RESEARCH | Article in Press

Demographic history and management practices shape the genomic landscape of New Zealand's feral Kaimanawa Horses

Arne Bielke · Siavash Salek Ardestani · Bilal Sharif · Pablo Orozco-ter Wengel · Robert Rodgers Fitak · Barbara Wallner · Kelly Wilson · Elmira Mohandesan

Received: 1 May 2026 / Accepted: 7 September 2026

Abstract

Background New Zealand's feral Kaimanawa Horses provide a unique opportunity to investigate how founder history, demographic processes, and human management influence genetic diversity and inbreeding in a free-ranging horse population. Here, we generate genome-wide SNP data and integrate these with previously generated mitochondrial and Y-chromosome markers from Kaimanawa Horses to compare them with 22 domestic breeds and characterize their genetic composition, population structure, genetic diversity, and demographic history, including changes in effective population size.

Results Kaimanawa Horses exhibit a complex genetic composition, with shared genetic components mainly associated with British pony, Thoroughbred, Arabian, and draft horse lineages, consistent with historical records of multiple breed contributions. The population shows reduced heterozygosity, elevated inbreeding, and an effective population size below 50, reflecting isolation, and recent demographic contraction. Population structure analyses identify two genetic subgroups, while Y-chromosome analysis reveals private paternal haplotypes absent from modern breed reference panels, indicating the retention of unique paternal diversity.

Conclusion Together, these findings provide the first genomic framework for understanding the genetic composition, diversity, and demographic history of the Kaimanawa horse population. These results demonstrate the value of genomic approaches for guiding conservation and management strategies, preserving genetic diversity, and supporting future research on feral horse populations.

Keywords Population genomics · Conservation genomics · Inbreeding · Runs of homozygosity · Feral horses · Kaimanawa · New Zealand

Background

Maintaining livestock genetic diversity is essential for long-term food security, making the conservation of farm animal genetic resources a major international priority [1]. Domestic horses (*Equus caballus*) have played important economic, cultural, and ecological roles throughout human history, and their relationship has generated new

possibilities for both species [2, 3]. Although the timing and geographical origins of horse domestication remain debated, modern horses largely descend from the DOM2 lineage that emerged approximately 4,000 years ago, and from which nearly all modern horses descend [4, 5]. Subsequent domestication involved repeated introgression of wild maternal mitochondrial (mtDNA) lineages [3, 6, 7], followed by intensive selective breeding and the establishment of modern breeds, most of which originated within the last ~ 200 years [8, 9]. These processes have reduced genetic diversity and increased inbreeding and genetic load in domestic horses [9, 10], with particularly strong effects on paternal Y-chromosome (Y-chr) diversity [11, 12]. The development of closed studbooks and intensive selection has contributed to an estimated 16% loss of genetic diversity during the past two centuries [9]. Of the 905 horse breeds recognized globally, approximately 10% are extinct, while many extant breeds remain at risk or lack formal conservation assessment [1].

As the wild ancestors of today's domestic horses are no longer extant and genetic diversity has declined in many managed breeds, feral and less intensively managed populations may retain valuable reservoirs of genetic variation. These include Przewalski's horses, feral horse populations, and locally adapted landrace-type horses maintained outside formal breeding systems.

Feral horses occur worldwide [13] and provide valuable systems for investigating how demographic history, environmental conditions, and natural selection shape genetic diversity and adaptation in the absence of intensive human-directed breeding. Ecological and behavioural studies have extensively characterized various feral horse populations, including their social organization, habitat use, movement patterns, and population dynamics and growth [14–17]. Previous studies using microsatellite and mtDNA markers have further demonstrated that feral horse populations can retain unique genetic variation, population structure, and maternal lineages that may have been reduced or lost in intensively managed breeds [18, 19]. However, genome-wide investigations of feral horses remain limited and have focused primarily on North American populations, including American Mustangs and the Canadian Sable Island and Alberta Foothills horses [20–22].

Expanding genome-wide analyses across diverse feral horse populations is therefore important for understanding how distinct demographic histories, environmental conditions, and management practices shape genetic diversity and adaptive potentials. As feral populations are founded from diverse sources and subsequently exposed to various levels of isolation, drift, and management, they provide valuable opportunities to investigate the long-term consequences of these processes. Previous genomic studies illustrate contrasting outcomes of these processes: Sable Island horses exhibit elevated inbreeding associated with founder effects and historical bottlenecks [22, 23], whereas Theodore Roosevelt National Park horses show signatures of recent management-related inbreeding and reduced genetic diversity [24]. Together, these studies demonstrate the importance of genome-wide data for evaluating the genetic status, and conservation value of feral horse populations.

The Kaimanawa Horses of New Zealand (Fig. 1) represent a particularly informative feral population. Descended from horses introduced by European settlers during the 19th century (including Thoroughbreds, Arabians, Clydesdales, Exmoor ponies, and Welsh ponies [25]), they have inhabited the Kaimanawa Ranges of the North Island under long-term geographic isolation and documented management [26]. Originally restricted to the Waiouru Military Training Area in central North Island within the Manawatū-Whanganui region (Fig. 2), the population was once at risk of extinction, prompting protective legislation in 1981. Following legal protection in the same year, subsequent population growth resulted in ecological concerns, including substantial impacts on native vegetation such as rare tussock grasses (shown in Fig. 1), vulnerable mistletoe, and the possibly locally extinct sand iris. These impacts led to the implementation of regulated population management through periodic musters (i.e., round-up), culling and rehoming programmes [27]. The Kaimanawa Wild Horse Advisory Group manages the population through annual musters and rehoming programmes, with the aim of maintaining approximately 300 horses. However, recent surveys estimate the population at around 450 individuals [28].

Based on historical records and long-term field observations, horses within the counted area of the Kaimanawa Ranges are distributed across three geographically separated regions (Fig. 2): Argo Valley (Zone 11/Lower

Fig. 1 Free-roaming Kaimanawa Horses in tussock grassland of the Kaimanawa Ranges, New Zealand. Photo credit: Kelly Wilson (All rights reserved)

14; current census size $n=44$ individuals), Southern Zones (Upper 14/Zone 15 /Home Valley/Zone 18/Zone 19; n=~290), and Zone 20 (n=~140) [28]. Current census size of 450 individuals, and the zone-specific estimates provided above, are based on 2025 documented data. Additional horses roam around in the isolated and uncounted regions (Zone 9/Māori Trust Land), which are not shown in Fig. 2. Natural barriers, including rivers, ravines, and mountain ranges, impede movement among these regions, and documented cross-region dispersal events are exceptionally rare [28]. Long-term field observations further indicate regionally stable family bands and geographically restricted phenotypes, such as the grey coat color confined to the Southern Zones following the release of a grey Arabian in the 1960th [29].

Our previous genetic study [30] of more than 100 Kaimanawa Horses used uniparental markers and demonstrated that, despite diverse contributions of various breeds, the present-day population is dominated by only six maternal and six paternal lines, some of which show stronger affinities to specific geographical zones. Y-chr analyses further indicated contributions from British Pony, Thoroughbred, and Arabian lineages. Here, we present the first genome-wide SNP analysis of 142 extant Kaimanawa Horses. This study investigates the genetic relationships between Kaimanawa Horses and global domestic breeds, characterizes their population structure and genetic diversity, and integrates genome-wide and uniparental markers to reconstruct their demographic and maternal and paternal histories. By examining the genetic consequences of population management, our findings provide insights that can inform conservation strategies and the long-term management of genetically distinct animal populations.

Materials and methods

Study area and sample collection

Hair follicle samples were obtained from 144 Kaimanawa Horses (69 males, 75 females; Table S1) through our collaboration with the Kaimanawa Heritage Horse Welfare Society (https://kaimanawaheritagehorses.org/) and private owners. Between 1991 and 2023 (most individuals originating from cohorts born 2000–2020), horses were mustered from Argo Valley (Zone 11/Lower 14), Southern Zones (Upper 14/Zone 15/Home Valley/Zone 18/Zone 19), and Zone 20 (Fig. 2). The geographical designations used here follow the established Kaimanawa management areas described in [28]. Argo Valley horses range between Lower 14 and Zone 11, separated by the Moawhango River. Although primarily associated with Argo Valley (Lower 14), the entire group may occasionally cross into Zone 11 to graze. Accordingly, all Argo Valley samples in this study were collected from horses in Argo Valley (Lower 14). While the Southern Zone comprises multiple management areas, the available location records assigned sampled horses only to Upper 14/Zone 15 and Home Valley, which are therefore presented separately throughout the manuscript. No samples were available from Zone 11, Zone 18, or Zone 19 because these

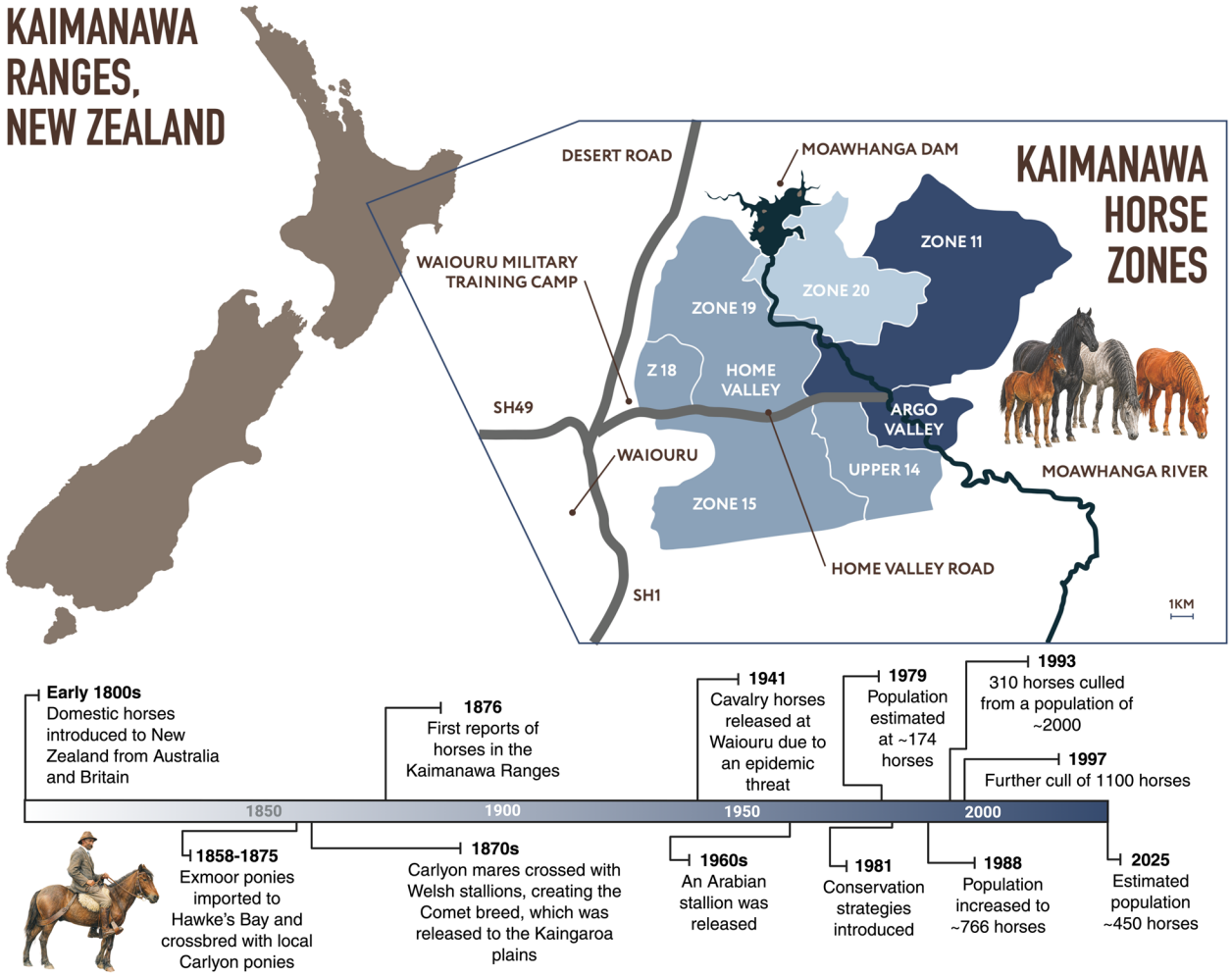


Fig. 2 Map of New Zealand showing the location of the Kaimanawa Range in the central North Island. The inset shows three zones colored by different shades of blue: Argo Valley (Zone 11/Lower 14); the Southern Zones (Upper 14/Zone 15/Home Valley/Zone 18/Zone 19); and Zone 20. The State Highways (SH1, SH49), Desert road and Moawhanga river are shown on the map. Current estimated population sizes are 44 horses in Argo Valley, ~290 in the Southern Zones, and ~140 in Zone 20 [28]. The accompanying timeline (bottom) details key milestones in the Kaimanawa Horse population history, spanning from the early 1800s through 2025

areas are not directly targeted during mustering, although occasional movement of horses between management areas cannot be excluded.

After professional taming, horses were rehomed in domestic settings, eligible for sampling. To minimize stress, samples were collected during New Zealand summer (December – February), when hair follicles are naturally relaxed [31]. Sampling was minimally invasive and performed by private owners following written instructions to gently pull only a few hairs at a time; if a horse appeared sensitive to tail hair collection, mane hair was used instead. Only 10–20 hair strands were removed per attempt, with multiple collections performed to obtain 40–60 follicles per horse. As the sampling procedure was non-terminal and only minimally invasive, neither euthanasia or anesthesia was required. The horses remained conscious throughout the procedure, and hairs were manually plucked by their private owners.

DNA Extraction, SNP genotyping and data processing

Genomic DNA was extracted from 20 to 30 hair follicles per horse ($n = 144$) using a magnetic bead-based protocol at Neogen Europe Ltd. Samples were genotyped on the Neogen GeneSeek Genomic Profiler (GGP) Equine v4 array (~75 K SNPs) following the manufacturer's instructions. The array includes ~71,000 markers evenly distributed across the genome, including autosomal and X-chr SNPs (~70,231) and uniparental markers (mtDNA: 1,187; Y-chr: 170).

Illumina GenomeStudio EquCab 3.0 Final Report files (top forward) were converted to PLINK v1.9 [32, 33] format (.ped/.map) using a custom bash script. SNPs associated with phenotypes (Equine v4 array annotation), mtDNA and sex chromosome markers, were excluded to retain exclusively neutral autosomal loci. Genome coordinates were updated from EquCab 2.0 to EquCab 3.0 using UCSC LiftOver ([34]; accessed March 3, 2023) with a minimum 95% base remapping and 100% alignment blocks or exons mapping; minimum hit and chain sizes were set to 0. Complete sample metadata and group assignments for all analyzed Kaimanawa Horse individuals are detailed in Table S1. More details about the specific reference genome assembly (EquCab 2.0 or EquCab 3.0) used for genotyping Kaimanawa Horses is provided in Table S2. The summary of data generation and downstream analytical workflow are illustrated in Fig. 3.

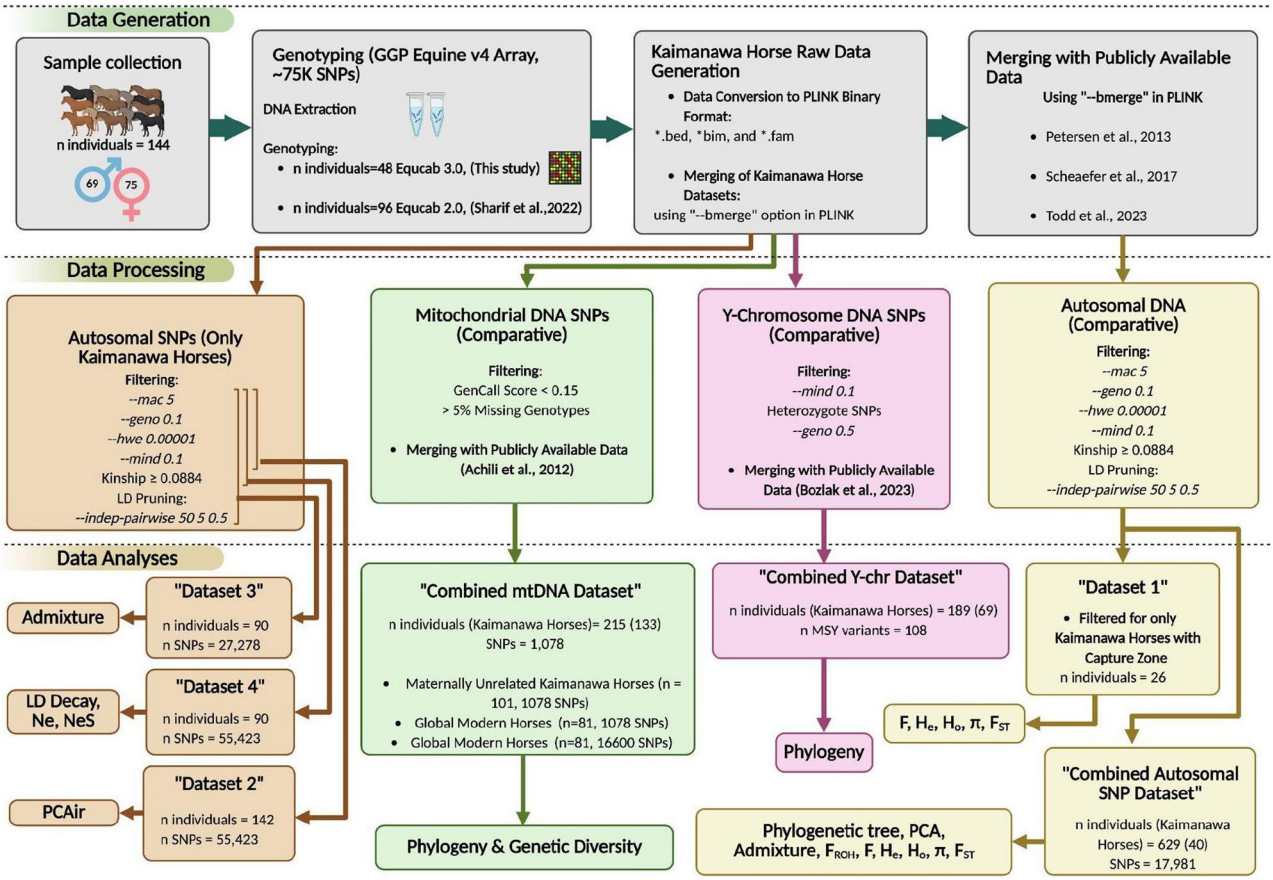


Fig. 3 Schematic flowchart outlining the workflow for data generation, processing, and downstream analysis across the eight datasets utilized in this study. Note: --mac: minor allele count; --geno: maximum per-SNP missingness; --hwe: Hardy-Weinberg equilibrium threshold; mind: maximum per-individual missingness; Kinship: kinship coefficient threshold used to remove related individuals; --indep-pairwise: linkage disequilibrium pruning parameters (window size, step size, r^2 threshold); F: inbreeding coefficient; H_O: observed heterozygosity; H_e: expected heterozygosity; F_{ST}: fixacity

Comparative genomic datasets of modern domestic horse breeds

For comparative genomic analyses, we integrated publicly available genomic data from domestic horse breeds selected based on documented historical records or a high likelihood of importation into New Zealand during the establishment of the Kaimanawa Horse population. The reference panel comprised genotype data from 23 selected breeds from [8] (EquCab2.0, 54 K SNP array), Welsh Ponies from [35] (EquCab2.0, 670 K array), and whole-genome sequencing (WGS) data from 12 selected breeds from [36] (EquCab3.0). Breed abbreviations, sample IDs, number of individuals per breed, data type, and sources are in Table S2. Datasets aligned to EquCab2.0 were converted to EquCab3.0 using UCSC liftOver (https://genome.ucsc.edu/cgi-bin/hgLiftOver; accessed 3 Mar 2025).

Genotype data filtering and relatedness estimation

To construct a working dataset, we merged (utilizing --bmerge option in PLINK v1.90) the Kaimanawa Horse SNP dataset with the comparative datasets and applied quality control using PLINK v1.90 [32, 33] to remove SNPs with: (i) minor allele count < 5 (--mac 5); (ii) call rate < 90% (--geno 0.1); (iii) SNPs deviating from Hardy-Weinberg equilibrium ($p < 10^{-5}$) (--hwe 0.00001) [8, 37, 38]; and (iv) individuals with > 10% missing data (--mind 0.1). During the quality control steps stated above, non-overlapping SNPs were removed as they did not satisfy the predefined call rate or minimum allele count thresholds. Two Kaimanawa Horse samples had more than 10% missing data and were thereafter excluded from the dataset, decreasing the total number of Kaimanawa Horse samples from 144 to 142. Relatedness was estimated using KING v2.3.1 [39]. Individual pairs with a kinship coefficient of ≥ 0.0884 (second-degree relatives or closer [40] were identified as related (Table S3). The remaining individuals were retained as an operationally unrelated dataset for downstream population genetic analyses to prevent contemporary family structures from distorting clustering and allele frequency estimates, and are hereafter referred to as 'unrelated'. To maximize the number of unrelated individuals, we applied NAToRA [40] to the kinship coefficients previously estimated by KING v2.3.1 [39]. Breeds represented by more than 40 individuals were randomly downsampled to 40 individuals to reduce substantial differences in sample size among breeds, thereby avoiding overrepresentation of highly sampled breeds and ensuring balanced comparisons in downstream analyses. Conversely, breeds with fewer than 15 individuals were excluded to maintain statistical robustness [41]. SNPs in linkage disequilibrium (LD) were pruned (--indep-pairwise 50 5 0.5), retaining one SNP per LD cluster. The resulting *"Combined Autosomal SNP Dataset"* (24 horse breeds; 629 individuals including 40 unrelated Kaimanawa Horses; 17,981 SNPs; Table S4) were further used for all comparative analysis. Although we applied a stringent LD pruning to construct the Combined Autosomal SNP Dataset, our filtering thresholds remain directly comparable to those used in several previous studies (e.g. [42, 43]), , thereby facilitating a direct comparison of our findings with the published literature.

In addition to the comparative Combined Autosomal SNP Dataset described above we compiled four analysis-specific datasets to perform population genetic analysis exclusively within the Kaimanawa Horse population. *Dataset 1*: Unrelated capture-zone LD-pruned ($n = 26$; 17,981 SNPs extracted from the Combined Autosomal SNP Dataset) used for among capture zone genetic diversity (F: individual inbreeding coefficient, H_o: observed heterozygosity, H_e: expected heterozygosity, nucleotide diversity: π) and genetic differentiation (F_{ST}). *Dataset 2*: Complete non-LD-prunned individuals ($n = 142$; 55,423 SNPs from the Kaimanawa Horse dataset) used for PC-AiR and consequently k-means clustering. *Dataset 3*: Unrelated LD-pruned ($n = 90$; 27,278 SNPs from the Kaimanawa Horse dataset) used for Admixture analysis. *Dataset 4*: Unrelated non-LD-prunned individuals ($n = 90$, 55,423 SNPs from the Kaimanawa Horse dataset) used for LD decay, N_e and change of N_e across generations (N_eS) calculation. Detailed information on different dataset can be found in Fig. 3 and Table S4.

Genetic diversity and inbreeding in domestic horse breeds

For each domestic breed, including Kaimanawa Horses, in the Combined Autosomal SNP Dataset, genetic diversity metrics (F, H_o, H_e) were computed (Table 1) in PLINK v1.90 (--het) [32, 33], and statistically analyzed in R v4.4.2 [44]. Average π was calculated using VCFtools v0.1.16 [45] (Table 1).

Moreover, Runs of Homozygosity (ROH) were identified in the Combined Autosomal SNP Dataset using the consecutive runs algorithm implemented in the R package detectRUNS v1.0.0 (consecutiveRUNS.run; R v4.4.2) [44, 46]. ROHs were defined using a minimum threshold of 30 SNPs per run (minSNP = 30), a maximum inter-SNP gap of 1 Mb (maxGap = 1,000,000 bp), a maximum of two missing SNPs (maxMissRun = 2), a maximum

Table 1 Breed names, abbreviations, representative colors and sample sizes used in this study. For each breed, summary statistics calculated from the Combined Autosomal SNP Dataset are provided, including sample size (n), inbreeding coefficient (F), expected heterozygosity (H_e), observed heterozygosity (H_o), and nucleotide diversity (π). Colors correspond to those used throughout the figures. Kaimanawa Horse estimates are additionally reported by capture zone. Kaimanawa Horse estimates are further subdivided by capture zone. F, H_e, and H_o were calculated using PLINK v1.9 [33, 34], and was calculated using VCFtools v0.1.16 [45]

Breed name	Abbreviation	Color	n	F	He	Ho	π
Akhal Teke	AKTE		15	0.049	0.327	0.311	0.314
Arabian	ARAB		39	0.102	0.328	0.294	0.308
Belgian	BELG		27	0.122	0.328	0.288	0.285
Clydesdale	CLYD		31	0.247	0.328	0.247	0.247
Exmoor Pony	EXPO		17	0.273	0.328	0.238	0.244
Fell Pony	FEPO		17	0.133	0.328	0.284	0.283
Finnhorse	FINN		24	0.067	0.328	0.306	0.302
Franchese Montagne	FRMO		29	0.067	0.328	0.306	0.302
French Trotter	FRTR		15	0.049	0.328	0.312	0.307
Hanovarian	HANO		18	-0.041	0.328	0.341	0.337
Icelandic	ICEL		26	0.122	0.328	0.288	0.288
Miniature Pony	MIPO		21	0.106	0.328	0.293	0.296
Mongolian	MONG		22	0.050	0.328	0.312	0.311
Morgan	MORG		40	0.067	0.328	0.306	0.314
Norwegian Fjord	NOFJ		18	0.155	0.328	0.277	0.276
Percheron	PERC		20	0.063	0.328	0.307	0.307
Quarter Horse	QUHO		40	-0.015	0.328	0.333	0.334
Saddlebred	SABR		23	0.058	0.328	0.309	0.305
Shire Horse	SHIR		22	0.169	0.328	0.272	0.272
Shetland Pony	SHPO		29	0.208	0.328	0.260	0.266
Standardbred	STBR		33	0.087	0.328	0.299	0.303
Thoroughbred	THBR		32	-0.006	0.328	0.330	0.324
Welsh Pony	WEPO		31	0.083	0.327	0.300	0.308
Kaimanawa Horses:							
All	KAHO		40	0.145	0.328	0.280	0.297
Capture Zone	Argo Valley		6	0.048	0.318	0.303	0.275
Capture Zone	Southern Zones (Home Valley)		5	0.039	0.318	0.306	0.282
Capture Zone	Southern Zones (Upper 14/ Zone 15)		6	0.025	0.318	0.310	0.294
Capture Zone	Zone 20		9	0.045	0.318	0.304	0.280

of one heterozygous SNP per run (maxOppRun = 1), and a minimum physical length threshold of 1,000 bp (minLength = 1,000 bp; default) [22]. For each individual, genomic inbreeding coefficients based on ROH (F_{ROH}) was calculated as proportion of the autosomal genome (2.36 Gb) covered by ROHs following [47]. To estimate the approximate timing of inbreeding, the mean ROH-based inbreeding coefficient (F_{ROH}) was calculated for five ROH length classes, grouped as short (0–2 Mb and 2–4 Mb), intermediate (4–8 Mb), long (8–16 Mb), and very long (> 16 Mb). Additionally, we calculated the average F_{ROH} for all detected ROHs regardless of length.

Phylogenetic relationships and pairwise genetic distances in domestic horse breeds

To investigate the phylogenetic relationships among modern horse breeds and Kaimanawa Horses (Combined Autosomal SNP Dataset), a Maximum Likelihood (ML) tree with 1,000 bootstrap replicates was generated using IQ-TREE v2.3.6 [48], employing the built-in model selection tool [49]. ML-trees were visualized in R v4.4.2 [44] using ggplot2 v4.0.0 [50] and ggtree v3.14.0 [51].

Pairwise F_{ST} values were estimated using the Combined Autosomal SNP Dataset following [52, 53] using StAMPP [54] with 1,000 bootstrap replicates and 95% confidence intervals. P-values were adjusted using the Benjamini-Hochberg False Discovery Rate (FDR) method [55]. Plots were generated using GENOVIS v1.0.5 [56].

Genetics clustering among domestic horse breeds

To characterize the autosomal genetic composition of Kaimanawa Horses relative to other modern horse breeds, we performed an unsupervised clustering analysis on the Combined Autosomal SNP Dataset using ADMIXTURE v1.3.0 software [57]. The analysis evaluated values of K ranging from 2 to 25 using 1,000 bootstrap replicates, where K represents the number of inferred genetic clusters. The optimal K was determined based on cross-validation (CV) error. Results were visualized in R v4.4.2 [44] using ggplot2 v4.0.0 [50]. To determine genetic similarities across breeds, we performed PCA using PLINK v1.9 [32, 33].

To examine the genetic similarities among all sampled Kaimanawa Horse individuals while explicitly accounting for relatedness, a Principal Components Analysis (PC-AiR) in Related Samples (PC-AiR) [58] was performed on Dataset 2. This approach allowed for the robust inclusion of all Kaimanawa Horses, including closely related individuals, without distorting the underlying population structure. Relatedness was first inferred using SNPRelate [59] following GENESIS guidelines [60], and incorporated into the PC-AiR [58]. Analyses were conducted in R v4.4.2 [44] using the GENESIS package v2.36.0 [60], SNPRelate v1.40.0 and GWASTools v1.52.0 [59].

Historical effective population size in Kaimanawa Horses

LD decay (mean r^2) in Dataset 4 was estimated up to 2 Mbp (minimum $r^2 = 0$) using PopLDdecay v3.31 [61]. The effective population size (N_e) was inferred with SNeP v1.11 [62] following [63]

$$Ne_{(t)} = (4f(C_t))^{-1}\left(E[r^2_{adj} \mid C_t]^{-1} - \alpha\right)$$

where $Ne_{(t)}$ is the estimated N_e at t generations ago, C_t is the recombination rate at generations t, r^2_{adj} is the sampling-bias-adjusted r^2, and α is the constant accounting for the occurrence of mutations. N_e was estimated using the following parameters: maximum SNP distance (-maxdist) 30,000,000 bp, minimum SNP distance (-mindist) 500,000 bp, the [64] method (-svedf), r^2 adjustment for population size (-samplesize) 90 unrelated Kaimanawa Horses), and recombination rate (-recrate) 1.246e-8.

The maximum SNP distance (30 Mbp) was used to capture recent N_e, and the recombination rate (1.246×10^{-8}) followed [65]; average generation time was set to 8.74 years [66]. The values of N_e for the K1 and K2 genetic subgroups, identified using K-means clustering of PC-Air results of Kaimanawa Horse population, were estimated

using identical filtering criteria to those applied to all unrelated individuals. However, the "-samplesize" option was specified as 21 for K1 and 69 for K2 to reflect group-specific sample sizes (Table S1). Additionally, the slope analysis of N_e values (N_eS) was performed using a custom R script to investigate N_e changes across generations, as explained by [67]. LD decay and N_e trends were generated and visualized in R v4.4.2 [44] with ggplot2 v4.0.0 [50].

Mitochondrial diversity in domestic horses

To investigate mtDNA haplogroups (HGs) and diversity, we extracted 1,187 SNPs from the GGP Equine v4 array dataset, combining newly generated genotypes from this study with previously published Kaimanawa Horse data [30]. SNP extraction and haplogroup assignment were performed using a custom-made R v4.4.2 [44] script. Subsequently, genotype calls with a GenCall score < 0.15 were set to missing, and individuals with > 5% missing genotypes were removed using a custom-made R v4.4.2 [44] script. The filtered mtDNA SNPs were converted to FASTA using the Biostrings v2.72.1 package [68] in R v4.4.2 [44]. Overlapping nucleotide positions between the GGP Equine mtDNA SNPs and the horse mtDNA reference genome (NC_001640.1) were converted to lowercase in the reference sequence using a custom bash script. A whole-mtDNA dataset (FASTA format) from a global dataset (79 domestic horses, two Przewalski's horses (NCBI accessions JN398377–457; [7], and one donkey (Equus asinus, NC_001788.1; [69]) was then aligned to the modified reference genome using MAFFT v7.526 [70]. The lowercase positions ($n = 1{,}187$) were extracted from the alignment and converted to FASTA using Biostrings v2.72.1 [68]. These sequences were combined with the Kaimanawa Horse mtDNA dataset, and SNPs with more than one missing genotype in the Kaimanawa population were removed. Following the steps above, we constructed a working dataset, hereafter denoted as the *"Combined mtDNA Dataset"* (Table S4).

mtDNA HGs relationships in Kaimanawa Horses in relation to the global modern horses were inferred using a Median-joining- (MJ) network in PopART v1.7 [71]. After excluding maternally related individuals (by nuclear π-hat ≥ 0.25 and identical mtDNA haplotype), diversity indices (h: number of haplotypes, S: number of polymorphic sites, Hd: haplotype diversity, π: nucleotide diversity, θW: Watterson's theta) were estimated for both Kaimanawa and global modern horses, using DnaSP v6.12.03 [10]. To obtain a finer-scale view of nucleotide diversity across the mtDNA, a sliding window analysis (window size = 200 bp, step size = 100 bp) was performed to estimate π in both populations using DnaSP v6.12.03 [10].

Y-chromosome diversity in domestic horses

The paternal Y-chr diversity was characterized using 170 male-specific region of the Y-chr (MSY) variants, newly genotyped in 26 male Kaimanawa Horses, using the GGP Equine v4 array and filtered to exclude: (i) individuals with > 10% missing data; (ii) variants showing heterozygous genotypes in males; and (iii) variants with > 5% missing data. After filtering, 150 MSY variants remained.

Genotypes at 37 Crown key MSY variants (6 Crown HGs determining key-variants; 26 Crown haplotype (HT) determining key-variants; 5 additional Crown variants) resolved in [72] were extracted and oriented to the LipY764 assembly [73, 74] if needed. In addition, 14 Crown key-variants not represented on the array were genotyped in Kaimanawa Horses using LGC KASP assays, following the protocol described by [73] and [75] to resolve the full Crown HT structure [76]. This includes variants in HGs daC_Ad-b* (rAF, rDT, eAPT, eAPU, fSA, eBZ, eAPX, qH), or daC_Am* (qCU, qCI, fTT, fRQ, eARK) or daC_Ao-a1D2 (qW).

This dataset was expanded with published data from 43 Kaimanawa males [30], resulting in a total of 69 Kaimanawa Horses, genotyped for 51 MSY variants. This dataset was further supplemented with published data from 120 male horses representing global Crown HTs and one Icelandic horse as outgroup [76]. Allelic states in 102 Crown key variants, and the six Crown determining variants were extracted for those 121 males, combined with the Kaimanawa dataset and the genotypes missing in the Kaimanawa Horses were imputed based on HG structure as described in [72, 77].

The resulting "Combined Y-chr Dataset" (Table S4) comprises 108 MSY variants across 69 Kaimanawa Horses, 120 horses from various Crown group breeds, and one non Crown Icelandic. HGs relationships were visualized using a MJ-network in NETWORK v10.2.0.0 (Fluxus Technology, Suffolk, UK) [78] with equal substitution weights and no external rooting.

Results

To characterize genetic diversity, genetic sharing with various domestic horse breeds, population structure, and inbreeding in Kaimanawa Horses, we combined genome-wide autosomal SNPdata with uniparental markers (mtDNA and Y-chr).

Integration with public datasets: combined autosomal SNP dataset

After quality control filtering and relatedness pruning, the Combined Autosomal SNP Dataset comprised 17,981 SNPs from 629 unrelated individuals, including 23 modern horse breeds ($n=589$) [8, 35, 36] and feral Kaimanawa Horses ($n=40$). Although not a formal breed, Kaimanawa Horses were included among modern breeds for analytical consistency. Breed names, abbreviations, sample sizes, and colors are listed in Table 1. Detailed information on the Kaimanawa Horses, all analysed individuals, and relatedness results are provided in Tables S1, S2, and S3, respectively.

Genetic diversity and inbreeding in domestic horse breeds

Within-breed genetic diversity was estimated using the Combined Autosomal SNP Dataset (17,981 SNPs) (Table 1). The Kaimanawa Horses exhibited a moderate level of genomic inbreeding ($F=0.145$), exceeding that observed in most cosmopolitan horse breeds but remaining lower than that of several isolated populations, including the Shetland Pony, Clydesdale, and Exmoor Pony. H_e was nearly identical across breeds (0.327–0.328), whereas H_o varied more widely, from 0.238 in Exmoor Ponies to 0.341 in Hanoverians. Kaimanawa Horses ($H_o = 0.280$) ranked within the upper third of breeds analysed. SNP-array–based estimates of the average π indicated intermediate genome-wide variation in Kaimanawa Horses ($\pi=0.297$), near the center of the distribution between Finnhorses (0.302) and Miniature Ponies (0.296) (Figure S1). Higher diversity was observed in Hanoverians (0.341), whereas Exmoor Ponies showed the lowest values (0.244). Because these estimates are derived from an ascertained biallelic SNP array, they represent relative diversity among breeds rather than absolute nucleotide diversity.

Using the Dataset 1, comprising the Combined Autosomal SNP dataset (17,981 SNPs), genetic diversity metrics were assessed among Kaimanawa Horses from different capture zones. Argo Valley horses showed the highest inbreeding ($F=0.048$) and lowest diversity ($H_e = 0.318$, $H_o = 0.303$, $\pi=0.275$), whereas individuals from Southern Zone (Upper 14/Zone 15) exhibited the lowest inbreeding ($F=0.025$) and highest diversity ($H_o = 0.310$, $\pi=0.294$) (Table 1; Figure. S2).

Genome-wide F_{ROH} value was the third highest in Kaimanawa Horses (0.156 ± 0.060), exceeded only by the Clydesdale (0.204 ± 0.035) and Exmoor Pony (0.213 ± 0.068) (Table S5). When partitioned by ROH length, F_{ROH} values were as follows: For short ROHs (0–4 Mb), Kaimanawa Horses show very low F_{ROH} (0–2 Mb = 0.000; 2–4 Mb = 0.008 ± 0.004), comparable to many other breeds. For intermediate ROHs (4–8 Mb), Kaimanawa Horses have a moderately elevated F_{ROH} (0.043 ± 0.015). This value is above the average of many breeds but lower than highly inbred breeds such as Exmoor Pony (0.076 ± 0.020), Clydesdale (0.073 ± 0.010), and Thoroughbred (0.061 ± 0.011). This suggests some contribution from inbreeding occurring several generations ago, but it is not exceptional. For long ROHs (8–16 Mb), Kaimanawa Horses exhibit a high F_{ROH} (0.057 ± 0.022), ranking third among all breeds, exceeded only by Exmoor Pony (0.083 ± 0.035) and Clydesdale (0.076 ± 0.025). When

it comes to very long ROHs (> 16 Mb), Kaimanawa Horses have the highest F_{ROH} of all breeds (0.049 ± 0.044). This is notably higher than Clydesdale and Exmoor Pony (0.040 ± 0.023, and 0.040 ± 0.030, respectively) and substantially exceeds most other breeds. Long ROH are indicative of very recent common ancestry and recent inbreeding, whereas progressively shorter ROH reflect more historical autozygosity and demographic processes [47, 79, 80].

Phylogenetic relationships and pairwise genetic distances in domestic horse breeds

The ML tree (Fig. 4a, Figure S3), constructed from the Combined Autosomal SNP Dataset, shows that nearly all Kaimanawa Horses form a distinct cluster positioned in close proximity to Quarter Horses, American Saddlebreds, and Morgan horses, except for one individual grouping with Hanoverian.

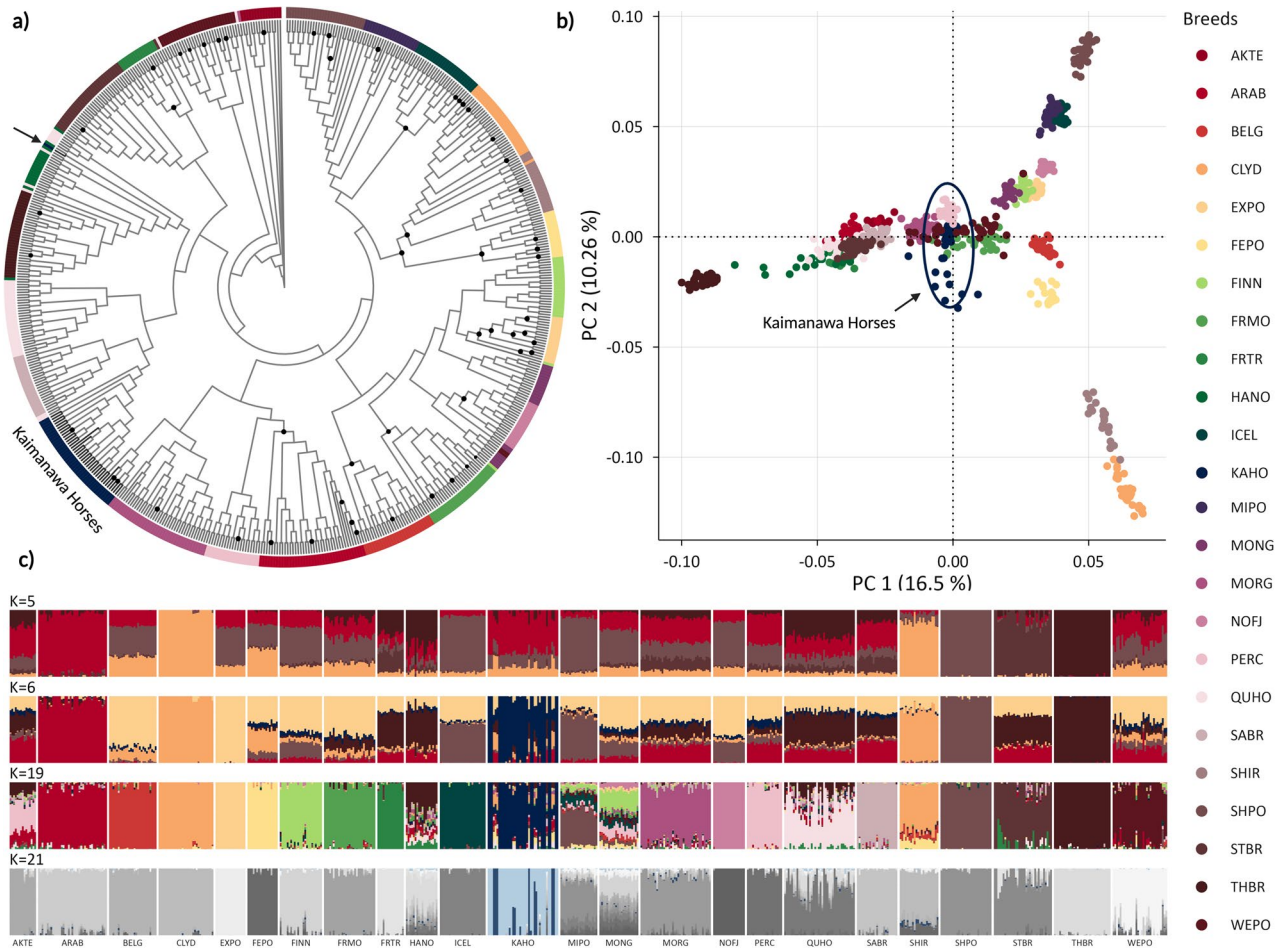


Fig. 4 Phylogenetics and population clustering analyses based on the Combined Autosomal SNP Dataset. **a** Circular Maximum Likelihood (ML) phylogenetic tree with equalized branch lengths. The outer ring shows breeds (Table 1), and node sizes represent bootstrap support values (90–100). The Arrow denotes a Kaimanawa Horse outlier, grouped with Hanoverian and Quarter Horses. **b** Principal component analysis (PCA) showing PC1 (16.5%) vs. PC2 (10.3%), with individuals colored by breed (Table 1). The blue ellipse represents a 95% confidence interval for Kaimanawa Horses. **c** Admixture analysis results for K = 5, 6, and 19, with colors representing inferred genetic components and the breed contributing most strongly to each component (Table 1). At K = 21, all components are shown in gray except those predominant in Kaimanawa Horses, which are highlighted in two shades of blue. The lowest cross-validation (CV) error was observed at K = 19. The ML tree was generated using VCF-KIT v0.2.6 [81] and visualized in R v4.4.2 [44] using ggplot2 [50] and ggtree [51]. PCA was performed using PLINK v1.9 [32, 33], and admixture analysis was performed using ADMIXTURE v1.3.0 [57]. Breed abbreviations and corresponding colors are provided on the right side of the figure

Pairwise F_{ST} values were calculated among all 24 breeds in the Combined Autosomal SNP Dataset (17,981 SNPs) (Figure S4). All pairwise comparisons were significant. The greatest differentiation was observed between Clydesdale and Exmoor Pony (F_{ST} = 0.20, p-value < 0.005). For Kaimanawa Horses, the highest differentiation occurred with Exmoor Ponies (F_{ST} = 0.14; p-value < 0.005), whereas the lowest was observed with Mongolian horses and Hanoverians (F_{ST} = 0.06; p-value < 0.005). The average F_{ST} between Kaimanawa Horses and all other breeds was 0.08.

Pairwise F_{ST} among Kaimanawa Horse individuals from different zones in Dataset 1 (Figure S5) revealed the greatest differentiation between Southern Zone (Upper 14/Zone 15) and Argo Valley (F_{ST} = 0.115; p-value < 0.005), whereas the lowest differentiation was observed between Argo Valley and Zone 20 (F_{ST} = 0.002; p-value < 0.005). Consistent with this interpretation, uniparental markers indicate that some zones are dominated by specific maternal and paternal lineages. All F_{ST} analyses were performed using an LD-pruned SNP dataset to avoid inflation of differentiation estimates from linked markers.

Genetic clustering among domestic horse breeds

In the PCA of the Combined Autosomal SNP Dataset, Kaimanawa Horses clustered near the center of the plot (Fig. 4b), overlapping within the 95% confidence interval (CI) with several light-draft, coldblood, riding and Pony horse breeds (French Trotter, Percheron, Morgan, and Welsh Pony). PC1, explaining 16.5% of the total variance, primarily separates Thoroughbreds (left) from Clydesdale and Shire horses (right). PC2, accounting for 10.26% of the variance, distinguishes Clydesdale and Shire horses (bottom right) from Shetland Ponies (top right). Additional principal component axes are shown in Figure S6.

Admixture analysis of the Combined Autosomal SNP Dataset (Fig. 4c) showed that at K = 5, Kaimanawa Horses shared genetic components with several domestic breeds, primarily Arabian, followed by Clydesdale, Shetland Pony, Thoroughbred, and Standardbred. From K = 6 onward, they began to form a distinct cluster. The optimal model (K = 19, lowest CV error; Figure S7) revealed mixed shared genetic components, with highest contributions from Clydesdale (0.037), Fell Pony (0.034), Thoroughbred (0.028), Quarter Horse (0.024), and Arabian (0.023) (Table S6). At K = 21, Kaimanawa Horses are divided into two genetic subgroups (Fig. 4c). The proportion of shared genetic components for K = 2–25 are shown in Figures S8–S11.

Among the Kaimanawa Horses, 27 individuals exhibited a high proportion (> 89%) of Kaimanawa genetic components. Four individuals showed intermediate proportions of Kaimanawa genetic components (60–84%), with the genetic components of any other breed not exceeding 7%. In contrast, nine individuals exhibited substantially lower proportions of Kaimanawa genetic components (5–46%) and greater contributions from other breed clusters, likely reflecting residual founder variation retained from the diverse domestic breeds that established the Kaimanawa population (Table S6). In PC-Air analysis (Figure S13), two of these nine individuals (PO_007_1, SH_044_1) are located in the intermediate position between the clusters, and two are on the border line of Upper 14/Zone 15 cluster (PA_032_1, TA_002_1).

We assessed within-population structure in Kaimanawa Horses using PC-Air [58] and ADMIXTURE v1.3.0 [57]. PC-Air of 142 individuals and 55,423 SNPs (Dataset 2) supported K = 2 clusters (Figure S12). PC1 (14.78%) and PC2 (4.85%) primarily separated capture zones (Figure S13). Argo Valley horses showed limited genetic variation and largely overlapped with individuals from Zone 20 and, to a lesser extent, Home Valley in the Southern Zones. In contrast, horses from Southern Zone (Upper 14/Zone 15) exhibited greater genetic variation and formed a distinct, non-overlapping cluster. Consistent with the PC-AiR pattern, K-means clustering assigned individuals from Argo Valley, Southern Zone (Home Valley), and Zone 20 to a single genetic cluster (K2), whereas Southern Zone (Upper 14/Zone 15) individuals formed a separate cluster (K1) (Table S1; Figure S13).

The admixture analysis performed on Dataset 3 identified K = 2 as the best-supported model based on the lowest CV error (Figures S14 and S15). This structure indicates spatial genetic differentiation among capture zones, consistent with restricted gene flow driven by terrain features and social band dynamics, as supported by long-term field observations [28]. K1 proportions were highest in Southern Zone (Upper 14/Zone 15; 0.750–1.000;

including individuals of unknown origin), intermediate in Southern Zone (Home Valley; 0.243–0.465), and lowest in Zone 20 (0.006–0.136). Very low K1 proportions (<0.006) were observed in individuals from Zone 20 and Argo Valley. Individual assignments are provided in Table S1.

Historical effective population size in Kaimanawa Horses

LD decay patterns calculated for Dataset 4 differed among groups, with the fastest decay observed in the full dataset and the slowest in K1 (Figure S16a, Table S7). N_e was estimated following [63] and [62] using individuals born between 1980 and 2022 (Figure S16b), assuming an average generation interval of 8.74 years reported for the Bardigiano breed [82]. Because pedigree data is currently unavailable for Kaimanawa Horses, generation interval was assumed from published estimates for semi-feral horses [82]. We note that generation intervals may vary among sex-specific pathways; for example, previous study [82] reported a value of 9.1 years for the mother–son pathway, while generation intervals of approximately 9–10 years have been reported in other equine breeds. Therefore, the adopted average value of 8.74 years should be considered an approximation.

Historical N_e declined from ~128 (17 generations ago) to 42 (one generation ago) (Table S8). Slope analysis of the inferred trajectory (N_eS) revealed non-uniform change through time, with the steepest (positive values) decline ~2 generations ago (N_eS = 0.43) and the slowest (negative values) at 12 and 17 generations ago ($N_eS = -0.33$) (Figure S16c). Estimated separately, N_e decreased from ~70 to 14 in K1 and from ~100 to 45 in K2 over the same period (i.e., 17 generations ago) (Table S8). The timing of N_e changes differed between groups: K1 showed the steepest decline 10 generations ago (N_eS = 0.25) and the slowest 16 generations ago ($N_eS = -0.40$), whereas K2 showed steepest changes at 7, 13, and 17 generations ago (N_eS = 0.25) and slowest at 6 and 12 generations ago ($N_eS = -0.25$) (Figure S16c).

Mitochondrial diversity in domestic horses

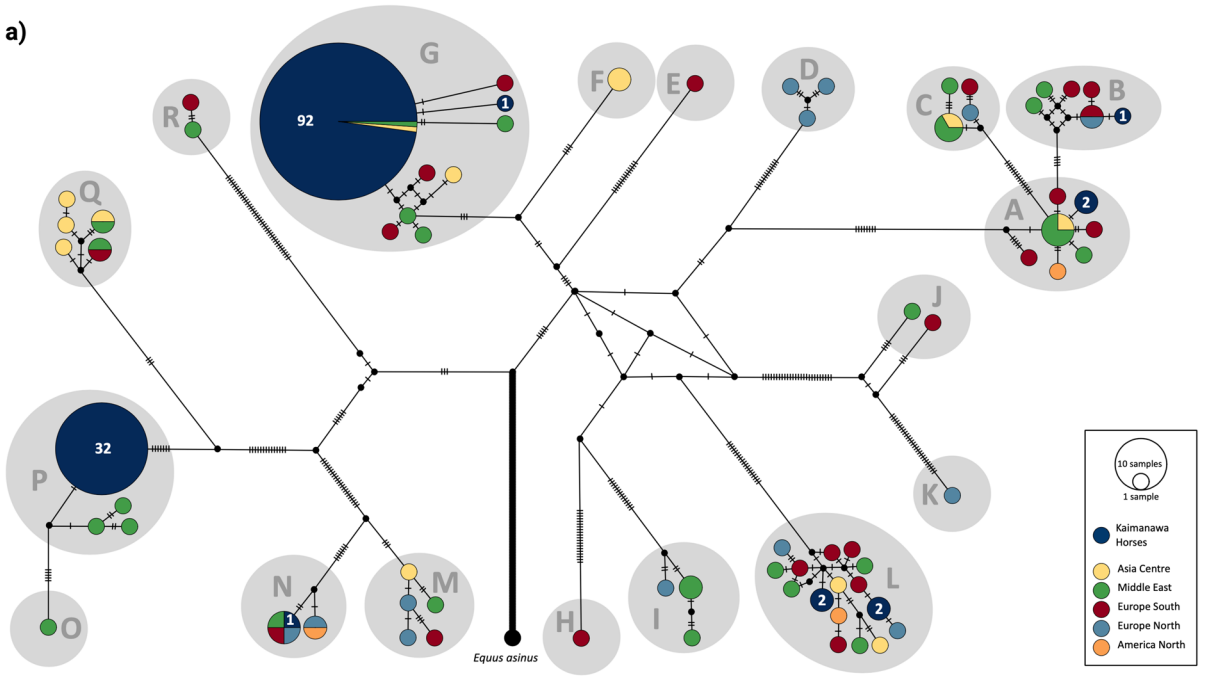
After quality control steps, the Combined mtDNA Dataset retained 1,078 SNPs from 133 Kaimanawa Horses, 79 domestic horses (Central Asia, Middle East, Europe, North America), two Przewalski's horses, and one donkey. MJ network analysis confirmed six previously defined HGs (A, B, G, L, N, P; Fig. 5a, Table S1), with Kaimanawa Horses dominated by HG-G (69.9%) and HG-P (24%). HG-A, B, L, and N each included fewer than four individuals. Compared with our previous findings [30], the new dataset ($n = 40$) added three individuals to HG-L (two different HTs), one to HG-P, and 36 to HG-G, including one distinct HT.

Among 101 maternally unrelated Kaimanawa Horses (from the Combined mtDNA Dataset), Hd was 0.489 ± 0.047, π was 0.019 ± 0.002, and θw was 0.020 ± 0.005 (Table S9). In comparison, a global modern horse mtDNA dataset [7] showed substantially higher diversity ($Hd = 0.994 \pm 0.003, \pi = 0.039 \pm 0.002, \theta w = 0.060 \pm 0.015$) (Table S9). While this comparison reflects global horse modern breeds diversity rather than each breed-level variation, it indicates that the Kaimanawa population represents a limited subset of the broader maternal diversity present in modern horses. Nucleotide diversity across mtDNA (Figure S17 and Table S10) confirmed reduced diversity, with peaks in COX3 ($\pi = 0.156$), ND5–ND6 ($\pi = 0.131$), 16 S rRNA ($\pi = 0.113$), and the Control Region/D-loop ($\pi = 0.116$).

Y-chromosome diversity in domestic horses

To investigate paternal ancestry in Kaimanawa Horses, we constructed a MJ network (Fig. 5b) using 108 MSY [72, 76] from 69 male Kaimanawa Horses and 121 reference horses representing diverse paternal lineages, including 120 Crown group horses (paternal lineages detected in the majority of modern domestic breeds) and one non Crown Icelandic horse as outgroup. Of the 108 variants, 37 were genotyped using the Illumina GGP Equine v4 array and 14 using LGC KASP assays [73, 75]; the remaining 57 were imputed based on HG topology as

https://doi.org/10.1186/s12864-026-13345-y



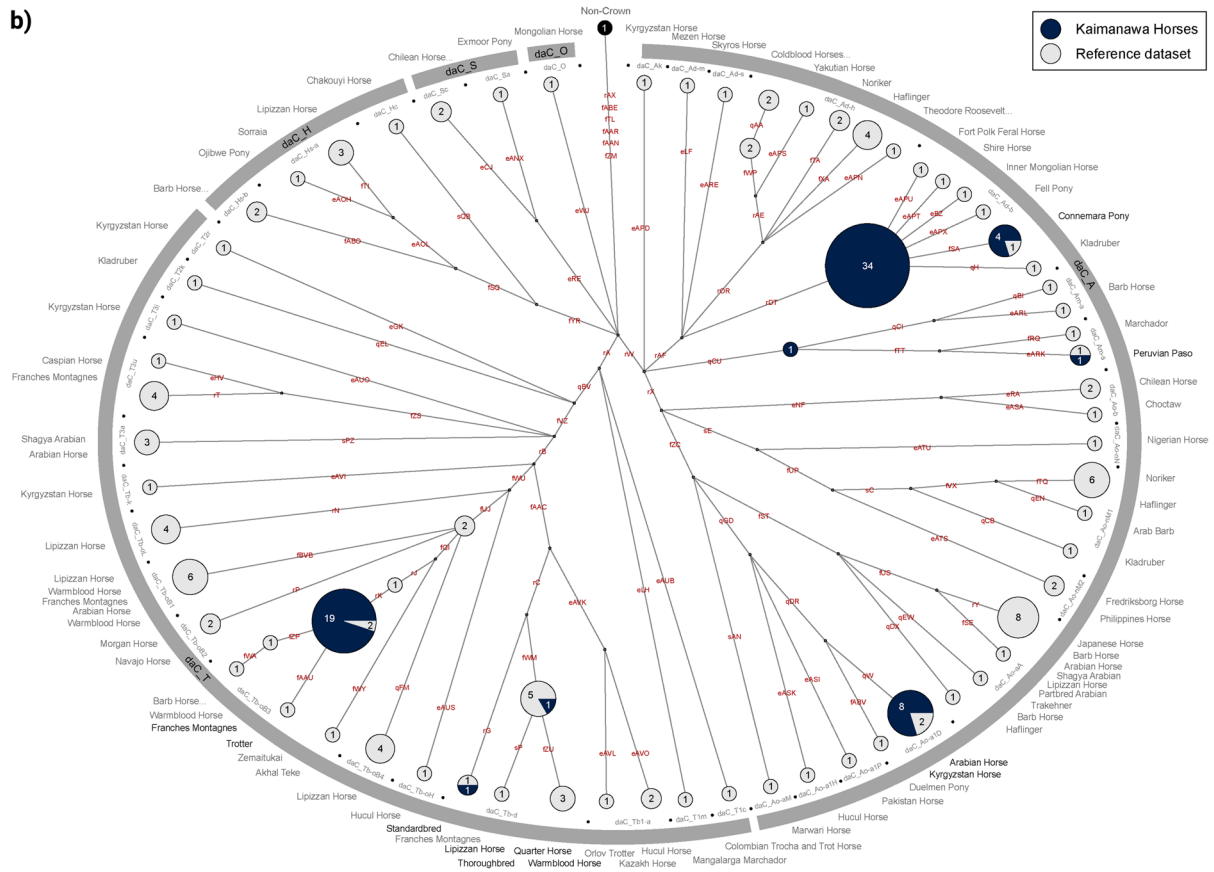


Fig. 5 **a** Median-Joining (MJ) network of mtDNA haplogroups (A–R) in 133 Kaimanawa Horses, 81 global modern horses [7] and one donkey (Equus asinus, NC_001788.1; (Xu et al.) [69]), based on 1,078 SNPs from the Combined mtDNA Dataset. Circles represent haplotypes, with size proportional to haplotype frequency and colors indicating geographic origin. The number of Kaimanawa Horse individuals per haplotype is shown, and branch labels indicate the number of nucleotide differences. **b** Median-Joining (MJ) network of Y-chromosome haplotypes based on 108 male-specific Y-chromosome variants from the Combined Y-chr Dataset, including 37 SNPs from the GGP Equine v4 array, 14 sub-lineage-specific SNPs from KASP genotyping, and 57 imputed variants. The network includes 69 related Kaimanawa Horses (this study: n = 26; (Sharif et al. 2022) [30]: n = 43) and 120 reference horses representing Crown group lineages (daC_), and one non Crown Icelandic Horse as an outgroup [30, 76]. Haplotypes are represented by circles, with circle size and labels indicating the number of individuals. Kaimanawa Horses are highlighted in dark blue. Breeds from the reference panel sharing the same haplotype as the Kaimanawa Horses are shown in black font, whereas all other reference breeds are shown in gray font

previously described [72, 77]. Details on the MSY markers, genotypes, variant classifications, ancestral/derived allele states, and inferred HGs are provided in Table S11.

Consistent with our previous study [30], six Crown HGs were detected in Kaimanawa Horses, dominated by daC_Ad-b, followed by daC_Tb-o, daC_Ao-a, daC_Tb-d, and daC_Am (Fig. 5b; Table S1 and S11). The most frequent lineage was daC_Ad-b* (n = 38; 55%), with four individuals further resolved to daC_Ad-bA (KASP variant: fSA), previously reported in British Isles riding and light-draft breeds with Spanish ancestry (e.g., Welsh B/D, Connemara Pony, Kerry Bog Pony). The second most common lineage was the Thoroughbred-associated daC_Tb-oB3b1 (n = 19; 27.5%), widely observed among modern breeds with Thoroughbred ancestry including Thoroughbred, Quarter Horse, Standardbred, Barb, Franches Montagnes, Shetland Pony, Trotter, and more.

Additional paternal lineages included the Arabian-associated daC_Ao-a1D2 (n = 8; 11.6%), as well as daC_Tb-dW* (n = 1; 1.4%) and daC_Tb-dM (n = 1; 1.4%), linked to Thoroughbred and Standardbred ancestry, respectively. Two individuals carried daC_Am*, a lineage associated with Spanish-influenced breeds such as Paso Fino, Paso Peruano, Barb, and Mangalarga Marchador [76]. KASP markers (variants: fTT and eARK) further resolved one of these to daC_Am-sP, a HT shared with Paso Peruano and Marajó horses [72].

Discussion

Genome-wide genetic relationships and founder origins of Kaimanawa Horses

The ML tree revealed clear genetic differentiation among the 24 horse breeds, with Kaimanawa Horses forming a distinct cluster closely associated with North American riding breeds, including Quarter Horses, American Saddlebreds, and Morgan horses. This relationship likely reflects historical introgression from 19th- and early 20th-century cavalry mounts and light utility horses that contributed to the founding population. One outlier, an 18-year-old bay-colored male Kaimanawa Horse, clustered with Hanoverian and Quarter Horses and showed minimal genetic similarity to the main Kaimanawa group. This individual carried rare mtDNA and Y-chr HGs (L and daC_Tb-dw, respectively), with related Y-chr sub-HGs previously reported in Hanoverian (daC_Tb-dW1) and Quarter Horses (daC_Tb-d, daC_Tb-oB1, and daC_Tb-oB3b) [72], suggesting that it may represent residual genetic variation from the heterogeneous founding population, or even though unlikely the result of undocumented admixture. However, historical or pedigree information would be required to confirm this interpretation.

PCA positioned Kaimanawa Horses intermediate among draft, riding, and pony breeds, reflecting their mixed genetic background. This pattern is consistent with historical records documenting the introduction of diverse horse types to New Zealand, including Welsh ponies, Thoroughbreds, Arabians, and other domestic breeds. Admixture analyses further supported contributions from hotblood (Arabian and Thoroughbred), coldblood (Clydesdale), pony (Shetland), and Standardbred lineages. While Kaimanawa Horses shared genetic components with several breeds at lower K values, they formed a distinct cluster from K ≥ 6, with higher-resolution analyses

revealing affinities to Clydesdale, Fell Pony, Thoroughbred, Quarter Horse, and Arabian breeds. The separation into two genetic components at higher K values was also supported by PC-AiR analyses.

Together, these findings support a multi-breed founder origin followed by differentiation through isolation and genetic drift [30, 83]. Although Arabian horses likely contributed to the founding population, the heterogeneous nature of the available Arabian reference panel limits direct historical interpretation. The detected Arabian-related component should therefore be considered a proxy for broader Arabian genetic influence rather than a representation of specific imported Arabian lineages. Given the diversity among Arabian subpopulations [42, 84] and potential changes in breed contributions over time, further comparisons using geographically and historically matched Arabian populations will be required to refine these inferences.

Maternal and paternal lineage contributions in Kaimanawa Horses

Despite expanding the previous mtDNA dataset by 43 individuals compared with [30], mitochondrial diversity remained low, indicating that the Kaimanawa population retains a restricted maternal lineage pool shaped by founder effects and prolonged genetic drift. HG-G and -P dominated the population, with two newly identified private haplotypes within these groups, while previously rare haplogroups (A, B, and N) were not detected, suggesting the loss of low-frequency maternal lineages over time.

Analysis of 108 MSY variants revealed diverse paternal ancestry across six sub-HGs, dominated by daC_Ad-b and daC_Tb-oB3b1. A subset of daC_Ad-b lineage (daC_Ad-bA) links to British and light-draft horses with Spanish-influenced ancestry (e.g., Welsh pony), while the prevalence of unresolved daC_Ad-b* suggests Kaimanawa-specific HTs. Additional Thoroughbred (daC_Tb-oB3b1) and Thoroughbred- and Arabian-associated lineages (daC_Tb-dW* and daC_Ao-a1D2) were also detected in low frequencies. Two individuals carried Spanish-influenced HTs daC_Am*, including one resolved as daC_Am-sP, consistent with shared ancestry with Paso-type breeds. Overall, the paternal background reflects predominantly coldblood, Spanish-influenced, and Thoroughbred origins, with minor Arabian input. Together, the contrasting maternal and paternal patterns suggest a heterogeneous founding population followed by lineage loss and drift.

Although historical records document Exmoor pony introductions to the Hawke's Bay region (Fig. 2), our genomic analyses provide little evidence for a substantial Exmoor contribution to the modern Kaimanawa population. Instead, Exmoor ponies showed the greatest genetic differentiation from Kaimanawa Horses among the evaluated breeds, suggesting that any historical introgression was limited, localized, or subsequently reduced by genetic drift.

Genetic diversity, population differentiation and inbreeding in Kaimanawa Horses

Across 24 breeds, Kaimanawa Horses show moderate genetic diversity and clear population structure, consistent with a feral origin, multiple founders, and long-term geographic isolation within the Kaimanawa Mountains. Bounded by State Highway 1, the Waipakihi Stream, the Rangitikei River, and surrounding sheep stations near Waiouru, this confined habitat likely contributed to the distinct admixture history and reduced genetic variability observed.

Pairwise F_{ST} values primarily reflect overall genetic differentiation among breeds rather than direct evidence of founder contributions to the Kaimanawa population. The highest divergence was observed between Kaimanawa Horses and Exmoor ponies, consistent with strong drift in this small, historically isolated breed. Conversely, the lowest F_{ST} values were observed with Mongolian and Hanoverian horses. As neither breed is documented among Kaimanawa founders, these affinities likely reflect shared ancestral variation and differences in demographic history rather than recent gene flow or founder input.

Although Kaimanawa Horses exhibited high overall genomic inbreeding (mean $F_{ROH} = 0.156 \pm 0.060$, ranking third among the evaluated breeds), the distribution of ROH across length classes indicates a predominantly recent origin of autozygosity. Low F_{ROH} values for short ROH segments (0–4 Mb) suggest limited historical inbreeding,

whereas the high contribution of long ROH (> 8 Mb) indicates recent demographic effects. Among the 24 breeds examined, Kaimanawa Horses ranked third for the 8–16 Mb class and highest for segments > 16 Mb, demonstrating an excess of long ROH relative to other breeds. This pattern suggests that recent population decline and/or prolonged reproductive isolation have been major drivers of genomic inbreeding in Kaimanawa Horses, outweighing the influence of more historical demographic processes.

Within-population genetic diversity and inbreeding in Kaimanawa Horses

Kaimanawa genetic subgroups differ markedly in population census size, genetic diversity, and management history. Argo Valley shows the highest inbreeding and lowest diversity, followed by Zone 20 and Southern Zone (Home Valley), whereas Southern Zone (Upper 14/Zone 15) exhibits the lowest inbreeding and highest diversity. Current census population sizes are ~44 (Argo Valley), ~140 (Zone 20), and ~290 (Southern Zones).

The elevated inbreeding in Argo Valley is consistent with its long-term geographic and social isolation, with limited movement across the Moawhango River and persistent family band structure. These patterns align with field observations [28] and are more likely the result of cumulative demographic processes, including prolonged isolation and historical population management that have constrained gene flow over time. For context, the most recent management (2025 muster) reduced the population from 67 to 44 individuals, and ongoing removal, rehoming, and immunocontraception (GonaCon) have further reduced foaling rates, (from ~20–25% to ~10% by 2025). Although these recent interventions are unlikely to be reflected in the genetic patterns observed here, they may further influence the genetic diversity and N_e of this subpopulation in the future.

In contrast, Southern Zone (Upper 14/Zone 15) exhibited the lowest inbreeding and highest genetic diversity among subpopulations. Limited removals since 2018 allowed this group to expand to ~420 individuals by 2025 before declining to ~290 following the April 2025 muster [28]. Its greatest pairwise F_{ST} differentiation from other zones likely reflects reduced gene flow caused by landscape fragmentation. The Moawhango River may further contribute to this structure by separating Argo Valley and Zone 20 from Southern Zones. Zone 20 and Southern Zone (Home Valley) showed intermediate levels of inbreeding and diversity, with Zone 20 declining from ~180 to ~140 individuals after the 2025 removals. Together, these patterns demonstrate the combined influence of population size, landscape structure, and management practices on genetic variation within the Kaimanawa population.

Effective population size decline and conservation implications

The LD-based N_e estimate for Kaimanawa Horses (~42) is lower than that reported for structured domestic breeds such as Asil Horses (N_e = 113; [85]), but comparable to small or semi-feral populations including Bardigiano (N_e = 39; [66]) and Sable Island horses (N_e = 48; [23]). Similar to these populations, Kaimanawa Horses have limited reproductive management, where social hierarchy and unequal reproductive success likely reduce the number of effective breeders. The discrepancy between census size (~450) and N_e (~48) suggests that reproduction is concentrated among a subset of individuals, increasing the effects of genetic drift and inbreeding and potentially constraining long-term adaptive potential.

Temporal LD-based analyses indicate a gradual N_e decline over approximately 17 generations, with a stronger reduction around 2008. Although this coincides with the major population reduction in 2009 (~40% decrease from ~500 to 300 horses), these estimates reflect cumulative demographic processes rather than the direct genetic impact of a single management event. The decline likely reflects the combined effects of founder events, prolonged small population size, and genetic drift, which promote allele loss and reduced genetic diversity [63]. These effects may be further intensified in free-ranging populations by unequal reproductive success and social structure [86, 87].

The overall N_e estimate assumes a single panmictic population; however, our analyses identify at least two genetically differentiated and partially isolated groups. If K1 and K2 represent partially isolated groups, the

effective number of breeders may be considerably lower than expected from the total census size, as the population is effectively partitioned into smaller genetic units. Therefore, even though separation between these groups may not be necessary for such a small, recently founded population, recognizing this genetic structure is important for informed management plans. Conservation strategies should balance connectivity between groups while preserving existing genetic variation, particularly to minimize the loss of rare alleles and further reductions in genetic diversity.

Geographic genetic structure and management implications

As mentioned above, across nuclear and uniparental markers, two distinct genetic subgroups (K1 and K2) with possibly different N_e were consistently identified, corresponding to geographic distribution and field observations. K1 is primarily associated with the Southern Zone (Upper 14/Zone 15), whereas K2 comprises Argo Valley, Southern Zone (Home Valley), and Zone 20. This subdivision is supported by differences in the frequencies of the dominant maternal (HG-P and -G) and paternal (daC_Ao-a1D2, daC_Ad-b*, and daC_Tb-oB3b1) lineages, as well as phenotypic differences, including the exclusive occurrence of grey horses in Upper 14/Zone 15 zone (K1 group). Together, these findings indicate restricted gene flow and long-term population subdivision. This pattern is consistent with the documented introduction of a grey Arabian stallion into the Southern Zones during the 1960s [29] and our previous observation that Arabian ancestry was predominantly associated with maternal HG-P [30]. However, the relationship between Arabian ancestry and grey coat colour was not formally evaluated in the present study.

Based on our previous uniparental analyses [30], and the current genome-wide SNP data, the Kaimanawa Horse population shows evidence of two partially isolated genetic groups. This information should be effectively communicated to the Department of Conservation in order to be incorporated into future management planning.

Interpreting population genomic analysis

Although the population genetic analyses consistently support a mixed founder origin for the Kaimanawa Horse population, each method captures different aspects of genetic history and therefore is not expected to produce identical patterns. Phylogenetic analysis reflects overall genomic similarity, PCA summarizes the principal axes of genetic variation, Admixture estimates the proportion of shared genetic components under different assumptions of population structure, pairwise F_{ST} quantifies allele frequency differentiation between populations, and mtDNA, and Y-chr HTs trace uniparental lineages. Consequently, these methods provide complementary rather than equivalent insights, particularly in a recently established feral population shaped by multiple founder breeds, subsequent admixture, isolation, and genetic drift.

Taken together, our analyses, based on the current sample set, support a multi-breed founder origin for the Kaimanawa Horse population, with variable contributions of founder lineages among individuals. The identified genetic components are associated with several British and European riding and draft horse lineages, as well as Arabian-related components, consistent with historical records. However, apparent affinities to individual reference breeds should not be interpreted as evidence of direct ancestry from these breeds, but rather as reflecting shared genetic components or genetic similarity resulting from historical admixture, incomplete representation of founder populations, and post-founding genetic drift.

Methodological Limitations

This study has several limitations, mainly associated with sampling, the unique characteristics of the feral horse population, the reduced marker density of the Combined Austosome SNP dataset, and the scarcity of genomic data available for New Zealand horse populations.

Firstly, sampling in this study was opportunistic rather than systematic, with individuals collected from horses permanently removed during musters conducted between 1998 and 2022. Samples were unevenly distributed across years and capture zones, meaning that the dataset may not fully represent the contemporary feral population. Furthermore, the limited number of samples available within specific time periods and geographic regions restricted our ability to robustly assess temporal trends and fine-scale spatial genetic structure. Nevertheless, no evidence of genetic clustering by muster year was detected, suggesting that the inclusion of older samples did not substantially influence the overall genomic patterns observed. Although rehoming decisions were not based on behavioural traits, adopter preferences for particular age or sex classes may have introduced some degree of sampling bias. Finally, because effective population size (N_e) was estimated using all unrelated individuals, the resulting estimates should be interpreted as reflecting the historical demographic dynamics of the sampled population rather than the contemporary N_e of recent generations.

Secondly, another limitation stems from the relatively low marker density of the Combined Autosomal SNP Dataset (approximately 18,000 SNPs), which decreased the resolution for detecting short ROH segments. Although the minimum physical length threshold was set to 1 kb, the requirement for at least 30 consecutive SNPs increased the effective minimum detectable ROH length to approximately 3–4 Mb. Consequently, the absence of ROHs and the resulting F_{ROH} value of zero in the 0–2 Mb category should be interpreted as a methodological limitation arising from the marker density and ROH-calling criteria, rather than as evidence that short ROHs are absent in Kaimanawa and domestic horses.

Finally, the lack of whole-genome reference data from Kaimanawa Horses across different geographical regions/zones may have limited the resolution of genetic component inference. However, this limitation was partly addressed by incorporating a comprehensive global reference panel encompassing the major breed groups historically associated with the formation of the Kaimanawa Horse population.

Future studies based on whole-genome sequencing data, which is currently underway, will be required to accurately characterize short ROH segments (<2 Mb) in Kaimanawa horses, as well as providing greater resolution of the population's evolutionary history and identify population-specific genetic variation.

Conclusion

This study provides the first autosomal SNP-based genomic characterization of the feral Kaimanawa Horse, establishing an important genomic resource for equine evolutionary and conservation research. By integrating genome-wide SNP and uniparental haplotype data, we reveal a complex genetic composition consistent with a multi-breed founder origin, with contributions from British and European riding and draft horse lineages, as well as Arabian-related genetic components. We further demonstrate distinct genetic subdivision and a demographic history shaped by isolation, genetic drift, and restricted gene flow. Despite elevated relatedness, inbreeding, and a small effective population size, Kaimanawa Horses retain moderate genetic diversity, supporting their conservation value and providing a foundation for future studies of adaptive variation and evidence-based genetic management.

Abbreviations

12S rRNA	Small subunit ribosomal ribonucleic acid
16S rRNA	Large subunit ribosomal ribonucleic acid
AI	Artificial Intelligence
AKTE	Akhal Teke
ANDE	Andean Horses
ARAB	Arabian
ATP6	ATP synthase F0 subunit 6

ATP8	ATP synthase F0 subunit 8
BELG	Belgian
bp	Base pairs
C	Recombination rate
CASP	Caspian
CLYD	Clydesdale
COX1-3	Cytochrome c oxidase subunits I, II, and III
CR / D-Loop	Control Region / Displacement Loop
CYTB	Cytochrome b
EXPO	Exmoor Pony
F	Individual inbreeding coefficient
FDR	Benjamini-Hochberg False Discovery Rate
FEPO	Fell Pony
FID	Family identifier
FINN	Finnhorse
FRMO	Franches-Montagnes
FROH	Inbreeding coefficient calculated from Runs of Homozygosity
FRTR	French Trotter
F_{ST}	Fixation index (population differentiation)
h	Total number of haplotypes
HAFL	Haflinger
HANO	Hanoverian
Hd	Haplotype diversity
H_e	Expected heterozygosity
HETHET	Proportion of SNPs heterozygous between two individuals
HG	Haplogroup
H_o	Observed heterozygosity
HT	Haplotype
IBS	Identical by state
IBS0	Proportion of SNPs where zero alleles are shared identical by state
ICEL	Icelandic
id	Identifier
IDB / IBD	Identical by descent
K	Number of hypothesized ancestral populations
KAHO	Kaimanawa Horses
kb	Kilo base pairs
KINSHIP	Kinship coefficient
LD	Linkage disequilibrium
LUSI	Lusitano
Mbp	Mega base pairs
MIPO	Miniature Pony
MJ	Median joining
ML	Maximum likelihood
MONG	Mongolian
MORG	Morgan
MSY	Male-specific region of the Y-chromosome
mtDNA	Mitochondrial deoxyribonucleic acid

na	Not applicable
ND1-6 / 4L	NADH dehydrogenase subunits
N_e	Effective population size
N_eS	Slope of the effective population size
NOFJ	Norwegian Fjord
ns	Not specified
PCA	Principal component analysis
PERC	Percheron
PRZE	Przewalski Horse
QUHO	Quarter Horse
r^2	Linkage disequilibrium decay
rRNA	Ribosomal ribonucleic acid
S	Total number of polymorphic sites
SABR	Saddlebred
SD	Standard deviation
SE	Standard error
SHIR	Shire Horse
SHPO	Shetland Pony
SNP	Single nucleotide polymorphism
STBR	Standardbred
STR	Short tandem repeat
THBR	Thoroughbred
tRNA	Transfer Ribonucleic Acid
tRNA-Ala	Transfer RNA for Alanine
tRNA-Arg	Transfer RNA for Arginine
tRNA-Asn	Transfer RNA for Asparagine
tRNA-Asp	Transfer RNA for Aspartic acid
tRNA-Cys	Transfer RNA for Cysteine
tRNA-Gln	Transfer RNA for Glutamine
tRNA-Glu	Transfer RNA for Glutamic acid
tRNA-Gly	Transfer RNA for Glycine
tRNA-His	Transfer RNA for Histidine
tRNA-Ile	Transfer RNA for Isoleucine
tRNA-Leu	Transfer RNA for Leucine
tRNA-Lys	Transfer RNA for Lysine
tRNA-Met	Transfer RNA for Methionine
tRNA-Phe	Transfer RNA for Phenylalanine
tRNA-Pro	Transfer RNA for Proline
tRNA-Ser	Transfer RNA for Serine
tRNA-Thr	Transfer RNA for Threonine
tRNA-Trp	Transfer RNA for Tryptophan
tRNA-Tyr	Transfer RNA for Tyrosine
tRNA-Val	Transfer RNA for Valine
v.	Version
WEPO	Welsh Pony
WGS	Whole Genome Sequence
Y-chr	Y-chromosome

θ_w	Watterson's theta based on segregating sites
π	Average nucleotide diversity per site

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1186/s12864-026-13345-y.

Acknowledgements The authors would like to thank the Kaimanawa Heritage Horse society (https://kaimanawaheritagehorses.org/) for their valuable collaboration on this project, providing us with permission to study New Zealand feral horses. We would like to thank Simone Frewin as representative of Kaimanawa Horse owners, facilitating the sampling procedure. We are grateful to all the Kaimanawa Horse owners for providing samples, photos and information on subject horses. Computational analyses were performed using the Life Science Compute Cluster (LiSC) at the University of Vienna. The project was hosted at the Department of Evolutionary Anthropology, University of Vienna, (Austria), and Konrad Lorenz Institute of Ethology, University of Veterinary Medicine, Vienna, (Austria). The authors thank the reviewers for their constructive comments and suggestions, which helped improve the quality of this manuscript.

Authors' contributions Conceptualization: E.M.; Methodology: A.B., S.S.A., M.B.S., B.W., R.R.F., P.O.-t., E.M.; Software: A.B., S.S.A., M.B.S.; Validation: A.B., S.S.A., M.B.S., B.W., E.M.; Formal Analysis: A.B., S.S.A., M.B.S.; Investigation: E.M.; Resources: E.M., K.W., B.W.; Data Curation: A.B., S.S.A.; E.M,; Writing, Original Draft Preparation: E.M., A.B., S.S.A.; Writing, Review and Editing: E.M., A.B., S.S.A., M.B.S., R.R.F., K.W., P.O.-t., and B.W.; Visualization: A.B., K.W.; Supervision: E.M,; Project Administration: E.M., K.W.; Funding Acquisition: E.M. All authors have read and agreed to the published version of the manuscript.

Funding This project was funded by the Austrian Science Fund (FWF; Grant No. P 37088-B; DOI: https://doi.org/10.55776/P37088) awarded to E.M, and the authors A.B. and S.S.A. were supported through the same FWF grant.

Data availability The SNP dataset generated and analyzed during the current study is available in the Figshare repository. To ensure the integrity of a complementary manuscript currently in preparation, the full dataset is currently hosted privately but is accessible to editors and reviewers via the following link: https://figshare.com/s/518a604db9506e9b9f2b. This dataset will be made fully public under a CC BY 4.0 license immediately upon publication. Additionally, specific data pertaining to the Y chromosome analyzed in this study are provided within the manuscript as Supplementary Table S7.

Declarations

Ethics approval and consent to participate The study was reviewed and approved by the Institutional Ethics and Welfare Committee of the University of Vienna (No. 2022-016) and the Kaimanawa Heritage Horse Welfare Society. Informed consent was obtained from all horse owners prior to the contribution of samples to this genetic study. All hair samples were imported into Austria following approvals (IDs: 2020−0.318.343, 2021−0.619.375; 2023−0.591.022) from the Federal Minister for Social Affairs, Health, Care and Consumer Protection (Der Bundesminister für Soziales, Gesundheit, Pflege und Konsumentenschutz).

Consent for publication Not applicable. As this study does not contain any individual person's data in any form (including any individual details, images, or videos), the requirement for consent for publication is not applicable. However, informed consent was obtained from each horse owner for the participation of their animals in this study and for the inclusion of the resulting genomic data in this publication.

Competing interests The authors declare no competing interests.

References

1. Scherf B, Pilling D. The second report on the state of the world's animal genetic resources for food and agriculture. 2015.
2. Hyland A. The horse in the ancient world. (No Title). 2003.
3. Librado P, Fages A, Gaunitz C, Leonardi M, Wagner S, Khan N, et al. The evolutionary origin and genetic makeup of domestic horses. Genetics. 2016;204:423–34.
4. Librado P, Gamba C, Gaunitz C, Der Sarkissian C, Pruvost M, Albrechtsen A, et al. Ancient genomic changes associated with domestication of the horse. Science. 2017;356:442–5.
5. Librado P, Khan N, Fages A, Kusliy MA, Suchan T, Tonasso-Calvière L, et al. The origins and spread of domestic horses from the Western Eurasian steppes. Nature. 2021;598:634–40.
6. Lippold S, Matzke NJ, Reissmann M, Hofreiter M. Whole mitochondrial genome sequencing of domestic horses reveals incorporation of extensive wild horse diversity during domestication. BMC Evol Biol. 2011;11:328.
7. Achilli A, Olivieri A, Soares P, Lancioni H, Hooshiar Kashani B, Perego UA, et al. Mitochondrial genomes from modern horses reveal the major haplogroups that underwent domestication. Proc Natl Acad Sci U S A. 2012;109:2449–54.
8. Petersen JL, Mickelson JR, Cothran EG, Andersson LS, Axelsson J, Bailey E, et al. Genetic diversity in the modern horse illustrated from genome-wide SNP data. PLoS ONE. 2013;8:e54997.
9. Fages A, Hanghøj K, Khan N, Gaunitz C, Seguin-Orlando A, Leonardi M, et al. Tracking five Millennia of horse management with extensive ancient genome time series. Cell. 2019;177:1419–e143531.
10. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, et al. DnaSP 6: DNA Sequence Polymorphism analysis of large data sets. Mol Biol Evol. 2017;34:3299–302.
11. Wallner B, Palmieri N, Vogl C, Rigler D, Bozlak E, Druml T, et al. Y Chromosome Uncovers the Recent Oriental Origin of Modern Stallions. Curr Biol. 2017;27:2029–35. .e5.
12. Wutke S, Sandoval-Castellanos E, Benecke N, Döhle H-J, Friederich S, Gonzalez J, et al. Decline of genetic diversity in ancient domestic stallions in Europe. Sci Adv. 2018;4:eaap9691.
13. Lever C. Naturalized mammals of the world. London, England: Prentice Hall; 1985.
14. McCort WD. Behavior of feral horses and ponies. J Anim Sci. 1984;58:493–9.
15. Aitken V, Kroef P, Pearson A, Ricketts W. Observations of feral horses in the southern Kaimanawa mountains. Report to the Wild Horse Committee, Palmerston North, New Zealand. 1979.
16. Linklater WL. Adaptive explanation in socio-ecology: lessons from the Equidae. Biol Rev Camb Philos Soc. 2000;75:1–20.
17. Cameron EZ, Linklater WL, Minot EO, Stafford KJ. Population dynamics 1994–1998 and management of Kaimaniwa wild horses.(Ed. J. Jasperse.) Science and Research Unit. Wellington, New Zealand: Science, Technology and Information Services, Department of Conservation; 2001.
18. Kobayashi I, Akita M, Takasu M, Tozaki T, Kakoi H, Nakamura K, et al. Genetic characteristics of feral Misaki horses based on polymorphisms of microsatellites and mitochondrial DNA. J Vet Med Sci. 2019;81:707–11.
19. King SRB, Schoenecker KA, Fike JA, Oyler-McCance SJ. Feral horse space use and genetic characteristics from fecal DNA. J Wildl Manage. 2021;85:1074–83.
20. Hendrickson SL. A genome wide study of genetic adaptation to high altitude in feral Andean Horses of the páramo. BMC Evol Biol. 2013;13:273.
21. Tollett CM, GENOMIC DIVERSITY AND ORIGINS OF. THE FERAL HORSES (Equus ferus caballus) OF SABLE ISLAND AND THE ALBERTA FOOTHILLS. 2018.
22. Colpitts J, McLoughlin PD, Poissant J. Runs of homozygosity in Sable Island feral horses reveal the genomic consequences of inbreeding and divergence from domestic breeds. BMC Genomics. 2022;23:501.
23. Uzans AJ, Lucas Z, McLeod BA, Frasier TR. Small N e of the isolated and unmanaged horse population on Sable Island. J Hered. 2015;106:660–5.
24. Thompson MA, McCann BE, Rhen T, Simmons R. Population genomics provide insight into ancestral relationships and diversity of the feral horses of Theodore Roosevelt National Park. Ecol Evol. 2024;14:e11197.
25. Mincham CJ. A social and cultural history of the New Zealand horse. PhD. Massey University; 2008.
26. Linklater WL, Cameron EZ, Minot EO, Stafford KJ. Feral horse demography and population growth in the Kaimanawa Ranges, New Zealand. Wildl Res. 2004;31:119–28.
27. Rogers GM, KAIMANAWA FERAL HORSES, AND THEIR ENVIRONMENTAL IMPACTS. N Z J Ecol. 1991;15:49–64.
28. Kelly Wilson KLF. Wild Kaimanawa Field Guide (2025 Edition). New Zealand: Kaimanawa Legacy Foundation; 2025.
29. Kelly Wilson KLF. Wild Kaimanawa Field Guide (2023 Edition). New Zealand: Kaimanawa Legacy Foundation; 2023.
30. Sharif MB, Fitak RR, Wallner B, Orozco-terWengel P, Frewin S, Fremaux M, et al. Reconstruction of the Major Maternal and Paternal Lineages in the Feral New Zealand Kaimanawa Horses. Anim (Basel). 2022;12. https://doi.org/10.3390/ani12243508.
31. O'Brien C, Darcy-Dunne MR, Murphy BA. The effects of extended photoperiod and warmth on hair growth in ponies and horses at different times of year. PLoS ONE. 2020;15:e0227115.

32. 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;81:559–75.
33. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4:s13742–015.
34. Hinrichs AS, Karolchik D, Baertsch R, Barber GP, Bejerano G, Clawson H et al. The UCSC Genome Browser Database: update 2006. Nucleic Acids Res. 2006;34 Database issue:D590–8.
35. Schaefer RJ, Schubert M, Bailey E, Bannasch DL, Barrey E, Bar-Gal GK, et al. Developing a 670k genotyping array to tag ~2 M SNPs across 24 horse breeds. BMC Genomics. 2017;18:565.
36. Todd ET, Fromentier A, Sutcliffe R, Running Horse Collin Y, Perdereau A, Aury J-M, et al. Imputed genomes of historical horses provide insights into modern breeding. iScience. 2023;26:107104.
37. Zhao S, Jing W, Samuels DC, Sheng Q, Shyr Y, Guo Y. Strategies for processing and quality control of Illumina genotyping arrays. Brief Bioinform. 2018;19:765–75.
38. Guo Y, He J, Zhao S, Wu H, Zhong X, Sheng Q, et al. Illumina human exome genotyping array clustering and quality control. Nat Protoc. 2014;9:2643–62.
39. 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.
40. Leal TP, Furlan VC, Gouveia MH, Saraiva Duarte JM, Fonseca PA, Tou R, et al. NAToRA, a relatedness-pruning method to minimize the loss of dataset size in genetic and omics analyses. Comput Struct Biotechnol J. 2022;20:1821–8.
41. Sjögren P, Wyöni P. Conservation genetics and detection of rare alleles in finite populations. Conserv Biol. 1994;8:267–70.
42. Cosgrove EJ, Sadeghi R, Schlamp F, Holl HM, Moradi-Shahrbabak M, Miraei-Ashtiani SR, et al. Genome diversity and the origin of the Arabian horse. Sci Rep. 2020;10:9702.
43. Poyato-Bonilla J, Sánchez-Guerrero MJ, Cervantes I, Gutiérrez JP, Valera M. Genetic parameters for canalization analysis of morphological traits in the Pura Raza Español horse. J Anim Breed Genet. 2021;138:482–90.
44. R Core Team. R: a language and environment for statistical computing (v. 4.4.2). Vienna, Austria: R Foundation for Statistical Computing; 2024.
45. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011;27:2156–8.
46. Biscarini F, Cozzi P, Gaspa G, Marras G. Others. Detect runs of homozygosity and runs of heterozygosity in diploid genomes. 2019.
47. McQuillan R, Leutenegger A-L, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, et al. Runs of homozygosity in European populations. Am J Hum Genet. 2008;83:359–72.
48. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32:268–74.
49. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14:587–9.
50. Wickham H. Getting started with ggplot2. In: Use R! Cham: Springer International Publishing; 2016. pp. 11–31.
51. Xu S, Dai Z, Guo P, Fu X, Liu S, Zhou L, et al. GgtreeExtra: Compact visualization of richly annotated phylogenetic data. Mol Biol Evol. 2021;38:4039–42.
52. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. Evolution. 1984;38:1358.
53. Wright S. The genetical structure of populations. Ann Eugen. 1949;15:323–54.
54. Pembleton LW, Cogan NOI, Forster JW. StAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. Mol Ecol Resour. 2013;13:946–52.
55. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J R Stat Soc Ser B Stat Methodol. 1995;57:289–300.
56. Salek Ardestani S, Mohandesan E. GENOVIS: a Python package for the visualization of population genetic analyses. BMC Genomics. 2026;27:190.
57. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009;19:1655–64.
58. Conomos MP, Miller MB, Thornton TA. Robust inference of population structure for ancestry prediction and correction of stratification in the presence of relatedness. Genet Epidemiol. 2015;39:276–93.
59. Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics. 2012;28:3326–8.
60. Gogarten SM, Sofer T, Chen H, Yu C, Brody JA, Thornton TA, et al. Genetic association testing using the GENESIS R/Bioconductor package. Bioinformatics. 2019;35:5346–8.
61. Zhang C, Dong S, Xu J-Y, He W-M, Yang T-L. PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. Bioinformatics. 2018;35:1786–8.
62. 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;6:109.

63. Corbin LJ, Liu AYH, Bishop SC, Woolliams JA. Estimation of historical effective population size using linkage disequilibria with marker data: Estimation of effective population size. J Anim Breed Genet. 2012;129:257–70.
64. Sved JA, Feldman MW. Correlation and probability methods for one and two loci. Theor Popul Biol. 1973;4:129–32.
65. Beeson SK, Mickelson JR, McCue ME. Exploration of fine-scale recombination rate variation in the domestic horse. Genome Res. 2019;29:1744–52.
66. Ablondi M, Dadousis C, Vasini M, Eriksson S, Mikko S, Sabbioni A. Genetic Diversity and Signatures of Selection in a Native Italian Horse Breed Based on SNP Data. Anim (Basel). 2020;10. https://doi.org/10.3390/ani10061005.
67. Pitt D, Bruford MW, Barbato M, Orozco-terWengel P, Martínez R, Sevane N. Demography and rapid local adaptation shape Creole cattle genome diversity in the tropics. Evol Appl. 2019;12:105–22.
68. Pagès H, Aboyoun P, Gentleman R, DebRoy S, Biostrings. Efficient manipulation of biological strings. R package version. 2019;2:10–18129.
69. Xu X, Janke A, Árnason Ú. The complete mitochondrial DNA sequence of the greater Indian rhinoceros, Rhinoceros unicornis, and the Phylogenetic relationship among Carnivora, Perissodactyla, and Artiodactyla (+Cetacea). Mol Biol Evol. 1996;13:1167–73.
70. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772–80.
71. Leigh JW, Bryant D, Popart. Full-feature software for haplotype network construction. Methods Ecol Evol. 2015;6:1110–6.
72. Radovic L, Remer V, Rigler D, Bozlak E, Allen L, Brem G, et al. The global spread of Oriental Horses in the past 1,500 years through the lens of the Y chromosome. Proc Natl Acad Sci U S A. 2024;121:e2414408121.
73. Felkel S, Vogl C, Rigler D, Dobretsberger V, Chowdhary BP, Distl O, et al. The horse Y chromosome as an informative marker for tracing sire lines. Sci Rep. 2019;9:6095.
74. Remer V, Bozlak E, Felkel S, Radovic L, Rigler D, Grilz-Seger G, et al. Y-Chromosomal Insights into Breeding History and Sire Line Genealogies of Arabian Horses. Genes. 2022;13. https://doi.org/10.3390/genes13020229.
75. He C, Holme J, Anthony J. SNP genotyping: the KASP assay. Methods Mol Biol. 2014;1145:75–86.
76. Bozlak E, Radovic L, Remer V, Rigler D, Allen L, Brem G, et al. Refining the evolutionary tree of the horse Y chromosome. Sci Rep. 2023;13:8954.
77. Lippold S, Xu H, Ko A, Li M, Renaud G, Butthof A, et al. Human paternal and maternal demographic histories: insights from high-resolution Y chromosome and mtDNA sequences. Investig Genet. 2014;5:13.
78. Bandelt HJ, Forster P, Röhl A. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol. 1999;16:37–48.
79. Kirin M, McQuillan R, Franklin CS, Campbell H, McKeigue PM, Wilson JF. Genomic runs of homozygosity record population history and consanguinity. PLoS ONE. 2010;5:e13996.
80. Ceballos FC, Joshi PK, Clark DW, Ramsay M, Wilson JF. Runs of homozygosity: windows into population history and trait architecture. Nat Rev Genet. 2018;19:220–34.
81. Cook DE, Andersen EC. VCF-kit: assorted utilities for the variant call format. Bioinformatics. 2017;33:1581–2.
82. Ablondi M, Vasini M, Beretti V, Superchi P, Sabbioni A. Exploring genetic diversity in an Italian horse native breed to develop strategies for preservation and management. J Anim Breed Genet. 2018;135:450–9.
83. Fleury W. Kaimanawa Wild Horse herd—draft management strategy. Wanganui, New Zealand: Department of Conservation; 1991.
84. Machmoum M, Boujenane I, Azelhak R, Badaoui B, Petit D, Piro M. Genetic diversity and population structure of Arabian horse populations using microsatellite markers. J Equine Vet Sci. 2020;93:103200.
85. Sadeghi R, Moradi-Shahrbabak M, Miraei Ashtiani SR, Schlamp F, Cosgrove EJ, Antczak DF. Genetic diversity of Persian Arabian horses and their relationship to other native Iranian horse breeds. J Hered. 2019;110:173–82.
86. Miller KA, Nelson NJ, Smith HG, Moore JA. How do reproductive skew and founder group size affect genetic diversity in reintroduced populations? Mol Ecol. 2009;18:3792–802.
87. Parreira BR, Gopalakrishnan S, Chikhi L. Effects of social structure on effective population size change estimates. Evol Appl. 2025;18:e70063.
88. Therneau TM, Grambsch PM. The cox model. Statistics for Biology and Health. New York, NY: Springer New York; 2000. pp. 39–77.

Authors and Affiliations

Arne Bielke[1,2] · Siavash Salek Ardestani[2,3] · Bilal Sharif[4,5] · Pablo Orozco-ter Wengel[6] · Robert Rodgers Fitak[7] · Barbara Wallner[8] · Kelly Wilson[9] · Elmira Mohandesan[1,2,3]

✉ Elmira Mohandesan
Elmira.Mohandesan@vetmeduni.ac.at

1 Department of Evolutionary Anthropology, University of Vienna, Vienna 1030, Austria

2 Human Evolution and Archaeological Sciences (HEAS), University of Vienna, Vienna, Austria

3 Department of Interdisciplinary Life Sciences, Konrad Lorenz Institute of Ethology, University of Veterinary Medicine Vienna, Vienna 1160, Austria

4 Centre for Palaeogenetics, Stockholm, Sweden

5 Department of Zoology, Stockholm University, Stockholm, Sweden

6 School of Biosciences, Cardiff University, The Sir Martin Evans Building, Museum Avenue, Cardiff CF10 3AX, Wales, UK

7 Department of Biology, Genomics and Bioinformatics Cluster, University of Central Florida, 4110 Libra Drive, Orlando, FL 32816, United States of America

8 Department of Biological Sciences and Pathobiology, Genetics and Animal Breeding, University of Veterinary Medicine, Vienna, Vienna 1210, Austria

9 Kaimanawa Legacy Foundation, 1505 Poihipi Rd, Oruanui, Taupo, New Zealand