

## Fallacy of Projected Gene - Language Correlation in South Asia

Sreenathan M\*

*\*Thunchathu Ezhuthachan Malayalam University, Vakkad, Tirur, Kerala, India*

*Received September 24<sup>th</sup>, 2020; Revised October 6<sup>th</sup>, 2020; Accepted October 8<sup>th</sup>, 2020*

### ABSTRACT

This paper addresses the theme of Gene-Language correlation popularly discoursed in South Asian context. Generally, the language orientation in genomic discourses of south Asia is popularly based on the well-established language families. With respect to the early dating of these language families, these families are not evidentially supporting the prehistoric linguistic background of South Asia. When the reported language affiliation against mtDNA evidences in terms of time depth is cross checked, there is significant mismatch in the current genomic discourses on language-mtDNA correspondences. To scientifically prove this mismatch haplogroup structure is crosschecked against the language attestation. It is found that the results are not only proving the obvious mismatch and also suggests there is huge level of Palaeolithic language lineage replacements. Genomic discourses are unable to express this gap instead they negotiate with the current status instead the genomic evidences meant to unravel the prehistory of peopling in South Asia.

**Keywords:** Gene-Language correlation, Prehistoric linguistic background, Haplogroup structure, Palaeolithic language lineage replacements, Genomic discourses

### INTRODUCTION

This study contradicts the popular gene-language attestation and enhances the premise that the Genealogical affiliation of languages does not presuppose genomic affiliation by providing evidences. The theme of this paper is a re-evaluation of linguistic and genomic corroboration popularly described in the context of south Asia and in particular Indian subcontinent. The genomic understanding of the region confirms one of the first regions to have been peopled by modern humans. Reconstructing Indian population history [1] suggests two major ancestral stream's influx in the Indian populace. To sum up, both ancient and recent genetic signatures are found in India. The present Indian population primordially comprises of genetic elements exclusive to ancient North Indians and Ancient South Indians and their admixture. In terms of linguistic manifestation, South Asia currently exhibits no Parallel strata to the suggested archaic genomic substratum. Suggesting a possibility of reconstructing language strata in this direction would be an interesting proposition especially in the context of gene-language coevolution hypothesis. It is argued that the genealogical affiliation of languages does not presuppose genomic affiliation between the member speech communities. Yet, the available discourses on genomic understanding of Indian population quite often exhibit a tendency to explain the mtDNA phylotree in corroboration with the associated current linguistic identities. Is it appropriate to assume the ancestral

genetic lineages of speech communities in correlation with their current linguistic status?

To understand the linguistic kinships there are various classification modes like genetic, typological and aerial are in vogue. But both typological traits and aerial features are generally excluded from establishing genetic grouping of languages. The genetic classification of languages is strictly based on linguistic genealogy. The family grouping of genetically related languages based on sound correspondences and lexical and grammatical similarities popularly referred to as genealogical classification. As far as India is concerned, there are six language families namely Austro-Asiatic (Munda and MonKhmer), Dravidian, Indo-Aryan (IA) and Tibeto-Burman, Andamanese and Ongan-Onge Jarawa (The Austronesian lineage). India has high language diversity (the number of languages) but it is poor in phylogenetic diversity (the number of language families). The less phylogenetic diversity reflects the Neolithic stage of India as the main language families have Neolithic lineages.

**Corresponding author:** Sreenathan M, Thunchathu Ezhuthachan Malayalam University, Vakkad, Tirur, Government of Kerala, 676502, India, Tel: +91-8129560028; E-mail: msreenathan@gmail.com

**Citation:** Sreenathan M. (2020) Fallacy of Projected Gene - Language Correlation in South Asia. *J Genet Cell Biol*, 4(1): 233-241.

**Copyright:** ©2020 Sreenathan M. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

The known Paleolithic linguistic lineages of India are Andamanese and Ongan language families. Besides, said language families, the presence of isolated linguistic remnant like the Nahali (Central India) are also noted. The Andamanese groups connect India with the prehistoric substratum. The entire linguistic scenario of India underlines the fact that the Andamanese groups were not colonized by the Neolithic expansion due to isolation and on the other hand, many of the mainland representatives of the old substrata got influenced heavily.

The present reconstructed protolanguages of Indo Aryan and Dravidian family of languages or any other family does not confirm the antiquity of ancient genomic substratum of peopling.

### Pre-historic linguistic groups of South Asia

There are Old Indo Aryan written sources and the history of investigations focused the arguments on the presence of non-Aryan elements in it. The Austro Asiatic and Dravidian connections are suggestively argued. Deep rooting of Dravidian connectivity is argued heavily by [2-13]. Also, there are arguments related to the Austro-Asiatic connection. Hock [10], Kuiper [11], Witzel and Lévi [13-15] have noted that the earliest foreign elements found in the Rigveda (RV) are of Austro-Asiatic in appearance. Dravidian origin of the dental-retroflex distinction is considered as an internal innovation [10], as post Rigvedic phenomenon [16], and diffusion of it from Pre-Mundaic language of Panjab [13].

The pre historic linguistic map was reconstructed by Southworth [17] showing “(1) languages belonging to known families: (a) Munda and related Austro-Asiatic languages, (b) Dravidian languages, and (c) Tibeto-Burman languages; (2) isolated languages of no known language families; (3) languages whose prior existence is inferred from traces left in existing languages. These are languages whose existence is inferred from traces (vocabulary and/or grammatical constructions) found in existing languages. Their prehistoric status is comparable to that of the isolated languages. (a) The ‘Indus’ language(s), which served as the source of numerous words, mainly names of plants, found in old Indo Aryan and early Dravidian; (b) ‘Meluhhan’ the source of some 40 ‘Indian’ words found in ancient Mesopotamian sources, referring to trade goods originating in the Indus Valley. This language may have been located in the hilly areas of Baluchistan, near to the Indus Valley [13]; (c) An unknown substrate language or group of languages, in the area of Bhili, Ahirani, Dangi, and Katkari (the region where Gujarat, Rajasthan, Madhya Pradesh, and Maharashtra adjoin each other) which has left its mark on the lexicon, and perhaps the grammatical structure, of these languages. I have provisionally dubbed this substrate ‘proto-Bhili’; (d) ‘Proto-Nilgiri,’ a pre-Dravidian substrate in the Nilgiris in South India [13]; (e) The Vedda substrate in Sri Lanka, inferred on the basis of loanwords and collocations in Sinhala; (f) Witzel also refers to ‘Central Asian substrate(s)’ – not shown here –

which may be the source of a number of words in early OIA as well as various Iranian languages [13] (g) Masica [18] posited a ‘Language X’ to account for agricultural words of unknown origin in Hindi-Urdu. Though Masica started with Hindi vocabulary in tracing the history of these words, the large majority of them are of general occurrence in Indo-Aryan. On the assumption that the ancestors of all Indo-Aryan languages passed through the Indus Valley during the OIA period, a source in that area seems most probable, and in the absence of evidence pointing to some other specific location, it seems reasonable to posit the ‘Indus’ languages as the source of this material. Of course, this language (group) may not have been confined to the Indus Valley region. Witzel [13] believes that the oldest stratum of these loanwords in the Rigveda is derived from Munda or related (and otherwise unknown) Austro-Asiatic (AA) languages which he designates as ‘Para-Munda’. Dravidian borrowings, according to Witzel [13], do not appear until the middle Rig Vedic period. Research in this front has not revealed any possibility of tracing Palaeolithic linguistic substratum rather they were only trying to dig out the past of Neolithic linguistic lineages.

The Indian specific mtDNA lineages M, R and U are confirmed as Pleistocene lineages bearing coalescent time at around 50,000 ybp [19-24]. The high frequencies of M observed across population suggest a deep founder effect of M in the evolution of Indian population. However, R and U too have their contribution in fixing the ancestral antiquity of Indians. The Y lineages of India are also detected as of pre-Holocene origin. Various studies on South Asian populations do not show any caste or language-specific distribution of major ancient lineages and also confirmed continent-specific distributions of certain mtDNA lineage groups [22,23,25] show that the genetic variation in India is characterized by a high Y-chromosome diversity, which is reflected by a greater correspondence with linguistic groups of India. Y-haplogroup structure suggests a common Pleistocene origin of Indian population as well as influx of subsequent migration.

In order to check how far linguistically unrelated populations are genetically united; we have attempted a haplogroup distribution analysis of 24 linguistically unrelated tribal groups which give us indications of ancestral unity among linguistically diverse groups. In this study we are looking at this correlation more diachronically in order to understand the archaic ethno history of linguistically unrelated population groups. It is envisaged that linguistically related populations are genetically close, thereby; linguistically unrelated populations are bound to show genetic differentiation. The 24 unrelated population groups surveyed\* in **Table 1**, substantiate such assumption is the subject of enquiry.

The above studied groups are clearly showing signs of divergence primarily in terms of presence of M2 and absence of M2. Out of 24 studied groups, 15 groups are carrying the Paleolithic baggage of M2. Remaining Nine groups are not

**Table 1.** Unrelated population groups surveyed.

Communities	Present language status	Haplogroups
<b>korku</b>	Austro Asiatic North Munda	M2 M4 M5 M6 M18 M30 M33 M38 M42
<b>Munda</b>	Austro Asiatic North eastern kherwarian	M2 M4 M5 M6 M18 M25 M31 M33 M35 M39 M40 M42
<b>Bettakuruba</b>	SouthDravidian	M2M4 M5 M35
<b>Madia</b>	Cendral Dravidian	M2 M3 M4 M6 M25 M35 M39
<b>Hillkolam</b>	CentralDravidian	M2 M3 M4 M5 M6 M18
<b>Malpaharia</b>	North Dravidian/ Indo Aryan	M2 M3 M4 M5 M6 M18 M33 M39 M41
<b>Jenukuruba</b>	South Dravidian	M3 M8_C_Z M9 M25 M36
<b>Kamar</b>	Dravidian unclassified/IndoAryan	M2 M3 M4 M5 M6 M12_G M36 M39 M40
<b>DungriBhil</b>	IndoAryan	M2 M3 M4 M5 M6 D M18 M30 M33
<b>Kathodi</b>	IndoAryan	M2 M3 M4 M5 M30 M35 M39
<b>Matakur</b>	IndoAryan	M2 M4 M30
<b>Katakur</b>	IndoAryan	M2 M4 M5 M12_G M39
<b>Sonalkachari</b>	IndoAryan	M2 M3 M4 M6 D M8_C_Z M9 M18 M49
<b>Paudibhuiya</b>	IndoAryan	M2 M3 M4 M5 M6 M12_G M18 M25 M31 M39 M40
<b>Katkari</b>	IndoAryan	M2 M3 M4 M12_G M39
<b>Andh</b>	Unclassified (IndoAryan)	M2 M3 M4 M5 M6 M30 M35 M39
<b>Nihal</b>	Language isolate	M2 M4 M5 M6 M12_G M18 M25 M30 M33 M35 M37
<b>Sherdukpan</b>	TibetoBurman (Bhotia group)	M3 M4 D
<b>Toto</b>	TB(Himalyan group)	M4 M6 D M18 M33 M35
<b>Gallong</b>	TB(NEFAGroup)	M3 M4 M6 D M8_C_Z M9 M12_G
<b>DirangMonpa</b>	TB(Bhotia group)	M3 M4 M6 D M8_C_Z M9 M12_G M28 M49
<b>Wancho</b>	TB(Naga group)	M4 M6 D M8_C_Z M9
<b>Lachungpa</b>	TB(Naga group)	M6 D M8_C_Z
<b>Lepcha</b>	<b>TB Central group</b>	M6 D M8_C_Z M9 M12_G M18 M33

*\*Data from Anthropological Survey of India*

haplogroups or these groups are representing a different stream. The occurrence of M3 is restricted in 15 groups which again show the pattern of M2 M3 (presently rated Indo Aryan groups like Dungri Bhil, Kathodi, Katkari, Andh, Kamar,

Paudibhuiya, Soanlkachari and among Dravidian designated groups Madia, Hillkolam and Malpaharia share this pattern) and M3(Tibeto Burman groups like Sherdukpan, Gallong, Dirang Monpa and Dravidian accounted Jenukuruba carries

this pattern). The influx of M4, which is the most widely distributed marker, is also revealing certain definite patterns. Twenty-one communities out of 24 communities share M4. Of which, it attests the pattern of sharing with M2 M3 carriers or sharing with M2 and M3 carriers separately or M4 alone. M2 M3 M4 - DungriBhil, Kathodi, Katkari, Andh, Kamar, sonalkachari and PaudiBhuiya (Presently Indo Aryan), Hillkolam, Malpaharia, and Madia (presently Dravidian)

M2M4 - Matakur, Katakur, Nihal of present Indo Aryan and Bettakuruba of Dravidian, and Korku and Munda of Austro Asaitic share this pattern

M3M4 – Tibeto Burman group like sherdukpan, Gallong, Diran Monpa attest this pattern.

M4 - Toto, Wancho (TibetoBurman groups)

The above pattern clearly reflects the ancestral haplogroup structure up to the influx of M4. This suggests in accordance with the age of haplogroups, language designation is not at all suitable. It reflects an erstwhile tongue common to the then existed ancestors who were predominated with these haplogroups. Haplogroup distributional divergence clearly indicates the trends of population divergence. Forthcoming analysis of distribution of each haplogroup and detecting of common distribution obviously show the divergence pattern clearly which in turn suggest the language discontinuity.

**Distribution of M5**

M2 M3 M4 M5 (Ancestral population of Dungribhil, kathodi, Andh, Kamar, Paudibhuiya, Hillkolam and Malpaharia)

M2 M4 M 5 (Ancestral population of Korku, Munda, \*Bettakuruba, Katakur and Nahal)

**Distribution of M6**

M2 M3 M4 M5 M6 (Ancestral population of DungriBhil, Andh, kamar, pauidibhuiya, Hillkolam and Malpahariya)

M2 M3 M4 M6 (Ancestral population of Madia and Sonalkachari)

M2 M4 M5 M6 (Ancestral population of Korku, Munda and Nahal)

M3 M4 M6 (Ancestral population of Gallong and DirangMonpa)

M4 M6 (Ancestral population of Toto and Wancho)

M6 (Ancestral population of Lachungpa and Lepcha)

**Distribution of Haplogroup D**

M2 M3 M4 M5 M6 D (Ancestral population of DungriBhil)

M2 M3 M4 M6 D (Ancestral population of Sonalkachari)

**M3 M4 D (SHERDUKPAN)\***

M3 M4 M6 D (Ancestral population of Gallong and DirangMonpa)

M4 M6 D (Ancestral population of Toto and Wancho)

M6 D (Ancestral population of Lepcha and Lachungpa)

**Distribution of M8\_C\_Z**

M2 M3 M4 M6 D M8\_C\_Z (Ancestral population of sonalkachari)

M3 M4 M6 D M8\_C\_Z (Ancestral population of Gallong and DirangMonpa)

M3 M8\_C\_Z (Ancestral population of Jenukuruba)

M4 M6 D M8\_C\_Z (Ancestral population of Wancho)

M6 D M8\_C\_Z (Ancestral population of LACHUNGPA and Lepcha)

**Distribution of M9**

M2 M3 M4 M6 D M8\_C\_Z M9 (Ancestral population of Sonalkachari)

M3 M4 M6 D M8\_C\_Z M9 (Ancestral population of Gallong and DirangMonpa)

M3 M8\_C\_Z M9 (Ancestral population of Jenukuruba)

M4 M6 D M8\_C\_Z M9 (WANCHO)

M6 D M8\_C\_Z M9 (Ancestral population of Lachungpa and Lepcha)

**Distribution of M12\_G**

M2 M3 M4 M5 M6 M12\_G (Ancestral population of kamar and pauidibhuiya)

M2 M4 M5 M12\_G (Ancestral population of katakur)

M2 M4 M5 M6 M12\_G (Ancestral population of Nahal)

M2 M3 M4 M12\_G (Ancestral population of Katkari)

M3 M4M6DM8\_C\_ZM9M12\_G (Ancestral population of GALLONG and DirangMonpa)

M6 D M8\_C\_Z M9 M12\_G (Ancestral population of Lepcha)

**Distribution of M18**

M2 M3 M4 M5 M6 D M18 (Ancestral population of Dungribhil)

M2 M3 M4 M5 M6 M12\_G M18 (Ancestral population of pauidibhuiya)

M2 M3 M4 M5 M6 M18 (Ancestral population of HILLKOLAM and Malpahariya)

M2 M4 M5 M6 M18 (Ancestral population of Korku and Munda)

M2 M4 M5 M6 M12_G M18 (Ancestral population of Nahal)	M2 M4 M5 M6 M12_G M18 M25 M30 M33 (Ancestral population of Nihal)
M4 M6D M18 (Ancestral population of Toto)	M4 M6D M18 M33 (Ancestral population of Toto)
M2 M3 M4 M6 D M8_C_Z M9 M18 (Ancestral population of Sonalkachari)	M6 D M8_C_Z M9 M12_G M18 M33 (LEPCHA)
M6 D M8_C_Z M9 M12_G M18 (Ancestral population of Lepcha)	<b>Distribution of M35</b>
<b>Distribution of M25</b>	M2 M3 M4 M6 M25 M35 (Ancestral population of Madia)
M2 M3 M4 M5 M6 M12_G M18 M25 (Ancestral population of paudibhuiya)	M2 M3 M4 M5 M6 M30 M35 (Ancestral population of Andh)
M2 M3 M4 M6 M25 (Ancestral population of Madia)	M2 M3 M4 M5 M30 M35 (Ancestral population of kathodi)
M2 M4 M5 M6 M12_G M18 M25 (Ancestral population of Nahal)	M2 M4 M5 M6 M12_G M18 M25 M30 M33 M35 (Ancestral population of Nahal)
M2 M4 M5 M6 M18 M25 (Ancestral population of Munda)	M2 M4 M5 M6 M18 M25 M31 M33 M35 (Ancestral population of Munda)
M3 M4M6DM8_C_ZM9M12_G M25 (Ancestral population of DirangMonpa)	M4 M6D M18 M33 M35 (TOTO)
M3 M8_C_Z M9 M25 (Ancestral population of Jenukuruba)	M2 M4 M5 M35 (BETTAKURUBA)
<b>Distribution of M30</b>	<b>Distribution of M36</b>
M2 M3 M4 M5 M30 (Ancestral population of kathodi)	M3 M8_C_Z M9 M25 M36 (JENUKURUBA)
M2 M3 M4 M5 M6 D M18 M30 (Ancestral population of Dugribhil)	M2 M3 M4 M5 M6 M12_G M36 (Ancestral population of kamar)
M2 M3 M4 M5 M6 M30 (Ancestral population of Andh)	<b>Distribution of M37</b>
M2 M4 M5 M6 M18 M30 (Ancestral population of Korku)	M2 M4 M5 M6 M12_G M18 M25 M30 M33 M35 M37 (NAHAL)
M2 M4 M30 (Matakur)	<b>Distribution of M38</b>
M2 M4 M5 M6 M12_G M18 M25 M30 (Ancestral population of Nahal)	M2 M3 M4 M12_G M38 (KATKARI)
<b>Distribution of M31</b>	M2 M4 M5 M6 M18 M30 M33 M38 (Ancestral population of Korku)
M2 M3 M4 M5 M6 M12_G M18 M25 M31 (Ancestral population of paudibhuiya)	<b>Distribution of M39</b>
M2 M4 M5 M6 M18 M25 M31 (Ancestral population of Munda)	M2 M3 M4 M6 M25 M35 M39 (MADIA)
<b>Distribution of M33</b>	M2 M3 M4 M5 M6 M30 M35 M39 (ANDH)
M2 M3 M4 M5 M6 D M18 M30 M33 (DUNGRIBHIL)	M2 M4 M5 M12_G M39 (KATAKUR)
M2 M3 M4 M5 M6 M18 M33 (Ancestral population of Malpahariya)	M2 M3 M4 M5 M30 M35 M39 (KATHODI)
M2 M4 M5 M6 M18 M30 M33 (Ancestral population of Korku)	M2 M4 M5 M6 M18 M25 M31 M33 M35 M39 (Ancestral population of Munda)
M2 M4 M5 M6 M18 M25 M31 M33 (Ancestral population of Munda)	M2 M3 M4 M5 M6 M18 M33 M39 (Ancestral population of Malpahariya)
	M2 M3 M4 M5 M6 M12_G M18 M25 M31 M39 (Ancestral population of paudibhuiya)
	M2 M3 M4 M5 M6 M12_G M36 M39 (Ancestral population of kamar)



**Distribution of M40**

M2 M3 M4 M5 M6 M12\_G M18 M25 M31 M39 M40  
(PAUDIBHUIYA)

M2 M3 M4 M5 M6 M12\_G M36 M39 M40 (KAMAR)

M2 M4 M5 M6 M18 M25 M31 M33 M35 M39 M40  
(Ancestral population of Munda)

**Distribution of M41**

M2 M3 M4 M5 M6 M18 M33 M39 M41  
(MALPAHARIYA)

**Distribution of M42**

M2 M4 M5 M6 M18 M30 M33 M38 M42 (KORKU)

M2 M4 M5 M6 M18 M25 M31 M33 M35 M39 M40 M42  
(MUNDA)

**Distribution of M49**

M3 M4M6DM8\_C\_ZM9M12\_G M25 M49  
(DIRANGMONPA)

M2 M3 M4 M6 D M8\_C\_Z M9 M18 M49  
(SONALKACHARI)

This analysis has shown the pattern that each ancestral group after divergence again undergo admixture and finally defined the present-day population structure. The general assumption is that Geographic proximities govern the pattern of gene flow while linguistic boundaries often restrict gene flow. From the above, it is clear that the 24 studied groups are distantly distributed at present. Yet they show some patterns of ancestral haplogroup sharing. It clearly indicates the present-day linguistic affiliation is a derived or latter adapted phenomenon. The original language in accordance with the age of the ancestral groups based on heritage haplogroups may not be continued in any of these population groups. Except Nihal, all these population groups are the speakers of Holocene immigrant language families like Austro Asiatic, Dravidian, Indo Aryan and Tibeto Burman. But their heritage haplogroups of their ancestral population is not derived during Holocene influx, thus it defines linguistic discontinuity. This study exposes the appropriate divergence time and later admixtures by showing the haplogroup structure. Our study vehemently argues the haplogroup distribution analysis can show linguistic prehistory as lost language lineages of the present day linguistically unrelated populations of India. The Paleolithic genetic continuity shown across language families confirms that Indian speech communities don't have corresponding Paleolithic linguistic lineages. Presence of M31 and M 32 among the Andaman negritos and their isolated linguistic status further confirm the Paleolithic linguistic ancestry. This obviously suggests that the attestation of current linguistic status to the genetic phylogeny of south Asia is mismatching and it obviously suggests language replacements.

**CONCLUSION**

This study was carried out to know the patterns of linguistic and genetic parallelism among the selected population groups in order to critically review the concept of language-gene co-evolution. It is undoubtedly displayed that the Language divergence and genetic divergence are not simultaneous and Paleolithic genetic continuity is not corresponded with their linguistic status. It suggests the existing peripheral understanding on languages is not a dependable trait to make attestation of the deep-rooted linguistic prehistory of populations. There is gap of information on Pleistocene linguistic lineages. The present mode of assigning language status to genetically archaic groups obviously has inherent limitations to express their pre-Neolithic linguistic lineages [26-69].

**REFERENCES**

1. Reich D, Thangaraj K, Patterson N, Price AL, Singh L, et al. (2009) Reconstructing Indian population history. *Nature* 461: 489-494.
2. Burrow T (1945) Some Dravidian words in Sanskrit. *TPS* 1945: 79-100.
3. Burrow T (1947a) Dravidian Studies vii: further Dravidian words in Sanskrit. *BSOAS* 12: 365-396.
4. Burrow T (1973) The Proto-Indo-Aryans. *JRAS* 1: 123-140.
5. Burrow T (1946) Loan words in Sanskrit. *TPS* 1-30.
6. Emeneau MB (1956) India as a linguistic area. *Language* 32: 3-16.
7. Emeneau MB (1966) The dialects of Old Indo-Aryan. *Ancient Indo-European Dialects*, Berkeley and Los Angeles, CA: University of California, pp: 123-138.
8. Emeneau MB (1954) Linguistic prehistory of India. *PAPS* 98: 282-292.
9. Burrow T, Emeneau MB (1984) *A Dravidian Etymological Dictionary* 2nd ed. Oxford: Clarendon Press. Oxford: Oxford University Press.
10. Hock HH (1975) Substratum influence on (Rig Vedic) Sanskrit? *Stud Ling Sci* 5: 76-125.
11. Kuiper FB (1955) Rigvedic loanwords. *Studia Indologica: Festschrift fü*.
12. Southworth FC (1974) Linguistic stratigraphy of north India. *Contact and Convergence in South Asian Languages*. *IJDL* (1974): 201-223.
13. Witzel M (1999) Substrate languages in old Indo-Aryan. *EJVS* 5: 1-67.
14. Levi S (1923) Pre-Aryan and Pre-Dravidian India. *J Asian* 203: 51-57.

15. Kuiper FB (1948) Proto-Munda words in Sanskrit. *Versions der Kon, Nederl, Academy of Sciences, Literature Department* 51: 3.
16. Deshpande MM, Hook PE (1979) Aryan and non-Aryan in India. Ann Arbor: The University of Michigan Center for South and Southeast Asian Studies, pp: 327.
17. Southworth FC (2005) *Linguistic Archaeology of South Asia*, Routledge Curzon, London, pp: xiv+369.
18. Masica CP (1979) Aryan and non-Aryan elements in north Indian agriculture. *Aryan and non-Aryan in India*, Ann Arbor: The University of Michigan Center for South & Southeast Asian Studies, pp: 55-152.
19. Mountain JL, Hebert JM, Bhattacharyya S, Underhill PA, Ottolenghi C, et al. (1995) Demographic history of India and mtDNA-sequence diversity. *Am J Hum Genet* 56: 979-992.
20. Kivisild T, Bamshad MJ, Kaldma K, Metspalu M, Metspalu E, et al. (1999) Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages. *Curr Biol* 9: 1331-1334.
21. Kivisild T, Kaldma K, Metspalu M, Parik J, Papiha SS, et al. (1999b) The place of the Indian mitochondrial DNA variants in the global network of maternal lineages and the peopling of the Old World, pp: 135-152.
22. Kivisild T, Rootsi S, Metspalu M, Mastana S, Kaldma K, 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.
23. 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.
24. Palanichamy MG, Agrawal S, Yao YG, Kong QP, Sun C, et al. (2006) Comment on "Reconstructing the origin of Andaman islanders". *Science* 311: 470.
25. Trivedi R, Sahoo S, Singh A, G Bindu H, 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.
26. Bamshad M, Kivisild T, Watkins WS, Dixon ME, Ricker CE, et al. (2001) Genetic evidence on the origins of Indian caste populations. *Genome Res* 11: 994-1004.
27. Barik SS, Sahani R, Prasad BV, Endicott P, Metspalu M, et al. (2008) Detailed mtDNA genotypes permit a reassessment of the settlement and population structure of the Andaman Islands. *Am J Phys Anthropol* 136: 19-27.
28. Barnabas S, Apte RV, Suresh CG (1996) Ancestry and interrelationships of the Indians and their relationship with other world populations: A study based on mitochondrial DNA polymorphisms. *Ann Hum Genet* 60: 409-422.
29. Basu A, Mukherjee N, Roy S, Sengupta S, Banerjee S, et al. (2003) Ethnic India a genomic view, with special reference to peopling and structure. *Genome Res* 13: 2277-2290.
30. Cavalli-Sforza LL, Menozzi P, Piazza A (1994) *The History and Geography of Human Genes* Princeton Univ. Press, Princeton.
31. Chandrasekar A, Kumar S, Sreenath J, Sarkar BN, Urade BP, et al. (2009) Updating phylogeny of mitochondrial DNA macrohaplogroup m in India dispersal of modern human in South Asian corridor. *PLoS One* 4: e7447.
32. Chaubey G, Metspalu M, Kivisild T, Villems R (2007) Peopling of South Asia investigating the caste-tribe continuum in India. *Bioessays* 29: 91-100.
33. Chaubey G, Karmin M, Metspalu E, Metspalu M, Selvi-Rani D, et al. (2008) Phylogeography of mtDNA haplogroup R7 in the Indian peninsula. *BMCE vol Biol* 8: 227.
34. Chaubey G, Metspalu M, Karmin M, Thangaraj K, Rootsi S1, et al. (2008) Language Shift by Indigenous Population: A Model Genetic Study in South Asia. *Int J Hum Genet* 8: 41-50.
35. David R (2009) Reconstructing Indian Population History. *Nature* 461: 489-494.
36. Fairservis WA (1992) *The Harappan Civilization and its Writing: A Model for the Decipherment of the Indus Script*. Delhi: Oxford & IBH.
37. *Diamond J, Belwood P (2003) Farmers and their languages. Science* 300: 597-603.
38. Kumar V, Reddy AN, Babu JP, Rao TN, Langstieh BT, et al. (2007) Y-chromosome evidence suggests a common paternal heritage of Austro-Asiatic populations. *BMC Evol Biol* 7: 47.
39. Lahr M, Foley R (1994) Multiple dispersals and modern human origins. *Evol Anthropol* 3: 4840.
40. Macaulay V, Hill C, Achilli A, Rengo C, Clarke D, et al. (2005) Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science* 308: 1034-1036.

41. Masica CP (1976) Defining a Linguistic Area: South Asia, Chicago. IL: University of Chicago Press, pp: 201-210.
42. Masica CP (1991) The Indo-Aryan Languages. Cambridge: Cambridge University Press.
43. McAlpin DW (1981) Proto-Elamo-Dravidian: The evidence and its implications, Philadelphia. PA: The American Philosophical Society, pp: 1-155.
44. Metspalu M, Romero IG, Yunusbayev B, Chaubey G, Mallick CB, 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.
45. Narang A, Jha P, Rawat V, Mukhopadhyay A, Dash D, et al. (2011) Recent admixture in an Indian population of African ancestry. *Am J Hum Genet* 89: 111-120.
46. Nei M, Roy Choudhury A (1982) Genetic relationship and evolution of human races. *Evol Biol* 14: 1-59.
47. Parpola A (1994) Deciphering the Indus Script, Cambridge University Press, pp: 114-135.
48. Renfrew C (2000) in *Archaeogenetics: DNA and the Population Prehistory of Europe*. (eds) Renfrew C & Boyle K (McDonald Institute, Cambridge, U.K.), pp: 3-12.
49. Choudhury RS, Roy S, Basu A, Banerjee R, Vishwanathan H, et al. (2001) Genomic structures and population histories of linguistically distinct tribal groups of India. *Hum Genet* 109: 339-350.
50. 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 USA* 103: 843-848.
51. Sengupta S, Zhivotovsky LA, King R, Mehdi SQ, Edmonds CA, 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.
52. Sreenathan M (2010) Language plurality and functional heterogeneity among people of India in Identity. *Cultural pluralism and State- South Asia in perspective*, Das N K & Rao V R (eds) Macmillan & AnSI.
53. Sreenathan M (2015) Compatibility of linguistic and genetic co evolution. *Indian J Phys Anthropol Hum Genet* 34: 2.
54. Sreenathan M (2015) Prehistory of personal pronouns of Andamanese family in studies on Indian languages and cultures. V I Subramoniam commemoration volume II, International school of Dravidian linguistics, Trivandrum, India.
55. Morris S (1952) Lexico statistic dating of Prehistoric Ethnic Contacts. *Proc Am Philos Soc* 96: 452-463.
56. Morris S (1953) Archaeological and Linguistic Chronology of Indo-European languages. *Am Anthropol* 55: 349-352.
57. Thangaraj K, Ramana GV, Singh L (1999) Y-chromosome and mitochondrial DNA polymorphisms in Indian populations. *Electrophoresis* 20: 1743-1747.
58. Thangaraj K, Singh L, Reddy AG, Rao VR, Sehgal SC, et al. (2003) Genetic affinities of the Andaman Islanders, a vanishing human population. *Curr Biol* 13: 86-93.
59. Thangaraj K, Chaubey G, Kivisild T, Reddy AG, Singh VK, et al. (2005) Reconstructing the origin of Andaman Islanders. *Science* 308: 996.
60. Thangaraj K, Chaubey G, Singh VK, Vanniarajan A, Thanseem I, et al. (2006a) In situ origin of deep rooting lineages of mitochondrial Macrohaplo group 'M' in India. *BMC Genomics* 7: 151.
61. Thangaraj K, Chaubey G, Reddy AG, Singh VK, Singh L, et al. (2006b) Unique origin of Andaman Islanders insight from autosomal loci. *J Hum Genet* 51: 800-804.
62. Thangaraj K, Nandan A, Sharma V, Sharma VK, Easwarkhanth M, et al. (2009) Deep rooting in-situ expansion of mtDNA Haplogroup R8 in South Asia. *PLoS One* 4: e6545.
63. Thangaraj K, Naidu BP, Crivellaro F, Tamang R, Upadhyay S, et al. (2010) The influence of natural barriers in shaping the genetic structure of Maharashtra populations. *PLoS One* 5: e15283.
64. Underhill PA, Kivisild T (2007) Use of y chromosome and mitochondrial DNA population structure in tracing human migrations. *Annu Rev Genet* 41: 539-564.
65. Underhill PA, Myres NM, Rootsi S, Metspalu M, Zhivotovsky LA, et al. (2010) Separating the post-Glacial coancestry of European and Asian Y chromosomes within haplogroup R1a. *Eur J Hum Genet* 18: 479-484.
66. Tripathy V, Nirmala A, Reddy MB (2008) Trends in molecular anthropological studies in India. *Int J Hum Genet* 8: 1-20.
67. Watkins WS, Bamshad M, Dixon ME, Bhaskara RB, Naidu JM, et al. (1999) Multiple origins of the mtDNA



- 9-bp deletion in populations of South India. *Am J Phys Anthropol* 109: 147-158.
68. Witzel M (1989) Tracing the Vedic dialects. *Dialects in Indo-Aryan literatures* Paris: Institut de Civilization Indienne, pp: 97-264.
69. Witzel M (1995) Early Indian history: Linguistic and textural parameters. *The Indo-Aryans of Ancient South Asia: Language, Material Culture and Ethnicity*, Berlin: Walter de Gruyter, pp: 85-125.