Man In India, 97 (1): 279–290

THE GENOME-WIDE ANALYSIS OF THE BHILS: THE SECOND LARGEST TRIBAL POPULATION OF INDIA

Gyaneshwer Chaubey, Periyasamy Govindaraj, Niraj Rai, George van Driem and Kumarasamy Thangaraj

The name of the Bhil population was mentioned in early literature of the Subcontinent, which suggests their presence in India since prehistoric times. Studies based on classical and genetic markers suggest a unique identity of the Bhil, who currently live in western and central Indian states. Our previous studies on two Bhil groups living in central and western Indian regions drew different conclusions on their origin. However, the first study on the Bhil of central India was based on haploid DNA and a few autosomal markers, whereas the second study on the western Bhil explored large number of autosomal SNPs. Therefore, in this study we have reconnoitered the inter-population and intra-population relationships of Bhil groups at four different geographical locations by using >95,000 autosomal SNPs. A combination of statistical analysis revealed that all Bhil populations are likely to have had a common source sharing a pan-Bhil ancestry. This common ancestry is clearly seen amongst the Bhil of Gujarat who turned out to show the lowest degree of admixture with their neighbours, whereas the Bhil of Rajasthan showed the highest diversity with extensive admixture with the surrounding populations. Both inter-population and intra-population comparison suggest a shared Bhil genome followed by chunks sharing with the Nihali population, a language community speaking a so-called language isolate.

Introduction

The classical division and endogamy of Indian society, both for caste and tribal populations, has undoubtedly played a major role in shaping the genetic landscape of the Subcontinent (Chaubey *et al*. 2007). Though each and every caste and tribe population of the Subcontinent is unique, a few tribal and caste populations stand out clearly either in terms of their cultural manifestation or phenotypical appearance and genetic make-up (Thangaraj *et al*. 2006; Chaubey *et al.* 2008, 2010, 2014; Sharma *et al.* 2012). The Bhil are one such group which show considerable divergence from the neighbouring clusters in classical and modern DNA studies (Bhatia and Rao 1986; Singh 1997; Maheshwari *et al.* 1986; Lim *et al.* 2012).

The Bhil are the second largest tribal population of India, only the Gond being more numerous. The Bhil are the largest tribal population of the Rajasthan-Gujarat region and the second largest in the Maharashtra-Madhya Pradesh region (Singh 1997). The Bhil are mentioned in the *Nāradapurāṇa*, *Rāmāyaṇa* and *Mahābhārata* (Ramadas 1925; Govind *et al.* 1961; Sharma and Tulasīdāsa 1967; Vidyarthi and Rai 1977). It has been suggested that the ethnonym Bhil is derived from Dravidian *vil* 'bow', a hunting accoutrement widely used by the tribes (Maheshwari *et al*. 1986). The Bhil make up various language communities dispersed across the states of Maharashtra, Gujarat, Madhya Pradesh and

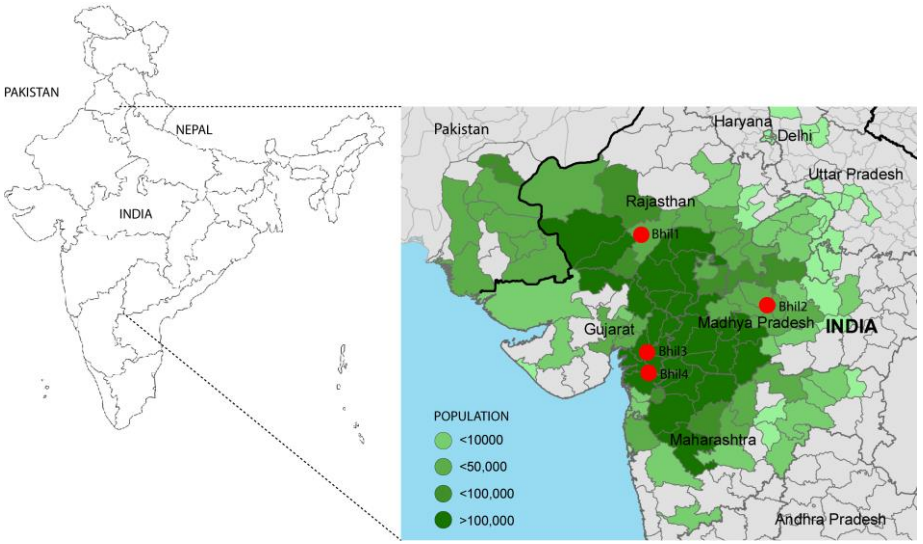

Figure 1: The geographical distribution of Bhils highlighting our sampling locations.

Rajasthan (Figure 1). In these communities, the Bhil speak over fifty different 'dialects', which are actually local tribal lects of a variety of local Indo-Aryan languages. Some of the Bhili languages are categorised as dialects of Gujarati, whereas others are categorised as dialects of particular languages of Rajasthan. Wagdi is the single most widely spoken Bhili lect. The Nihali have long been held to share both racial and cultural similarities with the geographically proximate Bhil (van Driem 2001). In 1954, Robert Shafer believed that the language isolate spoken by the Nihali might represent the only surviving remnant of the original linguistic stock of the linguistically Aryanised Bhil, whom Shafer identified with the *Niṣāda* of the Vedas (Shafer 1954). In 1880, James Campbell had rejected the equation of Bhil with the Vedic *Niṣāda* as simplistic, but he recorded the local view that the Nihali were looked upon by the Bhil as a type of Bhil, conversely linguistically assimilated Bhil considered themselves to be superior to the Nihali language community, which had retained its ancestral non-Indo-Aryan language (Campbell 1880).

With few exceptions, most of the genetic studies of the Subcontinent have focused on broader issues such as the initial peopling, the origin of different language groups, the relations between caste and tribe populations, the origin and migration of East and West Eurasian lineages etc. (Metspalu *et al.* 2004, 2011; Palanichamy *et al.* 2004; Chaubey *et al.* 2007,2011; Debnath *et al.* 2011; Xing *et al.* 2010; Chandrasekar *et al.* 2009; Reich *et al.* 2009; Moorjani *et al.* 2013). Our

previous study on the Bhil population of Central India suggested an affinity with the neighbouring Indo-European populations (Sharma *et al.* 2012). The Bhil exhibited a high frequency of haplogroup H1a-M82 and R1a-M17 on the Y chromosome, whereas M2, M3, M5, M30 and R5 were the most frequent mtDNA haplogroups. Notably haplogroups M37, M49, R31 and U1 were exclusively found amongst the Bhil population when compared with the neighbouring populations (Sharma *et al.* 2012). Conversely, our subsequent analysis on hundreds of thousands of autosomal markers in the Bhil population of Gujarat suggested their closer affinity with the Dravidian tribes and Munda speaking populations (Chaubey *et al.* 2015). This contrasting result would appear to suggest some genetic heterogeneity amongst the Bhil. However, the reduced genetic sharing of the Bhil population in Gujarat was striking in showing the lowest degree of genetic sharing with the neighbouring populations of Gujarat and Rajasthan. In order to resolve such issues as well as to obtain a high resolution view of the inter-population and intra-population relations of the Bhil population, we have analysed >95000 autosomal SNPs among four distinct Bhil groups coming from three Indian states (Fig. 1).

Results and Discussion

For the ease of understanding, we have renamed the Bhil of Rajasthan as Bhil1, the Bhil of Madhya Pradesh as Bhil2, and two Bhil populations of Gujarat as Bhil3 and Bhil4 (Fig. 1). We first measured the genetic distances (*F*st) with respect to other Eurasian groups. The intra-population comparison showed a closer affinity of Bhil1 to Bhil2 and of Bhil3 to Bhil4, whereas the genetic distances increased twofold when we compared Bhil1 or Bhil2 vs. Bhil3 or Bhil4 (0.0074 to 0.0154) (Fig. 2a). The *F*st analysis clustered four Bhil groups into two quite distinct Bhil super-groups. In the inter-population comparison, Bhil1 and Bhil2 were closer to Indian Indo-European, Gujarati and North Pakistani groups, whilst Bhil3 and Bhil4 were nearer to Indian Dravidian populations (Fig. 2a). The affinity of Bhil2 fit with our previous observations on haploid markers (Sharma *et al.* 2012). Therefore, the *F*st analysis showed a common ancestry of Bhil1 with Bhil2 and of Bhil3 with Bhil4. However, the large distance of Bhil1-Bhil2 as against Bhil3-Bhil4 suggests either a deep common pan-Bhil ancestry or otherwise a process of cultural assimilation whereby the integration of other tribal groups into the Bhil population may have erased such signatures.

We furthermore performed PCA and ADMIXTURE analyses with the settings described elsewhere (Metspalu *et al.* 2011). All the Bhil groups fell on the Indian cline, albeit some distance apart from one another (Fig. 2b).

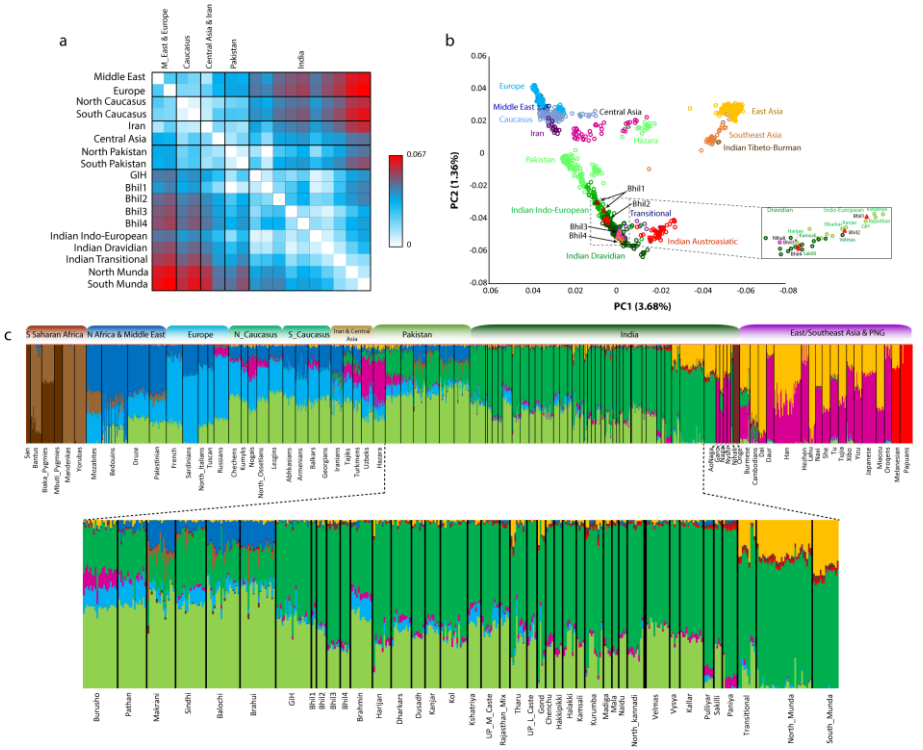

Figure 2: a) Regional comparison of population differentiation (*F*st) analysis of Bhil groups with the Indian and other Eurasian populations. b) PCA (Principle Component Analysis) of Eurasian populations showing the placement of Bhil groups over the South Asian cline; the mean value of populationwise clustering of Bhil and neighbours is zoomed-in inset figure. c) Individualwise ancestry proportion analysis inferred from ADMIXTURE representing ten ancestral populations of the world (K = 10).

Consistent with the *F*st analysis, the Bhil1 was found to be closer to the Indo-European groups, whereas Bhil3-Bhil4 were nearer to the Dravidian populations. Bhil2 attained an intermediate position between the Bhil1 population and the Bhil3-Bhil4 group (Fig. 2b). To gain a better understanding of population wise clustering, we plotted the mean PC values of the Bhil and their neighbouring populations. Bhil1 formed a cluster with Rajasthani and Gujarati populations, whilst Bhil2 clustered with the Indo-European tribes and Dravidian caste populations. Bhil3 and Bhil4 were likewise grouped together with the other Dravidian populations (Fig. 2b).

Table 1: The ancestral north indian (ANI) ancestry calculated in different Bhil groups

Population	% ANI (± SD)	Z
Bhil1	48.84 ± 1.77	-15.85
Bhil2	39.63 ± 1.02	-19.07
Bhil3	26.54 ± 1.17	-14.01
Bhil4	26.7 ± 1.32	-13.01

Our ADMIXTURE analysis was compatible with the *F*st and PCA analysis (Fig. 2c). All of the Bhil groups showed variable amounts of the two major components prevalent amongst South Asian populations. The estimation of ANI admixture was found to decrease in following order from Bhil1>Bhil2>Bhil4>Bhil3 (Table 1). It is interesting to note that the genetic placement of Bhi1 and Bhi2 was found to be largely consistent with their geographical pattern of settlement distribution. However, Bhil3 and Bhil4 showed a closer attraction towards the populations living to the South. Such observations challenge the scenario which suggests that the genetic canvas of the Subcontinent is mainly governed by the geographical landscape (Kivisild *et al.* 2003; Chaubey *et al.* 2008). The low level of heterozygosity and high *F*st values would lead us to reject assuming any recent migration of Bhil3 and Bhil4 from South.

We performed a outgroup *f3* test, using the Yoruba as the outliers (Fig. 3). All of the Bhil groups showed a significant level of allele sharing with other South Asian populations, whilst the level of allele sharing with non-South Asian populations was insignificant. The *f3* test supported the intrarelation of Bhils observed from previous analysis. It has also signalled the unexpected higher level of allele sharing between Bhil3-Bhil4 and Nihali (Fig. 3).

In order to gain better insight into the Bhil groups with regard to South Asian ancestry, we utilised fineSTRUCTURE analysis, which produced a co-ancestry matrix by using a haplotype-based approach (Lawson *et al.* 2012). In agreement with the previous analyses, the Bhil1 and Bhil2 populations has evidently received a high number of chunks from the neighbouring populations, whereas the Bhil3 and Bhil4 populations received more chunks from Dravidian and Munda speakers (Fig. 4a). In order to ascertain the major Indian donors to the Bhil population, we plotted the mean chunk count donated by the various Indian groups against all of the Bhil groups. Interestingly, we found that the

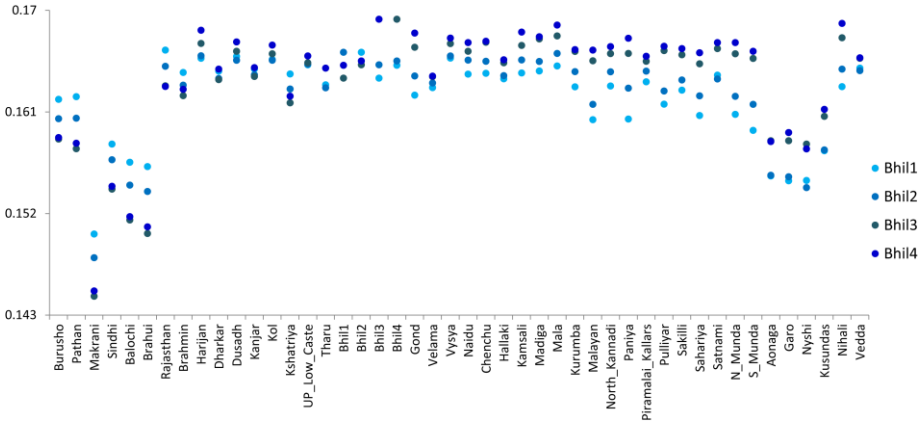

Figure 3: Outgroup *f3* statistics, showing the allele sharing of different Bhil groups with each other as well as with other South Asian populations.

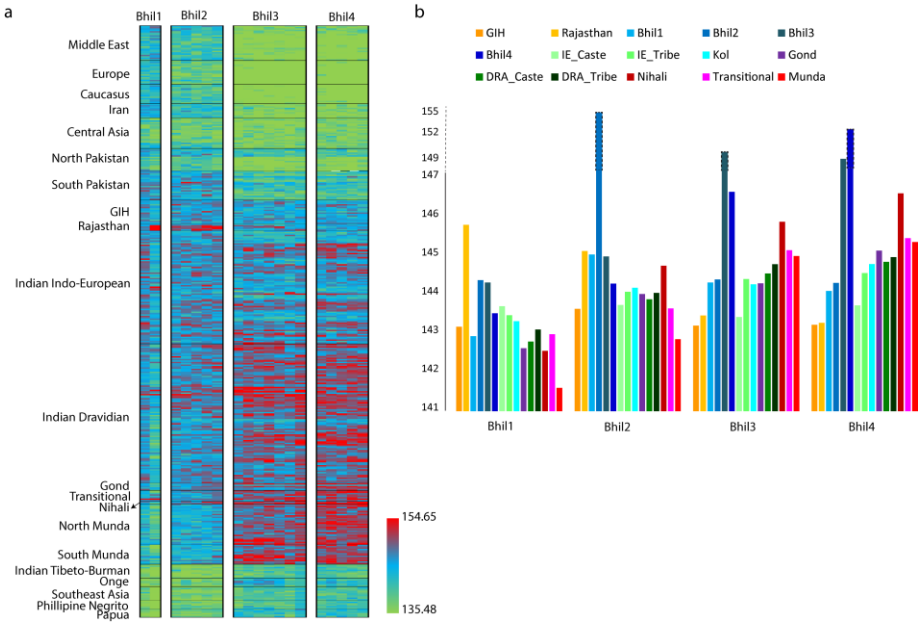

Figure 4: a) the co-ancestry matrix plot of chunkcount received by different Bhil groups from the other Indian and Eurasian populations; b) The number of chunks donated at inter and intra populations level for the Bhil groups from other Indian populations.

language isolate Nihali, residing in Central India, had donated a significantly higher number of chunks to the Bhil3 and Bhil4 than any other group had done (two tailed p value 0.039 for Bhil3 and for Bhil4 0.007), leaving out the chunks donated to each other (Fig. 4b). The populations Rajasthan, Bhil1, Bhil3 and Nihali were major chunk donors to Bhil2, whilst Rajasthan, Bhil2 and Bhil3 were the first and foremost donors to the population Bhil1. Therefore, a common ancestry with Nihali is clearly manifest amongst the Bhil2, Bhi3 and Bhil4 populations. An excessive amount of Nihali chunks is not evident in the Bhil1 group. Nevertheless the high number of chunks from Bhil3 and Bhil4 provide clear evidence for the view that all of the Bhil populations originated from a common pan-Bhil stock.

Overall our high resolution analysis for Bhil2 supported our previous observations obtained from haploid DNA analysis. The fineSTRUCTURE analysis revealed a most recent common ancestry of the Bhil with the Nihali population of central India. It is likely that the excessive number of chunks from Nihali to Bhil1 might have been lost secondarily due to a high level of admixture with local Rajasthani and Gujarati 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) (Li *et al.* 2008; Reich *et al.* 2009; Behar *et al.* 2010; Chaubey *et al.* 2011; Metspalu *et al.* 2011;Yunusbayev *et al.* 2011; Moorjani *et al.* 2013). All of the ethical guidelines were followed. The tribal and caste populations were grouped according to their language group. We grouped populations into a group which we labelled "transitional" when these were known to have undergone language shift in recent time (Kumar *et al.* 2008; Chaubey *et al.* 2008). 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 (Purcell *et al.* 2007). We used PLINK 1.07 (Purcell *et al.* 2007) 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 (Patterson *et al.* 2006) and ADMIXTURE (Alexander *et al.* 2009), we thinned the dataset by removing one SNP of any pair, in strong LD r2>0.4, in a window of 200 SNPs (sliding the window by 25 SNPs at a time).

We performed PC analysis using the *smartpca* programme of the EIGENSOFT package in the default settings (Patterson *et al.* 2006) in order to capture genetic variability described by the first five components. The fraction of 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 and remained virtually identical within 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 *F*st distance measure by following the methods as described by Cockerham and Weir (1984), and Phylip (Felsenstein 1993) and MEGA (Tamura *et al.* 2013) were used to construct the tree. To investigate the outgroup allele sharing of Bhil groups with Eurasian populations, we computed *f3* statistics (Patterson *et al.* 2012), taking the Yoruba as outgroups. For haplotype-based analysis (fineSTRUCTURE) (Lawson *et al.* 2012), first samples were phased with Beagle 3.3.2 (Browning and Yu 2009). A co-ancestry matrix was constructed using ChromoPainter (Lawson *et al.* 2012), and fineSTRUCTURE was used to perform an MCMC iteration using 10000000 burning runtime and 10000 MCMC samples.

Address for communication

Gyaneshwer Chaubey, Evolutionary Biology group, Estonian Biocentre, Riia 23, Tartu 51010, Estonia; *Periyasamy Govindaraj, Niraj Raj* and *Kumarasamy Thangaraj*, CSIR, Centre for Cellular and Molecular Biology, Uppal Road, Hyderabad 500007, India; *George van Driem,* Institut für Sprachwissenschaft, Universität Bern, Länggassstrasse 49, 3000 Bern 9, Switzerland.

Acknowledgements

This study was supported by the Estonian Personal grants PUT-766 and Estonian Institutional Research grants IUT24-1. GC acknowledges the financial support from European Union European Regional Development Fund through the Centre of Excellence in Genomics to Estonian Biocentre and University of Tartu by Tartu University grant (PBGMR06901). KT was supported by the Council of Scientific and Industrial Research, Government of India (GENESIS: BSC0121) and (BSC 0208). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Computational analyses were carried out at the High Performance Computing Center, University of Tartu and the Core Computing Unit of the Estonian Biocentre.

References

Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res 2009;19(9):1655-1664.

Behar DM, Yunusbayev B, Metspalu M, Metspalu E, Rosset S, Parik J, Rootsi S, Chaubey G, Kutuev I, Yudkovsky G, Khusnutdinova EK, Balanovsky O, Semino O, Pereira L, Comas D, Gurwitz D, Bonne-Tamir B, Parfitt T, Hammer MF, Skorecki K, Villems R. The genome-wide structure of the Jewish people. Nature 2010;466(7303):238-242.

Bhatia H, Rao V. Genetic Atlas of Indian Tribes. Institute of Immunohaematology, New Delhi: Indian Council of Medical Research 1986;:242-254.

Browning BL, Yu Z. Simultaneous genotype calling and haplotype phasing improves genotype accuracy and reduces false-positive associations for genome-wide association studies. Am J Hum Genet 2009;85(6):847-861.

Campbell JM. Gazetteer of the Bombay Presidency (Vol. XII: Khándesh). 1880.

Chandrasekar A, Kumar S, Sreenath J, Sarkar BN, Urade BP, Mallick S, Bandopadhyay SS, Barua P, Barik SS, Basu D, Kiran U, Gangopadhyay P, Sahani R, Prasad BVR, Gangopadhyay S, Lakshmi GR, Ravuri RR, Padmaja K, Venugopal PN, Sharma MB, Rao VR. Updating phylogeny of mitochondrial DNA macrohaplogroup m in India: dispersal of modern human in South Asian corridor. PloS one 2009;4(10):e7447.

Chaubey G, Metspalu M, Karmin M, Thangaraj K, Rootsi S, Parik J, Solnik A, Rani D, Singh V, Naidu B. Language shift by indigenous population: a model genetic study in South Asia. International Journal of Human Genetics 2008;8(1/2):41.

Chaubey G. The demographic history of India: A perspective based on genetic evidence (http://hdl.handle.net/10062/15240) [PhD]. [Estonia]: Universitatis Tartuensis; 2010.

Chaubey G, Karmin M, Metspalu E, Metspalu M, Selvi-Rani D, Singh VK, Parik J, Solnik A, Naidu BP, Kumar A, Adarsh N, Mallick CB, Trivedi B, Prakash S, Reddy R, Shukla P, Bhagat S, Verma S, Vasnik S, Khan I, Barwa A, Sahoo D, Sharma A, Rashid M, Chandra V, Reddy AG, Torroni A, Foley RA, Thangaraj K, Singh L, Kivisild T, Villems R. Phylogeography of mtDNA haplogroup R7 in the Indian peninsula. BMC Evol Biol 2008;8:227.

Chaubey G, Metspalu M, Choi Y, Mägi R, Romero IG, Soares P, van Oven M, Behar DM, Rootsi S, Hudjashov G, Mallick CB, Karmin M, Nelis M, Parik J, Reddy AG, Metspalu E, van Driem G, Xue Y, Tyler-Smith C, Thangaraj K, Singh L, Remm M, Richards MB, Lahr MM, Kayser M, Villems R, Kivisild T. Population Genetic Structure in Indian Austroasiatic speakers: The Role of Landscape Barriers and Sex-specific Admixture. Mol Biol Evol 2011;28(2):1013-1024.

Chaubey G, Metspalu M, Kivisild T, Villems R. Peopling of South Asia: investigating the caste-tribe continuum in India. Bioessays 2007;29(1):91-100.

Chaubey G, Singh M, Crivellaro F, Tamang R, Nandan A, Singh K, Sharma VK, Pathak AK, Shah AM, Sharma V, Singh VK, Selvi Rani D, Rai N, Kushniarevich A, Ilumäe AM, Karmin M, Phillip A, Verma A, Prank E, Singh VK, Li B, Govindaraj P, Chaubey AK, Dubey PK, Reddy AG, Premkumar K, Vishnupriya S, Pande V, Parik J, Rootsi S, Endicott P, Metspalu M, Lahr MM, van Driem G, Villems R, Kivisild T, Singh L, Thangaraj K. Unravelling the distinct strains of Tharu ancestry. Eur J Hum Genet 2014;22(12):1404-1412.

Chaubey G, Kadiyan A, Bala S, Rao V (2015) Genetic affinity of the Bhil, Kol and Gond mentioned in epic Ramayana. PloS one 10 (6) : p e0127655.

Cockerham CC, Weir BS. Covariances of relatives stemming from a population undergoing mixed self and random mating. Biometrics 1984;40(1):157-164.

Debnath M, Palanichamy MG, Mitra B, Jin JQ, Chaudhuri TK, Zhang YP. Y-chromosome haplogroup diversity in the sub-Himalayan Terai and Duars populations of East India. J Hum Genet 2011;56(11):765-771.

van Driem G. Languages of the Himalayas. Brill Leiden; 2001.

Felsenstein J. Documentation of PHYLIP (Phylogeny Inference Package) version 3.5c. Seattle: University of Washington; 1993.

Govind H, Bhatt, P. Index of Valmiki Ramayana in two Volumes. The Maharaja Sayajirao University of Baroda, India 1961

Kivisild T, Rootsi S, Metspalu M, Mastana S, Kaldma K, Parik J, Metspalu E, Adojaan M, Tolk HV, Stepanov V, Gölge M, Usanga E, Papiha SS, Cinnioğlu C, King R, Cavalli-Sforza L, Underhill PA, Villems R. The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. Am J Hum Genet 2003;72(2):313-332.

Kumar V, Reddy A, Babu P, Rao TN, Thangaraj K, Reddy A, Singh L, Reddy BM. Molecular Genetic Study on the Status of Transitional Groups of Central India: Cultural Diffusion or Demic Diffusion? BMC Biol 2008;8(1/2):31.

Lawson DJ, Hellenthal G, Myers S, Falush D. Inference of population structure using dense haplotype data. PLoS Genet 2012;8(1):e1002453.

Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, Ramachandran S, Cann HM, Barsh GS, Feldman M, Cavalli-Sforza LL, Myers RM. Worldwide human relationships inferred from genome-wide patterns of variation. Science 2008;319(5866):1100-1104.

Lim SS, Vos T, Flaxman AD, Danaei G, Shibuya K, Adair-Rohani H, Amann M, Anderson HR, Andrews KG, Aryee M, Atkinson C, Bacchus LJ, Bahalim AN, Balakrishnan K, Balmes J, Barker-Collo S, Baxter A, Bell ML, Blore JD, Blyth F, Bonner C, Borges G, Bourne R, Boussinesq M, Brauer M, Brooks P, Bruce NG, Brunekreef B, Bryan-Hancock C, Bucello C, Buchbinder R, Bull F, Burnett RT, Byers TE, Calabria B, Carapetis J, Carnahan E, Chafe Z, Charlson F, Chen H, Chen JS, Cheng ATA, Child JC, Cohen A, Colson KE, Cowie BC, Darby S, Darling S, Davis A, Degenhardt L, Dentener F, Des Jarlais DC, Devries K, Dherani M, Ding EL, Dorsey ER, Driscoll T, Edmond K, Ali SE, Engell RE, Erwin PJ, Fahimi S, Falder G, Farzadfar F, Ferrari A, Finucane MM, Flaxman S, Fowkes FGR, Freedman G, Freeman MK, Gakidou E, Ghosh S, Giovannucci E, Gmel G, Graham K, Grainger R, Grant B, Gunnell D, Gutierrez HR, Hall W, Hoek HW, Hogan A, Hosgood HD, Hoy D, Hu H, Hubbell BJ, Hutchings SJ, Ibeanusi SE, Jacklyn GL, Jasrasaria R, Jonas JB, Kan H, Kanis JA, Kassebaum N, Kawakami N, Khang YH, Khatibzadeh S, Khoo JP, Kok C, Laden F. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012;380(9859):2224-2260.

Maheshwari J, Kalakoti B, Lal B. Ethnomedicine of Bhil tribe of Jhabua District, MP. Ancient science of life 1986;5(4):255.

Metspalu M, Kivisild T, Metspalu E, Parik J, Hudjashov G, Kaldma K, Serk P, Karmin M, Behar DM, Gilbert MTP, Endicott P, Mastana S, Papiha SS, Skorecki K, Torroni A, Villems R. 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 2004;5:26.

Metspalu M, Romero IG, Yunusbayev B, Chaubey G, Mallick CB, Hudjashov G, Nelis M, Mägi R, Metspalu E, Remm M, Pitchappan R, Singh L, Thangaraj K, Villems R, Kivisild

T. Shared and unique components of human population structure and genome-wide signals of positive selection in South Asia. Am J Hum Genet 2011;89(6):731-744.

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.

Palanichamy M, Sun C, Agrawal S, Bandelt HJ, Kong QP, Khan F, Wang CY, Chaudhuri T, Palla V, Zhang YP. Phylogeny of mtDNA macrohaplogroup N in India based on complete sequencing: implications for the peopling of South Asia. Am J Hum Genet 2004;75:966-978.

Patterson N, Price AL, Reich D. Population structure and eigenanalysis. PLoS Genet 2006;2(12):e190.

Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y et al. (2012) Ancient admixture in human history. Genetics 192: 1065-1093.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81(3):559-575.

Ramadas G. The Aboriginal Tribes in the Ramayana. Man in India 1925;5:1-2.

Reich D, Thangaraj K, Patterson N, Price AL, Singh L. Reconstructing Indian population history. Nature 2009;461(7263):489-494.

Shafer R. Ethnography of Ancient India. Wiesbaden: Otto Harrassowitz; 1954.

Sharma G, Tamang R, Chaudhary R, Singh VK, Shah AM, Anugula S, Rani DS, Reddy AG, Eaaswarkhanth M, Chaubey G, Singh L, Thangaraj K. Genetic affinities of the central Indian tribal populations. PloS one 2012;7(2):e32546.

Sharma K, Tulasīdāsa. Tulasidas Ramayana. 1967. 241 p.

Singh KS. People of India. Oxford: Oxford University Press; 1997. 1266 p.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol 2013;30(12):2725-2729.

Thangaraj K, Chaubey G, Reddy AG, Singh VK, Singh L. Unique origin of Andaman Islanders: insight from autosomal loci. J Hum Genet 2006;51(9):800-804.

Vidyarthi LP, Rai BK. The Tribal Culture of India. Concept Publishing Company; 1977. 487 p.

Xing J, Watkins WS, Hu Y, Huff CD, Sabo A, Muzny DM, Bamshad MJ, Gibbs RA, Jorde LB, Yu F. Genetic diversity in India and the inference of Eurasian population expansion. Genome Biol 2010;11(11):R113.

Yunusbayev B, Metspalu M, Järve M, Kutuev I, Rootsi S, Metspalu E, Behar DM, Varendi K, Sahakyan H, Khusainova R, Yepiskoposyan L, Khusnutdinova EK, Underhill PA, Kivisild T, Villems R. The Caucasus as an Asymmetric Semipermeable Barrier to Ancient Human Migrations. Mol Biol Evol 2012; 29:359-365.