Mystery solved: how ancient DNA reframed the case of the Austroasiatic, Mundari-speaking population of India

Barbara Verardo

Introduction

This essay revolves around the Mundari-speaking population of the South Asian sub-continent and the long-standing debate about their migratory paths. It shows how the debate evolved from earlier linguistics-based theories to parental markers research and how it was ultimately resolved by the introduction of ancient DNA studies.

About 10 million people in eastern and central India are Mundari-speakers, mostly located in forested or hilly areas of the so-called "tribal belt' of India, the Chota Nagpur plateau. Munda (or Mundari) languages are considered to be a sub-group of the wider Austroasiatic language family[1]. This family constitutes an interesting case study, as most of its 168 languages are spoken by minority groups, without official recognition, and most of it 117 million speakers practice hunting-and-gathering or small-scale, subsistence agriculture. In addition to Munda languages, it also includes the Khasi and Nicobarese languages of north-east India, the Mon or Talaing languages of Lower Myanmar, the Palaung and Wa languages of Upper Myanmar, the Khmer or Cambodian languages of Indochina, and probably also the Cham languages of Vietnam and the Sakai and other dialects of Malaysia.

This essay is divided in three sections. The first section adopts linguistics analytical lens to introduce the debate and to outline the earliest migratory hypotheses. It also touches upon the highly-contested topic of the correspondence between language and genetics: how far do language boundaries correlate with genetic ones? The second section explores the contributions by Mitochondrial DNA and Y Chromosome studies to the debate. It highlights how maternal and paternal lineages and how this appear to reflect the dichotomic view of earlier linguistic hypotheses. The third section shows how the introduction of ancient DNA analyses is reframing the entire debate. Rather than talking about migratory waves from or to India, it suggests the existence of a Proto-Austroasiatic native population already resident from Central India to mainland Southeast Asia at least 10,500 years ago and sharing a common ancestry. Neolithic migrations of East Asian farmers moving south eventually changed the genetic landscape of South East Asia, except for remote pockets in Malaysia which still share some ancestry with the Mundas of India. These results seem to confirm the autochthonous claims of the Mundari-speaking communities of India, who are also defined as adivasis ("sons of the soil").

[1] On its turn, the Austroasiatic language family is part of the so-called Austric linguistic family, together with Austronesian languages. Originally from southern China, the Austric linguistic family then spread to many areas of Indonesia and the Pacific, up to the Hawaiian Islands in the North, New Zealand in the South and Madagascar, by Indonesian immigrants.

Section 1: Linguistics-based hypotheses about Mundari-speaking migratory paths

The "origin" of the Munda-speaking groups of India constitutes a long-standing debate. At the beginning of last century, W. Schmidt[2] proposed the theory of their migration from Southern China through South-East Asia, via Myanmar and Assam. By contrast, other scholars consider them to be originating in North-West India, through the Tibetan passes that lead to Himachal Pradesh and Nepal, to then continue south and across the immense Ganga valley[3].

Both theories are based on linguistic considerations: the languages spoken in Myanmar and by the Khasi in North-East India belong to the same Austro-Asiatic family and seem to indicate the itinerary followed by the Munda in their migration towards Eastern and Central India. B.H. Hodgson, a scholar of Indo-Tibetan languages, at the end of the last century identified significant lexical affinities between the Munda languages and Classical Tibetan. J.H. Hutton, anthropologist and British Raj administrator, was in favor of this second theory in 1935[4].

Until the beginning of this century, the debate about Munda migration was based solely on linguistic affinities, which do not necessarily imply "ethnic identity". It is possible, in fact, that small tribal groups that lived for a long time in contact with the Munda, probably even in a state of dependence on them, adopted their language: this is what happened to the Juang, originally nomadic hunters and food gatherers, as well as to the Asura, — who, after being driven away by the Munda, were later accepted and their language replaced by the Munda language, the Oraons, and others. Vice-versa, there are many Mundari-speaking groups that have adopted or are in the process of adopting non-Mundari languages through time, like the Gonds, or even members of the locally-perceived as "original" Munda groups of India, in the Tamar and Ranchi area of Jharkhand, whose younger generations have replaced their Mundari language with the local "Sadri" dialect[5].

---

[2] Schmidt, Wilhelm (1906). "Die Mon–Khmer-Völker, ein Bindeglied zwischen Völkern Zentralasiens und Austronesiens ('[The Mon–Khmer Peoples, a Link between the Peoples of Central Asia and Austronesia')". Archiv für Anthropologie. 5: 59–109. Schmidt was a German missionary who proposed the Austric language family frame for the first time.

[3] Some tribal populations still live in the valley, including the Chero, who originally belonged to the Munda group, although they have now become Hinduized and have replaced their language with a neo-Aryan dialect.

[4] "(...) since there are two distinct families of Austro-Asiatic languages in India, it seems plausible the existence of two different migration routes from Central Asia, one around the western edge of the Himalayas, the other along the eastern part. The first route is connected to the Mundaric languages, the second to the Mon-Khmer ones, confined to an area of Assam, while the inhabitants of the Nicobar Islands in the Bay of Bengal use a language intermediate between Munda and Mon-Khmer". In: Hutton, J. H. 1935. Mon and Munda in India and Beyond, in Indian Census Report, Vol I – No. 2 Section 155.

[5] It would also be important to note the critical role played by ritual practices for "ethnic" self-identification, but this goes beyond the scope of this paper.

Section 2: Are DNA studies confirming the linguistics-based hypotheses?

Recent DNA-based technological innovations have reignited this debate and the search for Munda migratory routes. Overall, uniparental markers have contributed significantly to the debate but have also generated apparently contradictory results. While mtDNA studies seem to support the India to East Asia migratory route, Y-chromosome studies support the opposite East Asia to India route (Zhang 2015). In this sense, they still seem to reflect the linguistic-based dichotomic debate about the "origins" of the Munda-speaking communities, as the paragraphs below attempt to show in greater detail.

Overall, mtDNA studies trace an expansion of Austroasiatic speakers from India eastwards towards South East Asia. Mitochondrial DNA (mtDNA) allows the tracing of maternal lineages back in time, as mitochondrias are inherited only from the mother and there is usually no change in mtDNA from parent to offspring. Some such studies refute Southern China as the primary origin of the groups. For instance, after sampling individuals from 15 Mundari groups and 7 so-called transitional groups (former Austroasiatic speakers who now speak a different language) from Chota Nagpur and surrounding areas, Kumar et al. (2006) compared the results with existing data from South East Asia and Africa[6]. This study provided evidence for distinct genetic origins of the wider Austroasiatic family and multiple African, non-Asian sources of migration of the Mundari population to India around 50.000 – 70.000 BP (Kumar 2006: 466). Other mDNA studies only link these migrations to the expansion of rice cultivation from Southern China but during Neolithic times and not earlier. What most of MtDNA studies seem to agree upon is the view that "the exclusive South-Asian maternal ancestry pinpointed their origin in South Asia" (Singh, P.P. 2020: 529)[7] - thus possibly confirming the second linguistic-based theory mentioned above of a north-west Indian origin of the language.

By contrast, Y-Chromosome studies have identified Southern China or Southeast Asia as sources of Austroasiatic-speaking populations, who later migrated into India. While Kumar (2006: 467)'s subsequent analysis of Y-haplogroups confirmed his earlier mtDNA-based hypothesis of Mundari-speaking groups been the first to arrive in India[8], Singh PP (2021) concluded that "the overwhelming East/Southeast-Asian-specific paternal ancestry associated with haplogroup O2a-M95 suggested a dispersal from East to West" (Singh, P.P. 2021: 529). This view also coincides with earlier studies (see Cavalli-Sforza et al 1994), as well as more

[6] Kumar V, Langsiteh BT, Biswas S, Babu JP, Rao TN, Thangaraj K, Reddy AG, Singh L, Reddy BM. Asian and non-Asian origins of Mon-Khmer- and Mundari-speaking Austro-Asiatic populations of India. Am J Hum Biol. 2006 Jul-Aug;18(4):461-9. doi: 10.1002/ajhb.20512. PMID: 16788903.

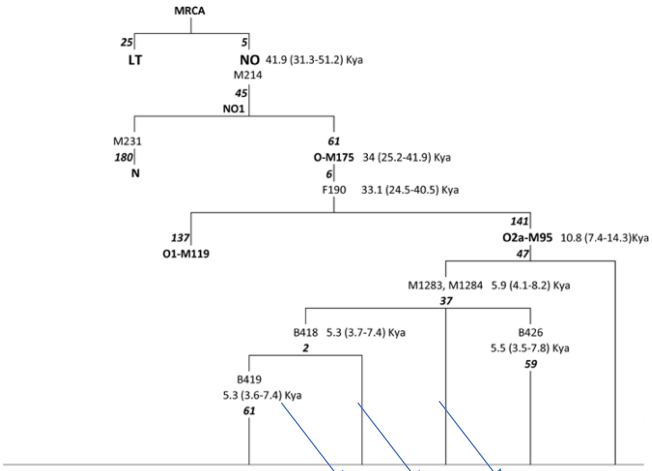
[7] Singh PP, Vishwakarma S, Sultana GNN, Pilvar A, Karmin M, Rootsi S, Villems R, Metspalu M, Behar DM, Kivisild T, van Driem G, Chaubey G. Dissecting the paternal founders of Mundari (Austroasiatic) speakers associated with the language dispersal in South Asia. Eur J Hum Genet. 2021 Mar;29(3):528-532. doi: 10.1038/s41431-020-00745-1. Epub 2020 Oct 21. PMID: 33087879; PMCID: PMC7940493.

[8] Though this study seems to be still unpublished.

recent ones, which identified the Y-Chromosomal haplogroup O2a1-M95 as the primary paternal lineage in Austro-Asiatic populations and confirmed either a Southern China or a South East Asia origin of Austroasiatic speakers, that only later migrated into India.

More recent studies have refined the timeline of these migrations. The increasing resolution of the Y-chromosomal tree and Y-STR (short tendem repeat)-based coalescent calculations allowed Singh, P.P et al. to re-analyse the haplogroup O2a-M959 in 2021. The new results showed that the haplogroup originated approximately 12.000 years ago in East/Southeast Asia and reached India during the last 5000 years. These results would seem to contradicts earlier views which identified the timing of arrival of the Austroasiatic population in India to more than 10,000 years ago[10]; while confirming Chaubey's more recent recent Y chromosome analyses that point to 4300 (+-200) years ago[11].

The same studies also confirmed earlier suggestions of multiple, more fragmented waves of migrations. While Chaubey (2007) identified a single male ancestor, Singh (2021) identified three "founders" as carrying haplogroup O2a-M95 in South Asia (Figure 1). This would seem to suggest that "the dispersal of Austroasiatic speakers to South Asia was not associated with the migration of a single clan or a drifted population" (Singh 2021: 528). In this sense, these results confirm Kumar's suggestion of multiple genetic origins and fragmented waves of migration. Critically, recent ground-breaking analyses of Ancient DNA have also confirmed this view.



POPULATION	N	M95	B419	B418	M1284	B426	M95*
Asur	88	29	-	15 (0.17)	0 (0.0)	7 (0.07)	7 (0.07)
Birhar	62	24	-	6 (0.09)	4 (0.06)	4 (0.06)	10 (0.16)
Ho	45	9	-	4 (0.08)	2 (0.04)	1 (0.02)	2 (0.04)
Mahali	32	5	-	1 (0.031)	0 (0.0)	0 (0.0)	4 (0.12)
Mawasi	39	22	-	15 (0.38)	3 (0.07)	2 (0.05)	2 (0.05)
Santhal	20	5	-	4 (0.2)	0 (0.0)	1 (0.05)	0 (0.0)
Baiga	65	23	-	2 (0.03)	3 (0.04)	3 (0.04)	15 (0.23)
Total North Munda	**351**	**117**	-	**47 (0.13)**	**12 (0.03)**	**18 (0.05)**	**40 (0.11)**
Bonda	42	23	-	4 (0.09)	0 (0.0)	2 (0.04)	17 (0.40)
Gadaba	27	6	-	1 (0.03)	0 (0.0)	1 (0.03)	4 (0.14)
Juang	54	16	-	6 (0.11)	3 (0.05)	3 (0.05)	4 (0.07)
Kharia	37	3	-	0 (0.0)	0 (0.0)	1 (0.02)	2 (0.05)
Total South Munda	**160**	**48**	-	**11 (0.06)**	**3 (0.01)**	**7 (0.04)**	**27 (0.17)**
Khasi Maghalaya	21	8	-	3 (0.14)	1 (0.04)	1 (0.04)	3 (0.14)
Bangladesh	680	69	16 (0.02)	1 (0.001)	11 (0.01)	40 (0.05)	1 (0.001)
Manipur	225	153	3 (0.01)	0 (0.00)	8 (0.03)	141 (0.62)	1 (0.004)

Figure 1 - The three Munda "founders". Source: Singh. et al. 2021. "The number of variants of a branch is shown in bold italics. For each branch, the total number of individuals of a population as well as the frequencies are shown. The M95 is M95x(B418, B419, B426, M1284)."*

[9] Singh, P.P., Vishwakarma, S., Sultana, G.N.N. et al. Dissecting the paternal founders of Mundari (Austroasiatic) speakers associated with the language dispersal in South Asia. Eur J Hum Genet 29, 528–532 (2021). Op. cit.

[10] Zhang, X. *et al.* Y-chromosome diversity suggests southern origin and Palaeolithic backwave migration of Austro-Asiatic speakers from eastern Asia to the Indian subcontinent. *Sci. Rep.* 5 (2015); Arunkumar, G. *et al.* A late Neolithic expansion of Y chromosomal haplogroup 02a1-M95 from east to west: Late Neolithic expansion of O2a1-M95. *J. Syst. Evol.* 53, 546-560 (2015); and Riccio, M. *et al.* The Austroasiatic Munda population from India and its enigmatic origin: a HLA diversity study. *Hum. Biol.* 83, 405-435 (2011).

[11] Chaubey G, Metspalu M, Kivisild T, Villems R. Peopling of South Asia: investigating the caste-tribe continuum in India. Bioessays. 2007 Jan;29(1):91-100. doi: 10.1002/bies.20525. PMID: 17187379.

Section 3 –How ancient DNA is reframing the debate

The analysis of ancient DNA is providing new answers to the old debate about Austroasiatic migrations. Critically, it is also reframing the entire debate, showing how the question itself was misplaced. Ancient DNA analyses show that Austroasiatic languages emerged out of a proto-language spoken by native Indian and South East Asians who shared a common ancestry earlier than 10,500 years ago. This ancestry then shifted through inter-breeding with Indo-European and Dravidian populations in India, and with Southern China/East Asian groups in South East Asia. Interestingly, the geographical and cultural isolation of native groups both in India and Malaysia may explain why their proto-language persisted through time – eventually giving rise to the Mundari and Mon-Khmer branches of the Austroasiatic language. This is explained more in detail below.

First, new data show the existence of a Proto-Austroasiatic native population extending from Central India to mainland Southeast Asia at least 10,500 years ago. Results from Tagore (2021)'s extensive analysis shows a negligible presence of East Asian genetic components not only in extant Indian genomes but also in the two oldest among the 46 ancient South East Asian genomes included in his study: "they must have lived in times when the migration of East Asian farmers had only just began and therefore had minimum East Asian ancestry" (Tagore 2021: 13)[12]. In other words, rather that talking about migrations of Austroasiatic speakers to or from India and South East Asia, present-day Austroasiatic languages emerged from speakers of a proto language that extended from Central India to Malaysia. A common ancestor was identified between extant Indian and Malaysian samples dating back to about 10,500 years ago.

Second, East Asian elements start to appear in Indian genomes only with the Neolithic expansion of agriculture from southern China/East Asia. As East Asian farmers migrated south towards South East Asia, the local Austroasiatic hunters-and-gatherers started to decline as a consequence while a new East Asian ancestry was introduced through inter-breeding, reaching as far as Malaysia. With regard to India, the Mundas were probably settled in the far eastern part of the country at the time of the ANI (Ancestral North Indian) -ASI (Ancestral South Indian) mixture (1900 to 4200 BP[13]), and thus somehow exposed to the same influences – though marginally[14].

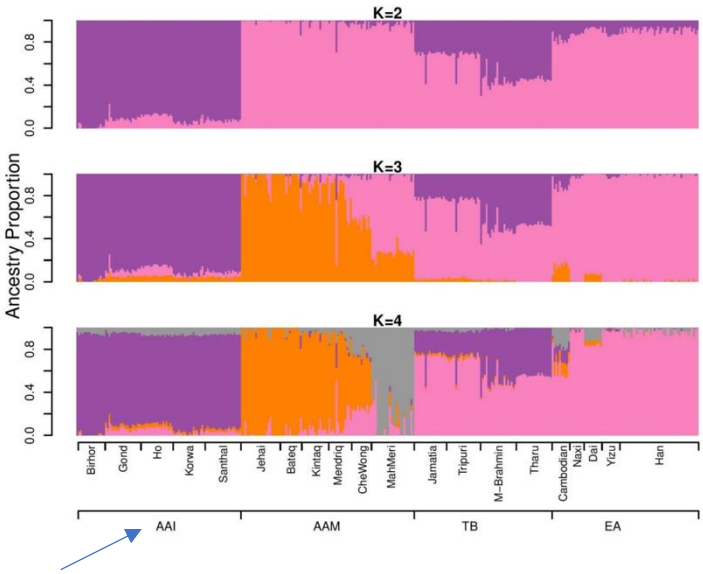
[12] Tagore, D., Aghakhanian, F., Naidu, R. et al. Insights into the demographic history of Asia from common ancestry and admixture in the genomic landscape of present-day Austroasiatic speakers. BMC Biol 19, 61 (2021). https://doi.org/10.1186/s12915-021-00981-x.

[13] Moorjani P, Thangaraj K, Patterson N, Lipson M, Loh PR, Govindaraj P, Berger B, Reich D, Singh L. Genetic evidence for recent population mixture in India. Am J Hum Genet. 2013 Sep 5;93(3):422-38. doi: 10.1016/j.ajhg.2013.07.006. Epub 2013 Aug 8. PMID: 23932107; PMCID: PMC3769933.

[14] Chabuey's analysis showed the absence of one ancestral component among the Munda samples, which was however present in Indo-European and Dravidian ones – and which he concluded must have spread at the time of the ANI -ASI mixture and was avoided by the Mundas being located further east.

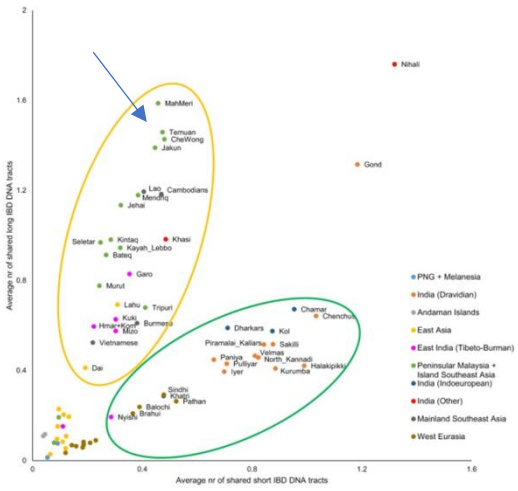
Third, East Asian ancestry decreased again in post-Neolithic times. Present-day Indian ancestry shows very little East Asian elements, possibly a residual of ancient admixture or more recent admixtures with Tibeto-Burman populations living in North-East Asia (Figure 2).

Figure 2 ADMIXTURE analysis of Indian, Malaysian. Tibeto-Burman and East Asian genomes. Tagore, D. et al. 2021.



These results seem to contradict earlier genetic analyses, according to which one quarter of extant Munda genetic ancestry was shared with Malaysia's Bateq, Jehai, Kintaq and Mendriq populations[15] (Figure 3).

Figure 3 Plotted average counts of IBD segments shared with Munda speakers. Source: Tatte, K. et al. 2019.



While this latter study was quite comprehensive, analysing 102 genomes of Munda-speaking individuals, together with 978 genomes from 72 populations from the South-Asian sub-continent, Southeast Asia and East Asia, tellingly it did not include ancient DNA.

[15] Tätte K, Pagani L, Pathak AK, Kõks S, Ho Duy B, Ho XD, Sultana GNN, Sharif MI, Asaduzzaman M, Behar DM, Hadid Y, Villems R, Chaubey G, Kivisild T, Metspalu M. The genetic legacy of continental scale admixture in Indian Austroasiatic speakers. Sci Rep. 2019 Mar 7;9(1):3818.

Conclusions

Technological progress has played a critical role in the investigation and reframing of old debates regarding migrations of Austroasiatic speakers, and more specifically of Mundari-speaking sub-groups. Earlier investigations based on linguistics have been complemented by analyses based first on uniparental markers and then on genome-wide genotype data. More recently, ancient DNA analyses have entered the picture, shedding new exciting light on the topic and reframing the old debate with new questions and new approaches.

Results from ancient DNA studies seem to support "sons of the soil" claims of the Mundari people, who are also widely known and referred to as Adivasis (sons of the soil). Their oral traditions and their spiritual world trace their origins back to the north of India. Their burial megaliths, which are unique of tribal communities in India and still practiced by Munda-speaking groups, are used as evidence of clanic land rights and hold an enormous yet unexplored potential to trace back migratory routes. Similarly, names of clan spirits include a chronological sequence of geographical/sacred locations occupied by previous generations of clan members, tracing their migratory paths through the forest[16]. Relative sizes of graveyards could also give an idea of timelines while excavations could provide archaeological and more DNA evidence. To date, however, these cultural lens seem not to have been used to support or triangulate DNA-based data.

The adoption of cultural lens and awareness of the local socio-cultural contexts may also help to correctly identify genetic samples. For instance, Tagore (2021)'s ancient DNA study failed to include any samples from the Munda group, the second most numerous among Mundari-speaking groups (see Supplementary Table 1a: Classification of populations used in his study). Along the same lines, the categorization of genetic samples as belonging to the various Munda-speaking groups (Mundas, Hos, Santhals, Birhors, and others) seems to assume rigid social/"ethnic" group boundaries between these them - while in reality, such boundaries are flexible and can be "crossed" though specific rituals not only at the individual level but also at community level. The author of this essay conducted ethnographic fieldwork in a forested Mundari-speaking community of Northern India and witnessed first-hand the process of transformation of a "Ho" group into a "Munda" one. Such transformations, as well as similar phenomena like intermarriages among members of different Mundari-speaking groups, are becoming increasingly common in semi-urban and urban areas and among the educated youth. This cultural awareness seems especially relevant when genetic studies revolve around marginal native groups like many Austroasiatic speaking groups, as their languages and "ethnic identities" are rarely formalized and constantly renegotiated.

[16] Verardo, Barbara (2003) *Rebels and devotees of Jharkhand: Social, religious and political transformations among the adivasis of northern India.* PhD thesis, London School of Economics and Political Science.

REFERENCES

Arunkumar, G., et al., 2015. A late Neolithic expansion of Y chromosomal haplogroup O2a1-M95 from east to west: Late Neolithic expansion of O2a1-M95. *Journal of Systematics and Evolution*, 53, pp.546-560.

Cavalli-Sforza, L.L., et al., 1994. *The history and geography of human genes*. Italy: Princeton University Press.

Chaubey, G., Metspalu, M., Kivisild, T. & Villems, R., 2007. Peopling of South Asia: Investigating the caste-tribe continuum in India. *BioEssays*, 29(1), pp.91-100.

Hutton, J.H., 1935. Mon and Munda in India and Beyond. *Indian Census Report*, Vol. 1, No. 2, Section 155.

Kumar V, Langsiteh, B.T., Biswas, S., Babu, J.P., Rao, T.N., Thangaraj, K., Reddy, A.G., Singh, L., Reddy, B.M. 2006. Asian and non-Asian origins of Mon-Khmer- and Mundari-speaking Austro-Asiatic populations of India. Am J Hum Biol. Jul-Aug;18(4):461-9.

Moorjani, P., Thangaraj, K., Patterson, N., et al., 2013. Genetic evidence for recent population mixture in India. *American Journal of Human Genetics*, 93(3), pp.422-438. doi:

Narasimhan, V.M., Patterson, N.J., Moorjani, P., et al. 2018. The Genomic Formation of South and Central Asia. bioRxiv.

Riccio, M., et al., 2011. The Austroasiatic Munda population from India and its enigmatic origin: A HLA diversity study. *Human Biology*, 83, pp.405-435.

Schmidt, W., 1906. Die Mon–Khmer-Völker, ein Bindeglied zwischen Völkern Zentralasiens und Austronesiens ('The Mon–Khmer Peoples, a Link between the Peoples of Central Asia and Austronesia'). *Archiv für Anthropologie*, 5, pp.59–109.

Singh, P.P., Vishwakarma, S., Sultana, G.N.N., et al., 2021. Dissecting the paternal founders of Mundari (Austroasiatic) speakers associated with the language dispersal in South Asia. *European Journal of Human Genetics*, 29(3), pp.528-532.

Tagore, D., Aghakhanian, F., Naidu, R., et al., 2021. Insights into the demographic history of Asia from common ancestry and admixture in the genomic landscape of present-day Austroasiatic speakers. *BMC Biology*, 19, p.61.

Tätte, K., Pagani, L., Pathak, A.K., et al., 2019. The genetic legacy of continental scale admixture in Indian Austroasiatic speakers. *Scientific Reports*, 9(1), p.3818.

Vagheesh M. Narasimhan et al., 2019. The formation of human populations in South and Central Asia. *Science* 365, pp. 74-87.

Verardo, B., 2003. Rebels and devotees of Jharkhand: Social, religious and political transformations among the Adivasis of northern India. PhD thesis. London School of Economics and Political Science. Available at: https://etheses.lse.ac.uk/2521/.

Zhang, X., et al., 2015. Y-chromosome diversity suggests southern origin and Palaeolithic backwave migration of Austroasiatic speakers from eastern Asia to the Indian subcontinent. *Scientific Reports*, 5.