

RESEARCH ARTICLE

Genetic Affinity of the Bhil, Kol and Gond Mentioned in Epic *Ramayana*

Gyaneshwer Chaubey^{1*}, Anurag Kadian², Saroj Bala³, Vadlamudi Raghavendra Rao⁴

1 Estonian Biocentre, Riia23, Tartu 51010, Estonia, **2** 5 Ror Colony, Behind Sector 7, Karnal, Haryana132001, India, **3** Institute of Scientific Research on Vedas, I-SERVE Delhi Chapter, C-6 / 302, Clarion the Legend, Gurgaon 122011, India, **4** Department of Anthropology, University of Delhi, North Campus, Delhi 110007, India

* gyanc@ebc.ee



OPEN ACCESS

Citation: Chaubey G, Kadian A, Bala S, Rao VR (2015) Genetic Affinity of the Bhil, Kol and Gond Mentioned in Epic *Ramayana*. PLoS ONE 10(6): e0127655. doi:10.1371/journal.pone.0127655

Academic Editor: Kumarasamy Thangaraj, Centre for Cellular and Molecular Biology, INDIA

Received: September 7, 2014

Accepted: April 17, 2015

Published: June 10, 2015

Copyright: © 2015 Chaubey et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Data have been deposited to Figshare: <http://dx.doi.org/10.6084/m9.figshare.1391920>. The data are also available at the data repository of the Estonian Biocentre: www.ebc.ee/free_data.

Funding: The authors received no specific funding for this work. GC is supported by Center of Excellence of Estonian Biocentre and Institutional Grant IUT24-1.

Competing Interests: Co-author Dr. Gyaneshwer Chaubey is a PLOS ONE Editorial Board member. This does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials.

Abstract

Kol, Bhil and Gond are some of the ancient tribal populations known from the *Ramayana*, one of the Great epics of India. Though there have been studies about their affinity based on classical and haploid genetic markers, the molecular insights of their relationship with other tribal and caste populations of extant India is expected to give more clarity about the the question of continuity vs. discontinuity. In this study, we scanned >97,000 of single nucleotide polymorphisms among three major ancient tribes mentioned in *Ramayana*, namely Bhil, Kol and Gond. The results obtained were then compared at inter and intra population levels with neighboring and other world populations. Using various statistical methods, our analysis suggested that the genetic architecture of these tribes (Kol and Gond) was largely similar to their surrounding tribal and caste populations, while Bhil showed closer affinity with Dravidian and Austroasiatic (Munda) speaking tribes. The haplotype based analysis revealed a massive amount of genome sharing among Bhil, Kol, Gond and with other ethnic groups of South Asian descent. On the basis of genetic component sharing among different populations, we anticipate their primary founding over the indigenous Ancestral South Indian (ASI) component has prevailed in the genepool over the last several thousand years.

Introduction

Knowledge about the past comes through different disciplines where researchers look at history through different lenses. And in many cases, these interdisciplinary studies land on the same conclusions [1,2]. However, in case of India, investigations from different disciplines have historically been highly contrasting [3,4]. India, also known as a 'land of spiritual heritage', has a deep history of civilisation, which is embedded in to multiple oral, traditional and written records. Much of this knowledge is rooted in oldest scriptures, the *Vedas*, which are four in number, namely *Rigveda*, *Yajurveda*, *Samaveda* and *Atharvaveda*. Then, there are *Puranas*, *Upanishads*, *Brahmanas* and *Aranyakas*, of which *Vedas* are said to be the precursors [5]. There is no consensus among historians regarding the date of compilation of the *Vedas* as well as the historical dates for the various *Puranas*, *Upanishads* and epics [6–10]. A comparative

analysis of such mythological sources may provide a concensus about the structuring of the ancient societies and rituals. More recently, some scholars have provided strong evidence about the chronology of these events hinting at a deep-rooted civilization, developing indigenously for over several thousand years [8,11–15].

Our survey on mythological sources has revealed detailed information about the ancient Indian society structure as well as relations of different tribal and caste groups and their rituals [10,11,16–18]. In many of these literary sources, names of various castes and tribal groups have been mentioned, including those of several surviving tribal groups (e.g. Bhil or Bheel, Kol, Gond, Savara, Oraon, Kirata, Ahirs, Nagas etc) [17–23]. It is already evident that during the *Ramayana* era, Indian society was well-stratified [16,17,21,23–26]. The Bhil, Kol and Gond are three major Indian tribes that have been widely acknowledged in the epic *Ramayana*, particularly in the portions known as the *Ayodhyakanda*, *Aranyakanda* and *Kishkindhakanda* [19,20,22–27]. It should be emphasised here that, Gond and Bhil are the top two tribal populations of modern India in terms of population size [28].

The Bhils are primarily from Central India and speak the Bhil language [28]. They have significant presence in states of Gujarat, Madhya Pradesh, Chhattisgarh, Maharashtra and Rajasthan as well as in the northeastern state of Tripura. Bhils are further divided into a number of endogamous territorial divisions, which in turn have a number of clans and lineages [22]. The Kol tribe in Uttar Pradesh is found mainly in the districts of Mirzapur, Varanasi, Banda and Allahabad [28]. It is the largest tribe found in the state Uttar Pradesh. They are said to have migrated from Central India some five centuries ago [28]. The Kol are further divided into a number of exogamous clans, such as the Rojaboria, Rautia, Thakuria, Monasi, Chero and Barawire. The Gond people are spread over the states of Madhya Pradesh, eastern Maharashtra (Vidarbha), Chhattisgarh, Uttar Pradesh and Telangana [28]. With over four million people, they are the largest tribe in Central India. They speak the Gondi language, which is related to present Dravidian language family [28,29].

More than 25 years of genetic research on Indian tribal and caste populations involving classical markers to mtDNA/Y chromosome and more recently autosomes, have indicated complex demographic history of the subcontinent [3,30–39]. Alongwith debate over initial peopling of the subcontinent, the major hot topic now shifted towards the population expansion and admixture during and after Neolithic times [37–40]. However, large number of individuals as well as genetic markers are required to reach any firm conclusions. Thus, the strict endogamy and social structure make South Asia much more complex, unlike to Europe, where genetic analysis of a population can predict the genetic structure of immediate neighbor with some confidence. In recent years, there has been an increase in the number of in-depth genetic studies focussing on the genetic structure of the populations of India [35,37,40–48], but none of them have related specific tribal populations mentioned in the traditional literatures.

Therefore, in the present study, we make an attempt to evaluate two schools of thought emerging from the current scenario. The first school suggests that the tribal people are the aboriginal inhabitants, while the later migrants, i.e., the Dravidians followed by the Aryans have pushed them back in to small pockets in South India [49–52]. According to this school, the caste system was established by the aforementioned later migrants [11,50,52,53]. The alternative hypothesis advocates that all the caste and tribal populations of India have Paleolithic roots and share a common origin [3,15,33,54–60]. The differentiation observed in modern South Asian populations is mainly derived by strict endogamy, long term isolation and several evolutionary forces. More specifically, relying on each other, first, we seek to investigate the continuity vs. discontinuity of the genetic thread connecting the different populations of India. Second, keeping in mind the pivotal information extracted from *Ramayana*, we look specifically into the question: whether and to what extent the three major tribes (Bhil, Kol and Gond)

share their genetic ancestry among them as well as with other contemporary caste and tribal populations?

Material and Methods

This study was performed using control samples collected, genotyped and published for various population studies conducted in the last few years (S1 Table) [37–39,46,61–63]. All the ethical guideline have been followed. The tribal and caste populations grouped according to their language group. We grouped populations in to “Transitional” who have known information of language change in recent time [64,65]. A check for closely related individuals was carried out within each population study by calculating average *identity by state* (IBS) scores for all pairs of individuals [66]. We used PLINK 1.07 [66] in order to filter our dataset to include only SNPs on the 22 autosomal chromosomes with minor allele frequency >1% and genotyping success >99%. As background linkage disequilibrium (LD) can affect both PCA [67] and ADMIXTURE [68], we thinned the dataset by removing one SNP of any pair, in strong LD $r^2 > 0.4$, in a window of 200 SNPs (sliding the window by 25 SNPs at a time).

We performed PC analysis using *smartpca* programme (with default settings) of the EIGENSOFT package [67] in order to capture genetic variability described by the first 5 components. The fraction of the total variation described by a PC is the ratio of its eigenvalue to the sum of all eigenvalues. In the final settings, we ran Admixture with a random seed number generator on the LD-pruned dataset twenty-five times at $K = 2$ to $K = 12$. Since the top values of the resulting log-likelihood scores were stable (virtually identical) within the runs of each K from $K = 2$ to $K = 10$, we can claim that convergence at global maximum was achieved. Thus, we omitted runs at $K = 11$ to $K = 12$ from further analysis.

Mean pairwise differences between different population groups were computed using *Fst* distance measure by following the methods as described by Cockerham and Weir [69], Phylip [70] and MEGA [71] were used to construct the tree. The Plink software [66] was used to calculate the genetic diversity and to find the 25 nearest-neighbours for the Bhil, Kol and Gond individuals. To investigate the derived allele sharing of Bhil, Kol and Gond with the Eurasian populations, we computed *f3* statistics [37], taking African as an outgroup. For haplotype-based analysis (fineSTRUCTURE) [72], we made two different runs—first by taking all the Eurasian populations and second exclusively on the Central Asian, Pakistani and Indian populations. For the fineSTRUCTURE analysis, first samples were phased with Beagle 3.3.2 [73]. A coancestry matrix was constructed using ChromoPainter [72], fineSTRUCTURE was used to perform an MCMC iteration using 10000000 burning runtime and 10000 MCMC samples. A tree was built using fineSTRUCTURE with the default settings. All these information are plotted for the Bhil, Kol and Gond as a recipient of number of chunks from one another as well as from other ethnic group.

Results and Discussion

We combined hundreds of thousands of autosomal markers generated from different studies (S1 Table) [37–39,46,61–63] and specifically looked into the population structure of Indian groups mentioned in classical literature. To find out the population clustering, we first ran the *Fst* (population differentiation) algorithm [69] and drew a tree [70,71], rooting out the African populations (S1 Fig). All the Indian populations, except the present Tibeto-Burman speaking populations, are well separated from other continental populations and form a major cluster comprising present populations speaking Indo-European, Dravidian and Austroasiatic (Munda) languages (S1 Fig). The Pakistani populations are scattered in different clusters, where few of them (Sindhi, Pathan and Burusho) cluster loosely with Indians; Hazaras show an

affinity toward Central Asians, and Balochi, Brahui and Makrani confirm an intermediate position because of shared recent African ancestry and gene flow [38,74,75]. The Bhil, Kol and Gond showed a closer affinity among them as well as with the extent Indo-European, Transitional and Munda speaking populations (Fig 1a and S1 Fig).

To get more deeper insight, we have used PCA (principle component analysis)[67] and ADMIXTURE [68], analysis using the same parameters as in our previous studies [38,39,45]. These analyses strengthened the inferences drawn from the *Fst* analysis. The PCA on Eurasians placed Indian populations between East and West Eurasia (Fig 2a). The cline of Indian subcontinent ranges from Pakistani populations (closer to West Eurasians) to Indian Munda groups (closer to East Eurasians). Departing from its geographical position, Bhil was clustering together with Scheduled castes and Scheduled tribe populations of Uttar Pradesh (Harijan), Andhra Pradesh (Kamsali) and Karnataka (North Kannadi) states. Kol is joined with the neighbouring populations alongwith the Indian-cline, while Gond was deflating away from the Indian cline by uniting with the Munda speakers (Fig 1b). Further, we assessed the proportion of individual-wise ancestry drawn from a given number of inferred populations (K) using a maximum-likelihood based approach implemented in ADMIXTURE.

Consistent with previous observations [37,38], the South Asian populations' genome are mainly made-up of two major components, which are distributed across the length and breadth of the subcontinent (Fig 1c). Alongwith these two major components, there are four minor components over the periphery of the subcontinent—the European and the Middle eastern components can be seen in Pakistani and northwest Indian populations, whilst the East/Southeast Asian components are present in nearby Munda and Tibeto-Burman speakers. (Fig 1c). The geographical distribution of the dark green component (ASI or Ancestral South Indian- unique to the subcontinent) was largely limited to the Indian subcontinent, and can be seen among all the populations of the subcontinent albeit in variable amount, whereas the second major component (light green: ANI or Ancestral North Indian (now ANE- Ancestral North Eurasian [76])) was shared with Central Asia, the Caucasus, Middle East and Europe (Fig 1c). The geographical origin of light green component (ANI or ANE) is so far unclear and more research is needed from unsampled area as well as from ancient DNA; however, the time of spread of this component from its origin place (either of any; the Caucasus, Near East, Indus Valley, or Central Asia) has happened more than 12.5 thousand years before [38], which is significantly earlier than the purported expansion of Dravidians and Aryans languages from outside the subcontinent. Notably, the Andaman Islanders are not the only population carrying the ASI component exclusively, as was suggested before [37]. Austroasiatic speakers (more precisely, the South Munda) of the subcontinent also seem to possess the ASI component in near unadulterated form (Fig 1c). More research with complete genome analysis would be required to clear the geographic center of the ANE component; however, it is evident from the present analysis that the dark green component (ASI) can be considered as a connecting thread for all the Indian populations (Fig 1c). Taken together, these results support the second hypothesis suggesting that all Indians, irrespective of their caste or tribal affiliations, share a common genetic ancestry, which is undoubtedly founded over the indigenous ASI component.

Our second question revolved around the three tribal populations mentioned in the ancient epic, their genome composition and affiliation with the surrounding caste and tribal populations. Based on information from *Ramayana*, we have considered these tribal populations to be ancient inhabitants of India, surviving from the times of the Stone Age [19,23]. If we assume that their genome carry the signature of peopling of ancient time, the assessment of their genomes and comparison with modern populations would test the scenario of continuity vs. discontinuity of prehistoric heritage. In case of continuity, we should see largely similar genome composition among contemporary caste and tribal populations of modern India. On the other

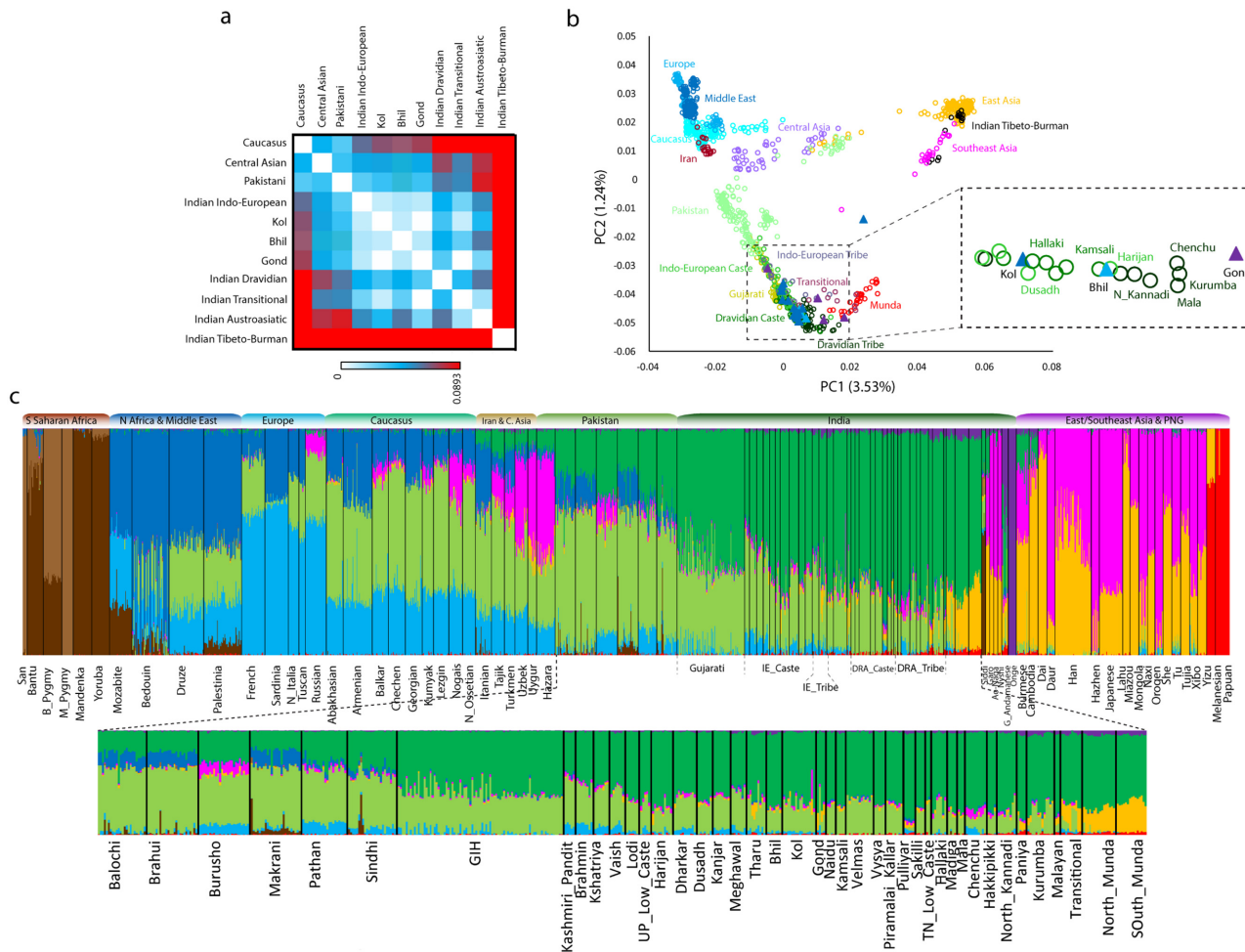


Fig 1. a) Regionwise population differentiation (F_{st}) analysis of Bhil, Kol and Gond with the Indian and other regional populations **b)** PCA (Principle Component Analysis) of Eurasian populations showing the placement of Bhil, Kol and Gond populations over the South Asian cline; the mean value of populationwise clustering of Bhil, Kol and Gond is zoomed-in inset figure. IE- Indo-European, DRA- Dravidian **c)** Individualwise ancestry proportion analysis inferred from ADMIXTURE representing ten ancestral populations of the world ($K = 10$).

doi:10.1371/journal.pone.0127655.g001

hand, in case of discontinuity, these tribal populations should show a unique genome composition or they should emerge as an outliers in our cluster based analysis. Our extended analysis on F_{st} , PCA and ADMIXTURE showed similar genome composition of these tribal populations, carrying both the ANE and ASI components (Fig 1a–1c, Table 1 and S1 Fig). We also calculated the genetic diversity of these populations with their neighbours (Table 2). The diversity of Kol, Bhil and Gond didn't show any significant deviation from their neighbouring extent Indo-European, Dravidian and Munda speaking groups.

The shared drift statistics analysis (f_3) suggested that most of the derived alleles of Bhil, Kol and Gond are overwhelmingly shared with Indian caste and tribal populations (S2 Fig). Gond, Dravidian tribes and Austroasiatic (Munda) groups shared the highest derived allele with Bhils. Indo-European castes, Gond and Dravidian tribes were closest with Kol. Whereas, Munda, transitional and Dravidian tribal groups shared the peak derived alleles with Gonds.

The haplotype based fineSTRUCTURE [72] analysis showed that the studied populations (Bhil, Kol and Gond) received nearly all of their chunks from the Indian closeby populations (Fig 2a and S3 Fig). Leaving out the number of chunks coming from the same population,

Table 1. The ANE and ASI admixture proportions of Bhil, Kol and Gond with respect to other South Asian groups.

Ethnic group	ANE	ASI
Pakistan	49.49 ± 10.88	23.78 ± 9.08
Bhil	19.81 ± 0.97	70.17 ± 1.13
Kol	25.06 ± 5.35	64.64 ± 9.84
Gond	15.55 ± 15.47	67.12 ± 11.71
Indian IE	33.57 ± 6.83	57.77 ± 8.29
Indian DRA	20.55 ± 9.6	70.08 ± 6.30
Transitional	9.88 ± 8.81	66.59 ± 2.92
Indian AA	0.69 ± 2.06	66.97 ± 4.94

IE—Indo-European, DRA—Dravidian, AA—Austroasiatic (Munda)

doi:10.1371/journal.pone.0127655.t001

chunk donors for Bhil and Kol were coming from all the major Indian ethnic groups, while for the Gond, Indian Transitional and Munda groups were the major chunk donors. More specifically the haplotype based sharing analysis is in congruent with the f_3 statistics. The fineSTRUCTURE clustering analysis revealed 37 clusters when we have included Iranian, Central Asian, Pakistani, Indian and Cambodian populations (S4 Fig). Most of the Indian populations unite in Indian specific clusters except Kashmiri Pandits and few Gujarati individuals who fell together with the Sindhi and Pathan individuals in Pakistani specific clade. Our targeted populations are dispersed in various clades. All the Bhil individuals form a tight cluster with the individuals mainly from Dravidian caste, few Indo-European and Transitional individuals. Most of the Kol and Gond individuals show a higher level of variation by falling in to distinct clusters. To make an individual-wise comparison, we plotted top twenty five closest neighbours of studied populations (Fig 2b). It was expected that any population members would be closest to themselves and thereafter to members of other populations, which was also pertinent in the present study. Consistent with the above observations drawn from F_{st} , PCA, ADMIXTURE, f_3 statistics and fineSTRUCTURE, there is no signature of large scale population replacement in the Indian subcontinent.

In conclusion, our high resolution analysis portraying the three ancient tribal populations, strongly rejects any incoming genetic signal of large scale recent (during the post-Neolithic) migration either of the present Dravidian or the Indo-European speaking populations to the subcontinent. We also concluded that the Indian populations preserve strong genetic

Table 2. The genetic diversity calculations of Bhil, Kol and Gond with respect to other South Asian groups.

Ethnic group	Genetic Diversity % (1-iBS) ± SD
Pakistan	26.89 ± 0.38
Bhil	25.26 ± 0.15
Kol	25.69 ± 0.32
Gond	25.66 ± 0.43
Indian IE	26.09 ± 0.34
Indian DRA	26.14 ± 0.47
Indian AA	25.23 ± 0.36

IE—Indo-European, DRA—Dravidian, AA—Austroasiatic (Munda)

doi:10.1371/journal.pone.0127655.t002

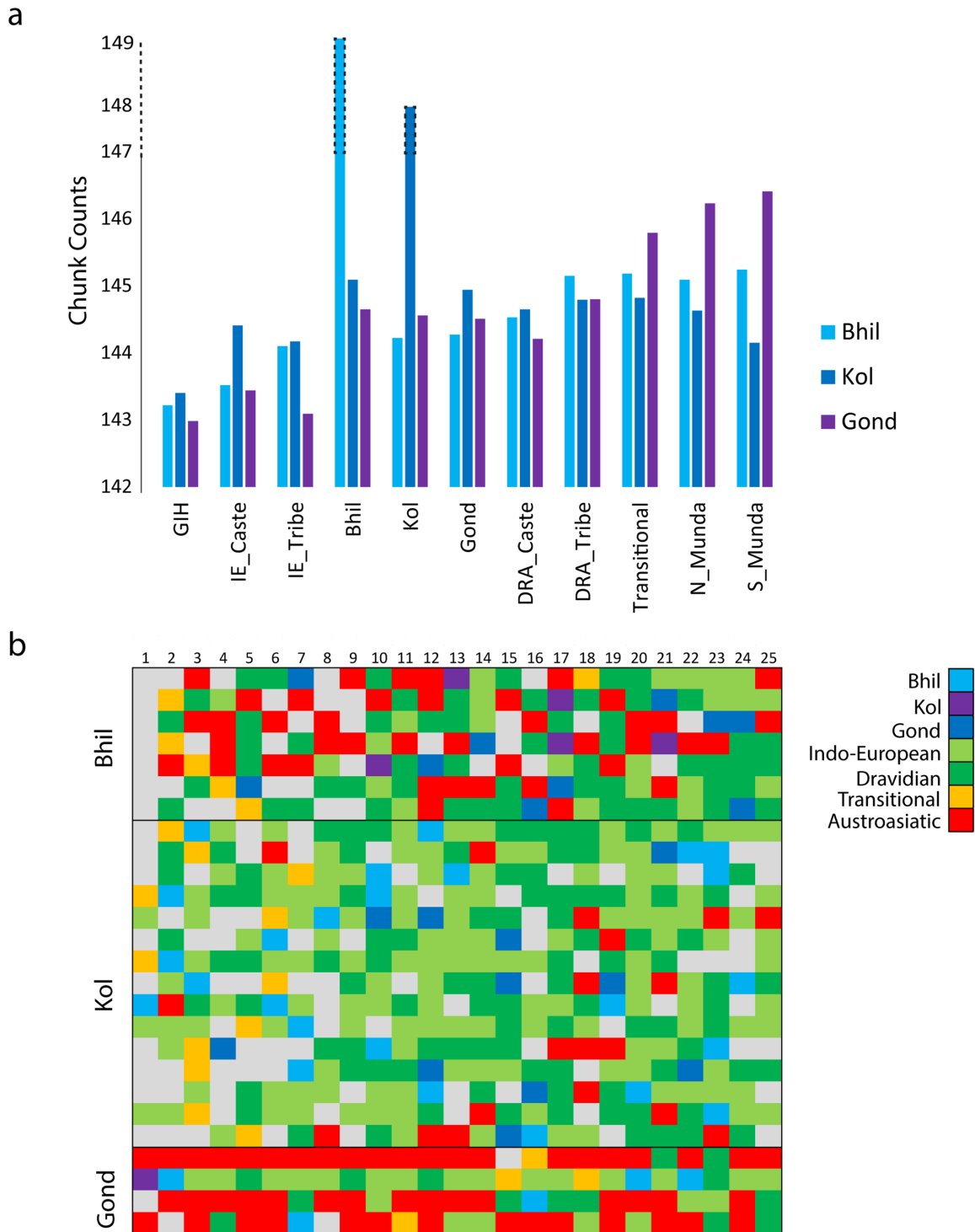


Fig 2. a) The number of chunks donated at inter and intra populations level for the Bhil, Kol and Gond with respect to the Indian, Central Asian and the Pakistani populations. b) Plot of 25 nearest neighbors of Bhil, Kol and Gond individuals. The match population individuals are colored in grey.

doi:10.1371/journal.pone.0127655.g002

signatures in support of a common ancestry. The studied tribal populations do share large number of genome among themselves as well as from o caste and tribal populations. Notably, the placement of various populations along the Indian cline is not solely governed by the geography, but also by the caste-tribe interaction and various other selectional forces. These patterns point to a complex demographic history of the subcontinent which has been shaped in-situ by admixture events at different time scale, as well as by intricate geographical heterogeneity and long term effect of several evolutionary forces.

Supporting Information

S1 Fig. Neighbour Joining (NJ) tree world populations inferred from *Fst* distances of genome-wide data. In the inset, the heatmap showing the inter and intra regional genetic affinity of the three tribal populations under investigation.

(TIF)

S2 Fig. The plot of shared drift obtained by the $f_3 = (\text{Yoruba; Bhil/Kol/Gond, X})$. The f_3 values are plotted on Y axis against the X- targeted populations on X axis. C_Asia- Central Asia, IN_IE_Caste- Indian Indo-European Caste, IN_IE_Tribe—Indian Indo-European Tribe, IN_DRA_Caste- Indian Dravidian Caste, IN_DRA_Tribe—Indian Dravidian Tribe, IN_AA- Indian Austroasiatic (Munda).

(TIF)

S3 Fig. Co-ancestry matrix plotted from fineSTRUCTURE analysis, showing the chunks donated by other Eurasian populations to the Bhil, Kol and Gond populations.

(TIF)

S4 Fig. Placement of Kol, Bhil and Gond individuals over the 37 clades obtained from the fineSTRUCTURE analysis.

(TIF)

S1 Table. The details of the populations (number of individuals and number of SNPs, used in the present study).

(XLSX)

Author Contributions

Conceived and designed the experiments: GC. Performed the experiments: GC AK SB VRR. Analyzed the data: GC. Contributed reagents/materials/analysis tools: AK SB VRR. Wrote the paper: GC AK SB VRR.

References

1. Cavalli-Sforza LL, Feldman MW (2003) The application of molecular genetic approaches to the study of human evolution. *Nat Genet* 33 Suppl: 266–275. PMID: [12610536](#)
2. HUGO Pan-Asian SNP Consortium, Abdulla MA, Ahmed I, Assawamakin A, Bhak J, Brahmachari SK et al. (2009) Mapping human genetic diversity in Asia. *Science* 326: 1541–1545. doi: [10.1126/science.1177074](#) PMID: [20007900](#)
3. Chaubey G, Metspalu M, Kivisild T, Villems R (2007) Peopling of South Asia: investigating the caste-tribe continuum in India. *Bioessays* 29: 91–100. PMID: [17187379](#)
4. Boivin N (2007) Anthropological, historical, archaeological and genetic perspectives on the origins of caste in South Asia. *The Evolution and History of Human Populations in South Asia*: 341–361.
5. Chaubey G. **The demographic history of India: A perspective based on genetic evidence** (<http://hdl.handle.net/10062/15240>). *PhD. Universitatis Tartuensis, Evolutionary Biology*; (2010).
6. Raychaudhuri H *Political history of ancient India*. Genesis Publishing Pvt Ltd; UK (2006).

7. Kazanas N (1999) The R. gveda and Indo-Europeans. Annals of the Bhandarkar Oriental Research Institute, Pune, India.
8. Bala Saroj and Mishra Kulbhusan (Ed):*Historicity of Vedic and Ramayan Eras*. Institute of Scientific Research on Vedas, Hyderabad, India; (2012).
9. Wheeler M (1979) Harappan Chronology and the Rig Veda. In *Ancient Cities of the Indus*, edited by Possehl G. L.. New Delhi: Vikas Publishing House Pvt Ltd.
10. Indian Heritage Research Foundation *Encyclopedia of Hinduism*. Mandala Publishing Group; USA (2013).
11. Tiwari SK *Tribal roots of Hinduism*. Sarup & Sons; New Delhi, India (2002).
12. Frawley D (Ed):*The Rig Veda and the history of India*. (2001), Rig Veda Bharata Itihasa. Aditya Prakashan. India.
13. Talageri S. *The Rigveda: a historical analysis*. Aditya Prakashan, India; (2000)
14. Achar BN (2000) A CASE FOR REVISING THE DATE OF VEDANGA JYOTISA. *Indian Journal of History of Science* 35: 73–183.
15. Lal B. *Deep Roots of Indian Civilization*. New Delhi: Aryan Book International; India (2009)
16. **Srimad Valmiki-Ramayana**. Gita Press Gorakhpur, India
17. **Gita Press, Gorakhpur India** [<http://www.gitapress.org/>]
18. Thapar R. *Ancient Indian Social History*. Orient Blackswan; USA (1978).
19. Ramadas G (1925) The Aboriginal Tribes in the Ramayana. *Man in India* 5: 1–2.
20. Sharma K, Tulasīdāsa. *Tulasidas Ramayana*. (1967)
21. Kumar BB. *Naga Identity*. Concept Publishing Company; India (2005)
22. Vidyarthi LP, Rai BK. *The Tribal Culture of India*. Concept Publishing Company; India (1977)
23. Ananda W P Guruge. *The Society of the Ramayana*. Abhinav Publications; India (1991)
24. Poddar HP. **Sri Ramcharitmanas**. Gita Press Gorakhpur, India.
25. Govind H, Bhatt P (1961) Index of Valmiki Ramayana in two Volumes. The Maharaja Sayajirao University of Baroda, India
26. Bhatt G, Vaidya P, Divanji P, Mankad D, Jhala G, Shah U (1975) Critical Edition of Complete Valmiki Ramayana (Seven Volumes),. The Maharaja Sayajirao University of Baroda, India
27. Chattopadhyaya B. *Studying Early India*. Orient Blackswan; India (2003)
28. Singh KS. *People of India*. Oxford: Oxford University Press; India (1997)
29. Russell RV. *The tribes and castes of the Central Provinces of India*. Macmillan and Co., limited; (1916)
30. Bhatia H, Rao V (1986) Genetic Atlas of Indian Tribes. Institute of Immunohaematology, New Delhi: Indian Council of Medical Research: 242–254.
31. Mastana SS, Papiha SS (1994) Genetic structure and microdifferentiation among four endogamous groups of Maharashtra, western India [published erratum appears in *Ann Hum Biol* 1994 Jul-Aug;21 (4):398]. *Ann Hum Biol* 21: 241–262. PMID: [8060113](https://pubmed.ncbi.nlm.nih.gov/8060113/)
32. Mountain JL, Hebert JM, Bhattacharyya S, Underhill PA, Ottolenghi C, Gadgil M et al. (1995) Demographic history of India and mtDNA-sequence diversity. *Am J Hum Genet* 56: 979–992. PMID: [7717409](https://pubmed.ncbi.nlm.nih.gov/7717409/)
33. Kivisild T, Rootsi S, Metspalu M, Mastana S, Kaldma K, Parik J 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. PMID: [12536373](https://pubmed.ncbi.nlm.nih.gov/12536373/)
34. Sengupta S, Zhivotovsky LA, King R, Mehdi SQ, Edmonds CA, Chow CE et al. (2006) Polarity and temporality of high-resolution y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of Central Asian pastoralists. *Am J Hum Genet* 78: 202–221. PMID: [16400607](https://pubmed.ncbi.nlm.nih.gov/16400607/)
35. Xing J, Watkins WS, Hu Y, Huff CD, Sabo A, Muzny DM et al. (2010) Genetic diversity in India and the inference of Eurasian population expansion. *Genome Biol* 11: R113. doi: [10.1186/gb-2010-11-11-r113](https://doi.org/10.1186/gb-2010-11-11-r113) PMID: [21106085](https://pubmed.ncbi.nlm.nih.gov/21106085/)
36. Chandrasekar A, Kumar S, Sreenath J, Sarkar BN, Urade BP, Mallick S et al. (2009) Updating phylogeny of mitochondrial DNA macrohaplogroup m in India: dispersal of modern human in South Asian corridor. *PLoS one* 4: e7447. doi: [10.1371/journal.pone.0007447](https://doi.org/10.1371/journal.pone.0007447) PMID: [19823670](https://pubmed.ncbi.nlm.nih.gov/19823670/)
37. Reich D, Thangaraj K, Patterson N, Price AL, Singh L (2009) Reconstructing Indian population history. *Nature* 461: 489–494. doi: [10.1038/nature08365](https://doi.org/10.1038/nature08365) PMID: [19779445](https://pubmed.ncbi.nlm.nih.gov/19779445/)

38. Metspalu M, Romero IG, Yunusbayev B, Chaubey G, Mallick CB, Hudjashov G 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. doi: [10.1016/j.ajhg.2011.11.010](https://doi.org/10.1016/j.ajhg.2011.11.010) PMID: [22152676](https://pubmed.ncbi.nlm.nih.gov/22152676/)
39. Chaubey G, Metspalu M, Choi Y, Mägi R, Romero IG, Soares P 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. doi: [10.1093/molbev/msq288](https://doi.org/10.1093/molbev/msq288) PMID: [20978040](https://pubmed.ncbi.nlm.nih.gov/20978040/)
40. Moorjani P, Thangaraj K, Patterson N, Lipson M, Loh PR, Govindaraj P et al. (2013) Genetic evidence for recent population mixture in India. *Am J Hum Genet* 93: 422–438. doi: [10.1016/j.ajhg.2013.07.006](https://doi.org/10.1016/j.ajhg.2013.07.006) PMID: [23932107](https://pubmed.ncbi.nlm.nih.gov/23932107/)
41. Krithika S, Maji S, Vasulu TS (2009) A microsatellite study to disentangle the ambiguity of linguistic, geographic, ethnic and genetic influences on tribes of India to get a better clarity of the antiquity and peopling of South Asia. *Am J Phys Anthropol* 139: 533–546. doi: [10.1002/ajpa.21018](https://doi.org/10.1002/ajpa.21018) PMID: [19278019](https://pubmed.ncbi.nlm.nih.gov/19278019/)
42. Thangaraj K, Naidu BP, Crivellaro F, Tamang R, Upadhyay S, Sharma VK et al. (2010) The influence of natural barriers in shaping the genetic structure of maharashtra populations. *PLoS one* 5: e15283. doi: [10.1371/journal.pone.0015283](https://doi.org/10.1371/journal.pone.0015283) PMID: [21187967](https://pubmed.ncbi.nlm.nih.gov/21187967/)
43. Sharma G, Tamang R, Chaudhary R, Singh VK, Shah AM, Anugula S et al. (2012) Genetic affinities of the central Indian tribal populations. *PLoS one* 7: e32546. doi: [10.1371/journal.pone.0032546](https://doi.org/10.1371/journal.pone.0032546) PMID: [22393414](https://pubmed.ncbi.nlm.nih.gov/22393414/)
44. Gazi NN, Tamang R, Singh VK, Ferdous A, Pathak AK, Singh M 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. doi: [10.1371/journal.pone.0075064](https://doi.org/10.1371/journal.pone.0075064) PMID: [24130682](https://pubmed.ncbi.nlm.nih.gov/24130682/)
45. Chaubey G, Singh M, Crivellaro F, Tamang R, Nandan A, Singh K et al. (2014) Unravelling the distinct strains of Tharu ancestry. *Eur J Hum Genet* 22: 1404–1412. doi: [10.1038/ejhg.2014.36](https://doi.org/10.1038/ejhg.2014.36) PMID: [24667789](https://pubmed.ncbi.nlm.nih.gov/24667789/)
46. International HapMap 3 Consortium, Altshuler DM, Gibbs RA, Peltonen L, Dermitzakis E, Schaffner SF et al. (2010) Integrating common and rare genetic variation in diverse human populations. *Nature* 467: 52–58. doi: [10.1038/nature09298](https://doi.org/10.1038/nature09298) PMID: [20811451](https://pubmed.ncbi.nlm.nih.gov/20811451/)
47. Metspalu M, Kivisild T, Metspalu E, Parik J, Hudjashov G 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. PMID: [15339343](https://pubmed.ncbi.nlm.nih.gov/15339343/)
48. Mastana SS (2014) Unity in diversity: an overview of the genomic anthropology of India. *Ann Hum Biol* 41: 287–299. doi: [10.3109/03014460.2014.922615](https://doi.org/10.3109/03014460.2014.922615) PMID: [24932744](https://pubmed.ncbi.nlm.nih.gov/24932744/)
49. Shaffer J. The Indo-Aryan Invasions: Cultural Myth and Archaeological Reality. In *The People of South Asia: The Biological Anthropology of India, Pakistan and Nepal*. Edited by Lukacs J. London: Plenum; (1984) 77–90.
50. Renfrew C (1991) The coming of the Aryans to Iran and India and the cultural and ethnic identity of the Dasas. By Asko Parpola. (*Studia Orientalia*, Vol. 64.) pp. 195–302, 33 figs Helsinki, The Finnish Oriental Society, 1988. *Journal of the Royal Asiatic Society (Third Series)* 1: 106–109.
51. Bryant E, Patton LL (Ed): *The Indo-Aryan controversy*. RoutledgeCurzon; (2005)
52. Thanseem I, Thangaraj K, Chaubey G, Singh VK, Bhaskar LVKS et al. (2006) Genetic affinities among the lower castes and tribal groups of India: inference from Y chromosome and mitochondrial DNA. *BMC Genet* 7: 42. PMID: [16893451](https://pubmed.ncbi.nlm.nih.gov/16893451/)
53. Witzel M. **Central Asian roots and acculturation in Indian subcontinent: linguistic and archaeological evidence from Western Central Asia, the Hindukush and northwestern Indian subcontinent for early Indo-Aryan language and religion.** In *Linguistics, Archaeology and the Human Past*. Edited by Osada T. Kyoto, Japan: Research Institute for Humanity and Nature, Kyoto; (2005) 87–211.
54. Poliakov L. *The Aryan Myth*. New York: Basic Books; (1974)
55. Agrawal D (1995) Demise of the Aryan Invasion Theory. Hindu Sevak Kendra, Mumbai, India
56. Lal B. *The Earliest Civilization of South Asia: Rise, Maturity, and Decline*. Aryan Books International; New Delhi, India (1997)
57. Lal BB, Saraswat K. *The Homeland of the Aryans: Evidence of Rigvedic Flora and Fauna & Archaeology*. Aryan Books International New Delhi, India (2005)
58. Kennedy K. **Have Aryans been Identified in the Prehistoric Skeletal Record from South Asia? Biological Anthropology and Concepts of Ancient Races.** In *The Indo-Aryans of Ancient South Asia*. Edited by Erdosy G. New Delhi: Munshiram Manoharlal; (1995) 33–66.
59. Sahoo S, Singh A, Himabindu G, Banerjee J, Sitalaximi T et al. (2006) A prehistory of Indian Y chromosomes: evaluating demic diffusion scenarios. *Proc Natl Acad Sci U S A* 103: 843–848. PMID: [16415161](https://pubmed.ncbi.nlm.nih.gov/16415161/)

60. Trivedi R, Sahoo S, Singh A, Bindu G, Banerjee J et al. (2008) Genetic Imprints of Pleistocene Origin of Indian Populations: A Comprehensive Phylogeographic Sketch of Indian Y-Chromosomes. *Int J Hum Genet* 8: 97–118.
61. Li JZ, Absher DM, Tang H, Southwick AM, Casto AM et al. (2008) Worldwide human relationships inferred from genome-wide patterns of variation. *Science* 319: 1100–1104. doi: [10.1126/science.1153717](https://doi.org/10.1126/science.1153717) PMID: [18292342](https://pubmed.ncbi.nlm.nih.gov/18292342/)
62. Yunusbayev B, Metspalu M, Järve M, Kutuev I, Rootsi S et al. (2012) The Caucasus as an Asymmetric Semipermeable Barrier to Ancient Human Migrations. *Mol Biol Evol* 29: 359–365. doi: [10.1093/molbev/msr221](https://doi.org/10.1093/molbev/msr221) PMID: [21917723](https://pubmed.ncbi.nlm.nih.gov/21917723/)
63. Behar DM, Yunusbayev B, Metspalu M, Metspalu E, Rosset S et al. (2010) The genome-wide structure of the Jewish people. *Nature* 466: 238–242. doi: [10.1038/nature09103](https://doi.org/10.1038/nature09103) PMID: [20531471](https://pubmed.ncbi.nlm.nih.gov/20531471/)
64. Chaubey G, Metspalu M, Karmin M, Thangaraj K, Rootsi S et al. (2008) Language shift by indigenous population: a model genetic study in South Asia. *International Journal of Human Genetics* 8: 41.
65. Kumar V, Reddy A, Babu P, Rao TN, Thangaraj K et al. (2008) Molecular Genetic Study on the Status of Transitional Groups of Central India: Cultural Diffusion or Demic Diffusion? *BMC Biol* 8: 31.
66. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81: 559–575. PMID: [17701901](https://pubmed.ncbi.nlm.nih.gov/17701901/)
67. Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. *PLoS Genet* 2: e190. PMID: [17194218](https://pubmed.ncbi.nlm.nih.gov/17194218/)
68. Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 19: 1655–1664. doi: [10.1101/gr.094052.109](https://doi.org/10.1101/gr.094052.109) PMID: [19648217](https://pubmed.ncbi.nlm.nih.gov/19648217/)
69. Cockerham CC, Weir BS (1984) Covariances of relatives stemming from a population undergoing mixed self and random mating. *Biometrics* 40: 157–164. PMID: [6733226](https://pubmed.ncbi.nlm.nih.gov/6733226/)
70. Felsenstein J *Documentation of PHYLIP (Phylogeny Inference Package) version 3.5c*. Seattle: University of Washington; (1993)
71. Tamura K, Stecher G, Peterson D, Filipiński A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* 30: 2725–2729. doi: [10.1093/molbev/mst197](https://doi.org/10.1093/molbev/mst197) PMID: [24132122](https://pubmed.ncbi.nlm.nih.gov/24132122/)
72. Lawson DJ, Hellenthal G, Myers S, Falush D (2012) Inference of population structure using dense haplotype data. *PLoS Genet* 8: e1002453. doi: [10.1371/journal.pgen.1002453](https://doi.org/10.1371/journal.pgen.1002453) PMID: [22291602](https://pubmed.ncbi.nlm.nih.gov/22291602/)
73. Browning BL, Yu Z (2009) Simultaneous genotype calling and haplotype phasing improves genotype accuracy and reduces false-positive associations for genome-wide association studies. *Am J Hum Genet* 85: 847–861. doi: [10.1016/j.ajhg.2009.11.004](https://doi.org/10.1016/j.ajhg.2009.11.004) PMID: [19931040](https://pubmed.ncbi.nlm.nih.gov/19931040/)
74. Qamar R, Ayub Q, Mohyuddin A, Helgason A, Mazhar K et al. (2002) Y-chromosomal DNA variation in Pakistan. *Am J Hum Genet* 70: 1107–1124. PMID: [11898125](https://pubmed.ncbi.nlm.nih.gov/11898125/)
75. Quintana-Murci L, Chaix R, Wells RS, Behar DM, Sayar H 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. PMID: [15077202](https://pubmed.ncbi.nlm.nih.gov/15077202/)
76. Lazaridis I, Patterson N, Mittnik A, Renaud G, Mallick S et al. (2014) Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* 513: 409–413. doi: [10.1038/nature13673](https://doi.org/10.1038/nature13673) PMID: [25230663](https://pubmed.ncbi.nlm.nih.gov/25230663/)