

Eastern Himalaya: The Cradle of Ethnogenesis

Language, genes and the peopling of Asia

Delivered by George van Driem

13 July 2014, Kathmandu

The Himalaya runs over 3600 kilometres from the Hazarahjat Highlands in the west to the Liangshan in the east. The Himalaya forms no natural watershed, and many of the rivers are of greater antiquity than the mountains themselves. The Kali Gandaki bisects the Himalaya into two halves of roughly equal length. The Eastern Himalaya is the half which runs eastward from the Dhaulagiri across the Himalaya, sub-Himalaya, Meghalaya, lower Brahmaputra basin and associated hills tracts, the eastern Tibetan plateau and Indo-Burmese borderlands into the Chinese provinces of Yunnan and Sichuan.

As a cradle of ethnogenesis, the Eastern Himalaya served both as staging area and principal thoroughfare in the peopling of Asia following the emergence of anatomically modern humans from Africa. New scientific insights from historical linguistics and population genetics enable us to reconstruct the founding dispersals of language families of Eastern Eurasia and Oceania which ultimately originated in the Eastern Himalaya. In presenting this new account, it is necessary first to dispel two antiquated scholarly ideas, one which still lives on in the popular imagination and another which survives in laggardly quarters of the linguistic community.

The Mongoloid myth

As a species, we have always been obsessed with how we look and appear to be similar or different from one another. The ancient Hindu caste system and apartheid in South Africa were just two of many systems based on our perceptions of caste, tribe and race. Even before the Portuguese made landfall in Japan in 1542, Europeans were trying to come to grips with the human phenotypical diversity which they observed in people whom they met on their voyages. Today we understand that in scientific terms there is actually no such thing as race. We are all members of one large human family. The relationship between genes, their phenotypical expression and pleiotropic interplay is inordinately complex. Our individual differences tend often to be larger than the differences between groups.

Long before the discovery of the molecular mechanisms underlying genetics, scholars resorted to superficial classifications in their attempts to understand human diversity. Classification was conducted on the basis of somatology, which involved crude observations about external appearance. On the basis of the descriptions in Dutch and Russian accounts of peoples in other parts of the world, the German scholar Christoph Meiners (1747-1810) set up a classification of races based on what he imagined were the racial prototypes of mankind. His cogitations were published posthumously in three volumes. The 'Mongoloid race' was designated by Meiners as one of the main races of mankind:

In physiognomy and physique the Mongol diverges as much from the usual form as does the Negro. If any nation merits being recognised as a racial prototype, then it should rightfully be the Mongol, who differs so markedly from all other Asian peoples in his physical and moral nature.

Meiners described the cruelty of the invading hordes led by Genghis Khan as inherent to the 'moral nature' of the Mongoloids, conveniently overlooking the historically well-documented cruelties of Western people. His classification gave rise to the Mongoloid myth. If the Mongols were the primordial tribe from which all peoples of the Mongoloid race descended, then it was logical to think that the homeland of all Mongoloids lay in Mongolia.

I have often been told by people in Nepal and northeastern India that their ancestors came from Mongolia. Some adorn their lorries, vans and motorcycles with captions like 'Mongol' or 'Mongolian'. When I ask them why, they tell me that they are members of the Mongoloid race or मंगोलजात *maṅgol jāti*, whose ancestors, as the name tells us, originated in Mongolia. I do not have the heart to tell them that the idea was dreamt up by a German scholar at the beginning of the 19th century, who was imaginatively trying to make sense of human diversity, although he had no specialist knowledge to do so.

People in the West suffer from the same obsolete ideas. A friend of mine from Abkhazia, who happens to be a renowned linguist, was travelling in the United States of America with a colleague from the Republic of Georgia. Their rented car was pulled over by a police officer. The heavily armed man in uniform demanded to see my friend's driving licence and asked, "Are you folks Arabs?" The policeman pronounced the word 'Arabs' with an American accent as *ay-rabz*.

Abkhazia and Georgia lie in the Caucasus, and my friend responded, "No, Sir, we are both Caucasians." This response somehow displeased the police officer, who asserted, "I am a Caucasian!" My friend coolly responded, "No, Sir, you are not a Caucasian, and you do not look particularly Caucasian. *We* are Caucasians". The exasperated policeman spluttered, "...but I am white!" My friend ended up having to explain where the Caucasus Mountains lay and who the Caucasians were. He did not bother to explain that the idea that Europeans were Caucasian originated with Meiners. Like the Mongoloid, the Caucasoid was another one of Meiner's racial prototypes.

Americans who apply for a driving licence, take a Scholastic Aptitude Test or fill in any number of other official forms are asked to specify their race. A person of European ancestry often checks a box saying that he or she is a 'Caucasian'. Some people are baffled by the choices of race on offer, which differ from one form to another. They are asked to decide whether they are 'coloured' or belong to some other 'race'. The topic of race is taboo in the US, but American society is riddled with antique modes of thinking about race and very much in denial about widely held racist assumptions. The US has no monopoly on such thinking, however.

The Sino-Tibetan myth

The Sino-Tibetan or Indo-Chinese myth likewise has its roots in the now defunct scholarly fashion of 'scientific' racism. Sino-Tibetan also owes its longevity to the fact that every age sees many scholars whose ignorance does not make them less prolific than their more knowledgeable colleagues. The Sino-Tibetan episode is all the more shameful because the Tibeto-Burman or Trans-Himalayan language family had already been recognised in 1823.

Julius von Klaproth identified the language family comprising Tibetan, Burmese, Chinese and all languages demonstrably relatable to these three. The Tibeto-Burman family which he had demonstrated was accepted not just on the Continent, but also in the British Isles (e.g. Hodgson 1857, Cust 1878, Forbes 1878, Houghton 1896).

Like Julius von Klaproth, Jean Jacques Huot in Paris and Max Müller in Oxford stressed that language and biological ancestry were two different things. Yet there were those who confounded language and race. In 1850, Heymann Steinthal wrote that language typology was a measure of the "instinctive self-awareness" of a language community. He claimed that: "Language differences reflect differences in the level of consciousness between different peoples." He qualified typological differences in language structure as "physiological".

Steinthal set up an evolutionary hierarchy of successive stages of language types, reflecting "the level of development of linguistic consciousness". He distinguished 12 levels from the most complex, represented by Sanskrit, to the most simple. He relegated Chinese and Thai to the lowest rung of the evolutionary ladder based on their 'monosyllabicity' and lack of inflection. Steinthal's language typology inspired scholars to argue that Chinese and Thai must be close relatives and that neither was close to Tibeto-Burman. Ostensibly, Chinese and Siamese mediated a rudimentary, less evolved way of thinking. In reality, Chinese was a defining member of Klaproth's Tibeto-Burman family, and Klaproth had already recognised that Thai belonged to another language family than Chinese.

In 1854, the French count Arthur de Gobineau published the four-volume *Essay on the Inequality of Human Races*, in which he argued for the inferiority or superiority of particular races based on the structure of their languages. To reconcile the technological advancement of Chinese civilisation with its low rung on the ladder of language evolution, Gobineau invented a distinction between so-called male and female races. As one might expect, the count imagined that 'male races' possessed a richer and more precise vocabulary than 'female races', whose languages were full of vague notions. To the count's mind, the Chinese 'race' was in some sense 'male' despite the inferior status which he imputed to its language.

In 1858, Ernest Renan, who would later become president of the Linguistic Society of Paris, wrote:

Is the Chinese language, with its inorganic and incomplete structure not the very image of the dryness of spirit and callousness of heart that characterises the Chinese race? ...Sufficient for the needs of daily life,

for describing manual skills, for a light literature of no sophistication, for a philosophy that is nothing more than the pretty but never elevated expression of mere common sense, the Chinese language excludes all philosophy, all science and all religion in the sense in which we understand these terms.

Steinthal's racist language typology caught on in Britain too. John Beames, who wrote the first grammar of Magar in 1870, was an adherent. For Beames, Chinese represented the most primitive stage of language development, but he promoted English and French to the highest rung of the evolutionary ladder, placing them even above Sanskrit. Beames introduced the term 'analytic', still in use amongst language typologists today, to describe English and French. His enhancements were approved by James Byrne, who in 1885 argued that "the causes which have determined the structure of language" lay in the varying "degrees of quickness of mental excitability possessed by different races of men".

Steinthal was German, but his ideas were popular in France and Britain. His thinking was strongly opposed by German linguists, since scholars following the tradition of Wilhelm von Humboldt rejected the racist paradigm. August Pott and Max Müller argued that the relationship between language structure and thought was subtle, intricate and not simplistic. Pott wrote a hefty point-by-point refutation of Gobineau's work, and the writings of the French count were largely forgotten in Germany. Yet, after the First World War, Gobineau's writings were rediscovered by Ludwig Schemann and Franz Hahne. Tragically, this time the count's cogitations were given a warm reception, and his theories were incorporated into the official ideology of Germany's National Socialist Party.

In the 19th century, racist linguistics took Chinese out of Klaproth's original Tibeto-Burman family and put Chinese into a separate branch together with Thai. The favoured family tree of the racist language typologists was Indo-Chinese, and in 1924 this phylogenetic model was renamed Sino-Tibetan. In 1938, Berkeley anthropologist Alfred Kroeber started the Sino-Tibetan Philology Project. His use of the new name Sino-Tibetan helped to deflect criticism against the Indo-Chinese model. Ironically, after the Cultural Revolution, Chinese scholars imported Sino-Tibetan from America and enshrined this family tree as linguistic orthodoxy in China. Today an increasing number of Chinese linguists have begun to feel uncomfortable with Sino-Tibetan, as they begin to discover the model's Sinophobic legacy as well as the fact that no evidence exists for this tree.

Since the 1970s, the Sino-Tibetan model has been defended from Berkeley by Jim Matisoff, who inherited the family tree from his mentor in the 1960s and never questioned it. Sino-Tibetan was challenged and refuted by various scholars, but Matisoff continued to act as the *Fidei Defensor*, assailing any scholar who questioned the tree. After years of resistance, Matisoff came to realise that the Sino-Tibetan model was wrong. Since his retirement, he publicly recanted on three occasions, acknowledging Sino-Tibetan to be a false tree. Today Matisoff goes in and out of denial, and in an attempt to save face several of his former students continue to defend Sino-Tibetan despite an inability to adduce evidence.

The history of linguistics is strewn with false 'Sino' theories that were founded upon methodologically flawed comparisons, bewilderment about the historical grammar of Chinese and inadequate knowledge of Trans-Himalayan languages: Sino-Tibetan (Przyluski 1924), Sino-Yeniseian (Schmidt 1926), Sino-Caucasian (Bouda 1950), Sino-Burman (Ramstedt 1957), Sino-Indo-European (Pulleyblank 1966), Sino-Himalayan (Bodman 1973), Sino-Austronesian (Sagart 1993), Sino-Kiranti (Starostin 1994), Sino-Mayan (Jones 1995) and Sino-Uralic (Gao 2008). None of these models is supported by sound evidence, and they all represent false language family trees.

The legacy of racist language typology misled many linguists for decades even though an informed view was readily available to any linguist who carefully read the history of the field and scrutinised the available evidence dispassionately. In 2004, the neutral geographical term Trans-Himalayan was introduced for Klapproth's Tibeto-Burman, which after 181 years still turned out to be the most well informed model of the language family.

The name Trans-Himalayan reflects the fact that the world's second most populous language family straddles the Himalayan range. Most speakers of Trans-Himalayan languages today live to the north and east of the Himalaya, but most of the over 300 different languages and three fourths of Trans-Himalayan subgroups are located to the south of the Himalayan divide.

Words of caution on language and genes

Numerous scholars since the early 19th century have stressed that language and biological ancestry were two different things. There were always others too, like Sir William Jones, who persisted in confusing language and race. Throughout time, people have been inclined to speak the language spoken by their parents, but the language which we happen to speak today may very well not be our parents' language. Since genes are invariably inherited by offspring from their biological parents, a probabilistic correlation may exist between language and genes in human populations, although this need not necessarily be so.

The past took a very long time, and there are many slices of the past. So a chronologically layered view of ethnolinguistic prehistory is essential. The famous EPAS1 gene which enables Tibetans to live healthy lives at high altitude without having to fabricate excessive amounts of haemoglobin is known to be shared exclusively with the extinct Denisovans, a Palaeolithic people who lived in the Altai mountains of Siberia. Like the Neanderthals, this extinct variety of human is not really entirely extinct because the Denisovans interbred with the ancestors of many existing populations, not just with the ancestors of the Tibetans. A small percentage of DNA is shared between Denisovans and other Asian populations and native Australians as well.

When an ancestral highland Asian population interbred with the Denisovans, these people did not yet speak a language related to Tibetan, and ethnolinguistically they were not yet Tibetan. That was long ago, and linguistically reconstructible prehistory by comparison relates to more recent slices of prehistory. Not only is the time depth accessible to historical linguistics shallower than the time depth accessible to human genetics, but the spread of language families also happens to be a more

recent phenomenon than the spread of our anatomically modern ancestors outside of Africa. Language families represent the maximal time depth accessible to historical linguistics because the relatedness of languages belonging to a recognised language family represents the limit of what linguists can empirically demonstrate.

Historical linguistics and human population genetics present two distinct windows on the past. Molecular genetic findings can shed light on ethnolinguistic prehistory and its unrecorded sociolinguistic dimensions. Correlations exist between chromosomal markers and language, but these relationships should not be confused with identity. The correlation of a particular genetic marker with the distribution of a certain language family must not be simplistically equated with populations speaking particular languages.

Moreover, other factors that must be taken into account include the potential skewing effects of natural selection, gene surfing, recurrent bottlenecks during range expansion and the sexually asymmetrical introgression of resident genes into incursive populations. Factors such as ancient population structure and possible ancient Y-chromosomal introgression also affect inferences and interpretations based on any single Y-chromosomal locus when attempting to reconstruct migrations and elucidate the geographical origins of populations.

Even with all these caveats in place, we must remain aware of all provisos built into our inferences and working hypotheses. Only then may we undertake to interpret ethnolinguistic phylogeography from a linguistically informed perspective.

Father Tongues

In the 1990s, population geneticists found that it was easier to find correlations between the language of a particular community and paternally inherited markers on the Y chromosome than between language and maternally inherited markers in the mitochondrial DNA of a speech community. This Father Tongue correlation was described by a Swiss-Italian team in 1997, even before the appearance of the first Y-chromosomal tree in 2000. Today we have an even higher resolution picture of the Y-chromosomal haplogroup tree and the world's paternal lineages.

Paternally inherited polymorphisms were inferred to be markers for linguistic dispersals, and correlations between Y-chromosomal markers and language could point towards male-biased linguistic intrusions. The Father Tongue correlation is ubiquitous but not universal. Its preponderance allows us to deduce that a mother teaching her children their father's tongue must have been a prevalent and recurrent pattern. It is reasonable to infer that some mechanisms of language change may be inherent to this pathway of transmission.

There are a number of reasons why we might expect this outcome. Initial human colonisation of any part of the planet must have involved both sexes in order for a population of progeny to establish itself. Once a population is in place, however, subsequent migrations could have been gender-biased. Male intruders could impose their language whilst availing themselves of the womenfolk already in place. Sometimes male intruders slaughtered resident males and their offspring, but

sometimes they formed an elite and consequently enjoyed preferential access to spouses, reared more offspring and propagated their genes.

By contrast, correlations between maternal lineages and linguistic phylogeography have proved underwhelming. Populations exist which form local exceptions to the Father Tongue correlation, such as the Hungarians and the Balti in northern Pakistan, but the aetiology of these cases is readily explicable. The correlations observed do not always make a precise fit, and correlation must not be confused with identity.

The Father Tongue correlation suggests that linguistic dispersals were, in most parts of the world, posterior to initial human colonisation and that many linguistic dispersals were predominantly male-biased intrusions. Our paternal ancestry only represents a very small segment of our ancestry, but emerging autosomal findings appear to corroborate the reconstructions presented here. These patterns are observed worldwide.

The spread of Niger-Congo languages closely patterns with Y-chromosomal haplogroups. The martial, male-biased historical spread of Han Chinese during the sinification of southern China, recounted in detail in the Chinese chronicles, is just as faithfully reflected in the genetic evidence. A common ancestry between native Americans and indigenous Altaians is based preponderantly on shared Y-chromosomal heritage and is not as well reflected in mitochondrial lineages. The saliency of Y-chromosomal haplogroups in tribal and caste populations in India contrasts with the comparatively featureless antiquity of the mitochondrial landscape. In Europe, the language isolate Basque is the sole surviving linguistic vestige of Palaeolithic European hunter-gatherers, whose predominant paternal lineage was haplogroup I. Even Basques have seen their original paternal heritage diluted by more recent Y-chromosomal lineages subsequently introduced into Europe, perhaps ultimately originating from the Western Himalaya.

The spread of various Y-chromosomal R subclades may be linked to the dispersal of Indo-European from an original homeland in the Pontic-Caspian steppe, but the unfolding story of these R lineages is complex. In an epoch anterior to the expansion of Indo-European from the Pontic Caspian, an older pre-Indo-European homeland could have lain in the Western Himalaya, as suggested by the presence of the ancestral clade R* in Indian populations.

The Y-chromosomal lineage L shows a diversity of subclades on the Iranian plateau and marks a patrilingual dispersal of Elamo-Dravidian from Bactria and Margiana. One of these haplogroup L subclades is likely to be correlated with the patrilingual spread of Dravidian from the Indus Valley into southern India. Haplogroup Q traces the paternal spread of the Greater Yenisseian linguistic phylum. Yet this exciting tale about the Western Himalaya will have to wait for another occasion to be told.

From the Eastern Himalaya to Lappland

The Eastern Himalaya served as the cradle of ethnogenesis for a number of major language families, the molecular tracers of which survive today as the paternal lineages N (M231) and O

(M175). These two linguistic phyla are Uralo-Siberian and East Asian. The geographical locus of the ancestral haplogroup NO (M214) lay in the Eastern Himalaya. After the two Y-chromosomal lineages N and O split up between 30,000 and 20,000 years ago, the spatial dynamics of the two haplogroups diverged greatly. The ancient populations bearing haplogroups N and O underwent expansions between 18,000 and 12,000 years ago.

The bearers of haplogroup N set out for East Asia just after the Last Glacial Maximum and then moved north in a grand counterclockwise sweep, braving ice and tundra and gradually migrating across northern Eurasia as far west as Lapland. Y-chromosomal haplogroup N marks the paternal spread of Uralo-Siberian, comprising communities speaking Uralic, Yukagir, Eskimo-Aleut, Nivkh and Chukotko-Kamchatkan languages.

The absence of haplogroup N in the Americas and its prevalence throughout Siberia allow us to infer that the paternal lineage N spread northward after the paternal founder lineages had already established themselves in the Americas. The Greater Yeniseian haplogroup Q must have expanded across Siberia and colonised the Americas by way of Beringia, where it became the predominant paternal lineage, before Y-chromosomal N lineages replaced it in the sparsely populated north.

The N lineages differentiated into N* (M231), N1 (M128), N2 (P43) and N3 (Tat). The most prevalent haplogroup N3 is widespread throughout the Uralo-Siberian area, spreading as far west as Scandinavia. Yet the ancestral haplogroup N* is still found in the highest frequency at the eastern end of the Eastern Himalaya, i.e. in northern Burma, Yunnan and Sichuan. Haplogroup N1 is particularly frequent in the Altai region and to a lesser extent in Manchuria, and N2 shows an especially high frequency on both the Yamal and Tamyр peninsulas in northern Siberia.

The East Asian linguistic phylum

Julius von Klaproth was able to distinguish the contours of many of the known Asian language families. Five families form part of the East Asian linguistic phylum: Trans-Himalayan, Hmong-Mien, Krađai, Austronesian and Austroasiatic. Later generations of linguists began to discern possible long-distance relationships between the recognised families. In 1901, Gustave Schlegel argued that Krađai was related to Austronesian. Schlegel's theory was taken up by Paul Benedict in 1975, but Benedict's 'Austro-Thai' was no more than an ingredient in his misconceived 'Japanese-Austro-Tai' theory.

In 2005, Weera Ostapirat became the first to present methodologically sound linguistic evidence that Krađai and Austronesian formed coordinate branches of a single Austro-Tai family. Ostapirat envisages an ancient migration from what today is southern China across the Taiwan Strait to Formosa, where the Austronesian language family established itself. The Krađai proto-language remained behind on the mainland. Much later, the Formosan exodus set in motion the spread of Malayo-Polynesian throughout the Philippines, the Malay peninsula, the Indonesian archipelago, Madagascar and Oceania. By uniting Austronesian and Krađai in an Austro-Tai family, Ostapirat has effectively reduced the number of East Asian language families from five to four.

Since the beginning of the 20th century, historical linguists have been attempting to unite the East Asian language families on purely linguistic grounds. In 1906, Wilhelm Schmidt proposed an 'Austric' macrofamily, uniting Austroasiatic and Austronesian. In 2005, Lawrence Reid envisaged an even larger macrofamily, proposing that Austric "may eventually need to be abandoned in favour of a wider language family which can be shown to include both Austronesian and Austroasiatic languages but not necessarily as sisters of a common ancestor".

August Conrady in 1916 and Kurt Wulff in 1934 each proposed a superfamily consisting of Austroasiatic, Austronesian, Kra-dai and Tibeto-Burman. Subsequently, Robert Blust in 1996 and Ilia Peiros in 1998 proposed an 'Austric' superfamily comprising Austroasiatic, Austronesian, Kra-dai and Hmong-Mien. In 2001, a year before his death, Stanley Starosta proposed the East Asian linguistic phylum encompassing Kra-dai, Austronesian, Tibeto-Burman, Hmong-Mien and Austroasiatic. Starosta's evidence was meagre, yet compelling in being primarily morphological in nature. The ancient morphological processes shared by the families of this phylum were an agentive prefix *<m->, a patient suffix *<-n>, an instrumental prefix <s-> and a perfective prefix *<n->. The East Asian word was ostensibly disyllabic and exhibited the canonical structure CVCVC.

As a theory of linguistic relationship, Starosta's East Asian theory lies on the horizon of what might be empirically demonstrable in historical linguistics. This hypothesis will remain our best linguistically informed conjecture until better linguistic evidence can be accrued to support or overturn the model.

Eastern Himalayan homeland

The East Asian linguistic phylum consists of four language families: Trans-Himalayan, Hmong-Mien, Austroasiatic and Austro-Tai. The populations speaking these languages today are not just characterised by a preponderance of the Y-chromosomal lineage O (M175). Language communities of the four families are each characterised by a particular subclade of haplogroup O, suggesting a paternal spread of these language families and a probable time depth for the East Asian linguistic phylum that is coeval with the antiquity of haplogroup O itself. As temperature and humidity increased after the Last Glacial Maximum, haplogroup O split up into the subclades O1 (MSY2.2), O2 (M268) and O3 (M122).

The three subclades can be putatively assigned to three geographical loci along an east-west axis without any claim to geographical precision. Whereas haplogroup O1 moved to the drainage of the Pearl River and its tributaries, the bearers of haplogroup O2 moved to southern Yunnan, whilst bearers of haplogroup O3 remained in the Eastern Himalaya. The O2 clade split into O2a (M95) and O2b (M176). Asian rice may have first been domesticated roughly in the area hypothetically imputed to O2 south of the central Yangtze.

The bearers of the subclade O2a became the ancestors of the Austroasiatics, who spread initially to the Salween drainage in northeastern Burma, to northern Thailand and to western Laos. In

time, the Austroasiatics would spread as far as the Mekong delta, the Malay peninsula and the Nicobars. Later, early Austroasiatics would introduce both their language and their paternal lineage to indigenous peoples of eastern India, whose descendants are today's Munda language communities.

Meanwhile, the bearers of the fraternal subclade O2b spread eastward, where they introduced rice agriculture to areas downstream south of the Yangtze. The bearers of the O2b haplogroup continued to sow seed as they continued to move ever further eastward, but they left no linguistic traces. This paternal lineage moved as far as the Korean peninsula and represents the second major wave of peopling attested in the Japanese genome. Yet the Japanese speak a language of the Altaic linguistic phylum, and the peopling of Japan is a distinct episode of prehistory.

At the dawn of the Holocene in the Eastern Himalaya, haplogroup O3 gave rise to the ancestral Trans-Himalayan paternal lineage O3a3c (M134) and the original Hmong-Mien paternal lineage O3a3b (M7). The bearers of haplogroup O3a3c stayed behind in the Eastern Himalaya, whilst bearers of the O3a3b lineage migrated east to settle in areas south of the Yangtze. On their way, the early Hmong-Mien encountered the ancient Austroasiatics, from whom they adopted rice agriculture.

The interaction between ancient Austroasiatics and the early Hmong-Mien not only involved the sharing of rice agriculture technology, but also left high frequencies of haplogroup O2a in today's Hmong-Mien and haplogroup O3a3b in today's Austroasiatic populations. The Austroasiatic paternal contribution to Hmong-Mien populations was modest, but the Hmong-Mien paternal contribution to Austroasiatic populations in Southeast Asia was significant. However, the incidence of haplogroup O3a3b in Austroasiatic communities of the Subcontinent is undetectably low. Subsequently, the Hmong-Mien continued to move eastward, as did bearers of haplogroup O2b.

Even further east, the O1 (MSY2.2) paternal lineage gave rise to the O1a (M119) subclade, which moved from the Pearl River to the Min river drainage in the Fujian hill tracts and then across the Taiwan Strait. Formosa consequently became the homeland of the Austronesians. The Malayo-Polynesian expansion via the Philippines into insular Southeast Asia must have entailed the introduction of Austronesian languages by bearers of haplogroup O1a to resident communities, whose original Austroasiatic paternal haplogroup O2a alongside other older paternal lineages would remain dominant even after linguistic assimilation. Similarly, Malagasy is an Austronesian language, but the Malagasy people trace their biological ancestries equally to Borneo and the African mainland.

Back in the Eastern Himalaya, the paternal spread of Trans-Himalayan is preserved in the distribution of Y-chromosomal haplogroup O3a3c (M134). The centre of phylogenetic diversity of the Trans-Himalayan language family is rooted squarely in the Eastern Himalaya, with outliers trailing off towards the loess plains of the Yellow River basin in the northeast. Initially, Trans-Himalayan expanded through Sichuan and Yunnan, north and northwest across the Tibetan

plateau, westward across the Himalaya and southward into the Indo-Burmese borderlands.

On the Brahmaputran plain, early Tibeto-Burmans encountered Austroasiatics, who had preceded them. The ancestral Trans-Himalayan paternal lineage O3a3c also spread from the Eastern Himalaya in a northeastern direction across East Asia to the North China plain. Subsequently, at a shallower time depth, the Tibeto-Burman paternal lineage O3a3c spread south from the Yellow River basin into southern China, beginning with the Han expansion during the Qin dynasty in the third century BC. The Trans-Himalayan paternal lineage O3a3c is intrusively present in the Korean peninsula and beyond, although the Evenki and other Uralo-Siberian populations predominantly retain the paternal lineage N.

The Eastern Himalaya furnished the ultimate cradle for the ethnogenesis of the various Uralo-Siberian and East Asian language families. Language and genes tell us what we might also have deduced from basic facts of geography. In the hoary past, when our ancestors emerged from Africa on their way to East Asia, Southeast Asia, Oceania, Siberia, the Americas and even Lapland, many of them first passed through the Eastern Himalaya and crossed the Brahmaputra.

This lecture was organised in collaboration with Himal Southasian.