Genome-wide variation reveal that goats were introduced into Asia via multiple migrations

Yonezawa, Takahiro ; Wu, Jiaqi ; Masuko, Ryo ; Iso, Kenta ; Nomura, Yuto ; Tabata, Risa ; Masaoka, Maho ; Ayin ; Kawaguchi, Fuki ;…

(Citation)
Scientific Reports,16(1):163

(Issue Date)
2025-12-21

(Resource Type)
journal article

(Version)
Version of Record

(URL)
https://hdl.handle.net/20.500.14094/0100499278

scientific reports

OPEN

Genome-wide variation reveal that goats were introduced into Asia via multiple migrations

Takahiro Yonezawa[1,13], Jiaqi Wu[1,2,13], Ryo Masuko[3], Kenta Iso[3], Yuto Nomura[3], Risa Tabata[3], Maho Masaoka[3], Ayin[3], Fuki Kawaguchi[3], Shinji Sasazaki[3], Aisaku Arakawa[4], Koh Nomura[5], Yukimizu Takahashi[5], Eiji Kobayashi[4], Manoj Kumar Shah[6], Muhammad Omar Faruque[7], Joseph S. Masangkay[8], Meirat Bakhtin[9], Polat Kazymbet[9], Tashi Dorji[10], Muhammad Ihsan Andi Dagong[11], Sri Rachma Aprilita Bugiwati[11], Johannes A. Lenstra[12] & Hideyuki Mannen[3]✉

In recent world-wide studies on the autosomal genetic diversity of goats, Asian goats were represented only by Southwest Asian, Pakistani and Chinese breeds. We have collected 55 K genome-wide SNP genotypes for 12 South/Southeast Asian and 2 central Asian goat populations, and inferred the origin and evolutionary history of Asian goats based on the population genomic analyses. Breed relationships, diversity clines, and coancestry patterns revealed two distinct migration routes separated by the Himalayan mountains: a northern route (Kazakhstan–Mongolia–Xinjiang) and a southern route (Bangladesh–Indochina). These routes tentatively parallel major human migration events across Eurasia. The migrations of goats converge into the Indochina goat populations, which then became the ancestors of the Philippine and Indonesian goats. Previous data on Y-chromosomal haplogroups indicate within the first group a separate migration of cashmere goats in eastern and northern China. Similarly, the southern route has been followed by two subsequent waves of goats, the first carrying the mitochondrial B haplogroup and in eastern Indochina associated with that Katjang type, and a later wave carrying exclusively the mitochondrial A haplogroup and associated in western Indochina with the Indian lop-eared trait with a roman convex facial profile. Haplogroup B in Indochina and Indonesia seems to be associated with tropical adaptation, whereas the Y1AB haplotype in northern China occurs at high frequency in cashmere goats, suggesting potential adaptation to arid environments. Together, these patterns point to a complex demographic history and diverse adaptive trajectories in Asian goats.

Goats (*Capra hircus*) are one of the most important livestock species, which has contributed to rural development in economically marginal areas by their high tolerance of harsh environments such as dryness and wetting. Goats are thought to have been domesticated first for meat supply, and later also for milk and wool[1,2]. Archaeological and molecular evidence indicate that domestication of goat took place 10,500–9500 years ago in the Taurus mountains in Turkey, the West Zagros mountains in Iran but also Baluchistan in Pakistanis[3–9].

Studies of goat mitochondrial DNA (mtDNA) revealed six divergent maternal lineages (haplogroups A, B, C, D, F and G), but these show only modest geographic differentiation. Haplogroup A is distributed worldwide[10], whereas the haplogroup B are predominant in Southeast Asia, with the frequency increasing toward the southeast[11,12]. The mtDNA haplogroup distribution of wild bezoar in the Middle East indicated

[1]Graduate School of Integrated Sciences for Life, Hiroshima University, Higashi-Hiroshima, Japan. [2]Department of Molecular Life Science, Tokai University School of Medicine, Isehara, Japan. [3]Graduate School of Agricultural Science, Kobe University, Kobe 657-8501, Japan. [4]Institute of Livestock and Grassland Science, National Agriculture and Food Research Organization, Tsukuba, Japan. [5]Faculty of Agriculture, Tokyo University of Agriculture, Atsugi, Japan. [6]National Swine Research Program, Nepal Agricultural Research Council, Dhankuta, Nepal. [7]Animal Breeding and Genetics, Bangladesh Agricultural University, Mymensingh, Bangladesh. [8]University of the Philippines, Los Banos, Philippines. [9]Institute of Radiobiology and Radiation Protection, Astana Medical University, Astana, Kazakhstan. [10]International Centre for Integrated Mountain Development, Kathmandu, Nepal. [11]Faculty of Animal Science, Hasanuddin University, Makassar, South Sulawesi, Indonesia. [12]Faculty of Veterinary Medicine, Utrecht University, Utrecht, The Netherlands. [13]Takahiro Yonezawa and Jiaqi Wu contributed equally to this study. ✉email: mannen@kobe-u.ac.jp

weak phylogeographic structure[8,9,13], but showed that bezoars with haplogroups A and B in Anatolia and North Iran, respectively, are the main maternal ancestors of Neolithic domestic goat with minor contributions of other haplogroups in central and east Iran. Ancient goat genome from Pre- and Post-Neolithic sites throughout the Near East, confirmed that domestication involved different bezoar populations and showed that after the Neolithic goat with haplogroup A moved to central and east Iran[13].

In contrast to the mtDNA haplotypes, the domestic Y-chromosomal haplotypes Y1AA, Y1AB, Y1B, Y2A, and Y2B variants have clearly different geographic ranges[14,15]. These haplotypes are related to available haplotypes from Anatolian and Iranian bezoars (CaY1, CaY1AA, CaY2 and CaY2B), but there are no haplotypes shared by the bezoars and the domestic populations. Y1AB and Y2A are found worldwide, whereas the Y1B is mainly in Northern Europe. Both Y1AA and Y2B have expanded in East and southeast Asia, respectively, whereas Y2A is dominant in west, east and south Africa[15], suggesting a numbers of severe male bottleneck during the colonization of Eurasia and Africa.

This is in line with the patterns autosomal 55 K SNP[16] and WGS[17] studies, which showed a strong regional partitioning of the of diversity. However, Southeast Asian goats (Fig. 1a)—key hubs for connecting South and East Asia—have not been included in these studies, leaving gaps in understanding the origin of mtDNA haplogroup B in Southeast Asia and the divergence of northern/southern migration routes. Here, we evaluate the diversity several south and southeast Asian goat populations. In combination with the mtDNA and Y-chromosomal diversity, this allows a reconstruction of separate eastward migration groups, which correlate with a phenotypic differentiation. Unlike uniparental markers (mtDNA/Y-chromosome) that only reflect single-lineage history, genome-wide SNPs enable us to capture comprehensive genetic structure. In addition to the established analyses by PCA, model-based clustering (ADMIXTURE) and genetic distances of breeds visualized in a tree or a Neighbor-Net graph, our newly developed Reduced Representation Admixture analysis (RRAA) further mitigates inbreeding bias in structure analysis.

Results

Phylogenetic and structure analyses of Asian goat populations

Prior to detailed analyses, we evaluated the genetic structure of Asian goats from a phylogeographic perspective (see Fig. S1 and Supplementary Text). Goats from Central and Northern Asia exhibited relatively homogeneous population structures at the country or provincial level (Fig. S1A). In contrast, goats from Southern and Southeast Asia displayed highly complex patterns (Fig. S1B). Consequently, for subsequent analyses, we defined genetically homogeneous populations according to the following the scheme described below. Cambodian goats were divided into Cambodia-M in a remote mountainous area east of the Mekong and Cambodia-P in the plain west of the Mekong, which clearly differ in morphology and in the frequency of the mtDNA haplogroup B (Fig. 1a)[11]. Similarly, Myanmar goats were divided into the Myanmar and Myanmar-Shan populations and Nepalese goats into the Nepal and Nepal-Hill populations, whereas a part of the Bangladesh goats were combined with those of Bhutan (Fig. S1B and Table S1). In this phylogenetic tree (Fig. S1B), the Bangladeshi goats were divided into three groups: one was closely related to the Bhutanese goats, and another was closely related to the Myanmar goats. Consequently, part of the Bangladeshi population was merged with the Bhutanese population to form the Bhutan–Bangladesh population, while another part was integrated into the Myanmar population. These integrations were justified because the pairwise F_{ST} values (0.008–0.019) were below the threshold for population differentiation ($F_{ST} > 0.05$), indicating genetic continuity. Similarly, a subset of the Bhutanese goats was also integrated into the Myanmar population ($F_{ST} = 0.04$) (Table S1). When defining genetically homogeneous populations, several exceptional individuals were excluded (see Supplementary Text S1).

PCA, ADMIXTURE, Neighbour-Net and TreeMix graphs of Eurasian and African goats (Figs. S2, S3, S4 and S5) consistently separate the goats from Europe, Africa and Asia, confirming previous studies[13,16]. The TreeMix graph suggests migration within Asian as well as within European/African populations at $m \leq 25$. Accordingly, we focused on bezoars and Asian domestic goats in subsequent analyses.

Based on the previously defined population scheme, each individual in the Asian dataset was assigned to its respective population. To reduce sampling bias between populations, each population was further subsampled to include no more than 30 individuals. The final Asian dataset of 474 individuals from 23 populations and 21,656 SNPs was analyzed by PCA, ADMIXTURE, a Neighbour-Net graph and TreeMix trees (Figs. 2, S2) consistently differentiate goats according to geographic origin. The PCA and the Neighbor-Net graph link the populations from northern Indochina (Myanmar-Shan, Laos Vietnam) to Jiangxi and Yunnan from southern China. In both analysis the Cambodia-M has an extreme position, which reflects its low heterozygosity by genetic isolation in a mountain region (Fig. 3). To exclude that this has disproportionate influence on the PCA pattern, we carried out supervised PCA[20] by calculating the PCs while excluding the Cambodia-M (Fig. 2a) and thus reducing the effects of local inbreeding event. However, this changed only the position of Cambodia-M and not the relative positions of the other populations.

Genetic structure analysis removing inbreeding bias

The results of the population structure analysis of Asian goats using ADMIXTURE are shown in Fig. S6. Cross-validation errors indicated that the model with $K = 13$ was the most appropriate. However, since the aim of this study is to understand the fundamental processes that have shaped the population structure of Asian indigenous goats, we focus our discussion on models assuming a smaller number of ancestral populations.

The genetic isolation of Cambodia-M also reflects the Admixture pattern. In the ADMIXTURE plots of $K = 4$ a separate genomic component is assigned to this population (Fig. 2b). However, because of its genetic isolation Cambodia-M is not a plausible ancestor of the surrounding populations that share with Cambodia-M the green color (Fig. 2b, $K = 4$), although as unique group of individuals in Hardy–Weinberg Equilibrium (HWE) it is inferred to be an ancestral component[21].

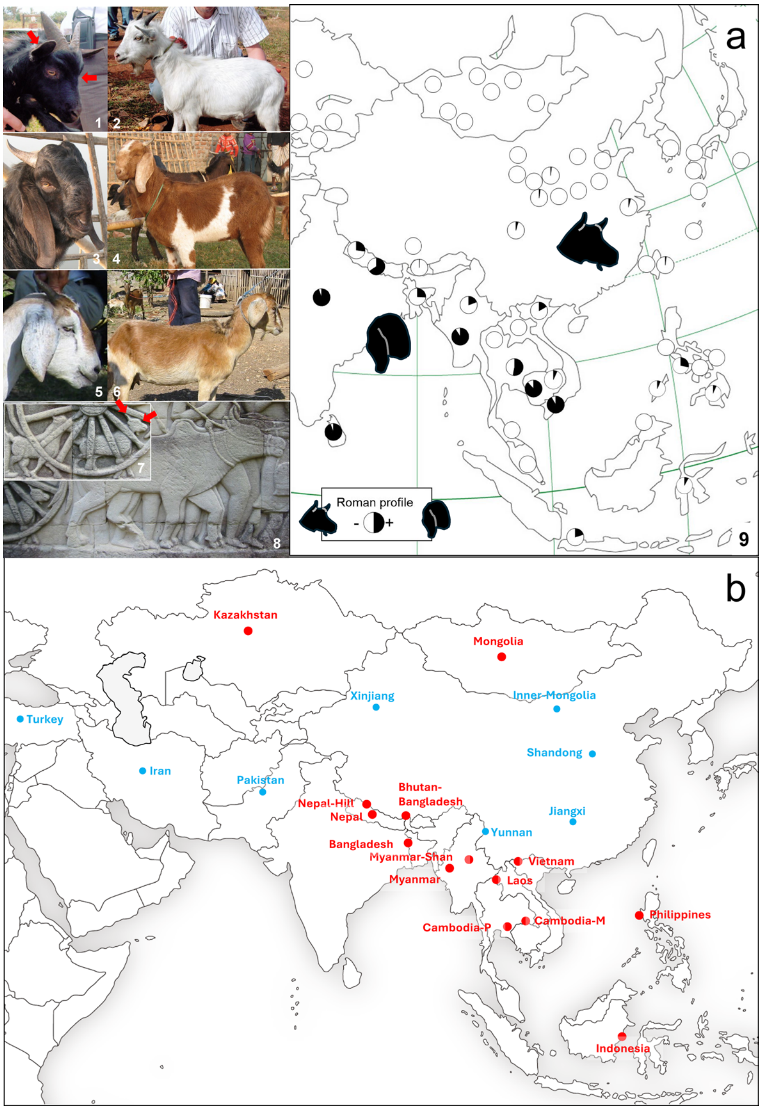

Fig. 1. Asian goats used in this study. (**a**) Morphology of Southeast Asian goats. **1, 2**: the Cambodian native goat in the mountain area (Cambodia-M) with prick (standing) ears and a concave facial profile (Red allows). **3, 4**: Nubian type of Indian Shirohi breed with pendent ears and a Roman convex facial profile. **5, 6**: the crossbred type of the Cambodian native goat in plain area (Cambodia-P). **7, 8**: Stone relief of goats in the Bayon, a Khmer temple at Angkor in Cambodia from the late twelfth century to early thirteenth century, showing prick ears, concave facial profile and small body size (Red allows). **9**: Distribution of the Roman facial profile in East Asia[11,18,19]. Quantitative phenotypic data showed that Katjang-type goats (mainly haplogroup B) in Cambodia-M had a mean body withers height of 51.2 ± 4.7 cm (adult female, n = 19) and 54.7 ± 4.1 cm (adult male, n = 2), and ear dropping ratio (drop/flat) of 0.176 (female and male, n = 99), while Roman convex-profile goats (mainly haplogroup A) in Cambodia-P had a mean body withers height of 57.3 ± 4.7 cm (adult female, n = 52) and 57.7 ± 5.0 cm (adult male, n = 27), and ear dropping ratio (drop/flat) of 0.978 (female and male, n = 99). (**b**) Red and blue circles indicate approximate locations of goat populations genotyped in this (red) and previous (blue) studies. The map was generated using Microsoft PowerPoint 2019 based on a free map website by 3kaku-K (https://www.freemap.jp/item/world/world1.html).

We consider this as an example of the "inbreeding bias", which confounds the identification of ancestral genomic components[22]. We reasoned that the inbreeding bias has the same effect as the "sampling bias"[22], by which populations with a large sample size may also at low K values be identified as ancestral genomic components. Therefore, reducing the sample size of the inbred populations is expected to counteract the effect of the inbreeding bias. Thus, we carried out the Admixture analysis 100 times with randomly reduced population sizes of the inbred populations (see Materials and Methods) and, after selection of runs with the same clustering patterns, averaged the percentage of genomic composition. The results of this reduced representation Admixture (RRAA) are shown in Figs. 2b and S7. At K = 2 and 3, the normal ADMIXTURE and RRAA results are identical, showing Cambodian-M as part of a southeast Asian gene pool. At K = 4, the normal ADMIXTURE shows

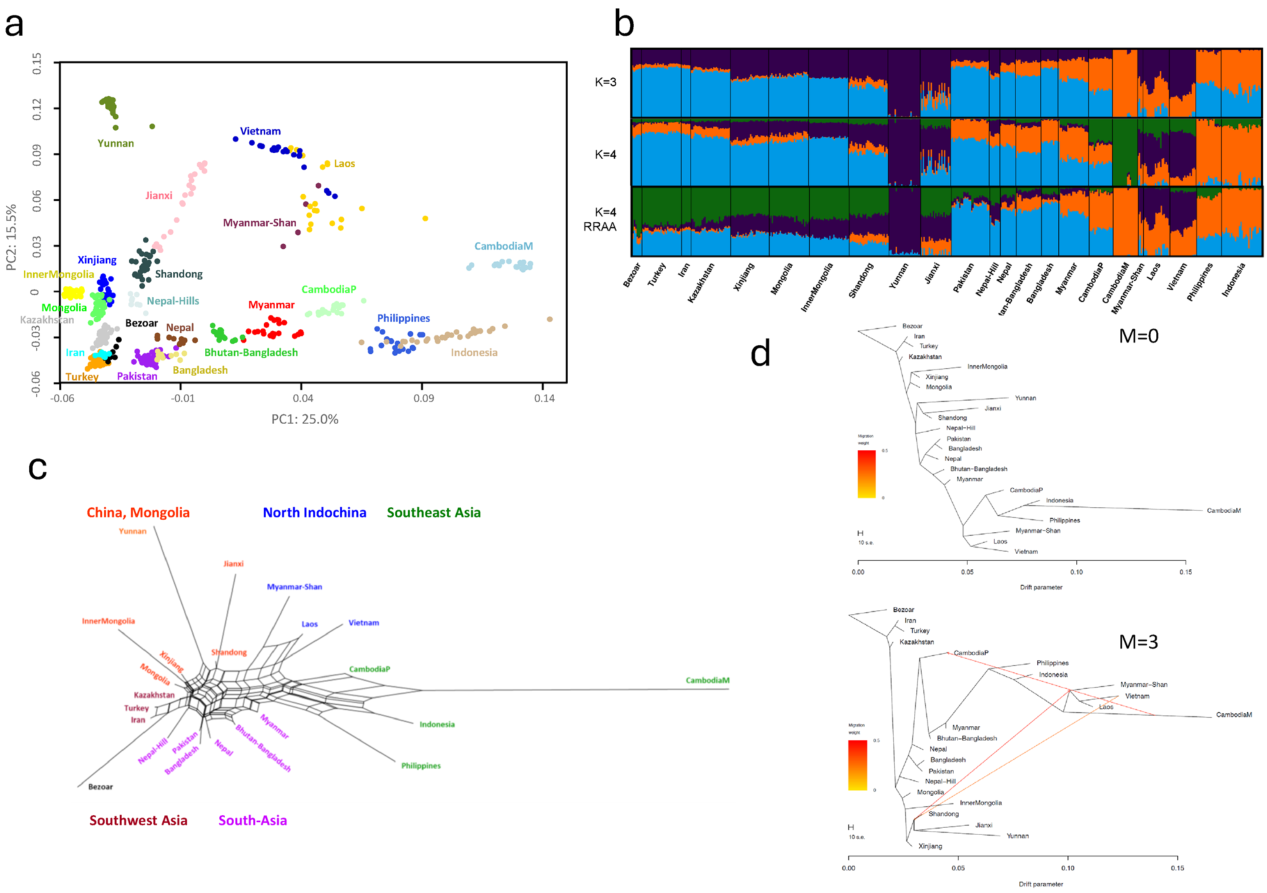

Fig. 2. Population genomic structuring and relationship among Asian goats. (**a**) Supervised Principal component analysis (svPCA) on Asian domestic goats and bezoars. Parts of Bangladesh and Bhutanese population were taken in Myanmar population. Furthermore, another part of Bangladesh population was merged with the remaining Bhutan population and treated as Bhutan-Bangladesh population (Table S1). First and second principal components are shown. (**b**) Model-based clustering by the ADMIXTURE program ver. 1.3 (K = 3, K = 4) and RRAA (K = 4), (**c**) Neighbor network of Asian goats. The edge lengths are proportional to the pairwise mean Fst distances between populations. (**d**) TreeMix graphs of the Asian goats without migration event (upper), with 3 migration events, which was selected as the best model by the optM (lower) and with bezoars as root of the tree. TreeMix at M = 3 was shown to adequately explain the Asian propagation routes. The migration edges are colored in accordance with their migration weights. Scale bar indicates the drift parameter.

influence of several populations to be influenced by Cambodia-M (Figs. 2b and S7). The RRAA pattern shows for Cambodia-M more plausibly the same pattern as the profiles for K = 2 and K = 3. Further, the dark-violet component in the ADMIXTURE and RRAA plots confirms an influence of Chinese goats on the Myanmar-Shan, Laos Vietnam population (Fig. 2b), which is also in agreement with the major migrations indicated by TreeMix (Fig. 2d).

Genetic clines and propagation route of Asian goats

To obtain additional evidence for genetic clines, we calculated for all pairs *i*, *j*. of Asian goat populations, the coancestry index f_3 (Turkish goats; *i*, *j*), which estimates the common ancestry of populations *i* and *j* after their divergence from the Turkish goats. These coancestries correlate negatively but incompletely with the genetic distances between populations *i* and *j*. Heat maps of the F_{ST} genetic distance and the f_3 coancestries (Fig. S8) show that the latter shows the most regular pattern, indicating that, just as the svPCA and RRAA, f_3 coancestries de-emphasize local effects as inbreeding and better shows the overall pattern.

Plotting the coancestries on a geographic map Fig. 3 visualizes a clear cline. Theoretically, for three populations A, B and C along the cline, an increase of coancestry relative to the outgroup O, (f_3 (O; B, C) > f_3 (O; A, B)) indicates a shared history of genetic drift of B and C after their divergence from A and before the divergence of B and C. Conversely, equal coancestries of A, B and C indicates a more stationary gene pool of A, B and C. In Fig. S9, we show the outgroup f₃ statistic between pairs of populations, regressed against the longitude of the more westerly population in each pair. A significant correlation was observed across all Asian populations ($r^2 = 0.32$, $p = 3.71 \times 10^{-19}$), reflecting the dispersal from West Asia to Eastern Eurasia. Furthermore, considering the population structure and phylogenetic relationships shown in Fig. 2, separate analyses of Central–Northern

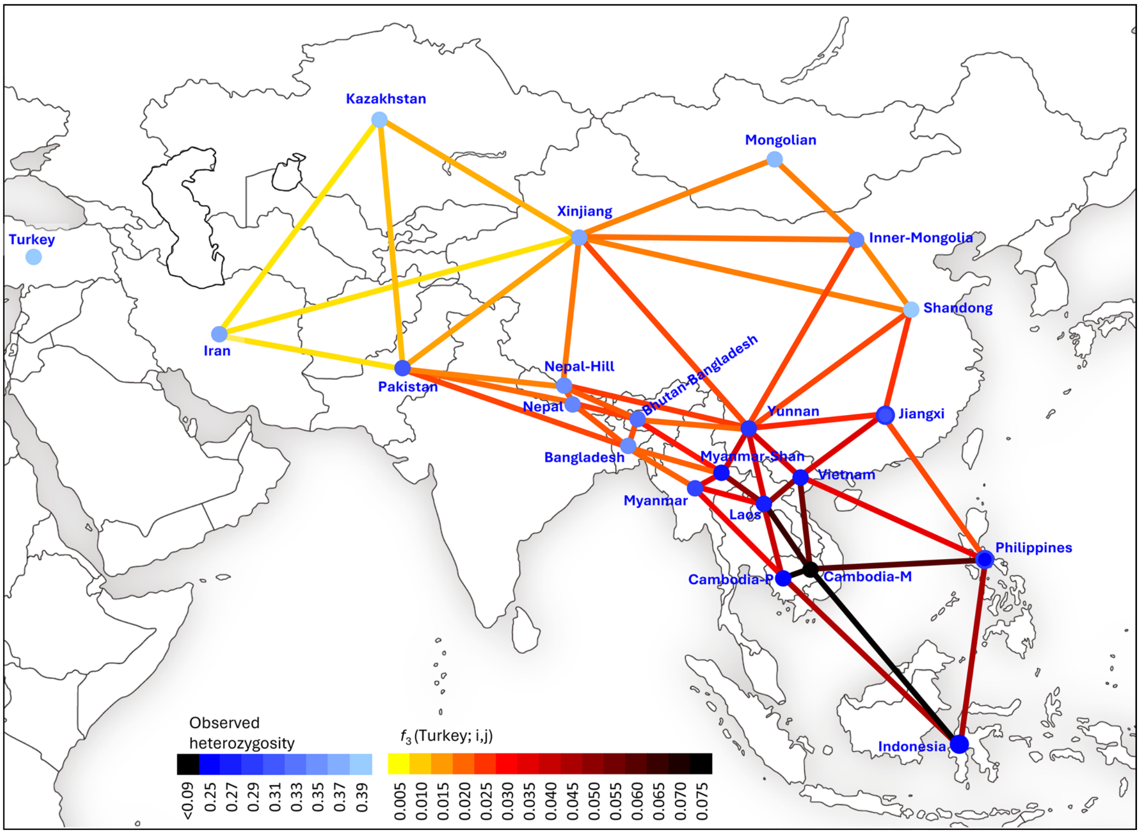

Fig. 3. Genetic clines of observed heterozygosity of populations and f3-based coancestries of pairs of populations since their divergence from Turkish goats. The map was generated using Microsoft PowerPoint 2019 based on a free map website by 3kaku-K (https://www.freemap.jp/item/world/world1.html).

Asia and Southern Asia revealed stronger correlations ($r^2 = 0.737$, $p = 5.23 \times 10^{-9}$ for Central–Northern Asia; $r^2 = 0.677$, $p = 4.26 \times 10^{-17}$ for Southern Asia). These results suggest distinct northward and southward genetic clines, with the effect sizes increasing when populations are analyzed separately. Thus, Fig. 3 indicates migrations from west to east and southeast with Indonesia and the Philippines as endpoints. This is more quantitatively shown in Fig. S10 for hypothetical migrations in northern China, a migration from Iran via Yunnan and Cambodia-M to southeastern Asia and a migration along the coast via Cambodia-P to the southeast. The large coancestries of Laos and Vietnam with both Indonesia and the Philippines indicate that goats in both island regions originate from east Indochina.

Figure 3 also shows the observed heterozygosities of populations. This follows the same trend as the coancestries but have a more outspoken north–south cline. The high heterozygosities in Mongolia and northern China as well as in north India together with homogeneous coancestries indicates a more stationary gene pool. However, long F_{ST} genetic distances (Fig. 2c) show a separate local development of Mongolian goats.

Discussion

Our population-genetic analyses of Asian goat genotypes consistently show a separation of northern and southern Asian goat populations. This most likely reflect different major propagation routes of Asian goats separated by the Himalayan mountains. The PCA pattern (Fig. 2a) suggests that the northern route mediated the spreading of Central-East Asian population goats through Kazakhstan, Mongolia, Xinjiang to the Chinese Middle Plain. The southern route follows the coast of the Indian Ocean. These migrations converge in the Indochina goat populations, which appear to be the ancestors of the Indonesian and Philippine Island goat populations. In contrast to the genetic landscape in north China and west Asia, we observe in southwest Asia sharp genetic clines of heterozygosities and coancestries with long genetic distances between populations (Figs. 2c, 3, S9).

The different migrations can tentatively be linked to human migrations. In the period of early domestication, expansion of livestock and crop plants may have occurred by the sharing of farm techniques, rather than large-scale migrations from a central core of populations across Southwest Asia[13,23–25]. However, the northern goat migration route overlaps geographically and temporally with the expansion of Yamnaya pastoralists (2500–2000 BCE), suggesting a potential association (e.g., driven by pastoral practices due to co-movement)[26] . This route also corresponds to the later northern Silk (Oasis) Road. The southern migrations may have coincided with the West Asian populations during the Early Harappan Period (3700–2800 BCE) of Indus Valley civilization[27]. Later (1500 BCE), Indo-Aryans migrated into the Indian subcontinent[28,29] and expanded the range of the Indo-

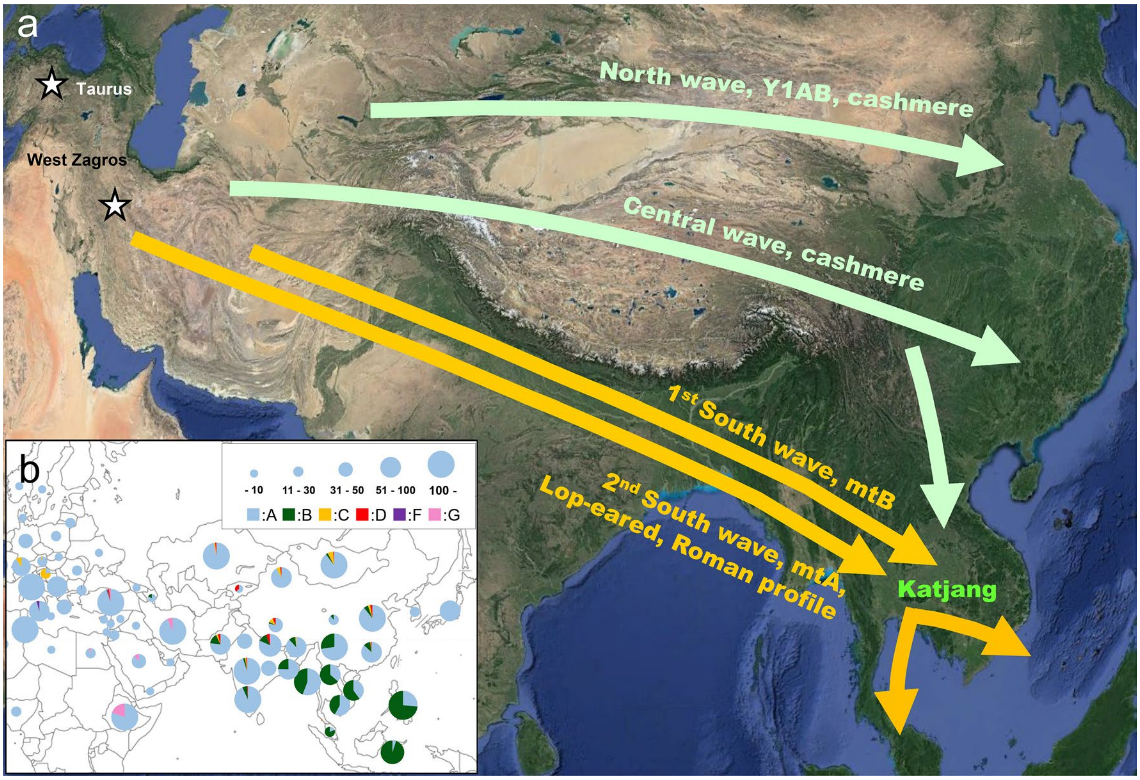

Fig. 4. Proposed migration routes of the Asian goats. (**a**) The star marks show the proposed domestication centers of goats[8,9] Light green allows indicate northern migration routes (Kazakhstan–Mongolia–Xinjiang), and yellow allows southern migration routes (Bangladesh-Indochina). (**b**) Distribution of the mtDNA haplogroups[18]. The map was generated using Microsoft PowerPoint 2019 based on a free map by Google Earth Pro (https://www.google.com/intl/ja/earth/about/versions/).

European language family. The migrations of goat may also have coincided with the immigration of taurine cattle into North China and of indicine cattle into South China, respectively[30,31].

The migrations from China to Indochina may have been mediated by the north–south migration events in East Asia after about 5000 years ago[32,33], the spread of the Sino-Tibetan language family from the Yangtze River area to Southeast Asia about 4000 years ago and the expansion of farmers around the Yellow River into southern China (about 2000–3600 years ago)[34,35].

Interestingly, previous data on the distribution of uniparental markers (mtDNA, Y-chromosomal) allow to refine the reconstruction of the history of Asian goats. The Y-chromosomal haplotype Y1AB is predominant in China north of the Yellow River. With only few exceptions all Y1AB carrying males are from cashmere goats[15,36]. This indicates a separate expansion of cashmere goats with predominantly the Y1AB haplotype. As noted above for the northern Chinese goats, the high heterozygosities, short genetic distance and modest levels of coancestry indicate that several cashmere goat populations in China constitute a relatively undifferentiated gene pool, which has diverged less from the southwest Asian goats than southwest Asian, African and European goats (Figs. 2c, S4).

In contrast most male goats south of the Yellow River and from India to the east carry haplotypes Y1AA or Y2B. In China and Southeast Asian countries, these are all from populations kept for the meat and have the small size of the Katjang goats[37], which has been depicted in a stone relief in Cambodia from around AD 1200 (Fig. 1a). The origin of these goats is not clear, but the small size probably contributes to the tropical climate adaptation and fits the requirements of smallholders[36]. This may apply as well to for central and eastern African dwarf goats.

The distribution of mtDNA haplogroups as markers of maternal lineages may also indicate historic propagation routes. MtDNA haplogroup A is dominant in all domestic goats worldwide and also in southern Asia, where it spread by both the northern and the southern routes. However, haplogroup B has a high frequency in Indochina and is even almost fixed in Indonesia (Fig. 4b)[18]. In Southwest Asia, haplogroup B has a low frequency in Iranian bezoars[9] and ancient Iranian goats[38] but is not found in Turkish goats. Therefore, the appreciable frequencies of haplogroup B in several southeast-Asian countries[18] indicates its descendance from a hypothetical Iranian goat population with at least a substantial haplogroup B frequency. Our data on autosomal diversity show that in Indochina the haplogroup B carrying goats have been influenced by south Chinese goats (Fig. 2)[39], although it is not clear if this is related to the spread of the Katjang phenotype (Fig. 1).

Remarkably, haplogroup B has not been observed in modern Iranian domestic goats and has a low frequency in India, where haplogroup A is now predominant. This indicates an expansion of haplogroup A carrying goats, which in Iran and India replaced the haplogroup B carrying goats[13]. The distribution of the goat mtDNA haplotype A[18] (Fig. 4b), suggests that spread of this haplotype coincided with the spread of Indo-European languages 1500 BCE by the Indo-Aryan migrations (see above). Haplotype A may also be associated with the morphology of the lop-eared sheep with large drooping ears and a roman convex facial profile (Fig. 1a). These traits are also observed in the Middle East and in South Africa (e.g., the Boer goats) and in the popular Anglo-Nubian breed[36]. Lop-eared goats are thought to have emerged in the region from Iran to India about 3500 years ago[40], which indeed coincides with the Indo-Aryan migrations. Together with the lop-eared/katjang intermediates, lop-eared sheep are typical for several Southwest Asian and southern Asian goat populations (Fig. 1a)[1–8]. As normal in livestock, a favored phenotype most likely spreads by importing males, which would explain why eastern and southern Indochinese lop-eared goats have retained a high haplogroup B frequency. Figure 4a shows the hypothetical migrations that are resolved by our present results. The 50 K goat SNP chip used in this study was designed based on population genomic data from six globally sampled goat breeds, including Katjang, and therefore it is unlikely that ascertainment bias for Southeast Asian goats is particularly stronger than that for populations from Europe or other regions. Nonetheless, analyses based on SNP chips designed from a limited number of breeds are inevitably affected by ascertainment bias to some extent. Consequently, further validation using whole-genome data and more detailed analyses will be necessary. Especially, future research would be important collecting ancient DNA samples from key regions (such as the Southern and Western Himalayas, and Indochina), with the collection time dating back to the Neolithic—Bronze Age (10,000 BC–2000 BC), in order to directly trace the migration time of goats and combine the archaeological goat remains with human remains to verify their common migration situation. Following the expansion or replacement of causative DNA variants would allow a functional interpretation of the historic developments. At the same time, livestock populations are subject to strong artificial selection, which requires careful consideration when reconstructing their evolutionary history from genomic data. In general, selection acts on specific loci and their surrounding regions, whereas demographic processes affect the genome as a whole. However, under strong artificial selection or during geographic dispersal, extreme inbreeding can occur, causing the effects of selection to appear across the entire genome. Consequently, standard principles commonly applied to wild populations or human groups may not always hold for livestock populations (Supplementary Text S2). The RRAA method proposed in this study accounts for such strong inbreeding and minimizes its impact on the analyses, making it an important tool for accurately understanding the evolutionary history of livestock. We conclude that the results of this study may contribute to our understanding of the multifaceted migration, expansion, and relationships of human language families in Southeast Asia.

Methods

Sample collection

We used DNA samples of 270 native goats obtained from 11 countries: Philippines (n = 20), Myanmar (n = 20), Laos (n = 20), Vietnam (n = 20), Cambodia (n = 40), Bangladesh (n = 20), Bhutan (n = 20), Nepal (n = 20), Kazakhstan (n = 30), Mongolia (n = 30) and Indonesia (n = 30) (Fig. 1b, Table S1). To avoid sampling related individuals, we first conducted interviews with farmers to exclude individuals with known parent–offspring or full-sibling relationships. The sample locations were apart each other (area, district, village, farm, hearing from famer and etc.) as possible.

Genotyping and data pruning

All samples were genotyped using Illumina goat SNP 50 K Bead chip (53,347 SNPs) (Illumina, Inc. San Diego, CA, USA). We combined these data with the previously published goat 50 K SNP data sets (Fig. 1b)[16,41], but without the cosmopolitan Angora, Boer, Saanen or the New World and the Oceanian populations. This resulted in genotypes of 3176 domestic goats from 42 countries and seven bezoars from Iran. To further consider the impact of a biased sample size on the estimation of population structure, etc., we subsampled from the Pakistani population (367 non-cosmopolitan goats) to 60 goats. After excluding SNPs with a call rate < 0.99 (i.e., missing rate > 1%) and individuals with a missing rate > 1% (with the exception of our own four Cambodian goats, which had up to 4% missing data), we performed LD pruning using PLINK (indep-pairwise: window size 50 kb, step size 5, r^2 threshold 0.04). The resulting global dataset comprised 21,931 SNPs and 2,839 individuals, representing 47 wild or domestic populations from 42 countries. Taking account of the phylogenetic structure (Fig. S1), we re-defined the Asian populations with the exclusions of the exceptional individuals. We then selected for an Asian datasets goats and bezoars from Asian countries with ≤ 30 goats per population. After excluding SNPs with > 0.01 missing individuals or a minor allele frequency in domestic goats of < 5% site and linkage disequilibrium pruning (plink indep-pairwise 50–5-0.04), the final dataset consisted of 21,656 SNPs with 474 goats/bezoars.

Genetic diversity and population structure analysis

Observed and expected homozygosities were calculated using Plink 1.9[42]. For detection of outliers and for assessing the differentiation within and across breeds, we visualized allele-sharing distance between individuals in Neighbor-joining trees (Fig. S1).

Pairwise weighted F_{ST} distances between breeds were calculated by VCFTOOLS 0.1.16[43] and visualized in Neighbour-Net graphs[44] inferred with the SPLITSTREE[45]. Probabilistic principal component analysis[46] (pPCA) was done with the R package PCAMETHOD version 3.14[47]. Supervised PCA (svPCA), in which the coordinated were calculated for a selection of the samples[20] was carried using Plink 1.9. Phylogenetic relationships reflecting possible admixture events were inferred using TreeMix 1.13[48] with m = 0–10 migration edges, using bezoars as the outgroup[38]. The optM package was employed to identify the optimal number of migration events[49]. Genetic population structure and admixture of 474 Asian goats from 23 populations were estimated using ADMIXTURE version 1.3.0[50] with K = 1–40 inferred ancestors. However, by the inbreeding bias[22], the most inbred populations are identified as ancestral cluster. Because overrepresented populations are at low k-values also inferred to be an ancestral cluster[22] (the sampling bias), we can, conversely, suppress the inbreeding bias by reducing the representation of the inbred breeds. This is the rationale of the RRAA (reduced representation Admixture analysis) procedure. We implemented it by randomly selecting three individuals from Cambodia-M (the most inbred population, observed heterozygosity = 0.11) and ten individuals each from Indonesia, Inner Mongolia, Jiangxi, Nepal, Philippines, Shandong, Turkey, Vietnam, Xinjiang, and Yunnan (all with observed heterozygosity > 0.2, based on preliminary results). Admixture was run for 100 iterations of such random selections, and the inferred genomic composition was averaged for each individual across replicates. The clustering patterns were > 90% consistent across iterations, confirming the stability of the method.

Data availability

The datasets generated and analysed during the current study are available in the Figshare repository, (https://doi.org/10.6084/m9.figshare.29035598). The code developed for the Reduced Representation Admixture Analysis (RRAA) and the datasets used in this study are available at the following repository: https://github.com/wujiaqi06/RRAA.

Received: 30 April 2025; Accepted: 14 November 2025

References

1. Sherratt, A. *Plough and Pastoralism: Aspects of the Secondary Products Revolution*. 261–305 (Cambridge University Press, 1981).
2. Sherratt, A. The secondary exploitation of animals in the Old World. *World Archaeol.* **15**, 90–104 (1983).
3. Meadow, R. H. *The origins and spread of agriculture and pastoralism In Eurasia*. 390–412 (Smithsonian Institution Press, 1996).
4. Peters, J., Helmer, D., Von Den Driesch, A. & Saña Segui, M. Early animal husbandry in the Northern Levant. *Paléorient* **25**, 27–47 (1999).
5. Zeder, M. A. & Hesse, B. The initial domestication of goats (*Capra hircus*) in the Zagros mountains 10,000 years ago. *Science* **287**, 2254–2257 (2000).
6. Zeder, M. A. Animal domestication in the Zagros: A review of past and current research. *Paléorient* **25**, 11–25 (1999).
7. Zeder, M. A. A metric alanalysis of a collection of modern goats (Capra hircus aegargus and Capra hircus hircus) from Iran and Iraq: implications for the study of caprine domestication. *J. Archaeol. Sci.* **28**, 61–79 (2001).
8. Luikart, G. et al. Multiple maternal origins and weak phylogeographic structure in domestic goats. *Proc. Natl. Acad. Sci. U S A* **98**, 5927–5932 (2001).
9. Naderi, S. et al. The goat domestication process inferred from large-scale mitochondrial DNA analysis of wild and domestic individuals. *Proc. Natl. Acad. Sci. U S A* **105**, 17659–17664 (2008).
10. Naderi, S. et al. Large-scale mitochondrial DNA analysis of the domestic goat reveals six haplogroups with high diversity. *PLoS ONE* **2**, e1012 (2007).
11. Lin, B. Z. et al. Molecular phylogeography and genetic diversity of East Asian goats. *Anim. Genet.* **44**, 79–85 (2013).
12. Tabata, R. et al. The Eurasian Steppe is an important goat propagation route: A phylogeographic analysis using mitochondrial DNA and Y-chromosome sequences of Kazakhstani goats. *Anim. Sci. J.* **90**, 317–322 (2019).
13. Daly, K. G. et al. Ancient goat genomes reveal mosaic domestication in the Fertile Crescent. *Science* **361**, 85–88 (2018).
14. Waki, A., Sasazaki, S., Kobayashi, E. & Mannen, H. Paternal phylogeography and genetic diversity of East Asian goats. *Anim. Genet.* **46**, 337–339 (2015).
15. VarGoats, C. et al. Geographical contrasts of Y-chromosomal haplogroups from wild and domestic goats reveal ancient migrations and recent introgressions. *Mol. Ecol.* **31**, 4364–4380 (2022).
16. Colli, L. et al. Genome-wide SNP profiling of worldwide goat populations reveals strong partitioning of diversity and highlights post-domestication migration routes. *Genet. Sel. Evol.* **50**, 58 (2018).
17. Denoyelle, L. et al. Genetic variations and differential DNA methylation to face contrasted climates in small ruminants: An analysis on traditionally-managed sheep and goats. *Front. Genet.* **12**, 745284 (2021).
18. Mannen, H. et al. Indonesian native goats (Capra hircus) reveal highest genetic frequency of mitochondrial DNA haplogroup B in the world. *Anim Sci J* **91**, e13485 (2020).
19. Mannen, H. In *Native Livestock in Asia* (ed The Society for Researches on Native Livestock (ed.)) 281–299 (The University of Nagoya Press, 2009).
20. Ciani, E. et al. On the origin of European sheep as revealed by the diversity of the Balkan breeds and by optimizing population-genetic analysis tools. *Genet. Sel. Evol.* **52**, 25 (2020).
21. Pritchard, J. K. & Donnelly, P. Case-control studies of association in structured or admixed populations. *Theor. Popul. Biol.* **60**, 227–237 (2001).
22. Lenstra, J. A. et al. Molecular tools and analytical approaches for the characterization of farm animal genetic diversity. *Anim. Genet.* **43**, 483–502 (2012).
23. Zeder, M. A. The Origins of Agriculture: New Data New Ideas. *Curr. Anthropol.* **52**, S221–S235 (2011).
24. Vigne, J. D. & Helmer, D. Was milk a "secondary product" in the Old World Neolithisation process? Its role in the domestication of cattle, sheep and goats. *Anthropozoologica* **42**, 9–40 (2007).
25. Fuller, D. Q., Willcox, G. & Allaby, R. G. Cultivation and domestication had multiple origins: arguments against the core area hypothesis for the origins of agriculture in the Near East. *World Archaeol.* **43**, 628–652 (2011).
26. Narasimhan, V. M. *et al.* The formation of human populations in South and Central Asia. *Science* **365**, aat7487 (2019).
27. Kenoyer, J. M. & Meadow, R. H. In *Corpus of Indus Seals and Inscriptions* Vol. 3 (eds A. Parpola, B.M. Pande, & Koskikallo P.) 44–58 (Suomalainen Tiedeakatemia, Helsinki., 2010).
28. Parpola, A. in *Archaeology and Language, III: Artefacts, languages and texts* (ed M. Spriggs) (Routledge, 1999).
29. Bryant, E. *The Quest for the Origins of Vedic Culture: The Indo-Aryan Migration Debate*. (Oxford University Press, 2001).
30. Chen, N. et al. Whole-genome resequencing reveals world-wide ancestry and adaptive introgression events of domesticated cattle in East Asia. *Nat. Commun.* **9**, 2337 (2018).
31. Zhang, M., Yan, S., Pan, W. & Jin, L. Phylogenetic evidence for Sino-Tibetan origin in northern China in the Late Neolithic. *Nature* **569**, 112–115 (2019).
32. Cai, X. et al. Human migration through bottlenecks from Southeast Asia into East Asia during Last Glacial Maximum revealed by Y chromosomes. *PLoS ONE* **6**, e24282 (2011).
33. Yang, M. A. et al. Ancient DNA indicates human population shifts and admixture in northern and southern China. *Science* **369**, 282–288 (2020).
34. Wang, C. C. et al. Genomic insights into the formation of human populations in East Asia. *Nature* **591**, 413–419 (2021).

35. Kutanan, W. et al. Complete mitochondrial genomes of Thai and Lao populations indicate an ancient origin of Austroasiatic groups and demic diffusion in the spread of Tai-Kadai languages. *Hum. Genet.* **136**, 85–98 (2017).
36. Porter , V., Alderson, L., Hall, S. J. G. & Sponenberg, D. P. *Mason's World Encyclopedia of Livestock Breeds and Breeding*. ISBN-13: 978 1-84592-466-8. (CABI, 2016).
37. Gao, J. et al. Genomic characteristics and selection signatures in Indigenous Chongming White Goat (*Capra hircus*). *Front Genet* **11**, 901 (2020).
38. Daly, K. G. *et al.* Herded and hunted goat genomes from the dawn of domestication in the Zagros Mountains. *Proc Natl Acad Sci U S A* **118** (2021).
39. Peng, W. et al. Analysis of world-scale mitochondrial DNA reveals the origin and migration route of East Asia Goats. *Front. Genet.* **13**, 796979 (2022).
40. Epstein, H. *The Origin of Domestic Animals of Africa*. (Africana Publ. Corp, 1971).
41. Berihulay, H. et al. Genetic diversity and population structure in multiple Chinese goat populations using a SNP panel. *Anim. Genet.* **50**, 242–249 (2019).
42. Purcell, S. et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
43. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
44. Bryant, D. & Moulton, V. Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Mol. Biol. Evol.* **21**, 255–265 (2004).
45. Huson, D. H. & Bryant, D. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23**, 254–267 (2006).
46. Tipping, M. E. & Bishop, C. M. Probabilistic principal component analysis. *J. R. Statist. Soc. B* **61**, 611–622 (1999).
47. Stacklies, W., Redestig, H., Scholz, M., Walther, D. & Selbig, J. pcaMethods–a bioconductor package providing PCA methods for incomplete data. *Bioinformatics* **23**, 1164–1167 (2007).
48. Pickrell, J. K. & Pritchard, J. K. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* **8**, e1002967 (2012).
49. Fitak, R. R. OptM: Estimating the optimal number of migration edges on population trees using Treemix. *Biol Methods Protoc* **6**, bpab017 (2021).
50. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).

Acknowledgements

This work was supported in part by JSPS KAKENHI Grant Numbers 25304038 and 17H04643, 19K22367 and 21KK0122. I wish to thank The Society for Researches on Native Livestock for helping field research work. The authors would like to explain special thanks the late Dr. K. Nozawa for providing long-standing observations of goat morphological distribution in East Asia, related to Fig. S1.

Author contributions

H.M., T.Y. and J.A.L conceived and designed the experiments and supervised the project. R.M., K.I., Y.N., R.T., M.M., A., E.K., F.K. and S.S. performed the genetic analyses. J.W., R.M., T.Y., A.A. and H.M. performed the phylogenetic analyses. K.N., Y.T., M.K.S., M.O.F., J.S.M., M.B., P.K., T.D., M.I.A.D. and S.R.A.B. conducted the field investigation and collected the samples. T.Y., J.A.L. and H.M. wrote the manuscript. All of the authors discussed the results and contributed to the final manuscript. Lenstra, JA played a fundamental role in this paper, including defining the research direction, structuring the paper, and developing the RRAA method.

Declarations

Competing interests

The authors declare no competing interests.

Ethical approval

This study is reported in accordance with ARRIVE guidelines (https:// arriv eguid elines. org). All experiments were carried out according to the Kobe University Animal Experimentation Regulations, and all protocols were approved by the Institutional Animal Care and Use Committee of Kobe University and by Association for the Promotion of Research Integrity (Tokyo, Japan) (Approval Number: AP0000436777). All blood samples collections were approved by animal owners with signed informed consent.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-29204-x.

Correspondence and requests for materials should be addressed to H.M.

Reprints and permissions information is available at www.nature.com/reprints.