

ARTICLE

Reconstructing the population history of the largest tribe of India: the Dravidian speaking Gond

Gyaneshwer Chaubey^{*1}, Rakesh Tamang^{2,3}, Erwan Pennarun¹, Pavan Dubey⁴, Niraj Rai⁵, Rakesh Kumar Upadhyay⁶, Rajendra Prasad Meena⁷, Jayanti R Patel⁴, George van Driem⁸, Kumarasamy Thangaraj⁵, Mait Metspalu¹ and Richard Villems^{1,9}

The Gond comprise the largest tribal group of India with a population exceeding 12 million. Linguistically, the Gond belong to the Gondi–Manda subgroup of the South Central branch of the Dravidian language family. Ethnographers, anthropologists and linguists entertain mutually incompatible hypotheses on their origin. Genetic studies of these people have thus far suffered from the low resolution of the genetic data or the limited number of samples. Therefore, to gain a more comprehensive view on ancient ancestry and genetic affinities of the Gond with the neighbouring populations speaking Indo-European, Dravidian and Austroasiatic languages, we have studied four geographically distinct groups of Gond using high-resolution data. All the Gond groups share a common ancestry with a certain degree of isolation and differentiation. Our allele frequency and haplotype-based analyses reveal that the Gond share substantial genetic ancestry with the Indian Austroasiatic (ie, Munda) groups, rather than with the other Dravidian groups to whom they are most closely related linguistically.

European Journal of Human Genetics (2017) 25, 493–498; doi:10.1038/ejhg.2016.198; published online 1 February 2017

INTRODUCTION

The linguistic landscape of India is composed of four major language families and a number of language isolates and is largely associated with non-overlapping geographical divisions. The majority of the populations speak Indo-European languages, which cover a large geographical area including northern and western India.^{1,2} Dravidian languages are spoken primarily in southern India with some exceptions, eg, Brahui in Pakistan, Kurukh–Malto in eastern India and Gondi–Manda languages in central India. Austroasiatic language speakers are scattered in pockets mainly towards eastern and central regions, whereas Tibeto-Burman language speakers are found along the Himalayan fringe and in the Northeast of the subcontinent.^{1,2} The genetic ancestry of Austroasiatic and Tibeto-Burman speakers in the subcontinent strongly correlates with the language. However, geography supersedes when we focus on the Indo-European and Dravidian languages.^{3,4}

The geographical distribution of languages in India is largely non-overlapping.⁵ However, eastern central India presents an amalgam of three major language groups.^{6,7} This region is home to more than 30% of South Asia's tribal populations, some of whom still practise hunting and gathering subsistence strategies.^{8,9} Geographically, the rivers Narmada and Tapti act as abundant water sources, and the mountain ranges Vindhya and Satpura act as a significant geographical barrier to casual interaction with adjoining regions. The complexity of the geography and the fact that this area has historically lain outside of the main thoroughfares of commercial and cultural exchange between the subcontinent's major *Hochkulturen* have rendered this region a fringe area, where from Neolithic and Chalcolithic times the local

material cultures, as preserved in the archaeological record, were comparatively less developed.^{10–12} The combination of the more rudimentary technological level of development of the resident populations and geographical remoteness may have facilitated the gradual admixture and assimilation of incursive populations willing to adapt to the subsistence strategies practised locally, while impeding the bearers of technologically more advanced cultural assemblages.¹⁰ Previous studies have reported language shift among many populations living in this region (viz. Bathudi, Bhuiyan, Kanwar, Pando and Mushar) and referred to them as Transitional.^{13,14} Nevertheless, these studies have also indicated that the process of language shift did not always greatly alter the genetic make-up of the local populations. The picture that is beginning to emerge from various genetic studies is that resident populations practising hunting and foraging and speaking now lost tongues adopted cultural influences and adapted linguistically as well as technologically to more advanced populations from other parts of South and Southeast Asia.^{4,7,15} In a similar vein, the linguistic assimilation of the local Munda populations in adjacent areas to the Austroasiatic language family provides a stunning case of language shift correlated with an exclusively male-biased linguistic intrusion from an area with a technologically more advanced level of cultural development.¹⁶

Unlike the caste populations in India, there are very few tribes with total population sizes ranging in millions. Among all the central Indian tribes, Gond is the most populous tribe and has a well-defined clan structure.⁸ With a population size of over 12 million, they are mainly found in eastern central India (Supplementary Figure 1). The time of the existence of Gond in the subcontinent is not known with certainty.

¹Evolutionary Biology Group, Estonian Biocentre, Tartu, Estonia; ²Department of Zoology, University of Calcutta, Kolkata, India; ³Department of Genetics, Osmania University, Hyderabad, India; ⁴Narsimhbhai Patel Dental College and Hospital, Sankhal Chand Patel University, Visnagar, India; ⁵Bharat Addhyan Kendra, Faculty of Arts, Banaras Hindu University, Varanasi, India; ⁶Prakash Hospitals and Trauma Centre, Mau, India; ⁷CSIR-Centre for Cellular and Molecular Biology, Hyderabad, India; ⁸Institute of Linguistics, University of Bern, Bern, Switzerland; ⁹Department of Evolutionary Biology, Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia

*Correspondence: Dr G Chaubey, Evolutionary Biology Group, Estonian Biocentre, Riia23b, Tartu 51010, Estonia. Tel: +372 73 75050; Fax: +372 74 20194; E-mail: gyanc@ebc.ee
Received 22 June 2016; revised 1 November 2016; accepted 14 December 2016; published online 1 February 2017

However, they are mentioned in the epic *Ramayana*, and four of their kingdoms are dated to between 1300 and 1600 AD.¹⁷ By the medieval period, these kingdoms had assimilated so much religious and cultural influence from neighbouring Hindu culture that the Gond societies had become a socially more hierarchically structured tribal population.

Different groups of the extended Gond population speak Gondi, Konda, Kui, Kuvi, Pengo and Manda, all languages of the South Central branch of the Dravidian language family.^{8,17} Linguistically, the Gondi–Manda subgroup shares its most recent common ancestry with Telugu that is mainly spoken in the state of Andhra Pradesh, including Telangana.^{18,19} Ethnographical studies by Robert von Heine-Geldern^{20,21} had suggested that a subset of Dravidian populations represented by the various Gond linguistic communities as well as the local ancestral component of the Munda populations collectively represent an older layer of peopling of the Indian subcontinent. This theory was adopted by Grigson,²² who proposed that the Gonds were an originally ‘pre-Dravidian’ or what he called ‘proto-Australoid’ population that had been modified by considerable Dravidian element. Christoph von Fürer-Haimendorf^{23–26} conducted studies on the Gond and their closely allied Dravidian linguistic communities, which led him to view these peoples as remnants of an earlier primordial population that had been linguistically assimilated.

Work on the mitochondrial DNA of Gond population groups has shown that the majority of their maternal gene pool falls into South Asian specific clades with a few haplotypes belonging to the haplogroups M2, R7, M40 and M45 shared with the Austroasiatic populations.^{4,7,27,28} The Y chromosomal and autosomal studies have suggested their deeply rooted South Asian ancestry.^{29,30} However, previous genetic studies relied on either low-resolution data or studied only a single Gond group.^{4,7,27–30} Therefore, in the present study, we extracted genome-wide SNP data (>95 K), of 18 Gond samples from two recent publications.^{31,32} These 18 samples represent four distinct geographical locations, spanning three Indian states: three samples each of Gond1 and Gond3 from Madhya Pradesh and five samples each of Gond2 from Chhattisgarh and Gond4 from Uttar Pradesh (Supplementary Figure 1). We first explored the relation of the different Gond groups in respect to a wider Eurasian context and then evaluated their genomic diversity at the intra and inter-population level. Furthermore, we evaluated the population interaction and gene flow across the overlapping linguistic phyla in this region.

MATERIALS AND METHODS

Present analyses were performed on the merged data published in various genome-wide studies^{16,31–40} (Supplementary Table 1). This study was approved by the ethical committee of the CSIR-CCMB, India. The tribal and caste populations were grouped according to their linguistic affiliation. We renamed four Gond groups as Gond1, Gond2, Gond3 and Gond4 (Supplementary Figure 1). Gond1 and Gond3 are from Madhya Pradesh, Gond2 is from the Chhattisgarh state and Gond4 group from Uttar Pradesh (Supplementary Figure 1). We grouped populations that were known to have undergone language shift in recent time as Transitional.¹⁴ Plink 1.9 was used for the data curation, management and IBS (Identity-by-State) calculations.⁴¹ To remove background linkage disequilibrium (LD) that can affect both principal component analysis (PCA)⁴² and ADMIXTURE,⁴³ we thinned the data set 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 the *smartpca* programme of the EIGENSOFT package with the default settings⁴⁴ to capture genetic variability described by the first five components. In the final settings, we ran ADMIXTURE⁴³ with a random seed number generator on the LD-pruned data set 25 times from $K=2$ to $K=12$. We have used the methods described earlier^{16,31} and found $K=9$ to

be the best K . Given the result of the PC and ADMIXTURE analyses, we have removed one outlier sample from Gond1 and Gond2 groups for further population-based analysis. The outgroup f_3 statistics⁴⁴ was calculated as $f_3 = (\text{Gond1}/\text{Gond2}/\text{Gond3}/\text{Gond4}, X; \text{Yoruba})$, where X was another Indian populations. To plot the alleles sharing of Gonds and other Indian populations with Dravidian vs Austroasiatic groups, we took the Paniya population as a representative of Dravidian and the Bonda (South Munda) population as a representative of Indian Austroasiatic. The selection of these populations was based on their outlier position and highest ASI (Ancestral South Indian) ancestry. To investigate the gene flow among different Indian populations, D statistics were used by taking African Yorubas as an outlier.⁴⁴ We constructed the maximum likelihood (ML) tree of Indian populations considering four migration events using *treemix*⁴⁵ with the -k4 flag; 25 replicates were made to assure convergence. For haplotype-based analysis (*fineSTRUCTURE*),⁴⁶ samples were phased with *Beagle 3.3.2*.⁴⁷ A co-ancestry matrix was constructed using *ChromoPainter*,⁴⁶ and *fineSTRUCTURE* was used to perform an MCMC iteration using 10 M burning runtime and 100 000 MCMC samples. The number of samples and SNPs used for each of the analyses have been listed in Supplementary Table 1.

RESULTS AND DISCUSSION

To explore the variability and visualise the genetic structure of the four Gond groups, we first performed PCA. The majority of the Gond samples were shifted away from the Indo-European-Dravidian cline^{31,37} (Figure 1a). Gond groups showed a gradient of affinity with Austroasiatic (Munda) populations from Gond2 being closest and Gond1 furthest to them, whereas Gond3 and Gond4 clustering together in between (Figure 1a).

ADMIXTURE⁴³ was applied to the pruned data set to visualise the multicomponent genetic structure of Gond (Figure 1b). The best-supported³¹ clustering ($K=9$) Admixture showed k4 (dark green) as predominant component among Gond groups (Figure 1b). The k7 ‘light green’ component was trifling compared with any Indo-European or non-Gondi Dravidian populations. Consistent with PCA, one sample from Gond1 and Gond2 showed deviation from the general pattern of genetic structure among the Gond. It is striking that the Gond groups were more similar in their ancestry component composition to the North Munda group than to their linguistic neighbours (Figure 1b). Therefore, the ADMIXTURE analysis suggest evidence for overwhelming North Munda (Austroasiatic) affinity with all the four Gond groups as well as gene flow between Gond1 and Dravidian or (and) Indo-European speakers. In contrast with many central Indian indigenous populations (Bhil, Kol), the proportion of Austroasiatic specific component is significantly (two-tailed P -value < 0.0001) higher in each of the Gond groups. Such observations point out a significant difference in the admixture process between the Munda and Gond groups as compared with the admixture of Kol, Bhil,⁴⁸ Nihali and others with the Munda groups.

To have a better understanding of genome sharing of the Gonds with the extent of other Indian populations, we applied the haplotype-based analysis *fineSTRUCTURE*.⁴⁶ This programme generates a co-ancestry matrix using *ChromoPainter*⁴⁶ and compares the haplotypes of each and every individual with one another. On the basis of haplotype sharing among the studied groups, we compared the mean chunk counts donated by Eurasian populations with various Gond groups (Figure 2a). Consistent with the PCA and ADMIXTURE analysis, two of the outlier samples showed a different pattern. Hence they were excluded from any population-based comparison. As expected from PCA and ADMIXTURE analyses, all Gonds received the majority of the chunks from South Asian populations when compared with other Eurasians. Among the South Asians, Munda, the

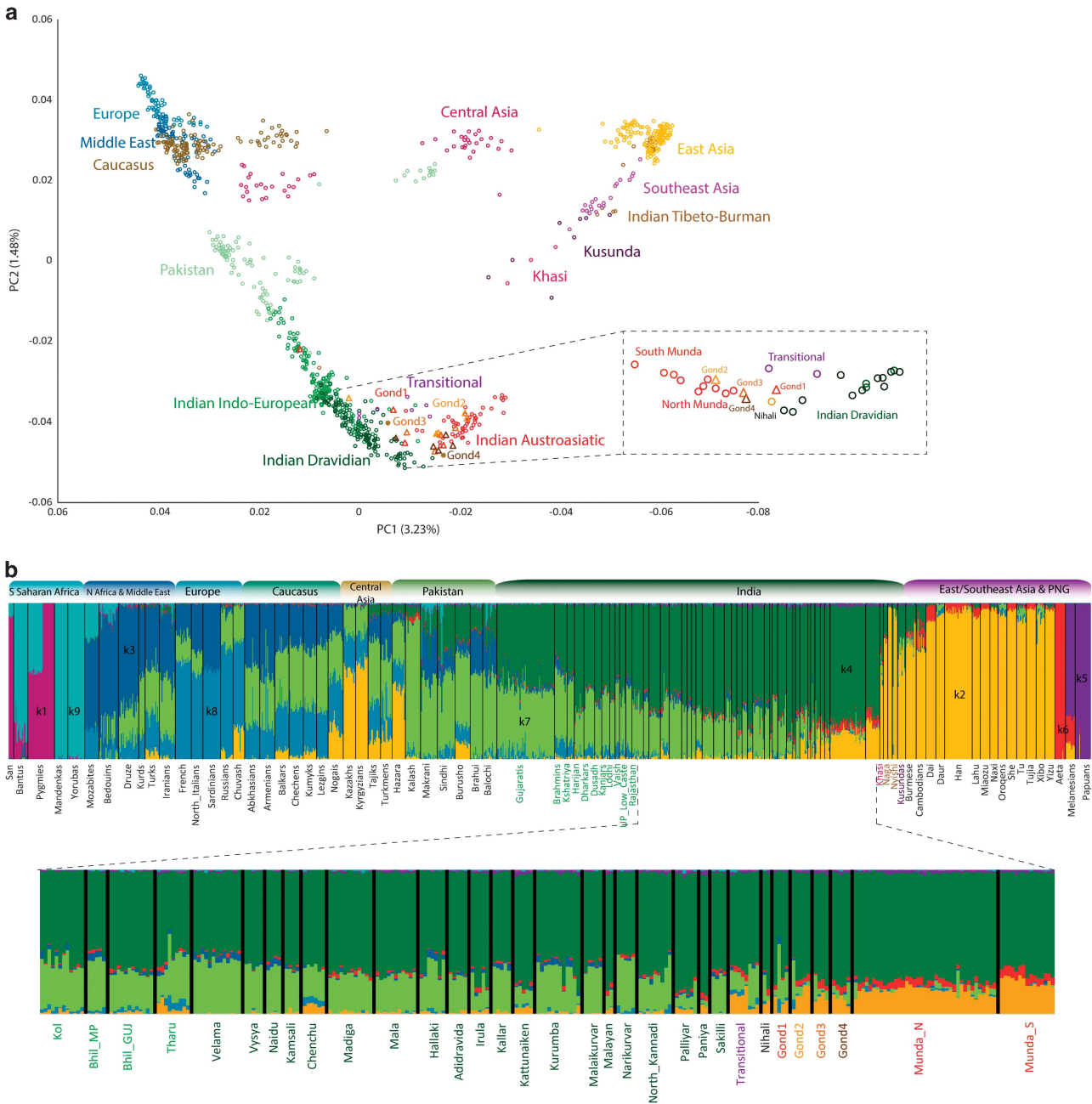


Figure 1 (a) Principal component analysis (PCA) of the combined autosomal SNP data of individuals from Eurasia. The inset picture showed the plot of mean eigen values of Gond and their genetic neighbours. (b) Plot of population-wise unsupervised ADMIXTURE analysis ($K = 9$) of world population with a zoom-in of various Indian populations including Gonds. The colour codes of the Indian populations have been given according to their linguistic affiliation shown in Figure 1a. Bhil_GUJ, Bhihs from Gujarat; Bhil_MP, Bhihs from Madhya Pradesh; Munda_N, North Munda group; Munda_S, South Munda group.

Transitional group and the Gond themselves were the major chunk contributors (Figure 2a). It is interesting to note that the Gond populations received significantly lower number of chunks from Dravidians (two-tailed P -value < 0.0001) than from the Munda speakers who are closest to them linguistically. The excess amount of allele sharing between Gond and Munda populations is also evident in IBS analysis (Supplementary Figure 2) as well as by the outgroup f_3 statistics (Figure 2b). We have also estimated the D -values.⁴⁴ When we

filtered the top 10 D -values of gene flow for each of the Gond sets, we found similar results supporting the extensive gene flow among Gond and Munda groups (Table 1).

The striking genetic affinity of Gond with Austroasiatic (Munda) populations is consistent in all our analyses (Figures 1 and 2 and Table 1). One reason for such closeness could be the process of language shift, which is common and reported among several populations of this region.^{13,14} However, it is noteworthy that the populations reported to have undergone language shift are numerically

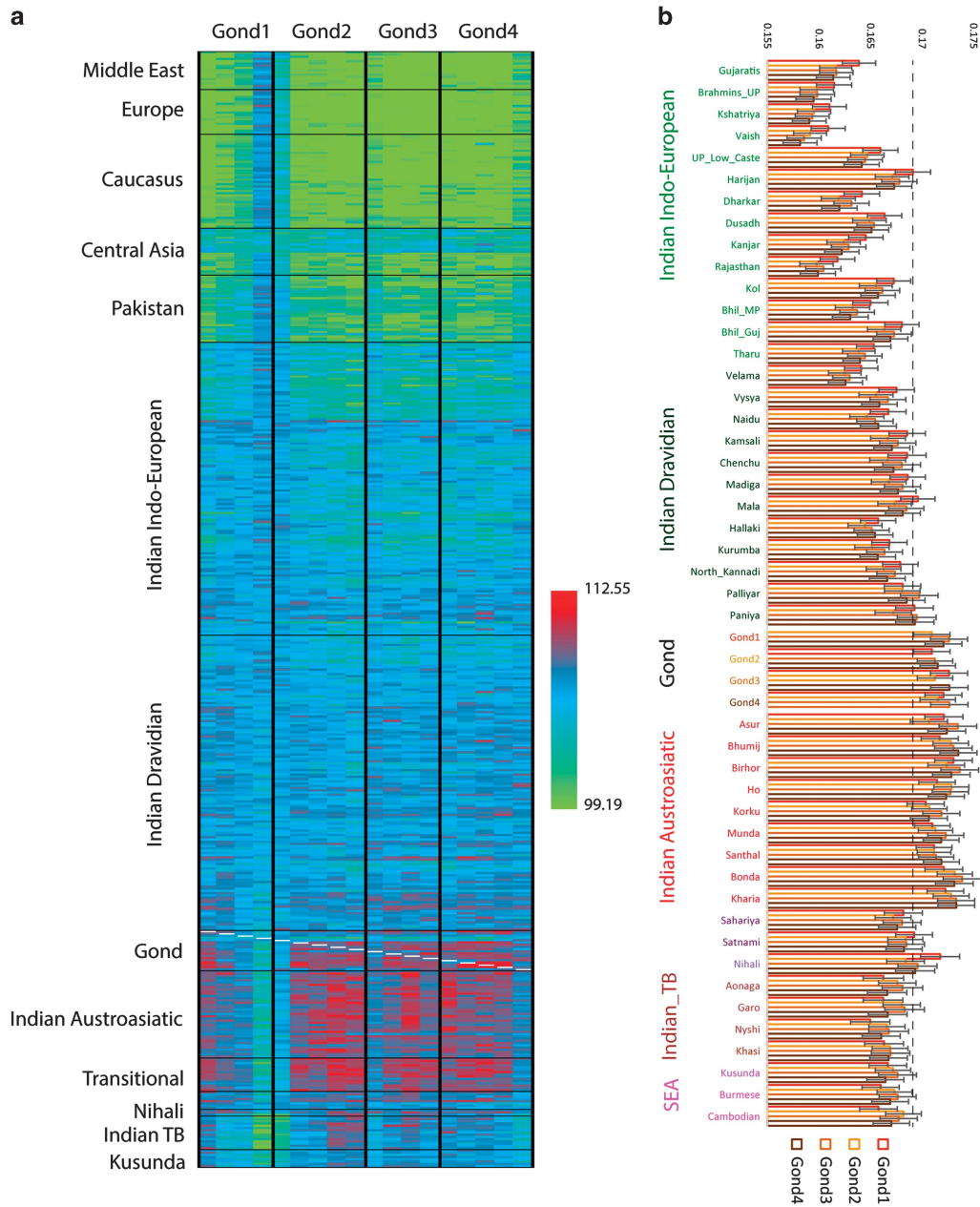


Figure 2 (a) Plot of mean sharing of chunk counts donated by Eurasian populations to the Gonds; (b) plot of shared drift f_3 analysis. The values were calculated as $f_3 = (X, Y; Yoruba)$, where X was another Indian population and Y a different Gond group. The colour codes of the populations followed their linguistic affiliation. Indian_TB, Indian Tibeto-Burman; SEA, Southeast Asian.

smaller and do not cover a vast geographical area such as that of the Gond. Moreover, the total number of Gond is equal to the number of Austroasiatic speakers of India.⁴⁹ By considering the case of language shift we modelled the scenario considering Gond originally as an Austroasiatic population, which has recently changed its language to Dravidian. In this case we should expect largely similar amount of chunks donated by an outlier distant Austroasiatic population (Bonda) to Gonds and their present Austroasiatic (both North and South Munda) neighbours. However, this was not the case in our analysis, and we observed significantly higher Bonda chunks among North and South Munda neighbours than any Gond group (Table 2). Hence, this

weakens the case for any recent language shift of Gond from Austroasiatic speakers and suggests a distinct genetic identity of the Gonds.

To compare the gene flow of Gond with Munda and Dravidian populations, two outlier populations, one from each group, Bonda (South Munda) and Paniya (Dravidian), were selected as distinct representatives of these language groups (see Materials and Methods section). The D statistics showed a significant level of gene flow between Gond and Munda groups when compared with Telugu speakers (Table 3 and Supplementary Table 2). However, the Gond showed largely similar levels of gene flow from both North and South

Table 1 The top 10 values of D statistics showing the gene flow between Gonds and other Indian populations

Gond1			Gond2			Gond3			Gond4		
Population	D-value	Z-score	Population	D-value	Z-score	Population	D-value	Z-score	Population	D-value	Z-score
Birhor	-0.0143	-6.004	Bonda	-0.0135	-7.459	Bonda	-0.0141	-7.474	Bhumij	-0.0159	-8.103
Gond3	-0.0132	-5.925	Ho	-0.0127	-6.827	Birhor	-0.0136	-5.799	Kharia-Juang	-0.0155	-8.808
Kharia-Juang	-0.0125	-6.438	Kharia-Juang	-0.0127	-7.388	Asur	-0.0131	-5.777	Bonda	-0.0151	-8.367
Bonda	-0.0122	-6.069	Bhumij	-0.0125	-6.571	Kharia-Juang	-0.0128	-6.628	Birhor	-0.0144	-6.316
Asur	-0.0121	-4.813	Birhor	-0.0116	-5.469	Bhumij	-0.0122	-6.184	Gond3	-0.0139	-6.543
Gond4	-0.012	-5.395	Gond4	-0.0097	-5.033	Ho	-0.0115	-5.796	Asur	-0.0134	-6.058
Nihali	-0.0113	-4.218	Munda	-0.0092	-4.715	Gond4	-0.0113	-5.508	Ho	-0.0133	-7.236
Bhumij	-0.0112	-5.163	Gond3	-0.0091	-4.449	Gond1	-0.0111	-4.844	Gond1	-0.0126	-5.673
Ho	-0.0107	-5.284	Santhal	-0.0089	-5.032	Munda	-0.0105	-4.955	Munda	-0.0121	-6.111
Santhal	-0.01	-5.082	Gond1	-0.0084	-3.884	Korku-Mawasi	-0.0095	-3.964	Santhal	-0.0121	-6.826

D, GondX,Yoruba; Kusunda, Y; X, Gond1-4; Y, other Indian populations.

Table 2 The result of two-tailed P-values when counting the donated chunks from Bonda (South Munda) to Gonds vs their neighbouring North and South Munda populations

Group	Gond1	Gond2	Gond3	Gond4
Korku-Mawasi (North Munda)	<0.0001	0.0001	0.0079	0.0273
Kharia-Juang (South Munda)	<0.0001	<0.0001	<0.0001	<0.0001

Table 3 The level of significance of gene flow between Gond, Telugu and Munda populations

Gp1	Gp2	Gp3	D-value	Z-score
Gond1	N_Munda	Telugu ^a	0.0096	9.7
Gond2	N_Munda	Telugu ^a	0.0143	16.017
Gond3	N_Munda	Telugu ^a	0.0138	14.593
Gond4	N_Munda	Telugu ^a	0.0141	15.849
Gond1	S_Munda	Telugu ^a	0.0108	8.548
Gond2	S_Munda	Telugu ^a	0.0167	14.147
Gond3	S_Munda	Telugu ^a	0.016	13.211
Gond4	S_Munda	Telugu ^a	0.0159	13.479
Gond1	N_Munda	S_Munda	-0.0012	-1.228
Gond2	N_Munda	S_Munda	-0.0025	-2.653
Gond3	N_Munda	S_Munda	-0.0022	-2.282
Gond4	N_Munda	S_Munda	-0.0019	-1.959
N_Munda	Gond1	S_Munda	-0.0088	-5.244
N_Munda	Gond2	S_Munda	-0.0083	-6.291
N_Munda	Gond3	S_Munda	-0.0068	-4.616
N_Munda	Gond4	S_Munda	-0.0072	-5.464

Abbreviations: N_Munda, North Munda; S_Munda, South Munda.

^aPopulations (Velama, Vysya, Naidu, Kamsali, Chenchu, Madiga, Mala) living in Andhra Pradesh and Telangana states and speaking Telugu (a branch of Dravidian language). D, Gp1,Yoruba;Gp2,Gp3.

Munda groups. Conversely, gene flow between North Munda and South Munda was significantly higher when compared with the Gond groups (Table 3).

We have plotted the shared drift values (calculated via f_3 statistics) of extant Indian populations with respect to the outlier Bonda (South Munda) vs Paniya (Dravidian) populations (Supplementary Figure 3). As both of the populations carried high amounts of ASI ancestry, we

should expect a linear trend of population assemblage. The excess of Paniya or Bonda related alleles in a particular population would place it towards that axis, away from the central line. We observed a deviation of the Gond groups from their linguistic neighbours in the direction of the Austroasiatic populations (Supplementary Figure 3). The digression of Tharu is also evident here, supporting our previous conclusion, suggesting that up to one half of their genome would be East Asian specific.⁵⁰ Interestingly, the f_3 statistics plot also revealed a clear-cut distinction of the Gond from their neighbours, which include Transitional and Nihali populations, in sharing the different proportions of Munda and Dravidian alleles (Supplementary Figure 3).

To visualise the affinity of Gond with other Indian populations and infer potential migration events, we drew a ML tree by using the method applied in *treemix*.⁴⁵ In the ML tree, all the Gond groups cluster with the western side of the Austroasiatic cluster (Supplementary Figure 4a), in consonance with a similar trend observed previously in the PCA plot (Figure 1a). With four migration events, substantial gene flow among the populations living in the central Indian region including Gonds is being revealed (Supplementary Figure 4b) supporting the notion that the central Indian region served as a selective melting pot for various populations speaking different tongues. The effect of the geography, language or ethnicity, which are major factors in other geographical regions, is minimised by the fact that eastern central India has acted as a marginal sink area. In this respect, eastern central India differs from regions such as Central Asia, where the genetic landscape was significantly shaped by the intrusion of Turkic nomads,⁵¹ with a contrasting example of the Caucasus region.³⁵

In conclusion, our extensive analysis of genome-wide genetic diversity on various Gond groups has revealed that all the Gond groups shared extensive portions of their genomes within the group as well as with North and South Munda groups. The distinctive gene flow patterns observed suggest a different population history of the Gond groups than that of their neighbouring populations. Within the overall South Asian landscape, the eastern central Indian region, with multiple language groups, is exceptional, where geography is not the major determinant correlating with genetic variation. Hence, our wide-ranging investigation on the Gond and their neighbours living in central India has shown population interaction and gene flow between various language groups transgressing the linguistic barrier by linguistic assimilation of resident populations to small but technologically more developed incursive groups.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This study was supported by Estonian Personal grants PUT-766 (GC). RV, MM and GC acknowledge financial support from the 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), and Estonian Institutional Research grants IUT24-1. KT was supported by the Council of Scientific and Industrial Research, Government of India (GENESIS: BSC0121) and (EpiHed: BSC 0118). All the analyses was performed in the High Performance Computer Center of University of Tartu, Estonia.

- 1 Lewis MP (ed): *Ethnologue: Languages of the World*. SIL International: Dallas, TX, 2009. Available at: <http://www.ethnologue.com/>
- 2 Driem G, van: *Languages of the Himalayas*. Leiden: Brill, 2001.
- 3 Indian Genome Variation Consortium: Genetic landscape of the people of India: a canvas for disease gene exploration. *J Genet* 2008; **87**: 3–20.
- 4 Chaubey G, Karmin M, Metspalu E *et al*: Phylogeography of mtDNA haplogroup R7 in the Indian peninsula. *BMC Evol Biol* 2008; **8**: 227.
- 5 Majumder PP: The human genetic history of South Asia. *Curr Biol* 2010; **20**: R184–R187.
- 6 Russell RV: *The Tribes and Castes of the Central Provinces of India*. Macmillan and Co., limited: London, UK, 1916.
- 7 Sharma G, Tamang R, Chaudhary R *et al*: Genetic affinities of the central Indian tribal populations. *PLoS One* 2012; **7**: e32546.
- 8 Singh KS: *People of India*. Oxford University Press: Oxford, 1997.
- 9 Ministry of Tribal Affairs. Government of India. Available at: <http://tribal.nic.in/> (accessed on March 2016).
- 10 Petraglia MD, Allchin B: *The Evolution and History of Human Populations*. In: Petraglia MD, Allchin B (eds). South Asia: Springer Verlag, 2007, pp 464.
- 11 Allchin B, Allchin FR (eds): *Origins of a Civilization*. Viking: New Delhi, 1997.
- 12 Jarrige JF: *South Asian Archaeology*. In: Allchin B (ed). Cambridge University Press: Cambridge, 1981, pp 21–28.
- 13 Chaubey G, Metspalu M, Karmin M *et al*: Language shift by indigenous population: a model genetic study in South Asia. *Int J Hum Genet* 2008; **8**: 41.
- 14 Kumar V, Reddy A, Babu P *et al*: Molecular genetic study on the status of transitional groups of central India: cultural diffusion or demic diffusion? *Int J Hum Genet* 2008; **8**: 31.
- 15 Thangaraj K, Sridhar V, Kivisild T *et al*: Different population histories of the Mundari- and Mon-Khmer-speaking Austro-Asiatic tribes inferred from the mtDNA 9-bp deletion/insertion polymorphism in Indian populations. *Hum Genet* 2005; **116**: 507–517.
- 16 Chaubey G, Metspalu M, Choi Y *et al*: Population genetic structure in Indian Austroasiatic speakers: the role of landscape barriers and sex-specific admixture. *Mol Biol Evol* 2011; **28**: 1013–1024.
- 17 Koreti S: Socio-cultural history of the Gond tribes of middle India. *Int J Soc Sci Human* 2016; **6**: 288.
- 18 Krishnamurti B: *The Dravidian Languages*. Cambridge University Press: Cambridge, UK, 2003.
- 19 Fuller D: Examining the farming/language dispersal hypothesis. In: Bellwood P, Renfrew C (eds): *The McDonald Institute for Archaeological Research*, Cambridge, 2003.
- 20 von Heine-Geldern R: Kopffjagd und Menschenopfer in Assam und Birma und ihre Ausstrahlungen nach Vorderindien. *Mitteilungen der Anthropologischen Gesellschaft in Wien* 1917; **XXXVII**: 1–65.
- 21 von Heine-Geldern R: Science and scientists in The Netherlands Indies. In: Pieter Honig, Verdoorn Frans (eds): *Board for The Netherlands Indies*. New York: Surinam and Curaçao of New York City, 1945, pp 129–167.
- 22 Grigson WV: *The Maria Gonds of Bastar*. London: Oxford University Press, 1938.
- 23 von Fürer-Haimendorf C: New aspects of the Dravidian problem. *Tamil Cult* 1953; **2**: 127–135.
- 24 von Fürer-Haimendorf C: *The Chenchus: Jungle Folk of the Deccan*. London: Macmillan, 1943.
- 25 von Fürer-Haimendorf C: *The Reddis of the Bison Hills: a Study in Acculturation-Aboriginal Tribes of Hyderabad*. London: Macmillan, 1945.
- 26 von Fürer-Haimendorf C: *The Raj Gonds of Adilabad: a Peasant Culture of the Deccan', The Aboriginal Tribes of Hyderabad*. London: Macmillan, 1948.
- 27 Mittal B, Tripathy V, Aruna M *et al*: Mitochondrial DNA variation and substructure among the tribal populations of Andhra Pradesh, India. *Am J Hum Biol* 2008; **20**: 683–692.
- 28 Baig MM, Khan AA, Kulkarni KM: Mitochondrial DNA diversity in tribal and caste groups of Maharashtra (India) and its implication on their genetic origins. *Ann Hum Genet* 2004; **68**: 453–460.
- 29 Trivedi R, Sahoo S, Singh A *et al*: Genetic imprints of Pleistocene origin of Indian populations: a comprehensive phylogeographic sketch of Indian Y-chromosomes. *Int J Hum Genet* 2008; **8**: 97–118.
- 30 Chaubey G, Kadian A, Bala S, Rao VR: Genetic Affinity of the Bhil, Kol and Gond mentioned in epic Ramayana. *PLoS One* 2015; **10**: e0127655.
- 31 Metspalu M, Romero IG, Yunusbayev B *et al*: Shared and unique components of human population structure and genome-wide signals of positive selection in South Asia. *Am J Hum Genet* 2011; **89**: 731–744.
- 32 Moorjani P, Thangaraj K, Patterson N *et al*: Genetic evidence for recent population mixture in India. *Am J Hum Genet* 2013; **93**: 422–438.
- 33 Li JZ, Absher DM, Tang H *et al*: Worldwide human relationships inferred from genome-wide patterns of variation. *Science* 2008; **319**: 1100–1104.
- 34 Behar DM, Yunusbayev B, Metspalu M *et al*: The genome-wide structure of the Jewish people. *Nature* 2010; **466**: 238–242.
- 35 Yunusbayev B, Metspalu M, Järve M *et al*: The Caucasus as an asymmetric semipermeable barrier to ancient human migrations. *Mol Biol Evol* 2012; **29**: 359–365.
- 36 International HapMap 3 Consortium, Altshuler DM, Gibbs RA *et al*: Integrating common and rare genetic variation in diverse human populations. *Nature* 2010; **467**: 52–58.
- 37 Reich D, Thangaraj K, Patterson N, Price AL, Singh L: Reconstructing Indian population history. *Nature* 2009; **461**: 489–494.
- 38 Rasmussen M, Guo X, Wang Y *et al*: An Aboriginal Australian genome reveals separate human dispersals into Asia. *Science* 2011; **334**: 94–98.
- 39 Raghavan M, Skoglund P, Graf KE *et al*: Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* 2014; **505**: 87–91.
- 40 Migliano AB, Romero IG, Metspalu M *et al*: Evolution of the pygmy phenotype: evidence of positive selection from genome-wide scans in African, Asian, and Melanesian pygmies. *Hum Biol* 2013; **85**: 251–284.
- 41 Chang CC, Chow CC, Tellier LC *et al*: Second-generation PLINK: rising to the challenge of larger and richer datasets. *BMC Biol* 2015; **4**: 1–16.
- 42 Patterson N, Price AL, Reich D: Population structure and eigen analysis. *PLoS Genet* 2006; **2**: e190.
- 43 Alexander DH, Novembre J, Lange K: Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 2009; **19**: 1655–1664.
- 44 Patterson N, Moorjani P, Luo Y *et al*: Ancient admixture in human history. *Genetics* 2012; **192**: 1065–1093.
- 45 Pickrell JK, Pritchard JK: Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet* 2012; **8**: e1002967.
- 46 Lawson DJ, Hellenthal G, Myers S, Falush D: Inference of population structure using dense haplotype data. *PLoS Genet* 2012; **8**: e1002453.
- 47 Browning SR, Browning BL: Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet* 2007; **81**: 1084–1097.
- 48 Chaubey G, Govindraj P, Rai N, van Driem G, Thangaraj K: The Genome-Wide Analysis of the Bhiils: the Second Largest Tribal Population of India. *Man in India* 2017; **97**: 279–290.
- 49 Census of India. Available at: <http://www.censusindia.net/>, 2011.
- 50 Chaubey G, Singh M, Crivellaro F *et al*: Unravelling the distinct strains of Tharu ancestry. *Eur J Hum Genet* 2014; **22**: 1404–1412.
- 51 Yunusbayev B, Metspalu M, Metspalu E *et al*: The genetic legacy of the expansion of Turkic-speaking nomads across Eurasia. *PLoS Genet* 2015; **11**: e1005068.

Supplementary Information accompanies this paper on European Journal of Human Genetics website (<http://www.nature.com/ejhg>)