The Genetic Ancestry of Modern Indus Valley Populations from Northwest India

Ajai K. Pathak,[1,*] Anurag Kadian,[2] Alena Kushniarevich,[1,3] Francesco Montinaro,[1,4] Mayukh Mondal,[1] Linda Ongaro,[1] Manvendra Singh,[5] Pramod Kumar,[6] Niraj Rai,[7] Jüri Parik,[1] Ene Metspalu,[1] Siiri Rootsi,[1] Luca Pagani,[1,8] Toomas Kivisild,[1,9] Mait Metspalu,[1] Gyaneshwer Chaubey,[1,10,*] and Richard Villems[1]

The Indus Valley has been the backdrop for several historic and prehistoric population movements between South Asia and West Eurasia. However, the genetic structure of present-day populations from Northwest India is poorly characterized. Here we report new genome-wide genotype data for 45 modern individuals from four Northwest Indian populations, including the Ror, whose long-term occupation of the region can be traced back to the early Vedic scriptures. Our results suggest that although the genetic architecture of most Northwest Indian populations fits well on the broader North-South Indian genetic cline, culturally distinct groups such as the Ror stand out by being genetically more akin to populations living west of India; such populations include prehistorical and early historical ancient individuals from the Swat Valley near the Indus Valley. We argue that this affinity is more likely a result of genetic continuity since the Bronze Age migrations from the Steppe Belt than a result of recent admixture. The observed patterns of genetic relationships both with modern and ancient West Eurasians suggest that the Ror can be used as a proxy for a population descended from the Ancestral North Indian (ANI) population. Collectively, our results show that the Indus Valley populations are characterized by considerable genetic heterogeneity that has persisted over thousands of years.

Introduction

The earliest evidence of farming-based economies in South Asia has been traced back to Mehrgarh, Pakistan ~9 kya.[1,2] From there, farming and a settled way of life spread farther east, laying foundations for the later Indus Valley civilization (3300−1300 BCE). Climatic reconstruction and other studies suggest that the decline of the Indus Valley civilization in the Bronze Age was most likely driven by a long-term drought, which might have triggered a movement of its inhabitants eastward toward the Gangetic Plain in about 2300 BCE.[3–8]

Contemporary populations of this region vary in their rituals and display diverse ethnic backgrounds.[9–14] The eastern Indus Basin, part of the early Vedic India (c. 2000 to c. 600 BCE), comprises the historical Kurukshetra[15,16] (now a district in the Haryana state). It adjoins Northwest (NW) India, which is the homeland of various ethnic communities whose long-term occupation of the area has been described in many Vedic and Hindu scriptures.[17–22]

Previous genetic studies have revealed a higher West Eurasian affinity among Northwest Indian and Pakistani (PNWI) populations than among South and East Indians.[23–35] Furthermore, some recent ancient DNA (aDNA) studies have suggested that the major West Eurasian genetic contributions in South Asia derive from Neolithic Iranians and early Bronze Age steppe populations.[36,37] Other studies[38,39] suggest contributions from the Middle and Late Bronze Age steppe populations in South Asia, together with a Chalcolithic or Bronze Age Central Asian admixture scenario. Nevertheless, despite major breakthroughs in our ability to test models of the genetic history of populations with aDNA, the lack of genome-wide data from Northwest India (NWI) hinders our understanding of present-day genetic variation in the Indus Valley region.

To fill this gap, we have performed a genome-wide study of 45 samples from four NW Indian ethnic groups whose long-term presence in the Indus Valley region has been historically attested: their names—Ror, Gujjar, Jat, and Kamboj—are explicitly mentioned in ancient Vedic scriptures. In addition, we used previously published genomic data for 20 individuals from the Khatri population of Punjab.[40] From the newly sampled populations, we generated mtDNA (190 individuals) and Y chromosome (248 individuals) data. We contextualized our data with 1,984 modern and 661 ancient Eurasian genomes from published sources (Tables S1 and S2). We set out to assess the extent of genetic heterogeneity among PNWI populations with regard to distinct genetic ancestries, as well as the amount of more recent ancestry (haplotype) sharing within and among neighboring regions. Furthermore, we also investigated the relationships between the PNWI groups, a set of ancient West Eurasians, and recent aDNA sources from South Asia.[37–39,41–43]

[1]Estonian Biocentre, Institute of Genomics, University of Tartu, Tartu 51010, Estonia; [2]5 Ror Colony, behind Sector 7, Karnal, Haryana 132001, India; [3]Institute of Genetics and Cytology, National Academy of Sciences of Belarus, Minsk 220072, Belarus; [4]Department of Zoology, University of Oxford, South Parks Road, OX1 3PS Oxford, UK; [5]Max-Delbrueck Centre for Molecular Medicine in the Helmholtz Association, Berlin-Buch 13125, Germany; [6]Applied Molecular Biology Laboratory, School of Life Sciences, Jawaharlal Nehru University, New Delhi 110029, India; [7]Birbal Sahni Institute of Palaeosciences, Lucknow 226007, India; [8]APE Lab, Department of Biology, University of Padova, Padova 35131, Italy; [9]Department of Human Genetics, KU Leuven, Leuven 3000, Belgium; [10]Cytogenetics Laboratory, Department of Zoology, Banaras Hindu University, Varanasi 221005, India

*Correspondence: ajaipathak@gmail.com (A.K.P.), gyaneshwer.chaubey@bhu.ac.in (G.C.)

https://doi.org/10.1016/j.ajhg.2018.10.022.

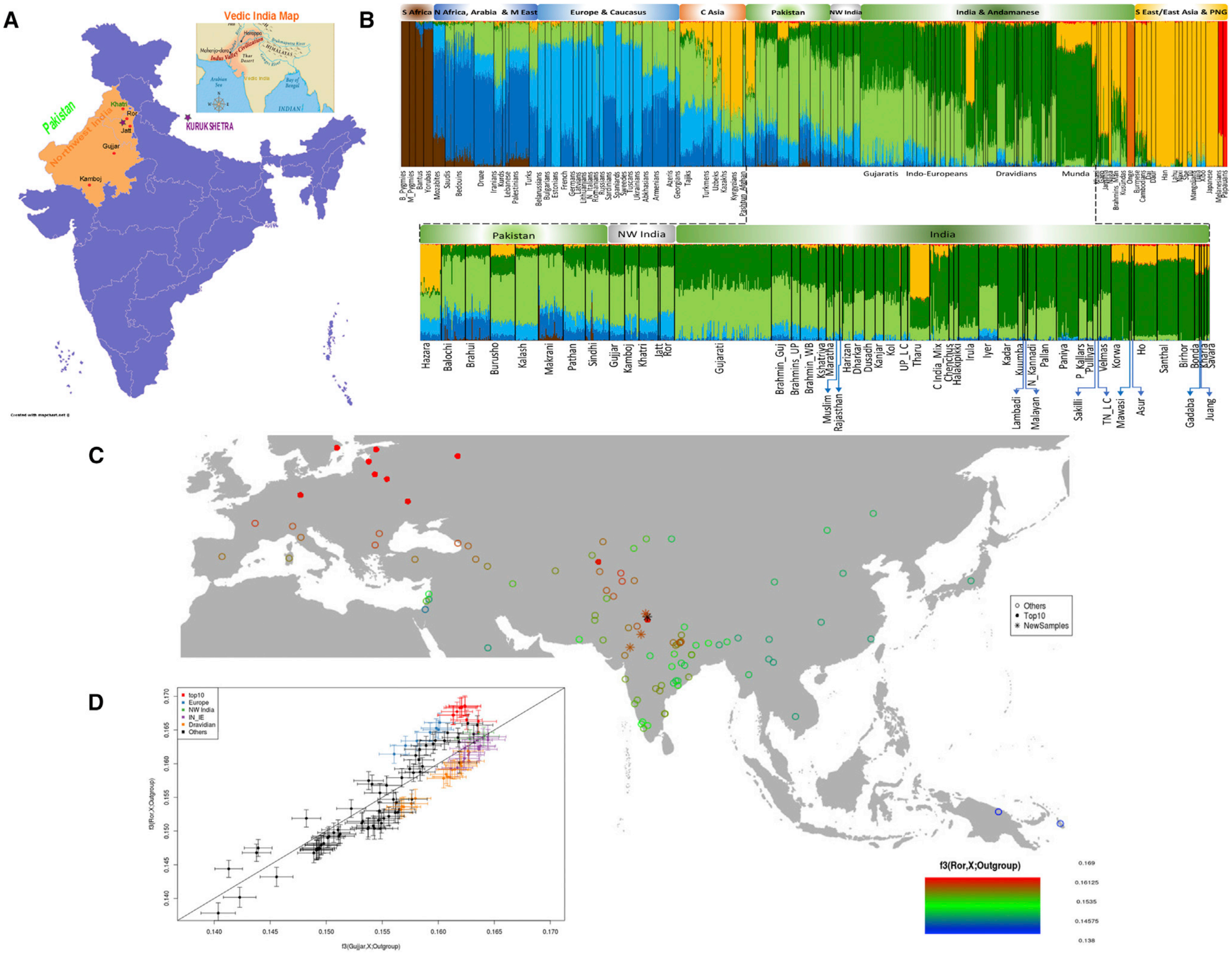


Figure 1. Sampling Locations, ADMIXTURE, and Shared Drift in Northwest India

(A) The geographic distribution and sampling locations of newly reported modern samples from Northwest India. An inset shows a map of Vedic India. Dots denote the samples studied, and green color indicates samples from published literature.

(B) Results of ADMIXTURE analysis at K8 ancestral components with global populations. The populations are ordered geographically in a bar plot. The genetic structure of new Northwest Indians is shown with a zoom-in on South Asia. The abbreviations are as follows: S Africa, sub-Saharan Africa; N Africa, North Africa; C Asia, Central Asia; NW India, Northwest India; C India_Mix, Central India Mix (Gond individuals together with one individual each from Bhunjia and Bengali); S East Asia, Southeast Asia; PNG, Papua New Guinea; Munda, Indian Austroasiatic speakers; UP_LC, Uttar Pradesh low-caste groups; TN_LC, Tamilnadu low-caste groups; N_Kannadi, North Kannadi; P_Kallars, Piramalai Kallars.

(C) Outgroup f_3 (Ror, X; Yoruba) gradient map, showing the affinity of the Ror to Eurasian populations. Red color indicates populations that have a high affinity with the Ror, green indicates groups that have a medium affinity, and blue indicates groups that have the least affinity with the Ror. The filled red circle shows the top 10 populations that share the highest amount of drift with the Ror. The star indicates the newly sampled population group, and the black star refers to the location of the Ror population.

(D) A scatterplot for outgroup f_3 (Ror, X; Yoruba) versus outgroup f_3 (Gujjar, X; Yoruba) plots the relative shared drift of West Eurasian populations with the Ror against the drift shared with the Gujjar. The top 10 populations sharing the most drift with the Ror are in red, and they are mostly from Europe. Other populations include Indian Austroasiatic, Central Asian, Pakistani, and Middle Eastern population groups. Abbreviations are as follows: NW India, Northwest India; IN_IE, Indian Indo-Europeans; and Dravidians, Indian Dravidians. Error bars represent jack-knife standard errors.

Material and Methods

Sampling

Blood or saliva samples were collected from 254 individuals residing in NWI, mainly from the Haryana and Rajasthan states. Sampling of the Ror population was carried out within a 100 km radius from the historical Kurukshetra. Other sampled populations also come from an area within 100–400 km of Kurukshetra (Figure 1A). The presence of the Kamboj, Gujjar, Ror, and Jat populations in the region dates back to the early historical and prehistorical period.[44–49] They represent different occupational caste populations who practiced agriculture and pastoralism. The fifth group we studied, the Khatri, is one of the few in the area with roots as a merchant community[50] (Supplemental Material and Methods). All subjects, who voluntarily participated in the study, were healthy adults who were selected through interviews carefully designed so that unrelated individuals would be chosen. Informed consent containing a signature and a left-thumb impression was collected from each participant. The project was carried out in accordance with the guidelines approved by the Ethical Committees of the University of Tartu, Estonia and the Banaras Hindu University (BHU), India.

Genotyping and Quality Control

DNA was purified from either whole blood or saliva cells via the standard phenol and chloroform extraction procedure.[51]

We genotyped 45 samples, including 15 Ror and 1 Jat from Haryana, and 15 Gujjar and 14 Kamboj from Rajasthan (Figure 1A and Table S1) with the Illumina HumanOmniExpress array for 730K SNPs as per the manufacturer's specifications. We analyzed the newly generated data together with similar data, from previously published sources, for 1,984 modern individuals across the globe[29,40,52–63] (Table S1). To evaluate the genetic affinities of PNWI groups with ancient source populations, we merged the modern dataset with 661 ancient genomes, mainly from West Eurasia and South Asia, which are geographically and temporally relevant in the context of the West Eurasian contribution to modern South Asians (Table S2).[37–39,41–43] In addition, for mtDNA coding and control region polymorphisms, we genotyped 190 individuals from NWI and assigned haplogroups according to the phylotree mtDNA tree Build 17 (Table S3). For Y chromosome genotyping, we genotyped 248 NWI samples by using either sequencing or PCR-RFLP to identify 37 binary haplogroup-informative Y chromosome markers and classified them into the respective haplogroups (Table S4).

For autosomal analyses, we processed the genome-wide SNP dataset by using PLINK v1.9.[64] We included only SNPs on the 22 autosomal chromosomes with minor allele frequency $> 1\%$ and removed all SNPs and samples with $>3\%$ missing data. One individual from each first- and second-degree relative pair detected with KING[65] was removed at random.

mtDNA and Y Chromosome Data Analysis

To explore the relationships of population groups, we performed principal-component analysis (PCA) on the matrix of haplogroup frequencies by using *prcomp* in R (Figure S1). We limited the populations to the geographical range surrounding PNWI groups and removed outliers from a zoomed landscape. The sample sets include earlier-published data from literature.[25,66–68]

Genome-wide SNP Data Analyses

We calculated mean pairwise F_{ST} values between PNWI groups and the regional population groups of West Eurasia (Table S5) by using the approach of Weir and Cockerham.[69,70] The Jat sample was not included in the F_{ST} analysis because there was only one sample from the population.

We carried out PC analysis by using the *smartpca* software (with default settings) implemented in the EIGENSOFT package[71] to capture genetic variability described by the first five principal components (PCs); the two most informative are discussed in the text (Figure S2).

For PCA with merged aDNA data, we projected relevant ancient samples on the PCA space and applied default parameters (with the added options of lsqproject: YES, numoutlier: 0, and autoshrink: YES) (Figure S3). We used two population sets as projection scaffolds; the first included present-day Eurasians, and the other included present-day South Asians.

We used the model-based clustering algorithm ADMIXTURE[72] to infer genomic ancestral components in PNWI in a global context. In the final settings, calculations for each of the tested ancestral clusters ($K = 2$ to $K = 15$) were repeated 25 times. The lowest cross-validation error parameter was observed at $K = 12$; however, we didn't observe any significant difference of cross-validation above $K = 8$ (Figure S4). Because both PCA[71] and structure-like analyses[72] might be affected by background linkage disequilibrium (LD), we thinned the marker set by pruning out SNPs in strong LD (pairwise genotypic correlation $r^2 > 0.4$) in a window of 200 SNPs (the window slid by 25 SNPs at a time).

We calculated f statistics by using the ADMIXTOOLS[73] programs *qp3Pop* and *qpDstat* with f4mode: YES. To investigate derived allele sharing between PNWI groups and modern or ancient Eurasian populations, we computed outgroup f_3 statistics[73] in the form of (Pop1, X; Yoruba), where Pop1 is a PNWI group or aDNA, X is a South Asian or West Eurasian population, and Yoruba is the outgroup (Figures S5–S7). We calculated D statistics[73] for various population combinations to assess gene flow among different modern populations and allele sharing between modern South Asians and published ancient sources (Figures S8 and S9 and Tables S8, S9, and S10).

We used *qpWave*[32,74] to test whether a set of "left" populations were related via N ancestry streams to a set of "right" populations; then we used *qpAdm*[42] to estimate ancestry proportions in a test population (PNWI) originating from a mixture of N "reference" populations[37–39] by exploiting shared genetic drift with a set of "outgroup" populations[37–39] (Table S16). The populations we included in a model that compared plausible "reference" populations were Early Bronze Age Yamnaya and Middle to Late Bronze Age from the Steppe region, as well as Neolithic farmers from Iran (Iran_N) and Onge. In order to differentiate between the Early Bronze Age Yamnaya (Steppe_EMBA) and Middle to Late Bronze Age (Steppe_MLBA) groups, we used the Neolithic Anatolian farmers (Anatolia_N) in addition to other outgroups[37] (Table S16) because Steppe_MLBA populations carry an Anatolian/European agriculturist component, but Steppe_EMBA populations do not.[39]

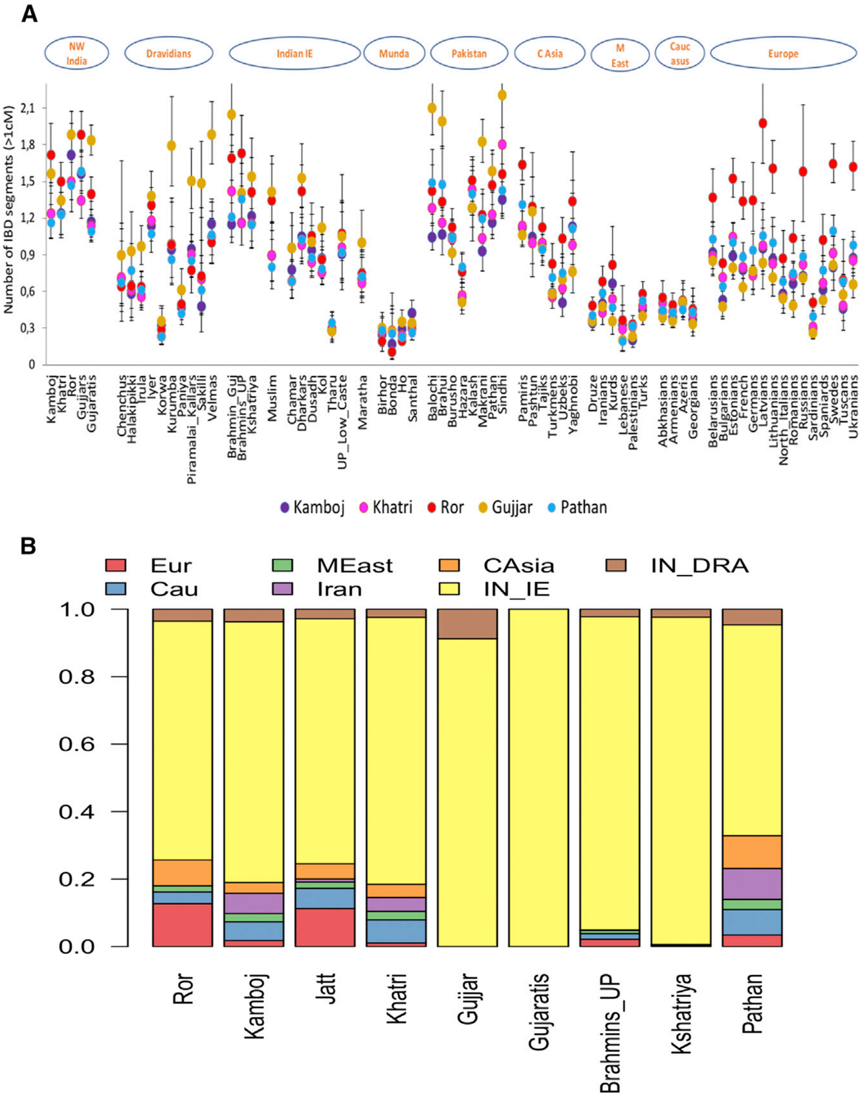
We applied ALDER[75] to compute a weighted LD statistic and to infer the date of admixture on the basis of exponential decay of linkage disequilibrium and were thus able to approximate the time of admixture between NWI and their neighboring regional ethnic groups. We used contemporary West Eurasian and South Asian populations as putative admixing source populations (Figure S10 and Table S18).

We constructed a maximum likelihood (ML) tree for a set of global populations by using *TreeMix* v.1.12[76] in order to place PNWI in a global context (Figure S11). We analyzed runs of homozygosity (RoHs) by using PLINK v.1.9[64] to investigate the parental relatedness among PNWI populations (Figure S12). RoHs were defined as a minimum of 50 consecutive SNPs in three different window sizes (1,000, 2,500, and 5,000 kb)—such that adjacent regions were fewer than 1,000 kb apart and the intraregional density of SNP coverage was no more than 1 SNP per 50 kb—allowing one heterozygous and five missing calls per window.[77,78] Because the total length and number of RoH segments varied considerably, we calculated the mean for each population.

Figure 2. IBD Sharing and Ancestry Profile of PNWI Populations

(A) Average number of IBD segments per pair of individuals for each Northwest Indian and Pathan ethnic group from Pakistan. Error bars represent 95% confidence intervals.

(B) Population-based ancestry estimates for PNWI and neighboring groups from the North Indian Gangetic Plain were inferred by CHROMOPAINTER via an NNLS-based analysis. NWI and Pakistani populations were excluded from the donor groups. Abbreviations are as follows: IN_IE, Indian Indo-Europeans; Pak, Pakistan; Eur, Europe; M East, Middle East; Cau, Caucasus; and IN_DRA, Indian Dravidians.



We used the refined IBD algorithm implemented in BEAGLE v4.0[79,80] to detect IBD segments that were shared by PNWI populations and a set of Eurasian populations. Refined IBD was run with default settings, and IBD segments longer than 1 cM were analyzed. IBD sharing was estimated as the average number of IBD segments per pair of individuals for each PNWI population (Figure 2A). 95% confidence intervals (CIs) for the average number of IBD segments were calculated with respect to Gujaratis and according to Kushniarevich et al..[62]

To perform chromosome painting, we used CHROMOPAINTER[81] on data phased with SHAPEIT v2.[82] We set up the -n and -M parameters after running the software's EM option on a small subset of the populations and five randomly selected chromosomes (3, 7, 10, 17, and 22), as described in Montinaro et al.[83] The estimated values for the two parameters were n = 526.701 and M = 0.00046. Recent ancestry sharing between closely related groups can hide distant relationships in CHROMOPAINTER analysis. We therefore performed two different analyses to take a more balanced approach:[84] the first one excluded NWI and Pakistani groups from donors (Figure 2B), and the second one kept all samples as donors (Figures S13 and S14A). The median number of SNPs for the inferred chunks was ~11 for both analyses (means = 13.56 and 12.79; 95% CI = 10.00–30.77 and 9.85–27.07, respectively).

We used a non-negative least-squares (NNLS)[85–87] "ancestry profile" method described in previous studies[83,88] to compare the copying vectors of the PNWI and North Indian Indo-European (NI_IE) populations. We modeled each target PNWI and NI_IE population as a mixture of haplotypes that best fit the painting profile of different donor populations. This approach not only accounts for deviation in sample size across donor groups but also considers the fact that human population groups are inherently related and thus that most haplotypes are shared. However, if true donor groups are not included as a result of a sampling issue, this method is likely to choose the "closest" among available sampled groups. Thus, groups recognized via this approach should be considered as the most related populations among the available ones. We have also calculated standard errors across the NNLS results by applying the jack-knife method (Table S19).

Results

Northwest Indian Populations in the Context of Other South Asians and West Eurasians

We first visualized the genetic structure of populations from NWI in the context of other South Asian and Eurasian populations by using PCA, F_{ST}, and ADMIXTURE analyses (Figures S2 and 1B and Table S5). In PCA, the NWI populations were placed between Indo-European (IE) populations from Pakistan and North India, and they fell within the North-South gradient that differentiates Europe from South Asia in principal component 2 (PC2) (Figure S2A). A similar pattern was observed in F_{ST} (Figure S2B) and ADMIXTURE (Figure 1B) results. Among South Asian populations, we detected consistently lower F_{ST} values between NWI and Pakistani groups (compared to all groups, the Ror are closer to the Pathan, and the Khatri are closer to the Sindhi) than between NWI and their North Indian neighbors (Figure S2B and Table S5). These observations are supported by PCA (Figure S2A inset) and by the presence of a significantly higher proportion (Wilcoxon test p value < 0.05) of the European-like light-blue component for the Ror and Jat, akin to the Pathan and Kalash, in ADMIXTURE (Figure 1B). Therefore, in contrast with uniparental DNA,[26,27] our autosomal data suggest a close genetic affinity between populations currently residing east and west of the Indus.

Demographic PNWI History Based on Allele Frequency and Haplotype-Sharing Analyses

Outgroup f_3 analysis in the form of (PNWI, X; Yoruba) showed that the Ror (and Jat) have distinct, high genetic similarity to modern Europeans (Figures 1C, 1D, and S5), far higher than the similarity observed in other NWI populations, such as the Gujjar (Figures 1D and S5). Among an extended set of South Asians, this pattern was repeated only in the Pathan population from Pakistan (Figure S5). This observation was further confirmed by D statistics, wherein the Pathan and Kalash share a higher proportion of alleles with the Ror than with any other group from NWI and NI_IE (Tables S7 and S8). *TreeMix* results (Figure S11) also indicate that the Kalash and Ror share the same branch.

Specifying various modern populations from South Asia, West Eurasia, and aDNA groups as likely sources, we used three population tests (f_3 statistics) to explore putative admixture patterns[73] in PNWI. We report only source combinations with a negative f_3 statistics value (Z score < -3) (Table S6).

We further investigated the Ror group's high affinity with modern Europeans at the haplotype-sharing level by performing identity by descent (IBD) (Figure 2A), CHROMOPAINTER (Figure S13), and NNLS (Figure 2B) analyses.

Refined IBD analysis highlights the general trend whereby the sharing of IBD segments declines as one moves along the cline from PNWI and NI_IE toward Dravidian and Indian Austroasiatic (IN_AA) groups (Figure 2A). Strikingly, among all PNWI groups studied, the Ror demonstrate the highest number of IBD segments shared with Europeans and Central Asians, whereas the Gujjar share a higher number of IBD segments with local Indian Indo-Europeans and Dravidians than do other PNWI groups (Figure 2A).

In CHROMOPAINTER analysis, as expected, the Ror (and Jat) exhibited a significantly higher number of chunks received from Europeans than do other NWI populations studied (t test, p value < 0.01). The excess sharing between the Ror and Europeans was made evident by the NNLS ancestry-profiling method, which we used to report the ancestry proportions of seven regional groups. Furthermore, the same analysis supports our previous observation suggesting a high degree of heterogeneity among PNWI groups. Populations such as the Ror and Jat possess more European and less Indian Indo-European (IN_IE) ancestry than other PNWI groups; however, they differ from each other in their Central Asian and Caucasus ancestry (Figure 2B). Interestingly, relative to other groups of NWI, the Khatri have a higher proportion of Caucasus ancestry, along with substantial Central Asian and Middle Eastern ancestry. Conversely, the Gujjar stand out as having the more ancestry from IN_IE groups and Dravidians than other PNWI groups do. To test whether population inbreeding could be responsible for the observed patterns, we analyzed RoHs in the genomes of PNWI groups, along with those of other neighboring Indian and West Eurasian populations. The Ror showed the smallest average number of RoHs (Figure S12), suggesting a higher effective population size (rather than inbreeding) or higher level of gene flow from other groups.

Furthermore, a putative West-South Eurasian admixture date for the Ror ~50 generations (~1,500 years) back, inferred by ALDER (Table S18), is corroborated by the lack of documented recent contacts between European populations and the Ror. This rules out any major impact of the recent colonial regime in India on the Ror population. Thus, the observed excess of a West Eurasian genetic component in the Ror group is most likely due to ancient migrations in the region.

Ancient Contributions to the Genome of Modern PNWI Populations

West Eurasian ancestry has been described as a composition of four main ancient components: Eastern European hunter-gatherers (EHG), Caucasus hunter-gatherers (CHG), Iran_N, and Anatolia_N.[36,41–43] However, EHG and CHG together are often associated in an ancestry from the steppe belt,[37] the Steppe_EMBA, which has been suggested as a major source of ancient admixture during Bronze Age population movements in West Eurasia and South Asia.[42,43] Recent aDNA studies, through the sampling of surrounding regions and analysis of the first samples from South Asia, have contributed significantly to our understanding of ancient population dynamics in South Asia. The new sampling mainly covers Neolithic (IranTuran_N) and Bronze Age (IranTuran_BA) individuals from the eastern Iran-Turkmenistan region, hunter-gatherers from West Siberia forest zone (WestSiberia_HG), Copper Age individuals from Botai, Kazakhstan (Botai), and further Steppe_MLBA individuals. Moreover, the new sampling includes Chalcolithic Namazga (Namazga_CA) and Iron Age (Turkmenistan_IA) individuals from Turkmenistan; Bactria-Margiana Archaeological Complex (BMAC) individuals and Bronze Age outliers from BMAC and eastern Iran (Indus_Diaspora [best known as Indus_Periphery]); and the first two ancient groups from South Asia, Iron Age or prehistorical samples (SPGT) and early historical period samples (SouthAsia_H) from the Swat Valley, Pakistan (Table S2).[38,39]

We first performed PCA to see the position of our PNWI group relative to these relevant ancient samples. We found that the NWI groups clustered near the Pakistani groups, close to the newly extracted proximal (temporally and geographically close) ancient sources (the Namazga_CA, Indus_Periphery, BMAC, SPGT, and SouthAsia_H individuals) and to other distal (temporally and geographically distant) ancient sources (Figure S3). We also observed a tight clustering of PNWI groups with the first ancient South Asian sources from the Swat Valley (the prehistorical SPGT and early historical SouthAsia_H individuals) and the Bronze Age outliers from BMAC region (the Indus_Periphery), who supposedly had a close connection with the ancient Indus Valley people as a result of their temporal and geographic vicinity (Figure S3).

We then used f_3 and D statistics to assess ancient West Eurasian contributions to modern PNWI populations (Figures S6–S9 and Tables S9 and S10). Analysis revealed that the Ror display more genetic components related to EHG, Anatolia_N, CHG, Steppe_EMBA, and Steppe_MLBA than any other South Asian population, as well as a higher affinity with SPGT, SouthAsia_H, and BMAC than other PNWI groups. At the same time, the affinity that the Ror exhibit with Iran_N, Namazga_CA, and Indus_Periphery is identical to that exhibited by their immediate geographic neighbors (Figures S6B, S7, S8B, and S9 and Tables S9 and S10). Higher West Eurasian ancestry in the genomes of modern Ror people could be due to ancient or recent admixture with sources west or north of the Indus Valley. The excess of EHG ancestry in the Ror population, compared to modern Iranians (Figure S8D), seems to rule out admixture with Iranian sources, hence pointing to a Central Asian or Steppe-related population as the most likely West Eurasian source (Table S6) in the Ror population. The Ror are also distinguished as the only South Asian group that is significantly closer to Neolithic Anatolians than to Neolithic Iranians (Z score $> +3$) (Figures S6B and S8B and Table S9). However, because of a lack of support from *qpAdm,* we present this as a tentative result.

Furthermore, we explored the allele sharing of PNWI groups relative to a set of two ancient sources by using D statistics in the form of D (pop1, Yoruba; pop2, pop3), where Yoruba is the outgroup, pop1 is a South Asian population, pop2 is an aDNA source, and pop3 is another aDNA source or a modern Dravidian population. These D tests revealed a general trend among South Asian populations of higher affinity (Z score $> +3$) with Steppe_EMBA than with either Steppe_MLBA or Chalcolithic Namazga_CA (Table S9). Interestingly, we observed that PNWI groups exhibited a trend of equal allele sharing when their affinity to available ancient sources from the Copper Age to Middle-Late Bronze Age Central Asia was compared, except for a visible closeness of the Ror, Jat, Kalash, and Pathan groups to Steppe_MLBA rather than to Indus_Periphery (or Indus_Diaspora) people (Z score < -3) (Table S9). In contrast, NI_IE resembled Dravidians in that the group had a higher affinity with Indus_Periphery rather than with Steppe_MLBA, Namazga_CA, and BMAC. However, by the Iron Age or prehistorical time, the scenario changed; NI_IE caste groups, similarly to PNWI group, share an equal number of alleles with the prehistorical SPGT and Indus_Periphery groups, whereas NI_IE tribes, Dravidians, and IN_AA people are closer to the Indus_Periphery group (Z score < -3) (Table S9). Finally, we compared the affinity between South Asians and both the Iron Age SPGT or early historical SouthAsia_H and modern Dravidian (Paniya) individuals separately in respective D statistics; we observed that PNWI and NI_IE castes are closer to SPGT and SouthAsia_H than to the Paniya (Z score $> +3$). This is unlike the IN_AA and NI_IE tribal people, who show a clear affinity with the Paniya (Z score < -3) (Table S9).

A previous ancient-DNA study has suggested that the Iran_N and Steppe_EMBA groups are the best proxies for the ancient West Eurasian component in South Asians. The study also suggested that most South Asians can be modeled as a mixture of these two groups but also have Onge- and Han-related ancestries,[37] a method sometimes referred to as distal modeling. However, other more recent aDNA studies have suggested that the Steppe_MLBA, grouped together with other ancient sources such as the Onge and either Namazga_CA or Indus_Periphery, offer a better fit in proximal modeling than do the Steppe_EMBA.[38,39]

We used *qpAdm* to explore how the distal models (Iran_N, either Steppe_MLBA or Steppe_EMBA, and Onge) and proximal models (Namazga_CA or Indus_Periphery, Steppe_MLBA, and Onge) fit in the case of our studied PNWI groups (Tables S11–S16). We observed that a model with three source populations (Iran_N, Steppe_MLBA, and Onge) fits with the data for the majority of PNWI populations (p value > 0.05 and low standard errors of admixture proportion estimates) (Figure 3 and Table S11). The only exceptions were the Burusho, Hazara, Kalash, and ancient SPGT groups, who could not be modeled from these three sources. Similar to the early historical SouthAsia_H group the NWI and NI_IE groups have high proportions of Steppe_MLBA ancestry, but they have higher proportions than the Pakistanis (except Pathan) and the Dravidian groups do. The Ror and Jat peoples stand out for having the highest proportion of Steppe_MLBA ancestry (~63%). The proportion of Steppe ancestry in the Ror is similar to that observed in present-day Northern Europeans.[42] We also observed that when we applied the "Pearson correlation" to the Steppe ancestry inferred in Europeans by Haak et al.,[42] the higher IBD sharing between the Ror people and Europeans was significantly and positively correlated with increasing Steppe ancestry in Europeans (Figure S15). Interestingly, when we used Steppe_EMBA instead of Steppe_MLBA in the distal model (Table S12), the Kalash and Iron Age Swat Valley SPGT data offered a good fit with the model, indicating a plausible Yamnaya-like impact in the Early-Middle Bronze Age; the effects of this impact might have persisted in some South Asian populations. In fact, we found that the model with Iran_N, Steppe_EMBA, and Onge works equally well for all modern and ancient South Asians. To

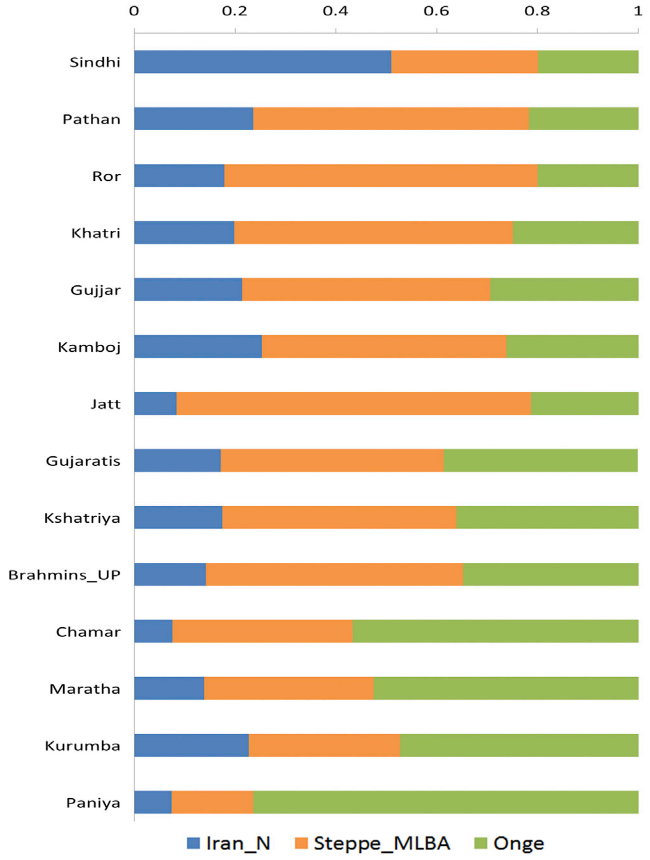


Figure 3. Proportions of Ancient Ancestry in South Asian Populations

qpAdm plot indicating proportions of ancestry made up of ancient sources (Iran_N, Steppe_MLBA, and Onge) among different South Asian populations.

test whether Steppe_MLBA or Steppe_EMBA fits better for modeling South Asians in the distal model, we added Neolithic Anatolians as a separate outgroup to the "right" list (Tables S13 and S16). This was motivated by the presence of a Neolithic Anatolian or European Early Neolithic component in Steppe_MLBA but not in Steppe_EMBA.[39] Our *qpAdm* results demonstrate that although most PNWI groups have significant Steppe_MLBA ancestry along with Steppe_EMBA, the NI_IE group from the Gangetic Plain and Dravidian South Asians show a significant Steppe_EMBA component but not a Steppe_MLBA component. However, prehistorical and early historical ancient South Asian individuals have a higher proportion of Steppe_MLBA than Steppe_EMBA.

To clarify the issue of plausible biases introduced by differences in the sample size of reference ancient groups, we replicated our analyses by using an equal number of individuals for both Steppe_EMBA and Steppe_MLBA, and we performed *D* statistics and *qpAdm* tests for cross-validation (Table S17). We observed no significant differences between the results.

Because various analyses (Figures 1B, S2A, and S9 and Tables S7 and S8) had highlighted either the Ror or Kalash peoples as having the highest proportion of ancient North Indian (ANI) ancestry combined with their terminal position on the South Asian cline (Figure S11), we used *D* statistics in the form of *D* (Yoruba, Test; Ror, Kalash) to further evaluate their relative affinity with worldwide populations (test). We observed that the Onge and other Indian populations with a high proportion of the ASI component share significantly more alleles (Z score > 4) with the Ror than with the Kalash but that the opposite is true for Georgians, who share significantly more alleles with the Kalash than with the Ror, indicating a higher proportion of the ANI component in the former (Table S7). These results suggest that the Ror might be used as a plausible alternative proxy for ANI in the demographic modeling of South Asians. The modeling might benefit from the reduced genetic drift in the Ror compared to the Kalash (Figure S11), although the Ror group harbors a small fraction of additional ASI component (~1%).

Y Chromosome and mtDNA Diversity in Northwest Indian Populations

In PC analysis of mtDNA and Y chromosomes, the NWI population fit on the broader North-South cline, consistent with the genome-wide analyses (Figures S1A and S1B). In mtDNA analysis, a substantial part (~37%–51%) of the maternal lineages in NWI is West Eurasian (R0, R2, R2′JT, T, HV, I, J, K, U3, U7, U9, and W) (Table S3), in agreement with the results of an earlier study.[25] The Y chromosome profiles of NWI revealed a high proportion (41%–76%) of South-Asian-specific lineages (C-M356, H-M69, R2-M124, and L-M11); the Gujjar stand out because they incorporate ~76% of these lineages. Other haplogroups (J2-M172 and R1a1-M17) are also present at a substantial frequency (20%–55%) (Table S4). Markedly, we observed that, among neighboring NWI groups, the Ror carry the highest proportion (about 58%) of South-Asian-specific maternal lineages (M18, M2, M3, M4, M5, M6, R5, and U2), and they carry more West Eurasian paternal lineages (J2 and Q) than other groups of the region. A high proportion of West Eurasian lineages in both uniparental loci is thus broadly consistent with the results based on autosomal loci.

Discussion

In this study, we have investigated the genetic relationships among contemporary populations of NWI in the context of neighboring populations from South Asia and West Eurasia, also considering the influence of ancient West Eurasian sources.

Genetic Ancestry Components of Northwest Indian and Pakistani Populations

Evidence from genome-wide genotype data (Figures 1B, 2, and S5 and Table S8) and uniparental markers (Figure 1) revealed Northwest Indians (east of the Indus) to be intermediate between Pakistanis (west of the Indus) and North

Indian Indo-European (NI_IE) speaking populations from the Gangetic Plain. Additionally, the genomic sharing between NWI populations and NI_IE populations from the Gangetic Plain, bolstered both by the results of analyses done by IBD (Figure 2A) and CHROMOPAINTER (Figures 2B and S13) and by their similar level of allele sharing with most ancient sources as shown by *D* statistics (discussed in the next section), establishes a noticeable genetic affinity between NWI and their contemporary neighbors on either side of the Indus Valley. This contrasts with earlier observations based on mtDNA and Y chromosome data.[27] Broadly speaking, these results could be compatible with archaeological evidence suggesting that people had high mobility within the region during the prehistoric and historical time. This mobility could include the migration of the Indus people toward the Gangetic Plain after the demise of the Indus Valley Civilization which is suggested by archaeological evidence.[3–7]

On the other hand, our data also reveal substantial intraregion heterogeneity (Figures 1, 2, S2, S5, and S13 and Tables S5, S7–S9, and S10). For instance, the Ror and Jat peoples, together with the Pakistani Pathan, share genetic ancestry pointing to their possible connection with Central Asians. The genetic relatedness of the Khatri and Sindhi may agree with both peoples' having been recognized as vital merchant communities of early modern India.[89] Among the populations of NWI, the elevated similarity of the Gujjar to local Indian populations and their lower affinity with West Eurasians may relate to their historically documented affinity with various extant South Asian ethnic communities.[90–93] High genetic differentiation among NW Indian populations suggests long-term population structure within the region.

Ancient West Eurasian Components in PNWI Populations

Previous claims of the widespread distribution of an ancient West Eurasian component in the Indian sub-continent, either through distal or proximal sources,[36–39] are well supported by our f_3, *D* statistics and *qpAdm* results (Figures 3 and S6–S9 and Tables S9, S10, and S11–S16). The observation that PNWI populations share more alleles with external sources from different time periods, including Mesolithic (EHG, CHG), Neolithic (Anatolia_N, Iran_N), Bronze (Steppe_EMBA, Steppe_MLBA), Copper (Namazga_CA, BMAC), and Iron Age (Turkmenistan_IA) groups than do other South Asian populations (Figures S7–S9 and Tables S9 and S10) can be explained by the geographic position of the Indus Valley as the gateway to the Indian sub-continent for any episode of gene flow from the west.

The higher affinity and admixture of PNWI populations with Neolithic Iranians and Anatolians (Figures S6C, S7B, and S8B and Tables S9 and S10), coupled with the substantial Middle Eastern component (dark blue, Figure 1B) and the significant influx of the Middle-East-related male lineage J2-M172 (Table S4) into the Indian sub-continent through the Northwest corridor,[94] might agree with earlier archaeological work that took place at Mehargarh and that suggested the plausible influence from the Zagros or Levant region on the first evident settled way of life in South Asia.[95,96]

A higher level of European ancestry in the Ror and Jat compared to other South Asians (Figures 1, 2, S2, S5, and S13 and Tables S5–S8) makes these two populations outliers within the broader Northwest South Asian landscape. This could be indicative of either a possible recent gene flow from a population related to Europe or to ancient West-Eurasian-related influx, which would agree with previous studies on adaptation, wherein the Ror and Jat have stood out for their high frequency of the lactase persistence allele (*LCT-13910T*) and the light-skin-color gene variant (*SLC24A5*).[70,97] We also report that, relative to other South Asians, the Ror group has high shared drift with the EHG and Steppe_EMBA groups, higher allele sharing with the Steppe_MLBA group, and higher affinity with the Iron Age (prehistorical) and early historical first South Asian ancient sources (Figures S6A, S6B, S7, S8A, S8D, and S9 and Tables S9 and S16). We find this indicative of multiple plausible influxes of Steppe-like ancestry into the Ror group, as well as their close connection with prehistorical to early historical South Asia.

The Ror display more affinity with Neolithic Anatolians than with Neolithic Iranians (Figures S6C and S8B and Table S8), whereas other South Asians in our dataset show almost equal allele sharing with both Neolithic aDNAs. Such an affinity might also explain the higher frequency of the light-skin-color variant (*SLC24A5* allele rs1426654) in the Ror because a higher frequency of this allele has also been found both in Neolithic Anatolians and CHG.[43,98] The Ror have an affinity with Anatolian Neolithic farmers and a closeness to the Pathan and Central Asians (Figures 2, S2, S6C, and S8B and Tables S5, S8, and S10). These facts, taken together with a gradient of affinity with ancient Steppe_MLBA or Steppe_EMBA, CHG, EHG, and Anatolia_N groups that decreases from Central Asia to the Ror, point to a possible contact with Central Asian and/or Steppe peoples that took place earlier than ~1,500 years ago, as suggested by the ALDER result (Table S18). *qpAdm* results consistently indicate a higher proportion of Steppe_MLBA ancestry in NWI populations than in other Indians. The Ror stand out in South Asia as the population with the highest proportion of Steppe ancestry (Figures 3 and S9 and Tables S10, and S11–S15), which could plausibly be linked to the finding that, among the South Asian groups, the Ror have the highest affinity with both the Neolithic Anatolians and northern Europeans. Interestingly, such a prominent West Eurasian link is not supported by mtDNA evidence in the Ror, perhaps hinting at a male-biased admixture scenario in the Ror from the Central Asia/Steppe region. Such a hypothesis is bolstered by the higher frequencies of the Y chromosomal haplogroups J2 and Q (Tables S3 and S4).

Among extant populations, both the Kalash and Ror groups stand out because they have the highest proportions of the ANI component, which can be modeled as a mixture of Iranian Neolithic and either Early-Middle (in case of the Kalash) or Middle-Late (in case of the Ror) Bronze Age Steppe ancestries. Although quantitatively the Kalash might have the highest ANI proportion, the Ror appear to be an important alternative to the Kalash as a proxy for ANI in demographic modeling in the absence of relevant ancient DNA data from India; this is due to diversity within the Steppe component as well as the high level of drift in the Kalash population.[99]

In summary, we demonstrate a higher proportion of genomic sharing between PNWI populations and ancient EHG and Steppe-related populations than we observe in other South Asians. We report that the Ror are the modern population that is closest to the first prehistorical and early historical South Asian ancient samples near the Indus Valley, and they also harbor the highest Steppe-related, EHG, and Neolithic Anatolian ancestry. However, compared to other adjoining groups, the Ror show less affinity with the Neolithic Iranians. The Ror population can plausibly be used as an alternative proxy for ANI in future demographic modeling of South Asian populations. Collectively, our results point out that the PNWI groups have high allele sharing with the region surrounding the ancient Indus Valley and that the PNWI region is an area of rich diversity in population dynamics and one where neighboring groups might harbor divergent genetic ancestries from multiple admixture events.

Accession Numbers

The data for the 45 sequences reported in this paper are available from the Gene Expression Omnibus of the National Centre for Biotechnology Information (GEO: GSE119653) and the data repository of the Estonian Biocentre (www.ebc.ee/free_data).

Supplemental Data

Supplemental Data include 15 figures, 19 tables, and Supplemental Material and Methods and can be found with this article online at https://doi.org/10.1016/j.ajhg.2018.10.022.

Acknowledgments

We thank the Ror, Gujjar, Kamboj, and Jat communities for their support of this study and all individual volunteers for donating their samples. R.V. thanks the Swedish Collegium for Advanced Studies for support during his sabbatical stay in Uppsala. We thank Lehti Saag, Bayazit Yunusbayev, Hovhannes Sahakyan, Jose Rodrigo Flores Espinoza, and Erwan Pennarun for useful discussions and assistance. We thank Tuuli Reisberg for her assistance in genotype data curation. We also thank Mari Järve for language editing. All data analyses were performed at the High-Performance Computer Centre of the University of Tartu, Estonia (http://www.hpc.ut.ee). Support was provided by the European Union through the European Regional Development Fund Centre of Excellence for Genomics and Translational Medicine (project no. 2014–2020.4.01.15-0012) to the EBC-IG and by the Estonian Institutional Research grant IUT24-1 (to A.K.P., R.V., M. Metspalu, S.R., E.M., and T.K.). M. Metspalu was supported by Estonian Research Council grant PRG 243. M. Mondal was supported by the European Union through the European Regional Development Fund (project no. 2014-2020.4.01.16-0030). G.C. is supported by National Geographic Explorer grant HJ3-182R-18, A.K. was supported by Estonian Personal grant PUT 1339, and L.P. was supported by the European Union through the European Regional Development Fund (project no. 2014-2020.4.01.16-0024, MOBTT53). A.K.P. was supported by the European Social Fund's Doctoral Studies and Internationalization Programme DoRa. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Declaration of Interests

The authors declare no competing interests.

Received: July 20, 2018
Accepted: October 25, 2018
Published: December 6, 2018

Web Resources

The 1000 Genomes Project, http://www.internationalgenome.org/home

HapMap3, https://www.sanger.ac.uk/resources/downloads/human/hapmap3.html

PhyloTree, http://www.phylotree.org/

References

1. Jarrige, J.-F. (1981). Economy and society in the Early Chalcolithic/Bronze Age of Baluchistan: New perspectives from recent excavations at Mehrgarh. South Asian Archaeol. *1979*, 93–114.
2. Costantini, L. (1984). The Beginning of Agriculture in the Kachi Plain: The Evidence of Mehrgarh. In South Asian Archaeology 1981, B. Allchin, ed. (Cambridge University Press), pp. 29–33.
3. Misra, V.N. (2001). Prehistoric human colonization of India. J. Biosci. *26* (4, Suppl), 491–531.
4. Tripathi, J.K., Bock, B., Rajamani, V., and Eisenhauer, A. (2004). Is Rhiver Ghaggar, Saraswati? Geochemical constraints. Curr. Sci. *87*, 1141–1144.
5. Gupta, A.K., Anderson, D.M., Pandey, D.N., and Singhvi, A.K. (2006). Adaptation and human migration, and evidence of agriculture coincident with changes in the Indian summer monsoon during the Holocene. Curr. Sci. *90*, 1082–1090.
6. Madella, M., and Fuller, D.Q. (2006). Palaeoecology and the Harappan civilisation of South Asia: A reconsideration. Quat. Sci. Rev. *25*, 1283–1301. https://doi.org/10.1016/j.quascirev.2005.10.012.
7. Brooke, J.L. (2014). Climate change and the course of global history: A rough journey (Cambridge University Press).
8. Dutt, S., Gupta, A.K., Wünnemann, B., and Yan, D. (2018). 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.

9. Thapar, B.K. (1979). A Harappan Metropolis beyond the Indus Valley. In Ancient Cities of the Indus, G.L. Possehl, ed. (Carolina Academic Press), pp. 196–202.
10. Possehl, G.L. (1982). Harappan civilization: A contemporary perspective (Aris & Phillips).
11. Shaffer, J.G., and Lichtenstein, D.A. (1989). Ethnicity and change in the Indus valley cultural tradition. Old Probl. New Perspect. Archaeol. South Asia *2*, 117–126.
12. Mughal, M.R. (1990). The Decline of the Indus Civilization and the Late Harappan Period in the Indus Valley. Lahore Museum Journal *3*, 1–22.
13. Possehl, G.L. (1990). Revolution in the urban revolution: The Emergence of the Indus urbanization. Annu. Rev. Anthropol. *19*, 261–282.
14. Nath, A. (1998). Rakhigarhi: A Harappan metropolis in the Saraswati-Drishadvati divide. Puratattva *28*, 39–45.
15. Benedetti, G. (2014). The chronology of Puranic kings and Rigvedic rishis in comparison with the phases of the Sindhu-Sarasvati civilization. In Sindu-Sarasvati Civilization: New Perspectives, N. Rao, ed. (D.K. Printworld (P) Ltd.), pp. 220–246.
16. Kenoyer, J.M. (2006). Cultures and societies of the Indus tradition. In Historical Roots in the Making of 'the Aryan, R. Thapar, ed. (National Book Trust), pp. 21–49.
17. Chidbhavananda, S. (1965). The Bhagavad Gita: Original Stanzas (Tapovanam Publishing House).
18. Purana, B. (1971). Samvat 2037 (Gita Press. Gorakhpur).
19. Lal, P. (1980). Mahabharata of Vyasa (Asia Book Corporation of America).
20. O'Flaherty, W.D. (1981). The Rig Veda: An anthology. One hundred and eight hymns, selected, translated, and annotated (Penguin Books).
21. Dange, S.S. (1984). The Bhagavata Purana: Mytho-Social Study (Ajanta Publications).
22. Sitaramiah, V., and Sītārāmayya, V. (1972). Valmiki Ramayana (Sahitya Akademi).
23. Kivisild, T., Rootsi, S., Metspalu, M., Mastana, S., Kaldma, K., Parik, J., Metspalu, E., Adojaan, M., Tolk, H.-V., Stepanov, V., et al. (2003). The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. Am. J. Hum. Genet. *72*, 313–332.
24. Metspalu, M., Kivisild, T., Metspalu, E., Parik, J., Hudjashov, G., Kaldma, K., Serk, P., Karmin, M., Behar, D.M., Gilbert, M.T.P., et al. (2004). 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*, 26.
25. Quintana-Murci, L., Chaix, R., Wells, R.S., Behar, D.M., Sayar, H., Scozzari, R., Rengo, C., Al-Zahery, N., Semino, O., Santachiara-Benerecetti, A.S., et al. (2004). Where west meets east: The complex mtDNA landscape of the southwest and Central Asian corridor. Am. J. Hum. Genet. *74*, 827–845.
26. McElreavey, K., and Quintana-Murci, L. (2005). A population genetics perspective of the Indus Valley through uniparentally-inherited markers. Ann. Hum. Biol. *32*, 154–162.
27. Thangaraj, K., Naidu, B.P., Crivellaro, F., Tamang, R., Upadhyay, S., Sharma, V.K., Reddy, A.G., Walimbe, S.R., Chaubey, G., Kivisild, T., and Singh, L. (2010). The influence of natural barriers in shaping the genetic structure of Maharashtra populations. PLoS ONE *5*, e15283.
28. Xing, J., Watkins, W.S., Hu, Y., Huff, C.D., Sabo, A., Muzny, D.M., Bamshad, M.J., Gibbs, R.A., Jorde, L.B., and Yu, F. (2010). Genetic diversity in India and the inference of Eurasian population expansion. Genome Biol. *11*, R113.
29. Metspalu, M., Romero, I.G., Yunusbayev, B., Chaubey, G., Mallick, C.B., Hudjashov, G., Nelis, M., Mägi, R., Metspalu, E., Remm, M., et al. (2011). 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.
30. Sharma, G., Tamang, R., Chaudhary, R., Singh, V.K., Shah, A.M., Anugula, S., Rani, D.S., Reddy, A.G., Eaaswarkhanth, M., Chaubey, G., et al. (2012). Genetic affinities of the central Indian tribal populations. PLoS ONE *7*, e32546.
31. Gazi, N.N., Tamang, R., Singh, V.K., Ferdous, A., Pathak, A.K., Singh, M., Anugula, S., Veeraiah, P., Kadarkaraisamy, S., Yadav, B.K., et al. (2013). Genetic structure of Tibeto-Burman populations of Bangladesh: evaluating the gene flow along the sides of Bay-of-Bengal. PLoS ONE *8*, e75064.
32. Moorjani, P., Thangaraj, K., Patterson, N., Lipson, M., Loh, P.-R., Govindaraj, P., Berger, B., Reich, D., and Singh, L. (2013). Genetic evidence for recent population mixture in India. Am. J. Hum. Genet. *93*, 422–438.
33. Chaubey, G., Kadian, A., Bala, S., and Rao, V.R. (2015). Genetic affinity of the Bhil, Kol and Gond mentioned in epic ramayana. PLoS ONE *10*, e0127655.
34. Chaubey, G., Ayub, Q., Rai, N., Prakash, S., Mushrif-Tripathy, V., Mezzavilla, M., Pathak, A.K., Tamang, R., Firasat, S., Reidla, M., et al. (2017). "Like sugar in milk": Reconstructing the genetic history of the Parsi population. Genome Biol. *18*, 110.
35. Reich, D., Thangaraj, K., Patterson, N., Price, A.L., and Singh, L. (2009). Reconstructing Indian population history. Nature *461*, 489–494.
36. Broushaki, F., Thomas, M.G., Link, V., López, S., van Dorp, L., Kirsanow, K., Hofmanová, Z., Diekmann, Y., Cassidy, L.M., Díez-Del-Molino, D., et al. (2016). Early Neolithic genomes from the eastern Fertile Crescent. Science *353*, 499–503.
37. Lazaridis, I., Nadel, D., Rollefson, G., Merrett, D.C., Rohland, N., Mallick, S., Fernandes, D., Novak, M., Gamarra, B., Sirak, K., et al. (2016). Genomic insights into the origin of farming in the ancient Near East. Nature *536*, 419–424.
38. Damgaard, P. de B., Martiniano, R., Kamm, J., Moreno-Mayar, J.V., Kroonen, G., Peyrot, M., Barjamovic, G., Rasmussen, S., Zacho, C., Baimukhanov, N., et al. (2018). The first horse herders and the impact of early Bronze Age steppe expansions into Asia. Science *360*, 6396.
39. Narasimhan, V.M., Patterson, N.J., Moorjani, P., Lazaridis, I., Mark, L., Mallick, S., Rohland, N., Bernardos, R., Kim, A.M., Nakatsuka, N., et al. (2018). The genomic formation of South and Central Asia. bioRxiv. https://doi.org/10.1101/292581.
40. Basu, A., Sarkar-Roy, N., and Majumder, P.P. (2016). Genomic reconstruction of the history of extant populations of India reveals five distinct ancestral components and a complex structure. Proc. Natl. Acad. Sci. USA *113*, 1594–1599.
41. Allentoft, M.E., Sikora, M., Sjögren, K.-G., Rasmussen, S., Rasmussen, M., Stenderup, J., Damgaard, P.B., Schroeder, H., Ahlström, T., Vinner, L., et al. (2015). Population genomics of bronze age Eurasia. Nature *522*, 167–172.
42. Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., Brandt, G., Nordenfelt, S., Harney, E., Stewardson, K., et al. (2015). Massive migration from the steppe was a source for Indo-European languages in Europe. Nature *522*, 207–211.

43. Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S.A., Harney, E., Stewardson, K., Fernandes, D., Novak, M., et al. (2015). Genome-wide patterns of selection in 230 ancient Eurasians. Nature *528*, 499–503.
44. Davids, T.W.R. (1903). Buddhist India (Putnam).
45. Elliott, H.M. (1859). Memoirs on the History, Folk-Lore, and Distribution of the Races of the North Western Provinces of India (Lond. Trübner).
46. Smith, V.A. (1924). The Early History of India, from 600 BC to the Muhammadan Conquest including the Invasion of Alexander the Great (Clarendon Press).
47. Law, B.C. (1924). Ancient Mid-Indian Ksatriya Tribes*Volume I* (Thacker, Spink & Co.).
48. Raychaudhuri, H. (2006). Political history of ancient India: From the accession of Parikshit to the extinction of the Gupta dynasty (Cosmo Publications).
49. Mishra, K.C. (1987). Tribes in The Mahabharata: A socio-cultural study (National Publishing House).
50. Levi, S. (1999). The Indian merchant diaspora in early modern central Asia and Iran. Iran. Stud. *32*, 483–512.
51. Mathew, C.G.P. (1984). The isolation of high molecular weight eukaryotic DNA. Nucleic Acids (Springer), pp. 31–34.
52. Li, J.Z., Absher, D.M., Tang, H., Southwick, A.M., Casto, A.M., Ramachandran, S., Cann, H.M., Barsh, G.S., Feldman, M., Cavalli-Sforza, L.L., and Myers, R.M. (2008). Worldwide human relationships inferred from genome-wide patterns of variation. Science *319*, 1100–1104.
53. Behar, D.M., Yunusbayev, B., Metspalu, M., Metspalu, E., Rosset, S., Parik, J., Rootsi, S., Chaubey, G., Kutuev, I., Yudkovsky, G., et al. (2010). The genome-wide structure of the Jewish people. Nature *466*, 238–242.
54. Altshuler, D.M., Gibbs, R.A., Peltonen, L., Altshuler, D.M., Gibbs, R.A., Peltonen, L., Dermitzakis, E., Schaffner, S.F., Yu, F., Peltonen, L., et al.; International HapMap 3 Consortium (2010). Integrating common and rare genetic variation in diverse human populations. Nature *467*, 52–58.
55. Chaubey, G., Metspalu, M., Choi, Y., Mägi, R., Romero, I.G., Soares, P., van Oven, M., Behar, D.M., Rootsi, S., Hudjashov, G., et al. (2011). Population genetic structure in Indian Austroasiatic speakers: The role of landscape barriers and sex-specific admixture. Mol. Biol. Evol. *28*, 1013–1024.
56. Rasmussen, M., Guo, X., Wang, Y., Lohmueller, K.E., Rasmussen, S., Albrechtsen, A., Skotte, L., Lindgreen, S., Metspalu, M., Jombart, T., et al. (2011). An Aboriginal Australian genome reveals separate human dispersals into Asia. Science *334*, 94–98.
57. Yunusbayev, B., Metspalu, M., Järve, M., Kutuev, I., Rootsi, S., Metspalu, E., Behar, D.M., Varendi, K., Sahakyan, H., Khusainova, R., et al. (2012). The Caucasus as an asymmetric semipermeable barrier to ancient human migrations. Mol. Biol. Evol. *29*, 359–365.
58. Behar, D.M., Metspalu, M., Baran, Y., Kopelman, N.M., Yunusbayev, B., Gladstein, A., Tzur, S., Sahakyan, H., Bahmanimehr, A., Yepiskoposyan, L., et al. (2013). No evidence from genome-wide data of a Khazar origin for the Ashkenazi Jews. Hum. Biol. *85*, 859–900.
59. Di Cristofaro, J., Pennarun, E., Mazières, S., Myres, N.M., Lin, A.A., Temori, S.A., Metspalu, M., Metspalu, E., Witzel, M., King, R.J., et al. (2013). Afghan Hindu Kush: where Eurasian sub-continent gene flows converge. PLoS ONE *8*, e76748.
60. Fedorova, S.A., Reidla, M., Metspalu, E., Metspalu, M., Rootsi, S., Tambets, K., Trofimova, N., Zhadanov, S.I., Hooshiar Kashani, B., Olivieri, A., et al. (2013). Autosomal and uniparental portraits of the native populations of Sakha (Yakutia): Implications for the peopling of Northeast Eurasia. BMC Evol. Biol. *13*, 127.
61. Raghavan, M., Skoglund, P., Graf, K.E., Metspalu, M., Albrechtsen, A., Moltke, I., Rasmussen, S., Stafford, T.W., Jr., Orlando, L., Metspalu, E., et al. (2014). Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. Nature *505*, 87–91.
62. Kushniarevich, A., Utevska, O., Chuhryaeva, M., Agdzhoyan, A., Dibirova, K., Uktveryte, I., Möls, M., Mulahasanovic, L., Pshenichnov, A., Frolova, S., et al.; Genographic Consortium (2015). Genetic heritage of the Balto-Slavic speaking populations: A synthesis of autosomal, mitochondrial and Y-chromosomal data. PLoS ONE *10*, e0135820.
63. Yunusbayev, B., Metspalu, M., Metspalu, E., Valeev, A., Litvinov, S., Valiev, R., Akhmetova, V., Balanovska, E., Balanovsky, O., Turdikulova, S., et al. (2015). The genetic legacy of the expansion of Turkic-speaking nomads across Eurasia. PLoS Genet. *11*, e1005068.
64. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. Gigascience *4*, 7.
65. Manichaikul, A., Mychaleckyj, J.C., Rich, S.S., Daly, K., Sale, M., and Chen, W.-M. (2010). Robust relationship inference in genome-wide association studies. Bioinformatics *26*, 2867–2873.
66. Qamar, R., Ayub, Q., Mohyuddin, A., Helgason, A., Mazhar, K., Mansoor, A., Zerjal, T., Tyler-Smith, C., and Mehdi, S.Q. (2002). Y-chromosomal DNA variation in Pakistan. Am. J. Hum. Genet. *70*, 1107–1124.
67. Grugni, V., Battaglia, V., Hooshiar Kashani, B., Parolo, S., Al-Zahery, N., Achilli, A., Olivieri, A., Gandini, F., Houshmand, M., Sanati, M.H., et al. (2012). Ancient migratory events in the Middle East: New clues from the Y-chromosome variation of modern Iranians. PLoS ONE *7*, e41252.
68. Derenko, M., Malyarchuk, B., Bahmanimehr, A., Denisova, G., Perkova, M., Farjadian, S., and Yepiskoposyan, L. (2013). Complete mitochondrial DNA diversity in Iranians. PLoS ONE *8*, e80673.
69. Weir, B.S., and Cockerham, C.C. (1984). Estimating F-statistics for the analysis of population structure. Evolution *38*, 1358–1370.
70. Gallego Romero, I., Basu Mallick, C., Liebert, A., Crivellaro, F., Chaubey, G., Itan, Y., Metspalu, M., Eaaswarkhanth, M., Pitchappan, R., Villems, R., et al. (2012). Herders of Indian and European cattle share their predominant allele for lactase persistence. Mol. Biol. Evol. *29*, 249–260.
71. Patterson, N., Price, A.L., and Reich, D. (2006). Population structure and eigenanalysis. PLoS Genet. *2*, e190.
72. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Res. *19*, 1655–1664.
73. Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., and Reich, D. (2012). Ancient admixture in human history. Genetics *192*, 1065–1093.
74. Reich, D., Patterson, N., Campbell, D., Tandon, A., Mazieres, S., Ray, N., Parra, M.V., Rojas, W., Duque, C., Mesa, N., et al. (2012). Reconstructing Native American population history. Nature *488*, 370–374.
75. Loh, P.-R., Lipson, M., Patterson, N., Moorjani, P., Pickrell, J.K., Reich, D., and Berger, B. (2013). Inferring admixture histories

of human populations using linkage disequilibrium. Genetics *193*, 1233–1254.

76. Pickrell, J.K., and Pritchard, J.K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet. *8*, e1002967.
77. McQuillan, R., Leutenegger, A.-L., Abdel-Rahman, R., Franklin, C.S., Pericic, M., Barac-Lauc, L., Smolej-Narancic, N., Janicijevic, B., Polasek, O., Tenesa, A., et al. (2008). Runs of homozygosity in European populations. Am. J. Hum. Genet. *83*, 359–372.
78. Joshi, P.K., Esko, T., Mattsson, H., Eklund, N., Gandin, I., Nutile, T., Jackson, A.U., Schurmann, C., Smith, A.V., Zhang, W., et al. (2015). Directional dominance on stature and cognition in diverse human populations. Nature *523*, 459–462.
79. Browning, S.R., and Browning, B.L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am. J. Hum. Genet. *81*, 1084–1097.
80. Browning, B.L., and Browning, S.R. (2013). Improving the accuracy and efficiency of identity-by-descent detection in population data. Genetics *194*, 459–471.
81. Lawson, D.J., Hellenthal, G., Myers, S., and Falush, D. (2012). Inference of population structure using dense haplotype data. PLoS Genet. *8*, e1002453.
82. Delaneau, O., Marchini, J., and Zagury, J.-F. (2011). A linear complexity phasing method for thousands of genomes. Nat. Methods *9*, 179–181.
83. Montinaro, F., Busby, G.B.J., Pascali, V.L., Myers, S., Hellenthal, G., and Capelli, C. (2015). Unravelling the hidden ancestry of American admixed populations. Nat. Commun. *6*, 6596.
84. van Dorp, L., Balding, D., Myers, S., Pagani, L., Tyler-Smith, C., Bekele, E., Tarekegn, A., Thomas, M.G., Bradman, N., and Hellenthal, G. (2015). Evidence for a common origin of blacksmiths and cultivators in the Ethiopian Ari within the last 4500 Years: Lessons for clustering-based inference. PLoS Genet. *11*, e1005397.
85. Lawson, C.L., and Hanson, R.J. (1995). Solving least squares problems (Philadelphia: SIAM).
86. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.
87. R Core Team (2013). R: A language and environment for statistical computing (R Foundation for Statistical Computing).
88. Leslie, S., Winney, B., Hellenthal, G., Davison, D., Boumertit, A., Day, T., Hutnik, K., Royrvik, E.C., Cunliffe, B., Lawson, D.J., et al.; Wellcome Trust Case Control Consortium 2; and International Multiple Sclerosis Genetics Consortium (2015). The fine-scale genetic structure of the British population. Nature *519*, 309–314.
89. Levi, S.C. (2002). The Indian Diaspora in Central Asia and Its Trade, 1550–1900 (Brill).
90. Bingley, A.H. (1978). Caste, tribes & culture of Rajputs (Ess Publications).
91. Ibbetson, D. (1987). Landmarks in Indian Anthropology: Punjab Castes, Races, Castes, and Tribes of the People of Punjab (Cosmo Publications).
92. Khatra, P.S., and Sharma, V. (1992). Socio-economic issues in the development of nomadic Gujjars. Indian J. Agric. Econ. *47*, 448–449.
93. Singh, D.E. (2012). Islamization in Modern South Asia: Deobandi Reform and the Gujjar Response (Walter de Gruyter).
94. Singh, S., Singh, A., Rajkumar, R., Sampath Kumar, K., Kadarkarai Samy, S., Nizamuddin, S., Singh, A., Ahmed Sheikh, S., Peddada, V., Khanna, V., et al. (2016). Dissecting the influence of Neolithic demic diffusion on Indian Y-chromosome pool through J2-M172 haplogroup. Sci. Rep. *6*, 19157.
95. Jarrige, C. (1995). Mehrgarh: Field reports 1974-1985, from Neolithic times to the Indus Civilization (Department of Culture and Tourism, Government of Sindh).
96. Gangal, K., Vahia, M.N., and Adhikari, R. (2010). Spatio-temporal analysis of the Indus urbanization. Curr. Sci. *98*, 846–852.
97. Basu Mallick, C., Iliescu, F.M., Möls, M., Hill, S., Tamang, R., Chaubey, G., Goto, R., Ho, S.Y., Gallego Romero, I., Crivellaro, F., et al. (2013). The light skin allele of SLC24A5 in South Asians and Europeans shares identity by descent. PLoS Genet. *9*, e1003912.
98. Jones, E.R., Gonzalez-Fortes, G., Connell, S., Siska, V., Eriksson, A., Martiniano, R., McLaughlin, R.L., Gallego Llorente, M., Cassidy, L.M., Gamba, C., et al. (2015). Upper Palaeolithic genomes reveal deep roots of modern Eurasians. Nat. Commun. *6*, 8912.
99. Ayub, Q., Mezzavilla, M., Pagani, L., Haber, M., Mohyuddin, A., Khaliq, S., Mehdi, S.Q., and Tyler-Smith, C. (2015). The Kalash genetic isolate: ancient divergence, drift, and selection. Am. J. Hum. Genet. *96*, 775–783.

The American Journal of Human Genetics, Volume *103*

Supplemental Data

The Genetic Ancestry of Modern Indus Valley Populations from Northwest India

Ajai K. Pathak, Anurag Kadian, Alena Kushniarevich, Francesco Montinaro, Mayukh Mondal, Linda Ongaro, Manvendra Singh, Pramod Kumar, Niraj Rai, Jüri Parik, Ene Metspalu, Siiri Rootsi, Luca Pagani, Toomas Kivisild, Mait Metspalu, Gyaneshwer Chaubey, and Richard Villems

Contents

Supplemental Text

Indus Valley Civilization and its antiquity

Alongside those developed in ancient Egypt and Mesopotamia, the Indus Valley Civilization was among the greatest early civilizations of the Old World during the early- to mid-third millennium BCE[1–3]. The Indus Valley Civilization extending across present-day Northwest Pakistan deep into Northwestern India had major settlements established on the plains between Indus and Ganges basin, with an urban-phase beginning ~4.6 kya BP[3,4]. This Civilization thrived in the Indus River basin [5,6] that flows through the whole Pakistan, and along a network of recurrent, monsoon-fed rivers that on one occasion flowed near Ghaggar-Hakra river, which is also known as Sarasvati (now dried up) in northwest India and eastern Pakistan[3,7,8].

Though so far, the least studied among the three, the emerging archaeological evidence suggests that the Indus Valley one (earlier often called narrowly as Harappan), might have been the most widespread encircling an area of ~1 million square km and the largest of the Bronze Age developed civilization[2,9]. During its peak, this Civilization possibly has had a population of about five million[10].

Archaeologists link the urbanism of the Indus Valley with the expansion of five huge settlements considered as cities, and several minor built-ups that are categorized by typical architectural elements and material culture[1,3,11]. The Indus Valley Civilization is also known as the Harappan Civilization after Harappa a place in the Punjab province of British India and now in Pakistan, where the first excavation had happened[12]. Mohenjo-daro was another main site of Indus Valley Civilization, discovered soon after the Harappa, whereas, other major urban centers were Dholavira, Ganeriwala, and Rakhigarhi[3].

Mehrgarh, a Neolithic site (~7000 BCE to c. 2500 BCE) located at the west of the Indus river valley, was considered to be one of the earliest center of the civilization with an indication of early farming and cattle herding in South Asia[13–16] and shift from foraging to settled agriculture and domestication[17]. Though a recent study links the spread of Neolithic farming into Northwest India from the Near East, they also found a decent signal for the indigenous domestication of barley, and zebu cattle at Mehrgarh similar to the humped Indian aurochs that is frequently depicted on Indus seals[15,18].

There are also archaeological traces of trade seals, decked with animals and mythical existences, demonstrating the Indus people's involvement in booming trade with faraway lands like Sumer in Mesopotamia[21]. Excavation of archaeological artifacts from Iraq, Central Asia, and Afghanistan e.g., the occurrence of Indus seals in Central Asia besides Iraq, and portable Indus beads, pottery across Iranian plateau, Oman and Mesopotamia, display the extended Indus trade networks[2,5].

The abrupt fall of the Indus Valley Civilization, which flourished from 2600 BCE until 1900 BCE, is highly debatable issue. Many theories like floods, tectonic shifts, droughts, and foreign invaders have been invoked to explain the remarkable sudden and entire disappearance of this thriving civilization[5,9,22–24]. Many scholars proposed that climate must have had an indirect effect on the Indus Valley[25,26] facilitating the drying-up of mighty river Sarasvati (Ghaghra Hakkar). It has been also proposed that these events triggered the migration of Indus people towards Gangetic plain [22,23], along with the theory of strategic local shifts in agriculture during extended droughts in 2200 BC by Madella and Fuller[27]. Nevertheless, none of the theories could solve this puzzle satisfactorily so far.

Studied Populations:

Ror

Ror is a Kshatriya community from Northwest India, especially from Haryana state that counts almost 750,000 in number all over India[28]. They have their roots primarily from an area, extending across the Gujarat – Rajasthan border[28]. Historically the Ror people were found concerted around Ror (Sukkur) in Sindh region of Pakistan[28] but, now they found mainly in Haryana, Rajasthan, Himachal Pradesh, western Uttar Pradesh and Uttaranchal in India. H. A. Rose describes the Ror as fine, stalwart men in his famous work *"A Glossary of the Tribes and Castes of Punjab and North-West Frontier Provinces"*[29]. A.C.L. Carlleyle also reports about the finding of a Ror warrior image at the site of Kaga Ror or Kagarol in the Archaeological Survey of India Report for the year 1871-72[30]. Ror Mahasabha claims that Sir Denzil Ibbetson classified Ror as one of the "Other dominant tribes" of United Punjab during the British rule in India[31]. In fact, the Ror people vary from other neighboring groups in their lifestyle like ethnic wears etc.

Conferring to the Indian Puranic tradition, Ror claim to belong to the solar race of *Ikshvaku Kul* and connect themselves to Lord *Sri Rama*, the hero of epic *Ramayana*.

Ror clans ruled from the Rori, also known by names of Roruka and Rorik since ancient times. It was the capital of Sindh for years and has been mentioned as of the most important Indian cities during seventh century B.C.[32]. During first few centuries and a couple of centuries prior to the Christian era, Ror have ruling seats of power in Gujarat, Rajasthan, Haryana, and Sindh; and afterward, in the 5th century AD, they consolidated their influence in the entire region from Afghanistan to Kannauj in India during the time of rai Dewaji[34,35].

Gujjar (Gurjar)

Gujjar also known as Gurjar, a pastoral ethnic group of India and Pakistan that used to be nomadic earlier [36]. The other synonymous of Gujjar are *Gurjjar, Gurjara, Gojar and Gūjar* [37–39].

The term Gujjar is derived from the Sanskrit word gurjara, which many scholars refer to the word Gurjaratra or Gujarat. The Gurjar origin is debatable, Cunningham[40] places them among Scythian tribe who conquered the Kabul around 100 BC and came to India and established the Gujjara or Gurjara kingdom while Smith[41] traces their origin to the white Huns who came to India around AD 465 as nomadic hordes.

In the 6th to 12th century, the Gujjar were primarily classified into Kshatriya and Brahmins many of whom are converted to Islam during Muslim rule in South Asia[42,43]. In some states in India, Gujjar are classified under the Other Backward Class (OBC)[44], Hindu Gujjars are now assimilated into several vernas. Many Gujjars link themselves with *Ramayana*[45], claiming that they are related with legendary hero *Lakshmana* brother of *Lord Rama*[46,47]. They claim to be Suryavanshi Kshatriya [48].

Gujjar of Northwestern India have ethnic affinities with Jats, Rajputs, Ahirs and belongs to one ethnic group[49,50]. Gujjars are widely spread over Jammu-Kashmir, Punjab, Chandigarh, Haryana, Himachal Pradesh, Rajasthan, Gujarat, Madhya Pradesh, Uttar Pradesh, Uttaranchal, and Delhi[36,51,52]. Muslim Gujjars are the prominent tribe in Pakistan occupied Kashmir and Pakistan and contributes almost twenty percent of the country's entire population[53]. The Gujjar population size in India is expected to be around 1,073,000[54].

Kamboj

Kamboj or Kamboh are an ethnic group from Northwest India, which may relate to the ancient Kamboja of Iron Age India and mentioned repeatedly in ancient Sanskrit texts and epigraphs[55–58]. The earliest reference to the Kambojas are recorded in *Pāṇini's* work (5th century BCE). Later they were referred in *Manusmirti* (2nd century BCE) and the *Mahabharata* (1st century BCE), describing Kambojas as former Warrior kshatriyas who could not stand by Hindu sacred rituals[59]. The ancient Kambojas were supposed of Indo-Iranian origin[56], however, they are also described as Indo-Aryans[57,58,60] or royal clan of the Sakas[61] time to time. The kingdom Kambojas territories were positioned beyond Pakistan, Gandhara, Afghanistan placing in Tajikistan, Uzbekistan, and Kyrgyzstan, and the *Edicts of Ashoka* (3rd century BCE) refers the Kamboja controlled territory an independent entity from the Mauryan empire in which it was located[59]. Clans of Kambojas together with Sakas, Pahlvas and the Yavanas, migrated from Central Asia during 2nd and 1st centuries BCE. They entered Afghanistan and India, setting up independent territories in Western and South-western India[62,63].

In Punjab, Haryana and Rajasthan region of India, many Kamboj frequently use Kamboj as their surname in lieu of the sub-caste or the gotra name, while their Muslim counterparts in Pakistan mostly use Kamboh as the last name instead of gotra name. A sizable number of Muslim Kamboh are also found in doab of Uttar Pradesh, India. The total population size of Kamboj Hindu, Kamboj Sikhs, and Kamboj Muslims are 508000, 546000 and 30000 respectively in India (Joshua Report). Kamboj of Indian sub-continent are having an ethnic affinity with Khatris, Rajputs, Brahmins, Jats, Aryans, and others.

Khatri

Khatri, a caste population in India and Pakistan, are mostly from the Punjab region. Additionally, they also live in Delhi, Uttar Pradesh, Madhya Pradesh, Himachal Pradesh, Chandigarh, Gujarat, Haryana, Rajasthan, Maharashtra and J & K. During the Mughal Empire, Khatris played a significant role in India's trans-regional trade[64]. Khatris have also been described among the "utmost vital merchant communities of early modern India"[65]. Khatri also claim Kshatriya status and regardless of individual ranking within the community, they believed themselves to be of pure Vedic stock and superior to Rajputs[66]. Although the Indian and the British authorities did not accept the Khatri claim of Kshatriya status since large number of mainstream Khatris were engaged in mercantile occupations[67]. Khatri of different regions follow different professions. Khatri claim that they were warriors who took to trade[68]. Historically, Khatri were formerly involved in the knitting of silk saris, and later some of them turned out to be merchants[52]. Khatri in India have a population size of 2,303,000 and follow mainly Hinduism and Sikhism[69].

Jat

Jat (also spelled as "Jatt" and "Jaat") people are an ethnic agricultural group that has apparently settled in Northwest South Asia for several thousand years[70]. The term Jat is supposedly derived from Jatta, a broad word used for cattle gazers and camel breeders, moving in a group – jatha. Jats have been known as zamindars (landowner) since the period of Mogul emperor Akbar in 16th century[71].

Jat have been identified as related to the Indo-Scythians that entered India between 200 BC and 600 AD[41]. Greek historians *Pliny* and *Ptolemy* suggested the movement of the Jat people from the banks of Oxus river to India around a century before Christ. They spoke Indo-Aryan languages and considered to be amongst the oldest ancient people of India, who settled in the fertile plains of Punjab along the Indus River turning into a pastoral and farmer community[71], few scholars believed Jat related to the Rajput tribes[50,72].

Jat largely belong to three main religious faiths, Hindu, Muslim, Sikhs and mostly are inhabited in Indian states of Haryana, Punjab, Delhi, Rajasthan, Uttar Pradesh and also Punjab, Sindh provinces of Pakistan[70,71,73]. the people in India were numbered to be around 82.5 million in 2012 and are classified as OBCs (other backward castes) in some states of India though, though only the Rajasthani Jats are entitled to OBC category by the central government of India[74].

Supplementary Figures

Figure S1. Principal Component Analysis (PCA) using haplogroup frequencies in South Asians, Iranians and Central Asians (A) mtDNA PC plot and **(B) Y-chromosome PC plot** triangle indicate new samples from Northwest Indian (NWI) and Gujjar_NW refers to the new Gujjar sampled from NWI.

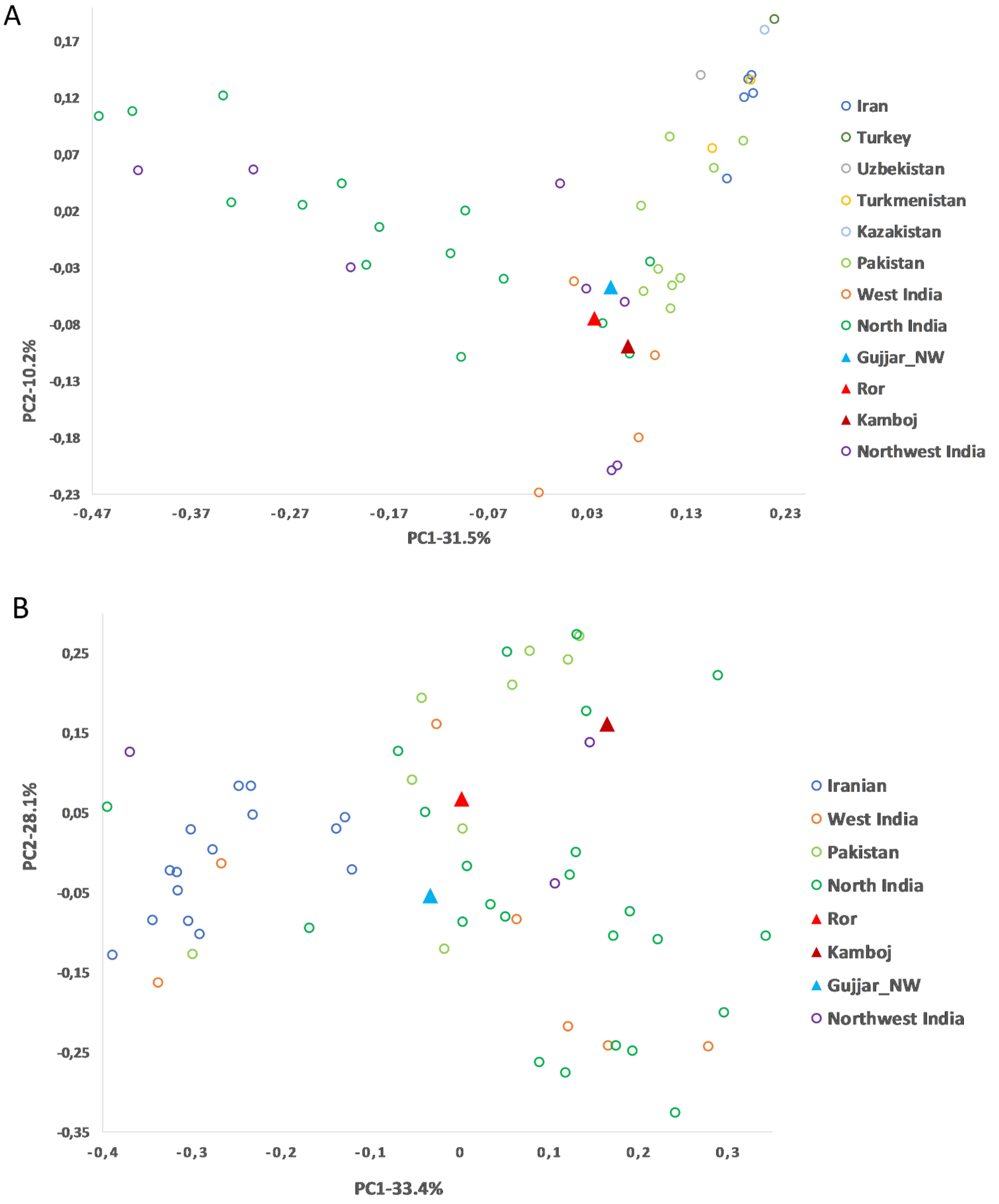


Figure S2. Genome-wide structure and genetic differentiation of Northwest India (NWI) (A) PCA of NWI with Eurasians PCA of NW India and other Eurasian populations. Mean eigenvalues of South Asians are shown in the inset plot, and **(B) Mean population pairwise *F*ST** was computed between Northwest Indians and world populations, given in diagonal. Indian IE_Caste, Indian Indo-European castes; Indian IE_Tribe, Indian Indo-European tribes; Indian IE_Rel gr, Muslim; Munda, Indian Austro Asiatic (Indian_AA) speakers.

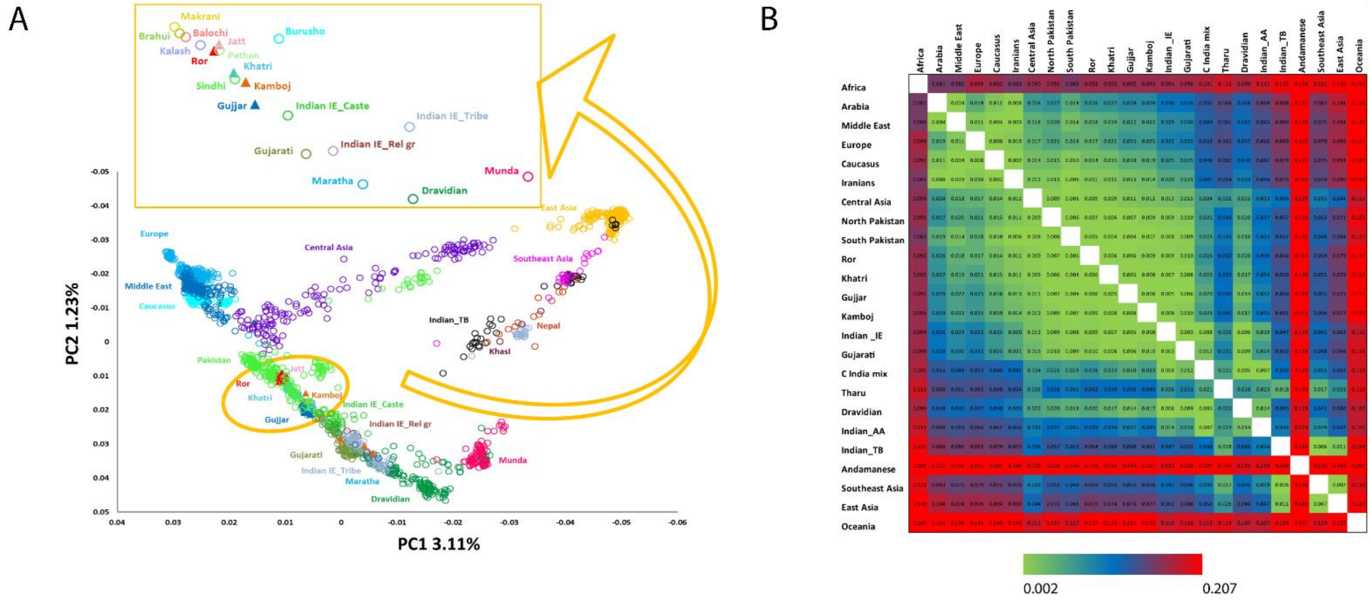


Figure S3. Genetic structure of PNWI in context of ancient DNA: (A) PC plot for modern Eurasians with a subset of ancient sources from Steppe, Iran, Anatolia, Central Asia, and South Asia, and **(B) PC plot for modern South Asians** with a subset of ancient sources.

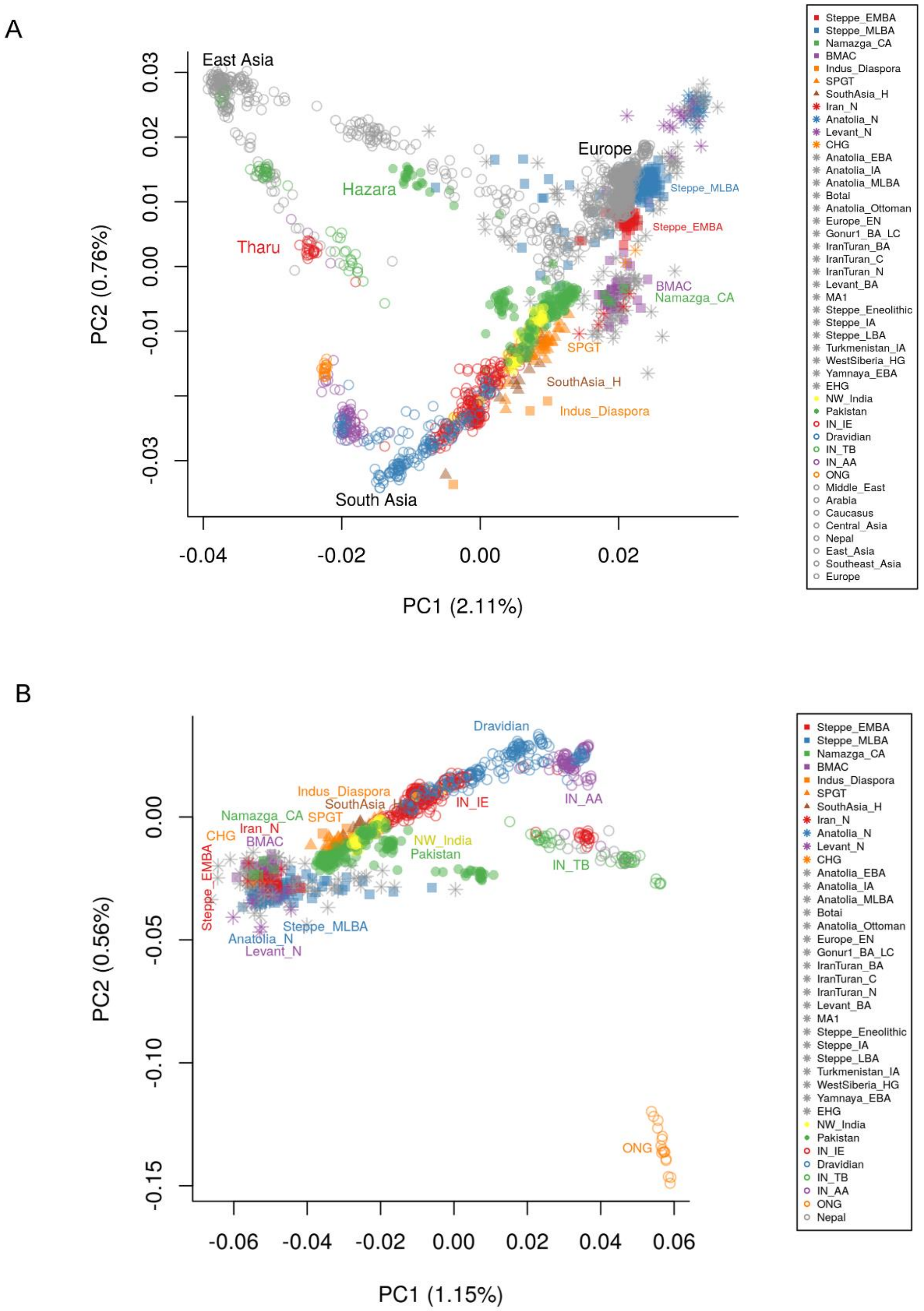


Figure S4. Cross-Validation for ADMIXTURE run plot showing cross-validation (CV) indexes of all runs of the Admixture analysis.

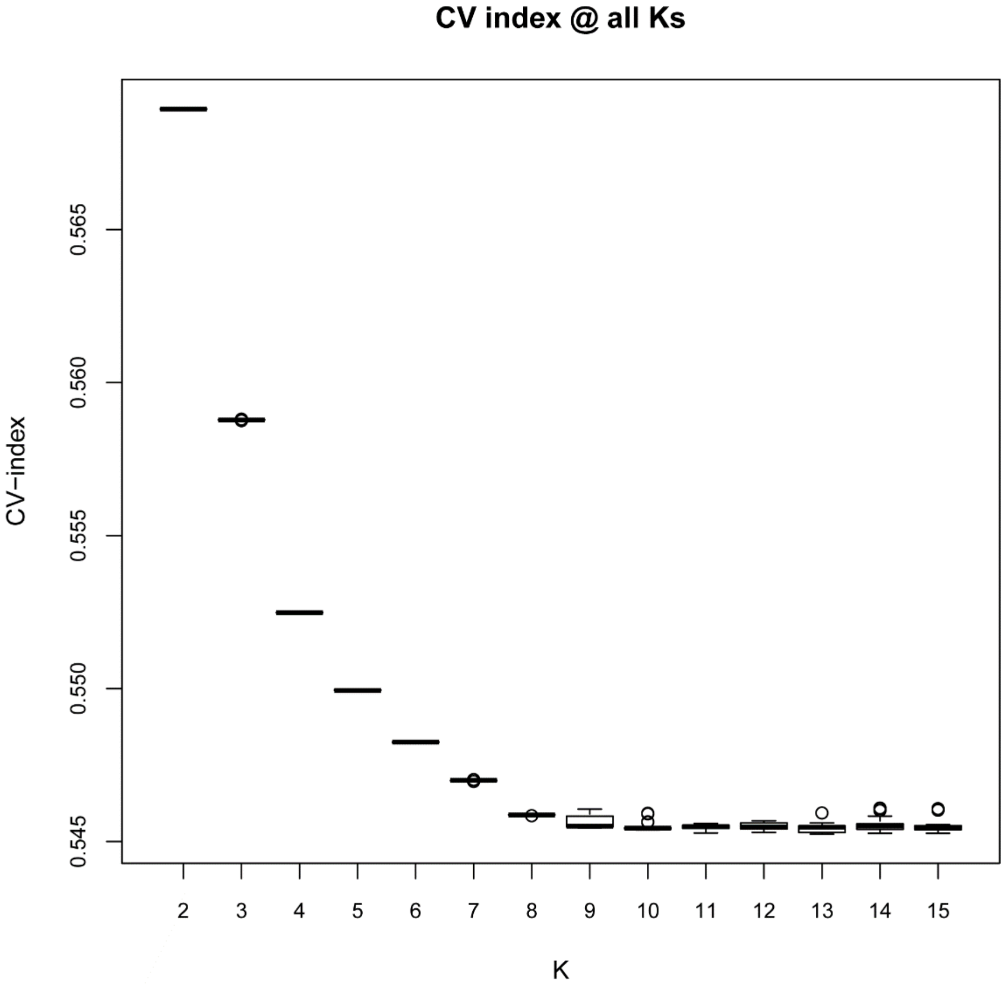
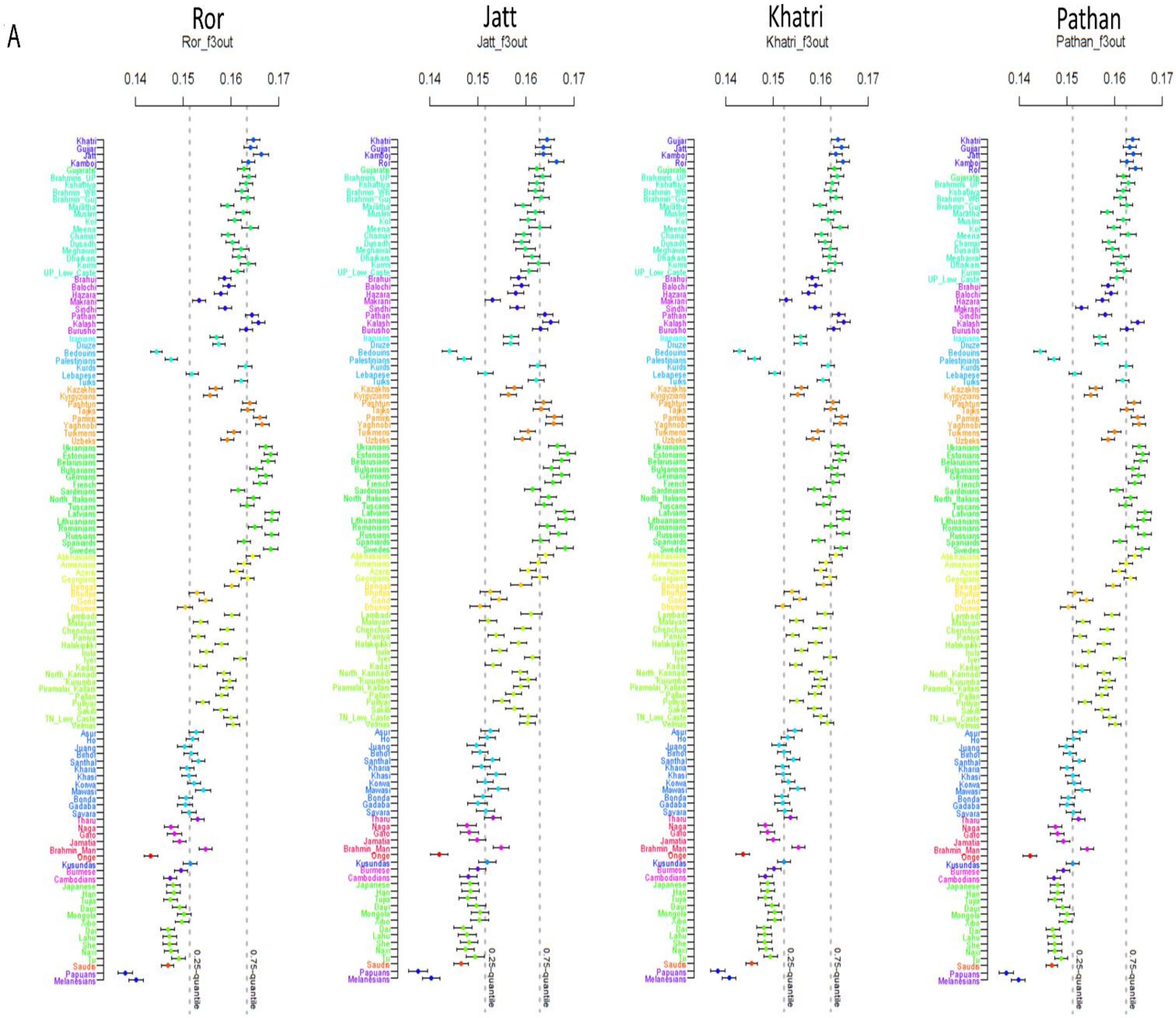


Figure S5. Derived allele sharing between modern PNWI and Eurasian groups (A) and **(B) plots of derived allele sharing** obtained for PNWI and Indian Gujaratis using Outgroup *f*3 (NWI/Pakistan/Gujaratis, X; Yoruba) test, where X is any other modern Eurasian population. Error bars represent jack-knife standard errors.



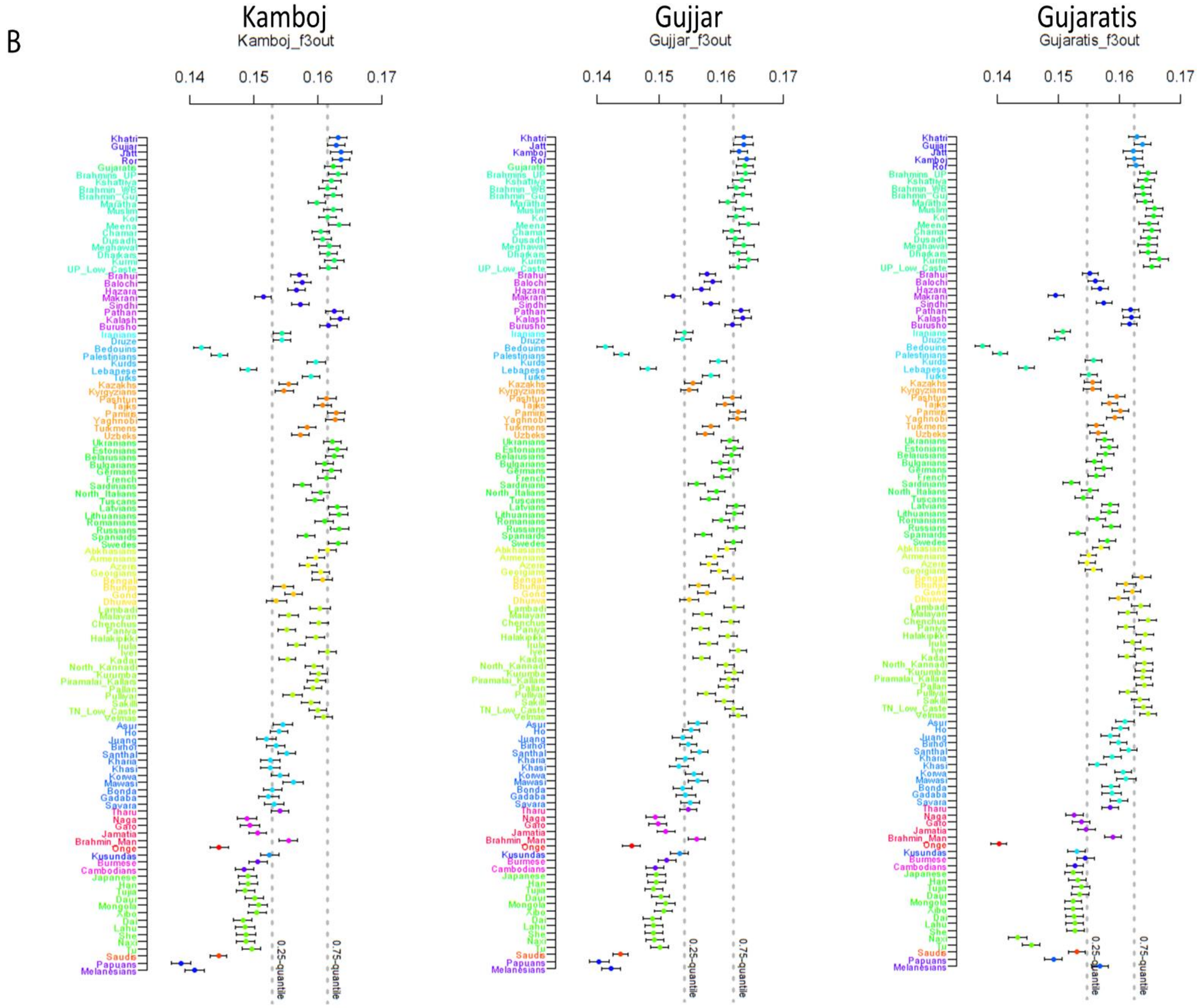
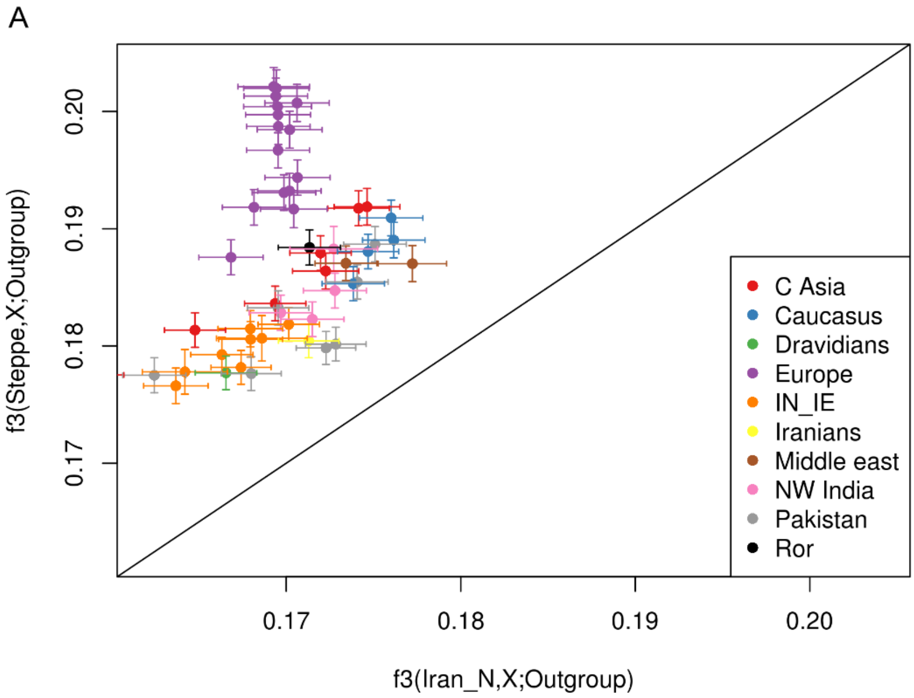
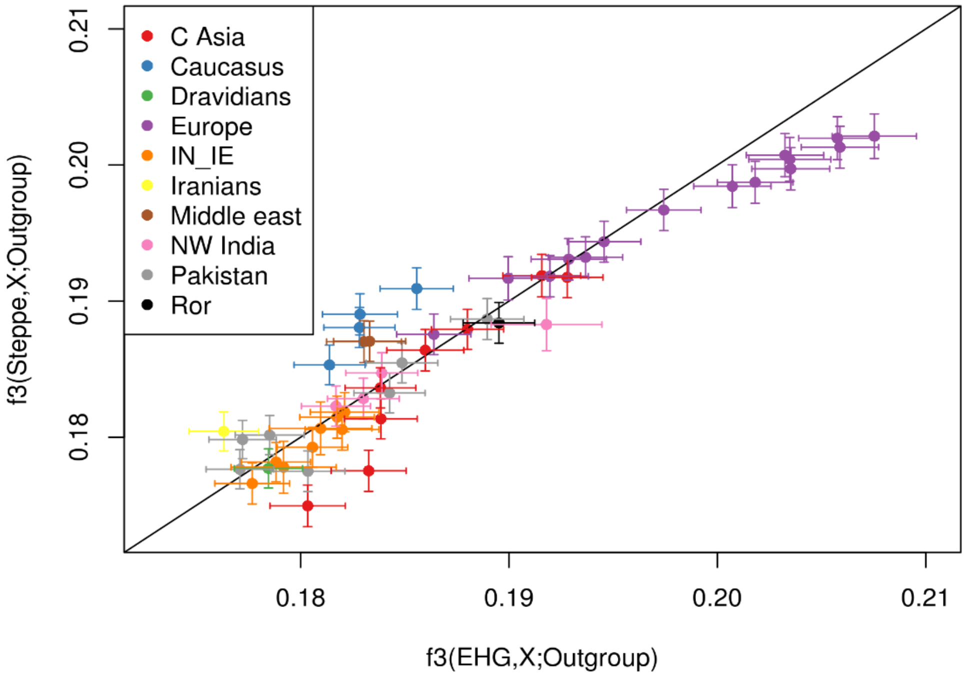

B
Kamboj
Kamboj_f3out
0.14
0.15
0.16
0.17
0.25-quantile
0.75-quantile
Khatri
Gujjar
Jatt
Ror
Gujarati
Brahmins_UP
Kshatriya
Brahmin_WB
Brahmin_Guj
Maratha
Muslim
Koli
Meena
Chamar
Dusadh
Meghawal
Dhankars
Kurmi
UP_Low_Caste
Brahui
Balochi
Hazara
Makrani
Sindhi
Pathan
Kalash
Burusho
Iranians
Druze
Bedouins
Palestinians
Kurds
Lebanese
Turks
Kazakhs
Kyrgyzians
Pashtun
Tajiks
Pamiris
Yaghnobi
Turkmens
Uzbeks
Ukranians
Estonians
Belarusians
Bulgarians
Germans
French
Sardinians
North_Italians
Tuscans
Latvians
Lithuanians
Romanians
Russians
Spaniards
Swedes
Abkhasians
Armenians
Azeris
Georgians
Bengali
Bhunjia
Gond
Dhurwa
Lambadi
Malayan
Chenchus
Paniya
Halakipikki
Irula
Iyer
Kadar
North_Kannadi
Kurumba
Piramalai_Kallars
Pallan
Pulliyar
Sakilli
TN_Low_Caste
Velmas
Asur
Ho
Juang
Birhor
Santhal
Kharia
Khasi
Korwa
Mawasi
Bonda
Gadaba
Savara
Tharu
Naga
Garo
Jamatia
Brahmin_Man
Onge
Kusundas
Burmese
Cambodians
Japanese
Han
Tujia
Daur
Mongola
Xibo
Dai
Lahu
She
Naxi
Tu
Saudis
Papuans
Melanesians
Gujjar
Gujjar_f3out
0.14
0.15
0.16
0.17
0.25-quantile
0.75-quantile
Khatri
Jatt
Kamboj
Ror
Gujarati
Brahmins_UP
Kshatriya
Brahmin_WB
Brahmin_Guj
Maratha
Muslim
Koli
Meena
Chamar
Dusadh
Meghawal
Dhankars
Kurmi
UP_Low_Caste
Brahui
Balochi
Hazara
Makrani
Sindhi
Pathan
Kalash
Burusho
Iranians
Druze
Bedouins
Palestinians
Kurds
Lebanese
Turks
Kazakhs
Kyrgyzians
Pashtun
Tajiks
Pamiris
Yaghnobi
Turkmens
Uzbeks
Ukranians
Estonians
Belarusians
Bulgarians
Germans
French
Sardinians
North_Italians
Tuscans
Latvians
Lithuanians
Romanians
Russians
Spaniards
Swedes
Abkhasians
Armenians
Azeris
Georgians
Bengali
Bhunjia
Gond
Dhurwa
Lambadi
Malayan
Chenchus
Paniya
Halakipikki
Irula
Iyer
Kadar
North_Kannadi
Kurumba
Piramalai_Kallars
Pallan
Pulliyar
Sakilli
TN_Low_Caste
Velmas
Asur
Ho
Juang
Birhor
Santhal
Kharia
Khasi
Korwa
Mawasi
Bonda
Gadaba
Savara
Tharu
Naga
Garo
Jamatia
Brahmin_Man
Onge
Kusundas
Burmese
Cambodians
Japanese
Han
Tujia
Daur
Mongola
Xibo
Dai
Lahu
She
Naxi
Tu
Saudis
Papuans
Melanesians
Gujaratis
Gujaratis_f3out
0.14
0.15
0.16
0.17
0.25-quantile
0.75-quantile
Khatri
Gujjar
Jatt
Kamboj
Ror
Brahmins_UP
Kshatriya
Brahmin_WB
Brahmin_Guj
Maratha
Muslim
Koli
Meena
Chamar
Dusadh
Meghawal
Dhankars
Kurmi
UP_Low_Caste
Brahui
Balochi
Hazara
Makrani
Sindhi
Pathan
Kalash
Burusho
Iranians
Druze
Bedouins
Palestinians
Kurds
Lebanese
Turks
Kazakhs
Kyrgyzians
Pashtun
Tajiks
Pamiris
Yaghnobi
Turkmens
Uzbeks
Ukranians
Estonians
Belarusians
Bulgarians
Germans
French
Sardinians
North_Italians
Tuscans
Latvians
Lithuanians
Romanians
Russians
Spaniards
Swedes
Abkhasians
Armenians
Azeris
Georgians
Bengali
Bhunjia
Gond
Dhurwa
Lambadi
Malayan
Chenchus
Paniya
Halakipikki
Irula
Iyer
Kadar
North_Kannadi
Kurumba
Piramalai_Kallars
Pallan
Pulliyar
Sakilli
TN_Low_Caste
Velmas
Asur
Ho
Juang
Birhor
Santhal
Kharia
Khasi
Korwa
Mawasi
Bonda
Gadaba
Savara
Tharu
Naga
Garo
Jamatia
Brahmin_Man
Onge
Kusundas
Burmese
Cambodians
Japanese
Han
Tujia
Daur
Mongola
Xibo
Dai
Lahu
She
Naxi
Tu
Saudis
Papuans
Melanesians

Figure S6. Comparing Outgroup *f_3* for PNWI groups with ancient sources
(A) Steppe_EMBA vs Iran_N scatter plot of Outgroup *f_3* (Iran_N, X; Yoruba) on x-axis vs Outgroup *f_3* (Steppe_EMBA, X; Yoruba as outgroup) on y-axis indicating the distinct pattern of PNWI in their affinity to Steppe highlighting Ror and Jat the closest ones, **(B) Steppe_EMBA vs EHG scatter plot** of Outgroup *f_3* (EHG, X; Yoruba) on x-axis vs Outgroup *f_3* (Steppe_EMBA, X; Yoruba as Outgroup) on y-axis indicating the distinct pattern for PNWI and Ror (and Jat) in their higher affinity to Steppe and EHG than other Indian and **(C) Anatolia_N vs Iran_N scatter plot** of Outgroup *f_3* (Iran_N, X; Yoruba) on x-axis vs Outgroup *f_3* (Anatolia_N, X; Yoruba) highlighting Ror as the only group in their relative higher affinity to Neolithic Anatolian than that of other South Asians. Here X refers to different West Eurasian and South Asian populations. Steppe_EMBA, Early and Bronze Age Steppe; Iran_N, Iran_Neolithic; EHG, Eastern Hunter Gatherer; Anatolia_N, Anatolia_Neolithic. Error bars represent jack-knife standard errors.

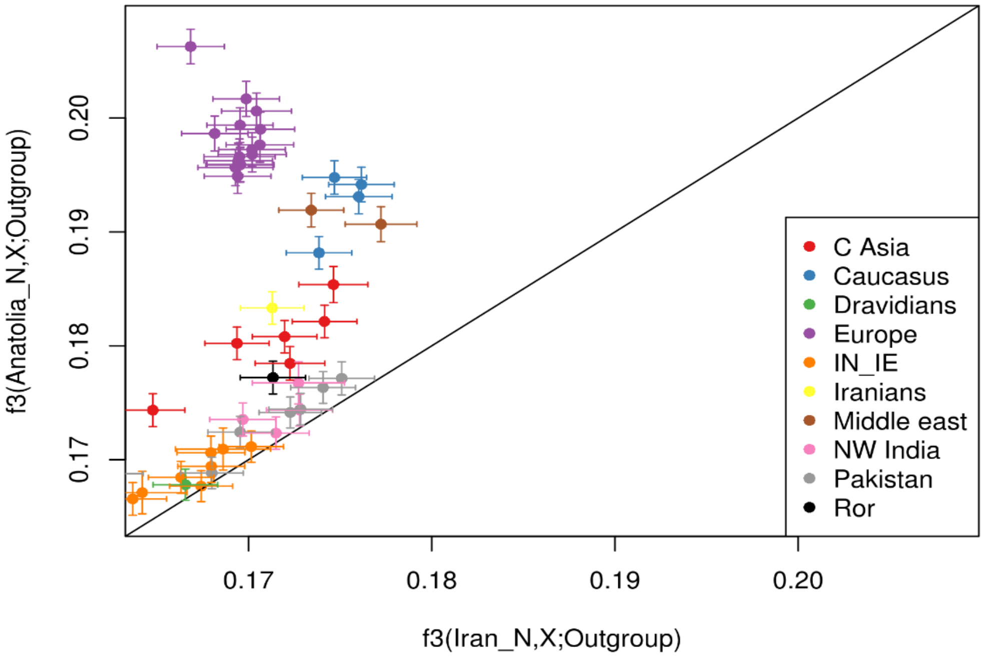


B



● C Asia
● Caucasus
● Dravidians
● Europe
● IN_IE
● Iranians
● Middle east
● NW India
● Pakistan
● Ror
0.18
0.19
0.20
0.21
f3(EHG,X;Outgroup)
0.18
0.19
0.20
0.21
f3(Steppe,X;Outgroup)

C



● C Asia
● Caucasus
● Dravidians
● Europe
● IN_IE
● Iranians
● Middle east
● NW India
● Pakistan
● Ror
0.17
0.18
0.19
0.20
f3(Iran_N,X;Outgroup)
0.17
0.18
0.19
0.20
f3(Anatolia_N,X;Outgroup)

Figure S7. Outgroup f_3 for South Asians and West Eurasians with ancient sources (A) EHG, **(B) Anatolia_N**, **(C) Iran_N** and **(D) Steppe_EMBA** Derived allele sharing obtained by Outgroup f_3 (Ancient Sources, X; Yoruba) test, where Ancient sources are EHG/Anatolia_N/Iran_N/Steppe_EMBA and X is any population from South Asia and West Eurasia. Error bars represent jack-knife standard errors.

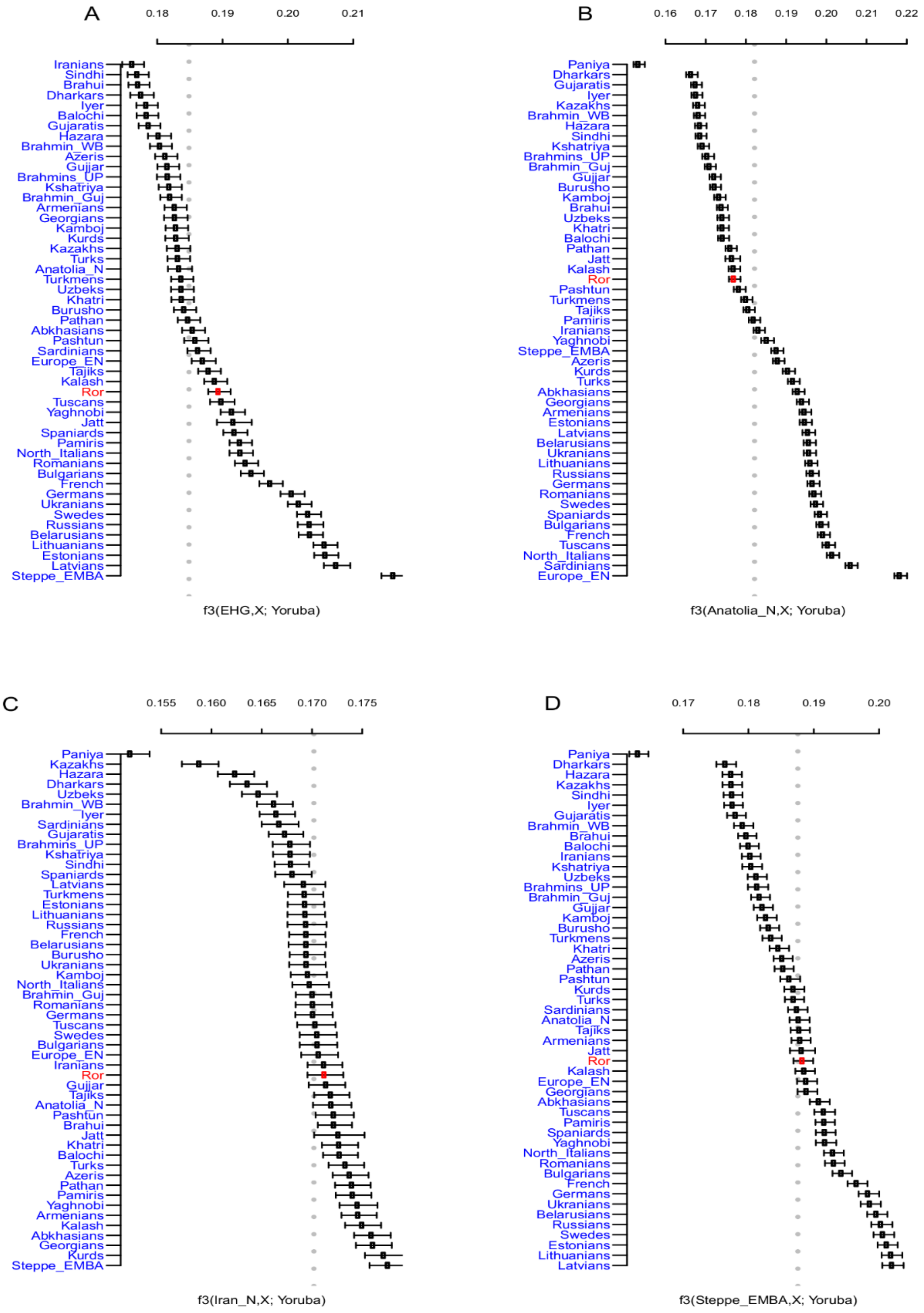
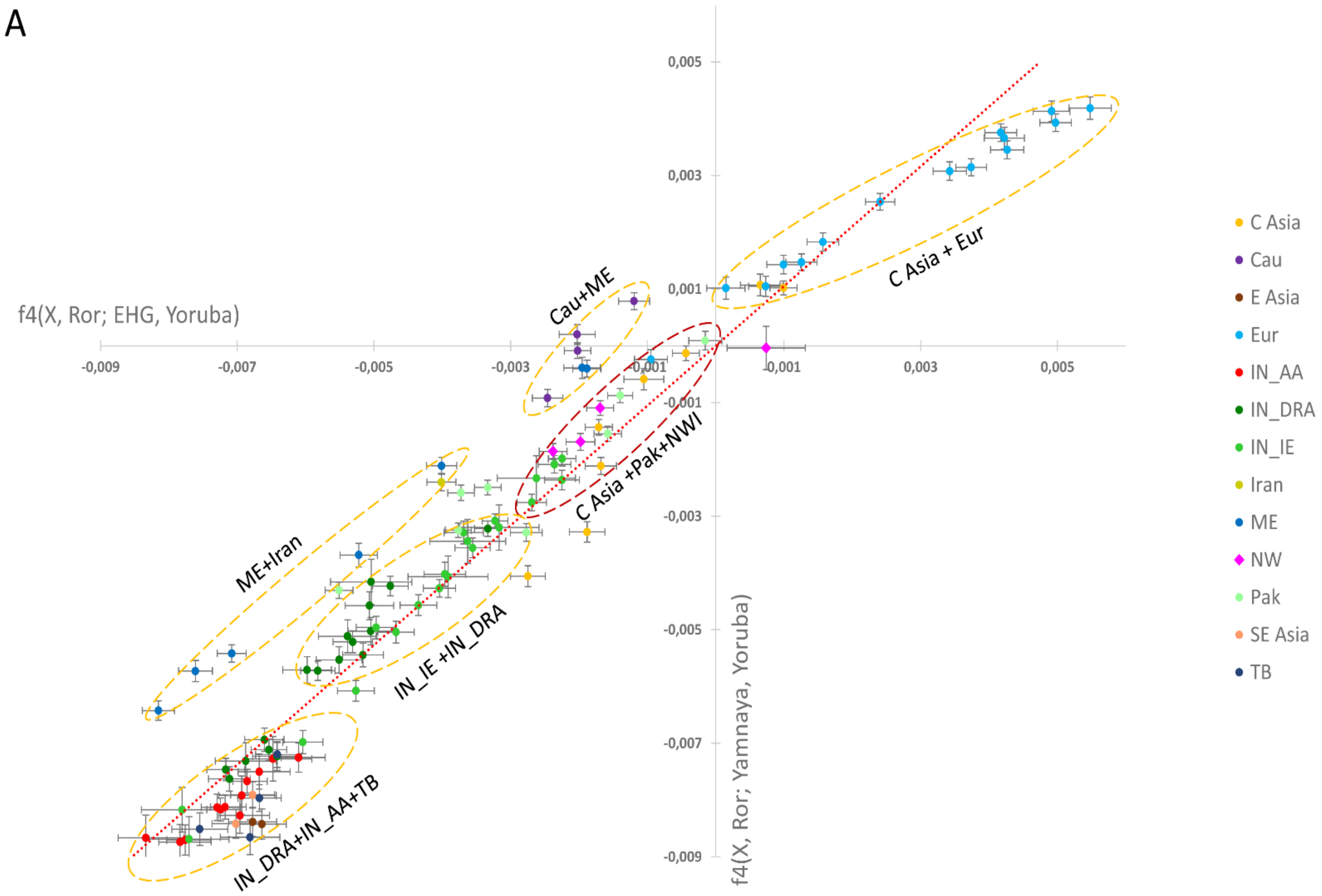
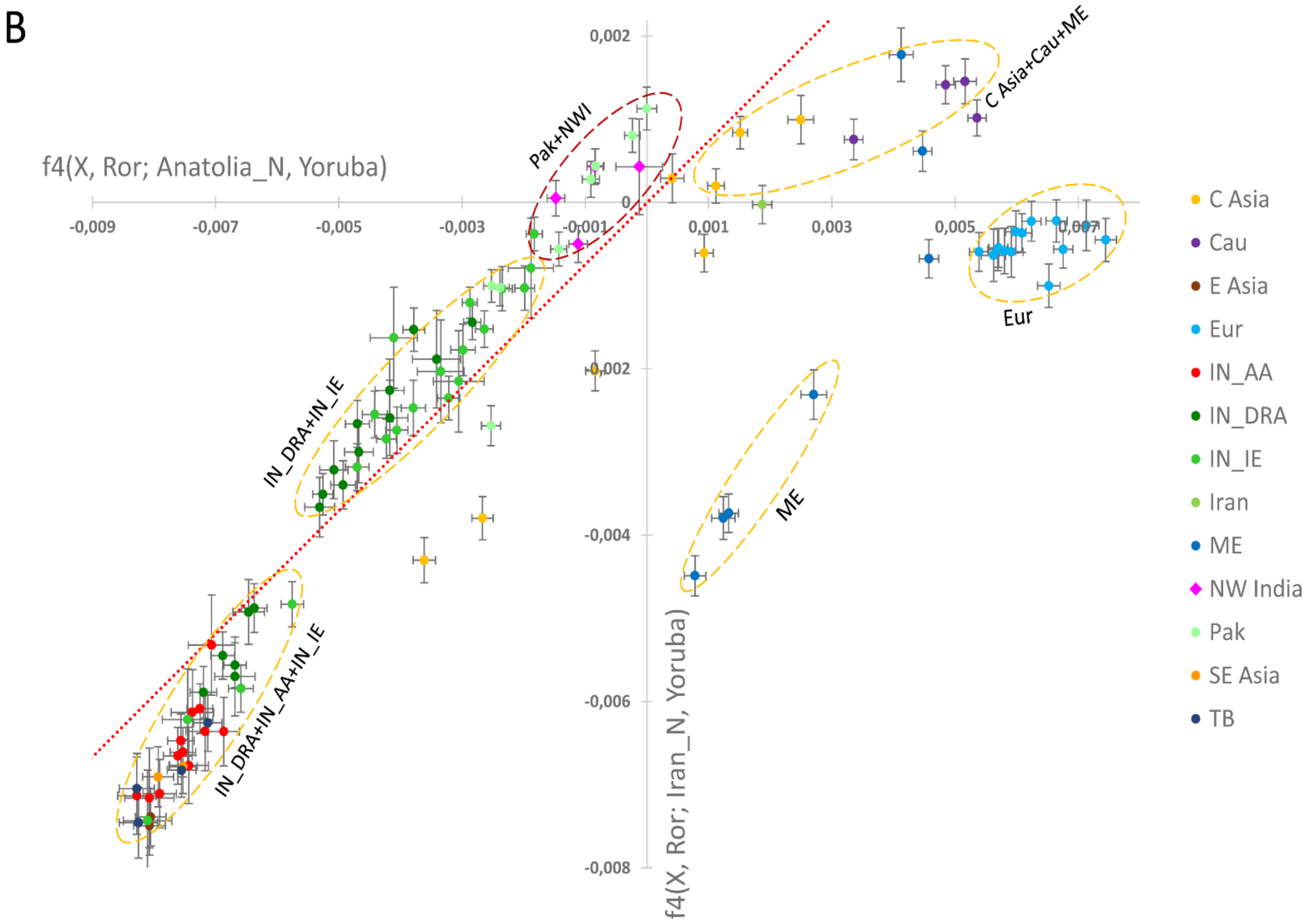
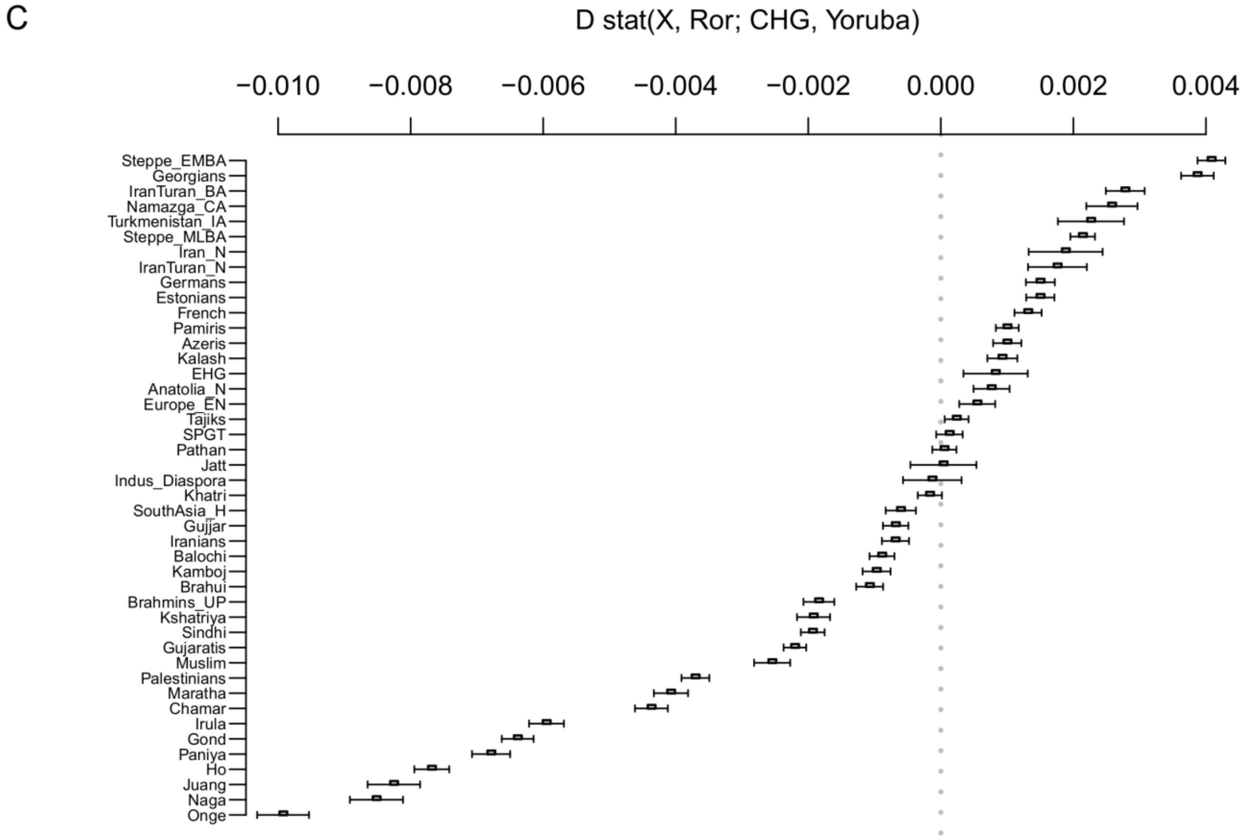


Figure S8. *D* statistics of South Asians with distant ancient sources (A) EHG vs Steppe_EMBA scatter plot comparing *D*(X, Ror; EHG, Yoruba) vs *D*(X, Ror; Steppe_EMBA, Yoruba), where X is a South Asian/West Eurasian, showing distinct affinity-based clustering of PNWI, striking out Ror and Jat in their highest allele sharing to EHG than Steppe_EMBA (or Yamnaya) in South Asia, **(B) Anatolia_N vs Iran_N scatter plot** comparing *D*(X, Ror; Anatolia_N, Yoruba) vs *D*(X, Ror; Iran_N, Yoruba) where X is a South Asian/West Eurasian, showing distinct affinity-based clustering of PNWI highlighting Ror in their highest allele sharing to Anatolia_N than Iran_N, in South Asia, **(C) *D* stat with CHG** *D*(X, Ror; CHG, Yoruba) reports distinct affinity of PNWI to CHG, Ror and Jat are the closest to CHG among South Asians, **(D) *D* stat with EHG** *D*(X, modern Iran; EHG, Yoruba) indicate the relative allele sharing of West Eurasians and South Asians to EHG when compared to modern Iranians. ME, Middle East; E Asia, East Asia; IN_IE, Indian Indo-European; NW and NW India, Northwest India; SE Asia, South East Asia; TB, Tibeto-Burman; Pak, Pakistan. Error bars represent jack-knife standard errors.




B
f4(X, Ror; Anatolia_N, Yoruba)
f4(X, Ror; Iran_N, Yoruba)
-0,009
-0,007
-0,005
-0,003
-0,001
0
0,001
0,003
0,005
0,007
0,002
-0,002
-0,004
-0,006
-0,008
Pak+NWI
C Asia+Cau+ME
IN_DRA+IN_IE
IN_DRA+IN_AA+IN_IE
ME
Eur
● C Asia
● Cau
● E Asia
● Eur
● IN_AA
● IN_DRA
● IN_IE
● Iran
● ME
◆ NW India
● Pak
● SE Asia
● TB


C
D stat(X, Ror; CHG, Yoruba)
−0.010
−0.008
−0.006
−0.004
−0.002
0.000
0.002
0.004
Steppe_EMBA
Georgians
IranTuran_BA
Namazga_CA
Turkmenistan_IA
Steppe_MLBA
Iran_N
IranTuran_N
Germans
Estonians
French
Pamiris
Azeris
Kalash
EHG
Anatolia_N
Europe_EN
Tajiks
SPGT
Pathan
Jatt
Indus_Diaspora
Khatri
SouthAsia_H
Gujjar
Iranians
Balochi
Kamboj
Brahui
Brahmins_UP
Kshatriya
Sindhi
Gujaratis
Muslim
Palestinians
Maratha
Chamar
Irula
Gond
Paniya
Ho
Juang
Naga
Onge

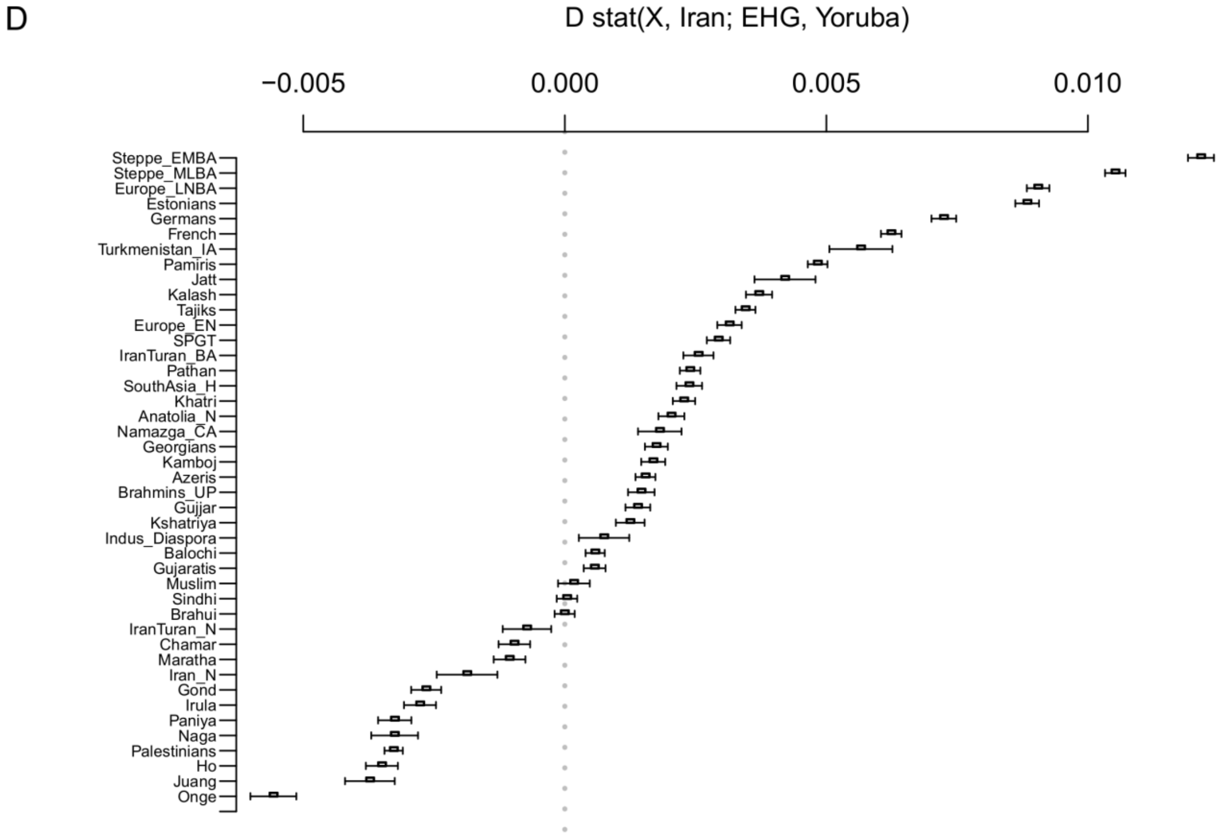
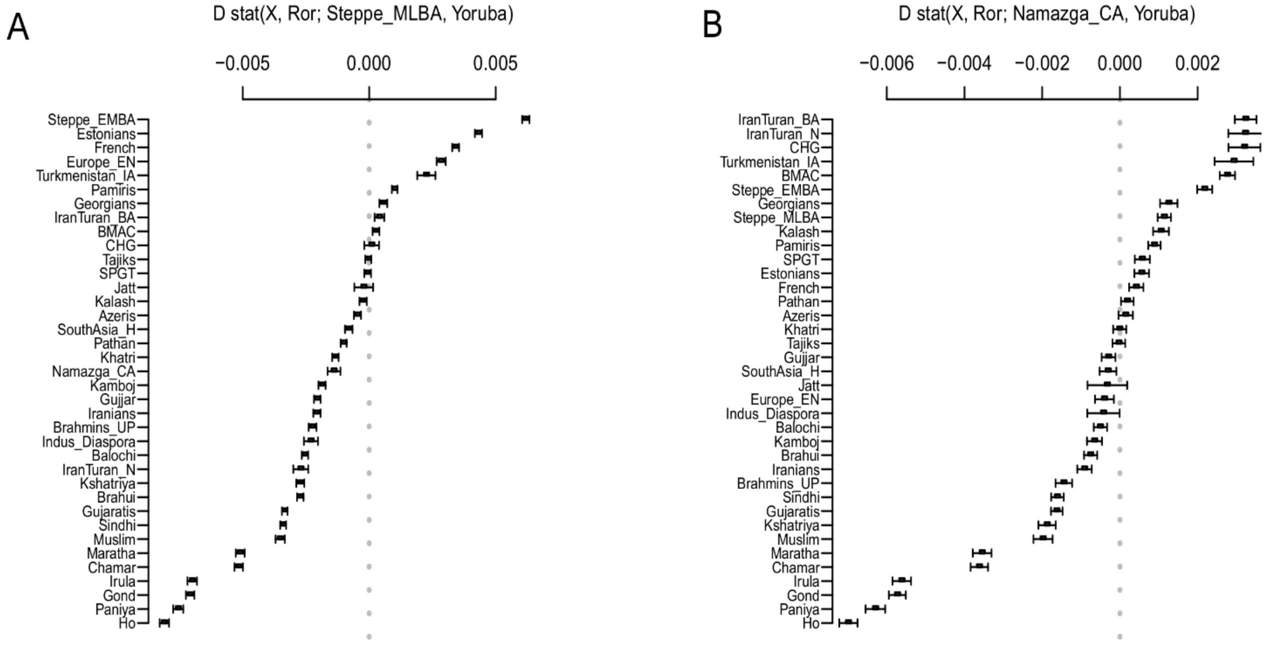

D
D stat(X, Iran; EHG, Yoruba)
−0.005
0.000
0.005
0.010
Steppe_EMBA
Steppe_MLBA
Europe_LNBA
Estonians
Germans
French
Turkmenistan_IA
Pamiris
Jatt
Kalash
Tajiks
Europe_EN
SPGT
IranTuran_BA
Pathan
SouthAsia_H
Khatri
Anatolia_N
Namazga_CA
Georgians
Kamboj
Azeris
Brahmins_UP
Gujjar
Kshatriya
Indus_Diaspora
Balochi
Gujaratis
Muslim
Sindhi
Brahui
IranTuran_N
Chamar
Maratha
Iran_N
Gond
Irula
Paniya
Naga
Palestinians
Ho
Juang
Onge

Figure S9. *D* statistics comparing allele sharing of South Asians with proximal ancient sources. We used *D* stat in form *D*(X, Ror; aDNA, Yoruba) where Yoruba X is a modern South Asian/West Eurasian/ancient group, aDNA is any of Steppe_MLBA/Namazga_CA/SPGT/Indus_Periphery/SouthAsia_H **(A) *D* stat with Steppe_MLBA** shows distinct affinity of NWI to Steppe_MLBA, highlighting Ror the closest to Steppe_MLBA among South Asians, **(B) *D* stat with Namazga_CA** showing equivalent affinity of NWI to Namazga_CA, striking out Kalash in their highest allele sharing to Namazga_CA than others, **(C) *D* stat with SPGT** showing distinct affinity of NWI to SPGT, striking out Ror and Kalash in their highest allele sharing to SPGT than any other South Asians, **(D) *D* stat with Indus_Periphery (Indus_Diaspora)**showing equivalent affinity of all NWI populations to Indus_Periphery, **(E) *D* stat with SouthAsia_H** showing distinct affinity of NWI to SouthAsia_H, striking out Ror and Kalash in their highest allele sharing than other South Asians. Used acronyms: Steppe_MLBA, Middle to Late Bronze Age Steppe; Namazga_CA, Chalcolithic Namazga; Indus_Diaspora, Indus_Periphery; SPGT, Iron Age to pre-historical South Asia; SouthAsia_H, Early historical South Asia; BMAC, Bactria Margiana Complex; Turkmenistan_IA, Iron Age Turkmenistan; IranTuran_N, Neolithic IranTuran; IranTuran_BA, Bronze Age IranTuran. Error bars represent jack-knife standard errors.



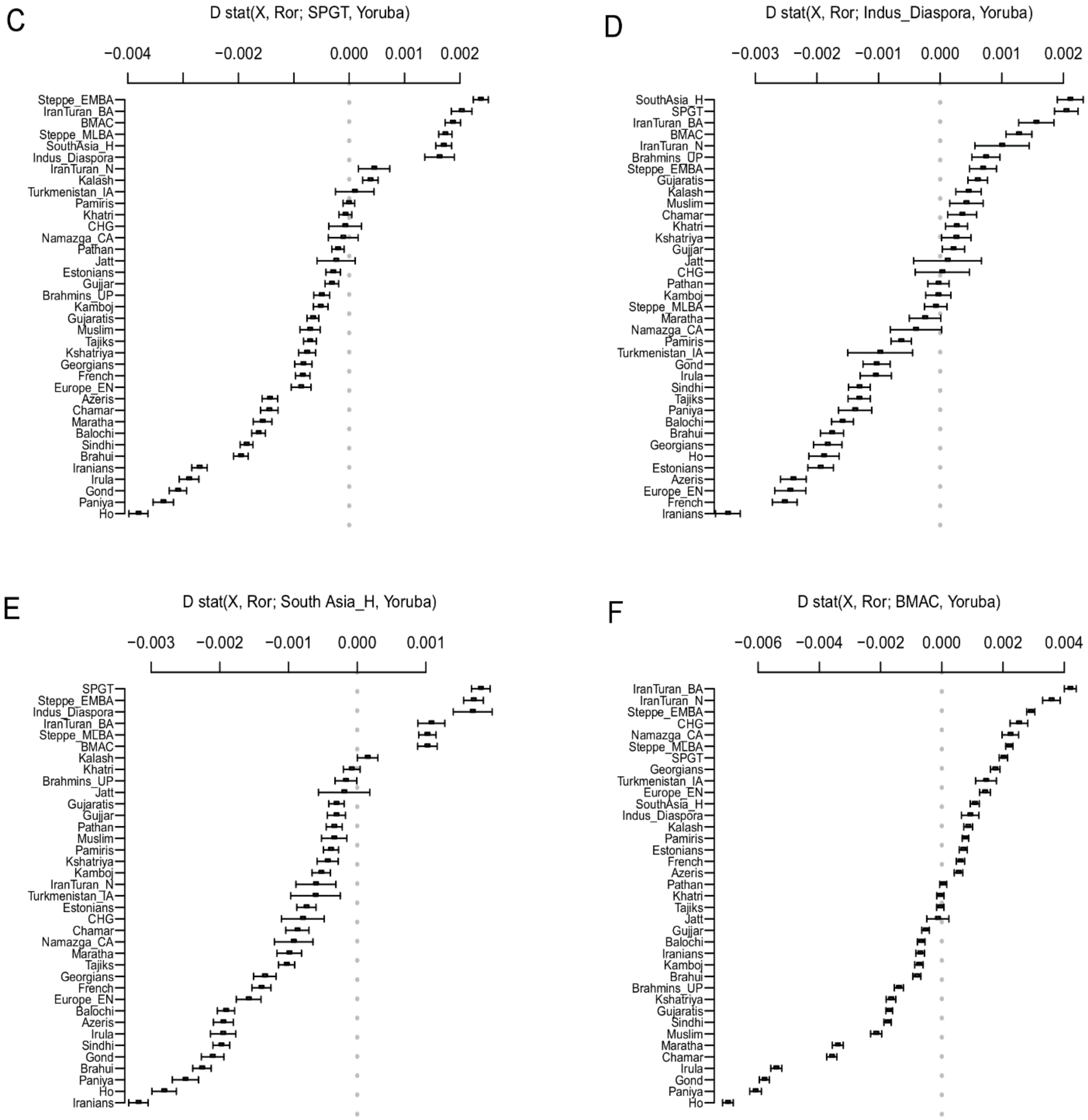

C
D stat(X, Ror; SPGT, Yoruba)
−0.004
−0.002
0.000
0.001
0.002
Steppe_EMBA
IranTuran_BA
BMAC
Steppe_MLBA
SouthAsia_H
Indus_Diaspora
IranTuran_N
Kalash
Turkmenistan_IA
Pamiris
Khatri
CHG
Namazga_CA
Pathan
Jatt
Estonians
Gujjar
Brahmins_UP
Kamboj
Gujaratis
Muslim
Tajiks
Kshatriya
Georgians
French
Europe_EN
Azeris
Chamar
Maratha
Balochi
Sindhi
Brahui
Iranians
Irula
Gond
Paniya
Ho
D
D stat(X, Ror; Indus_Diaspora, Yoruba)
−0.003
−0.002
−0.001
0.000
0.001
0.002
SouthAsia_H
SPGT
IranTuran_BA
BMAC
IranTuran_N
Brahmins_UP
Steppe_EMBA
Gujaratis
Kalash
Muslim
Chamar
Khatri
Kshatriya
Gujjar
Jatt
CHG
Pathan
Kamboj
Steppe_MLBA
Maratha
Namazga_CA
Pamiris
Turkmenistan_IA
Gond
Irula
Sindhi
Tajiks
Paniya
Balochi
Brahui
Georgians
Ho
Estonians
Azeris
Europe_EN
French
Iranians
E
D stat(X, Ror; South Asia_H, Yoruba)
−0.003
−0.002
−0.001
0.000
0.001
SPGT
Steppe_EMBA
Indus_Diaspora
IranTuran_BA
Steppe_MLBA
BMAC
Kalash
Khatri
Brahmins_UP
Jatt
Gujaratis
Gujjar
Pathan
Muslim
Pamiris
Kshatriya
Kamboj
IranTuran_N
Turkmenistan_IA
Estonians
CHG
Chamar
Namazga_CA
Maratha
Tajiks
Georgians
French
Europe_EN
Balochi
Azeris
Irula
Sindhi
Gond
Brahui
Paniya
Ho
Iranians
F
D stat(X, Ror; BMAC, Yoruba)
−0.006
−0.004
−0.002
0.000
0.002
0.004
IranTuran_BA
IranTuran_N
Steppe_EMBA
CHG
Namazga_CA
Steppe_MLBA
SPGT
Georgians
Turkmenistan_IA
Europe_EN
SouthAsia_H
Indus_Diaspora
Kalash
Pamiris
Estonians
French
Azeris
Pathan
Khatri
Tajiks
Jatt
Gujjar
Balochi
Iranians
Kamboj
Brahui
Brahmins_UP
Kshatriya
Gujaratis
Sindhi
Muslim
Maratha
Chamar
Irula
Gond
Paniya
Ho

Figure S10. LD Decay Curve for ALDER run (A) and **(B) Ror** weighted admixture LD Decay curve for best representative ALDER results for Ror, **(C) Kamboj** weighted admixture LD Decay curve for best representative ALDER results for Kamboj, **(D) Gujjar** weighted admixture LD Decay curve for best representative ALDER results for Gujjar.

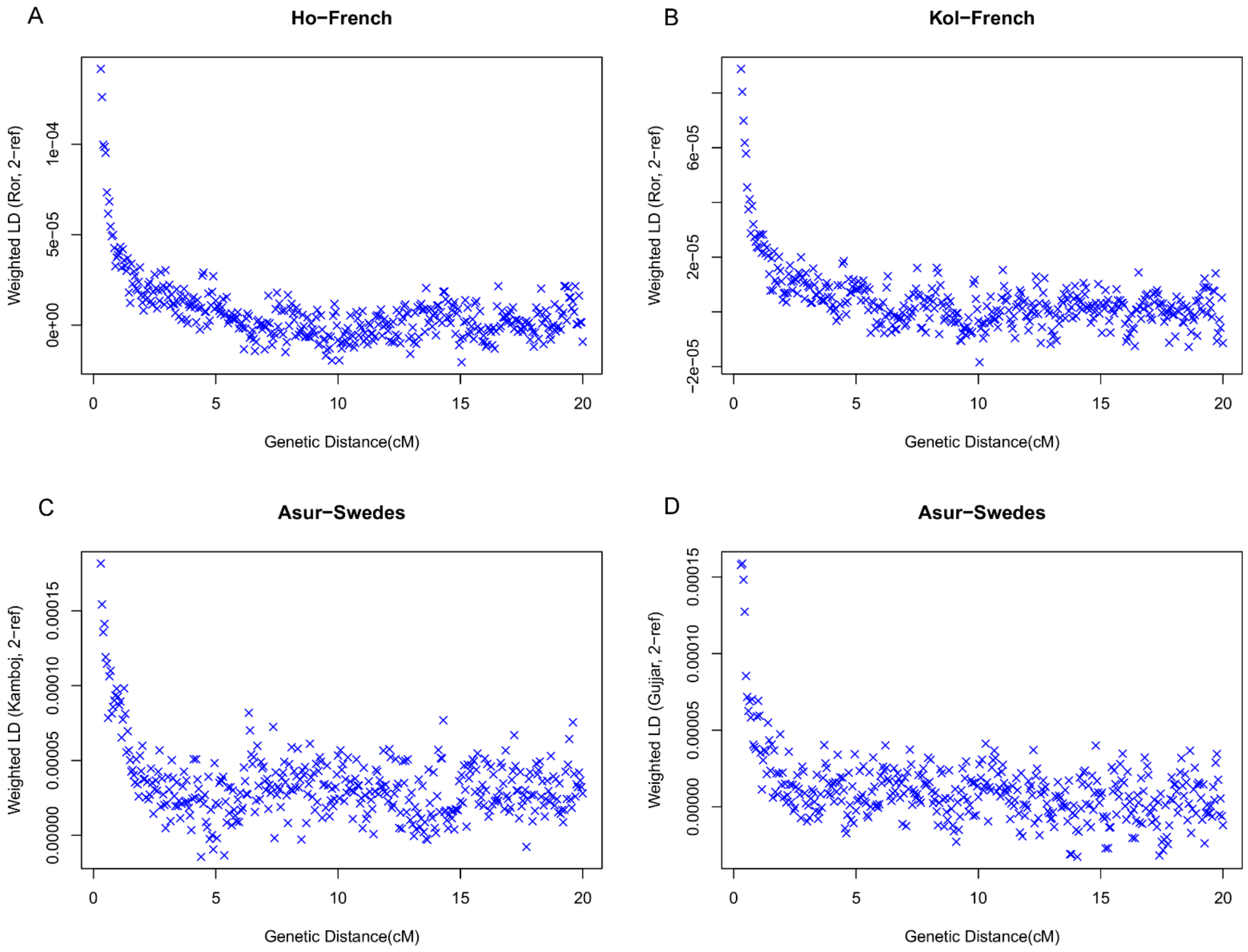


Figure S11. Maximum Likelihood tree of *TreeMix* relationship of Northwest Indian groups with neighboring Pakistani and other Eurasians inferred by *TreeMix* showing Ror and Kalash in the same clade.

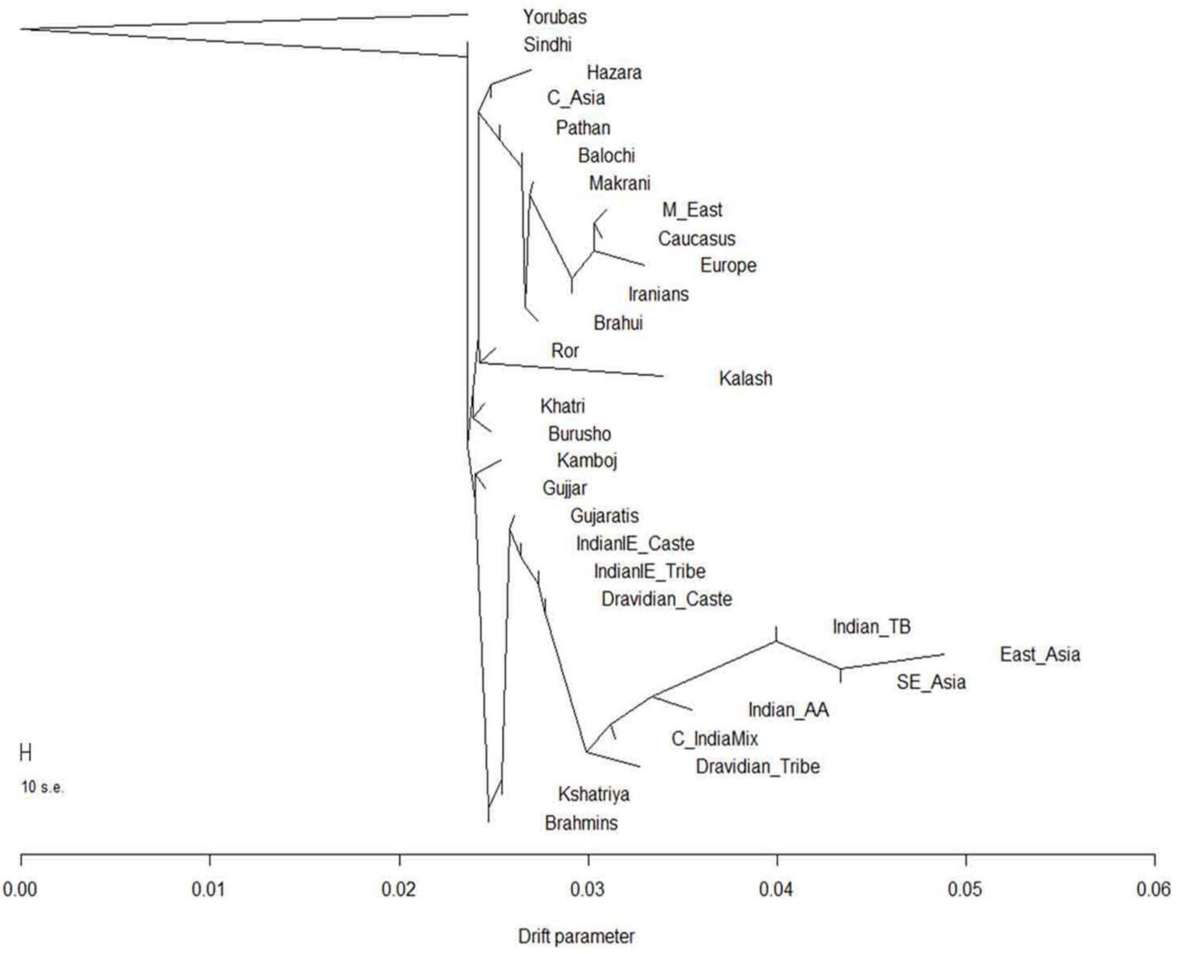


Figure S12. Runs of Homozygosity suggesting high density and population size for Ror the plot of Runs of Homozygosity (RoH) for Northwest Indians and other South Asians comparing the average length of RoH amongst them, using three different windows of 1000, 2500 and 5000 kb.

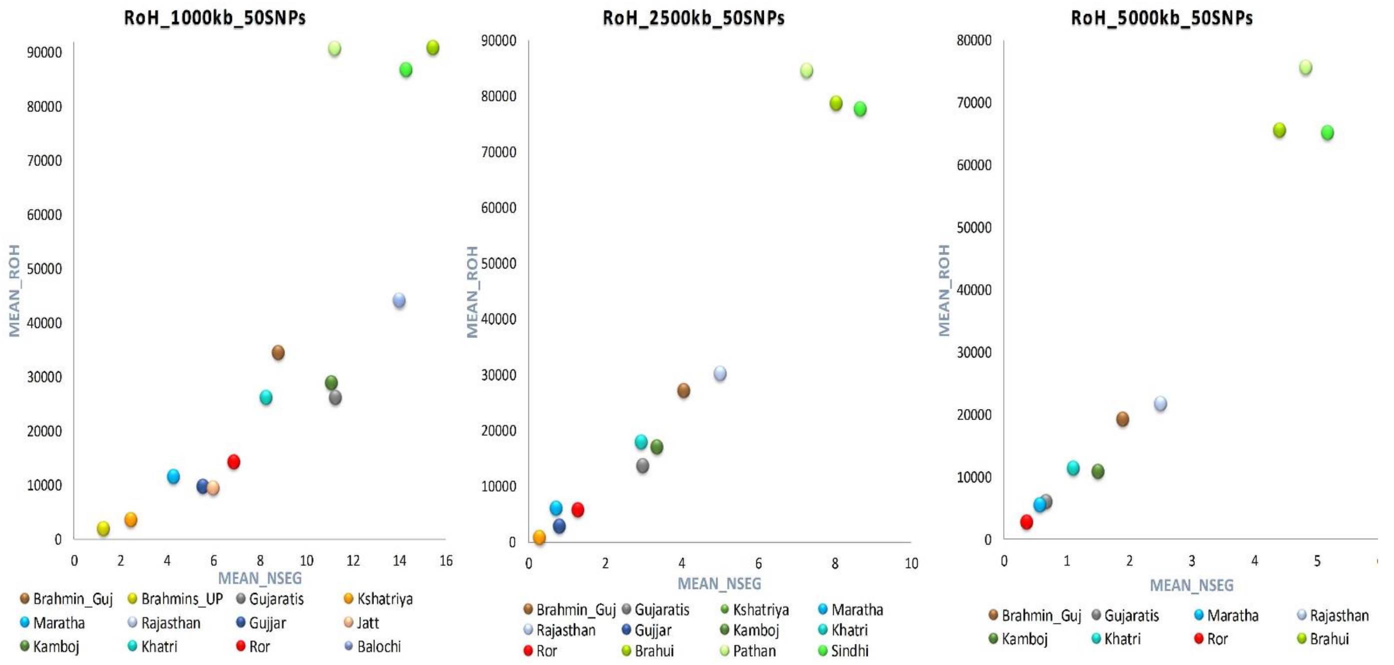


Figure S13. Mean Chunk received by PNWI and NI_IE (A) Chunks received from across Eurasia number of average Chunks received by PNWI and adjoining Gujaratis from across Eurasia and **(B) Chunks shared among PNWI individuals** number of average Chunks shared among PNWI and Gujarati is shown on the right.

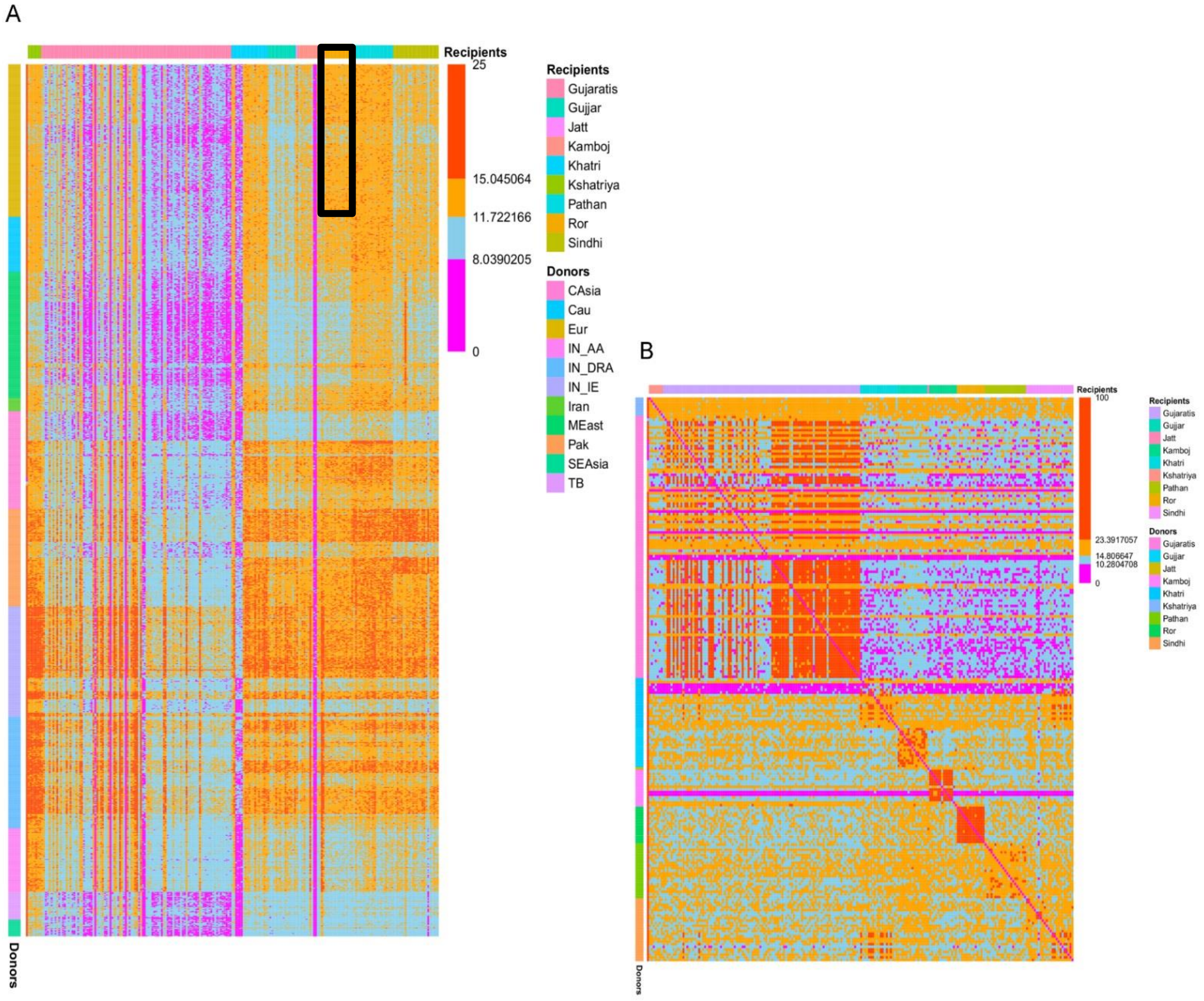


Figure S14. NNLS based Ancestry estimated for PNWI and neighboring populations (A) All populations as donor population-based ancestry estimates for PNWI and neighboring groups from North Indian Gangetic Plain inferred by CHROMOPAINTER using NNLS-based analysis including all populations as donor. **and (B) UNLINKED model** Ancestry estimates for PNWI and neighboring groups from North Indian Gangetic Plain inferred by UNLINKED CHROMOPAINTER run using NNLS-based analysis removing PNWI groups from donors. IN_IE, Indo-Europeans from India; Pak, Pakistan; Eur, Europe; M East, Middle East; Cau, Caucasus; IN_DRA, Indian Dravidians.

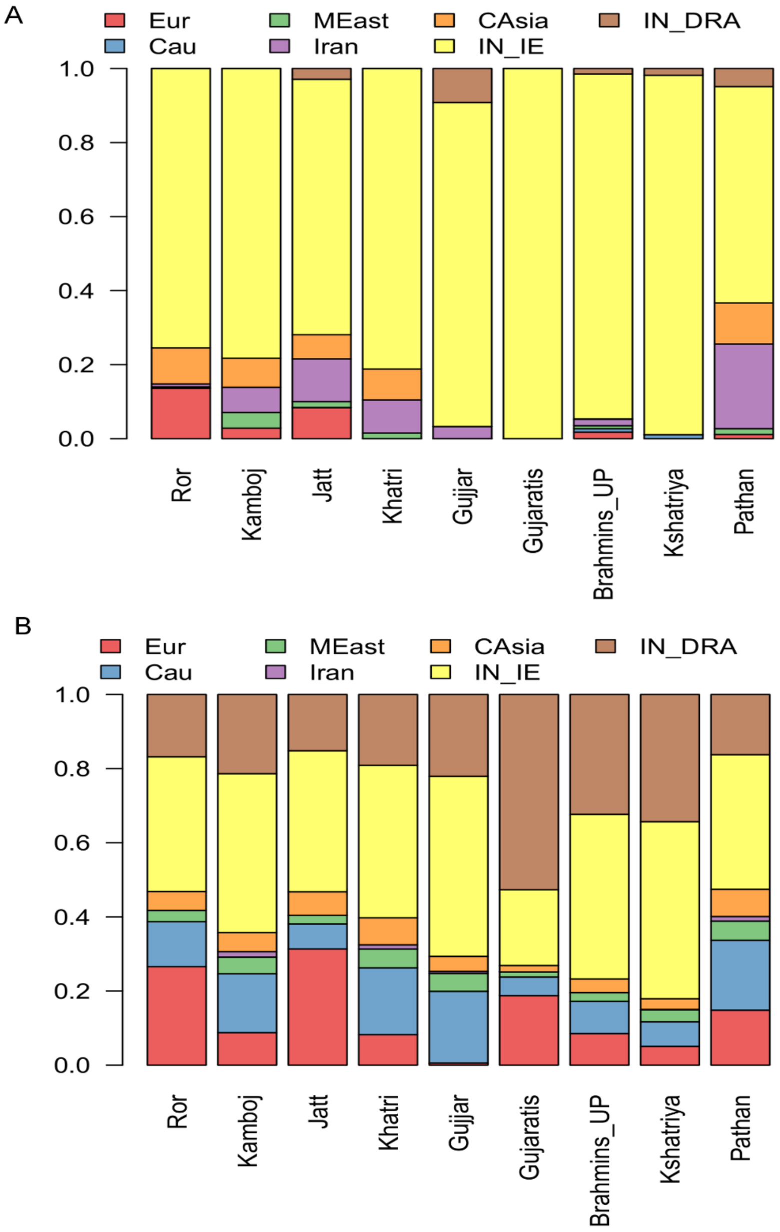
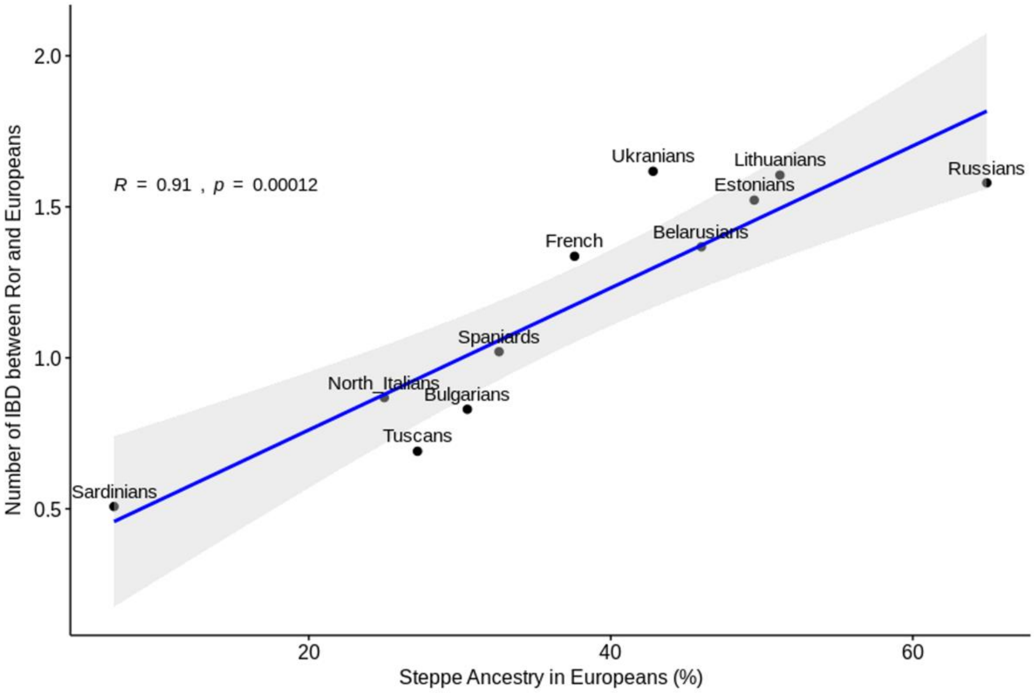


Figure S15. Correlation of IBD sharing of Ror with Steppe Ancestry in Europeans the scatter plot for correlation using “Pearson” method indicating the higher number of IBD segment shared between Ror and Europeans is positively correlated with the increasing Steppe ancestry in modern Europeans. Gray shade shows the 95% Confidence Interval.



Supplemental Tables

Table S1. Details of modern populations and number of overlapping SNPs used for genome-wide data analyses of the study

Separate file: Table_S1.xls

Table S2. Detail information of ancient samples used in various analyses of this study

Separate file: Table_S2.xls

Table S3. Mitochondrial genotyping in sampled Northwest Indians

(A). Haplogroup wise mtDNA frequency

Population	n	D	F	M9	M	M18	M1	M2	M3	M30	M31	M33	M34	M4	M40	M5	M52	M57	M6	R5	U2	N	R	R0	R2	R2'JT	T	H	HV	HV2	I	J	K	U3	U5	U7	U9	W
Gujjar	44	0.068	0.000	0.023	0.068	0.000	0.000	0.000	0.091	0.045	0.000	0.000	0.000	0.000	0.023	0.045	0.045	0.000	0.000	0.045	0.068	0.000	0.000	0.000	0.045	0.045	0.068	0.023	0.000	0.000	0.000	0.068	0.000	0.000	0.023	0.182	0.000	0.023
Ror	104	0.010	0.010	0.019	0.067	0.000	0.000	0.010	0.115	0.125	0.000	0.010	0.010	0.029	0.000	0.125	0.010	0.000	0.000	0.000	0.096	0.010	0.000	0.038	0.000	0.000	0.038	0.010	0.038	0.010	0.000	0.029	0.000	0.010	0.019	0.077	0.010	0.077
Kamboj	42	0.048	0.000	0.024	0.048	0.024	0.000	0.024	0.095	0.024	0.000	0.000	0.000	0.000	0.000	0.095	0.000	0.024	0.024	0.000	0.071	0.048	0.048	0.000	0.000	0.000	0.024	0.048	0.000	0.000	0.024	0.024	0.024	0.000	0.000	0.262	0.000	0.000

(B). Regional mtDNA haplogroups the acronyms we used are: C Asia, Central Asia; E Asia, East Asia; SE Asia, Southeast Asia; S Asia, South Asia; W Eurasia, West Eurasia.

Population	n	C Asia	E Asia + SE Asia	S Asia	W Eurasia
Gujjar	41	0.073	0.024	0.390	0.512
Ror	96	0.010	0.031	0.583	0.375
Kamboj	36	0.056	0.028	0.444	0.472

Table S4. Y-chromosomal genotyping in sampled Northwest Indians

(A). Y haplogroups frequency tabling the occurrence of main haplogroups

Population	n	C-M356	H-M69	R2-M124	L-M11	J2-M172	R1a1-M17	Q-M242
Ror	139	0.06	0.08	0.08	0.29	0.13	0.35	0.01
Kamboj	52	0.04	0.15	0.06	0.17	0.06	0.52	0.00
Gujjar	57	0.00	0.07	0.09	0.63	0.02	0.19	0.00

(B). Regional haplogroups we display the occurrence of major regional Y chromosomal haplogroups in newly sampled individuals

Population	n	S Asian	W Eurasian	R1a	C Asian/ C Siberian
Ror	139	0.52	0.13	0.35	0.01
Kamboj	52	0.42	0.06	0.52	0.00
Gujjar	57	0.79	0.02	0.19	0.00

Table S5. The population wise F_{ST} values based on autosomal dataset of studied populations

Separate file_Table S5.xls

Table S6. Formal test of admixture for PNWI and Gangetic Plain populations we used f_3 *test* in form of f_3 (X, Y; PNWI/NI_IE) where X is related to ancient or modern groups from West Eurasia and Y is a modern South Asian population (Paniya) representing the Ancestral South India (ASI). Steppe_EMBA = Early Bronze Age Steppe, Steppe_MLBA= Middle Late Bronze Age Steppe, Anatolia_N = Neolithic Anatolia, IranTuran_N = Neolithic Iran Turan, Iran_N = Neolithic Iran, BMAC = Bactrian Margiana Complex, CHG = Caucasus Hunter-Gatherers, WHG = Western Hunter-Gatherers, EHG = Eastern Hunter-Gatherers, Indus_Diaspora = Indus_Periphery, SPGT = Iron Age South Asians, SouthAsia_H = Early historical South Asia.

source1	source2	Target	*f3*	std err	Z score
Steppe_MLBA	Paniya	Ror	-0.007603	0.000315	-24.165
Steppe_EMBA	Paniya	Ror	-0.008151	0.000343	-23.768
French	Paniya	Ror	-0.007271	0.000324	-22.475
Anatolia_N	Paniya	Ror	-0.006213	0.000433	-14.353
BMAC	Paniya	Ror	-0.003462	0.000337	-10.269
Iranians	Paniya	Ror	-0.003012	0.000324	-9.296
EHG	Paniya	Ror	-0.00684	0.000756	-9.045
Namazga_CA	Paniya	Ror	-0.004222	0.000576	-7.332
CHG	Paniya	Ror	-0.004933	0.000678	-7.279
WHG	Paniya	Ror	-0.002781	0.000626	-4.439
IranTuran_N	Paniya	Ror	-0.002564	0.000646	-3.968
Iran_N	Paniya	Ror	-0.002897	0.00112	-2.587
Kalash	Paniya	Ror	0.001265	0.000388	3.261
SPGT	Paniya	Ror	0.004528	0.00034	13.334
Indus_Diaspora	Paniya	Ror	0.010129	0.00071	14.271
SouthAsia_H	Paniya	Ror	0.007152	0.000375	19.07
BMAC	Paniya	Gujjar	-0.006989	0.000297	-23.54
Steppe_EMBA	Paniya	Gujjar	-0.007728	0.000334	-23.157
Steppe_MLBA	Paniya	Gujjar	-0.006815	0.000309	-22.084
French	Paniya	Gujjar	-0.007148	0.000336	-21.283
Iranians	Paniya	Gujjar	-0.005664	0.000298	-19.019
Anatolia_N	Paniya	Gujjar	-0.007043	0.000397	-17.72
Namazga_CA	Paniya	Gujjar	-0.008219	0.000565	-14.546
CHG	Paniya	Gujjar	-0.008741	0.00063	-13.874
IranTuran_N	Paniya	Gujjar	-0.008236	0.000624	-13.2
Iran_N	Paniya	Gujjar	-0.008134	0.001101	-7.387
Kalash	Paniya	Gujjar	-0.002188	0.000328	-6.677
EHG	Paniya	Gujjar	-0.004514	0.000757	-5.962
Ror	Paniya	Gujjar	-0.000864	0.000295	-2.926
WHG	Paniya	Gujjar	-0.001178	0.00062	-1.899
SPGT	Paniya	Gujjar	0.000318	0.000295	1.077
Indus_Diaspora	Paniya	Gujjar	0.004152	0.000668	6.219
SouthAsia_H	Paniya	Gujjar	0.002968	0.000325	9.132
Steppe_EMBA	Paniya	Kamboj	-0.005587	0.00035	-15.986
French	Paniya	Kamboj	-0.005505	0.000353	-15.597
Steppe_MLBA	Paniya	Kamboj	-0.004983	0.000325	-15.311
Anatolia_N	Paniya	Kamboj	-0.005769	0.00044	-13.122
BMAC	Paniya	Kamboj	-0.004049	0.000342	-11.827
Iranians	Paniya	Kamboj	-0.00321	0.000333	-9.641
Namazga_CA	Paniya	Kamboj	-0.004907	0.000594	-8.262
CHG	Paniya	Kamboj	-0.00571	0.000697	-8.196
IranTuran_N	Paniya	Kamboj	-0.004751	0.000644	-7.375
EHG	Paniya	Kamboj	-0.003596	0.000748	-4.807
Iran_N	Paniya	Kamboj	-0.004286	0.00111	-3.86
WHG	Paniya	Kamboj	-0.000397	0.00064	-0.62
Kalash	Paniya	Kamboj	0.00067	0.000376	1.778
Ror	Paniya	Kamboj	0.002141	0.000323	6.625
SPGT	Paniya	Kamboj	0.003248	0.000332	9.771
Indus_Diaspora	Paniya	Kamboj	0.007481	0.000697	10.728
SouthAsia_H	Paniya	Kamboj	0.005899	0.000362	16.274

Steppe_EMBA	Paniya	Khatri	-0.007207	0.000311	-23.18
French	Paniya	Khatri	-0.006927	0.000313	-22.153
Steppe_MLBA	Paniya	Khatri	-0.006301	0.000286	-22.037
BMAC	Paniya	Khatri	-0.005825	0.000279	-20.913
Anatolia_N	Paniya	Khatri	-0.006566	0.000403	-16.291
Iranians	Paniya	Khatri	-0.00455	0.000286	-15.924
Namazga_CA	Paniya	Khatri	-0.00673	0.000521	-12.924
CHG	Paniya	Khatri	-0.007331	0.000621	-11.812
IranTuran_N	Paniya	Khatri	-0.006125	0.0006	-10.216
EHG	Paniya	Khatri	-0.004702	0.000722	-6.51
Iran_N	Paniya	Khatri	-0.005892	0.001044	-5.642
Kalash	Paniya	Khatri	-0.000788	0.000322	-2.446
WHG	Paniya	Khatri	-0.00122	0.000591	-2.064
Ror	Paniya	Khatri	0.00115	0.000281	4.095
SPGT	Paniya	Khatri	0.002147	0.000267	8.038
Indus_Diaspora	Paniya	Khatri	0.006635	0.000654	10.144
SouthAsia_H	Paniya	Khatri	0.004834	0.00032	15.103
French	Paniya	Pathan	-0.009491	0.000281	-33.748
Steppe_MLBA	Paniya	Pathan	-0.008505	0.000255	-33.296
Steppe_EMBA	Paniya	Pathan	-0.009326	0.000296	-31.544
BMAC	Paniya	Pathan	-0.007356	0.000265	-27.779
Iranians	Paniya	Pathan	-0.006722	0.00026	-25.874
Anatolia_N	Paniya	Pathan	-0.009425	0.000384	-24.52
Namazga_CA	Paniya	Pathan	-0.008525	0.000538	-15.835
CHG	Paniya	Pathan	-0.009156	0.000592	-15.469
IranTuran_N	Paniya	Pathan	-0.007976	0.00058	-13.748
EHG	Paniya	Pathan	-0.006436	0.000682	-9.441
Iran_N	Paniya	Pathan	-0.008657	0.000986	-8.78
WHG	Paniya	Pathan	-0.003808	0.000556	-6.847
Kalash	Paniya	Pathan	-0.001983	0.000309	-6.426
Ror	Paniya	Pathan	0.000204	0.000262	0.778
SPGT	Paniya	Pathan	0.001166	0.000252	4.621
Indus_Diaspora	Paniya	Pathan	0.006603	0.000647	10.206
SouthAsia_H	Paniya	Pathan	0.004243	0.000309	13.726
BMAC	Paniya	Gujaratis	-0.00697	0.000242	-28.806
Steppe_EMBA	Paniya	Gujaratis	-0.007409	0.000261	-28.413
Steppe_MLBA	Paniya	Gujaratis	-0.006494	0.000239	-27.189
French	Paniya	Gujaratis	-0.007031	0.000262	-26.85
Iranians	Paniya	Gujaratis	-0.005871	0.00024	-24.454
Anatolia_N	Paniya	Gujaratis	-0.006482	0.000325	-19.939
Namazga_CA	Paniya	Gujaratis	-0.008034	0.00044	-18.256
IranTuran_N	Paniya	Gujaratis	-0.007774	0.000485	-16.019
Kalash	Paniya	Gujaratis	-0.004007	0.000252	-15.886
CHG	Paniya	Gujaratis	-0.007501	0.000508	-14.753
Ror	Paniya	Gujaratis	-0.002904	0.000205	-14.183
SPGT	Paniya	Gujaratis	-0.002103	0.000188	-11.214
EHG	Paniya	Gujaratis	-0.005673	0.00059	-9.613
Iran_N	Paniya	Gujaratis	-0.008246	0.000858	-9.609
WHG	Paniya	Gujaratis	-0.002584	0.000485	-5.328
SouthAsia_H	Paniya	Gujaratis	-0.000468	0.00021	-2.229
Indus_Diaspora	Paniya	Gujaratis	-0.000101	0.000497	-0.204
Steppe_EMBA	Paniya	Brahmins_UP	-0.011508	0.000402	-28.612
Steppe_MLBA	Paniya	Brahmins_UP	-0.010736	0.000394	-27.247
French	Paniya	Brahmins_UP	-0.010907	0.000415	-26.294
BMAC	Paniya	Brahmins_UP	-0.009068	0.000406	-22.336
Iranians	Paniya	Brahmins_UP	-0.008516	0.000403	-21.147
Anatolia_N	Paniya	Brahmins_UP	-0.01016	0.000483	-21.049
Namazga_CA	Paniya	Brahmins_UP	-0.009662	0.000663	-14.572
IranTuran_N	Paniya	Brahmins_UP	-0.009571	0.000703	-13.607
Kalash	Paniya	Brahmins_UP	-0.005703	0.000427	-13.354
CHG	Paniya	Brahmins_UP	-0.009513	0.00072	-13.206
Ror	Paniya	Brahmins_UP	-0.004646	0.000386	-12.036
EHG	Paniya	Brahmins_UP	-0.010244	0.000855	-11.982
WHG	Paniya	Brahmins_UP	-0.007495	0.000671	-11.162
SPGT	Paniya	Brahmins_UP	-0.003662	0.000391	-9.37
Iran_N	Paniya	Brahmins_UP	-0.010985	0.001215	-9.038
SouthAsia_H	Paniya	Brahmins_UP	-0.002036	0.00041	-4.971
Indus_Diaspora	Paniya	Brahmins_UP	-0.000647	0.000729	-0.888

Table S7. *D* statistics to test if Ror and Kalash belong to same clade we applied *D* stat in form of *D*(Yoruba,X; Ror, Kalash) where X is a global population.

Yoruba	pop1	pop2	pop3	Dstat	Z score
Yoruba	Ho	Ror	Kalash	-0.0048	-4.722
Yoruba	Onge	Ror	Kalash	-0.0057	-4.711
Yoruba	Gond	Ror	Kalash	-0.0045	-4.554
Yoruba	Paniya	Ror	Kalash	-0.0043	-4.047
Yoruba	Irula	Ror	Kalash	-0.0041	-3.952
Yoruba	Juang	Ror	Kalash	-0.0049	-3.857
Yoruba	Chamar	Ror	Kalash	-0.0034	-3.372
Yoruba	Kol	Ror	Kalash	-0.0026	-2.765
Yoruba	Tharu	Ror	Kalash	-0.0027	-2.649
Yoruba	Dharkars	Ror	Kalash	-0.0025	-2.569
Yoruba	Kshatriya	Ror	Kalash	-0.0024	-2.475
Yoruba	Burmese	Ror	Kalash	-0.0026	-2.353
Yoruba	Kurmi	Ror	Kalash	-0.0035	-2.272
Yoruba	Papuans	Ror	Kalash	-0.0026	-2.122
Yoruba	Naga	Ror	Kalash	-0.0026	-2
Yoruba	Brahmins_UP	Ror	Kalash	-0.0018	-1.887
Yoruba	Gujaratis	Ror	Kalash	-0.0017	-1.874
Yoruba	Japanese	Ror	Kalash	-0.0022	-1.866
Yoruba	Han	Ror	Kalash	-0.0018	-1.567
Yoruba	Estonians	Ror	Kalash	-0.0015	-1.516
Yoruba	Cambodians	Ror	Kalash	-0.0017	-1.458
Yoruba	Gujjar	Ror	Kalash	-0.0013	-1.413
Yoruba	Kurumba	Ror	Kalash	-0.0015	-1.328
Yoruba	Germans	Ror	Kalash	-0.0012	-1.165
Yoruba	French	Ror	Kalash	-0.0011	-1.145
Yoruba	Kamboj	Ror	Kalash	-0.0006	-0.643
Yoruba	Kazakhs	Ror	Kalash	-0.0006	-0.601
Yoruba	Hazara	Ror	Kalash	-0.0001	-0.094
Yoruba	North_Italians	Ror	Kalash	0.0001	0.132
Yoruba	Sindhi	Ror	Kalash	0.0003	0.302
Yoruba	Khatri	Ror	Kalash	0.0003	0.352
Yoruba	Pathan	Ror	Kalash	0.0009	0.978
Yoruba	Palestinians	Ror	Kalash	0.0012	1.404
Yoruba	Balochi	Ror	Kalash	0.0013	1.51
Yoruba	Yaghnobi	Ror	Kalash	0.002	1.852
Yoruba	Burusho	Ror	Kalash	0.0017	1.875
Yoruba	Makrani	Ror	Kalash	0.0019	2.203
Yoruba	Pamiris	Ror	Kalash	0.0021	2.325
Yoruba	Brahui	Ror	Kalash	0.0021	2.408
Yoruba	Armenians	Ror	Kalash	0.0025	2.702
Yoruba	Iranians	Ror	Kalash	0.0025	2.827
Yoruba	Azeris	Ror	Kalash	0.0026	2.878
Yoruba	Georgians	Ror	Kalash	0.0039	4.046

Table S8. Gene flow among PNWI and NI_IE and their relative affinity with Europe *D* statistics in form *D*(pop1, Yoruba; pop2, pop3), where pop1 is any South Asians/French and pop2 is Gujaratis/Pathan/Ror and pop3 is any of modern South Asians, show relative allele sharing of Pakistani, Northwest Indians and North Indians among themselves striking out the Ror's closeness with Pathan and Kalash and their highest affinity with Europe than other South Asians. PNWI = Pakistani and Northwest Indians, NI_IE = North Indian Indo-Europeans from Gangetic Plain, Brahmins_UP = Brahmins from Uttar Pradesh.

pop1	Yoruba	pop2	pop3	Dstat	Z score
Pathan	Yoruba	Gujaratis	Ror	-0.0057	-8.208
Pathan	Yoruba	Gujaratis	Kalash	-0.0066	-8.062
Pathan	Yoruba	Gujaratis	Khatri	-0.0044	-7.596
Pathan	Yoruba	Gujaratis	Gujjar	-0.0030	-5.439
Pathan	Yoruba	Gujaratis	Brahmins_UP	-0.0023	-3.218
Pathan	Yoruba	Gujaratis	Kamboj	-0.0017	-2.777
Pathan	Yoruba	Gujaratis	Jatt	-0.0046	-2.382
Pathan	Yoruba	Gujaratis	Kshatriya	0.0000	0.029
Pathan	Yoruba	Gujaratis	Sindhi	0.0081	14.447
Kalash	Yoruba	Gujaratis	Ror	-0.0083	-10.948
Kalash	Yoruba	Gujaratis	Khatri	-0.0063	-9.982
Kalash	Yoruba	Gujaratis	Pathan	-0.0063	-9.759
Kalash	Yoruba	Gujaratis	Gujjar	-0.0035	-5.673
Kalash	Yoruba	Gujaratis	Kamboj	-0.0032	-4.644
Kalash	Yoruba	Gujaratis	Jatt	-0.0068	-3.187
Kalash	Yoruba	Gujaratis	Brahmins_UP	-0.0023	-2.911
Kalash	Yoruba	Pathan	Ror	-0.002	-2.612
Kalash	Yoruba	Gujaratis	Kshatriya	-0.0003	-0.356
French	Yoruba	Ror	Jatt	0.0013	0.643
French	Yoruba	Ror	Kalash	0.0011	1.145
French	Yoruba	Ror	Pathan	0.0041	5.242
French	Yoruba	Ror	Khatri	0.0076	9.532
French	Yoruba	Ror	Kamboj	0.0102	11.075
French	Yoruba	Ror	Burusho	0.0108	13.669
French	Yoruba	Ror	Brahmins_UP	0.014	14.326
French	Yoruba	Ror	Gujjar	0.0129	14.994
French	Yoruba	Ror	Brahui	0.0131	15.605
French	Yoruba	Ror	Kshatriya	0.0182	17.795
French	Yoruba	Ror	Muslim	0.0227	19.351
French	Yoruba	Ror	Sindhi	0.0203	26.019
French	Yoruba	Ror	Makrani	0.0238	26.793
French	Yoruba	Ror	Gujaratis	0.0213	27.305
French	Yoruba	Ror	Maratha	0.0335	27.42
French	Yoruba	Ror	Chamar	0.0348	29.353
French	Yoruba	Ror	Irula	0.0469	35.408
French	Yoruba	Ror	Ho	0.0547	41.695

Table S9. *D* statistics for relative allele sharing of South Asians with a set of distal ancient sources we used *D* stat in form *D*(pop1, Yoruba; pop2, pop3), where pop1 is any South Asians and pop2 is Steppe_EMBA/Anatolia_N and pop3 is any of modern South Asians/Iran_N/EHG/Levant_N, highlighting Ror as the closest South Asian to Steppe_EMBA and Anatolia_N.

pop1	Yoruba	pop2	pop3	Dstat	Z score
Jatt	Yoruba	Steppe_EMBA	Iran_N	0.0045	7.314
Paniya	Yoruba	Steppe_EMBA	Iran_N	0.0032	8.369
Sindhi	Yoruba	Steppe_EMBA	Iran_N	0.0028	8.434
Gujjar	Yoruba	Steppe_EMBA	Iran_N	0.0031	8.905
Chamar	Yoruba	Steppe_EMBA	Iran_N	0.0033	9.181
Kshatriya	Yoruba	Steppe_EMBA	Iran_N	0.0036	9.392
Ho	Yoruba	Steppe_EMBA	Iran_N	0.0035	9.558
Gujaratis	Yoruba	Steppe_EMBA	Iran_N	0.0031	9.609
Pathan	Yoruba	Steppe_EMBA	Iran_N	0.0033	9.872
Khatri	Yoruba	Steppe_EMBA	Iran_N	0.0035	10.039
Brahmins_UP	Yoruba	Steppe_EMBA	Iran_N	0.0040	10.525
Kamboj	Yoruba	Steppe_EMBA	Iran_N	0.0038	10.789
Kalash	Yoruba	Steppe_EMBA	Iran_N	0.0040	11.209
Ror	Yoruba	Steppe_EMBA	Iran_N	0.0050	14.259
Ho	Yoruba	Steppe_EMBA	EHG	-0.0014	-4.948
Paniya	Yoruba	Steppe_EMBA	EHG	-0.0011	-3.508
Chamar	Yoruba	Steppe_EMBA	EHG	-0.0009	-2.866
Jatt	Yoruba	Steppe_EMBA	EHG	-0.0013	-2.499
Kshatriya	Yoruba	Steppe_EMBA	EHG	-0.0006	-1.967
Ror	Yoruba	Steppe_EMBA	EHG	-0.0005	-1.804
Gujaratis	Yoruba	Steppe_EMBA	EHG	-0.0004	-1.444
Brahmins_UP	Yoruba	Steppe_EMBA	EHG	-0.0003	-1.019
Kalash	Yoruba	Steppe_EMBA	EHG	-0.0003	-0.859
Kamboj	Yoruba	Steppe_EMBA	EHG	-0.0002	-0.729
Gujjar	Yoruba	Steppe_EMBA	EHG	0.0000	-0.152
Sindhi	Yoruba	Steppe_EMBA	EHG	0.0000	-0.036
Pathan	Yoruba	Steppe_EMBA	EHG	0.0000	0.079
Khatri	Yoruba	Steppe_EMBA	EHG	0.0000	0.127
Gujaratis	Yoruba	Anatolia_N	Iran_N	0.0000	-0.064
Gujjar	Yoruba	Anatolia_N	Iran_N	0.0001	0.237
Sindhi	Yoruba	Anatolia_N	Iran_N	0.0001	0.412
Chamar	Yoruba	Anatolia_N	Iran_N	0.0002	0.523
Kshatriya	Yoruba	Anatolia_N	Iran_N	0.0002	0.594
Paniya	Yoruba	Anatolia_N	Iran_N	0.0003	0.68
Khatri	Yoruba	Anatolia_N	Iran_N	0.0003	0.906
Kalash	Yoruba	Anatolia_N	Iran_N	0.0006	1.49
Jatt	Yoruba	Anatolia_N	Iran_N	0.0010	1.642
Pathan	Yoruba	Anatolia_N	Iran_N	0.0006	1.725
Brahmins_UP	Yoruba	Anatolia_N	Iran_N	0.0007	1.834
Ho	Yoruba	Anatolia_N	Iran_N	0.0007	1.944
Kamboj	Yoruba	Anatolia_N	Iran_N	0.0010	2.732
Ror	Yoruba	Anatolia_N	Iran_N	0.0017	4.572

Ho	Yoruba	Anatolia_N	EHG	-0.0042	-12.203
Ror	Yoruba	Anatolia_N	EHG	-0.0039	-12.066
Gujaratis	Yoruba	Anatolia_N	EHG	-0.0036	-11.83
Chamar	Yoruba	Anatolia_N	EHG	-0.0039	-11.598
Paniya	Yoruba	Anatolia_N	EHG	-0.0041	-11.392
Kshatriya	Yoruba	Anatolia_N	EHG	-0.0040	-11.196
Kalash	Yoruba	Anatolia_N	EHG	-0.0037	-10.796
Brahmins_UP	Yoruba	Anatolia_N	EHG	-0.0036	-10.491
Gujjar	Yoruba	Anatolia_N	EHG	-0.0031	-9.882
Khatri	Yoruba	Anatolia_N	EHG	-0.0031	-9.532
Kamboj	Yoruba	Anatolia_N	EHG	-0.0031	-9.125
Sindhi	Yoruba	Anatolia_N	EHG	-0.0027	-8.874
Pathan	Yoruba	Anatolia_N	EHG	-0.0028	-8.642
Jatt	Yoruba	Anatolia_N	EHG	-0.0048	-7.727
Jatt	Yoruba	Anatolia_N	Levant_N	0.0032	7.098
Ho	Yoruba	Anatolia_N	Levant_N	0.0019	7.185
Paniya	Yoruba	Anatolia_N	Levant_N	0.0021	7.328
Chamar	Yoruba	Anatolia_N	Levant_N	0.0023	8.122
Kshatriya	Yoruba	Anatolia_N	Levant_N	0.0026	9.168
Brahmins_UP	Yoruba	Anatolia_N	Levant_N	0.0028	10.154
Gujaratis	Yoruba	Anatolia_N	Levant_N	0.0026	10.552
Sindhi	Yoruba	Anatolia_N	Levant_N	0.0025	10.554
Gujjar	Yoruba	Anatolia_N	Levant_N	0.0028	10.7
Pathan	Yoruba	Anatolia_N	Levant_N	0.0028	11.051
Kamboj	Yoruba	Anatolia_N	Levant_N	0.0029	11.204
Ror	Yoruba	Anatolia_N	Levant_N	0.0031	11.289
Kalash	Yoruba	Anatolia_N	Levant_N	0.0031	11.541
Khatri	Yoruba	Anatolia_N	Levant_N	0.0030	11.765

Table S10. *D* statistics for relative allele sharing of South Asians with a set of proximal ancient sources we computed *D stat* in form of *D* (pop1, Yoruba; pop2, pop3), where pop1 is South Asians, pop2 is different aDNA sources and pop3 refers to aDNA samples or modern Dravidian displaying relative allele sharing of South Asians with neighboring (proximal) ancient sources from South Asia and Central Asia.

Separate file_Table S10.xls

Table S11. *qpAdm* with Steppe_MLBA we used *qpAdm* to deduce ancestry proportion from distal ancient sources (Iran_N, Steppe_MLBA and Onge) among South Asians. Only populations with p value > 0.5 have been reported leaving out Kalash, SPGT, Burusho, Hazara, IN_AA and IN_TB (p value < 0.5).

Target Pop	p value	Mix coeff1	Mix coeff2	Mix coeff3	std err1	std err2	std err3	Source1	Source2	Source3
Ror	0.295723	0.179	0.621	0.2	0.028	0.024	0.014	Iran_N	Steppe_MLBA	Onge
Khatri	0.0898528	0.198	0.553	0.249	0.03	0.025	0.015	Iran_N	Steppe_MLBA	Onge
Gujjar	0.266217	0.214	0.492	0.295	0.032	0.028	0.016	Iran_N	Steppe_MLBA	Onge
Kamboj	0.124754	0.253	0.485	0.263	0.032	0.027	0.017	Iran_N	Steppe_MLBA	Onge
Pathan	0.334088	0.236	0.547	0.217	0.027	0.024	0.014	Iran_N	Steppe_MLBA	Onge
Jatt	0.258625	0.084	0.703	0.213	0.08	0.07	0.038	Iran_N	Steppe_MLBA	Onge
Sindhi	0.116182	0.51	0.291	0.199	0.04	0.034	0.019	Iran_N	Steppe_MLBA	Onge
Balochi	0.777542	0.583	0.287	0.13	0.043	0.037	0.021	Iran_N	Steppe_MLBA	Onge
Brahui	0.355776	0.629	0.269	0.102	0.046	0.04	0.022	Iran_N	Steppe_MLBA	Onge
Makrani	0.47551	1	NA	NA	0.062	NA	NA	Iran_N	Steppe_MLBA	Onge
Brahmin_Guj	0.148345	0.211	0.489	0.3	0.03	0.025	0.015	Iran_N	Steppe_MLBA	Onge
Gujaratis	0.0905237	0.171	0.443	0.385	0.031	0.026	0.015	Iran_N	Steppe_MLBA	Onge
Brahmins_UP	0.220679	0.142	0.51	0.348	0.036	0.031	0.018	Iran_N	Steppe_MLBA	Onge
Kshatriya	0.173372	0.175	0.464	0.361	0.037	0.031	0.018	Iran_N	Steppe_MLBA	Onge
Chamar	0.0602319	0.076	0.357	0.567	0.039	0.035	0.02	Iran_N	Steppe_MLBA	Onge
Kanjars	0.734447	0.182	0.373	0.445	0.038	0.033	0.02	Iran_N	Steppe_MLBA	Onge
Kurmi	0.291761	0.083	0.51	0.408	0.076	0.067	0.04	Iran_N	Steppe_MLBA	Onge
Dharkars	0.53426	0.099	0.463	0.437	0.035	0.03	0.018	Iran_N	Steppe_MLBA	Onge
Muslim	0.0880746	0.108	0.443	0.449	0.046	0.04	0.022	Iran_N	Steppe_MLBA	Onge
Maratha	0.50061	0.139	0.336	0.525	0.04	0.036	0.02	Iran_N	Steppe_MLBA	Onge
Meena	0.256096	0.16	0.512	0.328	0.078	0.07	0.038	Iran_N	Steppe_MLBA	Onge
Meghawal	0.0920744	0.183	0.339	0.479	0.084	0.07	0.04	Iran_N	Steppe_MLBA	Onge
Kol	0.260324	0.063	0.408	0.528	0.034	0.031	0.017	Iran_N	Steppe_MLBA	Onge
Gond	0.210592	NA	0.28	0.72	NA	0.035	0.021	Iran_N	Steppe_MLBA	Onge
Kurumba	0.445845	0.227	0.3	0.473	0.048	0.042	0.025	Iran_N	Steppe_MLBA	Onge
Paniya	0.309749	0.074	0.162	0.764	0.044	0.039	0.023	Iran_N	Steppe_MLBA	Onge
Irula	0.138575	0.136	0.198	0.667	0.045	0.038	0.023	Iran_N	Steppe_MLBA	Onge
SouthAsia_H	0.629248	0.165	0.542	0.293	0.03	0.027	0.015	Iran_N	Steppe_MLBA	Onge
BMAC	0.53134	0.544	0.407	0.048	0.04	0.035	0.019	Iran_N	Steppe_MLBA	Onge
Namazga_CA	0.703087	0.588	0.412	NA	0.069	0.06	NA	Iran_N	Steppe_MLBA	Onge
Indus_Diaspora	0.0914973	0.305	0.378	0.316	0.064	0.056	0.03	Iran_N	Steppe_MLBA	Onge
Turkmenistan_IA	0.804688	0.307	0.648	0.045	0.072	0.066	0.035	Iran_N	Steppe_MLBA	Onge
IranTuran_BA	0.171678	0.52	0.445	0.035	0.046	0.04	0.022	Iran_N	Steppe_MLBA	Onge

Table S12. *qpAdm* with Steppe_EMBA we used *qpAdm* to deduce ancestry proportion from distal ancient sources (Iran_N, Steppe_EMBA and Onge) among South Asians. Only populations with p value > 0.5 have been reported leaving out Burusho, Hazara, IN_AA and IN_TB (p value < 0.5).

Target Pop	p value	Mix coeff1	Mix coeff2	Mix coeff3	std err1	std err2	std err3	Source1	Source2	Source3
Ror	0.862682	0.244	0.544	0.212	0.027	0.022	0.015	Iran_N	Steppe_EMBA	Onge
Kalash	0.627904	0.228	0.583	0.189	0.031	0.026	0.017	Iran_N	Steppe_EMBA	Onge
Khatri	0.185302	0.261	0.483	0.256	0.029	0.024	0.016	Iran_N	Steppe_EMBA	Onge
Gujjar	0.624598	0.269	0.429	0.302	0.03	0.025	0.017	Iran_N	Steppe_EMBA	Onge
Kamboj	0.215046	0.306	0.424	0.27	0.031	0.025	0.018	Iran_N	Steppe_EMBA	Onge
Pathan	0.242626	0.3	0.476	0.225	0.029	0.023	0.016	Iran_N	Steppe_EMBA	Onge
Jatt	0.380243	0.172	0.611	0.218	0.071	0.06	0.038	Iran_N	Steppe_EMBA	Onge
Sindhi	0.348933	0.532	0.263	0.205	0.037	0.03	0.02	Iran_N	Steppe_EMBA	Onge
Balochi	0.726245	0.617	0.249	0.135	0.042	0.034	0.022	Iran_N	Steppe_EMBA	Onge
Brahui	0.432531	0.658	0.236	0.106	0.044	0.036	0.023	Iran_N	Steppe_EMBA	Onge
Makrani	0.55936	0.947	0.053	NA	0.057	0.047	NA	Iran_N	Steppe_EMBA	Onge
Brahmin_Guj	0.577213	0.264	0.428	0.307	0.028	0.023	0.015	Iran_N	Steppe_EMBA	Onge
Gujaratis	0.833375	0.218	0.389	0.393	0.028	0.022	0.015	Iran_N	Steppe_EMBA	Onge
Brahmins_UP	0.487651	0.2	0.445	0.355	0.033	0.027	0.019	Iran_N	Steppe_EMBA	Onge
Kshatriya	0.693484	0.224	0.409	0.367	0.034	0.028	0.018	Iran_N	Steppe_EMBA	Onge
Chamar	0.512026	0.113	0.319	0.569	0.034	0.029	0.02	Iran_N	Steppe_EMBA	Onge
Kanjars	0.956662	0.223	0.327	0.45	0.036	0.03	0.021	Iran_N	Steppe_EMBA	Onge
Kurmi	0.141331	0.155	0.433	0.411	0.068	0.058	0.04	Iran_N	Steppe_EMBA	Onge
Dharkars	0.184288	0.159	0.397	0.444	0.034	0.027	0.019	Iran_N	Steppe_EMBA	Onge
Muslim	0.4431	0.158	0.39	0.453	0.04	0.033	0.022	Iran_N	Steppe_EMBA	Onge
Maratha	0.779805	0.178	0.292	0.53	0.036	0.031	0.02	Iran_N	Steppe_EMBA	Onge
Meena	0.497844	0.211	0.454	0.335	0.071	0.061	0.038	Iran_N	Steppe_EMBA	Onge
Meghawal	0.140849	0.221	0.298	0.481	0.075	0.06	0.041	Iran_N	Steppe_EMBA	Onge
Kol	0.556904	0.114	0.353	0.533	0.031	0.026	0.017	Iran_N	Steppe_EMBA	Onge
Gond	0.178967	NA	0.264	0.736	NA	0.03	0.021	Iran_N	Steppe_EMBA	Onge
Kurumba	0.75005	0.256	0.268	0.476	0.044	0.037	0.025	Iran_N	Steppe_EMBA	Onge
Paniya	0.420394	0.092	0.143	0.765	0.04	0.033	0.023	Iran_N	Steppe_EMBA	Onge
Irula	0.294562	0.154	0.177	0.669	0.04	0.032	0.023	Iran_N	Steppe_EMBA	Onge
SPGT	0.768161	0.271	0.496	0.233	0.027	0.022	0.015	Iran_N	Steppe_EMBA	Onge
SouthAsia_H	0.949565	0.23	0.47	0.3	0.028	0.025	0.016	Iran_N	Steppe_EMBA	Onge
BMAC	0.264358	0.599	0.348	0.053	0.04	0.033	0.021	Iran_N	Steppe_EMBA	Onge
Indus_Diaspora	0.320086	0.336	0.341	0.323	0.058	0.049	0.03	Iran_N	Steppe_EMBA	Onge
Turkmenistan_IA	0.907793	0.383	0.567	0.05	0.067	0.06	0.036	Iran_N	Steppe_EMBA	Onge
IranTuran_BA	0.238759	0.57	0.387	0.042	0.043	0.037	0.024	Iran_N	Steppe_EMBA	Onge
Namazga_CA	0.684163	0.64	0.36	NA	0.065	0.054	NA	Iran_N	Steppe_EMBA	Onge

Table S13. *qpAdm* with both Steppe_EMBA and Steppe_MLBA we used *qpAdm* to deduce ancestry proportion from distal ancient sources (Iran_N, Steppe_EMBA, Steppe_MLBA and Onge) among South Asians. Only populations with p value > 0.5 have been reported leaving out Burusho, Hazara, BMAC, IN_AA and IN_TB (p value < 0.5).

Target Pop	p value	Mix coeff1	Mix coeff2	Mix coeff3	Mix coeff4	std err1	std err2	std err3	std err4	Source1	Source2	Source3	Source4
Makrani	0.554654	0.968	NA	NA	0.032	0.067	NA	NA	0.028	Iran_N	Steppe_EMBA	Steppe_MLBA	Onge
Brahui	0.484494	0.629	NA	0.265	0.106	0.05	NA	0.122	0.02	Iran_N	Steppe_EMBA	Steppe_MLBA	Onge
Balochi	0.133865	0.672	0.239	NA	0.089	0.045	0.099	NA	0.019	Iran_N	Steppe_EMBA	Steppe_MLBA	Onge
Sindhi	0.426382	0.523	0.262	NA	0.215	0.045	0.101	NA	0.019	Iran_N	Steppe_EMBA	Steppe_MLBA	Onge
Pathan	0.492786	0.26	0.205	0.314	0.221	0.03	0.069	0.075	0.013	Iran_N	Steppe_EMBA	Steppe_MLBA	Onge
Kalash	0.599629	0.215	0.486	0.112	0.187	0.036	0.087	0.093	0.016	Iran_N	Steppe_EMBA	Steppe_MLBA	Onge
Ror	0.950983	0.218	0.351	0.224	0.208	0.03	0.07	0.076	0.013	Iran_N	Steppe_EMBA	Steppe_MLBA	Onge
Khatri	0.236179	0.236	0.317	0.192	0.254	0.033	0.073	0.08	0.014	Iran_N	Steppe_EMBA	Steppe_MLBA	Onge
Gujjar	0.616238	0.248	0.274	0.18	0.298	0.035	0.079	0.086	0.015	Iran_N	Steppe_EMBA	Steppe_MLBA	Onge
Kamboj	0.235878	0.281	0.236	0.218	0.266	0.036	0.081	0.09	0.016	Iran_N	Steppe_EMBA	Steppe_MLBA	Onge
Jatt	0.381769	0.163	0.554	0.066	0.217	0.085	0.201	0.218	0.036	Iran_N	Steppe_EMBA	Steppe_MLBA	Onge
Gujaratis	0.917428	0.218	0.389	NA	0.393	0.033	0.068	NA	0.014	Iran_N	Steppe_EMBA	Steppe_MLBA	Onge
Brahmins_UP	0.542499	0.21	0.446	NA	0.344	0.039	0.089	NA	0.017	Iran_N	Steppe_EMBA	Steppe_MLBA	Onge
Kshatriya	0.783016	0.23	0.409	NA	0.361	0.04	0.09	NA	0.017	Iran_N	Steppe_EMBA	Steppe_MLBA	Onge
Chamar	0.620047	0.107	0.317	NA	0.577	0.041	0.101	NA	0.018	Iran_N	Steppe_EMBA	Steppe_MLBA	Onge
Dharkars	0.268219	0.153	0.396	NA	0.45	0.042	0.093	NA	0.018	Iran_N	Steppe_EMBA	Steppe_MLBA	Onge
Muslim	0.476656	0.17	0.392	NA	0.438	0.047	0.103	NA	0.02	Iran_N	Steppe_EMBA	Steppe_MLBA	Onge
Maratha	0.875488	0.181	0.292	NA	0.526	0.044	0.102	NA	0.019	Iran_N	Steppe_EMBA	Steppe_MLBA	Onge
Meena	0.638788	0.215	0.455	NA	0.33	0.081	0.198	NA	0.034	Iran_N	Steppe_EMBA	Steppe_MLBA	Onge
Meghawal	0.181145	0.245	0.3	NA	0.455	0.094	0.211	NA	0.04	Iran_N	Steppe_EMBA	Steppe_MLBA	Onge
Kol	0.697869	0.115	0.353	NA	0.532	0.037	0.085	NA	0.016	Iran_N	Steppe_EMBA	Steppe_MLBA	Onge
Gond	0.217194	NA	0.253	NA	0.747	NA	0.102	NA	0.02	Iran_N	Steppe_EMBA	Steppe_MLBA	Onge
Kurumba	0.777743	0.243	0.266	NA	0.491	0.052	0.125	NA	0.023	Iran_N	Steppe_EMBA	Steppe_MLBA	Onge
Paniya	0.561081	0.094	0.144	NA	0.762	0.049	0.113	NA	0.022	Iran_N	Steppe_EMBA	Steppe_MLBA	Onge
Irula	0.295332	0.134	0.176	NA	0.69	0.049	0.106	NA	0.021	Iran_N	Steppe_EMBA	Steppe_MLBA	Onge
Namazga_CA	0.762098	0.642	0.358	NA	NA	0.079	0.196	NA	NA	Iran_N	Steppe_EMBA	Steppe_MLBA	Onge
Indus_Diaspora	0.386398	0.354	0.34	NA	0.306	0.071	0.169	NA	0.027	Iran_N	Steppe_EMBA	Steppe_MLBA	Onge
SPGT	0.177337	0.238	0.18	0.363	0.219	0.031	0.074	0.083	0.013	Iran_N	Steppe_EMBA	Steppe_MLBA	Onge
SouthAsia_H	0.884903	0.185	0.144	0.381	0.291	0.032	0.08	0.088	0.014	Iran_N	Steppe_EMBA	Steppe_MLBA	Onge
IranTuran_BA	0.0543647	0.531	NA	0.469	NA	0.051	NA	0.14	NA	Iran_N	Steppe_EMBA	Steppe_MLBA	Onge
Turkmenistan_IA	0.765259	0.334	NA	0.666	NA	0.08	NA	0.234	NA	Iran_N	Steppe_EMBA	Steppe_MLBA	Onge

Table S14. *qpAdm* using proximal model with Namazga_CA we used *qpAdm* to deduce ancestry proportion from distal ancient sources (Namazga_CA, Steppe_MLBA and Onge) among South Asians. Only populations with p value > 0.5 have been reported.

Target pop	p value	Mix coeff1	Mix coeff2	Mix coeff3	std err1	std err2	std err3	Source1	Source2	Source3
Ror	0.1103	0.388	**0.436**	0.176	0.103	0.066	0.042	Namazga_CA	Steppe_MLBA	Onge
Khatri	0.1128	0.578	0.244	0.178	0.078	0.047	0.038	Namazga_CA	Steppe_MLBA	Onge
Gujjar	0.4164	0.529	0.244	0.227	0.079	0.048	0.039	Namazga_CA	Steppe_MLBA	Onge
Kamboj	0.1433	0.54	0.243	0.217	0.093	0.055	0.045	Namazga_CA	Steppe_MLBA	Onge
Pathan	0.608	0.589	0.239	0.172	0.071	0.045	0.033	Namazga_CA	Steppe_MLBA	Onge
Sindhi	0.9338	0.58	0.218	0.202	0.066	0.042	0.031	Namazga_CA	Steppe_MLBA	Onge
Brahui	0.1223	0.699	0.179	0.122	0.072	0.049	0.031	Namazga_CA	Steppe_MLBA	Onge
Brahmins_UP	0.054	0.309	0.335	0.356	0.146	0.08	0.072	Namazga_CA	Steppe_MLBA	Onge
Chamar	0.0996	0.122	0.268	0.61	0.087	0.054	0.048	Namazga_CA	Steppe_MLBA	Onge
Kanjars	0.0865	0.262	0.289	0.449	0.123	0.067	0.065	Namazga_CA	Steppe_MLBA	Onge
Kurmi	0.4326	0.588	0.131	0.281	0.201	0.103	0.116	Namazga_CA	Steppe_MLBA	Onge
Maratha	0.3011	0.135	0.308	0.557	0.095	0.059	0.051	Namazga_CA	Steppe_MLBA	Onge
Meena	0.8675	0.333	0.301	0.365	0.191	0.114	0.098	Namazga_CA	Steppe_MLBA	Onge
Meghawal	0.2495	NA	0.288	0.712	NA	0.099	0.115	Namazga_CA	Steppe_MLBA	Onge
Gond	0.2566	NA	0.183	0.817	NA	0.047	0.047	Namazga_CA	Steppe_MLBA	Onge
Kurumba	0.0888	0.246	0.234	0.52	0.125	0.068	0.073	Namazga_CA	Steppe_MLBA	Onge
Paniya	0.3473	NA	0.185	0.815	NA	0.052	0.049	Namazga_CA	Steppe_MLBA	Onge
Irula	0.5916	0.129	0.159	0.713	0.076	0.048	0.047	Namazga_CA	Steppe_MLBA	Onge
Ho	0.0701	NA	0.111	0.889	NA	0.045	0.048	Namazga_CA	Steppe_MLBA	Onge
Asur	0.3203	NA	NA	1	NA	NA	0.072	Namazga_CA	Steppe_MLBA	Onge
Juang	0.1246	NA	NA	1	NA	NA	0.079	Namazga_CA	Steppe_MLBA	Onge
SouthAsia_H	0.5284	0.644	0.159	0.197	0.076	0.047	0.038	Namazga_CA	Steppe_MLBA	Onge

Table S15. *qpAdm* using proximal model with Indus_Periphery (Indus_Diaspora) we used *qpAdm* to deduce ancestry proportion from distal ancient sources (Indus_Diaspora, Steppe_MLBA and Onge) among South Asians. Only populations with p value > 0.5 have been reported.

Target pop	p value	Mix coeff1	Mix coeff2	Mix coeff3	std err1	std err2	std err3	Source1	Source2	Source3
Ror	0.641608	0.468	0.473	0.059	0.092	0.056	0.04	Indus_Diaspora	Steppe_MLBA	Onge
Kalash	0.386132	0.555	0.445	NA	0.129	0.078	NA	Indus_Diaspora	Steppe_MLBA	Onge
Khatri	0.072829	0.765	0.235	NA	0.13	0.078	NA	Indus_Diaspora	Steppe_MLBA	Onge
Gujjar	0.148636	0.717	0.214	0.069	0.131	0.08	0.056	Indus_Diaspora	Steppe_MLBA	Onge
Kamboj	0.130149	0.85	0.15	NA	0.134	0.081	NA	Indus_Diaspora	Steppe_MLBA	Onge
Pathan	0.076498	0.71	0.29	NA	0.12	0.073	NA	Indus_Diaspora	Steppe_MLBA	Onge
Jatt	0.37967	0.601	0.399	NA	0.232	0.14	NA	Indus_Diaspora	Steppe_MLBA	Onge
Brahmin_Guj	0.138981	0.675	0.234	0.091	0.129	0.078	0.055	Indus_Diaspora	Steppe_MLBA	Onge
Gujaratis	0.61345	0.56	0.214	0.226	0.109	0.065	0.047	Indus_Diaspora	Steppe_MLBA	Onge
Brahmins_UP	0.456242	0.429	0.34	0.23	0.141	0.083	0.062	Indus_Diaspora	Steppe_MLBA	Onge
Kshatriya	0.90496	0.553	0.247	0.2	0.123	0.074	0.054	Indus_Diaspora	Steppe_MLBA	Onge
Chamar	0.261509	0.307	0.218	0.475	0.167	0.097	0.072	Indus_Diaspora	Steppe_MLBA	Onge
Kanjars	0.518784	0.608	0.143	0.249	0.152	0.091	0.066	Indus_Diaspora	Steppe_MLBA	Onge
Kurmi	0.425804	0.199	0.385	0.416	0.269	0.161	0.115	Indus_Diaspora	Steppe_MLBA	Onge
Dharkars	0.165794	0.25	0.351	0.399	0.155	0.091	0.067	Indus_Diaspora	Steppe_MLBA	Onge
Muslim	0.662533	0.428	0.232	0.341	0.15	0.088	0.065	Indus_Diaspora	Steppe_MLBA	Onge
Maratha	0.742817	0.473	0.149	0.378	0.149	0.089	0.063	Indus_Diaspora	Steppe_MLBA	Onge
Meena	0.324146	0.445	0.313	0.242	0.304	0.182	0.128	Indus_Diaspora	Steppe_MLBA	Onge
Meghawal	0.555112	0.855	NA	0.146	0.285	NA	0.122	Indus_Diaspora	Steppe_MLBA	Onge
Kol	0.621467	0.256	0.267	0.477	0.123	0.073	0.053	Indus_Diaspora	Steppe_MLBA	Onge
Kurumba	0.688289	0.801	NA	0.199	0.179	NA	0.077	Indus_Diaspora	Steppe_MLBA	Onge
Paniya	0.74176	0.389	NA	0.611	0.171	NA	0.075	Indus_Diaspora	Steppe_MLBA	Onge
Irula	0.869405	0.476	NA	0.524	0.157	NA	0.068	Indus_Diaspora	Steppe_MLBA	Onge
SPGT	0.8029	0.616	0.338	0.046	0.098	0.06	0.042	Indus_Diaspora	Steppe_MLBA	Onge
SouthAsia_H	0.710365	0.469	0.363	0.168	0.105	0.063	0.045	Indus_Diaspora	Steppe_MLBA	Onge

Table S16. *qpWave* results

(A): *qpWave* p value for all models we used *qpWave* for various modellings of South Asian populations indicating the differentiation of sources by the listed outgroups.

p value	f4 rank	source1	source2	source3	source4	Outgrp1	Outgrp2	Outgrp3	Outgrp4	Outgrp5	Outgrp6	Outgrp7	Outgrp8	
4.13E-44	1	Iran_N	Steppe_MLBA	Onge	NA	Ust_Ishim	MA1	Kostenki14	Papuans	Chukchis	Karitiana	Mbuti_Pygmies	NA	distal model1
7.60E-52	1	Iran_N	Steppe_EMBA	Onge	NA	Ust_Ishim	MA1	Kostenki14	Papuans	Chukchis	Karitiana	Mbuti_Pygmies	NA	distal model2
2.52E-25	2	Iran_N	Steppe_MLBA	Onge	Steppe_EMBA	Ust_Ishim	MA1	Kostenki14	Papuans	Chukchis	Karitiana	Mbuti_Pygmies	Anatolia_N	distal model3
3.956E-10	1	Namazga_CA	Steppe_MLBA	Onge	NA	Ust_Ishim	MA1	Kostenki14	Switzerland_HG	Clovis	Natufian	NA	NA	proximal mode2
1.899E-08	1	Indus_Diaspora	Steppe_MLBA	Onge	NA	Ust_Ishim	MA1	Kostenki14	Dai	Chukchis	Karitiana	Mbuti_Pygmies	NA	proximal model2

(B): Differentiation of Iran_N, Steppe_EMBA or Steppe_MLBA and Onge the Outgroups (Ust_Ishim, MA1, Kostenki14, Papuans, Chukchis, Karitiana and Mbuti_Pygmies) could significantly distinguish the reference sources (p < 0.0001).

f4rank:	p-value	Sources
0	3.01E-70	Iran_N, Steppe_EMBA
1	7.60E-52	Iran_N, Steppe_EMBA, Onge

f4rank:	p-value	Sources
0	1.03E-55	Iran_N, Steppe_MLBA
1	4.13E-44	Iran_N, Steppe_MLBA, Onge

(C): Differentiation of Iran_N, Steppe_EMBA, Steppe_MLBA and Onge the Outgroups (Ust_Ishim, MA1, Kostenki14, Papuans, Chukchis, Karitiana, Mbuti_Pygmies and Anatolia_N) could distinguish the reference sources (p < 0.0001).

f4rank:	p-value	Sources
0	2.63E-69	Iran_N, Steppe_EMBA
1	7.22E-64	Iran_N, Steppe_EMBA, Steppe_MLBA
2	2.52E-25	Iran_N, Steppe_EMBA, Steppe_MLBA, Onge

Table S17. Cross-validation for the unbiasedness due to varying sample sizes

(A): *D* statistics we applied *D statistics* and *qpAdm* using equal number (52 each) of individuals from Steppe_EMBA and Steppe_MLBA to cross-verify if there is any biasedness in results due to different sample sizes for ancient sources like Steppe_EMBA and Steppe_MLBA in our earlier analyses and observed there is negligible deviation.

pop1	Yoruba	pop2	pop3	Dstat	Z score
Ror	Yorubas	Steppe_EMBA	Steppe_MLBA	0.00061	5.342
Jatt	Yorubas	Steppe_EMBA	Steppe_MLBA	0.00062	3.096
Gujjar	Yorubas	Steppe_EMBA	Steppe_MLBA	0.000759	6.632
Kamboj	Yorubas	Steppe_EMBA	Steppe_MLBA	0.000587	5.152
Khatri	Yorubas	Steppe_EMBA	Steppe_MLBA	0.000748	6.88
Pathan	Yorubas	Steppe_EMBA	Steppe_MLBA	0.000686	6.316
Kalash	Yorubas	Steppe_EMBA	Steppe_MLBA	0.000932	7.818
Sindhi	Yorubas	Steppe_EMBA	Steppe_MLBA	0.000742	6.927
Brahmins_UP	Yorubas	Steppe_EMBA	Steppe_MLBA	0.000645	5.582
Kshatriya	Yorubas	Steppe_EMBA	Steppe_MLBA	0.000635	5.417
Chamar	Yorubas	Steppe_EMBA	Steppe_MLBA	0.000467	3.982
Gujaratis	Yorubas	Steppe_EMBA	Steppe_MLBA	0.000684	6.424
Paniya	Yorubas	Steppe_EMBA	Steppe_MLBA	0.000234	1.934
Ho	Yorubas	Steppe_EMBA	Steppe_MLBA	0.000111	0.997
Ror	Yorubas	Indus_Diaspora	Steppe_MLBA	-0.002219	-8.358
Jatt	Yorubas	Indus_Diaspora	Steppe_MLBA	-0.001881	-3.822
Gujjar	Yorubas	Indus_Diaspora	Steppe_MLBA	0.000049	0.178
Kamboj	Yorubas	Indus_Diaspora	Steppe_MLBA	-0.000398	-1.405
Khatri	Yorubas	Indus_Diaspora	Steppe_MLBA	-0.000677	-2.532
Kalash	Yorubas	Indus_Diaspora	Steppe_MLBA	-0.001623	-5.53
Pathan	Yorubas	Indus_Diaspora	Steppe_MLBA	-0.001304	-4.947
Sindhi	Yorubas	Indus_Diaspora	Steppe_MLBA	-0.000133	-0.518
Brahmins_UP	Yorubas	Indus_Diaspora	Steppe_MLBA	0.000662	2.216
Kshatriya	Yorubas	Indus_Diaspora	Steppe_MLBA	0.000659	2.309
Chamar	Yorubas	Indus_Diaspora	Steppe_MLBA	0.003119	10.797
Gujaratis	Yorubas	Indus_Diaspora	Steppe_MLBA	0.00163	6.436
Paniya	Yorubas	Indus_Diaspora	Steppe_MLBA	0.003746	12.544
Ho	Yorubas	Indus_Diaspora	Steppe_MLBA	0.003761	13.317

(B): *qpAdm* we applied *qpAdm* for same number (52 each) of individuals from Steppe_EMBA and Steppe_MLBA to cross-verify if there is any biasedness in results due to different sample sizes for ancient sources like Steppe_EMBA and Steppe_MLBA in our earlier analyses and observed there is no significant deviation.

Target Pop	p value	Mix coeff1	Mix coeff2	Mix coeff3	Mix coeff4	std err1	std err2	std err3	std err4	Source1	Source2	Source3	Source4
Ror	0.942223	0.226	0.355	0.214	0.205	0.029	0.073	0.081	0.012	Iran_N	Steppe_EMBA	Steppe_MLBA	Onge
Khatri	0.29553	0.24	0.303	0.205	0.252	0.032	0.076	0.084	0.013	Iran_N	Steppe_EMBA	Steppe_MLBA	Onge
Gujjar	0.673977	0.253	0.274	0.177	0.297	0.034	0.084	0.093	0.014	Iran_N	Steppe_EMBA	Steppe_MLBA	Onge
Kamboj	0.257426	0.288	0.242	0.205	0.265	0.034	0.086	0.095	0.015	Iran_N	Steppe_EMBA	Steppe_MLBA	Onge
Jatt	0.537537	0.178	0.61	0	0.213	0.082	0.21	NA	0.033	Iran_N	Steppe_EMBA	Steppe_MLBA	Onge
Pathan	0.532103	0.271	0.207	0.305	0.217	0.03	0.073	0.08	0.012	Iran_N	Steppe_EMBA	Steppe_MLBA	Onge
Kalash	0.63285	0.22	0.496	0.098	0.186	0.035	0.09	0.098	0.015	Iran_N	Steppe_EMBA	Steppe_MLBA	Onge
Sindhi	0.263567	0.51	0.264	0	0.226	0.044	0.108	NA	0.017	Iran_N	Steppe_EMBA	Steppe_MLBA	Onge
Gujaratis	0.916619	0.216	0.39	0	0.395	0.032	0.072	NA	0.013	Iran_N	Steppe_EMBA	Steppe_MLBA	Onge
Brahmins_UP	0.580855	0.208	0.446	0	0.346	0.037	0.091	NA	0.016	Iran_N	Steppe_EMBA	Steppe_MLBA	Onge
Balochi	0.244612	0.668	0.236	0	0.096	0.045	0.107	NA	0.018	Iran_N	Steppe_EMBA	Steppe_MLBA	Onge

Table S18. ALDER dating for admixture in Northwest Indians we estimated the Admixture dates in newly sampled Northwest Indians using ALDER, results reveal admixture at around 1500 years ago in the Ror, quite recent admixture event in Gujjar and the very old event in Kamboj. We used 30 years per generation (gens) to convert dates in generations to years.

Test status	p-value	test pop	ref A	ref B	Z-score	Admixture Date (gens)
success	0.0075	Ror	Asur	French	4.37	43.68 +/- 8.59
success	5.6E-07	Ror	Kol	French	6.12	48.69+/- 7.42
success	7.6E-07	Ror	Ho	French	6.07	46.72+/-7.44
success	0.0017	Ror	Tharu	French	4.68	42.36 +/- 7.62
success	6.2E-05	Ror	Santhal	French	5.32	54.17 +/- 9.43
success	6.3E-05	Ror	Gond	French	5.32	59.56 +/- 10.02
success	9.3E-05	Ror	Kol	Germans	5.24	49.70 +/- 8.50
success	3.9E-05	Ror	Ho	Germans	5.4	49.06 +/- 8.47
success	0.0011	Ror	Tharu	Germans	4.78	40.12 +/- 7.67
success	0.015	Ror	Santhal	Germans	4.21	54.27 +/- 11.40
success	0.00023	Ror	Gond	Germans	5.08	63.09 +/- 9.79
success	0.0057	Gujjar	Asur	Swedes	4.43	15.04 +/- 3.40
success	0.028	Gujjar	Asur	Germans	4.07	13.20 +/- 3.24
success	0.0046	Kamboj	Asur	Estonians	4.47	144.23 +/- 22.54
success	0.008	Kamboj	Asur	North_Italians	4.35	130.33 +/- 20.26
success	0.0019	Kamboj	Asur	Swedes	4.66	148.00 +/- 23.06
success	0.044	Kamboj	Asur	Ukranians	3.97	147.76 +/- 22.30
success	0.017	Kamboj	Swedes	UP_Low_Caste	4.19	79.99 +/- 19.09

Table S19. standard error for NNLS with jack-knife method

(A): All populations as donor

	Eur	Cau	MEast	IN_IE	IN_DRA	Iran	CAsia
Ror	0.0066	0	0.0015	0.0092	0.0008	0.0213	0.0025
Kamboj	0.0038	0.004	0.0063	0.2039	0	0.0608	0.0071
Jatt	0.0051	0.0049	0.0019	0.0176	0.0034	0.0535	0.006
Khatri	0	0	0.004	0.0012	0	0.0307	0.0023
Gujjar	0	0	0	0.0255	0.0056	0.0147	0
Gujaratis	0	0	0	0.1634	0	0	0
Brahmins_UP	0.001	0.0027	0.0006	0.0298	0.0011	0.0126	0.0005
Kshatriya	7E-05	0.0031	0.0005	0.0409	0.0012	0.0035	0
Pathan	0.0004	0.0014	0.0021	0.0264	0.0011	0.0141	0.0037

(B): Without PNWI as donor

	Eur	Cau	MEast	IN_IE	IN_DRA	Iran	CAsia
Ror	0.0385	0.0153	0.0064	0.0302	0.0081	0	0.0042
Kamboj	0.0121	0.0582	0.0184	0.0266	0.0093	0.061	0.0059
Jatt	0.0331	0.0325	0.0064	0.0276	0.0036	0.043	0.0098
Khatri	0.0086	0.0717	0.0281	0.0797	0.0087	0.0768	0.0136
Gujjar	0.0054	0.0056	0.0001	0.0413	0.0044	0	0
Gujaratis	0	0	0	0.0704	0	0	0
Brahmins_UP	0.0145	0.0031	0.003	0.03	0.0053	0	0.0009
Kshatriya	0.0037	0.0022	0.0004	0.0537	0.0027	0	0
Pathan	0.0115	0.1609	0.0132	0.0392	0.0044	0.248	0.0196

(C): UNLINKED model

	Eur	Cau	MEast	IN_IE	IN_DRA	Iran	CAsia
Ror	0.0436	0.0182	0.0044	0.0383	0.0192	0	0.0107
Kamboj	0.0345	0.0157	0.0116	0.041	0.0318	0.0037	0.0063
Jatt	0.0846	0.0398	0.0075	0.0811	0.0481	0	0.0112
Khatri	0.0328	0.0133	0.0039	0.0398	0.0242	0.0038	0.0085
Gujjar	0.0114	0.0165	0.004	0.0468	0.0324	0.0052	0.0097
Gujaratis	0.0354	0.0106	0.0083	0.0314	0.0641	0	0.0056
Brahmins_UP	0.0349	0.0198	0.0056	0.0587	0.0541	0.0021	0.0063
Kshatriya	0.0254	0.0174	0.005	0.0631	0.0543	0.0038	0.0089
Pathan	0.0317	0.0177	0.0036	0.0406	0.0184	0.0026	0.0064

References:

1. Kenoyer, J.M. (1998). Kenoyer, J. M. Ancient Cities of the Indus Valley Civilization. (Oxford University, Oxford, 1998).

2. Possehl, G.L. (2002). Possehl, G. L. The Indus Civilisation: A Contemporary Perspective. (AltaMira, Lanham, MD, 2002).

3. Wright, R.P. (2010). The ancient Indus: urbanism, economy, and society.

4. Kenoyer, J.M. (2008). Kenoyer, J. M. in The Ancient City: New Perspectives in the Old and New World (eds J. Marcus & J. A. Sablof) 85–109 (SAR, Santa Fe, NM, 2008).

5. Lawler, A. (2008). Boring no more, a trade-savvy Indus emerges. Science *320*, 1276–1281.

6. Joseph, T. (2017). Who built the Indus Valley civilisation? The Hindu.

7. Giosan, L., Clift, P.D., Macklin, M.G., Fuller, D.Q., Constantinescu, S., Durcan, J.A., Stevens, T., Duller, G.A.T., Tabrez, A.R., Gangal, K., et al. (2012). Fluvial landscapes of the Harappan civilization. Proc. Natl. Acad. Sci. *109*, E1688–E1694.

8. Giosan, L., Fuller, D.Q., Nicoll, K., Flad, R.K., and Clift, P.D. (2013). Climates, Landscapes, and Civilizations. (Washington: Wiley).

9. Gangal, K., Vahia, M.N., and Adhikari, R. (2010). Spatio-temporal analysis of the Indus urbanization. Curr. Sci. 00113891 *98*, 846–852.

10. McIntosh, J. (2008). The ancient Indus Valley: new perspectives (Santa Barbara, Calif: ABC-CLIO).

11. Petrie, C.A. (2013). Petrie, C. A. in Oxford Handbook of Cities in History (ed P. Clark) 139–173 (Oxford University Press, Oxford, 2013). Oxf. Univ. Press 130–173.

12. Beck, R.B. (1999). McDougal Littell world history: patterns of interaction (Evanston, Ill.: McDougal Littell).

13. Jarrige, J.-F. (1986). Excavations at Mehrgarh-Nausharo. Pak. Archaeol. *10*, 63–131.

14. Jarrige, J.F. (2006). Mehrgarh Neolithic, Pragdhara No. 18.

15. Gallego Romero, I., Basu Mallick, C., Liebert, A., Crivellaro, F., Chaubey, G., Itan, Y., Metspalu, M., Eaaswarkhanth, M., Pitchappan, R., Villems, R., et al. (2012). Herders of Indian and European Cattle Share Their Predominant Allele for Lactase Persistence. Mol. Biol. Evol. *29*, 249–260.

16. Dixit, Y., Hodell, D.A., and Petrie, C.A. (2014). Abrupt weakening of the summer monsoon in northwest India~ 4100 yr ago. Geology *42*, 339–342.

17. Allchin, B., and Allchin, F.R. (1997). Origins of a civilization: the prehistory and early archaeology of South Asia (Viking Adult).

18. Gangal, K., Sarson, G.R., and Shukurov, A. (2014). The Near-Eastern Roots of the Neolithic in South Asia. PLOS ONE *9*, e95714.

19. Beck, R.B., and Beck, R.B. (2005). World history: Patterns of interaction (McDougal Littell Evanston, IL).

20. Tripathi, J.K., Bock, B., Rajamani, V., and Eisenhauer, A. (2004). Is Rhiver Ghaggar, Saraswati? Geochemical constraints. Curr. Sci. *87*, 1141–1144.

21. Gupta, A.K., Anderson, D.M., Pandey, D.N., and Singhvi, A.K. (2006). Adaptation and human migration, and evidence of agriculture coincident with changes in the Indian summer monsoon during the Holocene. Curr. Sci.-BANGALORE- *90*, 1082.

22. MacDonald, G. (2011). Potential influence of the Pacific Ocean on the Indian summer monsoon and Harappan decline. Quat. Int. *229*, 140–148.

23. Singh, G. (1971). The Indus valley culture. Archaeol. Ocean. *6*, 177–189.

24. Wright, R.P., Bryson, R.A., and Schuldenrein, J. (2008). Water supply and history: Harappa and the Beas regional survey. Antiquity *82*, 37–48.

25. Madella, M., and Fuller, D.Q. (2006). Palaeoecology and the Harappan Civilisation of South Asia: a reconsideration. Quat. Sci. Rev. *25*, 1283–1301.

26. rorraja.com History Of Ror Community | Ror Employees Association.

27. Rose, H.A., and Asian Educational Services (1990). A glossary of the tribes and castes of the Punjab and North-West Frontier Province. Vol. 2, Vol. 2, (New Delhi: Asian Educational Services).

28. Beglar, J.D., Carlleyle, A.C.L., and Cunningham, S.A. (1994). Report for the Year 1871-72. Delhi (Rahul Publishing House).

29. rormahasabha.org Ror Community | About Us.

30. Davids, T.W.R. (1903). Buddhist India, by TW Rhys Davids (Putnam; London).

31. Elliott, H.M. (1859). Memoirs on the History, Folk-Lore, and Distribution of the Races of the North Western Provinces of India. Lond. Trübner.

32. Elliot, H.M. (1869). Memoirs on the History, Folk-Lore, and Distribution of the Races of the North Western Provinces of India; being an amplified Edition of the original: Supplemental Glossary of India Terms By the late Henry M. Elliot. Edited, revised, and re-arranged by John Beames. In 2 Volumes. II (Trübner & Company).

33. peoplegroupsindia.com Gujjar – People Groups of India.

34. (2001). Encyclopaedia of the world Muslims: tribes, castes and communities (Delhi: Global Vision Pub. House).

35. (2011). Culture, institutions, and development: new insights into an old debate (New York: Routledge).

36. Bose, S.K. (2015). Boots, hooves, and wheels: and the social dynamics behind South Asian warfare (New Delhi, India: Vij Books India Pvt. Ltd).

37. Cunningham, A. (1871). Four Reports Made During the Years, 1862-63-64-65 (Printed at the Government Central Press).

38. Smith, V.A. (1924). The Early History of India, from 600 BC to the Muhammadan Conquest including the Invasion of Alexander the Great. revised by SM Edwardes (Oxford at the Clarendon Press (first published in 1914)).

39. Khatra, P.S., and Sharma, V. (1992). Socio-economic issues in the development of nomadic Gujjars. Indian J. Agric. Econ. *47*, 448–449.

40. Singh, D.E. (2012). Islamization in Modern South Asia: Deobandi Reform and the Gujjar Response (Walter de Gruyter).

41. Agrawal, S.P., and Aggarwal, J.C. (1991). Educational and social uplift of backward classes: at what cost and how?: Mandal Commission and after (New Delhi: Concept Pub. Co).

42. Chand, P. (1990). Indian archers Dhanushkas as Sanskrit literature depicts with a special reference to the Dhanaka family matrimonially connected with the Iksvaku kings Andhradesa 200 625 AD.

43. Puri, B.N. (1954). The history of the Gurjara-Pratihāras. University of Oxford.

44. Tripathi, R.S. (1989). History of Kanauj: to the Moslem conquest (Delhi: Motilal Banarsidass).

45. Shashi, S.S. (1979). The nomads of the Himalayas (Delhi: Sundeep).

46. Bingley, A.H. (1978). Caste, tribes & culture of Rajputs (Ess Ess Publications).

47. Ibbetson, D.C.J. (1981). Report on the Census of Punjab (Calcutta).

48. Verma, Y.K. (1954). Gurjara Itihas (in Hindi).

49. Singh, K.S. (1998). India’s communities (Delhi: Oxford Univ. Press).

50. Joshua Project Gujjars in Pakistan.

51. Joshua Project Gujjar in India.

52. Law, B.C. (1924). Ancient Mid-Indian Ksatriya Tribes - Vol. I (Calcutta: Thacker, Spink & Co.).

53. Dwivedi, R.K. (1977). A Critical study of Changing Social Order at Yuganta: or the end of the Kali Age. 1977 DD Kosambi Commem. Vol. Varanasi Banaras Hindu Univ.

54. Mishra, K.C. (1987). Tribes in The Mahabharata: A socio-cultural study (New Delhi: National Publishing House).

55. Raychaudhuri, H. (2006). Political history of ancient India: from the accession of Parikshit to the extinction of the Gupta dynasty (New Delhi: Cosmo Publications).

56. West, B.A. (2009). Encyclopedia of the peoples of Asia and Oceania (New York: Facts On File).

57. Majumdar, R.C., Pusalker, A.D., Majumdar, A.K., Ghose, D.K., and Dighe, V.G. (1966). The History and Culture of the Indian People: The struggle for empire.-2d ed (Bharatiya Vidya Bhavan).

58. Tapp, N., and Walker, A. (2001). The Tai World: A Digest of Articles from the Thai-Yunnan Project Newsletter (Department of Anthropology, Research School of Pacific and Asian Studies, The Australian National University).

59. Singh, M.R. (1972). Geographical data in the early Pur\={a}\d {n} as: a critical study. Punthi Pustak Calcutta.

60. Jayswal, K.P., and Lamotte, E. Kamboja Kingdom.

61. (2007). Global Indian diasporas: exploring trajectories of migration and theory (Netherlands: Amsterdam University Press).

62. Levi, S.C. (2002). The Indian Diaspora in Central Asia and Its Trade, 1550-1900 (Brill).

63. Syan, H.S. (2013). Sikh militancy in the seventeenth century: religious violence in Mughal and early modern India (London: Tauris).

64. Mclane (2002). Land and local kingship 18th century bengal. (Cambridge: Cambridge Univ Press).

65. McLeod, W.H., and McLeod, W.H. (2009). The A to Z of Sikhism (Lanham: Scarecrow Press).

66. Joshua Project Khatri (Hindu traditions) in India.

67. Mahal, D.G. (2015). Ancestral Roots of the Jats DNA Revelations (DGM Associates).

68. peoplegroupsindia.com Jat – People Groups of India.

69. Crooke, W. (1896). The tribes and castes of the North-Western Provinces and Oudh (Office of the superintendent of government printing).

70. wikipedia.org (2017). Jat people.

71. Chatterji, S. History repeats itself as yet another Central govt faces a Jat stir | india | Hindustan Times.

72. Li, J.Z., Absher, D.M., Tang, H., Southwick, A.M., Casto, A.M., Ramachandran, S., Cann, H.M., Barsh, G.S., Feldman, M., Cavalli-Sforza, L.L., et al. (2008). Worldwide Human Relationships Inferred from Genome-Wide Patterns of Variation. Science *319*, 1100–1104.

73. Behar, D.M., Yunusbayev, B., Metspalu, M., Metspalu, E., Rosset, S., Parik, J., Rootsi, S., Chaubey, G., Kutuev, I., Yudkovsky, G., et al. (2010). The genome-wide structure of the Jewish people. Nature *466*, 238–242.

74. Yunusbayev, B., Metspalu, M., Järve, M., Kutuev, I., Rootsi, S., Metspalu, E., Behar, D.M., Varendi, K., Sahakyan, H., Khusainova, R., et al. (2012). The Caucasus as an asymmetric semipermeable barrier to ancient human migrations. Mol. Biol. Evol. *29*, 359–365.

75. Behar, D.M., Metspalu, M., Baran, Y., Kopelman, N.M., Yunusbayev, B., Gladstein, A., Tzur, S., Sahakyan, H., Bahmanimehr, A., Yepiskoposyan, L., et al. (2013). No evidence from genome-wide data of a Khazar origin for the Ashkenazi Jews. Hum. Biol. *85*, 859–900.

76. Yunusbayev, B., Metspalu, M., Metspalu, E., Valeev, A., Litvinov, S., Valiev, R., Akhmetova, V., Balanovska, E., Balanovsky, O., Turdikulova, S., et al. (2015). The Genetic Legacy of the Expansion of Turkic-Speaking Nomads across Eurasia. PLOS Genet. *11*, e1005068.

77. Raghavan, M., Skoglund, P., Graf, K.E., Metspalu, M., Albrechtsen, A., Moltke, I., Rasmussen, S., Stafford, T.W., Orlando, L., Metspalu, E., et al. (2014). Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. Nature *505*, 87–91.

78. Kushniarevich, A., Utevska, O., Chuhryaeva, M., Agdzhoyan, A., Dibirova, K., Uktveryte, I., Möls, M., Mulahasanovic, L., Pshenichnov, A., Frolova, S., et al. (2015). Genetic Heritage of the Balto-Slavic Speaking Populations: A Synthesis of Autosomal, Mitochondrial and Y-Chromosomal Data. PLOS ONE *10*, e0135820.

79. Metspalu, M., Romero, I.G., Yunusbayev, B., Chaubey, G., Mallick, C.B., Hudjashov, G., Nelis, M., Mägi, R., Metspalu, E., and Remm, M. (2011). 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.

80. Cristofaro, J.D., Pennarun, E., Mazières, S., Myres, N.M., Lin, A.A., Temori, S.A., Metspalu, M., Metspalu, E., Witzel, M., King, R.J., et al. (2013). Afghan Hindu Kush: Where Eurasian Sub-Continent Gene Flows Converge. PLOS ONE *8*, e76748.

81. Chaubey, G., Metspalu, M., Choi, Y., Mägi, R., Romero, I.G., Soares, P., van Oven, M., Behar, D.M., Rootsi, S., and Hudjashov, G. (2011). Population genetic structure in Indian Austroasiatic speakers: the role of landscape barriers and sex-specific admixture. Mol. Biol. Evol. *28*, 1013–1024.

82. Consortium, I.H. 3 (2010). Integrating common and rare genetic variation in diverse human populations. Nature *467*, 52–58.

83. Chaubey, G., Ayub, Q., Rai, N., Prakash, S., Mushrif-Tripathy, V., Mezzavilla, M., Pathak, A.K., Tamang, R., Firasat, S., Reidla, M., et al. (2017). "Like sugar in milk": reconstructing the genetic history of the Parsi population. Genome Biol. *18*, 110.

84. Basu, A., Sarkar-Roy, N., and Majumder, P.P. (2016). Genomic reconstruction of the history of extant populations of India reveals five distinct ancestral components and a complex structure. Proc. Natl. Acad. Sci. *113*, 1594–1599.

85. Rasmussen, M., Guo, X., Wang, Y., Lohmueller, K.E., Rasmussen, S., Albrechtsen, A., Skotte, L., Lindgreen, S., Metspalu, M., Jombart, T., et al. (2011). An Aboriginal Australian genome reveals separate human dispersals into Asia. Science *334*, 94–98.

86. Pierron, D., Razafindrazaka, H., Pagani, L., Ricaut, F.-X., Antao, T., Capredon, M., Sambo, C., Radimilahy, C., Rakotoarisoa, J.-A., Blench, R.M., et al. (2014). Genome-wide evidence of Austronesian–Bantu admixture and cultural reversion in a hunter-gatherer group of Madagascar. Proc. Natl. Acad. Sci. *111*, 936–941.

87. Damgaard, P. de B., Martiniano, R., Kamm, J., Moreno-Mayar, J.V., Kroonen, G., Peyrot, M., Barjamovic, G., Rasmussen, S., Zacho, C., Baimukhanov, N., et al. (2018). The first horse herders and the impact of early Bronze Age steppe expansions into Asia. Science eaar7711.

88. Narasimhan, V.M., Patterson, N.J., Moorjani, P., Lazaridis, I., Mark, L., Mallick, S., Rohland, N., Bernardos, R., Kim, A.M., Nakatsuka, N., et al. (2018). The Genomic Formation of South and Central Asia. BioRxiv 292581.

89. Rasmussen, M., Sikora, M., Albrechtsen, A., Korneliussen, T.S., Moreno-Mayar, J.V., Poznik, G.D., Zollikofer, C.P.E., Ponce de León, M.S., Allentoft, M.E., Moltke, I., et al. (2015). The ancestry and affiliations of Kennewick Man. Nature *523*, 455.

90. Allentoft, M.E., Sikora, M., Sjögren, K.-G., Rasmussen, S., Rasmussen, M., Stenderup, J., Damgaard, P.B., Schroeder, H., Ahlström, T., and Vinner, L. (2015). Population genomics of bronze age Eurasia. Nature *522*, 167–172.

91. Fu, Q., Li, H., Moorjani, P., Jay, F., Slepchenko, S.M., Bondarev, A.A., Johnson, P.L.F., Aximu-Petri, A., Prüfer, K., de Filippo, C., et al. (2014). Genome sequence of a 45,000-year-old modern human from western Siberia. Nature *514*, 445.

92. Günther, T., Malmström, H., Svensson, E.M., Omrak, A., Sánchez-Quinto, F., Kılınç, G.M., Krzewińska, M., Eriksson, G., Fraser, M., Edlund, H., et al. (2018). Population genomics of Mesolithic Scandinavia: Investigating early postglacial migration routes and high-latitude adaptation. PLOS Biol. *16*, e2003703.

93. Jones, E.R., Gonzalez-Fortes, G., Connell, S., Siska, V., Eriksson, A., Martiniano, R., McLaughlin, R.L., Llorente, M.G., Cassidy, L.M., Gamba, C., et al. (2015). Upper Palaeolithic genomes reveal deep roots of modern Eurasians. Nat. Commun. *6*, 8912.

94. Lazaridis, I., Patterson, N., Mittnik, A., Renaud, G., Mallick, S., Kirsanow, K., Sudmant, P.H., Schraiber, J.G., Castellano, S., Lipson, M., et al. (2014). Ancient human genomes suggest three ancestral populations for present-day Europeans. Nature *513*, 409.

95. Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S.A., Harney, E., Stewardson, K., Fernandes, D., and Novak, M. (2015). Genome-wide patterns of selection in 230 ancient Eurasians. Nature *528*, 499–503.

96. Fu, Q., Posth, C., Hajdinjak, M., Petr, M., Mallick, S., Fernandes, D., Furtwängler, A., Haak, W., Meyer, M., Mittnik, A., et al. (2016). The genetic history of Ice Age Europe. Nature *534*, 200.

97. Lazaridis, I., Nadel, D., Rollefson, G., Merrett, D.C., Rohland, N., Mallick, S., Fernandes, D., Novak, M., Gamarra, B., and Sirak, K. (2016). Genomic insights into the origin of farming in the ancient Near East. Nature *536*, 419–424.