, and *LMOD1*. *MYL1* links this differentiated module to previous canine studies of musculoskeletal traits and supports further investigation of these regions in relation to breed specific morphology and locomotor biology [58,59].

The recurrent candidate genes identified by intersection analysis provide a complementary view of breed group differentiation. Five gene regions were shared across all comparisons involving Sheepdogs and Cattle Dogs, including *B3GALNT2, GNG4, LEMD3, TBCE,* and *WIF1*. Among these, *WIF1* has been reported in relation to canine ear morphology (chr10: 8027948) [60]. The largest recurrent signal was observed in comparisons involving French Bulldogs, with 66 genes shared across all four comparisons with the other selected FCI breed groups. Most of these genes were located on canine chromosomes 5 and 26, indicating that this signal was regionally concentrated. The chromosome 5 signal (chr5: 29243555-34607475, CFA5.32359028) is particularly relevant in light of previous canine genetic studies showing that skull shape variation, including the brachycephaly to dolichocephaly continuum, is influenced by major genomic regions and developmental genes [42,61].

Ancestry informative SNP panels for genetic breed assignment

Genetic breed assignment can provide useful information in settings where phenotypic information is unavailable, incomplete, or unreliable. This includes biological trace analysis, degraded DNA samples, pedigree verification, ancestry assignment, detection of unintended admixture, and genetic monitoring in breeding programs. In this context, we developed compact ancestry informative SNP panels for assigning individuals among the eight focal breeds included in this study. The performance of the SNP panels depended on both the number of markers and the analytical framework used for classification. PCA based on the 10 SNP panel showed partial overlap among several breeds in two dimensional space, indicating that visual inspection of PCA alone was insufficient for confident breed separation. However, when PCA coordinates and ADMIXTURE derived genetic component proportions were used as input features, both Random Forest and XGBoost achieved 100% accuracy in the present internal validation set. Compared with the smaller marker sets, the 200 SNP panel provided clearer PCA clustering and more distinct genetic component profiles, suggesting greater robustness for representing breed structure. These panels offer cost-effective tools for pedigree validation, breed verification, and genetic monitoring in breeding programs. Importantly, precise breed assignment also facilitates the detection of unintended crossbreeding and the monitoring of inbreeding levels, both of which are critical for maintaining healthy breeding populations [62-64].

Limitation

Several limitations should be considered when interpreting these results. First, the analyses of highly differentiated regions were based on a limited number of focal breeds within each FCI breed group. The use of SNP array data also limits the resolution of the analyses. Second, FST measures allele frequency differentiation between groups but does not determine the direction of allele frequency change. Therefore, the observed signals cannot by themselves distinguish selection from demographic processes such as drift, bottlenecks, or founder effects. Third, the ancestry informative SNP panels were evaluated for breed assignment among the eight focal breeds in the current dataset.

Their performance in closely related breeds, admixed dogs, village dogs, and independent populations remains to be tested.

Future studies combining larger and more balanced breed sampling, within breed comparisons of working, show, and pet lines, whole genome sequencing, and detailed phenotypic records will be necessary to clarify the genetic basis of breed related morphology, behavior, and performance traits.

Conclusions

This study provides a genome wide assessment of population structure, genomic inbreeding, and breed group differentiation among selected dog breeds. Together, these findings provide a genome wide overview of differentiation among selected dog breeds and identify candidate regions and marker sets that may support future studies of canine breed history, breed associated morphology, and genotype-based breed assignment.

Supplementary Materials

The following supporting information can be downloaded at:

Supplementary Figure S1 PHATE plot delineating paths of 9 canine lineages originating with gray wolf. Different colors represent distinct FCI groups, with most NAs corresponding to village dogs.

Supplementary Figure S2 Genome-wide distribution of window-based FST values across Group 1 and Group 2, Group 1 and Group 6, Group 1 and Group 8, Group 1 and Group 9, and Group 2 and Group 6.

Supplementary Fig. S3. Genome-wide distribution of window-based FST values across Group 2 and Group 8, Group 2 and Group 9, Group 6 and Group 8, Group 6 and Group 9, and Group 8 and Group 9.

Supplementary Fig. S4. PCA of eight dog breeds based on 10-SNP panels.

Author Contributions

Rongxing Wei and Jiaqian Le contributed equally to this work. Conceptualization, Project Administration, and Writing - Original Draft Preparation, Rongxing Wei and Jiaqian Le. Resources and Writing - Review and Editing, Chao Liu and Weian Du. Supervision and Funding Acquisition, Ling Chen and Weian Du. Methodology, Linying Ye. Software, Rongxing Wei. Validation, Jiarong Chang. Formal Analysis, Linyuan Guo. Investigation, Chang Su and Mingyue Zhao. Data Curation, Weibin Wu. All authors have read and agreed to the published version of the manuscript.

Funding information

This work was supported by the Jiangxi Provincial Program for Selected Candidates to Lead Key Research Projects (No. 20223BBG71020).

Ethics approval

This study was approved by the Ethics Committee of the Nanchang Police Dog Base (No. GNJ20240103). All DNA samples were obtained from dogs with the owners'

informed consent, in compliance with institutional and national animal welfare guidelines.

Data and model availability statement

The genomic datasets generated and analyzed during the current study are publicly available in the OMIX database under accession number OMIX016186 (BioProject: PRJCA061648).

Acknowledgements

We are grateful to all colleagues and volunteers who contributed support for this study.

Declaration of interest

The authors declare no conflict of interest.

References

1 Savolainen, P., Zhang, Y. P., Luo, J., Lundeberg, J. & Leitner, T. Genetic evidence for an East Asian origin of domestic dogs. *Science* **298**, 1610-1613, doi:10.1126/science.1073906 (2002).

2 Thalmann, O. *et al.* Complete mitochondrial genomes of ancient canids suggest a European origin of domestic dogs. *Science* **342**, 871-874, doi:10.1126/science.1243650 (2013).

3 Frantz, L. A. *et al.* Genomic and archaeological evidence suggest a dual origin of domestic dogs. *Science* **352**, 1228-1231, doi:10.1126/science.aaf3161 (2016).

4 Wang, G. D. *et al.* Out of southern East Asia: the natural history of domestic dogs across the world. *Cell Res* **26**, 21-33, doi:10.1038/cr.2015.147 (2016).

5 Zhou, Q. J. *et al.* A single-nucleus transcriptomic atlas of the dog hippocampus reveals the potential relationship between specific cell types and domestication. *Natl Sci Rev* **9**, nwac147, doi:10.1093/nsr/nwac147 (2022).

6 Sutter, N. B. & Ostrander, E. A. Dog star rising: the canine genetic system. *Nat Rev Genet* **5**, 900-910, doi:10.1038/nrg1492 (2004).

7 Rimbault, M. *et al.* Derived variants at six genes explain nearly half of size reduction in dog breeds. *Genome Res* **23**, 1985-1995, doi:10.1101/gr.157339.113 (2013).

8 Plassais, J. *et al.* Whole genome sequencing of canids reveals genomic regions under selection and variants influencing morphology. *Nat Commun* **10**, 1489, doi:10.1038/s41467-019-09373-w (2019).

9 Rudolph, T. E., Ramey, A. R., Cullen, J. N., Friedenberg, S. G. & Clark, L. A.

Identification of genetic variants associated with ear length in drop-eared dogs. *Sci Rep* **16**, 3105, doi:10.1038/s41598-025-33036-0 (2025).

10 Shearin, A. L. *et al.* The MTAP-CDKN2A Locus Confers Susceptibility to a Naturally Occurring Canine Cancer. *Cancer Epidemiology, Biomarkers & Prevention* **21**, 1019-1027, doi:10.1158/1055-9965.Epi-12-0190-t (2012).

11 Karyadi, D. M. *et al.* A Copy Number Variant at the KITLG Locus Likely Confers Risk for Canine Squamous Cell Carcinoma of the Digit. *PLOS Genetics* **9**, e1003409, doi:10.1371/journal.pgen.1003409 (2013).

12 Karlsson, E. K. *et al.* Genome-wide analyses implicate 33 loci in heritable dog osteosarcoma, including regulatory variants near CDKN2A/B. *Genome Biology* **14**, R132, doi:10.1186/gb-2013-14-12-r132 (2013).

13 Schiffman, J. D. & Breen, M. Comparative oncology: what dogs and other species can teach us about humans with cancer. *Philosophical Transactions of the Royal Society B: Biological Sciences* **370**, doi:10.1098/rstb.2014.0231 (2015).

14 Koch, Z. *et al.* Multi-omic analysis of canine aging uncovers conserved aging pathways. *Geroscience*, doi:10.1007/s11357-025-02029-2 (2025).

15 Das, S. *et al.* Comparative genomic and clinicopathological analysis uncovers contrasting molecular profiles of canine and human thyroid carcinomas. *Commun Biol* **9**, 4, doi:10.1038/s42003-025-09225-y (2025).

16 Zhang, S.-J. *et al.* Genomic evidence for the Holocene codispersal of dogs and humans across Eastern Eurasia. *Science* **390**, 735-740, doi:10.1126/science.adu2836 (2025).

17 Marsh, W. A. *et al.* Dogs were widely distributed across western Eurasia during the Palaeolithic. *Nature* **651**, 995-1003, doi:10.1038/s41586-026-10170-x (2026).

18 Bergström, A. *et al.* Genomic history of early dogs in Europe. *Nature* **651**, 986-994, doi:10.1038/s41586-026-10112-7 (2026).

19 Boyko, A. R. *et al.* A Simple Genetic Architecture Underlies Morphological Variation in Dogs. *PLOS Biology* **8**, e1000451, doi:10.1371/journal.pbio.1000451 (2010).

20 vonHoldt, B. M. *et al.* Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. *Nature* **464**, 898-902, doi:10.1038/nature08837 (2010).

21 Ostrander, E. A., Wayne, R. K., Freedman, A. H. & Davis, B. W. Demographic history, selection and functional diversity of the canine genome. *Nat Rev Genet* **18**, 705-720, doi:10.1038/nrg.2017.67 (2017).

22 Parker, H. G. *et al.* Genomic Analyses Reveal the Influence of Geographic Origin, Migration, and Hybridization on Modern Dog Breed Development. *Cell Reports* **19**, 697-708, doi:https://doi.org/10.1016/j.celrep.2017.03.079 (2017).

23 Marsden, C. D. *et al.* Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. *Proc Natl Acad Sci U S A* **113**, 152-157, doi:10.1073/pnas.1512501113 (2016).

24 Proschowsky, H. F. *et al.* A new future for dog breeding. *Animal Welfare* **34**, e1, doi:10.1017/awf.2024.66 (2025).

25 Dutrow, E. V., Serpell, J. A. & Ostrander, E. A. Domestic dog lineages reveal genetic

drivers of behavioral diversification. *Cell* **185**, 4737-4755.e4718, doi:10.1016/j.cell.2022.11.003 (2022).

26 Zhou, T. *et al.* Dog10K: an integrated Dog10K database summarizing canine multi-omics. *Nucleic Acids Research* **53**, D939-D947, doi:10.1093/nar/gkae928 (2025).

27 Shannon, L. M. *et al.* Genetic structure in village dogs reveals a Central Asian domestication origin. *Proceedings of the National Academy of Sciences* **112**, 13639-13644, doi:doi:10.1073/pnas.1516215112 (2015).

28 Patterson, N., Price, A. L. & Reich, D. Population Structure and Eigenanalysis. *PLOS Genetics* **2**, e190, doi:10.1371/journal.pgen.0020190 (2006).

29 Moon, K. R. *et al.* Visualizing structure and transitions in high-dimensional biological data. *Nature Biotechnology* **37**, 1482-1492, doi:10.1038/s41587-019-0336-3 (2019).

30 Street, K. *et al.* Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 477, doi:10.1186/s12864-018-4772-0 (2018).

31 Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, s13742-13015-10047-13748, doi:10.1186/s13742-015-0047-8 (2015).

32 Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* **19**, 1655-1664, doi:10.1101/gr.094052.109 (2009).

33 Chen, S. *et al.* AncestryPainter 2.0: Visualizing Ancestry Composition and Admixture History Graph. *Genome Biology and Evolution* **16**, evae249, doi:10.1093/gbe/evae249 (2024).

34 Sams, A. J. & Boyko, A. R. Fine-Scale Resolution of Runs of Homozygosity Reveal Patterns of Inbreeding and Substantial Overlap with Recessive Disease Genotypes in Domestic Dogs. *G3 Genes|Genomes|Genetics* **9**, 117-123, doi:10.1534/g3.118.200836 (2019).

35 Weir, B. S. & Cockerham, C. C. ESTIMATING F-STATISTICS FOR THE ANALYSIS OF POPULATION STRUCTURE. *Evolution* **38**, 1358-1370, doi:10.1111/j.1558-5646.1984.tb05657.x (1984).

36 Pickrell, J. K. & Pritchard, J. K. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet* **8**, e1002967, doi:10.1371/journal.pgen.1002967 (2012).

37 Wu, T. *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation* **2**, 100141, doi:https://doi.org/10.1016/j.xinn.2021.100141 (2021).

38 Conway, J. R., Lex, A. & Gehlenborg, N. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* **33**, 2938-2940, doi:10.1093/bioinformatics/btx364 (2017).

39 Yang, Q. *et al.* Fine-scale biogeographical ancestry inference in Southeast and East Asians via high-efficiency markers and machine learning approaches. *Frontiers in Ecology and Evolution* **Volume 13 - 2025**, doi:10.3389/fevo.2025.1572596 (2025).

40 Wang, G. D. *et al.* Signature of balancing selection at the MC1R gene in Kunming dog populations. *PLoS One* **8**, e55469, doi:10.1371/journal.pone.0055469 (2013).
41 Plassais, J. *et al.* Whole genome sequencing of canids reveals genomic regions under selection and variants influencing morphology. *Nature Communications* **10**, 1489, doi:10.1038/s41467-019-09373-w (2019).
42 Schoenebeck, J. J. *et al.* Variation of BMP3 contributes to dog breed skull diversity. *PLoS Genet* **8**, e1002849, doi:10.1371/journal.pgen.1002849 (2012).
43 Pongrácz, P. & Dobos, P. Behavioural differences and similarities between dog breeds: proposing an ecologically valid approach for canine behavioural research. *Biological Reviews* **100**, 68-84, doi:https://doi.org/10.1111/brv.13128 (2025).
44 Larson, G. *et al.* Rethinking dog domestication by integrating genetics, archeology, and biogeography. *Proc Natl Acad Sci U S A* **109**, 8878-8883, doi:10.1073/pnas.1203005109 (2012).
45 Parker, H. G. *et al.* Genetic Structure of the Purebred Domestic Dog. *Science* **304**, 1160-1164, doi:10.1126/science.1097406 (2004).
46 Letko, A. *et al.* Genomic diversity and population structure of the Leonberger dog breed. *Genetics Selection Evolution* **52**, 61, doi:10.1186/s12711-020-00581-3 (2020).
47 Bigi, D. *et al.* Investigating the population structure and genetic differentiation of livestock guard dog breeds. *Animal* **12**, 2009-2016, doi:https://doi.org/10.1017/S1751731117003573 (2018).
48 Suárez, N. M., Betancor, E., Fregel, R. & Pestano, J. Genetic characterization, at the mitochondrial and nuclear DNA levels, of five Canary Island dog breeds. *Animal Genetics* **44**, 432-441, doi:https://doi.org/10.1111/age.12024 (2013).
49 Freedman, A. H. *et al.* Demographically-Based Evaluation of Genomic Regions under Selection in Domestic Dogs. *PLOS Genetics* **12**, e1005851, doi:10.1371/journal.pgen.1005851 (2016).
50 Akey, J. M. *et al.* Tracking footprints of artificial selection in the dog genome. *Proceedings of the National Academy of Sciences* **107**, 1160-1165, doi:10.1073/pnas.0909918107 (2010).
51 Axelsson, E. *et al.* The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature* **495**, 360-364, doi:10.1038/nature11837 (2013).
52 Crisci, J. L., Poh, Y.-P., Mahajan, S. & Jensen, J. D. The impact of equilibrium assumptions on tests of selection. *Frontiers in Genetics* **Volume 4 - 2013**, doi:10.3389/fgene.2013.00235 (2013).
53 Slatkin, M. & Wiehe, T. Genetic hitch-hiking in a subdivided population. *Genetical Research* **71**, 155-160, doi:10.1017/S001667239800319X (1998).
54 Santiago, E. & Caballero, A. Variation After a Selective Sweep in a Subdivided Population. *Genetics* **169**, 475-483, doi:10.1534/genetics.104.032813 (2005).
55 Cagan, A. & Blass, T. Identification of genomic variants putatively targeted by selection during dog domestication. *BMC Evolutionary Biology* **16**, 10, doi:10.1186/s12862-015-0579-7 (2016).
56 Osiak-Wicha, C., Kras, K. & Arciszewski, M. B. Comparative Analysis of Muscle Fibers in Selected Muscles of Working and Companion Dog Breeds. *Animals*

(Basel) **14**, doi:10.3390/ani14243576 (2024).

57 Acevedo, L. M. & Rivero, J. L. New insights into skeletal muscle fibre types in the dog with particular focus towards hybrid myosin phenotypes. *Cell Tissue Res* **323**, 283-303, doi:10.1007/s00441-005-0057-4 (2006).

58 Baker, L. A. *et al.* Genome-wide association analysis in dogs implicates 99 loci as risk variants for anterior cruciate ligament rupture. *PLoS One* **12**, e0173810, doi:10.1371/journal.pone.0173810 (2017).

59 Lee, B. T. *et al.* Identification of genetic variants associated with anterior cruciate ligament rupture and AKC standard coat color in the Labrador Retriever. *BMC Genom Data* **24**, 60, doi:10.1186/s12863-023-01164-z (2023).

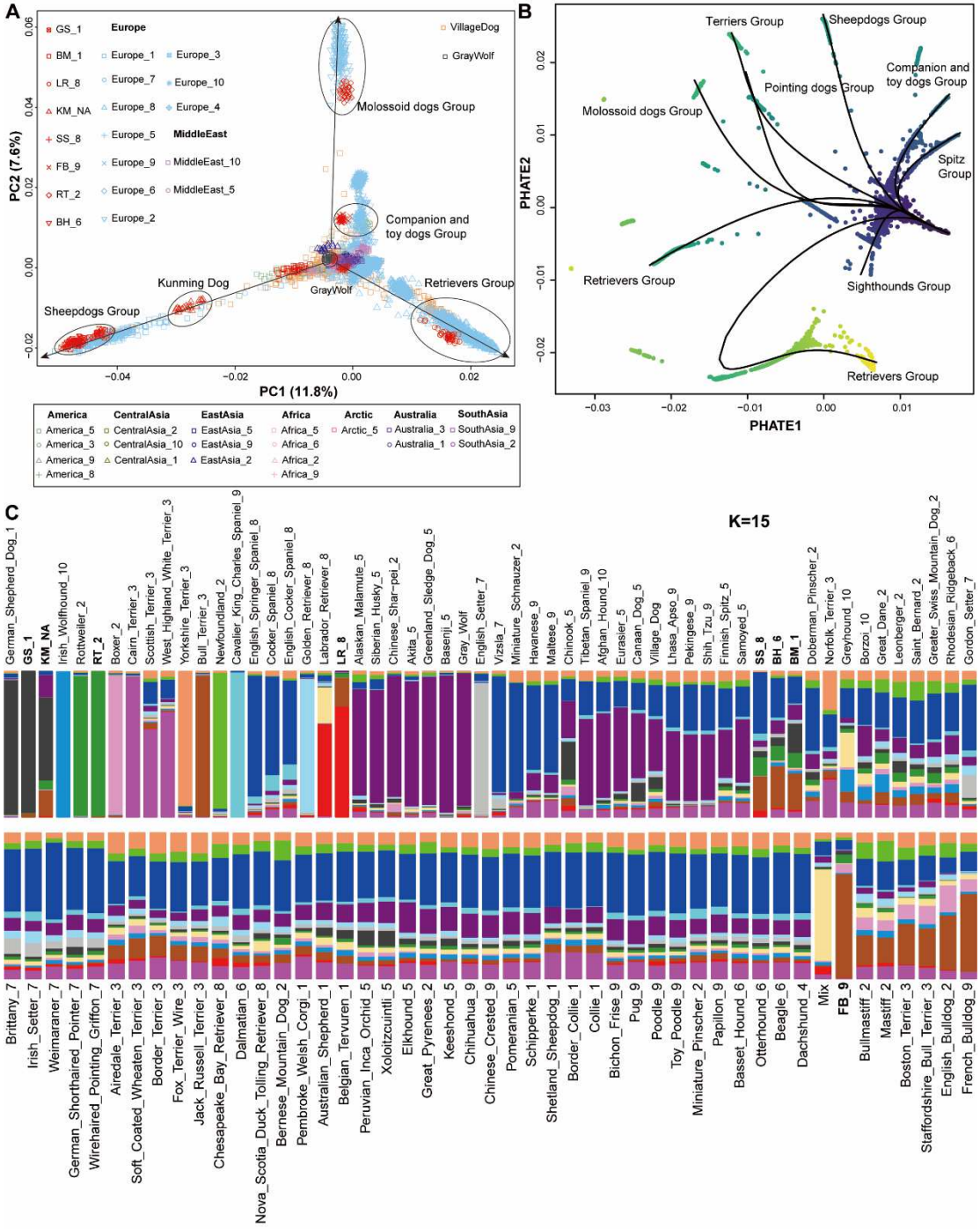
60 Morrill, K. *et al.* Ancestry-inclusive dog genomics challenges popular breed stereotypes. *Science* **376**, eabk0639, doi:10.1126/science.abk0639 (2022).

61 Mansour, T. A. *et al.* Whole genome variant association across 100 dogs identifies a frame shift mutation in DISHEVELLED 2 which contributes to Robinow-like syndrome in Bulldogs and related screw tail dog breeds. *PLOS Genetics* **14**, e1007850, doi:10.1371/journal.pgen.1007850 (2018).

62 Leroy, G. Genetic diversity, inbreeding and breeding practices in dogs: results from pedigree analyses. *Vet J* **189**, 177-182, doi:10.1016/j.tvjl.2011.06.016 (2011).

63 Kropatsch, R., Melis, C., Stronen, A. V., Jensen, H. & Epplen, J. T. Molecular Genetics of Sex Identification, Breed Ancestry and Polydactyly in the Norwegian Lundehund Breed. *Journal of Heredity* **106**, 403-406, doi:10.1093/jhered/esv031 (2015).

64 Mellanby, R. J. *et al.* Population structure and genetic heterogeneity in popular dog breeds in the UK. *Vet J* **196**, 92-97, doi:10.1016/j.tvjl.2012.08.009 (2013).



760

761 **Fig. 1. Dimensionality reduction and genetic ancestry of dogs. A**: Principal
762 component analysis based on genome wide SNP data. Newly genotyped focal breeds
763 are highlighted in red. Publicly available reference populations are labeled according
764 to geographic origin and FCI group classification when applicable. The numbers
765 following geographic labels indicate the corresponding FCI group. The full list of
766 population labels, geographic origins, FCI group assignments, and sample sizes is

provided in Table S2. **B**: PHATE plot delineating paths of 9 dog lineages originating from the gray wolf. Points are colored by maximum pseudotime in any lineage and plotted lines are principal curves generated with Slingshot. **C**: ADMIXTURE result (K=15) illustrating inferred genetic components of dogs. The numbers following geographic labels indicate the corresponding FCI group. FCI Group 1, Sheepdogs and Cattle Dogs; Group 2, Pinscher and Schnauzer, Molossoid and Swiss Mountain and Cattle Dogs; Group 6, Scent Hounds and Related Breeds; Group 8, Retrievers, Flushing Dogs and Water Dogs; Group 9, Companion and Toy Dogs.

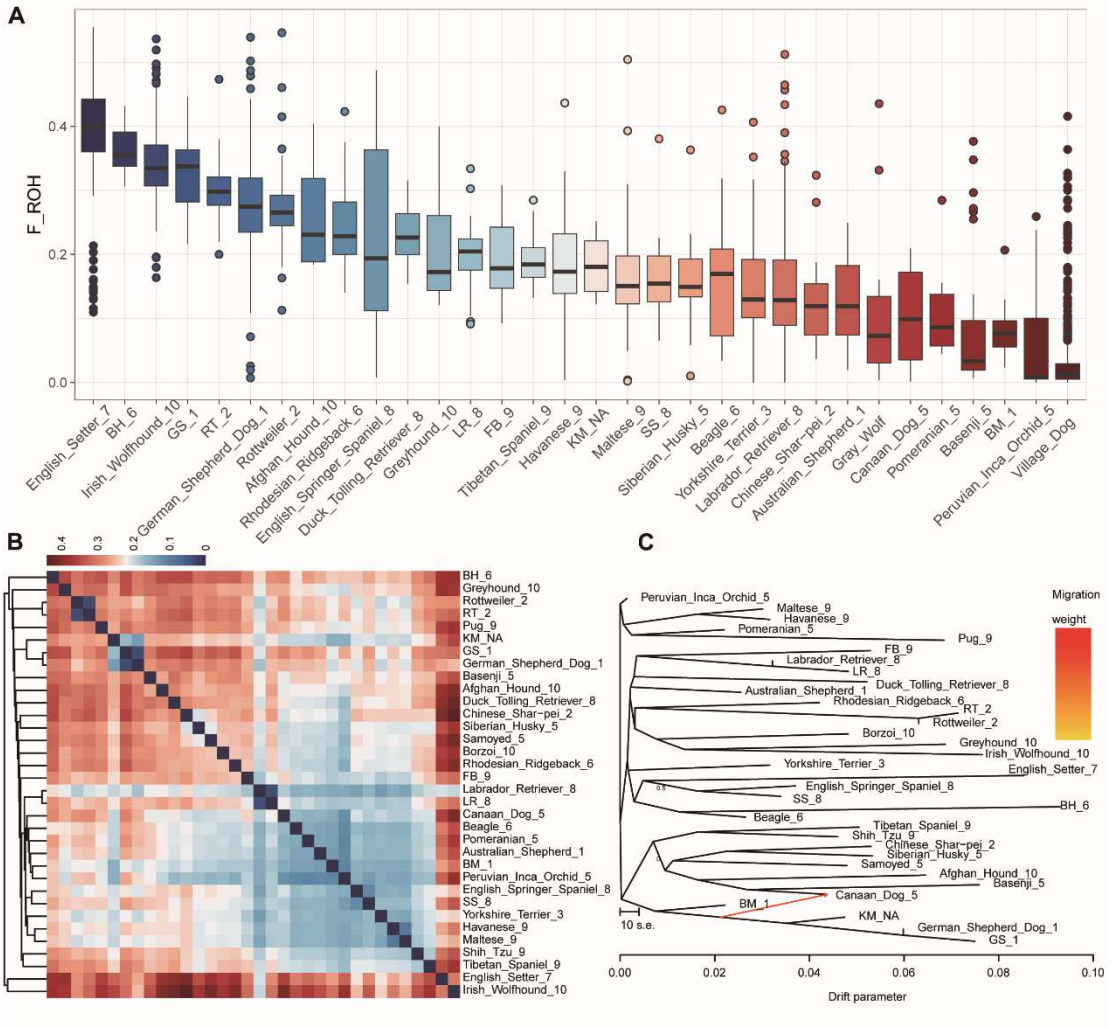


Fig. 2. Genomic homozygosity, population differentiation, and historical gene flow of dogs. A: Genome-wide inbreeding coefficients estimated from runs of homozygosity (F_ROH). **B:** Heatmap of pairwise genome-wide FST among dog breeds. **C:** Maximum-likelihood population tree inferred using TreeMix with one migration edges (Δm = 1). The numbers following geographic labels indicate the corresponding FCI group. FCI Group 1, Sheepdogs and Cattle Dogs; Group 2, Pinscher and Schnauzer, Molossoid and Swiss Mountain and Cattle Dogs; Group 6, Scent Hounds and Related Breeds; Group 8, Retrievers, Flushing Dogs and Water Dogs; Group 9, Companion and Toy Dogs.

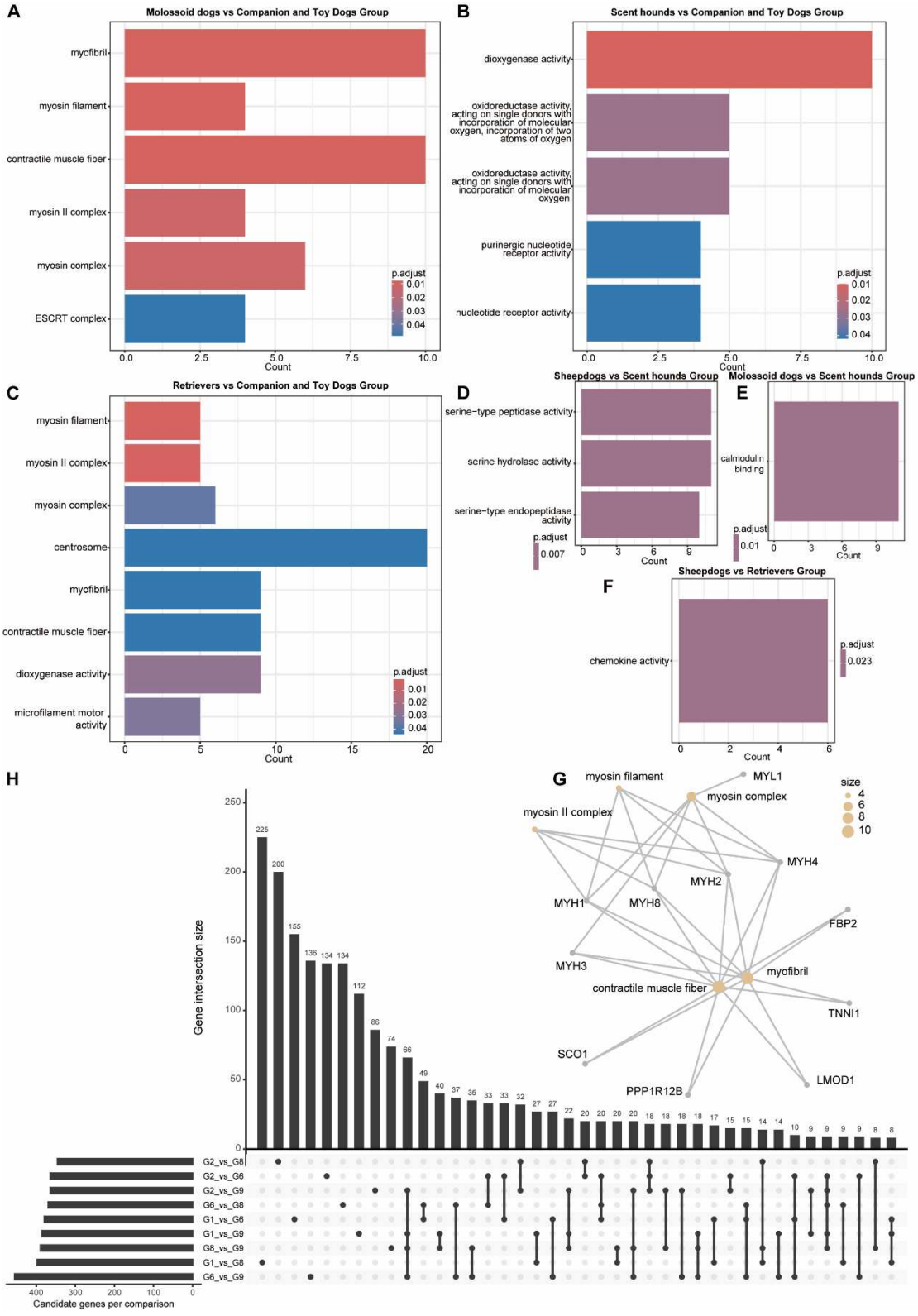


Fig. 3. Functional enrichment and intersection analysis of candidate genes located in highly differentiated genomic regions across FCI breed group comparisons. A to F show significantly enriched Gene Ontology terms for candidate genes identified from pairwise FST comparisons. **A**: Molossoid dogs versus

Companion and Toy Dogs Group. **B**: Scent Hounds versus Companion and Toy Dogs Group. **C**: Retrievers versus Companion and Toy Dogs Group. **D**: Sheepdogs versus Scent Hounds Group. **E**: Molossoid dogs versus Scent Hounds Group. **F**: Sheepdogs versus Retrievers Group. Bar length indicates the number of candidate genes annotated to each GO term, and color indicates the Benjamini Hochberg adjusted P value. **G**: Gene concept network showing relationships between muscle related GO terms and candidate genes detected in the Molossoid dogs versus Companion and Toy Dogs Group comparison. Node size indicates the number of connected genes or GO terms. **H**: UpSet plot showing intersections of candidate genes across the ten pairwise FST comparisons. The vertical bars indicate the number of shared candidate genes in each intersection, and the horizontal bars indicate the total number of candidate genes detected in each comparison. Only GO terms with adjusted P values below 0.05 are shown. FCI Group 1, Sheepdogs and Cattle Dogs; Group 2, Pinscher and Schnauzer, Molossoid and Swiss Mountain and Cattle Dogs; Group 6, Scent Hounds and Related Breeds; Group 8, Retrievers, Flushing Dogs and Water Dogs; Group 9, Companion and Toy Dogs.

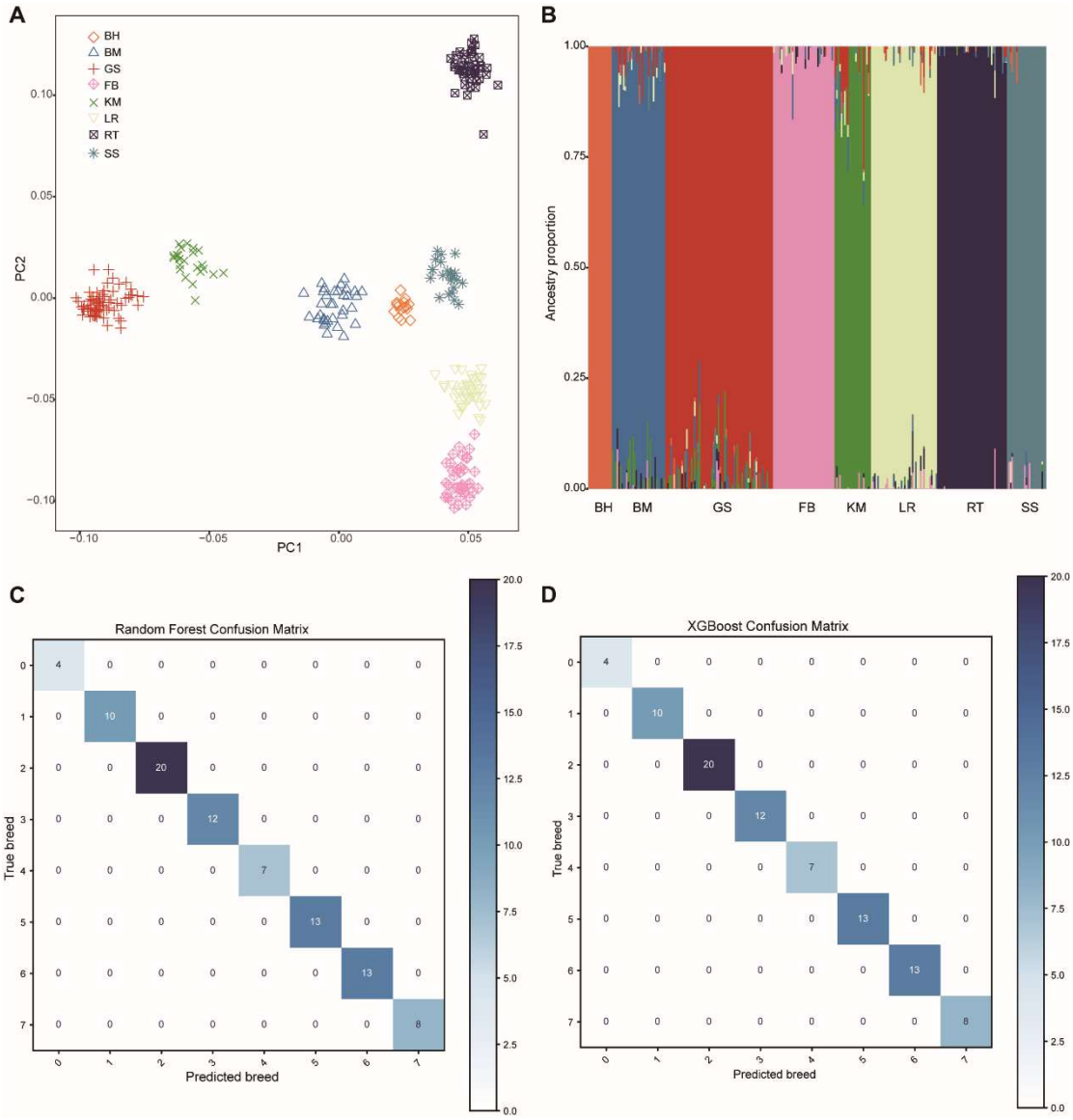


Fig. 4. Evaluation of ancestry informative SNP panels for breed assignment among the eight focal breeds.

A: PCA of eight dog breeds based on 200-SNP panel.

B: ADMIXTURE analysis based on the 200 SNP panel with K = 8

C: Confusion matrix for the Random Forest model based on the 10 SNP panel.

D: Confusion matrix for the XGBoost model based on the 10 SNP panel.

0=BH, 1=BM, 2=GS, 3=FB, 4=KM, 5=LR, 6=RT, 7=SS.

Supplementary Files

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

- Supplementarymaterial.docx
- SupplementaryTable.xlsx