Geographical distribution, genetic diversity, and environmental adaptations of dromedary camel breeds in Saudi Arabia

Mohanad A. Ibrahim[1,2,3], Marco Tolone[4], Mario Barbato[5], Faisal M. Alsubaie[2,6,7], Abdulwahed Fahad Alrefaei[6] and Mikhlid Almutairi[6]*

[1]Genalive Medical Laboratory, Riyadh, Saudi Arabia, [2]Ministry of Environment, Water, and Agriculture (MEWA), Riyadh, Saudi Arabia, [3]Data Science Program, King Abdullah International Medical Research Center, Riyadh, Saudi Arabia, [4]Department of Chemical, Biological, Pharmaceutical, and Environmental Sciences, University of Messina, Messina, Italy, [5]Department of Veterinary Science, Università degli Studi di Messina, Messina, Italy, [6]Zoology Department, College of Science, King Saud University, Riyadh, Saudi Arabia, [7]Genome Department, National Livestock and Fisheries Development Program, Riyadh, Saudi Arabia

The dromedary camel (*Camelus dromedarius*) in Saudi Arabia exhibits significant genetic diversity, driven by adaptation to diverse ecological niches such as deserts, mountains, and coastal areas. This study explores the genetic structure of these camel populations, correlating their genetic diversity with geographical regions rather than ecological classifications. Through whole-genome sequencing of 63 camel genomes, we identified substantial differences in heterozygosity and inbreeding across different ecotypes, particularly noting higher genetic diversity in mountainous populations and lower diversity in coastal populations. The study also revealed significant enrichment of specific gene sets associated with environmental adaptation, such as the HECT domain in desert populations, which is crucial for maintaining protein integrity under extreme conditions. Principal component and admixture analyses further highlighted the genetic distinctiveness of certain breeds, particularly the Awarik (beach ecotype), which showed signs of genetic isolation.

KEYWORDS

SNPs, dromedary camels, genetic diversity, Awarik, Majaheem, deserts

Introduction

The Arabian Peninsula is home to the dromedary camel (*Camelus dromedarius*), which has adapted to various ecological niches, including deserts, mountains, and coastal locations. This diversity of camel populations is noteworthy (1). Saudi Arabia provides an excellent environment for researching these camels' genetic variety and adaptation techniques because of its varied settings (2). According to genetic studies, ecological categories do not correspond as well with the genetic structure of Saudi Arabian camels as geographic areas do. This suggests a complicated connection between genetic variety and environmental adaptability (2). Due to this genetic difference, different breeds in the north, middle, and western areas differ from those in the southwest and southeast (3).

A thorough analysis of the whole genome sequences of dromedaries from the Arabian Peninsula has shown a mostly uniform gene pool with few geographical variations (4). The genetic similarity observed across camel populations can be attributed to historical trade routes and transit activities facilitating gene flow between geographically distant regions. These trade networks allowed the movement and interbreeding of camels across diverse areas,

TABLE 1 Sampling distribution and characteristics across different habitat groups.

Group	Breed	Sample count	Sample location	Sex (M/F)	Age range
Beach	Sahliah	12	Coastal regions near Jeddah and South Jazan	8 M / 4F	Juvenile to Adult
	Awarik	9	Coastal regions near Jeddah and South Jazan	5 M / 4F	Juvenile to Adult
Deserts	Sofor	8	Najad and its eastern areas	4 M / 4F	Juvenile to Subadult
	Shul	10	Najad and its eastern areas	5 M / 5F	Subadult to Adult
	Majaheem	11	Najad and its eastern areas	6 M / 5F	Subadult to Adult
Mountain	Awadi	4	Hijaz mountains and southwestern KSA	2 M / 2F	Juvenile to Subadult
	Haddana	9	Hijaz mountains and southwestern KSA	5 M / 4F	Juvenile to Adult

The table summarizes the distribution of sampled camel breeds across three ecological groups (beach, deserts, and mountain), detailing sample count, geographic location, sex distribution (M/F), and age range.

reducing genetic differentiation. Additionally, the fact that camel owners do not have structured breeding programs has led to unmanaged breeding, which makes genetic traits even more similar. This lack of selective breeding aligns with observations by Al Abri and Faye (5), who noted that historic and ongoing movements of camels for trade and other purposes have significantly shaped their genetic landscape. Although there has been some genetic mixing, there has also been some geographic-associated structuring that has been identified, resulting in the division of camels into three main groups: those from the North, Central, and West and Southwest and Southeast regions of the Arabian Peninsula (6).

In the Arabian Peninsula and beyond, camels have been of paramount importance in the socio-economic progress of human civilizations (7). Primarily domesticated about 3,000–4,000 years ago in the southern Arabian Peninsula, dromedaries enabled transportation and commerce in dry areas, promoting cultural and economic interactions among far-flung populations (8). Camels offer milk, meat, and wool and are adaptable in challenging desert environments. Rendering them highly useful to nomadic and semi-nomadic communities (9). Moreover, camels retain cultural importance, frequently participating in customary festivities and competitions firmly established in Arabian cultures (10).

About 35 million dromedary camels exist worldwide, primarily in the arid and semi-arid regions of Australia, Asia, the Arabian Peninsula, and Africa. Food and Agriculture Organization (11) with around 1.6 million camels distributed throughout different areas and bred for racing, milk production, and meat, Saudi Arabia has a sizable camel population, which is indicative of its strong tradition in camel husbandry and breeding (2). The nation has fourteen recognized camel breeds, each suited to a specific type of climate. The north and central parts of the Arabian Peninsula are home to desert breeds such as Magaheem, Wodeh, Sofor, and Shual (12). While beach breeds like Sahlia and Awarik are located in the west and southwest coastal regions, mountain breeds like Hadana and Awadi live in the western and southwestern mountainous areas (13).

This study's principal aim is to examine the genetic links across Saudi camel populations, particularly emphasizing groupings defined by geographical and environmental features, such as coastal, mountainous, and desert environments. To preserve and enhance this species, it is crucial to comprehend the genetic variety and organization of these many populations. For camel populations to be resilient and sustainable in the face of environmental and climatic difficulties, breeding plans should be based on insights from these kinds of studies (14).

Materials and methods

Sample collection

The camels studied in this work were sampled from three distinct ecological environments: coastal, desert, and mountainous regions (Table 1, Figure 1). Based on their origins and breed names—historically documented and officially recorded by the Ministry of Agriculture and Environment—they were categorized into specific groupings. The coastal group included the "Sahliah" and "Awarik" breeds, primarily found near the Jeddah and South Jazan coastlines. The desert group comprised the "Sofor," "Shul," and "Majaheem" breeds, which have evolved to endure the harsh desert climates of Najad and Riyadh. Lastly, the "Awadi" and "Haddana" breeds, well-adapted to the rugged terrain of the Hijaz Mountains in southwestern Saudi Arabia, comprised the mountain group. Each breed represents a small, proportional part of the overall subpopulation within its respective ecological zone. Sampling was carried out on several farms, and individuals were selected based on information supplied by the farmers to avoid, as much as possible, closely related animals.

DNA extraction and sequencing

Following Wu et al. (15), we conducted blood collection and extracted genomic DNA using a Maxwell system (Promega, United States) with Maxwell RSC cartridges. We quantified the DNA concentration using the QuantiFluor dsDNA System and a Quantus Fluorometer and confirmed its purity and integrity through gel electrophoresis. Sampling was carried out on several farms, and individuals were selected based on information provided by farmers to minimize the inclusion of closely related animals. The Beijing Genomics Institute (BGI) performed whole-genome sequencing (WGS), fragmenting DNA into approximately 300 bp fragments using Covaris technology, followed by end repair, addition of an 'A' base to the 3′ ends, and adapter ligation. DNA nanoballs (DNBs) were generated from amplified libraries using rolling circle amplification (RCA), then loaded onto patterned nanoarrays and sequenced through the ILLUMINA NovaSeq 6,000 platform. High-throughput paired-end reads were generated with an average sequencing depth of 30× per genome. We evaluated the sequencing quality using FastQC v0.12.0 (16) (see Supplementary Table S1 for a detailed summary of the metrics).

FIGURE 1
Map of Saudi Arabia. Red dots mark sampling locations. The labeled points represent key regions: Jeddah (Sahliah), Jazan (Awarik), Hijaz (Awadi, Haddana), and Najd (Majaheem, Shul, and Sofor).

Data analysis

The sequencing workflow, as illustrated in Supplementary Figure S1, began with assessing data quality using FastQC v0.12.0 (16) to detect adapter contamination and low-quality nucleotides. Data cleaning was performed using SOAPnuke (17), configured for paired-end read cleaning. Reads were removed if they contained over 50% adapter sequences, more than 50% of bases with a Phred quality score below 20, or at least 2% ambiguous bases ("N"). Cleaned reads were aligned to the CamDro3 reference genome (NCBI Assembly)[1] using BWA v0.7.18 (18), with the reference genome indexed using the bwa index command to enhance mapping efficiency. Alignments were stored in sequence alignment/map (SAM) format for subsequent analysis. Using SAMtools (19, 20), SAM files were converted into binary alignment/map (BAM) files sorted by genomic coordinates. Additional processing included marking shorter splits as secondary alignments (-M), enabling YML format output (-Y), and specifying read group information (-R) to ensure compatibility with downstream analyses.

Variant calling was performed using GATK HaplotypeCaller v4.5.0.0 (21) to identify high-confidence variants. The analysis applied specific criteria, including a minimum mapping quality of 20, a maximum of 6 alternative alleles, and a minimum base quality score of 20. These thresholds were based on the default recommendations from the GATK Best Practices pipeline (22). Variants were further filtered using stringent thresholds to exclude low-quality or potentially erroneous calls: quality by depth (QD) scores below 2.0, quality scores (QUAL) below 40.0, Fisher Strand (FS) values above 60.0, mapping quality (MQ) below 40.0, mapping quality rank sum (MQRankSum) below-12.5, and read position rank sum (ReadPosRankSum) below-8.0. These thresholds reflect widely accepted standards for distinguishing true variants from sequencing artifacts, as outlined in the GATK supported by Bahbahani et al. (23). To streamline processing, GVCF files generated for each sample were combined into a single consolidated file using GATK CombineGVCFs. The resulting file was then converted into VCF format using GATK GenotypeGVCFs, ensuring compatibility with downstream analyses, including those conducted using PLINK and other tools.

The discovered single nucleotide polymorphisms (SNPs) were refined for quality control using PLINK v1.9 software (24, 25). Variants with incomplete genotype data were eliminated by establishing a threshold to exclude SNPs with significant missing data (>0.05) with the --geno flag. A filter was subsequently applied based on Minor Allele Frequency (MAF) utilizing the --maf flag to select prevalent and informative SNPs by excluding less common variants (MAF <0.05). Linkage disequilibrium (LD) pruning was executed using the --indep-pairwise command with the settings (2000 k 1 0.5)

---

1 https://www.ncbi.nlm.nih.gov/assembly/GCF_000803125.2

to eliminate tightly connected SNPs. The selection of a 2000 kb window size and a r^2 threshold of 0.5 was determined by camel genomics, characterized by large linkage disequilibrium blocks, in unlike humans. Similar techniques have been utilized in livestock species, including cattle, where bigger LD blocks necessitate appropriately sized windows (26). Furthermore, the --chr parameter focused exclusively on autosomal chromosomes, so excluding non-autosomal variants from influencing the results. The --rel-cutoff flag was employed to eliminate closely related samples, adhering to the relatedness estimation methodology established by Purcell et al. (25) and Manichaikul et al. (27) in PLINK.

To evaluate genetic diversity, the level of genetic variation within individuals was determined by calculating observed heterozygosity (H_o) through a specific command (--het) in PLINK v1.9 (24, 25). This calculation assesses the overall genetic diversity by comparing the actual number of homozygous genotypes (O(HOM)) with the expected number (E(HOM)) in each individual, helping to estimate the inbreeding coefficient (F).

Runs of Homozygosity (ROH) were analyzed using the --homozyg command in PLINK. ROH detection relied on specific criteria, including the minimum number of SNPs, genomic density, and allowable gaps, to identify segments of homozygosity. ROHs were then categorized into three classes: small (<0.1 Mbp), medium (0.1–5 Mbp), and large (>5 Mbp), following thresholds widely used in livestock genetics (28, 29). These classification thresholds align with common patterns of genetic associations and inbreeding, making them suitable for camels given their genome size and linkage disequilibrium traits (26).

Effective population size (N_e) was estimated with SNeP v1.11 (30), incorporating default settings and specific adjustments, including sample size correction for unphased genotypes, mutation rate correction, and the Sved and Feldman (31) mutation rate modifier.

The analysis of population structure involved the utilization of two complementary methodologies to gain a comprehensive understanding. Principal Component Analysis (PCA) was conducted using PLINK v1.9 to identify the main genetic variation trends among different populations. Furthermore, a model-based clustering method was employed with ADMIXTURE v1.3.0 (32) to ascertain the probable count of ancestral populations (K). ADMIXTURE was run for K values ranging from 2 to 5, with cross-validation to identify the optimal K based on prediction error.

Selection signatures were identified using the XP-nSL statistic implemented in Selscan v2.0.2 (33, 34) through pairwise comparisons between camel groups using phased data without setting the alternate flag. The analysis was configured with a scale parameter of 20,000 for normalizing extended haplotype homozygosity (EHH) decay curves, a maximum gap parameter of 200,000 to account for the camel genome's larger LD blocks and recombination rates, and an EHH cutoff of 0.05 to reduce false positives while maintaining sensitivity. These parameters were chosen based on Selscan documentation and their effectiveness in species with similar genomic characteristics (35). To pinpoint important SNPs, XP-nSL scores were standardized, and variants with z-scores surpassing the critical value for a combined two-sided significance level of 0.05 were chosen. Extended selection windows were crafted by determining the average gap between adjacent SNPs and extending half of this distance on both sides to precisely depict genomic regions undergoing selection.

To annotate the selection windows, we used a reference GFF file to identify overlapping genes. The dataset was divided into three ecological groups and six populations. Genes were retained based on specific criteria to ensure robust and biologically meaningful results. Specifically, we isolated genes consistently under selection across all populations within an ecological group and retained only those present in every population of the group. Additionally, we extracted unique genes exclusive to one group but absent in the others, emphasizing ecological and population-specific adaptations. For functional analysis, significant Gene Ontology (G.O.) terms and pathways were identified using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) (36), which provided insights into the biological functions and processes associated with the selected genes. These criteria ensured that the annotation process focused on genes with consistent and ecologically relevant selection signals.

Results

Data analysis

Sequencing of 63 camel genomes generated ~385 million reads (~57 billion bases). After applying quality refinement filters, including the removal of reads with >50% bases below a Phred quality score of 20, the dataset was consolidated to 373 million high-quality reads (~56 billion bases), resulting in a minimal data loss of 2.95%, which did not compromise downstream analyses. The refined dataset exhibited an average G.C. content of 42.40% and a Q30 score of 97.07% across all samples, indicating high sequencing accuracy and reliability (Supplementary Table S1).

Using BWA v0.7.18 for alignment, the reads were mapped to the CamDro3 reference genome, achieving an average mapping rate of 99.81% and a correctly paired read ratio of 99.69%, underscoring the dataset's high reliability. The average sequencing depth was 21.5X, meaning the genome was covered more than five times in 92.87% of cases and at least one-fold in 93.74%. For more information, see Supplementary Table S2, which has detailed sequencing metrics for each sample.

Further analysis identified an average of 2,959,240 single nucleotide polymorphisms (SNPs) through variant calling, with a distribution of 2,870,314 homozygous SNPs and 88,927 heterozygous SNPs. This distribution underscores a significant prevalence of homozygous SNPs, constituting about 97.04% of the total SNPs, while heterozygous SNPs account for approximately 2.96%. Additionally, the transition-to-transversion ratio of nucleotide substitutions averaged 1.7027, highlighting the genetic variation within the sampled camel populations as depicted in Supplementary Table S3.

The initial dataset consisted of 63 samples and 12,932,844 variants. After the first stage of quality control, which involved filtering out variants with missing genotype data, the dataset was reduced to 4,502,956 variants, maintaining all 63 samples and achieving a genotyping rate of 0.63. Subsequent application of a minor allele frequency (MAF) filter further refined the dataset by removing 425,847 variants, resulting in 4,077,109 variants with a genotyping rate of ~1. Linkage disequilibrium pruning left 392,139 loci. Selecting autosomes exclusively further narrowed it down to 367,871 loci. Finally, removing related individuals led to a final

dataset comprising 34 samples and 367,871 variants, ensuring high-quality data for subsequent analyses. In summary, see Figure 2.

Genetic diversity

AWA

The H_o Table 2 across the populations ranges from 0.374 in the AWA population to 0.414 in the HAD population, indicating varying levels of genetic diversity. The HAD population shows the highest H_o at 0.414, while the AWA population exhibits the lowest at 0.374. The other populations, including MAJ, SAH, SHU, and SOF, have H_o values between 0.396 and 0.405.

The effective population size N_e varies significantly, with the MAJ population having the highest N_e of 37, followed by SAH with 24 and SOF with 23. The HAD population has the smallest N_e of 11, while the AWA and SHU populations have N_e values of 15 and 17, respectively.

The F_{ROH} values are generally low across all populations, with the AWA population having the highest value of 0.015 and the SAH and SOF populations showing the lowest values at 0.000. The remaining populations have values between 0.002 and 0.004.

The Principal Component Analysis (PCA) plot displays the genetic variation among the populations, with PC1 explaining 7.7% of the variance and PC2 explaining 5.6% (Figure 3). The plot reveals distinct clustering patterns: The desert populations (indicated by circles) are aligned along a single line on the right side of the plot, showing tight clustering along PC1. The mountain populations (indicated by triangles) cluster closely together along the positive side of PC2. In contrast, the beach populations (indicated by diamonds) are split into two groups: one group clusters closely with the mountain populations above, while the other group is positioned separately along the negative side of PC1, indicating distinct genetic differentiation within the beach ecotype.

The admixture analysis for K values ranging from 2 to 5 reveals the following genetic structure among the populations (Figure 4). At the AWA, the population is distinctly separated from the others, with less than 10% of the AWA ancestral component appearing in SAH and HAD and a negligible presence in MAJ, SHU, and SOF. At K = 3, SHU acquires a private cluster is shared with SOF and partially with MAJ. At K = 4, a fourth ancestral cluster emerges, further diversifying HAD and MAJ. Finally, the genetic structure becomes more complex, with additional ancestral components present across all populations.
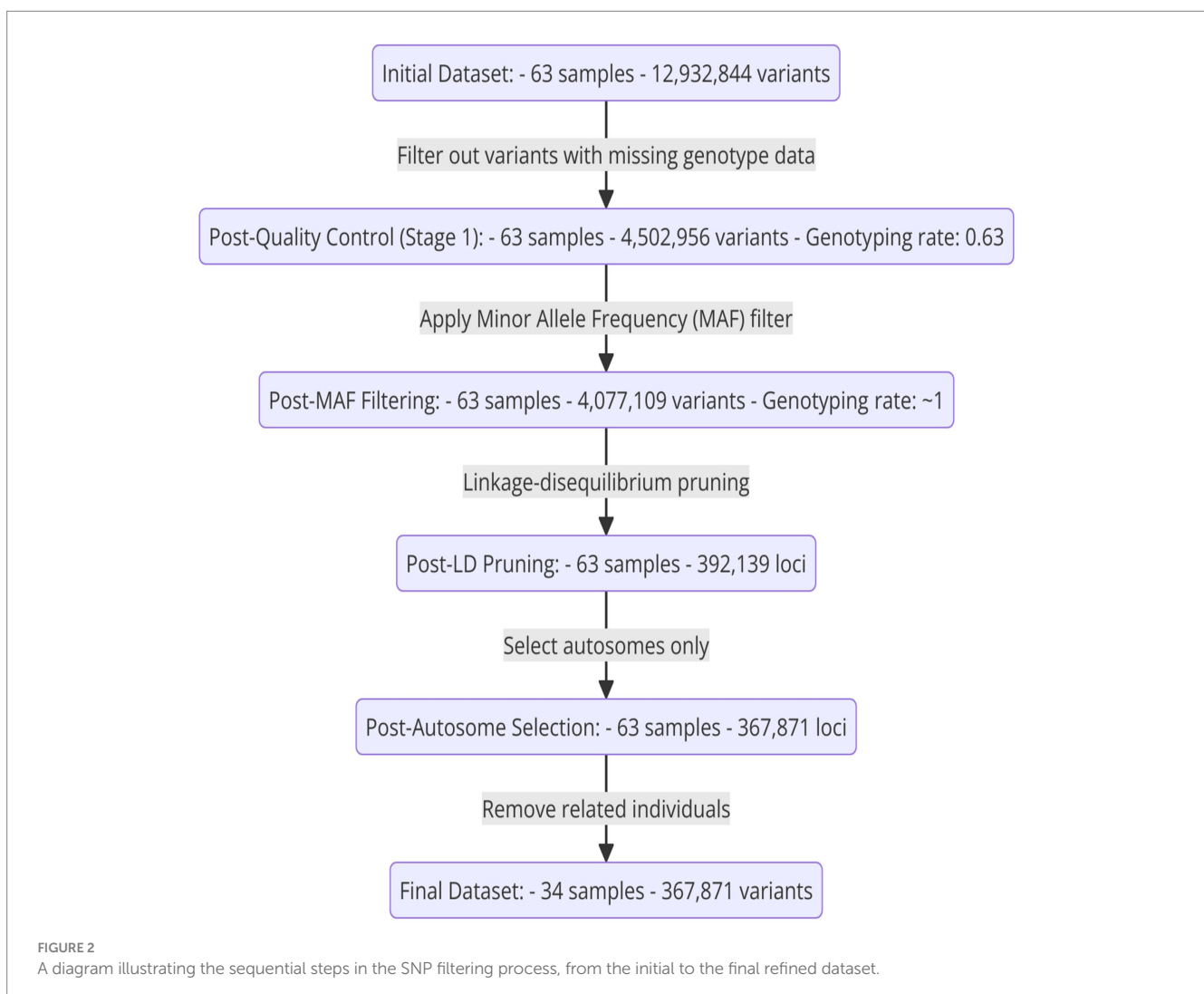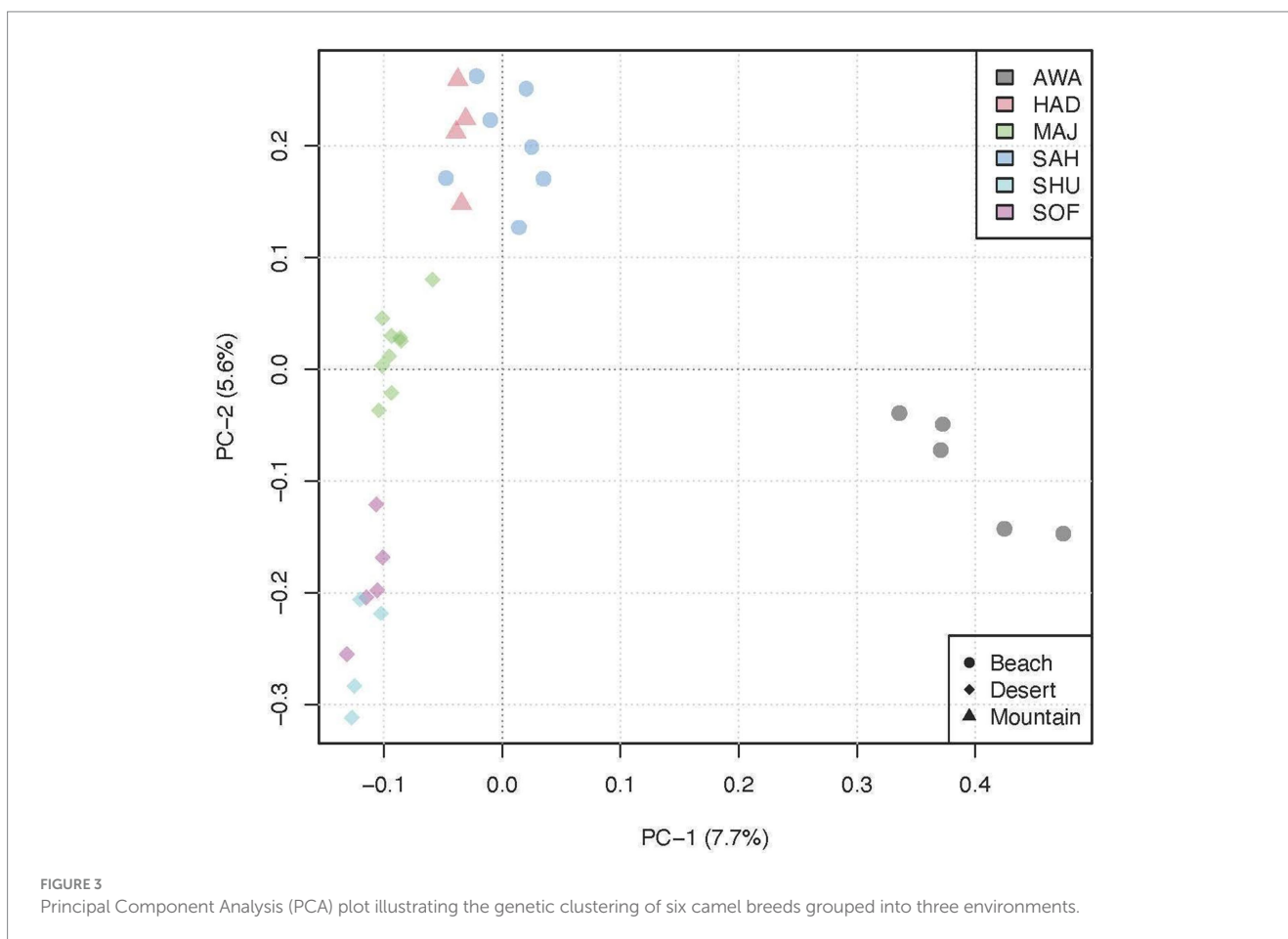

FIGURE 2
A diagram illustrating the sequential steps in the SNP filtering process, from the initial to the final refined dataset.

TABLE 2 Data on six camel populations including their acronyms, ecological groups (beach, mountain, and deserts), sample size N, observed heterozygosity H_o with standard deviation S.D., genomic inbreeding based on the proportion of the genome in runs of homozygosity F_{ROH} with S.D.; and effective population size N_e.

Population	Acronym	Ecotype	N	H_o (S.D.)	F_{ROH} (S.D.)	N_e
Awarik	AWA	Beach	5	0.374 (0.03)	0.015 (0.01)	15
Haddana	HAD	Mountain	4	0.414 (0.00)	0.003 (0.00)	11
Majaheem	MAJ	Deserts	9	0.400 (0.01)	0.002 (0.00)	37
Sahliah	SAH	Beach	7	0.405 (0.01)	0.000 (0.00)	24
Shul	SHU	Deserts	4	0.396 (0.01)	0.004 (0.01)	17
Sofor	SOF	Deserts	5	0.401 (0.01)	0.000 (0.00)	23

FIGURE 3
Principal Component Analysis (PCA) plot illustrating the genetic clustering of six camel breeds grouped into three environments.

The analysis of ROH across different camel ecotypes reveals distinct patterns in the size categories of ROHs (Table 3; Figure 5)—classified as large (greater than 5 Mb), medium (2–5 Mb), and small (1–2 Mb). For the beach ecotype, the data shows 3 large ROHs totaling 15.76 Mb, 113 medium ROHs totaling 295.96 Mb, and 468 small ROHs with a total length of 646.34 Mb. In the desert ecotype, there are 2 large ROHs with a combined length of 12.94 Mb, 52 medium ROHs totaling 134.97 Mb, and 271 small ROHs accumulating to 371.00 Mb. The mountain ecotype exhibits three large ROHs. Totaling 15.76 Mb, 115 medium ROHs with a total length of 299.59 Mb, and 431 small ROHs summing to 595.75 Mb.

Selection signatures were identified using the XP-nSL statistic, focusing on pairwise comparisons between camel groups. This method, which analyzes phased data to detect extended haplotype homozygosity, revealed genetic signals across the beach, deserts, and mountain camel populations. The results indicated that the beach group harbored 105 unique genes, the desert group had 30 unique genes, and the mountain group exhibited the most extensive selection, with 322 unique genes (Figure 6).

In addition to these unique signatures, there were also shared selection signals: 89 genes were common between the beach and mountain groups, 23 between the beach and desert groups, 63 between the desert and mountain groups, and 66 genes were shared across all three ecotypes.

The enrichment analysis, as detailed in Table 4, identified 21 genes in the deserts group with key functional annotations, including the
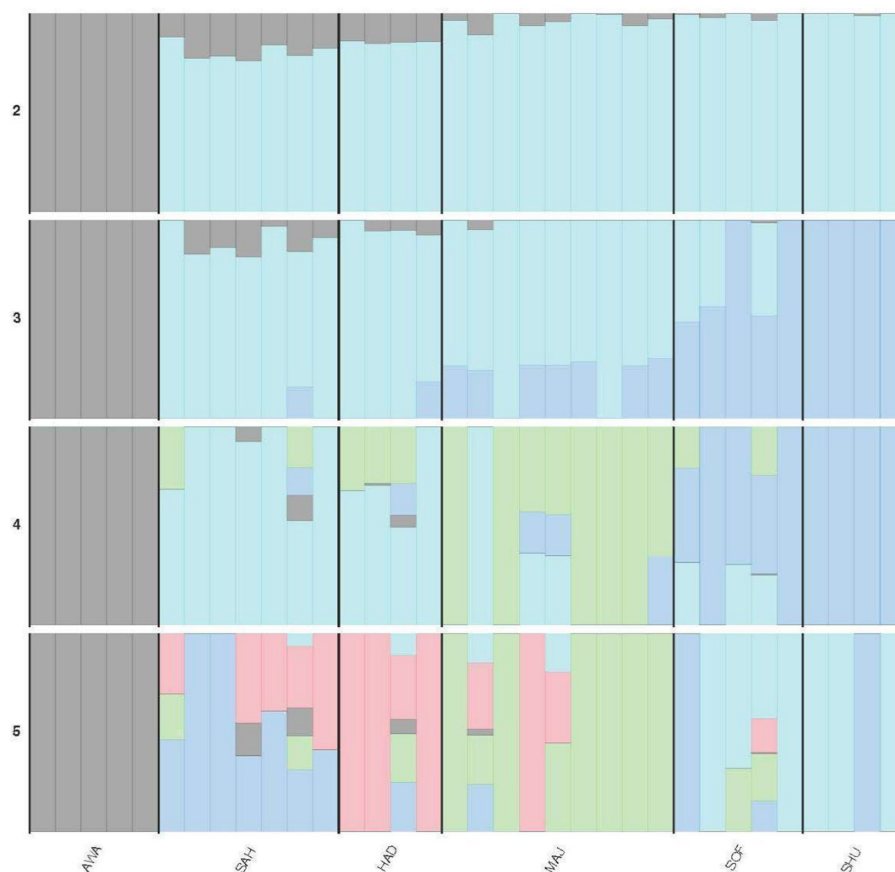
FIGURE 4
Admixture plot showing the genetic composition of six camel populations across K-values (K = 2 to 5). See Table 2 for population acronyms.

TABLE 3 Comparison of runs of homozygosity (ROH) categories across camel ecotypes.

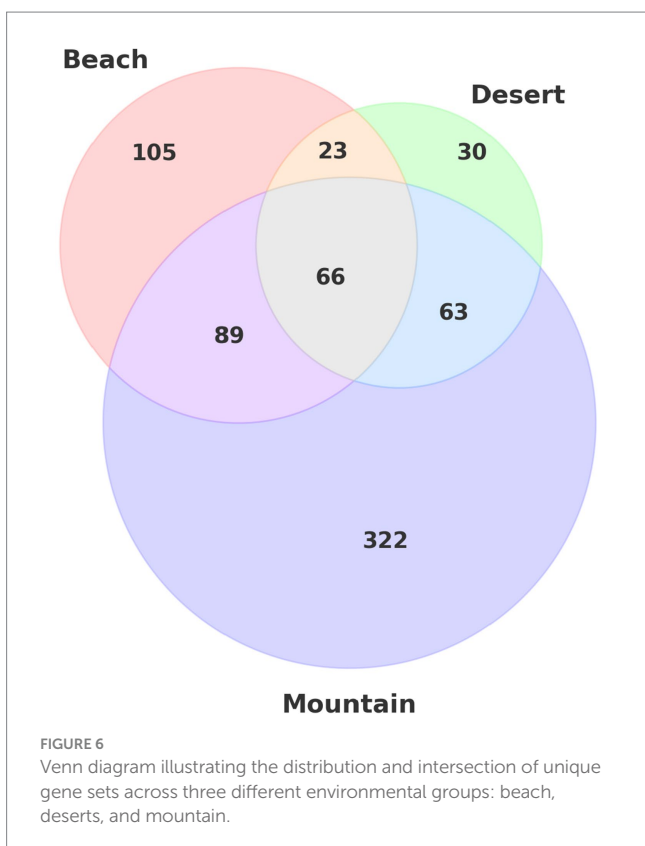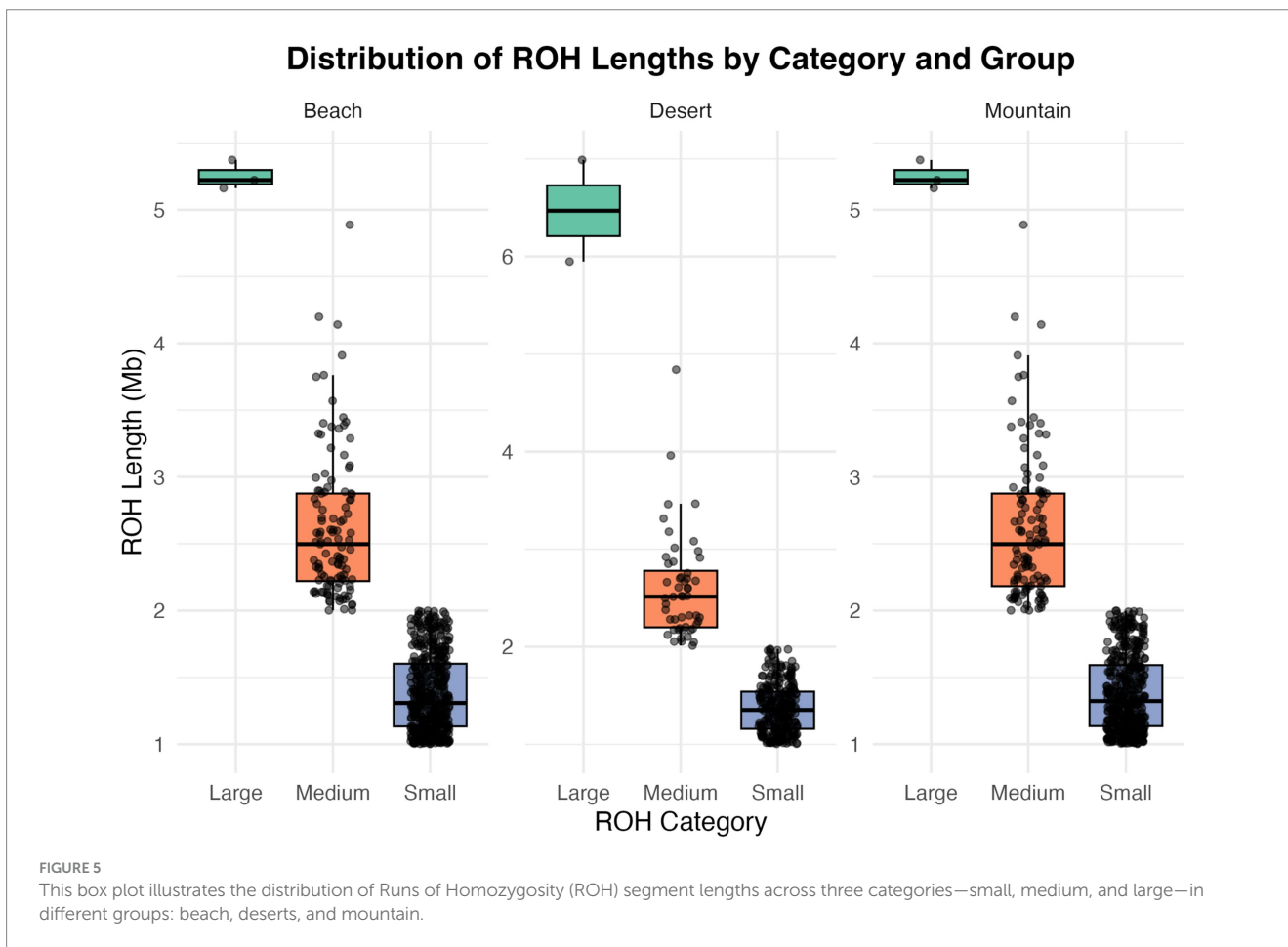| ROH category	Beach (Count)	Deserts (Count)	Mountain (Count)
Large	3	2	3
Medium	113	52	115
Small	468	271	431

HECT domain, basic and acidic residues, and disordered regions. This resulted in an enrichment score of 0.11 and significant *p*-values of 0.024 and 0.026. Table 5 highlights that the top enriched pathway in this group is Ubiquitin-ligase activity, with key genes such as HECTD4 and TMED5 noted in Table 6.

In the beach group, as shown in Table 4, 90 genes were identified with key annotations, such as the P.H. domain, HECT domain, and thyroid hormone synthesis, yielding an enrichment score of 1.864. Significant p-values for these annotations include 0.009 and 0.030 for the P.H. domain and 0.024 for the HECT domain. Table 5 highlights that the top enriched pathways in this group are Phosphatidylinositol signaling and thyroid hormone synthesis, with example genes like DAGLB and TMEM70 listed in Table 6.

The mountain group, as detailed in Table 4, revealed 274 genes with key annotations, including the B30.2/SPRY domain, HECT domain, and transferase activity, achieving the highest enrichment score of 2.655. Significant p-values include 0.000345 for the B30.2/SPRY domain, 0.000410 for the SPRY domain, and 0.024 for the HECT domain. Table 5 indicates that the top enriched pathways in this group are related to immune response and metabolic adaptations, with critical genes such as ITGB1, KIF3B, and TMEM70 highlighted in Table 6.

Discussion

Saudi Arabian dromedary camels vary genetically due to environmental constraints, breeding methods, and trading routes. These factors have changed camel genetics, improving their resilience and performance in harsh conditions. The genetic diversity of these populations is mostly uniform, although ecological and environmental changes have created geographic structure (6, 23, 37). Our dataset was identified in this study's methodology. After alignment and purification, 99.81% of reads were mapped to the reference genome. After comprehensive quality control and filtering, 34 samples and 367,871 variations

FIGURE 5
This box plot illustrates the distribution of Runs of Homozygosity (ROH) segment lengths across three categories—small, medium, and large—in different groups: beach, deserts, and mountain.

FIGURE 6
Venn diagram illustrating the distribution and intersection of unique gene sets across three different environmental groups: beach, deserts, and mountain.

were left, ensuring high-quality data for additional investigations (Supplementary Tables S1–S3).

Our study revealed significant genetic patterns across different camel populations, highlighting variations in heterozygosity, effective population sizes, and selection signatures. Principal Component Analysis (PCA, Figure 3) showed a clear genetic separation between desert breeds (MAJ, SHU, and SOF), which clustered closely on the first two principal components (PC-1 and PC-2), and beach breeds (SAH and AWA), which formed distinct clusters. Notably, the AWA breed emerged as an independent cluster on PC-1, likely due to higher levels of inbreeding and genetic drift. This observation is supported by its low heterozygosity ($H_o = 0.374$) and the highest inbreeding coefficient ($F_{ROH} = 0.015$) among all populations. These patterns may be influenced by geographic isolation or intentional breeding practices (6).

Admixture analysis (Figure 3) corroborates the PCA findings, revealing significant genetic mixing among breeds, such as SAH (Beach) and HAD (Mountain), suggesting historical gene flow or interbreeding. The K = 3 scenario revealed a desert-dependent genetic signature represented by the darker blue component, highlighting the unique genetic architecture of desert breeds. In contrast, the distinctiveness of AWA likely reflects its limited admixture and preservation of unique genetic traits, aligning with similar patterns observed in Omani dromedaries (23).

Runs of Homozygosity (ROH, Table 3; Figure 5) provided further insights into genetic architecture. Small ROHs (1–2 Mb), indicative of genetic drift and historical bottlenecks, dominated

TABLE 4 The table summarizes the recognized genes, key functional annotations, enrichment scores, significant *p*-values, and example genes for the deserts, beach, and mountain groups.

Group	Recognized genes	Key functional annotations	Enrichment score (key cluster)	Significant *p*-values	Example genes
Deserts	21	HECT domain, basic and acidic residues, disordered	0.11	0.024 (HECT domain), 0.026 (HECT domain)	HECTD4, TMED5
Beach	90	P.H. domain, HECT domain, thyroid hormone synthesis	1.864	0.009 (P.H. domain), 0.030 (P.H. domain), 0.024 (HECT domain)	DAGLB, TMEM70
Mountain	274	B30.2/SPRY domain, HECT domain, transferase activity	2.655	0.000345 (B30.2/SPRY domain), 0.000410 (SPRY_dom), 0.024 (HECT domain)	ITGB1, KIF3B

TABLE 5 The top enriched pathways for the deserts, beach, and mountain groups and the corresponding *p*-values.

Group	Top enriched pathways	*p*-values
Deserts	Ubiquitin-ligase activity	0.024 (HECT domain), 0.026 (HECT domain)
Beach	Phosphatidylinositol signaling, thyroid hormone synthesis	0.009 (P.H. domain), 0.033 (Thyroid hormone synthesis), 0.024 (HECT domain)
Mountain	Immune response (B30.2/SPRY domains), metabolic adaptations	0.000345 (B30.2/SPRY domain), 0.000410 (SPRY domain), 0.024 (HECT domain)

TABLE 6 The table highlights specific genes, their functions, and their significance within the deserts, beach, and mountain ecotypes.

Group	Gene	Function	Significance
Deserts	HECTD4	Ubiquitin-ligase activity	Essential for protein maintenance in extreme temperatures
Beach	TMED5	Protein trafficking within the cell	Vital for maintaining cellular functions under stress.
	DAGLB	Endocannabinoid biosynthesis	Manages stress from varying salinity and nutrient availability
	TMEM70	Mitochondrial function and energy metabolism	Critical for adapting to the dynamic energy demands of coastal environments
Mountain	ITGB1	Cell adhesion and signal transduction	Adaptations to the physical demands of mountainous terrain
	KIF3B	Intracellular transport along microtubules	Supports cellular functions under high-altitude stress

in the AWA population. Medium ROHs (2–5 Mb), linked to recent inbreeding or population substructure, were prevalent in HAD, while large ROHs (>5 Mb), indicative of very recent inbreeding, were distributed more evenly across beach and mountain populations but less so in desert breeds. These patterns reflect both ecological constraints and historical breeding practices.

The extremely low N_e values across populations emphasize their vulnerability to genetic diversity loss, with the MAJ population demonstrating the highest N_e (37), likely due to its broader distribution and higher gene flow. These findings align with previous research (6), emphasizing the role of historical admixture and environmental adaptation in shaping genetic structures.

Gene set enrichment and over-representation analyses highlight functional pathways critical for adaptation across ecotypes. For example, HECT family E3 ubiquitin ligases, prominent in desert camels, are pivotal for protein stability under oxidative stress, a condition prevalent in arid environments (38–43). Similarly, DAGLB and TMEM70, identified in beach ecotypes, play critical roles in lipid signaling and mitochondrial function, enabling adaptation to fluctuating salinity and nutrient availability (44–46). Immune response genes such as CX3CR1, IL6R, and CCR8, identified in desert camels, highlight

adaptations to pathogen-rich environments (47, 48). Similarly, genes linked to fertility, such as ESR1 and SPACA5, emphasize the reproductive resilience of camels in harsh climates, paralleling findings in African zebu cattle (49, 50).

Energy metabolism genes, such as ESRRG and CRTC1, were identified as critical for energy homeostasis in desert and racing camels, underscoring the importance of efficient metabolic regulation for survival and stamina (51, 52). Chondrogenesis-related genes, including CHSY1 and CRLF1, reflect the physical demands placed on camels historically used for transport and racing, highlighting their role in skeletal adaptation (53, 54).

Genes associated with milk production (PICALM) and running performance (NAA16) further illustrate selective pressures shaped by cultural and economic practices (55, 56). These findings echo evolutionary pressures observed in cattle and other domesticated livestock, where production and performance traits have undergone significant selection.

Our findings highlight the intricate interplay between ecological and anthropogenic factors in shaping the genetic structure of camel populations. The distinct genetic profiles observed across ecotypes reflect adaptations to environmental pressures and historical breeding practices. Functional insights into key genes provide a deeper understanding of the mechanisms underlying resilience and

performance in dromedary camels. These results, in line with prior studies ([15], [23]), underscore the importance of integrating genomic data with ecological and physiological studies to inform conservation and breeding strategies.

Data availability statement

The Genomic data have been deposited in NCBI repository, accession number PRJNA1219399. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1219399.

Ethics statement

The animal study protocol was approved by the Research Ethics Committee of King Saud University, Riyadh, Saudi Arabia (Ethics Reference No. KSU-SE-23-93, 19/10/2023). The study was conducted in accordance with the local legislation and institutional requirements.

Author contributions

MI: Conceptualization, Formal analysis, Writing – original draft, Writing – review & editing. MT: Investigation, Supervision, Writing – review & editing. MB: Methodology, Supervision, Validation, Writing – review & editing. AA: Project administration, Resources, Writing – review & editing. MA: Conceptualization, Data curation, Resources, Writing – original draft, Project administration. YA: Investigation, Resources, Writing – review & editing. FA: Conceptualization, Funding acquisition, Resources, Writing – review & editing.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2024.1490186/full#supplemetary-material

References

1. Burger PA, Ciani E, Faye B. Old World camels in a modern world – a balancing act between conservation and genetic improvement. *Anim Genet*. (2019) 50. doi: 10.1111/age.12858

2. Hossam Mahmoud A, Mohammed Abu-Tarbush F, Alshaik M, Aljumaah R, Saleh A. Genetic diversity and population genetic structure of six dromedary camel (*camelus dromedarius*) populations in Saudi Arabia. *Saudi J Biol Sci*. (2020) 27. doi: 10.1016/j.sjbs.2019.11.041

3. Al-Swailem AM, Al-Busadah KA, Shehata MM, Al-Anazi IO, Askari E. Classification of Saudi Arabian camel (*Camelus dromedarius*) subtypes based on RAPD technique. *J Food Agric Environ*. (2007) 5. doi: 10.1234/4.2007.749

4. Khalkhali-Evrigh R, Hafezian SH, Hedayat-Evrigh N, Farhadi A, Bakhtiarizadeh MR. Genetic variants analysis of three dromedary camels using whole genome sequencing data. *PLoS One*. (2018) 13. doi: 10.1371/journal.pone.0204028

5. Al Abri MA, Faye B. Genetic improvement in dromedary camels: challenges and opportunities. *Front Genet*. (2019) 10. doi: 10.3389/fgene.2019.00167

6. Almathen F, Bahbahani H, Elbir H, Alfattah M, Sheikh A, Hanotte O. Genetic structure of Arabian Peninsula dromedary camels revealed three geographic groups. *Saudi J Biol Sci*. (2022) 29. doi: 10.1016/j.sjbs.2021.11.032

7. Burger PA. The history of Old World camelids in the light of molecular genetics. *Trop Anim Health Prod*. (2016) 48. doi: 10.1007/s11250-016-1032-7

8. Legesse YW, Dunn CD, Mauldin MR, Ordonez-Garza N, Rowden GR, Gebre YM, et al. Morphometric and genetic variation in 8 breeds of Ethiopian camels (*Camelus dromedarius*). *J Anim Sci*. (2018) 96. doi: 10.1093/jas/sky351

9. Ahmad S, Yaqoob M, Hashmi N, Ahmad S, Zaman MA, Tariq M. Economic importance of camel: A unique alternative under crisis. *Pak Vet J*. (2010) 30

10. Khalaf S. Camel racing in the Gulf notes on the evolution of a traditional cultural sport. *Anthropos*. (1999) 94

11. Food and Agriculture Organization. (2018). *FAOSTAT*.

12. Mahmoud AH, Abou-tarboush FM, Alshaikh MA, Aljumaah RS. Genetic characterization and bottleneck analysis of Maghateer camel population in Saudi Arabia using microsatellite markers. *J King Saud Univ Sci*. (2020) 32. doi: 10.1016/j.jksus.2019.11.027

13. Harek D, Ikhlrf H, Bouhadad R, Sahel H, Cherifi Y, Djellout NE, et al. Genetic diversity status of camel's resources (*Camelus Dromedarius*. Linnaeus, 1758) in Algeria. *GABJ*. (2017) 1. doi: 10.46325/gabj.v1i1.87

14. Obšteter J, Jenko J, Pocrnic I, Gorjanc G. Investigating the benefits and perils of importing genetic material in small cattle breeding programs via simulation. *J Dairy Sci*. (2023) 106. doi: 10.3168/jds.2022-23132

15. Wu H, Guang X, Al-Fageeh MB, Cao J, Pan S, Zhou H, et al. Camelid genomes reveal evolution and adaptation to deserts environments. *Nat Commun*. (2014) 5. doi: 10.1038/ncomms6188

16. Andrews S. (2010). FastQC. *Babraham Bioinformatics*.

17. Chen Y, Chen Y, Shi C, Huang Z, Zhang Y, Li S, et al. SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *GigaScience*. (2018) 7. doi: 10.1093/gigascience/gix120

18. Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*. (2009) 25. doi: 10.1093/bioinformatics/btp324

19. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. *GigaScience*. (2021) 10. doi: 10.1093/gigascience/giab008

20. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics*. (2009) 25. doi: 10.1093/bioinformatics/btp352

21. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. (2010) 20. doi: 10.1101/gr.107524.110

22. Broad Institute. (n.d.). HaplotypeCaller. GATK best practices. Available at: https://gatk.broadinstitute.org/hc/en-us/articles/360037225632-HaplotypeCaller (accessed December 12, 2024)

23. Bahbahani H, Musa HH, Wragg D, Shuiep ES, Almathen F, Hanotte O. Genome diversity and signatures of selection for production and performance traits in dromedary camels. *Front Genet*. (2019) 10. doi: 10.3389/fgene.2019.00893

24. Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*. (2015) 4. doi: 10.1186/s13742-015-0047-8

25. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. (2007) 81. doi: 10.1086/519795

26. El Hou A, Rocha D, Venot E, Blanquet V, Philippe R. Long-range linkage disequilibrium in French beef cattle breeds. *Genet Sel Evol*. (2021) 53:63. doi: 10.1186/s12711-021-00657-8

27. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. *Bioinformatics*. (2010) 26:2867–73. doi: 10.1093/bioinformatics/btq559

28. Ferenčaković M, Sölkner J, Curik I. Estimating autozygosity from high-throughput information: effects of SNP density and genotyping errors. *Genet Sel Evol*. (2013) 45:42. doi: 10.1186/1297-9686-45-42

29. Purfield DC, Berry DP, McParland S, Bradley DG. Runs of homozygosity and population history in cattle. *BMC Genet*. (2012) 13:70. doi: 10.1186/1471-2156-13-70

30. Barbato M, Orozco-terWengel P, Tapio M, Bruford MW. SNeP: a tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. *Front Genet*. (2015) 6. doi: 10.3389/fgene.2015.00109

31. Sved J. A., Feldman M. W.. "Correlation and probability methods for one and two loci." (1973), pp. 129–132.

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

33. Ferrer-Admetlla A, Liang M, Korneliussen T, Nielsen R. On detecting incomplete soft or hard selective sweeps using haplotype structure. *Mol Biol Evol*. (2014) 31. doi: 10.1093/molbev/msu077

34. Szpiech ZA, Hernandez RD. Selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. *Mol Biol Evol*. (2014) 31. doi: 10.1093/molbev/msu211

35. Garduño López VI, Martínez-Rocha R, Núñez Domínguez R, Ramírez Valverde R, Domínguez Viveros J, Reyes Ceron A, et al. Genome-wide scan for selection signatures in Mexican Sardo negro zebu cattle. *PLoS One*. (2024) 19:e0312453. doi: 10.1371/journal.pone.0312453

36. Sherman BT, Hao M, Qiu J, Jiao X, Baseler MW, Lane HC, et al. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res*. (2022) 50. doi: 10.1093/nar/gkac194

37. Al Abri M, Al-Sumry HS, Kashmar M, Al-Ward H, Al-Azri SS. Assessing genetic diversity and defining signatures of positive selection on the genome of dromedary camels from the southeast of the Arabian Peninsula. *Front Vet Sci*. (2023) 10. doi: 10.3389/fvets.2023.1296610

38. Abdelnour SA, Swelum AA, Abd El-Hack ME, Khafaga AF, Taha AE, Abdo M. Cellular and functional adaptation to thermal stress in ovarian granulosa cells in mammals. *J Therm Biol*. (2020) 92. doi: 10.1016/j.jtherbio.2020.102688

39. Hoter A, Rizk S, Naim HY. Cellular and molecular adaptation of Arabian camel to heat stress. *Front Genet*. (2019) 10. doi: 10.3389/fgene.2019.00588

40. Marín I. Animal HECT ubiquitin ligases: evolution and functional implications. *BMC Evol Biol*. (2010) 10:56. doi: 10.1186/1471-2148-10-56

41. Metzger MB, Hristova VA, Weissman AM. HECT and RING finger families of E3 ubiquitin ligases at a glance. *J Cell Sci*. (2012) 125. doi: 10.1242/jcs.091777

42. Qian H, Zhang Y, Wu B, Wu S, You S, Sun Y. Structure and function of HECT E3 ubiquitin ligases and their role in oxidative stress. *J Transl Intern Med*. (2020) 8:71–9. doi: 10.2478/jtim-2020-0012

43. Qian J, Chen W, Ye Y. Regulation of proteasome degradation by ubiquitination and deubiquitination. *Front Mol Neurosci*. (2020) 13:140. doi: 10.3389/fnmol.2020.00140

44. Bahri H, Buratto J, Rojo M, Dompierre JP, Salin B, Blancard C, et al. TMEM70 forms oligomeric scaffolds within mitochondrial cristae promoting in situ assembly of mammalian ATP synthase proton channel. *Biochim Biophys Acta Mol Cell Res*. (2021) 1868. doi: 10.1016/j.bbamcr.2020.118942

45. Brar NK, Waggoner C, Reyes JA, Fairey R, Kelley KM. Evidence for thyroid endocrine disruption in wild fish in San Francisco Bay, California, USA. Relationships to contaminant exposures. *Aquat Toxicol*. (2010) 96. doi: 10.1016/j.aquatox.2009.10.023

46. Ruiz-Jarabo I, Martos-Sitcha JA, Barragán-Méndez C, Martínez-Rodríguez G, Mancera JM, Arjona FJ. Gene expression of thyrotropin-and corticotrophin-releasing hormones is regulated by environmental salinity in the euryhaline teleost *Sparus aurata*. *Fish Physiol Biochem*. (2018) 44. doi: 10.1007/s10695-017-0457-x

47. Coghill JM, Fowler DH, West ML, Morice WG, Blazar BR. CCR8 expression identifies CD4+ T cells enriched for regulatory T-cell activity. *Transplantation*. (2013) 85:620–5. doi: 10.1097/TP.0b013e3181638625

48. Nesargikar PN, Spiller B, Chavez R. The complement system: history, pathways, cascade, and inhibitors. *Eur J Microbiol Immunol*. (2012) 2:103–11. doi: 10.1556/EuJMI.2.2012.2.2

49. Matthews J, Gustafsson JÅ. Estrogen receptor and the future of estrogen therapy. *Nat Rev Cancer*. (2003) 3:802–11. doi: 10.1038/nrc1206

50. Bahbahani H, Salim B, Almathen F, Al Enezi F, Mwacharo JM, Hanotte O, et al. Signatures of positive selection in African Butana and Kenana dairy zebu cattle. *PLoS One*. (2018) 13:e0190446. doi: 10.1371/journal.pone.0190446

51. Altarejos JY, Montminy M. CREB and the CRTC co-activators: sensors for hormonal and metabolic signals. *Nat Rev Mol Cell Biol*. (2008) 12:141–51. doi: 10.1038/nrm2069

52. Eichner LJ, Giguère V. Estrogen-related receptors (ERRs): a new dawn in transcriptional control of mitochondrial gene networks. *Mol Endocrinol*. (2011) 25:585–93. doi: 10.1210/me.2010-0028

53. Stefanovic B, Stefanovic L. TGF-β induction of cytokine receptor-like factor 1 in chondrocytes regulates cartilage metabolism. *J Cell Physiol*. (2012) 227:901–9. doi: 10.1002/jcp.22892

54. Wilson DG, Phamluong K, Lin WY, Barck K, Carano RA, Diehl L, et al. Chondroitin sulfate synthase 1 (Chsy1) is critical for cartilage and limb development in mice. *Dev Dyn*. (2012) 241:1949–57. doi: 10.1002/dvdy.23895

55. Kelly SA, Bell TA, Selitsky SR, Buus RJ, Hua K, Weinstock GM, et al. A novel N(alpha)-acetyltransferase gene (NAA16) is associated with increased running endurance in mice. *Physiol Genomics*. (2014) 46:473–82. doi: 10.1152/physiolgenomics.00120.2013

56. Sanchez MP, Ramayo-Caldas Y, Wolf V, Laithier C, Martin P, Palhière I, et al. Genetic control of milk production and composition in dairy sheep through variants in PICALM and other candidate genes. *J Dairy Sci*. (2017) 100:9381–93. doi: 10.3168/jds.2017-12988