Layers in the sand: The genetic imprint of migration, culture, and Indus craft in the Thar desert

Rishabh Jain[a*], Sachin M. Rathod[a*], Pankaj Jha[e], Deepika Jangir[a] , Mohammed Faruq[f], Ganga Nath Jha[d], Mitali Mukerji[a,b,c,*]

[a] Department of Bioscience and Bioengineering, Indian Institute of Technology Jodhpur, Rajasthan, India

[b] School of Artificial Intelligence and Data Science, IIT Jodhpur, Rajasthan, India

[c]Centre for Emerging Technologies for sustainable development, IIT Jodhpur, Rajasthan, India

[d] Department of Anthropology, Vinoba Bhave University, Hazaribagh, Jharkhand, India

[e] Amrita Research Center, Amrita Vishwa Vidyapeetham, Faridabad,India

[f] CSIR Institute of Genomics and Integrative Biology, New Delhi, India.

[*]Equal contributions

Email: mitali@iitj.ac.in

Abstract

The Thar Desert of northwest India, despite its harsh ecology, has sustained settlement of ancient crafts and pastoral communities. Their persistence provides a unique opportunity to study how migration, ecology, and culture have shaped genetic diversity. We analyzed genome-wide SNP data from 176 individuals across eight occupational communities in four districts, alongside global and Indian populations and diverse ancient genomes. Population history, including ancestral migration routes, population structure, demography, admixture, and founder effects, was elucidated using diverse population genetic statistics methods. Thar groups occupy an intermediate position on the Indian north–south cline. Pastoralists and artisans (woodcarvers, Persian gold embosser) clustered with West Eurasian lineages, while potters and performers aligned with southern clines. Uniparental data confirmed heterogeneous Indian lineages. Signals of gene–culture was evident: the lactase persistence allele was higher in pastoralists but lower in gold-embossers despite shared ancestry. Noteworthy, the populations retains a high frequency of the European allele of SLC24A5 associated with lighter skin pigmentation despite the desert environment. Demographic analyses indicate admixture 60–80 generations ago and strong founder effects in certain groups, particularly tie-and-dye and Persian migrant artisans ~500–600 years ago. Ancient DNA comparisons confirmed continuity with

Indus-periphery and historical South Asian populations. The genetic landscape of the Thar is a palimpsest shaped by successive layers of settlement, migration, and cultural continuity. By establishing the first genomic baseline of its craft and pastoral communities, this study reveals how ecology and endogamy intersect with population history to generate a unique genetic landscape. These findings provide essential context for studying genetic risk, adaptation, and human resilience in extreme environments.

Keywords: Thar Desert, Genetic Diversity, Indus Valley, Gene-culture co-evolution, Artisan, pastoralism, founder

Introduction

Northwest India lies at a historic crossroad of ancient trade and migration routes connecting South Asia, Central Asia, and the Middle East. [1,2] Particularly, the Thar Desert region is characterised by resident populations from diverse ethnic groups with distinct socio-cultural practices, including endogamy, nomadism, and pastoralism. Kin-based occupational continuity is a hallmark of the Thar communities, which are renowned for their vibrant traditional crafts, including pottery, woodcraft, leatherwork, textiles, music, and various dance forms [3–7]. Many of these practices can be traced back to ancient times, particularly the Indus Valley Civilisation (IVC, c. 3300–1300 BCE), which thrived in the adjoining geographical regions encompassing Harappa and Mohenjo-daro in the West, Kalibanga in the North, and Dholavira and Lothal in the South of the Thar desert [8–10,11,12] . Recently, Harappan remains have been found in the Thar Desert, pointing to a broader geographical expansion of the civilisation [13].

The Thar Desert today is acknowledged as the world's most densely populated desert. Evident from the continuity of craft, farming, and pastoralism in extant populations, the migrants en route seem to have settled in the Thar despite extremely harsh and resource-scarce conditions [14,15]. The earliest evidence of farming, predominantly of wheat and barley, and early crafts like pottery could be traced to the excavations from Mehrgarh, the Neolithic sites (c. 7000 BCE) of South Asia [16] that share boundaries with the northwestern regions of Thar. These practices flourished during the IVC from 3300 to 1300 BCE, but declined due to climatic changes around 1900 BCE[17]. Populations subsequently migrated eastwards and settled in the Gangetic Plains, Gujarat, and further southward coastal regions[18,19,20–23 14,15,24].

A few genetic studies suggest that modern-day South Asian populations are primarily a mixture of two major ancestral components: Ancestral North Indian (ANI) and Ancestral South Indian (ASI). [19,25,26] However, the ANI-ASI cline does not fully explain the subcontinental genetic diversity. This is evident from earlier studies of the Indian Genome Variation Consortium (IGV) that provided the first comprehensive genetic landscape encompassing extensive linguistic, geographic, ethnic, and socio-cultural boundaries within India and the work of many others[27,28].

The Thar Desert's rich tapestry, woven from its ancient past and vibrant present, reflects a composite of ancestries shaped by early settlements, successive waves of migration, and diverse cultural and craft practices preserved within today's distinct endogamous communities. The Marwari dialect spoken across Rajasthan also reflects the linguistic spread of the Indo-European linguistic family[29,30]. Mass migrations of people in the last century could also have contributed significantly to demographic and social impacts on regions in the western part of Rajasthan. In addition, the harsh environment of the Thar Desert, including high temperatures, aridity, frequent sandstorms, and scarcity of resources[14,31,32], may have together shaped human adaptation and population structure. Craft communities in the Thar, with their kinship-based structures, distinctive cultural practices, and histories of migration from diverse regions, may have contributed to the unique genetic architecture of this desert population. This study aims to elucidate the genetic diversity of Thar populations, with a particular focus on craft communities, and to contextualize their genetic profiles. We report genetic insights into the peopling of the Thar from genome-wide data on 176 individuals spanning eight distinct occupational communities in four underexplored districts. By establishing a genomic baseline from the Thar, this study offers potential for risk stratification in native populations while augmenting efforts to uncover the genetic basis of adaptation to extreme desert environments.

Results

Thar desert populations along the global genetic cline

We studied the genetic landscape across the Thar Desert in northwestern India, within the broader context of South Asian and Eurasian groups. The study populations were carefully selected based on their traditional crafts and cultural practices that are maintained within distinct endogamous, kinship-based groups from rural, urban and seminomadic settings (Table S1). This included the Thar Urban leather artisans (TH-IJE-1-U), tie-dye artisans (TH-IJE-2-U) and the Migrant gold embossing

artists (TH-IJE-MU); the Pastoralist from Rural region TH-IJE-1-PR; the Semi-Nomad/Performing Artist (TH-IJE-1-SN); Potters (TH-IJE-2-R), Embroiderers and Weavers (TH-IJE-3-R) and Wood carvers/ Carpenters (TH-IJE-4-R). This approach not only provided authenticity but also allowed us to directly assess the correspondence between genetic structure and craft as well as culture-based genetic affiliations and origins.

Principal component analysis (PCA) with a variance of 3.35 % PC1 and 1.27 % PC2 reveals a genetic gradient extending from West Eurasia through South Asia to East Asia. The Thar craft populations occupy an intermediate position between groups along the western frontier and certain Indo-European–speaking populations from northwestern India, forming part of a north–south genetic cline (Figures 1, 3). This pattern of affinity is further supported by F_{st} values (Table S3, Figure 2), which indicate close genetic relationships between the Thar populations, western frontier groups, and populations from northern and western India. Nevertheless, some Thar groups deviate from this continuum, suggesting genetic differentiation within the region.

Consistent results were observed in both ADMIXTURE (Figures 1, 5a, b) and F_{st} analyses (Figure 2, Table S3). In particular, the pastoralist (TH-IJE-1-PR), migrant artists (TH-IJE-1-MU), and wooden craft communities (TH-IJE-4-R) display closer genetic affinities with Pathan and Sindhi populations, as well as Indo-European–speaking groups from the Gangetic Plains and Northwestern India (Figures 2 and 5a). By contrast, the potter community (TH-IJE-2-R) and the seminomadic group traditional folk dancers (TH-IJE-1-SN) are distantly positioned at the lower end of the north–south genetic cline.

ADMIXTURE analysis (based on the lowest CV error value and log likelihood), at K=9 (Figure 5 and Table S4), reveals further that the genetic structure of Thar populations reflects variable contributions from South Asian, Middle Eastern, and European ancestries. The northwestern populations share major ancestry components with several South Asian groups (colours- yellow and green in Figures 5a and 5b). The green component is highest in the Kalash population and is widely distributed across Central Asian, Middle Eastern, and European groups. Both TH-IJE-1-MU and TH-IJE-4-R harbor comparable levels of the European component (brown), consistent with their proximity to Indo-European–speaking populations of the Gangetic Plains and the northwestern frontier. This corroborates the observations from the admixture analysis (Figure 1d). TH-IJE-1-SN shows a higher proportion of Ancestral South Indian (ASI)-related ancestry, similar to the isolated Dravidian-speaking population DRSIP1, whereas the Kalash population displays the lowest levels of ASI but the highest levels of Ancestral North Indian

(ANI)-related ancestry. These observations are consistent with historical interactions along the northwestern frontier.

The population relationships and demographic events reconstructed using f_3 statistics that provide formal tests for admixture and shared genetic drift among populations reveal an interesting pattern. Outgroup f3 statistics for genetic drift with Yoruba as the outgroup reveal that the Thar subgroups—TH-IJE-4-R, TH-IJE-1-PR, and TH-IJE-1-MU, with some other Indo-European-speaking populations, exhibit elevated levels of shared genetic drift with West Eurasian populations. The strongest signals are observed with the Kalash, Pathan, Lezgin, Chechen, Belarusian, Lithuanian, and Scottish Orkney groups (z-score > 100) (Table S5, Figure 10), suggesting a clear affinity between these Thar groups, West Eurasians, and populations located along the northwestern frontier. In contrast, other subgroups like TH-IJE-1-SN and TH-IJE-2-R showed stronger levels of shared drift with Indian populations from the Gangetic plain Indo-European cline and Dravidian-speaking South Indian caste groups. Notably, TH-IJE-1-SN follow a drift-sharing pattern resembling that of Gujarati populations. These findings from the outgroup f3 analysis are further corroborated by D-statistics using qpDstat (Tables S6 and S8). Specifically, the Pathan and Kalash populations exhibit significantly higher levels of gene flow with the Thar individuals compared to most other North Indian groups. Contrastingly, TH-IJE-1-SN does not display significant allele sharing with West Eurasian populations when European and Pakistani populations, such as French and Georgians, Iranians, Pathans, Balochi and Kalash, were taken as reference populations (Table S6, Figure 11).

We also computed f3-statistics to evaluate the extent of formal admixture of Thar populations along with groups from the Gangetic Plains and Rajasthan with Ancestral North Indian and Ancestral South Indian lineages. Ancient and modern West Eurasian populations were used as proxies for the ANI component and DR-S-IP1 for the ASI component. We found that populations such as TH-IJE-1-MU, TH-IJE-1-U, TH-IJE-2-U, and TH-IJE-1-SN did not yield significantly negative f3-values, suggesting a stronger affinity toward a single ancestral source and limited evidence of direct ANI–ASI admixture. By contrast, groups including TH-IJE-4-R, IEWLP3, TH-IJE-3-R, and TH-IJE-1-PR from the Thar region exhibit significantly negative f3-values, consistent with admixture signals observed among Gangetic Plains populations and northwestern Indo-European–speaking groups (Table S7). These results suggest that while certain Thar populations retain genetic continuity with one predominant ancestry, others share admixture patterns characteristic of broader ANI–ASI interactions that have shaped the genetic landscape of North India.

To further estimate the proportion of the Ancestral North Indian (ANI) component, we applied the f4-ratio method of ADMIXTOOLS. Higher proportions of ANI ancestry were observed in populations such as the Kalash, Pathan, and northwestern groups IENLPX3 (Ror), IENLPX4 (Gujjar), and IENLPX5 (Kambhoj). Among the Thar groups, TH-IJE-4-R, TH-IJE-1-PR, and TH-IJE-1-MU also exhibit elevated ANI ancestry compared to most other South Asian populations. In contrast, TH-IJE-1-SN displayed a notably reduced proportion of shared ANI ancestry (Table S9).

Consistent patterns are observed in the maximum likelihood tree of TREEMIX v.1.13, which estimates population-specific drift and divergence. Thar populations cluster within the clade of Indian Indo-European caste groups. Compared to other Thar populations, TH-IJE-1-MU and TH-IJE-2-U exhibit higher levels of population-specific drift, as reflected by longer branch lengths relative to most South Asian populations, except Kalash. Furthermore, several northwestern subcontinental populations were positioned within the same clade as TH-IJE-1-MU, TH-IJE-1-PR, and TH-IJE-4-R. In contrast, TH-IJE-1-SN consistently remains more distantly related to this clade (Figure 6).

Analysis of identity-by-descent (IBD) using RefinedIBD patterns shows a reduction in sharing in a clinal gradient, declining westward from groups near the northwestern frontier and northward in the Indo-European groups of North India. This decline was even more pronounced with Dravidian and Austroasiatic groups from southern and eastern India, respectively. Notably, the migrant artist group from the western frontier (TH-IJE-1-MU) exhibited elevated IBD sharing with European populations, a pattern also observed in the Pathan population. Similarly, TH-IJE-1-PR and TH-IJE-4-R show increased IBD sharing with neighboring northwestern frontier populations such as Balochi and Sindhi, as well as with Indo-European groups from the northwest and Gangetic Plains, and to a lesser extent, with Dravidian populations (Figure 7). Within-population IBD analyses normalized by population size[33,34], we observe that individuals of the Thar population, such as TH-IJE-1-MU and TH-IJE-2-U, exhibit the highest IBD scores, comparable to Kalash and DRSIP1 amongst Indian populations (Table S10). This finding corroborates earlier accounts of increased inbreeding and strong genetic drift. Elevated levels of homozygosity were also observed in most artisan communities of the Thar Desert. Both the mean number and mean length of runs of homozygosity (ROH) segments >1 Mb were significantly higher compared to other Indian groups. These findings indicate reduced genetic diversity and recent population isolation (Figure 8).

To investigate the demographic history of the Thar populations, we applied complementary methods focusing on founder events and admixture timing. Using the ASCEND (Allele Sharing Correlation for Estimating Demography) algorithm, we inferred founder intensities ranging from 2.0% to 4.6% across Thar subgroups, indicating varying degrees of population bottlenecks and long-term endogamy (Figure 13a). Application of the HapNe-LD approach further identified recent bottleneck events in groups such as TH-IJE-2U, TH-IJE-1-MU, and TH-IJE-2R, with estimated timings within the last 400–500 years (Figure 13b). These findings suggest that certain Thar subgroups experienced relatively recent reductions in effective population size, likely linked to social isolation and strict endogamy. Furthermore, for the timing of West Eurasian admixture, we applied weighted linkage disequilibrium (LD)-based inference using ALDER. Admixture dates were successfully estimated by modelling the Thar populations as mixtures of two reference sources: modern West Eurasian groups and South Asian groups. Following earlier studies, Austroasiatic-speaking groups were used as proxies for AASI ancestry to represent the South Asian source. Under this, we inferred admixture events in Thar artisan populations, with most groups showing evidence of Steppe-related West Eurasian ancestry dating to approximately 60–80 generations before present. Notably, TH-IJE-2U exhibits a more recent admixture event, dated to ~30 generations before present (Table S11).

Complementary insights were obtained from ChromoPainter and finestructure analysis based on haplotype chunk counts. Individuals from TH-IJE-4-R, TH-IJE-1-PR, and TH-IJE-1-MU share a greater mean number of haplotype chunks with European and Caucasian populations compared to other Thar artisan groups. By contrast, TH-IJE-1-SN exhibits the lowest levels of haplotype sharing with these groups (Figure 12). Collectively, these results reveal a gradient of westward genetic affinity, with most Thar populations showing stronger haplotype sharing with populations along the northwestern frontier than with other South Asian groups, including Dravidian- and Austroasiatic-speaking communities.

Determination of ancestral proportions in Thar populations in a global context

Using PCA, outgroup-f_3, and D-statistics analyses, we found that the Thar populations exhibit stronger genetic affinity with EMBA groups than with later Steppe MLBA groups. Within the Thar, individuals from TH-IJE-4-R, TH-IJE-1-PR, TH-IJE-1-MU, and TH-IJE-2-U cluster more closely with proximal sources such as South_Asia-H, Indus Periphery, SPGT, Namazga Chalcolithic, and BMAC, distinguishing them from other Thar subgroups with higher indigenous components that cluster farther along the South Asian cline (Figures 4).

Outgroup-f_3 analyses in the form of f(test, ancient pop; Yoruba) showed that these Thar groups shared elevated genetic drift with ancient reference populations, including EEHG, CHG, Neolithic Iranians (Ganj Dareh_N), and Anatolia Neolithic, compared to other South Asian groups (Table S12, Figure 9).

We applied D-statistics in several configurations to explore allele sharing with ancient populations. Using D(test, Yoruba, AncientPop1, AncientPop2) and D(test, TH-IJE-[1PR/4R], AncientPop, Yoruba), we assessed affinities between South Asian groups, ancient sources, and proxies for ASI ancestry. Across most South Asian populations, the D-statistics produce broadly similar results. Western Steppe_EMBA show consistently stronger affinity compared to Central and Western Steppe_MLBA and Namazga_CA, and also exhibits stronger proximal affinities with SPGT, South_Asia-H, and Indus Diaspora populations (Table S12, Figure 9). The Thar groups TH-IJE-1-PR, TH-IJE-4-R, and TH-IJE-1-MU show allele sharing and affinities with Indus_Periphery at levels comparable to other northern and western Indian populations (Tables S13, S14). Among them, TH-IJE-1-PR and TH-IJE-4-R demonstrate enhanced affinity with Chalcolithic and Neolithic sites such as Namazga_CA and Ganj_Dareh_N, as well as other historical groups like SPGT, BMAC, and South_Asia-H, when compared with other South Asian populations (Table S12). Together with TH-IJE-1-PR and TH-IJE-4-R, TH-IJE-1-MU show greater affinity with EEHG than with present-day Iranians, suggesting contributions from a Central Asian or Steppe-related West Eurasian source rather than recent Iranian admixture (Table S13). Both TH-IJE-1-MU and TH-IJE-1-PR also display significant allele sharing with Turkmenistan_IA compared to SPGT, highlighting Central Asian links in their genetic profiles (Table S14). Notably, when comparing Indus and SPGT, TH-IJE-1-PR exhibit strong affinity with BMAC, whereas other South Asian groups, including Dravidian and Austroasiatic populations, appear closer to Indus_Periphery and SPGT. In addition, we detected a distinct affinity with Indus Diaspora groups when compared to BMAC, Namazga_CA, and SPGT, particularly in northwest Indian groups and Thar artisan populations. Within the Thar samples, TH-IJE-1-PR, TH-IJE-1-MU, and TH-IJE-4-R show comparable levels of allele sharing, while central Austroasiatic and Dravidian populations appeared closer to the Indus Periphery (Tables S13, S14). Individuals TH-IJE-1-PR and TH-IJE-4-R were among the closest South Asians to CHG and also carry stronger affinities to EEHG, Ganj_Dareh_N, Anatolia_N, and Steppe-related sources than most South Asians, placing them nearer to groups such as Pathan and Kalash. TH-IJE-1-PR, in particular, exhibit significantly higher affinity with BMAC than with South_Asia-H or SPGT, whereas TH-IJE-4-R and TH-IJE-1-MU show closer ties to South_Asia-H and SPGT (Tables S13 S14).

Admixture modelling using f_4 ratios by a model of f4(Mbuti, EEHG; Test, non-African Andamases) / f4(Mbuti, EEHG; Central_Steppe_MLBA, non-African Andamases) further supported these findings. We detected substantial Central Steppe MLBA ancestry (proxy for ANI-related ancestry when using Namazga CA / Indus periphery and AHG for a better fit model) in the Thar, ranging from 48.7 ± 1.9% to 34.6 ± 2.2%. This proportion was lowest in TH-IJE-1-SN and TH-IJE-2-R but higher in Pathan, Kalash, and in Thar groups such as TH-IJE-1-PR and TH-IJE-4-R. (Table S15) Using qpAdm[35,36] three-way admixture models (Western/Central Steppe MLBA, Indus Periphery cline, and AHG/non-African Andamases), we found that many populations fit the qpAdm model ($p > 0.01$), though some failed, suggesting that this framework may not fully capture the complexities of ancestries along the north–south cline. These estimated results show a consistent proportion, as mentioned in previously studied populations. Notably, TH-IJE-1-PR seems to have a predominantly ancient Iranian farmer-related proportion (~82%) with minimal AHG contribution (~1–5%), closely resembling the Kalash profile. By contrast, TH-IJE-1-SN shares genetic proportions that are similar to the Dravidian-speaking caste group (Table S16). Among the Thar individuals, TH-IJE-1-MU consistently reveal the highest proportions of Steppe ancestry, followed by the other Thar groups, as well as northern/western Indian populations such as IENLPX4, IENLP5, and IEWLP4. The observed steppe component was consistent compared to other South Asian populations and similar to the proportion comparable to Pathan and Kalash, with higher shared Iranian source genetic ancestry in TH-IJE-4R and TH-IJE1-PR. (Table S16).

Broadly, proportions from genetic sources in our artisan populations are within the observed range in other South Asian populations. When using the distal model, We found that our model fit showed $p > 0.05$ for most of the populations. Across both distal and proximal models, Bronze Age Steppe ancestry was found to be qualitatively higher in TH-IJE-1-MU (Table S17). Both distal and proximal qpAdm models reveal patterns consistent with previous reconstructions of South Asian population history and episodes of population movement from the Steppe and Central Asia into South Asia during the Bronze and Iron Ages, while also underscoring the unique position of Thar subgroups at the intersection of cultural and demographic exchanges across the Indo-Iranian frontier.

Gene flows mirror the genetic route of migrations from Europe to the Thar.

To assess whether present-day Thar populations carry genetic signatures that corroborate the insights from ancient migration and adaptation histories, we examined the distribution of functional alleles

underlying traits such as lactase persistence and skin pigmentation. We investigated the distribution of the lactase persistence-associated rs4988235 T allele (−13910*T) of a LCT gene, that enables the digestion of milk into adulthood. Our analysis revealed marked heterogeneity in its frequency among populations of the Thar. Consistent with broader South Asian patterns, the T allele generally declines in frequency from North to South and from West to East. However, within the Thar populations, we observed distinct local enrichments. Compared to the previously studied Ror (IENLPX3) population[37,38], which shows the highest frequency (71.4%), the groups from our study exhibit extensive heterogeneity, with the pastoralist populations having the highest frequency TH-IJE-1-PR (67.9%), followed by TH-IJE-4-R (40.5%), TH-IJE-3-R (29.0%), and TH-IJE-1-U (26.8%). Surprisingly, the least in TH-IJE-1-MU (Figure 14 a,b). These frequencies resemble those observed in European and West Eurasian pastoralist groups, where the allele is under strong positive selection due to reliance on dairy consumption[38–40].

In addition, we also examined the rs1426654 variant in the SLC24A5 gene, a major contributor to lighter skin pigmentation in Eurasian populations[41,42]. The derived A allele, which is nearly fixed in Europeans, also displays considerable frequency variation across the Thar region. It is most frequent in the IENLPX3, IENLPX5, IENLP5, TH-IJE-1-U and TH-IJE-1-PR groups and is also elevated in several other North Indian populations. Interestingly, some groups of these also show high frequencies of the lactase persistence allele, although the two traits vary independently, suggesting parallel but distinct histories of adaptation for dairy consumption and pigmentation. Conversely, groups such as TH-IJE-1-SN and TH-IJE-2-R exhibit intermediate A allele frequencies, similar to many Indo-European and tribal-affiliated populations (Figure 14 b).

Uniparental Genetic Analysis.

The uniparental markers of the Thar populations reveal a tapestry of ancestries, reflecting both South Asian roots and multiple waves of external influence[43–49].

On the maternal side, mitochondrial analysis revealed 55 distinct sub-haplogroups encompassing 5 of the 12 major global haplogroups. Over half of the individuals (51.1%) belonged to haplogroup M, the most widespread and ancient lineage of South Asia, while nearly one-third (31.8%) carried haplogroup U. The latter links South Asia to West Eurasian maternal ancestry. Smaller fractions of haplogroup H (5.7%) and the presence of lineages such as W, HV, T, J, K, and I highlight mixtures with West Eurasia, whereas F, Z, and N reflect East Eurasian connections. Remarkably, we also detected haplogroup X

(0.6%), a rare lineage with a wide but patchy distribution across Eurasia. Such diversity suggests that the maternal gene pool of the Thar has been shaped by multiple sources from different regions and times (Table S18).

On the paternal side, Y-chromosomal analysis also demonstrates heterogeneity. Among the Thar dwellers, the most common haplogroups were R1a (22.1%), associated with Indo-European expansions, and L1a (17.5%), more frequent in Dravidian-speaking groups. This corroborates the ANI-ASI continuity. H1a (11.2%) and R2a (13.9%) further underscore the presence of the indigenous population. At the same time, lineages such as J2a, J2b, and G2a point to introgression from Neolithic lineage and later movements from the Middle East and Central Asia, while C1b and Q1a trace connections extending into Central Asia and Siberia (Table S19).

Discussion

Our study shows genetic patterns among present-day craft artisan communities. The crafts of these communities in the northwest region are not just artistic expressions but also integral to their cultural identity and livelihoods, a legacy passed down through generations. We analyzed the broader genomic relationships and ancestry of these craft artisan communities and pastoral groups in the Thar Desert. We also analyzed their linkages with neighboring South Asian populations and global groups. Importantly, we did not classify the assigned populations based on oral history alone, but rather by their traditional occupations through endogamy and customary practices across generations.

Communities and artisan groups in the Thar illustrate historical connections through both craft and lineage that are also recognised by Government of India (https://handicrafts.nic.in/). For instance, TH-IJE-1-MU, the gold-embossing artisans of Bikaner, trace their ancestry to Persian craftsmen invited during the Mughal era, who introduced naqashi manoti techniques and carried a legacy of Central Asian artistry. This Central Asian thread is also indicated for the pastoralist group TH-IJE-1-PR, whose nomadic lifestyle recalls origins in Baluchistan and Afghanistan, later adapting to the Thar's ecology through long-distance seasonal mobility. The wooden craft group (TH-IJE-4-R), associated with the Rathakar, or chariot builders, doors/kawads. Their guild-based traditions in carving, measurement, and geometry echo Vedic building practices and are even mentioned in the Skandapurana. Similarly, the potters (TH-IJE-2-R) sustained lime kilns and pottery traditions engaged in antiquity. Their craft connects to the Namazga and Shahr-i Sokhta cultures and continues in the utilitarian pottery of the Indus Valley, also referenced in the *Yajurveda*. Collectively, these groups exemplify artisanship central to both

everyday life and ritual across South Asian history. Other Thar communities preserved equally distinct traditions. TH-IJE-3-R combines agriculture, cattle-herding, embroidery, and weaving, with some also engaged in skinning and tanning. TH-IJE-1-U, descended from shoemakers, specialised in leatherwork—producing swords, saddles, belts, footwear, and other goods for warrior clans. TH-IJE-1-SN, traditionally snake charmers and performers known as Gypsies of Rajasthan's Deserts, lived a nomadic life and sustained by trading snake venom, music, and dance, a heritage later recognised by UNESCO. Finally, TH-IJE-2-U, cloth printers and dyers renowned for tie-and-dye textiles. They adopted this profession during the Mughal period, with some members later undergoing religious conversions[50–56,57,58,3,9,59,52,60–63,52,60–62,64]. The settlement of these communities in the Thar Desert can be understood in relation to both environment and livelihood. While the region is harsh and resources are limited, its dry climate favored crafts that required low humidity, such as tie-dye textiles and leatherwork. Seasonal rainfall, pastoral migration, and animal husbandry provided an additional base for survival and exchange. Their craft traditions were maintained over centuries, with some techniques showing similarities to those known from the Indus Valley Civilization. At the same time, designs that bear influence from Eurasia and Central Asia reflect episodes of trans-regional movement. Royal patronage and local demand further supported the continuation of these practices. Together, these factors helped preserve cultural identity, sustain livelihoods, and tie the communities both to their desert setting and to wider networks of movement and interaction.

Present-day northwest India was historically influenced by successive waves of migration and incursions, including those of the Greeks, Scythians, Parthians, Kushans, and Huns. Taken together, these communities embody the Thar as a corridor of migration, cultural exchange, and ecological adaptation. Their hereditary crafts, anchored in oral traditions, Vedic references, Mughal patronage, and transregional linkages, remain inseparable from their genetic history[62,65–68]. Previous studies on populations from southern and central Rajasthan, including both tribal and caste groups, have reported higher average heterogeneity values based on autosomal STR analyses. These groups showed stronger affinities with Indian populations, particularly North Indians, rather than with populations from the western frontier. Some tribal groups from the northwest were also found to share greater genetic similarity with Dravidian populations. In addition, populations such as Gujjar and Kambhoj, located near ancient Rajasthan sites of civilisation, have been studied for their affinities with both ancient and modern populations. These populations are also included in our study for comparison and reference[37,69–7172,73].

From multiple genetic distance analyses, we found that craft artisan communities aligned with the Indo-European cline of South Asia, and there was substantial heterogeneity within the northwest region, which maintained separate and distinct population groups over a long period. Genetic differentiation was found to be lower in the population from the Gangetic Plain and the western frontier, that is, proximal to the Indus. We found that the Thar artisan population shows affinity with Sindhis; however, the populations like TH-IJE-4R, TH-IJE-1-PR, and TH-IJE-1-MU of the Thar artisans share more affinity with western frontier populations like Pathan, which shows a Central Asian connection. This is reflective of the close geographical proximity of Sindh to the Thar Desert, wherein Sindhis were merchants and traders of goods and crafts. This suggests that population structure experienced mobility along regional routes during the migration of Indo-Europeans toward the Gangetic Plain after the decline of civilization.

TH-IJE-4R, TH-IJE-1-PR, and TH-IJE-1-MU show a higher degree of West Eurasian genetic influence, comparable to or even exceeding that observed in other North Indian populations. This pattern aligns with previous studies[26,37,46,72,74–76] indicating that northwestern Indian groups and upper-caste populations across India share greater genetic similarity with populations from Central and West Asia, as well as parts of Europe. In contrast, within the Thar region, artisan communities such as TH-IJE-1-SN, TH-IJE-2R, TH-IJE-3R, TH-IJE-1U, and TH-IJE-2U demonstrate stronger genetic affinity with local Indian populations and diverse South Asian ethnic groups, while showing relatively lower affinity with West Eurasian populations.

This observation is further supported by the proportion of ANI (Ancestral North Indian), now recognised as ANE (Ancestral North Eurasian)[77], which was consistently higher in TH-IJE-4R, TH-IJE-1-PR, and TH-IJE-1-MU, comparable to levels observed in Kalash and Pathan populations, while being lowest in TH-IJE-1-SN. Given that ANI ancestry is widely shared with Central Asians, Caucasians, Middle Easterners, and parts of Europe, these results underscore substantial variation in West Eurasian genetic influence among artisan communities of the Thar region. This can be seen through Y-chromosome haplogroups, with a dominance of R1a among many craft artisans, reflecting ties to Steppe pastoralism and Indo-European expansions in India. The presence of L1a suggests connections to Neolithic migrations, possibly linked to the spread of Dravidian languages and agriculture from Iran into South Asia[78]. The occurrence of Middle Eastern and Central Asian associated haplogroups such as J2a, J2b, and G2a further reflects broader West and Central Asian linkages. In Continuity, the highest frequencies are observed in South Asian-related M haplogroups, indicating a

strong local or regional maternal lineage. However, haplogroups U7, W, and R2, which are associated with West Eurasian ancestry, are also consistently present in the three Thar populations studied, with a noted higher contribution in the population labelled TH-IJE-2-U. These West Eurasian haplogroups have spread into the northwestern region, suggesting gene flow from West Eurasian populations into this area. The high proportion of West Eurasian lineages in both uniparental and autosomal loci indicates a consistent pattern of genetic ancestry and reasoning, which is likely through migration or trade with West Eurasian groups, alongside a dominant South Asian genetic component.

The observed pattern of heterogeneous population structure and genetic drift in several Thar populations, particularly TH-IJE-1-MU and TH-IJE-2-U, reflects broader founder events documented across South Asia. These patterns may be linked to historical episodes of migration and religious conversion occurring within the last 400–500 years. The observed homozygosity patterns, corroborated by runs of homozygosity (ROH) and effective population size estimates, indicate isolation and genetic drift due to endogamy and geographical constraints. These demographic patterns show consolidation of artisan communities under guild-based production systems that became increasingly rigid during the Mughal period, when craft specialisation was institutionalised within hierarchical social structures.

Earlier genetic studies have shown that the populations of the Indian subcontinent fall along a broad cline shaped by two major ancestral components. Ancestral South Indians (ASI) derive from indigenous South Asian hunter-gatherers and ancient Iranian Neolithic farmers, while Ancestral North Indians (ANI) reflect ancestry from Middle to Late Bronze Age (MLBA) steppe pastoralists connected to Eurasia. Situated along a historical corridor in northwestern India, the Thar Desert populations show varying proportions of MLBA steppe ancestry, with some northwestern groups (NWI) carrying the highest levels reported in India to date[19,25,26]. These findings are supported by multiple lines of evidence, including D-statistics, outgroup f3-statistics, and admixture modeling with qpAdm.

The presence of these ancestral components likely reflects both proximal and distal migration events. Elevated allele sharing in NWI highlights the Thar's role as a genetic crossroads, shaped by gene flow from diverse external sources. Their ancestry shows connections to the Iranian plateau and interactions with steppe groups, interactions with Indus Valley populations, and later diasporas. In addition, further including Mesolithic hunter-gatherer ancestry (EEHG, CHG), Neolithic farmer contributions from Anatolia and Iran, and later Bronze, Copper, and Iron Age influences from the Central Asian steppe and adjacent regions[79,80]. These genetic inputs, particularly in Thar, align with the spread of traits like lactase

persistence (−13910*T allele), introduced via pastoralist migrations from the Middle East or Anatolia around 5,000–3,000 years ago[38,39]. The geographic distribution of lactase persistence across India exhibits a clear northwest-to-southeast declining pattern and heterogeneity in its frequency among populations of the Thar Desert. We found that IENLPX3 and TH-IJE-1PR show the strongest signals in South Asia, likely linked to their traditions of animal husbandry, agro-pastoralism, and dairy consumption. These groups share ancestry from the Anatolia Neolithic (Anatolia N) and Caucasus Hunter-Gatherers (CHG)[81]. In contrast, another pastoralist group from the same lineage, IENLPX4, shows a stronger affinity with Indian-specific ancestry, while the pastoralist group in our study displays greater West Eurasian-related affinity. Specifically, it leans more toward Neolithic Iranian and BMAC lineages (~20 to 25% Anatolia_N-related ancestry), which are geographically and temporally closer to the Indus Valley populations and the Indo-Iranian frontier. A complementary signal of West Eurasian influx is also reflected in the frequency of the SLC24A5 rs1426654 variant (associated with lighter skin pigmentation)[41,42]. This allele occurs at higher levels in most of the Thar population and in some Northwest Indo-European-speaking groups, but its frequency decreases substantially toward the south.

In conclusion, our study positions the Thar Desert as a northwestern corridor in South Asian prehistory that is at a remarkable genetic crossroads where geography, migration, craft traditions, cultural practices, and ecology converge. In this genome-wide study of 176 individuals from distinct craft artisan communities, we show that beyond social isolation, endogamy, and consanguinity, the artistic traditions integral to cultural identity and livelihood also contribute to genetic distinctiveness. The genomes of the artisan and pastoralist groups act as a palimpsest, preserving signals from successive demographic events. Their crafts and communities show higher genetic drift and population isolation, particularly during episodes of social and political change in the medieval, early modern, and colonial periods. Religious practices, occupational shifts driven by need, community separation, and the formation of new social groups through occupational mobility all contributed to their genetic distinctiveness. Endogamy, maintained for socioeconomic and cultural reasons, reinforced group boundaries and shaped their genetic structure in recent centuries. When integrated into broader Indian and global genomic datasets, the roots of these communities trace both to movements along the western corridor and to settlements from the Indian interior, reflected in their clinal genetic patterns. They preserve layered imprints of demographic episodes, from early Neolithic farming expansions to Bronze Age steppe-related movements and later exchanges. Thar Desert populations also show examples of a gene–culture co-evolution differentiated by selective pressures that leave genetic signatures on adaptive traits such as lactase persistence and skin pigmentation. The diversity of uniparental markers, spanning South Asian

lineages alongside inputs from West and Central Eurasia, further underscores the heterogeneity and fine-scale structure of the Thar. This study thus provides the first genomic baseline for the Thar Desert populations. It demonstrates how migration history, cultural identity, and ecological pressures collectively shape genetic structure. Beyond reconstructing population history, this baseline will be valuable for detecting convergent selective signatures in extreme environments and for risk stratification in health studies.

Methods and Materials

The study was carried out on a dataset of 176 individuals from eight different populations from the Thar region and diverse populations from Africa, Europe, Asia, South America, and Oceania. Genotype data were integrated from major published studies for a total of 2812 individuals. Depending on the analysis and availability of genotype data, different subsets were considered for population genomics studies and are further detailed below. The data from Thar was generated in-house after a detailed ethnography study (Table S1).

Ethnogenetic study in Thar

Ethnography of the native population from the Thar Ecoregion

Our study focuses on the craftsmen of the northwest provinces of the Indian subcontinent. This region spans across four districts: Jaisalmer, Jodhpur, Barmer, and Bikaner. The populations native to these regions practice traditional crafts and art forms that include pottery, terracotta, woodcraft and carving, gold embossing and painting, leather craft (mojri), tie and dye, and folk dance. We first reviewed historical records, archaeological data, and previous studies to identify native groups and their settlements[4,6,7,50–52,62,82–84]. We validated the selected population's indigenous status and cultural/linguistic affiliations through multiple data sources. This was followed by pilot surveys to identify the cluster and centre and to gather auxiliary information about the artisan communities. The surveys undertaken encompassed a variety of methods aimed at gathering in-depth insights into the craft-making communities. Observations were conducted to understand the craft-making processes, the types of crafts produced, the availability and types of markets, marketing strategies, and the social and economic status of these communities. Focus Group Discussions (FGDs) were organised, with eight sessions planned—one for each craft community—featuring senior representatives who shared their traditional histories. Similarly, Key Informant Interviews (KIIs) were carried out, involving both senior and

younger generations from each community to explore the evolution and changes in their traditional art forms. Additionally, eight case studies, one per community, focused on documenting the traditional histories as narrated by senior artisans. Lastly, a Genealogical Table (GT) of Artisan Families was developed to map key kinship links determined by alliance and descent, providing a detailed understanding of family relationships within the craft communities. These were captured using a questionnaire that was designed to assess the ethnic, socio-economic, and occupational background of the subjects, as well as capture their oral history and lifestyle. Detailed documentation also included digital capture of the design and patterns of diverse crafts (Table S1).

Briefly, clusters of families of different crafts that include textile and embroidery, tie/dye printing, leather, wood carving, and pottery, as well as natives who practice pastoralism or have semi-nomadic lifestyles, with distinct art forms included.

Ethics statement

All interactions with participants were conducted in accordance with the standard ICMR human ethics guidelines[85], and informed consent was secured from all subjects who participated in the project. Population identifiers as well as sample names were anonymized to safeguard their privacy and prevent stigmatization risk of isolated and endogamous communities.Identification of populations and sampling

Based on the ethnographic study, we identified eight native communities and one migrant community in the Thar Desert region. These communities are contrasting and diverse in terms of occupation, art forms, and sociocultural practices and are together referred to as Thar populations. A convention has evolved to label the population according to their locality and community practices in the samples. The population names were labelled as TH-IJE-X-(R/NR/MU/SN/U/PR), where TH-IJE indicates the study ID from the IIT Jodhpur ethnography study followed by the serial number of sampling population (X), and from Rural (R), Nomadic (N), MU (migratory urban), SN (semi-nomadic), U (urban), and Pastoralist (PR) (Figure 1) (Table S1,S2).

Genotyping and Integration of Thar with global datasets

A total of 185 individuals were genotyped using the commercially available Infinium Global Screening Array 24+ v.3 Kit (abbreviated GSA) + Multi‑Disease markers in the Illumina iScan System (110 V/220 V) according to the manufacturer's protocol using extracted DNA from peripheral blood of selected

population samples[86]. Genome Studio v2 (Illumina) was used for pre-QC of genotyped samples with the standard protocol for gentrain score, call rate, and cluster separation score[87]. After sample QC, further analysis was carried out on 176 samples. To increase the efficiency in genotyping calls, cluster files generated with Indian reference population data sets[88] were used. To compare our study populations with other global as well as Indian populations.and other previously studied populations, representing major continental populations[27,72,89–95,96]. We analysed the Thar genotype data (described above) together with published data on Indian and other continental populations generated on different array platforms. To harmonise the nomenclature with respect to geographic and linguistic affiliations of Indian populations we followed a convention of renaming the populations based on the convention followed in Indian Genome Variation Consortium[92]. The details of the populations, as well as their inclusion in analysis, intersecting sets of genotypes from different studies were included, are indicated in the (Table S2).

We also considered panel of variations from ancient sites reported in previous studies[19,35,81,97–99] identify multiple layers of ancient West Eurasian ancestry in South Asia, including Mesolithic Eastern European Hunter-Gatherers (EEHG) and Caucasus Hunter-Gatherers (CHG), Neolithic Iranian farmers (~9,000–7,000 years ago) from the Zagros, and Neolithic Anatolians. Steppe Early–Middle Bronze Age (EMBA) groups (Yamnaya-related) arose from EEHG and CHG. Additional ancestry sources include Neolithic/Bronze Age Iran-Turan, Siberian hunter-gatherers, Botai, Namazga, Bactria–Margiana Archaeological Complex (BMAC), Indus Periphery, Historical South Asia (SouthAsia_H), SPGT (likely Swat Protohistoric Grave Type) and India_Roopkund.

The genotyped data were merged using EIGENSOFT tools: CONVERTF and MERGEIT[100]. After applying the QC filter, viz., removal of individuals with low genotyping rates, genotyping error cutoff at 0.10, MAF 0.01, individual missingness per individual data 0.15, and HWE 1e-06, we obtained a total of 77561 variants, and 2812 people passed filters and QC (Dataset 1). Additionally, we pruned the data based on linkage disequilibrium (LD) using PLINK [101] using --indep-pairwise parameter with a window size of 200 SNPs, a step size of 50 SNPs, and an r^2 threshold of 0.4. After LD pruning, 71414 variants and 2812 people pass filters and QC. The pruned dataset (Dataset 2) was used for PCA and admixture analysis. We also created a merged subset Dataset 3 with 1240K panel[102] populations with 85393 variants for further analysis to understand the relatedness of Thar with its proximal and distal ancient sources from Eurasia and South Asia (Table S2).

Genome-wide autosomal analysis

We conducted genome-wide autosomal analyses to investigate the genetic structure of Thar populations in relation to neighboring and reference groups**.**

Principal Component Analysis (PCA):

PCA was performed at the individual level after pruning SNPs using SMARTPCA from the EIGENSOFT package[100], with default parameters (numoutlieriter: 0, lsqproject: Yes). Results were visualized in **R** using the *ggplot2* package (http://www.r-project.org/)[103].

Admixture analysis:

We carried out ADMIXTURE (Alexander et al., 2009) analysis after pruning SNPs in strong linkage disequilibrium with Plink v1.9[101]. ADMIXTURE was run with 5-fold cross-validation, varying the number of ancestral populations from K = 2 to K = 18, each with 25 replicates using different random seeds. The optimal K value was chosen based on the lowest cross-validation error and highest log-likelihood.

Fst analysis:

Pairwise Fst values were estimated between Thar populations and neighbouring groups using SMARTPCA in EIGENSOFT[100], with the parameters fstonly: YES and phylipoutname enabled. Block jackknife resampling was used to calculate standard errors.

Treemix:

In order to understand the relationships between populations in terms of common ancestry, genetic drift, and gene flow, we estimated a maximum-likelihood tree using TreeMix v1.13[104] based on genome-wide allele frequency data from Indian populations, where OANIP (Andamanese) was used as an outgroup to root the output tree based on the ML graph visualizing population splits and admixture events without any migration edges in order to provide a simple graph tree.

Pairwise outgroup f3 statistics and f4 statistics:

As ADMIXTURE is not a formal test for gene flow between populations, we conducted three-population admixture tests (f3 statistics)[36] to assess shared genetic drift and evolutionary history. In pairwise Outgroup f3, we considered Thar and its most adjoining region population (Thar/Pakistan/IEWLPX1 Pop2, Outgroup) test, where Pop2 are all the other neighbouring populations,

such as South Asians, Europeans, Middle Easterners and Caucasians from Dataset 3 and Pop2 as outgroup African populations (Yoruba). We inferred the extent of admixture and gene flow with neighbouring populations using the threshold significant value of the Z-score -2 to +2 range with jackknife standard errors for the test population.

We also employed f4 statistics using the qpdstat program in Admixtool 7.0.2[36,25] with f4mode=YES parameter and as f4 with many combinations to assess gene flow and investigate derived allele sharing between Thar and other populations. The proportion of ancestral North India was estimated from the f4 ratio estimation using AdmixTools 7.0.2[36,25], which analyses the correlation of allele frequency patterns to determine the proportion of Ancestral North Indian (ANI) ancestry in different Indian groups. The ANI ancestry proportion was calculated based on the model outlined by Moorjani et al[25]. We included Yoruba (YRI) as an outgroup and Andamanese (OANIP) as indigenous representatives from the Andaman Islands, along with two West Eurasian populations. To assess the uncertainty associated with ANI ancestry proportions, we benchmarked our studies with two distinct geographic extremes: the Abkhazian and Basque populations reported earlier.

Admixture coefficient modelling by qpAdm:

We used qpWave and qpAdm,[36] using the admixr[105] package from R, to estimate ancestry proportions in South Asians by modelling them as a mixture of 'reference' populations while utilising shared genetic drift with a set of 'outgroup' populations. In this analysis, the OANIP* group was taken as a proxy for the South Asian component, representing Ancient Hunter-Gatherers (AHG), since they lack recent West Eurasian ancestry. Using the 1240K-merged dataset, the test population was modeled with three reference groups: OANIPa* (n = 27; non-African Andamanese islanders), Mesolithic and Late Bronze Age Steppe-related groups (Central/Western Steppe-MLBA), and Indus Periphery/Diaspora (Indus Periphery) as the proximal model, while Namazga_CA was included as the distal model. To assess mixture proportions, we used a diverse set of modern and ancient populations; for distal, we used outgroups such as Ethiopia_4500BP. SG, ANE, PPN, EEHG, WEHG, Natufian, ESHG, and Dai. For proximal, we used Ethiopia_4500BP. SG, Anatolia_N, PPN, EEHG, WEHG, Sarazm_EN, ESHG, and Dai, including the fitted model as right outgroup populations, followed using[19,106] allowed us to explore the genetic structure of Thar natives and their ancestral affinities with different prehistoric populations.

Dating admixture event

We computed weighted linkage disequilibrium (LD) statistics using ALDER[107] to estimate the timing of admixture events in the Thar population. The method relies on the exponential decay of LD generated by admixture, with the decay rate providing an estimate of the time since admixture. To improve accuracy, the algorithm measures SNP correlations in the admixed population and applies weights based on allele frequency differences between reference populations. This approach enhances admixture-related LD signals while reducing the influence of background LD, thereby minimizing potential biases. For our analysis, we used West Eurasian and AASI populations as references. Standard errors were calculated using chromosome-based jackknifing, and a generation time of ~30 years was assumed for converting results into calendar years.

Founder age and effective population size estimation.

We used ASCEND v10.1.1[108], which measures allele sharing correlation between pairs of individuals across the genome to infer the age and strength of the founder event in the Thar population. We used a merged dataset with a 1,240k panel dataset[102], which includes ~138k SNPs. Mbuti samples were considered as the outgroup. The founder age was calculated assuming a generation time of 30 years. For inferring demographic history and estimating the effective population, we used HapNe[109] based on LD summary statistics and composite likelihood given in the default parameters.

ROH Analysis

To identify runs of homozygosity (ROHs) across the genome, we used PLINK v1.9[101]. Using the --homozyg function, we applied a sliding window approach to detect ROHs. Each window spanned 1000 kilobases (kb) and included a minimum of 100 SNPs to ensure reliable detection. To accommodate potential genotyping errors and missing data, we permitted up to one heterozygous call and five missing calls within each window. We retained only ROHs that were at least 1000 kb in length for further analysis. To maintain the reliability of our population-level results, we excluded family groups (FIDs) with fewer than four individuals. For each population, we calculated mean ROH statistics, such as the average number of ROH segments and their mean lengths. We focused on specific populations of interest, identified by their groups. The results were visualised using the ggplot2 package in the R software (http://www.r-project.org/).

Haplotype-based analyses

We phased genome-wide dense SNP data using SHAPEIT2[110] and used the Refined IBD and MergeIBD algorithms[33] for identity by descent (IBD) analysis with default settings on a phased dataset. Segments greater than 1 cm were analysed. Estimations were made on an individual basis, and the average IBD segment per pair of individuals was calculated by dividing the total number of segments by the possible combinations of pairs within each population. For ibd score within the population, we used the sample size for normalisation, and computed the raw IBD score only for the total length of IBD segments between 3 and 20 cM and used as mentioned in Nakatsuka et al 2017[34]

The phased genome-wide dense SNP data were used for ChromopainterV2 to estimate the coancestry matrix, followed by running FineSTRUCTURE[111] to perform Bayesian clustering and identify fine-scale population structure, and finally, the results were interpreted using the generated dendrogram and clustering assignments. We used ChromoPainterV2 and ran the analysis twice. First, we performed an initial run on a subset comprising 1/10 of the population (one individual per population), using 10 EM iterations. From this, we took the weighted average of the -n (mutation rate) and -m (switch rate) parameters to fix the mutation and switch rates for the full analysis. Specifically, we used the values inferred from a subset of chromosomes (1, 3, 7, 10, and 22) as described in[112]., which resulted in -n = 555.7575698993 and -m = 0.00033791836. Using these inferred values, we then ran the analysis on all chromosomes with all individuals, applying an all-to-all donor strategy. The results of the mean chunks heatmap were visualised using the ggplot2 package in the R software (http://www.r-project.org/).

Uniparental lineage analysis

We analysed maternal lineages using mtDNA from 176 individuals with the rCRS as a reference. mtDNA haplogroups were assigned with HaploGrep3[113] , based on PhyloTree Build 17 Forensic Update 1.2[113–115]. Y-chromosome haplogroups were determined using Y-LineageTracker[115] for the 63 male individuals out of the 176, following the ISOGG Y-DNA phylogenetic tree (2019–2020).

Ethical Clearance and Funding

This study was conducted in accordance with the ICMR guidelines (National Ethical Guidelines for Biomedical and Health Research involving Human Participants 2017). Ethical approval was obtained from the Institutional Ethics Committee, Indian Institute of Technology Jodhpur.

Acknowledgements

We gratefully acknowledge the support and funding support by the Jodhpur City Knowledge and Innovation Foundation (JCKIF) and seed grant to MM by IIT Jodhpur (JCKIF/Thar/Proj-01/2022, I/SEED/MTM/20220130). Genotyping support from Council of Scientific and Industrial Research – Institute of Genomics and Integrative Biology (CSIR-IGIB) is also acknowledged. We are particularly thankful to Dr. Pooja Sharma for their valuable technical support. We also express our gratitude to Dr. Arindam Mitra, National Institute of Biomedical Genomics (NIBMG), Kolkata, for assistance with preliminary genotype data processing.

We extend our thanks to the Department of Bioscience and Bioengineering, IIT Jodhpur, and to the students for their participation. We acknowledge the contributions of Dr. Manasi Mukherjee (Project Coordinator) and Dr. Ravi Pratap Singh for organizing the health camp during the survey. Thanks to Mr. Padam Singh and Mr. Jitendra Singh for their work as field assistants. Finally, we acknowledge the financial assistance received in the form of student fellowship from the University Grants Commission (UGC), and Ministry of Human Resource Development (MHRD), India.

Data availability

The dataset analyzed and presented in this study is available from the corresponding author upon reasonable request.

Declaration of generative AI and AI-assisted technologies in the writing process:

During the preparation of work , we used ChatGPT to improve the text's readability and language, we reviewed and edited the content as needed and took responsibility for it.

References

1. French, F. G. & Stamp, L. D. *The Indian Empire: Northwest India*. (1926).
2. Al., S. I. et. Historical study of the socio-economic system of ancient indus civilization. *Russ. Law J.* **11**, (2023).
3. Bharucha, R. *Rajasthan, an Oral History: Conversations with Komal Kothari*. (Penguin Global, 2003).
4. Singh, K. S. People of India: An Introduction. Preprint at (1992).

5. Kenoyer, J. M. 10. Interaction systems, specialised crafts and culture change: The Indus Valley Tradition and the Indo-Gangetic Tradition in South Asia. *The Indo-Aryans of Ancient South Asia* 213–257 (3 2020).
6. Nath, A. & Wacziarg, F. *Arts and Crafts of Rajasthan*. (University of Washington Press, 1987).
7. Smith, J. D. Aman Nath and Francis Wacziarg (ed.): Arts and crafts of Rajasthan. 229 pp. London: Thames and Hudson Ltd., 1987. £27.50. *Bull. Sch. Orient. Afr. Stud.* **51**, 624–625 (10 1988).
8. McIntosh, J. *The Ancient Indus Valley : New Perspectives*. 84 (ABC-CLIO, Santa Barbara, CA, 2008).
9. Habib, I. *A People's History of India 2: The Indus Civilization*. (2015).
10. Thapar, R., Allchin, B. & Allchin, R. The Rise of Civilization in India and Pakistan. *J. Am. Orient. Soc.* **104**, (1984).
11. Korisettar, R. Book Review: Irfan Habib, Prehistory and Irfan Habib, The Indus Civilization. *Stud. People's Hist.* **2**, (2015).
12. Jun. Historical Review of Mohenjo-Daro and Harappan Civilization in Pakistan. *Pacific International Journal* **5**, (2022).
13. Bhatia, V. Thar yields Harappan site, shifting Indus Valley map. *The Times Of India* https://timesofindia.indiatimes.com/city/jaipur/thar-yields-harappan-site-shifting-indus-valley-map/articleshow/122983224.cms (2025).
14. Dhir, R. P., Joshi, D. C. & Kathju, S. *Thar Desert in Retrospect and Prospect*. (Scientific Publishers, 2018).
15. Natural Resource Management in the Thar Desert Region of Rajasthan. *Natural Resource Management in the Thar Desert Region of Rajasthan* Preprint at https://doi.org/10.1007/978-3-031-34556-2 (2023).
16. Jarrige, J.-F. Mehrgarh Neolithic. *International Seminar on "First Farmers"* (2006).
17. Mughal, M. 1990. The Decline of the Indus Civilization and the Late Harappan Period in the Indus Valley, Lahore Museum Bulletin (Lahore), Vol. III (2): 1-17. Preprint at https://www.academia.edu/9109775/1990_The_Decline_of_the_Indus_Civilization_and_the_Late_Harappan_Period_in_the_Indus_Valley_Lahore_Museum_Bulletin_Lahore_Vol_III_2_1_17.
18. Kumar, L. *et al.* Genetic Affinities and Adaptation of the South-West Coast Populations of India. *Genome Biol. Evol.* **15**, (12 2023).
19. Narasimhan, V. M. *et al.* The formation of human populations in South and Central Asia. *Science* **365**, (9 2019).

20. Gupta, A., Anderson, D., Pandey, D. & Singhvi, A. Adaptation and human migration, and evidence of agriculture coincident with changes in the Indian summer monsoon during the Holocene. *Current Science* **90**, 1082–1090 (2006).
21. Madella, M. & Fuller, D. Q. Palaeoecology and the Harappan Civilisation of South Asia: a reconsideration. *Quat. Sci. Rev.* **25**, 1283–1301 (6 2006).
22. Dutt, S., Gupta, A. K., Wünnemann, B. & Yan, D. A long arid interlude in the Indian summer monsoon during ~4,350 to 3,450 cal. yr BP contemporaneous to displacement of the Indus valley civilization. *Quat. Int.* **482**, 83–92 (2018).
23. Costantini, L. Beginning of agriculture in the Kachi Plain: the evidence of Mehrgarh. 29–33 Preprint at https://ehrafarchaeology.yale.edu/document?id=aq40-001 (1984).
24. Blinkhorn, J., Achyuthan, H., Petraglia, M. & Ditchfield, P. Middle palaeolithic occupation in the thar desert during the upper pleistocene: The signature of a modern human exit out of Africa? *Quat. Sci. Rev.* **77**, (2013).
25. Moorjani, P. *et al.* Genetic evidence for recent population mixture in India. *Am. J. Hum. Genet.* **93**, 422–438 (9 2013).
26. Reich, D., Thangaraj, K., Patterson, N., Price, A. L. & Singh, L. Reconstructing Indian population history. *Nature* **461**, 489–494 (9 2009).
27. Basu, A., Sarkar-Roy, N. & Majumder, P. P. Genomic reconstruction of the history of extant populations of India reveals five distinct ancestral components and a complex structure. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 1594–1599 (2 2016).
28. Brahmachari, S. K. *et al.* Genetic landscape of the people of India: a canvas for disease gene exploration. *J. Genet.* **87**, 3–20 (2008).
29. Grierson, S. G. A. *The Languages of India: Being a Reprint of the Chapter on Languages*. (1903).
30. *Linguistic Survey of India: Rajasthan Part 1*. (2002).
31. *Climate of Rajasthan State*. (1988).
32. Sehgal, V. K. *Climate, Agriculture and Man Shrinking Biodiversity and Sustainability*. (Scientific Publishers, 2019).
33. Browning, B. L. & Browning, S. R. Improving the accuracy and efficiency of identity-by-descent detection in population data. *Genetics* **194**, 459–471 (2013).
34. Nakatsuka, N. *et al.* The promise of discovering population-specific disease-associated genes in South Asia. *Nat. Genet.* **49**, 1403–1407 (9 2017).
35. Haak, W. *et al.* Massive migration from the steppe was a source for Indo-European languages in

Europe. *Nature* **522**, 207–211 (6 2015).

36. Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192**, 1065–1093 (11 2012).
37. Pathak, A. K. *et al.* The Genetic Ancestry of Modern Indus Valley Populations from Northwest India. *Am. J. Hum. Genet.* **103**, 918 (12 2018).
38. Romero, I. G. *et al.* Herders of indian and european cattle share their predominant allele for lactase persistence. *Mol. Biol. Evol.* **29**, 249–260 (1 2012).
39. Kaufman, E. J. & Tan, C. White as milk: Biocentric bias in the framing of lactose intolerance and lactase persistence. *Sociol Health Illn* **44**, 1533–1550 (2022).
40. Bayoumi, R. *et al.* Positive selection of lactase persistence among people of Southern Arabia. *Am. J. Phys. Anthropol.* **161**, 676–684 (12 2016).
41. Mallick, C. B. *et al.* The Light Skin Allele of SLC24A5 in South Asians and Europeans Shares Identity by Descent. *PLoS Genet.* **9**, (11 2013).
42. Ali, M. *et al.* Characterizing the genetic differences between two distinct migrant groups from Indo-European and Dravidian speaking populations in India. *BMC Genet.* **15**, 86 (7 2014).
43. Chandrasekar, A. *et al.* Updating Phylogeny of Mitochondrial DNA Macrohaplogroup M in India: Dispersal of Modern Human in South Asian Corridor. *PLoS One* **4**, (3 2009).
44. Khurana, P. *et al.* Y Chromosome Haplogroup Distribution in Indo-European Speaking Tribes of Gujarat, Western India. *PLoS One* **9**, e90414 (3 2014).
45. Quintana-Murci, L. *et al.* Where West Meets East: The Complex mtDNA Landscape of the Southwest and Central Asian Corridor. *Am. J. Hum. Genet.* **74**, 827–845 (2004).
46. Metspalu, M. *et al.* Most of the extant mtDNA boundaries in South and Southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans. *BMC Genet.* **5**, (8 2004).
47. Zhao, Z., Khan, F., Borkar, M., Herrera, R. & Agrawal, S. Presence of three different paternal lineages among North Indians: a study of 560 Y chromosomes. *Ann Hum Biol* **36**, 46–59 (2009).
48. Pathak, A. K. *et al.* Human Y chromosome haplogroup L1-M22 traces Neolithic expansion in West Asia and supports the Elamite and Dravidian connection. *iScience* **27**, 110016 (2024).
49. Kivisild, T. *et al.* Deep common ancestry of indian and western-Eurasian mitochondrial DNA lineages. *Curr. Biol.* **9**, 1331–1334 (11 1999).
50. Singh, K. S. *People of India: Rajasthan*. (Popular Prakashan, 1998).
51. Davis, S. *The Bikaner School: Usta Artisans and Their Heritage*. (2008).
52. Ranjan, A. & Ranjan, M. P. *Handmade in India: A Geographic Encyclopedia of India Handicrafts*.

(Abbeville Publishing Group, 2009).

53. India. Office of the Registrar General. *Bibliography on Scheduled Castes, Scheduled Tribes, and Selected Marginal Communities of India: A-K Series.-[2]L-Z Series*. (1970).
54. Singh, K. S. *People of India*. (1992).
55. Saraswati, B. *Pottery-Making Cultures and Indian Civilization*. (Abhinav Publications, 1978).
56. Khairnar, R. Inconceivable creation: The kalbeliya dance from & its conception to worldwide identification. *ShodhKosh J. Vis. Per. Arts* **4**, (2023).
57. Sinha, R. Parallelism between indian kalbeliya folk dance and european roma:gypsy dance JNU. **27**, 2025 (01 2025).
58. Mathur, U. B. *Ethnographic Atlas of Rajasthan - Census 1961*. xvi,138 (Superintendent of Census Operations, Rajasthan, Rajasthan, India, 1969).
59. Ibbetson, S. D. & Maclagan. *Glossary of the Tribes and Castes of the Punjab and North West Frontier Province*. (Asian Educational Services, 1990).
60. Singh, U. A History of Ancient and Early Medieval India : From the Stone Age to the 12th Century. 677 (2018).
61. Coulter, C. R. & Turner, P. *Encyclopedia of Ancient Deities*. (Taylor & Francis, 2013).
62. Swarup, S. *Mughal Art: A Study in Handicrafts*. (1996).
63. Bhushan, J. Costumes And Textiles Of India. (1958).
64. Darbar, M. Report On The Census Of 1891, Volume Ii, The Castes Of Marwar. (1894).
65. Dhillon, B. S. & Dhillon, B. S. *History and Study of the Jats: With Reference to Sikhs, Scythians, Alans, Sarmatians, Goths, and Jutes*. (Gloucester, Ont. : Beta Publishers Incorporated, 1994).
66. UNESCO. *History of Civilizations of Central Asia: The Development of Sedentary and Nomadic Civilizations: 700 B.C. to A.D. 250*. (UNESCO Publishing, 1994).
67. Gokhale, B. G. *Ancient India: History and Culture*. (Popular Prakashan, 1995).
68. Kulke, H. & Rothermund, D. A history of India. 432 (2004).
69. Kumawat, R. K., Shrivastava, P., Shrivastava, D., Mathur, G. K. & Dixit, S. Genomic blueprint of population of Rajasthan based on autosomal STR markers. *Ann. Hum. Biol.* **47**, 70–75 (1 2020).
70. Kumawat, R. K. *et al.* Peopling of Rajasthan, India: Evaluating the gene flow from east and west. *Gene Rep.* **22**, (3 2021).
71. Dada, R. *et al.* Genetic sketch of the six population groups of Rajasthan: a study based on 12 autosomal loci. *Anthropol. Sci.* **119**, 259–264 (12 2011).
72. Metspalu, M. *et al.* Shared and unique components of human population structure and

genome-wide signals of positive selection in South Asia. *Am. J. Hum. Genet.* **89**, 731–744 (12 2011).

73. Pagani, L., Colonna, V., Tyler-Smith, C. & Ayub, Q. An Ethnolinguistic and Genetic Perspective on the Origins of the Dravidian-Speaking Brahui in Pakistan. *Man India* **97**, 267–278 (2017).
74. Kivisild, T. *et al.* The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. *Am. J. Hum. Genet.* **72**, 313–332 (2 2003).
75. McElreavey, K. & Quintana-Murci, L. A population genetics perspective of the Indus Valley through uniparentally-inherited markers. *Ann. Hum. Biol.* **32**, 154–162 (3 2005).
76. Xing, J. *et al.* Genetic diversity in India and the inference of Eurasian population expansion. *Genome Biol.* **11**, (11 2010).
77. Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409–413 (2014).
78. Underhill, P. A. *et al.* Separating the post-Glacial coancestry of European and Asian Y chromosomes within haplogroup R1a. *Eur J Hum Genet* **18**, 479–484 (2010).
79. Narasimhan, V. M. *et al.* The Genomic Formation of South and Central Asia. *bioRxiv* (2018) doi:10.1101/292581.
80. de Barros Damgaard, P. *et al.* The first horse herders and the impact of early Bronze Age steppe expansions into Asia. *Science* **360**, (6 2018).
81. Mathieson, I. *et al.* Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499–503 (12 2015).
82. *Census of India, 1991: Rajasthan. A Portrait of Population*. (1999).
83. *Report on the Census of 1891. Volume II. The Castes of Marwar. Illustrated. (By Babu Raghbir Sahai.)*. (1894).
84. Commissioner, I. C. *Census of India, 1921*. (1923).
85. *National Ethical Guidelines for Health and Health-Related Research, 2017*. (2018).
86. Miller, S. A., Dykes, D. D. & Polesky, H. F. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* **16**, 1215 (1988).
87. Zhao, S. *et al.* Strategies for processing and quality control of Illumina genotyping arrays. *Brief. Bioinform.* **19**, 765–775 (2018).
88. Bhattacharjee, E. *et al.* Genetic variants associated with spontaneous preterm birth in women from India: a prospective cohort study. *Lancet Reg Health Southeast Asia* **14**, 100190 (2023).
89. Behar, D. M. *et al.* The genome-wide structure of the Jewish people. *Nature* **466**, 238–242 (7

2010).

90. Li, J. Z. *et al.* Worldwide human relationships inferred from genome-wide patterns of variation. *Science* **319**, 1100–1104 (2 2008).
91. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (9 2015).
92. Indian Genome Variation Consortium. The Indian Genome Variation database (IGVdb): a project overview. *Hum Genet* **118**, 1–11 (2005).
93. Yunusbayev, B. *et al.* The caucasus as an asymmetric semipermeable barrier to ancient human migrations. *Mol. Biol. Evol.* **29**, 359–365 (2012).
94. Fairley, S., Lowy-Gallego, E., Perry, E. & Flicek, P. The International Genome Sample Resource (IGSR) collection of open human genomic variation resources. *Nucleic Acids Res* **48**, D941–D947 (2020).
95. Yi, W., Delroba, E., Zizzamia, B., Spera, S. & Yang, M. A. The AADR Visualizer: An ArcGIS Online Visualizer for ancient human DNA from the Allen Ancient DNA Resource. *Bioinformatics* **5**, (2025).
96. Mondal, M. *et al.* Genomic analysis of Andamanese provides insights into ancient human migration into Asia and adaptation. *Nat Genet* **48**, 1066–1070 (2016).
97. Me, A. *et al.* Population genomics of Bronze Age Eurasia. *Nature* vol. 522 167–172.
98. Lazaridis, I. *et al.* Genomic insights into the origin of farming in the ancient Near East. *Nature* **536**, 419–424 (7 2016).
99. Allentoft, M. E. *et al.* Population genomics of Bronze Age Eurasia. *Nature* **522**, 167–172 (6 2015).
100. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet* **2**, e190 (2006).
101. 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).
102. Mallick, S. *et al.* The Allen Ancient DNA Resource (AADR) a curated compendium of ancient human genomes. *Sci Data* **11**, 182 (2024).
103. R Core Team. R: A Language and Environment for Statistical Computing. Preprint at https://www.R-project.org/ (2021).
104. Pickrell, J. K. & Pritchard, J. K. Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *PLoS Genet.* **8**, (11 2012).
105. Petr, M., Vernot, B. & Kelso, J. admixr-R package for reproducible analyses using ADMIXTOOLS. *Bioinformatics* **35**, 3194–3195 (2019).

106. Kerdoncuff, E. *et al.* 50,000 years of evolutionary history of India: Impact on health and disease variation. *Cell* **188**, 3389–3404.e6 (2025).
107. Loh, P. R. *et al.* Inferring admixture histories of human populations using linkage disequilibrium. *Genetics* **193**, 1233–1254 (2013).
108. Tournebize, R., Chu, G. & Moorjani, P. Reconstructing the history of founder events using genome-wide patterns of allele sharing across individuals. *PLoS Genet* **18**, e1010243 (2022).
109. Fournier, R., Tsangalidou, Z., Reich, D. & Palamara, P. F. Haplotype-based inference of recent effective population size in modern and ancient DNA samples. *Nat Commun* **14**, 7945 (2023).
110. Delaneau, O., Marchini, J. & Zagury, J.-F. A linear complexity phasing method for thousands of genomes. *Nat Methods* **9**, 179–181 (2011).
111. Lawson, D. J., Hellenthal, G., Myers, S. & Falush, D. Inference of population structure using dense haplotype data. *PLoS Genet.* **8**, (1 2012).
112. Montinaro, F. *et al.* Unravelling the hidden ancestry of American admixed populations. *Nat. Commun.* **6**, (2015).
113. Schönherr, S., Weissensteiner, H., Kronenberg, F. & Forer, L. Haplogrep 3 - an interactive haplogroup classification and analysis platform. *Nucleic Acids Res* **51**, W263–W268 (2023).
114. van Oven, M. & Kayser, M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat* **30**, E386–94 (2009).
115. Chen, H., Lu, Y., Lu, D. & Xu, S. Y-LineageTracker: a high-throughput analysis framework for Y-chromosomal next-generation sequencing data. *BMC Bioinformatics* **22**, 1–15 (2021).

Figure 1. Genetic structure and affinities of populations from the Thar Desert, Rajasthan.
(a) Map showing the geographical and sampling locations of the newly studied population groups from four major districts of Rajasthan (Thar Desert region of India).
(b) Principal component analysis (PCA) of the merged dataset with worldwide reference populations (see Methods). The zoomed inset highlights the study populations from representative crafts, illustrating their placement within the north–south genetic ancestry cline of South Asia.
(c) Unsupervised clustering results from ADMIXTURE analysis at K = 9, displayed as a stacked bar plot. Global populations are ordered geographically (Africa, Arabia and the Middle East, Caucasus and Europe, Central Asia, Southeast and East Asia, Papua New Guinea, and South Asia). The zoomed inset focuses on South Asian populations, with study groups labelled (see Methods).
(d) Outgroup f3 statistics of the form f3(TH-IJE-4-R, X; Yoruba), represented as a geographical gradient map. Populations with the highest allele sharing are indicated in blue (high affinity), green (moderate affinity), and red (low affinity) to West Eurasian. Asterisk (✳) the top 15 populations showing the greatest amount of shared drift with TH-IJE-4-R.

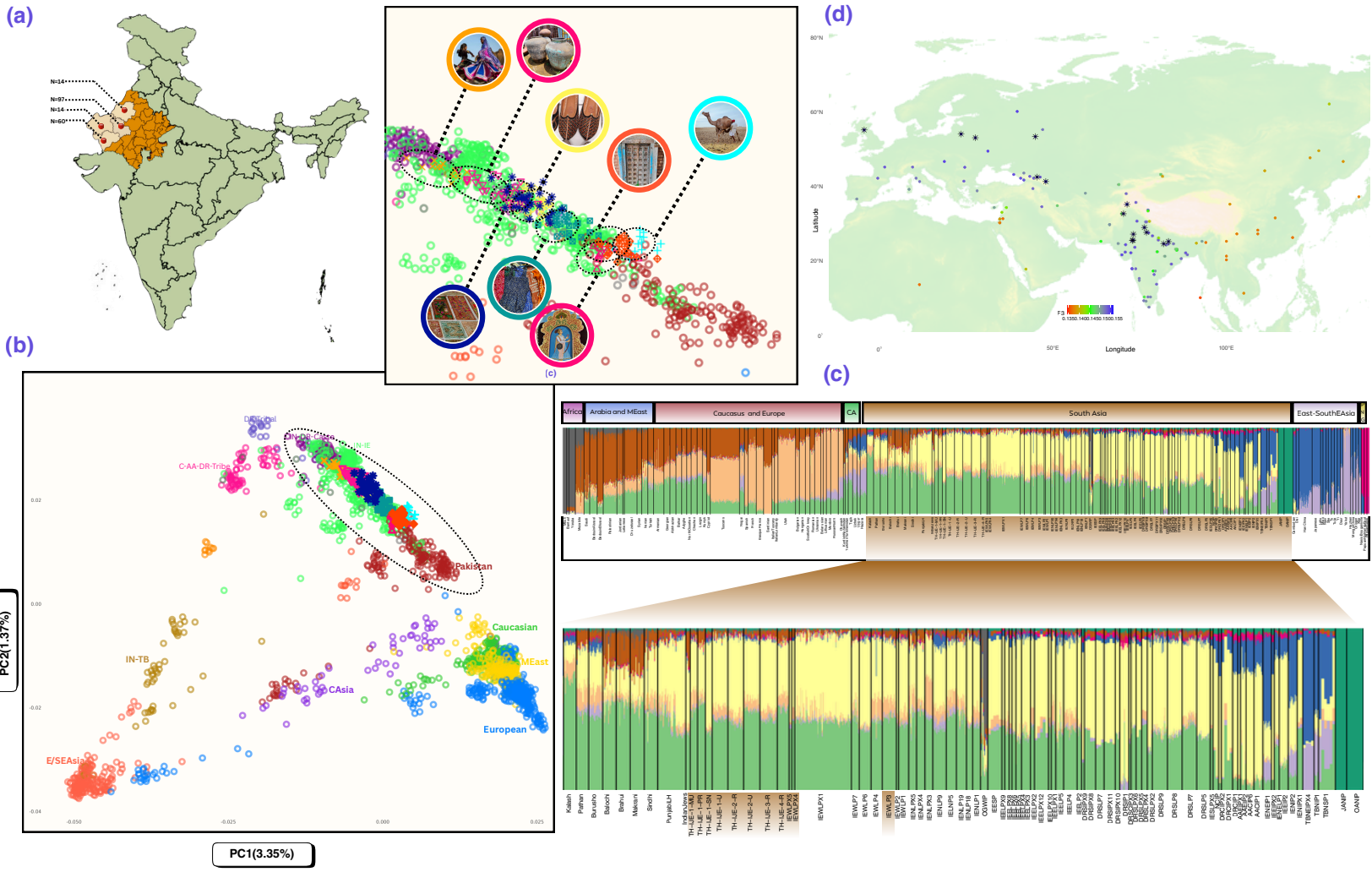


Figure 2. Pairwise population differentiation measured by Fst.
Matrix of pairwise Fst (fixation index) values calculated using SMARTPCA across all reference populations. Shading intensity reflects the magnitude of genetic divergence, as indicated by the color scale. Populations are ordered geographically and grouped by continent, with the newly sampled Thar Desert craft populations highlighted for comparison.

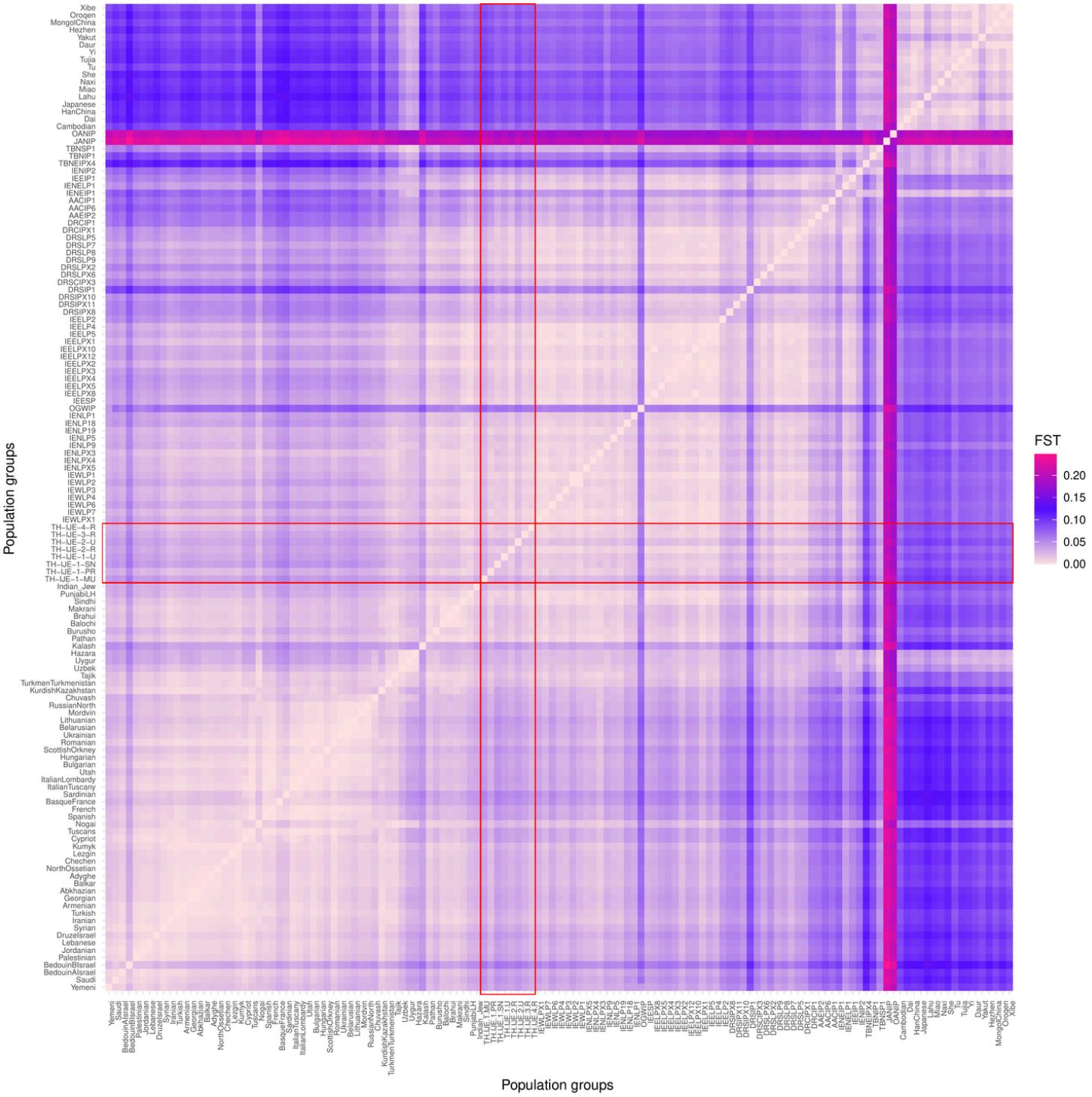


Figure: 3 Genome-wide principal component analysis (PCA) showing the genetic structure of Thar Desert craft populations in the context of modern South Asian and other Eurasian populations. Populations are ordered by geography and linguistic affiliation and are color-coded accordingly. The PCA highlights the distribution of study populations along the Ancestral North Indian (ANI)–Ancestral South Indian (ASI) cline, illustrating their relative positions within South Asian genetic diversity.

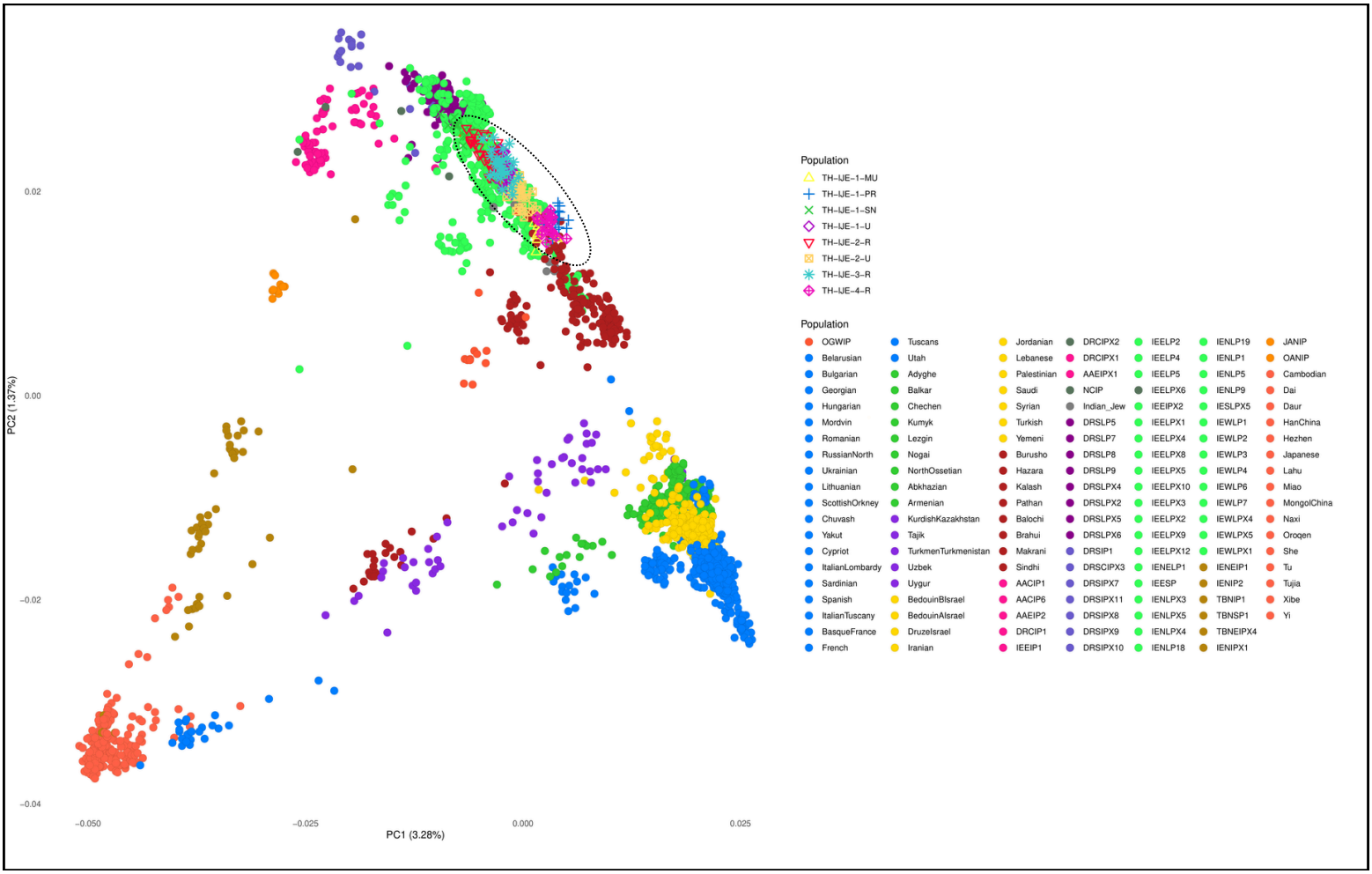


Figure 4: Genome-wide principal component analysis (PCA) illustrating the genetic structure of Thar Desert craft populations in the context of ancient DNA. The plot includes modern Eurasian populations alongside selected ancient sources from the Central and Western Steppe, Iran, Anatolia, Central Asia, and South Asia, highlighting the study populations in relation to both ancient and contemporary South Asian

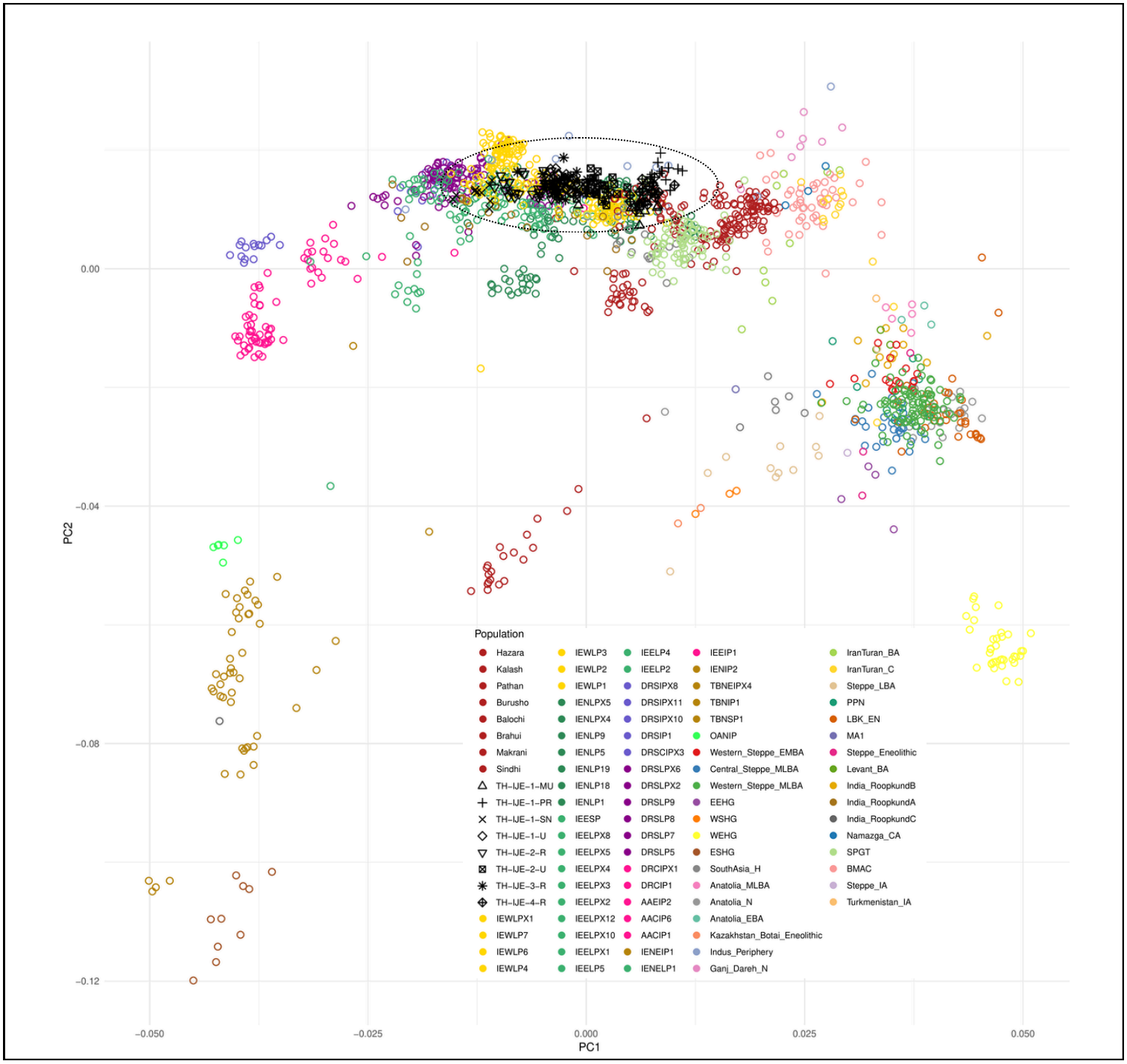


Figure 5. Genetic ancestry components inferred from ADMIXTURE analysis.
(a) Unsupervised clustering results at K = 9, displayed as a stacked bar plot. Global populations are ordered geographically (Africa, Arabia and the Middle East, Central Asia, Caucasus and Europe, Southeast and East Asia, Papua New Guinea(PNG), and South Asia). The inset highlights South Asian populations, with the newly studied Thar Desert groups labelled (see Methods).
(b) Cross-validation results for ADMIXTURE runs with K values ranging from 2 to 18, illustrating the hierarchical structure of genetic components across populations.
(c) ADMIXTURE results across multiple K values (K = 2–15), shown as stacked bar plots, illustrating the hierarchical emergence of ancestry components.

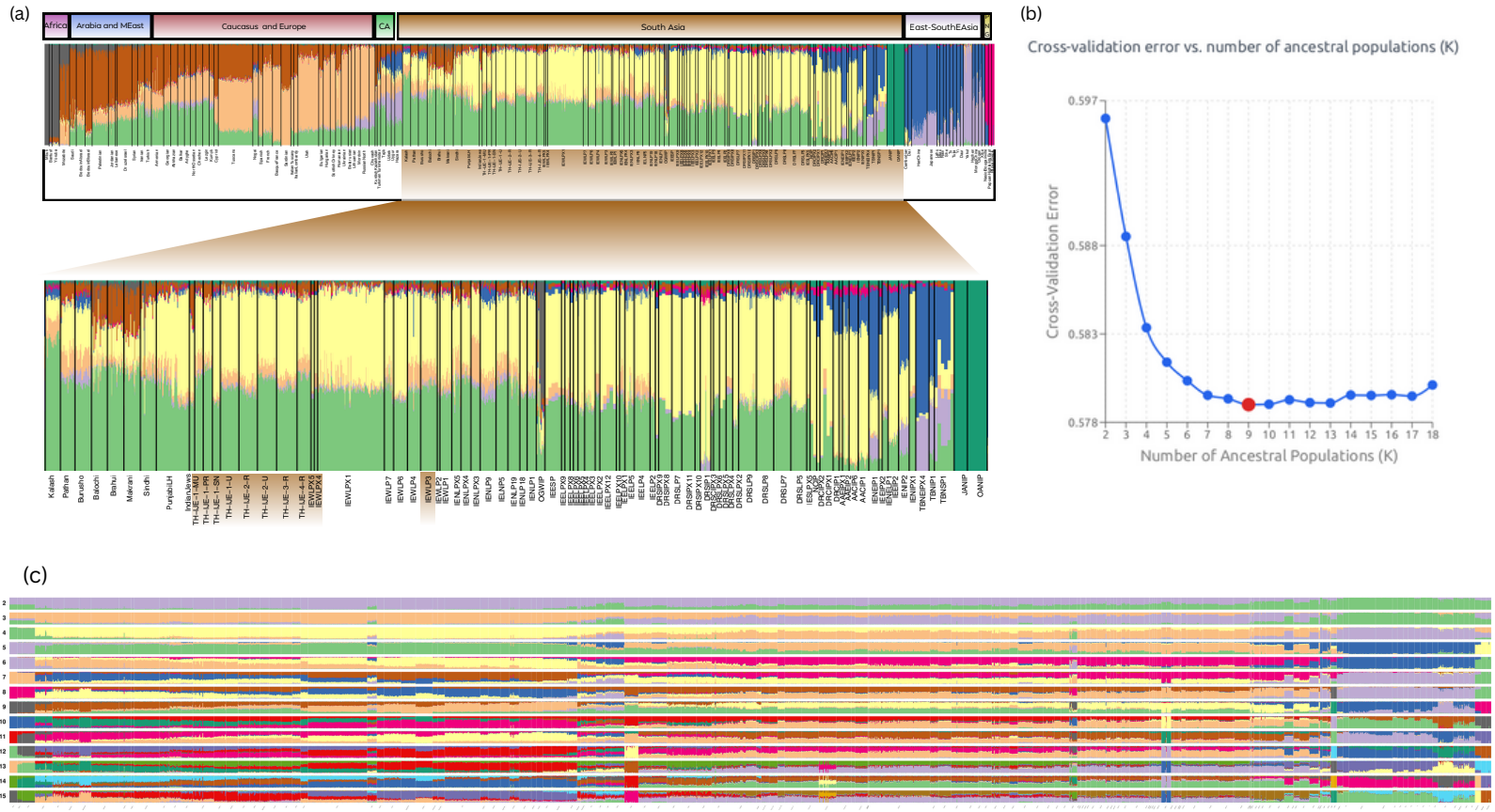


Figure 6: Maximum likelihood tree depicting the genetic relationships of Thar Desert populations with neighboring groups, including the Western Frontier and other Eurasian populations, as inferred by TreeMix. Branch lengths reflect population-specific drift, highlighting higher genetic drift observed in Thar populations relative to other groups.

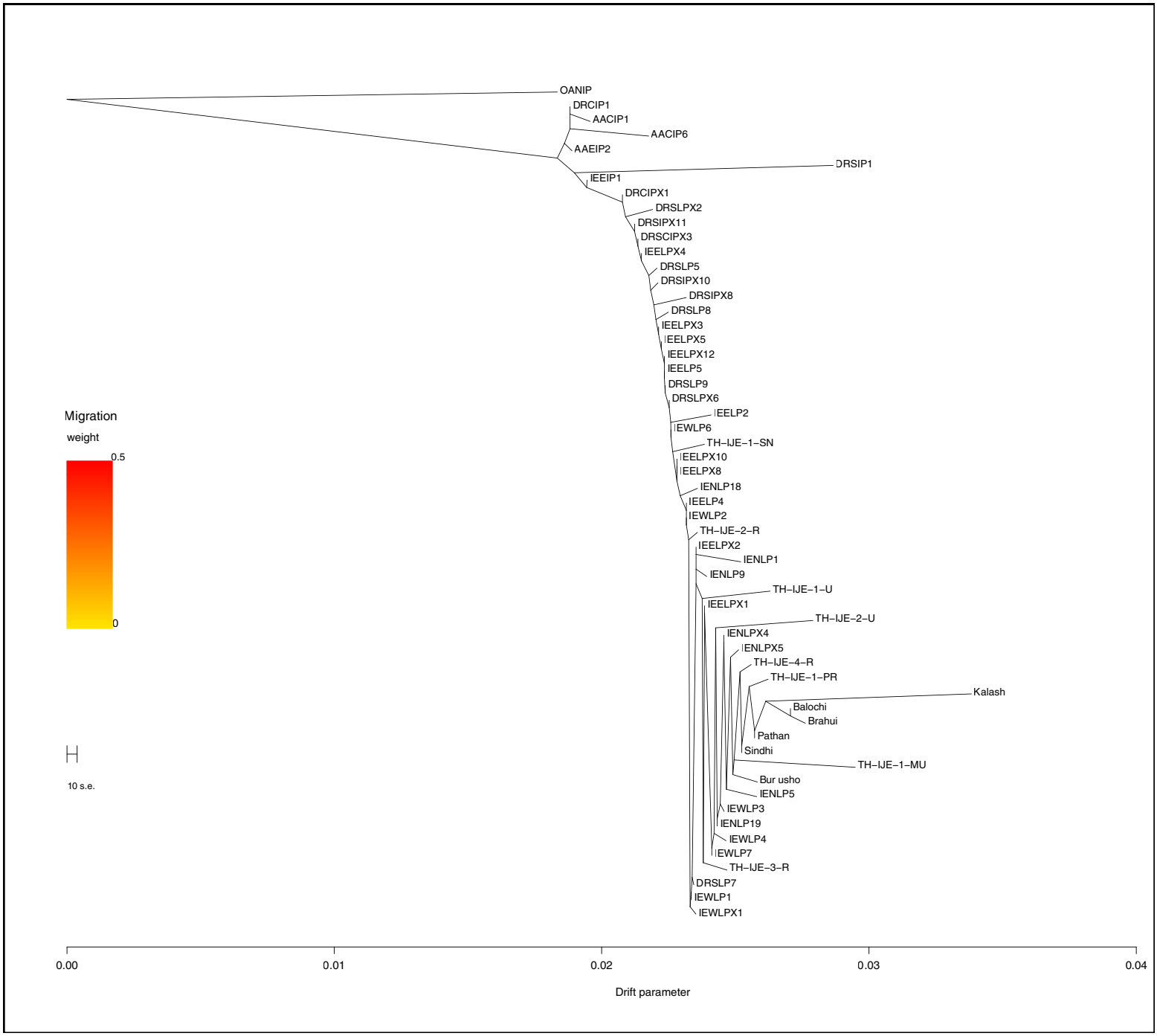


Figure 7: Average number of IBD (Identity-By-Descent) segments shared per pair of individuals in Thar Desert populations and the Pathan ethnic group from the Western Frontier, as inferred using refined IBD (see Methods).

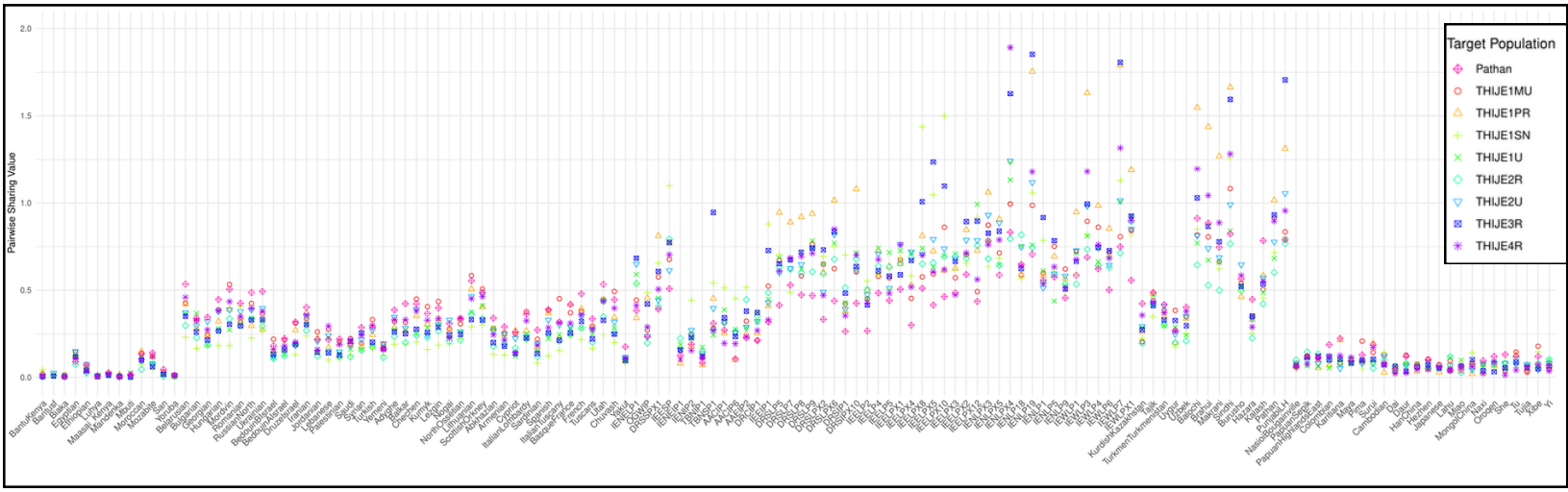


Figure 8: Runs of Homozygosity (RoH) in Thar Desert populations compared with other South Asian groups. The plot displays the mean number of RoH segments versus the mean segment length in kb. RoH were identified using PLINK v1.9 (see Methods). Populations are colour-coded and grouped by linguistic affiliation. The distribution patterns of recent inbreeding and population isolation in Thar populations relative to other South Asian groups.

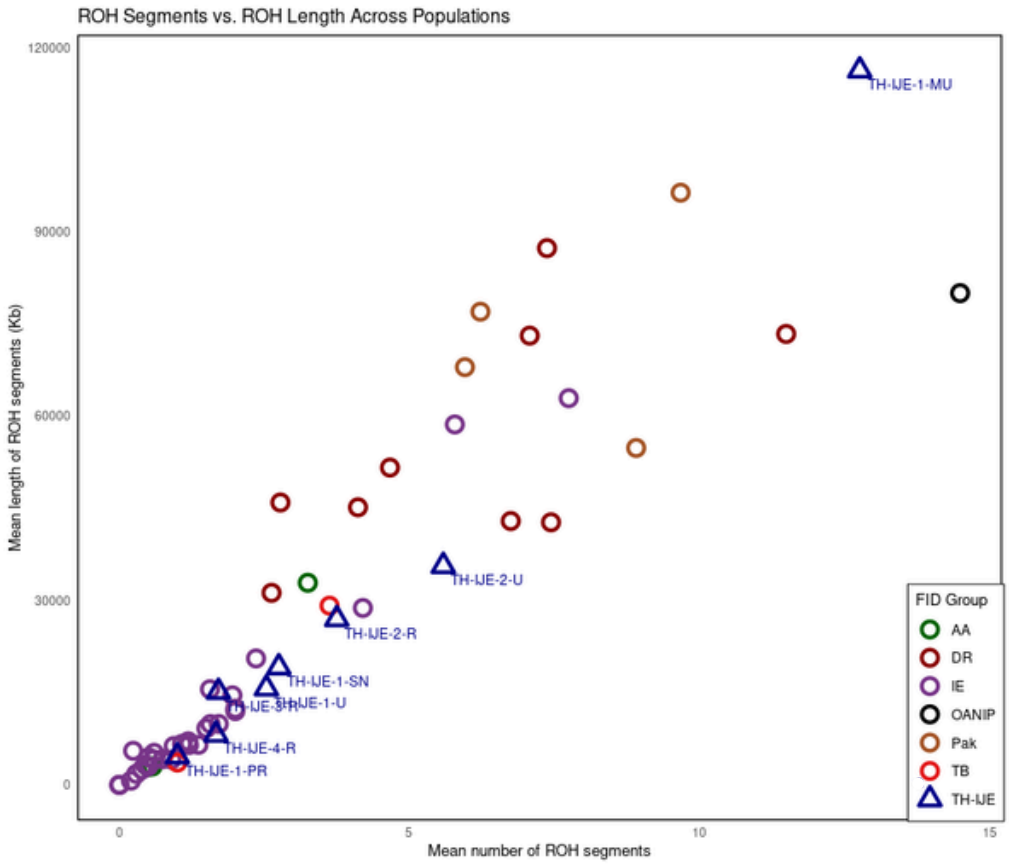


Figure 9: Outgroup f_3 statistics measuring shared genetic drift between populations relative to an outgroup, using ancient reference sources for South Asian and West Eurasian populations. Panels show results with (A) Anatolia_N, (B) EEHG, (C) Central_Steppe_MLBA, and (D) Ganj_Dareh_N. The statistic is calculated as f_3(Ancient Source, X; Yoruba), where X represents a South Asian or West Eurasian population. Error bars indicate jackknife standard errors. Colors correspond to Z-scores, with the gradient legend shown alongside.

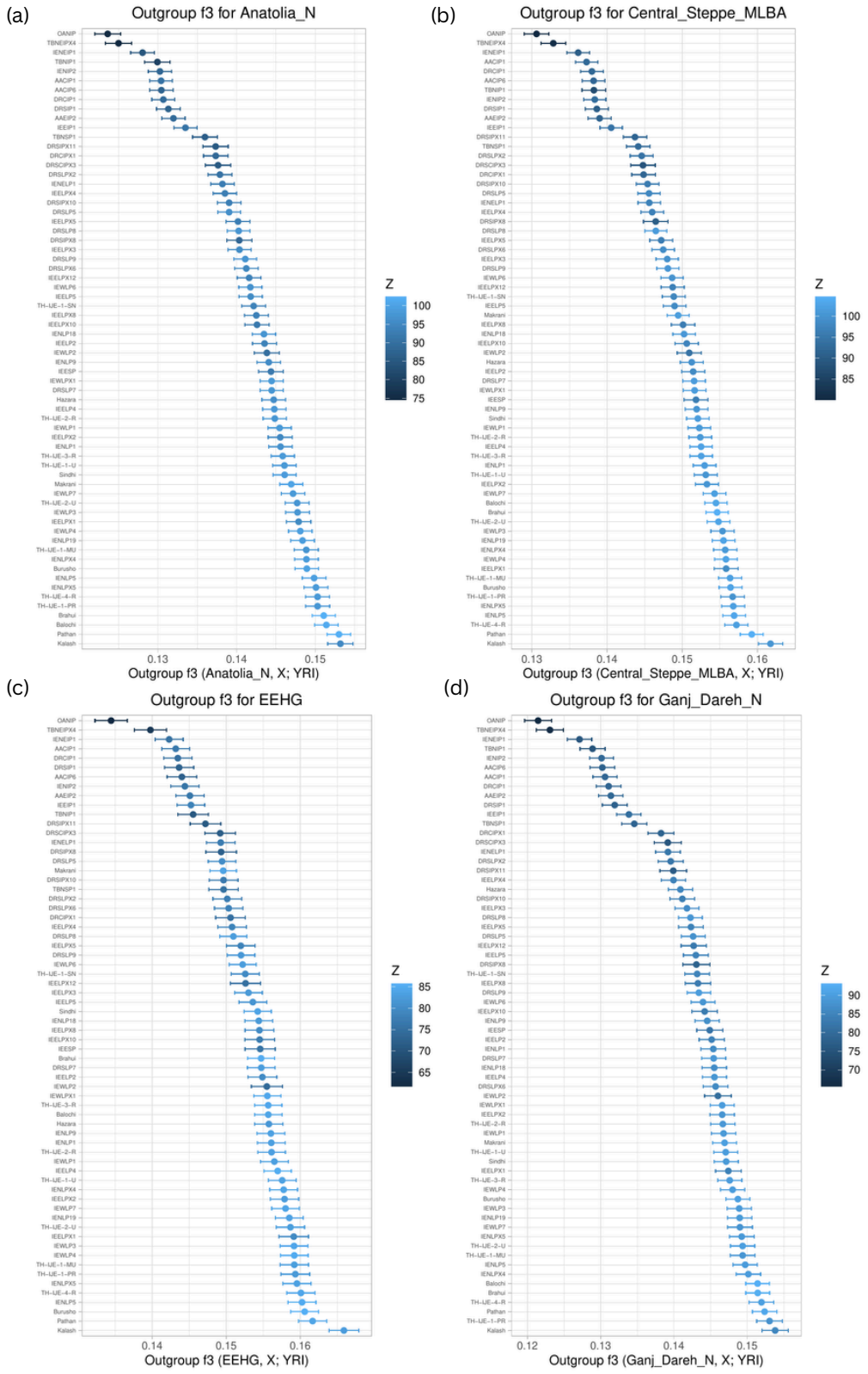


Figure 10: Derived allele sharing of modern Thar Desert populations with other Eurasian groups. Scatter plots show the results of the outgroup f_3 test of the form f_3(Thar or western frontier population, X; Yoruba), where X represents any other modern Eurasian population. Error bars indicate jackknife-derived standard errors, highlighting the relative allele sharing between the study populations and Western Eurasian groups.

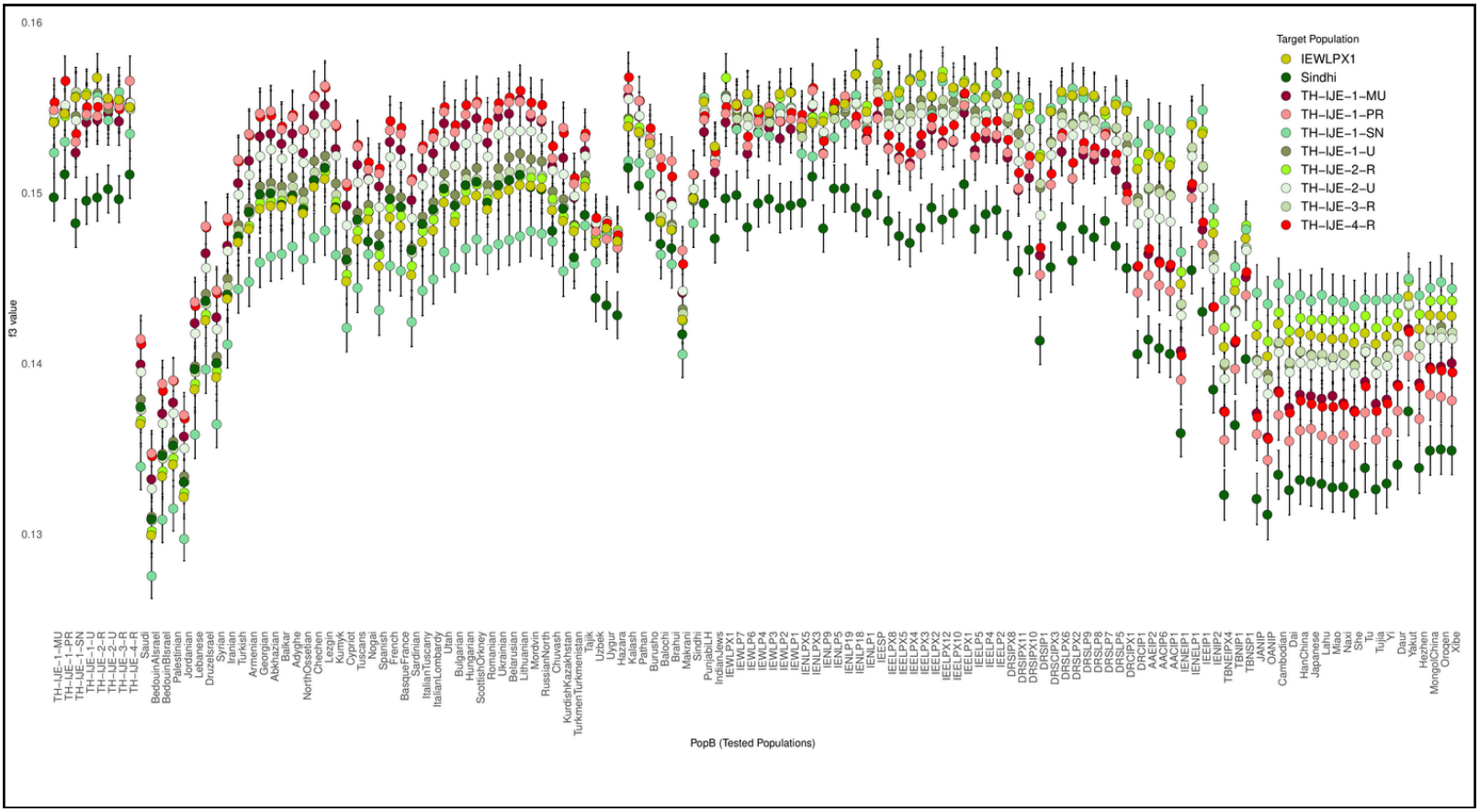


Figure 11. Gene flow and relative affinities of Thar Desert populations with Indo-European groups and Europe.
D-statistics of the form D(pop1, Yoruba; pop2, pop3) were computed, where pop1 corresponds to the Thar individual IEWLPX1, pop2 represents modern South Asian populations, and pop3 includes South Asian or West Eurasian reference groups. The results highlight patterns of allele sharing among the western frontier populations, North/West Indians, and Indo-European groups of northern India. The Thar craft communities show particular genetic proximity to Pathan and Kalash populations, and demonstrate their strongest affinities with European groups compared to other South Asians

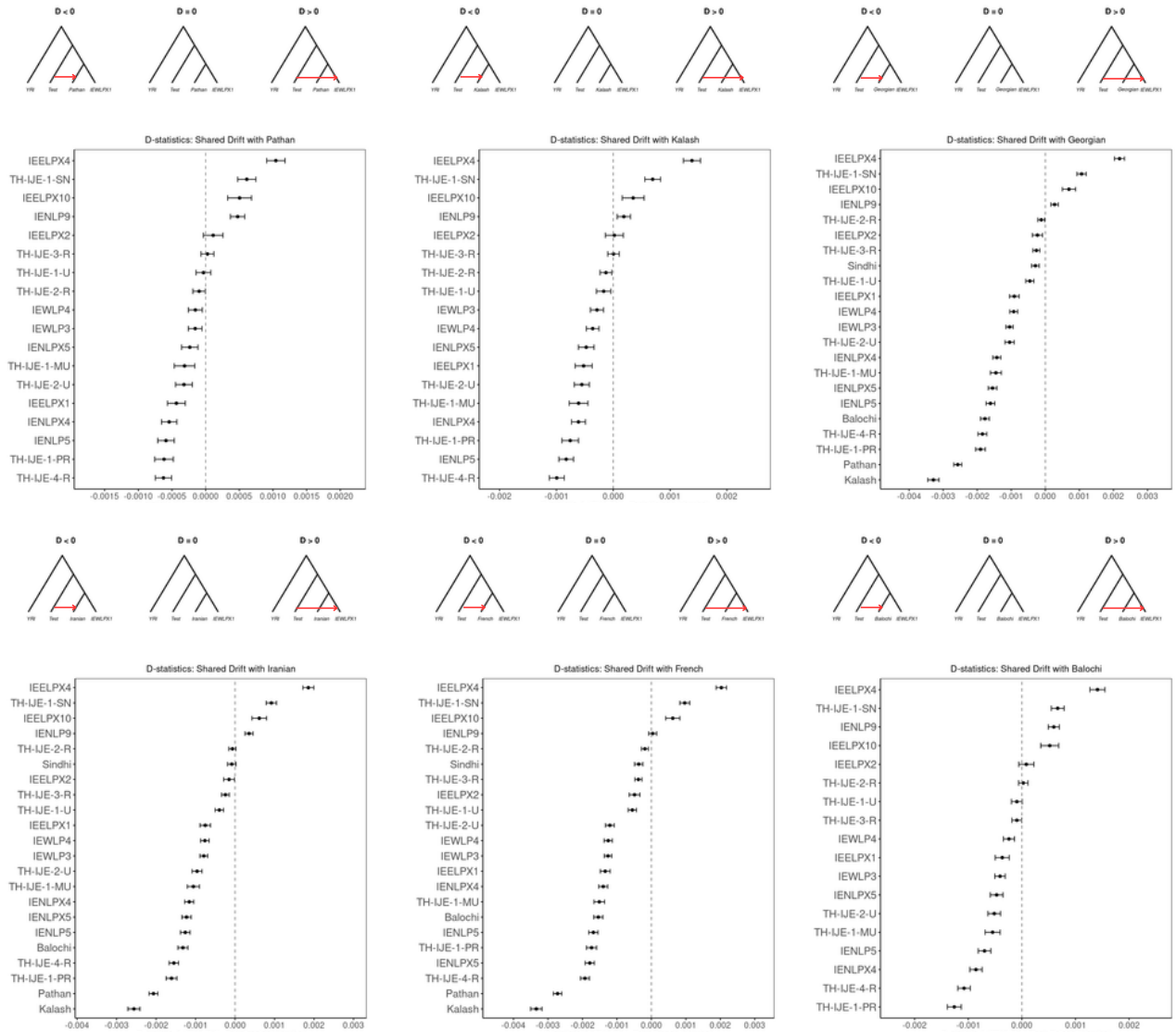


Figure 12: Heatmap showing the mean haplotype chunks received by Thar Desert populations and selected North/Western Indo-EuropeanA populations, as inferred using CHROMOPAINTER. Chunks represent shared genomic segments donated from populations across Eurasia, highlighting patterns of allele sharing and gene flow, particularly into the Thar populations. The color scale reflects the relative amount of shared ancestry, illustrating contributions from different Eurasian source populations to both the study and comparative groups.

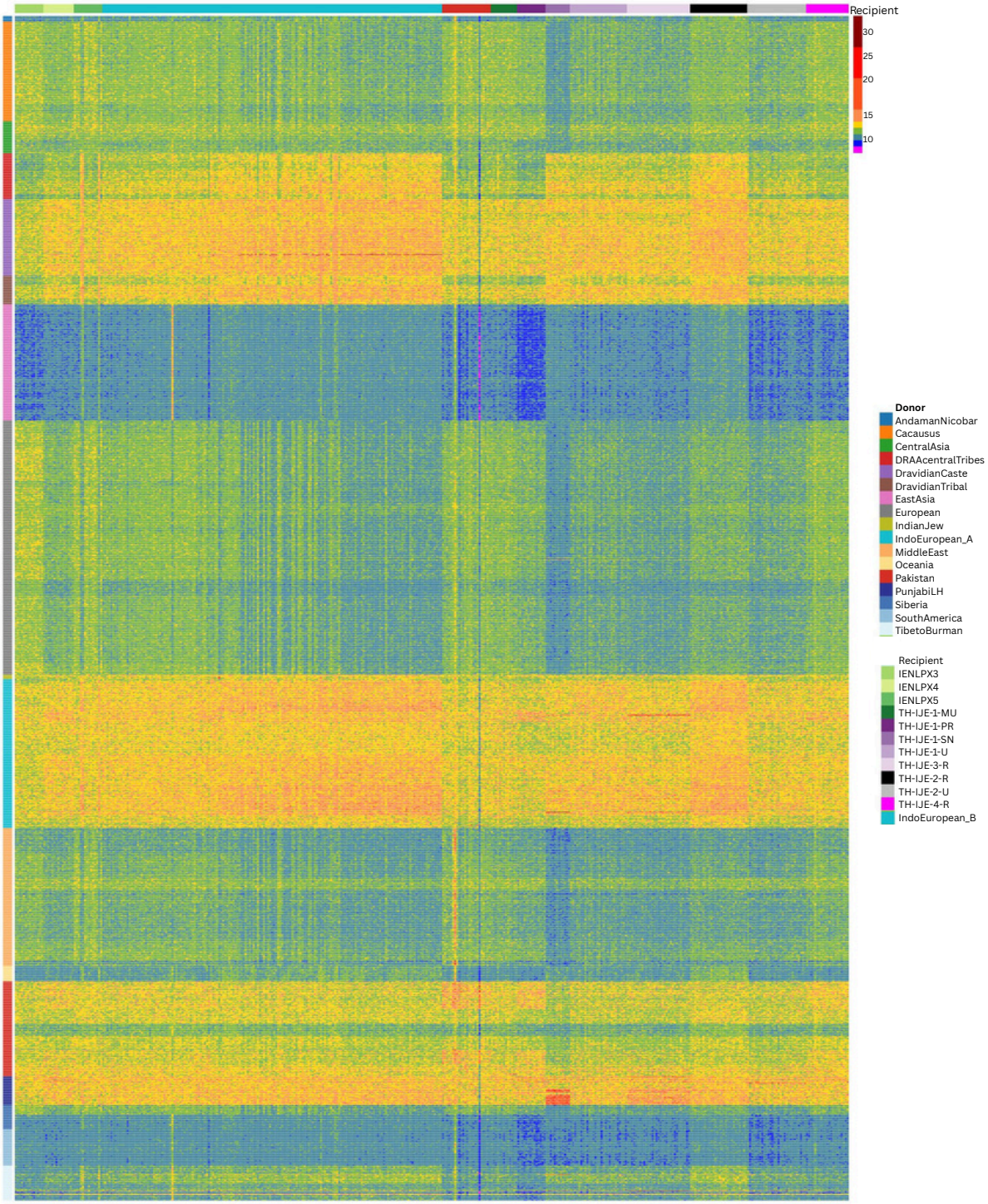
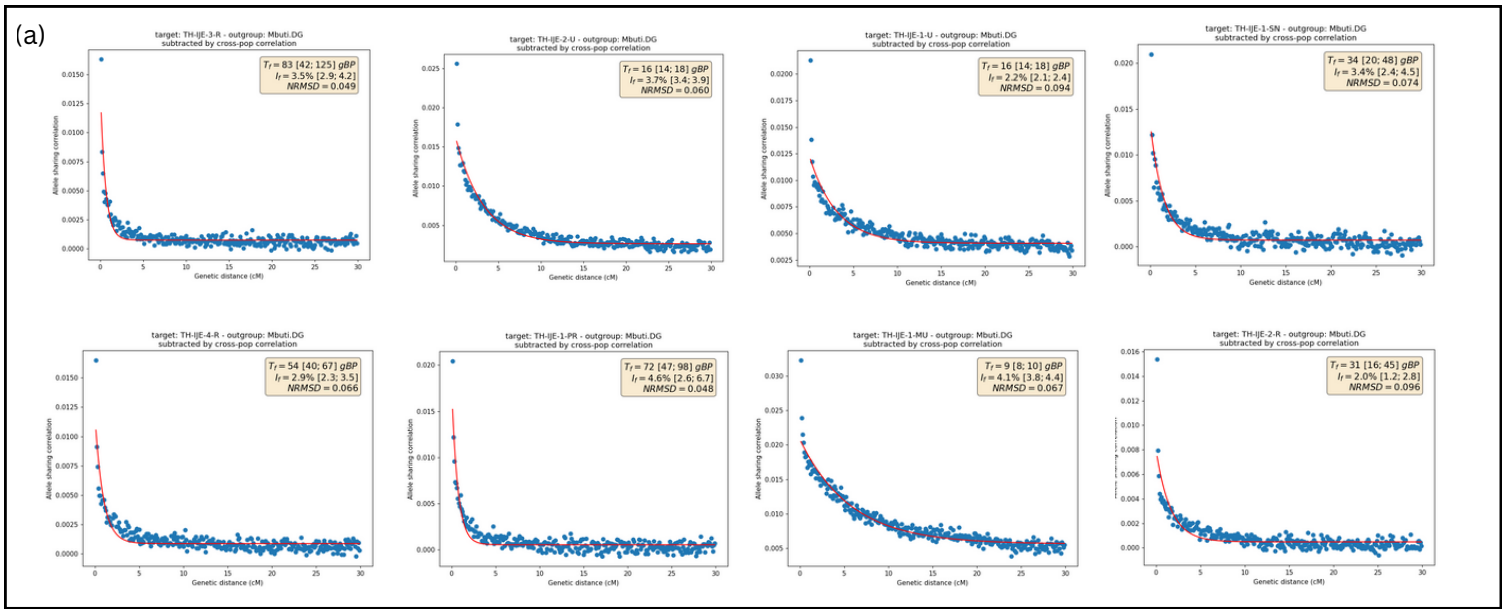


Figure 13: Demographic history of Thar Desert populations.
(a) Founder events inferred using ASCEND, showing the occurrence and magnitude in the Thar population. Plots depict the time of founder events (Tf) and their strength (If), highlighting any historical bottlenecks or reductions in population size.
(b) Historical effective population size (Ne) inferred using HapNe-IBD. The x-axis represents generations before present, and the y-axis shows Ne estimates. Light and dark shaded areas correspond to 95% and 50% confidence intervals, respectively, obtained via bootstrap quantiles, capturing recent changes in the effective population size of the Thar populations.



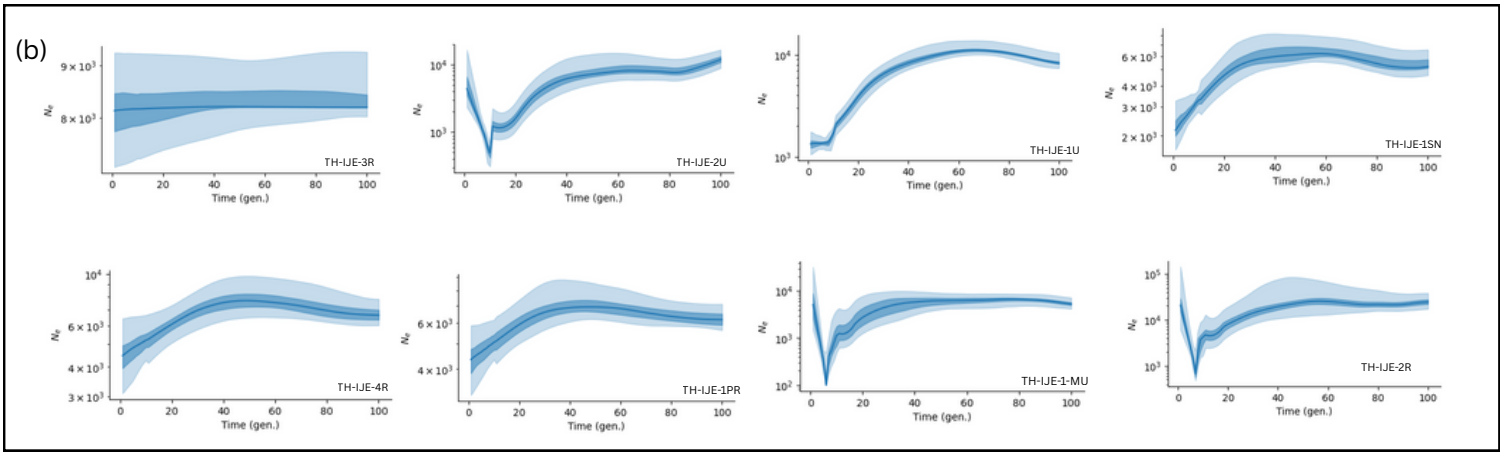
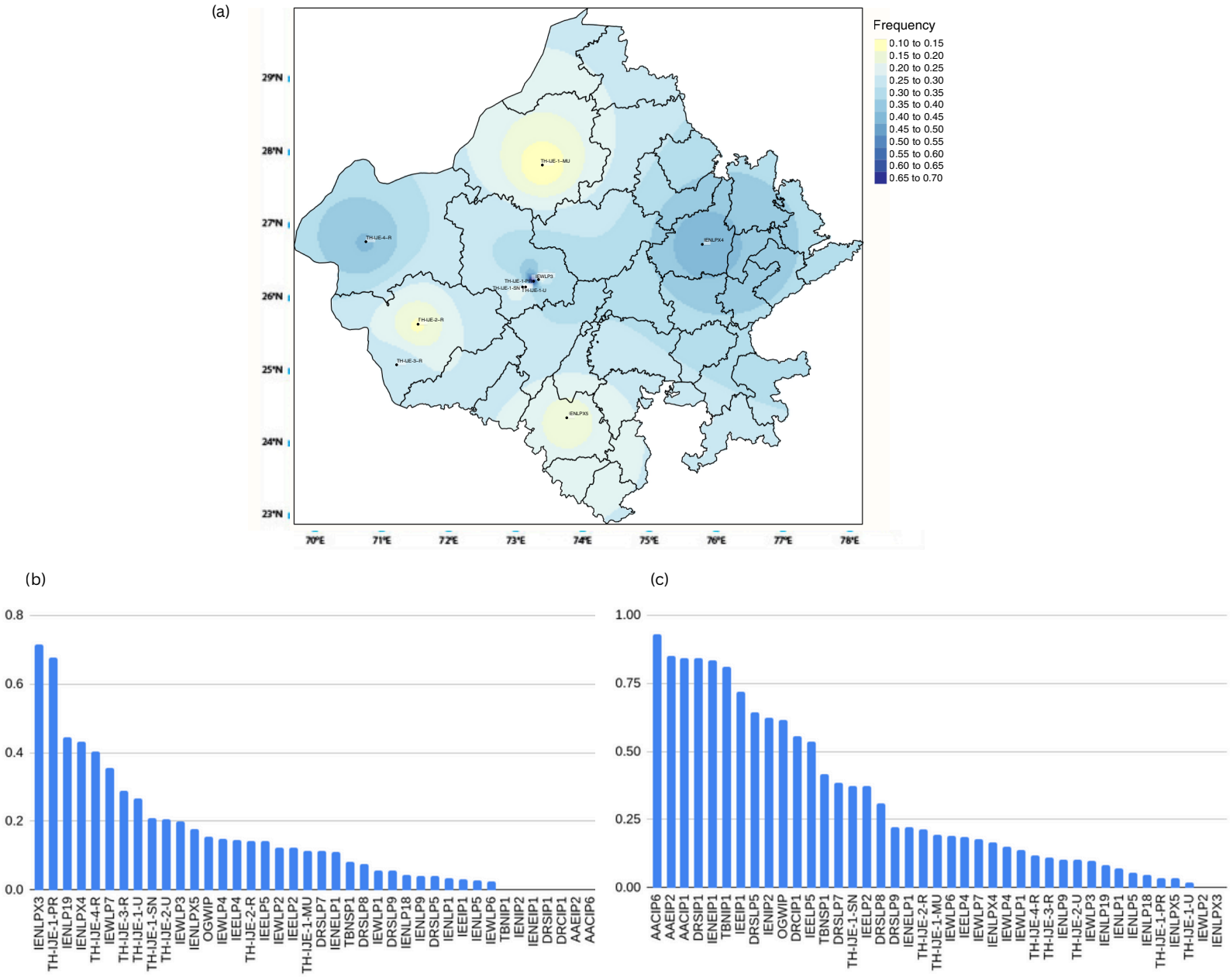



Figure 14: (a) Spatial frequency map. (b) Bar plot showing the distribution of minor allele frequencies of SNP rs4988235-13910*TIN LCT from Rajasthan populations. The colour gradient below each map represents the observed frequency from minimum to maximum. (c) Bar plot showing the frequency distribution of SNP rs1426654 in SLC24A5, with a skin pigmentation gradient (darker to lighter) indicated for Indian populations.