COMPATIBILITY OF LINGUISTIC AND GENETIC CO-EVOLUTION IN INDIA

M. Sreenathan[*]

ABSTRACT

India proffers enormous genetic and linguistic diversity. The aim of this paper is to investigate compatibility between linguistic and genetic patterns in order to verify genetic and linguistic co-evolution hypothesis in the Indian context. A complex population diversity map reveals the presence of multiple population migrations and admixtures. Various studies suggest linguistically and genetically homogenous, linguistically homogenous and genetically heterogeneous and genetically homogenous and linguistically heterogeneous patterns among Indian population. Despite the common ancestry, individual linguistic groups / family show substantial differences due to differential admixture with neighboring groups as well as isolation and drift. Identifying the compatibility between the embedded genetic and linguistic substratum and language shift among population is often challenging. Thus it is proposed to unravel systematically the correlation between distribution of Indian languages, mtDNA and Y chromosomes in order to understand patterns of convergence and divergence that defined linguistic and genetic continuities and discontinuities in Indian populations. Both Language and genetic phylogenies would throw light on human prehistory but its correlations may be challenging. This study suggests that, beyond the existing knowledge on language and family linkages, typological information should be given due consideration in interpreting genetic results based on language. Also, suggest the requirement of an interdisciplinary framework to accommodate non genomic along with genomic results to fix the compatibility of interdisciplinary results in prehistoric understanding.

BACKGROUND

Unraveling the history of peopling and interrelation of population groups no doubt has been attractive to many disciplines. Tools of knowing the prehistory of human kind have been contributed by many disciplines. Philology had tried to untie prehistory through comparative method which in turn made the world witness inter relationships of populations of distant areas. Also lexical statistical dating (glottochronology) approaches are used in linguistics to find out linguistic chronology of language divergence (Swadesh, 1952). Anthropology, Archeology, Paleontology, etc. too tried to map the traits of common and distinct, through

Dr. M. Sreenathan, Professor and Faculty Dean of Linguistics, Thunchath Ezhuthachan Malayalam University, Vakkad, Tirur, Malappuram, Kerala, Pin: 676502, India, *E-mail: msreenathan@gmail.com*

comparative method. Genomic mapping through Molecular biology is the latest, trying to trace human ancestry again through a kind of comparative method. However, at present, genomic evidences have succeeded in tracing the spread of anatomically modern humans originated in Africa. Without doubt, the genetic signatures have proved successfully the importance of India in the history of human dispersal.

South Asia is important in the history of Human migration. It is one of the first regions to have been peopled by modern humans. India, has been recognized with the history of waves of human dispersal, explored as a hotspot of genetic diversity in corroboration with cultural, linguistic, social and ecological diversity (Bamshad *et al.,* 2001; Barik *et al.,* 2008; Barnabas *et al.,* 1996; Basu *et al.,* 2003; Chandrasekar *et al.,* 2009; Chaubey *et al.,* 2007; Kivisild *et al.,* 1999; Kivisild *et al.,* 2003; Kumar *et al.,* 2006; Kumar *et al.,* 2007; Metspalu *et al.,* 2004; Metspalu *et al.,* 2011; Mountain *et al.,* 1995; Nei *et al.,* 1988; Palanichamy *et al.,* 2006; Rakesh *et al.,* 2012; Sahoo *et al.,* 2006; Thangaraj *et al.,* 2003, 2005, 2006a, 2006b, 2009, 2010; Watkins, 1999). Recent studies suggest two major ancestral contributions to most of the Indian sub-populations (Tamang *et al.,* 2012). To sum up, both ancient and recent genetic signatures are found in India. In situ origin of mtDNA clades are detected. Undoubtedly these studies establish the role of India in early human migration.

In this context, molecular results need corroborative evidences to justify its social relevance. Corroborative evidences from other disciplines like Linguistics, Archeology, Paleontology, Anthropology, Folklore studies, etc. are having immense significance in establishing the prehistoric substrata. In this study, an attempt is made to amalgamate linguistic and molecular evidences.

Understanding linguistic prehistory by applying molecular genetics is a significant part of archaeo-genetics. Numerous studies exposed strong associations between languages and genes among human populations at continental and global scales (Cavalli-Sforza *et al.,* 1994; Renfrew, 2000; Lahr and Foley, 1994). Both genes and languages will drift apart regularly over time, the former slowly, the latter much more quickly. Aerial distribution of genetic markers and their world language alliances promulgates the theory of parallel evolution of genes. Genes are not people and they can have distributional logic quite different from languages and cultures and hence Paleolithic continuity theory attests the view that genetic continuity does not imply linguistic continuity. The compatibility between genetic and linguistic relations thus subscribes the debate of gene-language continuity and discontinuity. In accordance, this correlation was hypothesized as part of farming expansions (*Jared Diamond and Peter Belwood,* 2003). Genetic and linguistic variation evolves in parallel after the genes and languages of farmers replace those of hunter-gatherers in the path of expansion. Many attempts have been made to establish genetic and linguistic links in India. The results so far have been not of much help to unravel gene – language connection. It is mainly because of the non-correlation between biological, cultural and linguistic categories.

LINGUISTIC ATTEMPTS TO PREHISTORY

First attempt from language study evolved when philology is the comparative method used in reconstruction of proto ancestors. Since 19thC this could establish many language families in the world. Reconstruction of linguistic ancestors (proto languages) by using historical comparative method is commonly considered as an end to understand the linguistic prehistory. In mid of 20thC, Glottochronology had developed lexicostatistic approach to mark linguistic chronology of language divergence. Distant level comparison became the current practice of understanding long range relationships between language families. Thus to some of the new generation, Linguists reconstructed proto languages are the beginning. Beyond the description of families, they are studying on interconnection between families to establish the proto language ancestry. Lexical and morphological correspondences of many Eurasian and African languages have led to the hypothesis that they derive from a common linguistic ancestor, Nostratic. Nostratic is a macro family of languages and is regarded as a hypothetical super ancestor of certain language families of Eurasia and Africa. It means Nostratic is a proto language of proto languages. Nostratic as a model of deep linguistic relationships has been developed including Indo-European, Uralic, Altaic, Kartvelian, Dravidian, and Afro Asiatic (Hemito Semitic) families.

Extension of the traditional comparative method is diversely practiced among the Nostraticists. Comparing resemblances is a method followed by Greenberg to show how the world languages are related by using dictionary comparison. Protolanguage comparison is the norm among the Moscovists. Tentative attempts of the above kind reveal lots of hopes to believe that human language sprung from common source. Ultimate areal comparison would be Proto-World, the hypothetical ancestor of all human languages. However, most linguists continue to remain skeptical on the accuracy of reconstruction. But paleo-linguists generally agree that Eurasia was originally populated (10000-15000 BP) by the Nostratic groups later to become Indo-European, Uralic, Altaic, Kartvelian, Dravidian, and Afro Asiatic (Hemito Semitic) families.

NOSTRATIC AND FARMING DISPERSAL

The enquiries on geo-linguistic patterning of language families based on genetic information clearly demarcates between the results of cultural and demographic process (Cavalli *et al.,* 1994; Chen *et al.,* 1995; Nettle and Harris, 2003). Gene frequencies of linguistic groups clearly attest whether the spread resulted as part of biological or linguistic evolution. The main demographic process associated with cultural change, demic diffusion, is the expansion into additional territories of a population whose size is increasing. Genetic variation in the world are largely reflecting populations expansion starting in Neolithic and permitted by an advanced subsistence technology, farming. Ability to produce food increased population densities. Populations then expanded outward in four major waves, with each wave propagating farming along with a protolanguage from which Indo-European,

Elamo-Dravidian, Afro-Asiatic, and Altaic later developed. Renefrew (1997) described the above language families are the outcome of the post Pleistocene farming dispersal after 10,000BP. Language resemblance on such a large scale may be due to either cultural exchange between sedentary populations or a demographic process whereby the speakers of a language move into different areas (Renefrew, 1992). Both phenomena have had an influence on the distribution of contemporary languages. The presence of systematic set of correspondences – phono semantic sets cannot be result of chance. Barbujani and Andrea Pilastro (1993) defined it as Nostratic demic diffusion (NDD) and called these language families the NDD families on the light of Gene-frequency within these linguistic families. Their study suggest that language diffusion was largely associated with population movements (Farming dispersals) rather than with purely cultural transmission (Sreenathan, 2010).

LINGUISTIC DIVERSITY OF INDIA

Four language families are generally recognized as having presence in India. The overall notion of a family indicates that it is a group of related languages and such related languages as descended from a common parent language. The four language families of India-Indo Aryan, Dravidian, Austro-Asiatic, and Tibeto-Burman – are understood as the surviving representatives of the erstwhile language families that originated in South Asia. 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. In fact, the latter must have subsumed many of the then existing languages spoken by the Paleolithic hunter gatherers. The spread of agriculture entailed replacement of some languages by other languages. Thus large numbers of minor languages spoken by hunting and gathering groups must have disappeared due to the advancement made during the Neolithic age. The present patterns of Indian languages indicate that farming communities gradually fragmented into small units and their language began to diverge and produced modern Indian languages. Besides, said language families, the presence of isolated linguistic remnants like the Nahali (Central India), Nilgiri languages, and the Andamanese family of languages connects India with prehistoric substratum. The entire linguistic scenario of India underlines the fact that the language isolates like the Andamanse 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 (Sreenathan, 2009).

PRE HISTORIC LINGUISTIC GROUPS OF SOUTH ASIA

South Asia offers a different population history. The late Pleistocene and early Holocene population history of South Asia is not thoroughly known which in turn defines the importance of linguistic archeology. There are pure historical and comparative linguistic methods in use to unravel the antiquity of languages. It has

been hoped that the decipherment of the Indus valley seal inscriptions would expose the early linguistic situation of South Asia. There are serious attempts to unravel the inscriptions (Fairservis, 1992, Parpola, 1994). Despite of many claims, yet there has been no generally accepted interpretation of the Indus script. Also, an argument exists that even if it is deciphered, may not provide the details of the then existed languages.

Script centric cultures focused the idea of knowing the past through written evidences. There are Old Indo Aryan written sources and the history of investigations focused the arguments on the presence of non-Aryan elements on it. The Austro Asiatic and Dravidian connections are suggestively argued. Deep rooting of Dravidian connectivity is argued heavily by Burrow (1945, 1947, 1973), Emeneau (1943, 1954), Emeneau and Burrow (1962), Hock (1975), Kuiper (1955, 1967), Southworth (1979), Witzel (1999). Also there are arguments related to the Austro Asiatic connection. Levi (1923), Kuiper (1948, 1955, 1991), Witzel (1999), Hock (1975) 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 (Kuiper, 1967) is considered as an internal innovation (Hock, 1975), as post Rigvedic phenomenon (Deshpande, 1979), and diffusion of it from Pre-Mundaic language of Panjab (Witzel, 1999). Another argument is centered on the evidence of loanwords in old Indo Aryan (OIA) (Burrow, 1945, 1947, 1973; Emeneau, 1943, 1954; Emeneau and Burrow, 1968). Syntactic influence of Dravidian on old Indo Aryan is also suggested by Kuiper (1967), Emenaue (1954), but Hock (1975) has refuted it. Witzel (1999) discussed it as of structural innovations in old Indo Aryan.

The pre historic linguistic map was reconstructed by Southworth (2005) 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. The existence of probable Munda/Austro-Asiatic words in OIA throughout the entire Vedic period *Inferred 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 (Southworth, 1988); (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 (Witzel, 1999); (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 (Zvelebil, 1990, Witzel, 1999); (e) The Vedda substrate in Sri Lanka,

inferred on the basis of loanwords and collocations in Sinhala (De Silva, 1972); (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 (1999) (g) Masica (1979) 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 (1999) 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, do not appear until the middle Rig Vedic period. The largest number of words in this group are those which occur in old Indo Aryan and in early Dravidian, and which do not appear to be original in either group; since they do not have the typical Munda characteristics, they may be provisionally considered to have originated in the otherwise unknown "Indus" languages. Kuiper has long held that the earliest identifiable foreign words in the Rigveda are of Munda or AA origin (1948, 1955, 1991), which would imply the presence of speakers of these languages in the Panjab as early as the second millennium BCE. The Rigveda alone contains more than 300 such words.

Andamanese language family alone qualifies the prehistoric languages in its true sense in India. Isolates like Nahali maintains near about 30% of exclusive vocabulary suggesting the vestiges of a prehistoric presence and Munda and Dravidian languages are also holding the prehistoric substratum. Prehistoric stages of languages can be connected with archeological evidences and molecular evidences. Members of the identified families of languages are historic while their proto ancestor may be of prehistoric.

Understanding linguistic prehistory by applying molecular genetics is significant. The mtDNA lineages (Kivisild *et al.,* 1999, 2003; Mountain *et al.,* 1995) are important in understanding human population movements and the language phylogeny across India (Roychoudhury *et al.,* 2001; Tripathy *et al.,* 2008; Trivedi *et al.,* 2008; Chaubey *et al.,* 2008). The connection of genetic history of the Indian subcontinent, with the languages has not been well documented. The language diversity remains controversial at the biological level as there are discordant results regarding the biogeographical separation between the major linguistic groups. Present day genetic patterns suggest a limitation to gene flow for some language groups, preventing population admixture. For instance, haplogroups such as M4 a, M40a, M45,R7 and R6a are regarded as Mudari specific, M2, M6a Austro Asiatic, M2a (Dravidian), M2b (TibetoBurman), M5 (Khasi) M6 (Nicobarese) and other haplogroups (Majumder, 2001; Metspalu, 2004; Baig *et al.,* 2004; Sun *et al.,* 2006; Kumar *et al.,*

2007). In recent years, number of studies has challenged the hypothesis of the Indian linguistic diversity representing genetic diversity. A review of existing genomic studies in general show trends of correlation between linguistic and genetic boundaries in India as Physical barriers that may have caused both genetic and linguistic boundaries. There are only linguistic and genetic boundaries but not physical ones or cases of linguistic boundaries have generated or enhanced genetic boundaries or both may be consequence of political, cultural, and social boundaries that have played a role similar to that of physical barriers. Vikal *et al.,* (2008) critically reviewed various molecular Anthropological studies in India till date and presented salient outcomes of these studies. Instead of repeatedly referring individual studies, this paper fixes the said review as a preface. Despite having differences in inferences, these studies succeeded in attesting some trends towards population affinities and routes of migration at the global level. Molecular debates on much perplexed views about initial peopling, tribe - caste dichotomy, Linguistic diversity etc. have significantly revolutionized our understanding on India. Deep rooted maternal affinity between tribe and caste population has suggested common origin and the present diversity is caused by the subsequent admixture and genetic drift. However, the basic clustering of maternal or paternal lineages shows a common trend that they are not clustered around particular languages.

The Paleolithic ancestry of contemporary Indians is one of the outcomes of many mtDNA studies on Indian population. The picture presented by many studies is the substantial genetic continuity between modern groups and the Paleolithic hunter-gatherers who inhabited the same region thousands of years ago. Against the earlier assumption, various demographic and evolutionary mechanisms may have made immense genetic contributions from migratory people that occurred during the Neolithic, and into the Bronze and Iron Ages. Indians descended from hunter gatherer populations who lived until Neolithic expansion. Paleolithic ancestry is seen as remaining relatively unaffected by later gene flow including any large-scale movements of farmers out of the Middle East during the Neolithic era. These agriculturalists are in fact presented as outsiders who left only limited genetic traces among contemporary Indians, who derive most of the ancestry from indigenous groups that adopted agricultural practice through diffusion process. Thus, the immigrant gene flow may have been not so significant even amounting to wholesale population replacement. Founder effects, genetic drift and bottlenecks also have had a dramatic impact. In addition, Darwinian principles of natural selection and resistance against disease may have changed the face of India over time, causing certain genetic groups to disappear while others have come to dominate the genetic landscape. It cannot be believed that these events, either alone or in combination, have resulted in a striking genetic discontinuity between past and present populations.

The corollary structure of the Indian antiquity is a Neolithic expression. The history of communication system is premeditated as a Neolithic lineage. The linguists have explored India as a linguistic area and in extended sense even India as part of

south Asian linguistic area on the basis of language typology and projected long time contact as its rationale. The structural dynamics of convergence and divergence have been found astound. This linguistic hypothesis of validations framed out of comparative linguistic tradition and language typology positioned India as a land of imported languages at one hand and the web of contact and pattern convergence at the other end. The prevailing awareness on Indian language phylogeny confirms the accordance of almost 100% of ethno linguistic groups under four language families (Austro-Asiatic, Dravidian, Indo Aryan and Tibeto-Burman). The Andamanese language family is the outlier. Austro-Asiatic speakers are dispersed mostly in the central and eastern parts of the country while the Tibeto-Burman speakers are concentrated in the northeastern part of the subcontinent. Dravidians are mainly confined to South India with some exceptions like Kurukh and Gondi speaking people, living in central and eastern parts of India. Indo-European is the most widely spread family, which covers North, Central and Western parts of India. The origin and dispersal of these language families remains unresolved. Reconstruction of hypothetical linguistic ancestors further landed in Europe and other parts of Asia. Linguistics thus helped India to accept as a land of immigrants and their localization. Equating language with people accentuated the dichotomy of immigrants and locals which they couldn't affix time and space unconvincingly. The pre-history of India thus stranded in the dynamics of immigrant time and associated cultural corroborations. A make belief circularity of ideas permits the Neolithic communication systems as the antique expression of India. The protagonist of prehistory quite often became language which tranquillized pre-history as history on language lineages of India despite the fact that except the outlier Andamanese language family, all major language families are Neolithic lineages. The profound episode as far as India is concerned, was not initially peopled during Neolithic times. Modern human remains dating back to the late Pleistocene (55000–25000 years before present, ybp) and by the middle Paleolithic period (50,000–20,000 ybp), are found in India, subscribing Allchin's view that "the descendants of early and middle Stone Age formed a considerable element in the late Stone Age population of the subcontinent". The genomic explorations have evidently projected India as a corridor for dispersal of modern humans in concordance with southern route hypothesis. Locating high percentage of the Pleistocene specific Indian mtDNA lineages and Y lineages among the Indian populace cutting across transvariable (language, environment, culture, caste, economy, etc.) suggest a founder genomic substratum underlying in the diverse population groups in India. This scene encapsulates the view that genetic continuity does not imply linguistic continuity. So far, the nation has been encountering the question of original inhabitants without substantial deep rooted evidences. The stratified understandings on peoples of India in terms of language have framed various questions. Who are the original inhabitants of India among the language isolates, Austro-Asiatic and Dravidian speaking populations? Many of the non-genomic studies (Risley, 1915; Rapson, 1955; Thapar, 1966; Thapar, 1995; Pattanayak, 1998; Pattanayak, 1998) and genomic studies (Roychoudhary *et al.,* 2001; Basu *et*

al., 2003; Kumar and Reddy, 2003) argued that Austro Asiatic groups are the earliest settlers. The comprehensive Y chromosome study (Trivedi *et al.,* 2008) support the view that Dravidians are the original settlers (Buxton, 1925 and Sarkar, 1958) by not only refuting the settler status of Austrics, but also any possibility that Austro-Asiatic speakers could have dispersed from India is also eliminated.

The genomic studies encompass the view that Indian populace has unison in maternal pattern and in general genetic phylogeny is not in congruence with language phylogeny or caste affiliations. The Indian specific mtDNA lineages M, R and U are confirmed as Pleistocene lineages bearing coalescent time at around 50,000 ybp (Mountain *et al.,* 1995; Kivisild *et al.,* 1999, 2000; Quintana Murci *et al.,* 1999; Metspalu *et al.*, 2004; Palanichamy *et al.,* 2004; Sun *et al.*, 2006). 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. The above genomic revelations on Indian's maternal and paternal ancestry since Paleolithic times emancipate the question of archeo-linguistic history of India. Nevertheless, the genomic studies too have added some amount of confusion by insensitively attesting language correspondences to genetic inferences. The current genomic expressions are obviously not found in favor of stratifying the peoples of India in terms of language at least for genomic interpretations as it project the view that the basic clustering of maternal or paternal lineages shows a common trend that they are not clustered around particular languages. 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 (Kivisild *et al.,* 2003; Metspalu *et al.,* 2004; Sun *et al.,* 2006; Torroni *et al.,* 2006). Trivedi *et al.,* (2008) 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.

The underpinning of this presumption is quite potential to look forward the Paleolithic communication traits. The present language phylogeny of India suggests a Neolithic punctuation and it clearly suggests assimilation of various Paleolithic residual remnants. Linguistically, locating Paleolithic traits within the established language families of India is not yet known; while Genomic disclosures on ancient lineages are no longer unconvinced. Without discrediting the view that there is fundamental incomparability between the results of diverse disciplines, one could utilize the genomic results for a better understanding of ethno linguistic prehistory of India. Genomic understanding of Indian population confirms that the present population is an extension of Pleistocene lineages instead of accommodating repeated Neolithic waves. Pleistocene continuity in genetic structure has not corresponded with Pleistocene communication systems. Then existed languages

are replaced or assimilated to the Holocene immigrant languages. The Andamanese languages remain as remnants of the Paleolithic language diversity.

The overall picture emerges as interface of genomic structure and language history of India is the replacement of Paleolithic languages/ communication systems. It suggests that Pleistocene genetic antiquity with Holocene expression made Indian diversity. In trying to determine the appropriate pigeon hole of the language diversity of India in terms Neolithic lineages may not reflect the past communication diversity of India since initial peopling. However at the outset, the emerging picture is about the Pleistocene founder population and their languages. No trace is available on such substratum except the Andamanese family of languages. The available genomic expressions indicate Pleistocene genetic lineages within the contemporary population. The present linguistic obsession with existing models when addressing the issue of prehistory of India is detrimental to the understanding of pre-history of Indian communication systems. There is gap of information on Pleistocene linguistic lineages/elements. The reconstruction of proto language obviously has inherent limitations to take us to Pleistocene.

The mtDNA lineages are important for understanding human population movements and the language phylogeny across India. The genetic relationships between populations may offer insights helpful in the reconstruction of their putative past linguistic affiliation. The Neolithic lineages are putatively associated with the dispersal of various immigrant language speakers through Indian sub-continent. As lineages M, N, and R reach a combined frequency of 92%, Indian peoples with Paleolithic affinity must have formed the main sources of ancestry (Tripathy *et al.,* 2008; Maji *et al.,* 2008). The presence of paleolithic genetic signatures indicates a substratum which has persisted despite language shift and sociopolitical change. Despite linguistic discontinuity, the endurance of genetic signatures provides some continuity and may be viewed as a response to external events which are driving social, cultural and linguistic change. This paper draws on a corpus of groups which have been dissimilarly affected by the processes of change. The presence of paleolithic genetic signatures indicates a substratum of which has persisted despite language shift and sociopolitical change. Regardless of linguistic discontinuity, endurance of genetic signatures provides some continuity and may be viewed as a response to external events which are driving social, cultural and linguistic change. This study examines whether Language divergence can be equated with genetic divergence with respect to Indian context. For assessing possible scenarios, we made reassessment of some established results.

mtDNA structure of Austro-Asiatic samples does not coincide with proposed linguistic divisions at micro level. Haplogroup distributions and genetic distances seem to show genetic continuity among the Santhal, Kharia and Munda; the second by Ho, Bareng Paroja, Bhumij and Rajbanshi, Dhurwa, Ollaro Gadaba, Pareng Gadaba, Bado Gadaba and Konda Paroja, Oraon, Riang and Khasi, Lanjia Saora and Khond, Nicobarese, Chenchu and Kamma (Kumar, 2003). But these groups are not exactly showing linguistic continuity as normally expected. Among them

Santal, Kharia and Munda form a sub-cluster. Santal and Munda are linguistically representing two Kherwarian sub-clusters of eastern Munda and Kharia is closer to Juang representing Southern Munda. Proto Munda features make these three related, yet Santali is apparently closer to Munda than their genetically closed Kharia. Ho, Bareng Paroja, Bhumij and Rajbhanshi form another genetic sub cluster. In this case, Ho is genetically closer to Bareng paroja than with Bhumij. Linguistic affiliation of this cluster reveals Ho and Bhumij are part of kherwarian sub cluster and Bareng paroja is part of southern Munda representing Gorum which is closer to Saora. In addition, most of the subgroups of paroja are regarded as Dravidian groups. Linguistically Parji is the tongue of all Parojas. The LSI (Linguistic Survey of India) listed Parji as a dialect of Gondi and 1961 census accommodated it as a mother tongue of Dhurwa, then it was notified as a sub group of Gond. Later Burrow and Bhattacharya (1970) identified it as a separate language and grouped under central Dravidian. The POI (People of India) displayed Parjas as parji speeking community. In this cluster, language affinity is not exactly corresponded to genetic distance. Rajbahanshi has completely undergone language replacement and they are presently representing as Indo-Aryan. Analysis of Gadaba sub cluster comprising of Ollaro Gadaba, Pareng Gadaba, BadoGadaba and Khonda Paroja show that they are not linguistically related. Linguistic Survey of India has treated Gadaba both as Dravidian and Austro-Asiatic. 1961 census upto 1991 recognized it as an Austro-Asiatic language. The POI found three language groups viz; Gadaba speakers of Andhra Pradesh, Bhatri speakers of Madhya Pradesh and Gutob speakers of Orissa. Gadaba, ethnologically one ethnic group but linguistically they belong to three families, the Gutob Gadaba of Orissa is of Munda, Bhatri Gadaba of Madhya Pradesh is of Indo-Aryan (influenced by Halbi) and Dravidian Gadaba of Andhra Pradesh. There are two dialects of Dravidian Gadaba recognized. Bhattacharya (1951) claimed, Ollari of Koraput as an independent Dravidian language. Konekar Gadaba of Salur in Andhra Pradesh is the other dialect established by Rao (1980). 1961 census mentioned Koraputi as a dialect of Parji. Lanjia Saora and Khond form one cluster. The LSI reported language of Khond as Khond / Khand / Kui and 1961 census included them as Khond/Kondh and having Kui and Kuvi as their mother tongues. The POI has identified Kondh in Andhra Pradesh and Orissa as Kuvi speakers. The POI also made distinction between two subgroups of Kondh found in Orissa, Dongria kondh and Sita kondh. Linguistically, Dongria kondh speak Kuvi and Sita kondh speak Kui. As per the POI, Jatapu of Andhra Pradesh and Kondh groups viz, Kondh, Dongriya Kondh and Kuvi Kandh of Orissa (Kuvi Kandh is not separately declared as a scheduled tribe) are Kuvi speakers. Besides Sita Kondh, scheduled tribes like Dal and Kandha Gauda of Orissa too are Kui speakers, as the POI noted. In fact, apart from the above mentioned Kuvi and Kui languages; three more languages have been identified among them. It means, Kondh being a single ethnic group speaks five languages viz; Kui, Kuvi, Pengo, Manda and Indi Awe (Ramakrishna Reddy has identified Indi Awe as a language, 2003). All these languages are Dravidian though due to intimate contact with Indo Aryan and Munda languages various amount of non-Dravidian features

diffused in it. Saora is treated both as Austro-Asiatic and Dravidian. Khasi and Riang genetically closer but linguistically belongs to Monkhmer of Austro-Asiatic and Tibeto-Burman respectively. Oraon is another diverged group in this cluster.

Kurukh is one of the major Dravidian languages spread in various states. The regional languages have influenced this language, though the general assumption is that this language is devoid of dialects. According to LSI, Kurukh language is known under various names in different localities such as Oraon, Kol, Kisan, Mahar, Kora, Dhangari, etc. 1961 census recorded the following dialect groups, Kurukh/Oraon: *Adibhasha–Kurukh/Oraon, Adibhasha Nagesia, Dhangri, Dhanwari, Kisan Kurukh/Oraon, Lohari-Kurukh/Oraon, Malhar, Nagesia.* The POI has reported them as Kurukh speaking community except in Maharashtra, where they have been recognized as Sadri speaking group.

The above analysis reveals some trends in genetic and linguistic discontinuity. But our hypothesis to check the premise that language divergence can be equated with genetic divergence in Indian context has proved without correspondences. Adopting new languages either as part of convergence, subordination or replacement is an ongoing process in India. The above described groups attest genetic affinity but not linguistic continuity. Chaube *et al.,* (2008) proved that language shift has virtually no influence on the genetic makeup of the respective groups which we too support. Our study suggests genetic and linguistic divergence doesn't simultaneously appear. For instance, Santal and Kharia are genetically closer but linguistically less close. This scenario attests that genetic divergence and linguistic divergence is not simultaneous.

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. For establishing ethnic and language correlations, ethnolinguistically targeted DNA sampling is required. Regarding the correlation between the genetic structure and linguistic orientation, Indian genomic studies maintain a dual view. Some studies subscribe the view that genetic structure and language orientation is coterminous and others attest virtually no influence of language shift in genetic structure. Both the views are contextually significant as they are substantiated with the results of a particular study. 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 here, substantiate such assumption is the subject of enquiry.

HAPLOGROUP SHARING AND DIVERGENCE PATTERN

The above studied groups are clearly showing signs of divergence primarily in terms of presence of M2 and absence of M2.

Table 1: Haplogroup structure based on present linguistic status of 24 communities

Communities	Present language status	Haplogroups
Korku	Austro Asiatic North Munda	M2 M4 M5 M6 M18 M30 M33 M38 M42
Munda	Austro Asiatic North-Eastern Kherwarian	M2 M4 M5 M6 M18 M25 M31 M33 M35 M39 M40 M42
Betta Kuruba	South Dravidian	M2 M4 M5 M35
Madia	Central Dravidian	M2 M3 M4 M6 M25 M35 M39
Hillkolam	Central Dravidian	M2 M3 M4 M5 M6 M18
Malpaharia	North Dravidian/ Indo Aryan	M2 M3 M4 M5 M6 M18 M33 M39 M41
Jenukuruba	South Dravidian	M3 M8_C_Z M9 M25 M36
Kamar	Dravidian unclassified/ Indo-Aryan	M2 M3 M4 M5 M6 M12_G M36 M39 M40
Dungri Bhil	Indo-Aryan	M2 M3 M4 M5 M6 D M18 M30 M33
Kathodi	Indo-Aryan	M2 M3 M4 M5 M30 M35 M39
Matakur	Indo-Aryan	M2 M4 M30
Katakur	Indo-Aryan	M2 M4 M5 M12_G M39
Sonal Kachari	Indo-Aryan	M2 M3 M4 M6 D M8_C_Z M9 M18 M49
Paudi Bhuiya	Indo-Aryan	M2 M3 M4 M5 M6 M12_G M18 M25 M31 M39 M40
Katkari	Indo-Aryan	M2 M3 M4 M12_G M39
Andh	Unclassified (Indo-Aryan)	M2 M3 M4 M5 M6 M30 M35 M39
Nihal	Language isolate	M2 M4 M5 M6 M12_G M18 M25 M30 M33 M35 M37
Sherdukpan	Tibeto-Burman (Bhotia group)	M3 M4 D
Toto	TB (Himalayan group)	M4 M6 D M18 M33 M35
Gallong	TB (NEFA group)	M3 M4 M6 D M8_C_Z M9 M12_G
Dirang Monpa	TB (Bhotia group)	M3 M4 M6 D M8_C_Z M9 M12_G M28 M49
Wancho	TB (Naga group)	M4 M6 D M8_C_Z M9
Lachungpa	TB (Naga group)	M6 D M8_C_Z
Lepcha	TB Central group	M6 D M8_C_Z M9 M12_G M18 M33

Out of 24 studied groups, 15 groups are carrying the Paleolithic baggage of M2. Remaining Nine groups are not having M2 signatures which suggest two trends, either M2 signatures are later overwhelmed by other 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 out of 24 communities share M4, and attests to the pattern of sharing with M2 M3 carriers or sharing with M2 and M3 carriers separately or M4 alone.

M2 M3 M4 – Dungri Bhil, Kathodi, Katkari, Andh, Kamar, Sonal kachari and Paudi Bhuiya (Presently Indo Aryan), Hillkolam, Malpaharia and Madia (presently Dravidian)

M2 M4 – Matakur, Katakur, Nihal of present Indo Aryan and Bettakuruba of Dravidian, and Korku and Munda of Austro-Asiatic share this pattern

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

M4 – Toto, Wancho (Tibeto-Burman 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 trends of population divergence. Forthcoming analysis of distribution of each haplogroup and detection of common distribution obviously show the divergence pattern clearly which in turn suggests the language discontinuity.

Distribution of M5

M2 M3 M4 M5 (Ancestral population of Dungri Bhil, Kathodi, Andh, Kamar, Paudi Bhuiya, Hillkolam and Malpaharia)

M2 M4 M5 (Ancestral population of Korku, Munda,*Bettakuruba, Katakur and Naha

Distribution of M6

M2 M3 M4 M5 M6 (Ancestral population of Dungri Bhil, Andh, kamar, Paudi Bhuiya, Hillkolam and Malpahariya)

M2 M3 M4 M6 (Ancestral population of Madia and Sonal Kachari)

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 Dungri Bhil)

M2 M3 M4 M6 D (Ancestral population of Sonal Kachari)

M3 M4 D (Ancestral population of Shedrkpan)*

M3 M4 M6 D (Ancestral population of Gallong and Dirang Monpa)

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 Sonal Kachari)

M3 M4 M6 D M8_C_Z (Ancestral population of Gallong and Dirang Monpa)

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 Sonal Kachari)

M3 M4 M6 D M8_C_Z M9 (Ancestral population of Gallong and Dirang Monpa)

M3 M8_C_Z M9 (Ancestral population of Jenukuruba)

M4 M6 D M8_C_Z M9 (Ancestarl population of 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 Paudi Bhuiya)

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 M4 M6 D M8_C_Z M9 M12_G (Ancestral population of Gallong and Dirang Monpa)

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 Dungri Bhil)

M2 M3 M4 M5 M6 M12_G M18 (Ancestral population of Paudi Bhuiya)

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 Nihal)

M4 M6D M18 (Ancestral population of Toto)

M2 M3 M4 M6 D M8_C_Z M9 M18 (Ancestral population of Sonal Kachari)

M6 D M8_C_Z M9 M12_G M18 (Ancestral population of Lepcha)

Distribution of M25

M2 M3 M4 M5 M6 M12_G M18 M25 (Ancestral population of Paudi Bhuiya)

M2 M3 M4 M6 M25 (Ancestral population of Madia)

M2 M4 M5 M6 M12_G M18 M25 (Ancestral population of Nahal)

M2 M4 M5 M6 M18 M25 (Ancestral population of Munda)

M3 M4M6DM8_C_ZM9M12_G M25 (Ancestral population of Dirang Monpa)

M3 M8_C_Z M9 M25 (Ancestral population of Jenukuruba)

Distribution of M30

M2 M3 M4 M5 M30 (Ancestral population of Kathodi)

M2 M3 M4 M5 M6 D M18 M30 (Ancestral population of Dungri Bhil)

M2 M3 M4 M5 M6 M30 (Ancestral population of Andh)

M2 M4 M5 M6 M18 M30 (Ancestral population of Korku)

M2 M4 M30 (Matakur)

M2 M4 M5 M6 M12_G M18 M25 M30 (Ancestral population of Nahal)

Distribution of M31

M2 M3 M4 M5 M6 M12_G M18 M25 M31 (Ancestral population of Paudi Bhuiya)

M2 M4 M5 M6 M18 M25 M31 (Ancestral population of Munda)

Distribution of M33

M2 M3 M4 M5 M6 D M18 M30 M33 (Ancestral population of Dungri Bhil)

M2 M3 M4 M5 M6 M18 M33 (Ancestral population of Malpahariya)

M2 M4 M5 M6 M18 M30 M33 (Ancestral population of Korku)

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

M2 M4 M5 M6 M12_G M18 M25 M30 M33 (Ancestral population of Nihal)

M4 M6D M18 M33 (Ancestral population of Toto)

M6 D M8_C_Z M9 M12_G M18 M33 (Ancestral population of Lepcha)

74

Distribution of M35

M2 M3 M4 M6 M25 M35 (Ancestral population of Madia)

M2 M3 M4 M5 M6 M30 M35 (Ancestral population of Andh)

M2 M3 M4 M5 M30 M35 (Ancestral population of Kathodi)

M2 M4 M5 M6 M12_G M18 M25 M30 M33 M35 (Ancestral population of Nahal)

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

M4 M6D M18 M33 M35 (Ancestarl population of Toto)

M2 M4 M5 M35 (Ancestral population of Bettakuruba)

Distribution of M36

M3 M8_C_Z M9 M25 M36 (Ancestral population of Jenukuruba)

M2 M3 M4 M5 M6 M12_G M36 (Ancestral population of kamar)

Distribution of M37

M2 M4 M5 M6 M12_G M18 M25 M30 M33 M35 M37 (Ancestral population of Nahal)

Distribution of M38

M2 M3 M4 M12_G M38 (Ancestral population of Katkari)

M2 M4 M5 M6 M18 M30 M33 M38 (Ancestral population of Korku)

Distribution of M39

M2 M3 M4 M6 M25 M35 M39 (Ancestarl population of Madia)

M2 M3 M4 M5 M6 M30 M35 M39 (Ancestral population of Andh)

M2 M4 M5 M12_G M39 (Ancestarl population of Katakur)

M2 M3 M4 M5 M30 M35 M39 (Ancestral population of Kathodi)

M2 M4 M5 M6 M18 M25 M31 M33 M35 M39 (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 Paudi Bhuiya)

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 (Ancestral population of Paudi Bhuiya)

M2 M3 M4 M5 M6 M12_G M36 M39 M40 (Ancestral population of 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 (Ancestral population of Malpahariya)

Distribution of M42

M2 M4 M5 M6 M18 M30 M33 M38 M42 (Ancestarl population of Korku)

M2 M4 M5 M6 M18 M25 M31 M33 M35 M39 M40 M42 (Ancestral population of Munda)

Distribution of M49

M3 M4M6DM8_C_ZM9M12_G M25 M49 (Ancestral population of Dirang Monpa)

M2 M3 M4 M6 D M8_C_Z M9 M18 M49 (Ancestral population of Sonal Kachari)

The above analysis of haplogroup distribution clearly attests the ancestral groups and their divergence. This analysis has shown the pattern that, each ancestral group after divergence again undergo admixture and finally depict the defined present day population structure. It means genetic divergence of ancestral population by genetic drift is eventually reduced by gene flow and its impact depends on the relationship of populations. 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 the heritage haplogroup of their ancestral population is not derived during Holocene influx, thus it defines linguistic discontinuity. It also suggests the genetic discontinuity in terms of divergence. To transform genetic variation into divergence times, common haplogroup pattern has to be understood in each case. This study exposes the appropriate divergence time and later admixtures by showing the haplogroup structure. Present study vehemently argues that haplogroup distribution analysis can show prehistory of 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. The Neolithic influx has replaced the Paleolithic linguistic lineages without correspondingly replacing genetic structure (Kivisild, 2003). The Neolithic admixture has produced genetically united population into linguistically diverse population as the results clearly indicate; linguistic barrier attested genetic differences are comparatively lower among the studied groups.

DISCUSSION

Assumptions on languages led the genomic interpretations of Indian populations.

Keeping the view of language and gene co evolution – The genomic studies tried to interpret results in corroboration with language families.

i)   Without taking the account of genetic heritage and linguistic heritage are not of the same nature.

ii) Without giving importance to the concept of India or even south Asia as a linguistic area.

iii) Possibility of language replacements was not given much importance.

The results so far have been not of much help to unravel gene – language connection and linguistic admixture pattern. It is mainly because of the impermanent behavior of languages.

The above analysis has undoubtedly proved that linguistic and genetic structures are not co-evolved among Indian population. Tamang *et al.,* (2012) has shown that India is inhabited by autochthonous as well as migrant populations. Among the autochthonous there are ancestral North Indian (ANI)', which shared genetic affinity with the populations of the Middle East, Central Asia and Europe and an 'ancestral South Indian', which has no relation with any population outside India (Reich *et al.,* 2009). Many deep rooted autochthonous haplogroups in India reveal deep rooted autochthonous haplogroups which in turn justify the antiquity of population across language families (Thangaraj *et al.,* 2003, 2006a, 2009; Chaubey *et al.,* 2008; Sharma *et al.,* 2012). Linguistically true to subscribe the view that Andamanese group represent the true ancestral south Indian. The presence of deep rooted autochthonous haplogroups suggests the admixture, which reflects the typological affinities of linguistic families in India. In case of linguistic affinity, Andamanese language family has not apparently established any connection with any of the four existing language groups of Africa today. The Pleistocene genetic antiquity and more of factor of isolation of the Andaman Negritos, together made their language, evolutionarily much significant in the discussion of origin and evolution of human languages in general and language diversity of Asia in particular. The first person pronoun signature of Anadamanese family further suggests that basically they are the representatives of an old stratum of nasal based pronoun paradigm (Sreenathan, 2015).

CONCLUSION

This study was carried out to know the patterns of linguistic and genetic continuity and discontinuity among the selected population groups in order to critically review the concept of language-gene co-evolution. It reveals that Indian tribal groups are maternally and culturally united but linguistically diverged. Language divergence and genetic divergence are not simultaneous among them. Paleolithic genetic continuity is not corresponded with the linguistic continuity. Despite of having

cultural continuity, language varies among them. It suggests the existing peripheral understanding on languages is not a dependable trait to establish the deep rooted prehistory of populations. Hence it demands that typological features and structural exclusiveness should be given emphasis in interpretation of language based genetic antiquity. This study suggests beyond the existing knowledge on language and family linkages, typological information should be given due consideration in interpreting genetic results based on language. Also, more trait based linguistic contribution would only help linguistic and genetic corroborative studies. A clear interdisciplinary framework needs to be invented to accommodate non genomic and genomic results in fixing the compatibility of interdisciplinary explanations of prehistory due to the antiquity of diffusion between groups and languages.

REFERENCES

Abbi A. Is Great Andamanese genealogically and typologically distinct from Onge and Jarawa. *Language Sciences.* 2009; 31: 791-812.

Abbi A. Endangered languages of the Andaman Islands (München Lincom Europa) (2006).

Bamshad M, Kivisild T, Watkins WS, Dixon ME, Ricker CE, Rao BB, *et al.* Genetic evidence on the origins of Indian caste populations. *Genome Res.* 2001; 11: 994-1004.

Barik SS, Sahani R, Prasad BV, Endicott P, Metspalu M, Sarkar BN, *et al.* Detailed mtDNA genotypes permit a reassessment of the settlement and population structure of the Andaman Islands. *Am. J. Phys. Anthropol.* 2008; 136: 19-27.

Barnabas S, Apte RV and Suresh CG. Ancestry and interrelationships of the Indians and their relationship with other world populations a study based on mitochondrial DNA polymorphisms. *Ann. Hum. Genet.* 1996; 60: 409-422.

Basu A, Mukherjee N, Roy S, Sengupta S, Banerjee S, Chakraborty M, *et al.* Ethnic India a genomic view, with special reference to peopling and structure. *Genome Res.* 2003; 13: 2277-2290.

Burrow T. Some Dravidian words in Sanskrit. *TPS*. 1945: 79–100.

Burrow T. Loan words in Sanskrit. *TPS*. 1946: 1–30.

Burrow T. Dravidian Studies vii: further Dravidian words in Sanskrit. BSOAS 1947a; 12:365-396.

Burrow T. The Sanskrit Language, London: Faber & Faber (1955).

Burrow T. The Sanskrit Language (2nd edition), London: Faber & Faber (1973).

Burrow T. The Proto-Indo-Aryans. JRAS. 1973; 1: 123-140.

Burrow T and Emeneau MB. A Dravidian Etymological Dictionary (2nd edition), Oxford: Clarendon Press. Oxford: Oxford University Press (1984).

Cavalli-Sforza LL, Menozzi P and Piazza A. The History and Geography of Human Genes (Princeton Univ. Press, Princeton) (1994).

Chandrasekar A, Kumar S, Sreenath J, Sarkar BN, Urade BP, Mallick S, *et al.* Updating phylogeny of mitochondrial DNA macrohaplogroup m in India dispersal of modern human in South Asian corridor. PLoS One. 2009;4:e7447.

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

Chaubey G, Karmin M, Metspalu E, Metspalu M, Selvi-Rani D, Singh VK, *et al.* Phylogeography of mtDNA haplogroup R7 in the Indian peninsula. *BMC Evol. Biol.* 2008; 8: 227.

Chaubey G, Metspalu M, Choi Y, Magi R, Romero IG, Soares P, *et al.* Population genetic structures in Indian Austroasiatic speakers the role of landscape barriers and sex-specific admixture. *Mol. Biol. Evol.* 2011; 28 : 1013-1024.

Chaubey G, Metspalu M, Monika K, Thangaraj K, Rootsi S, Parik J, *et al.* Language Shift by Indigenous Population: A Model Genetic Study in South Asia. *Int. J. Hum. Genet.* 2008; 8: 41-50.

Deshpande MM and PE Hook (eds) (1979). Aryan and non-Aryan in India, Ann Arbor: The University of Michigan Center for South and Southeast Asian Studies.

Emeneau MB. A Dravidian etymology of the Sanskrit proper name Nala," in University of California Publications in Classical Philology. 1943: 255-262.

Emeneau MB. Linguistic prehistory of India. PAPS. 1954; 98: 282-292.

Emeneau MB. India as a linguistic area Language. 1956; 32: 3-16.

Emeneau MB. The dialects of Old Indo-Aryan, in H. Birnbaum and J. Puhvel (eds), Ancient Indo-European Dialects, Berkeley and Los Angeles, CA: University of California, 1966: 123-138.

Fairservis WA. The Harappan Civilization and its Writing: A Model for the Decipherment of the Indus S*cript*, Delhi: Oxford & IBH (1992).

Hock HH. Substratum influence on (Rig Vedic) Sanskrit? Studies in the Linguistic Sciences. 1975; 5: 76-125.

Diamond J and Belwood P. Farmers and their languages. Science. 2003; 300: 597-603.

Kivisild T, Bamshad MJ, Kaldma K, Metspalu M, Metspalu E, Reidla M, *et al.* Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages. *Curr. Biol.* 1999; 9: 1331-1334.

Kivisild T, Rootsi S, Metspalu M, Mastana S, Kaldma K, Parik J, *et al.* 2003. The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. *Am. J. Hum. Genet.* 2003; 72: 313-332.

Kivisild T, Kaldma K, Metspalu M, Parik J, Papiha SS and Villems R. The place of the Indian mitochondrial DNA variants in the global network of maternal lineages and the peopling of the Old World, in Papiha *et al.* (eds.), 1999b; 135-152.

Kumar V, Reddy AN, Babu JP, Rao TN, Langstieh BT, Thangaraj K, *et al.* Y-chromosome evidence suggests a common paternal heritage of Austro-Asiatic populations. *BMC Evol. Biol.* 2007; 7: 47.

Kuiper FB. Proto-Munda words in Sanskrit. Verhandelingen der Kon. Nederl. Akademie van Wetenschappen, Afd. Letterkunds. 1948; 5: 3.

Kuiper FB. 1955 "Rigvedic loanwords," in O. Spies (ed.), *Studia Indologica: Festschrift für Willibald Kirfel* Bonn: Orientalistisches Seminar der Universität Bonn, pp. 137–85.

Kirfel B. A*ryans in the Rigveda*, Amsterdam and Atlanta, GA: Rodopi (1991).

Lahr M and Foley R. Multiple dispersals and modern human origins. *Evolutiona* y *Anthropology* 1994; 3: 4840.

Lévi S. Pré-aryen et pré-dravidien dans l'Inde. *Journal Asiatique.* 1923;203:1-57 (English translation by P. C. Bagchi in P. C. Bagchi, (1975) *Pre-Aryan and Pre- Dravidian in India*, Calcutta: Calcutta University Press.)

Macaulay V, Hill C, Achilli A, Rengo C, Clarke D, Meehan W, et al. Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science.* 2005; 308: 1034-1036.

Manoharan S. Linguistic peculiarities of Andamanese family of languages. *Indian Linguistics* 1986; 47: 25-32.

Manoharan S. A Descriptive and Comparative Study of the Andamanese Language. Anthropological Survey of India, Calcutta (1989).

Masica CP. Defining a Linguistic Area: South Asia, Chicago, IL: University of Chicago Press (1976).

Masica CP. Aryan and non-Aryan elements in north Indian agriculture, in M. M. Deshpande and P. E. Hook (eds), Aryan and non-Aryan in India, Ann Arbor: The University of Michigan Center for South and Southeast Asian Studies, 1979: 55-152.

Masica CP. The Indo-Aryan Languages, Cambridge: Cambridge University Press (1991).

McAlpin DW. Proto-Elamo-Dravidian: The Evidence and its Implications, Philadelphia, PA: The American Philosophical Society (1981).

Metspalu M, Kivisild T, Metspalu E, Parik J, Hudjashov G, Kaldma K, *et al.* 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, *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.

Mountain JL, Hebert JM, Bhattacharyya S, Underhill PA, Ottolenghi C, Gadgil M, *et al.* Demographic history of India and mtDNA-sequence diversity. *Am. J. Hum. Genet.* 1995; 56: 979-992.

Narang A, Jha P, Rawat V, Mukhopadhyay A, Dash D, Basu A, *et al.* Recent admixture in an Indian population of African ancestry. *Am. J. Hum. Genet.* 2011; 89: 111-120.

Nei M and Roychoudhury A. Genetic relationship and evolution of human races. *Evol. Biol.* 1982; 14: 1–59.

Palanichamy MG, Agrawal S, Yao YG, Kong QP, Sun C, Khan F, *et al.* Comment on reconstructing the origin of Andaman islanders. *Science.* 2006; 311: 470.

Parpola A. *Deciphering the Indus Script*, Cambridge University Press (1994).

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

Renfrew C. In *Archaeogenetics: DNA and the Population Prehistory of Europe*, eds. Renfrew, C. & Boyle K. (McDonald Institute, Cambridge, U.K.), 2000: 3-12.

Roychoudhury S, Roy S, Basu A, Banerjee R, Vishwanathan H, Usha Rani MV, *et al.* Genomic structures and population histories of linguistically distinct tribal groups of India. *Hum. Genet.* 2001; 109: 339-350.

Sahoo S, Singh A, Himabindu G, Banerjee J, Sitalaximi T, Gaikwad S, *et al.* A prehistory of Indian Y chromosomes evaluating demic diffusion scenarios. *Proc. Natl. Acad. Sci. USA.* 2006; 103: 843-848.

Sengupta S, Zhivotovsky LA, King R, Mehdi SQ, Edmonds CA, Chow CE, *et al.* 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.* 2006; 78: 202-221.

Sreenathan M and Bednarik RG. Traces of the ancients: Ethnographic vestiges of Pleistocene art, Rock Art Research. 2012; 29: 191-217.

Sreenathan M. Language plurality and functional heterogeneity among people of India in Identity, Cultural pluralism and State- South Asia in perspective, Das NK and Rao VR (eds) Macmillan and An.S.I. (2009).

Sreenathan M and Rao VR. Andamanese Mythical signatures linking Gondwana Mythology with Lauraesian cluster, Mother tongue. Journal of the Association for the study of language in prehistory. ASLIP, Harvard University (2008).

Sreenathan M, Rao VR and Bednarik R. Palaeolithic cognitive inheritance as aesthetic behaviour of the Jarawas of the Andaman Islands. *Anthropos.* 2008; 103: 367-392.

Sreenathan M. 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 (2015).

Southworth FC. Linguistic stratigraphy of north India, in FC. Southworth and ML. Apte (eds), *Contact and Convergence in South Asian Languages, IJDL* (special issue), 1974: 201-223.

Southworth FC. Linguistic Archaeology of South Asia, Routledge Curzon, London (2005).

Swadesh M. Lexico statistic dating of Prehistoric Ethnic Contacts proceedings of the American philosophical society (1952).

Swadesh M. Archaeological and Linguistic Chronology of Indo European languages. *American Anthropologist.* 1953; 55: 349-352.

Thangaraj K, Ramana GV and Singh L. Y-chromosome and mitochondrial DNA polymorphisms in Indian populations. *Electrophoresis.* 1999; 20: 1743-1747.

Thangaraj K, Singh L, Reddy AG, Rao VR, Sehgal SC, Underhill PA, *et al.* Genetic affinities of the Andaman Islanders, a vanishing human population. *Curr. Biol.* 2003; 13: 86-93.

Thangaraj K, Chaubey G, Kivisild T, Reddy AG, Singh VK, Rasalkar AA, *et al.* Reconstructing the origin of Andaman Islanders. *Science.* 2005; 308: 996.

Thangaraj K, Chaubey G, Singh VK, Vanniarajan A, Thanseem I, Reddy AG *et al.* In situ origin of deep rooting lineages of mitochondrial Macrohaplogroup 'M' in India. BMC Genomics. 2006a; 7: 151.

Thangaraj K, Chaubey G, Reddy AG, Singh VK and Singh L. Unique origin of Andaman Islanders insight from autosomal loci. *J. Hum. Genet.* 2006b; 51: 800–804.

Thangaraj K, Nandan A, Sharma V, Sharma VK, Eaaswarkhanth M, Patra PK, et al. Deep rooting in-situ expansion of mtDNA Haplogroup R8 in South Asia. *PLoS One.* 2009; 4: e6545.

Thangaraj K, Naidu BP, Crivellaro F, Tamang R, Upadhyay S, Sharma VK, *et al.* The influence of natural barriers in shaping the genetic structure of Maharashtra populations. *PLoS One.* 2010; 5: e15283.

Trivedi R, Sanghamitra S, Singh A, Hima Bindu G, Banerjee J, Tandon M, *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.

Underhill PA and Kivisild T. Use of Y-chromosome and mitochondrial DNA population structure in tracing human migrations. *Annu. Rev. Genet.* 2007; 41: 539-564.

Underhill PA, Myres NM, Rootsi S, Metspalu M, Zhivotovsky LA, King RJ, *et al.* Separating the post-Glacial co-ancestry of European and Asian Y chromosomes within haplogroup R1a. *Eur. J. Hum. Genet.* 2010; 18: 479-484.

Tripathy V, Nirmala A and Reddy BM. Trends in Molecular Anthropological Studies in India. *Int. J. Hum. Genet.* 2008; 8: 1-20.

Watkins WS, Bamshad M, Dixon ME, Bhaskara Rao B, Naidu JM, Reddy PG, *et al.* Multiple origins of the mtDNA 9-bp deletion in populations of South India. *Am. J. Phys. Anthropol.* 1999; 109: 147-158.

Witzel M. Tracing the Vedic dialects, in C. Caillat (ed.), Dialects dans les littératures Indo-Aryennes, Paris: Institut de Civilisation Indienne. 1989: 97-264.

Witzel M. Early Indian history: linguistic and textural parameters in G. Erdosy (ed.), The Indo-Aryans of Ancient South Asia: Language, Material Culture and Ethnicity, Berlin: Walter de Gruyter. 1995: 85-125.

Witzel M. Substrate languages in old Indo-Aryan. *EJVS.* 1999; 5: 1-67.