3. From the Dhaulagiri to Lappland, the Americas and Oceania

GEORGE VAN DRIEM

Introduction

The Indian subcontinent was the central staging area for our ancestors when anatomically modern humans emerged from Africa and entered Eurasia. Molecular genetic findings tell us that our ancestors emerged in waves (Rasmussen *et al.*, 2011). 75,000–62,000 years ago, the First Wave spilled out of Africa and followed a littoral route eastward across the Indian subcontinent and Southeast Asia.

The colonisation of Australia took place some 50,000 years ago and involved a major sea crossing from Asia. There is genomic evidence for some secondary gene flow between New Guinea and Australia some 8,000 years ago at the time of the Sahul land bridge, but this exchange pertains to a much later epoch. How the Australians reached the great continent down under remains a mystery. Were the initial groups of settlers marooned after having been haplessly swept out to sea, or were the first Australians intrepid mariners who essentially abandoned seafaring after having made landfall? Similarly, how did the ancestral Andamanese reach the Andaman Islands?

For a great many ancestral groups of modern peoples, however, not the littoral route but the Himalayan region played a pivotal role as a principal thoroughfare. The mid hills, the foothills and the Terai of the Himalayan corridor served as a conduit for not just one, but for various chronological layers of prehistory. 38,000 to 25,000 years ago, the Second Wave spilled out of Africa through the Levant. A small group branched off in Asia Minor and ventured into Europe but the main body of the wave swept across South Asia and headed to East Asia. Other than those headed for Europe, these migrating peoples interbred with populations of the First Wave whom they met on their way. It is an inescapable consequence of geography that when our anatomically modern ancestors emerged from

Africa and took this inland route on their way to East Asia, Southeast Asia, Oceania, Siberia, the Americas and even Lappland, many of these ancestors must at one point have passed through the Himalayan region and, on their way east, must have crossed the Brahmaputra.

In this long stretch of prehistory, the focus of this paper lies on just a subset of early Holocene episodes which led to the ethnolinguistic phylogeography which we observe in eastern Eurasia and Oceania today. The Eastern Himalaya can be identified as a cradle of ethnogenesis and a principal thoroughfare in the course of population prehistory. Geographically, the Eastern Himalaya can be demarcated as a well-defined area.

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. In fact, the Kali Gandaki, which runs right past the Dhaulagiri (8167m), is the deepest river valley on the face of the planet. This grand invagination at the very centre of the range 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.

New linguistics and genetic findings enable us to reconstruct the founding dispersals of major language families in Asia and Oceania. The Eastern Himalaya appears to have served as a cradle of ethnogenesis not just once, but at different time depths in the past. This new understanding helps further 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 myths of a Mongoloid race and a Sino-Tibetan language family tree still survive in modern discourse. Both paradigms are false and historically rooted in "scientific" racism. The two myths must be abandoned. At the same time, in studying languages and genes, correlations must not be confused with identity and a number of other caveats must be heeded. The remarkable finding that peoples and nations are observed more often to speak Father Tongues than Mother Tongues is explained. The evidence is presented which tells the tale of how the Eastern Himalaya served as the ultimate homeland to all the East Asian language families.

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 the 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 about 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.[1]

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 *Mangol jāti* (मंगोल जाति), 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 nineteenth 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 of his from the Republic of Georgia. Whilst driving a rented car, they were pulled over by a police officer. The obese and heavily armed man in uniform demanded to see my friend's driving licence and then asked them, "Are you folks Arabs?" The policeman spoke with a heavy American accent and pronounced the word Arabs as ['eIræ:bz]. Since Abkhazia and Georgia both lie in the Caucasus, 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 like a Caucasian. We are Caucasians'. The exasperated policeman spluttered, "…but …but I am White!"

In the aftermath, my friend had to explain to the American policeman where the Caucasus Mountains lay and who the Caucasians were. However, he did not go as far as to explain that the idea that Europeans are purportedly Caucasian originated with Meiners in 1813. Like the Mongoloid, the Caucasoid was another one of his racial prototypes. Americans who apply for a driving licence, take a Scholastic Aptitude Test (SAT) or fill in any number of other official forms are often asked to specify their race. A person of European ancestry often checks a box saying that he or she is a 'Caucasian'. Some people from Asia and Africa are baffled by these racial questions and by the choices of race on offer, which differ from one form to another, and then end up having to decide whether they are 'coloured' or belong to some other 'race'. Although the topic of race is taboo in America, American society is both riddled with antique modes of thinking about race and very much in denial about widely held racist assumptions. America 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 less knowledgeable scholars whose ignorance does not make them less prolific writers 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 confused 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 twelve 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 a 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 nineteenth 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 a "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.[2] 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-Yenisseian (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 are 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 Klaproth'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 **(Figure 1)**, but most of the over 300 different languages and three fourths of Trans-Himalayan subgroups are located to the south of the Himalayan divide **(Figures 2, 3)**.

Figure 3.1: Geographical Distribution of Trans-Himalayan Languages

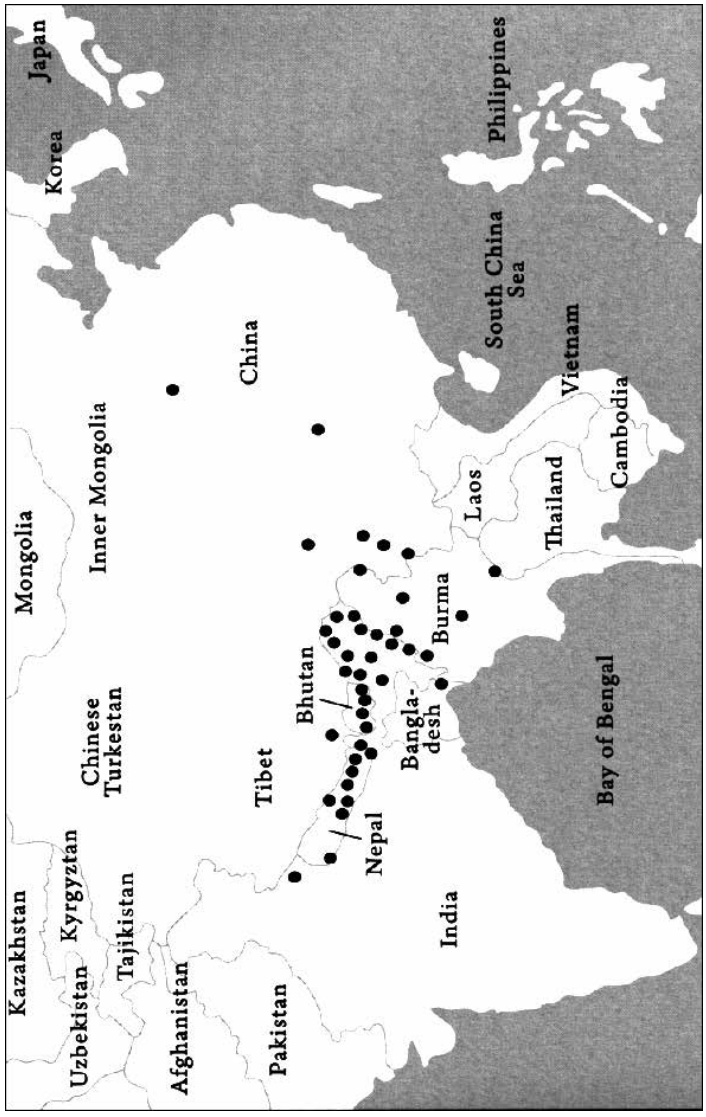

Figure 3.2: Geographical distribution of the major Trans-Himalayan subgroups

(Each dot represents not just one language but the putative historical geographical centre of each of the 42 major linguistic subgroups)

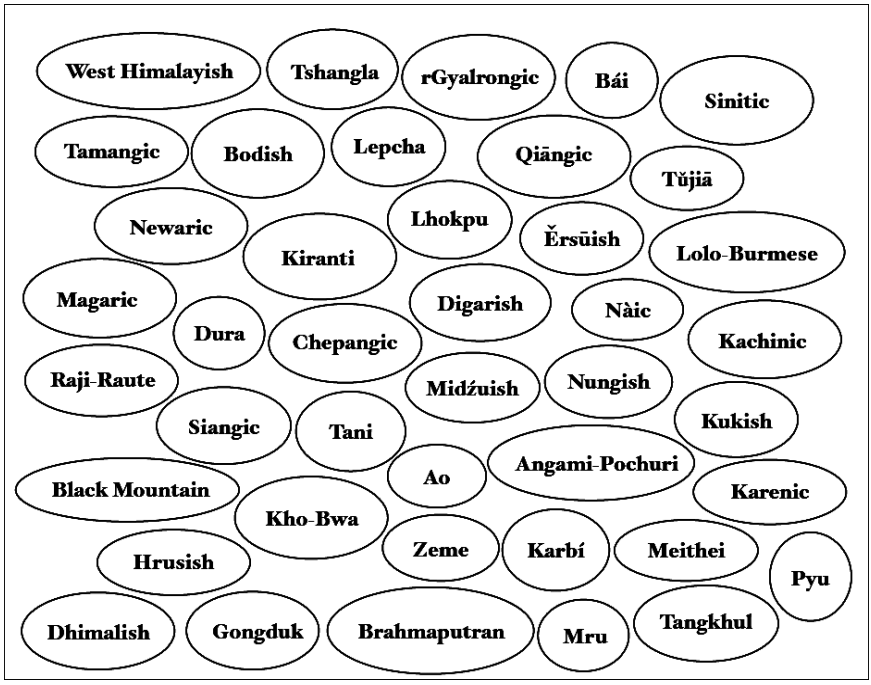

Figure 3.3: Thirty out of forty-two Trans-Himalayan subgroups lie south of the Himalayan divide, seven to the north and east, and five (Tshangla, Bodish, Nungish, Lolo-Burmese and Kachinic) straddle both flanks of the Himalaya

Words of Caution on Language and Genes

Numerous scholars since the early nineteenth 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 altitudes 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 humans is not really entirely extinct because the Denisovans interbred with the ancestors of many existing populations and 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, inter alia, 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 (Poloni *et al.* 1997, 2000), even before the appearance of the first Y-chromosomal tree in 2000 (Underhill *et al.* 2000, 2001). Today we have an even higher resolution picture of the Y-chromosomal haplogroup tree and the world's paternal lineages, e.g. Karafet *et al.* (2008).

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 intruders formed an élite 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 Hàn 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 being 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 perhaps preserves a vestige of a tracer for 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).[3] 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 **(Figure 4)**. The ancient populations bearing haplogroups N and O underwent expansions 18,000–12,000 years ago.

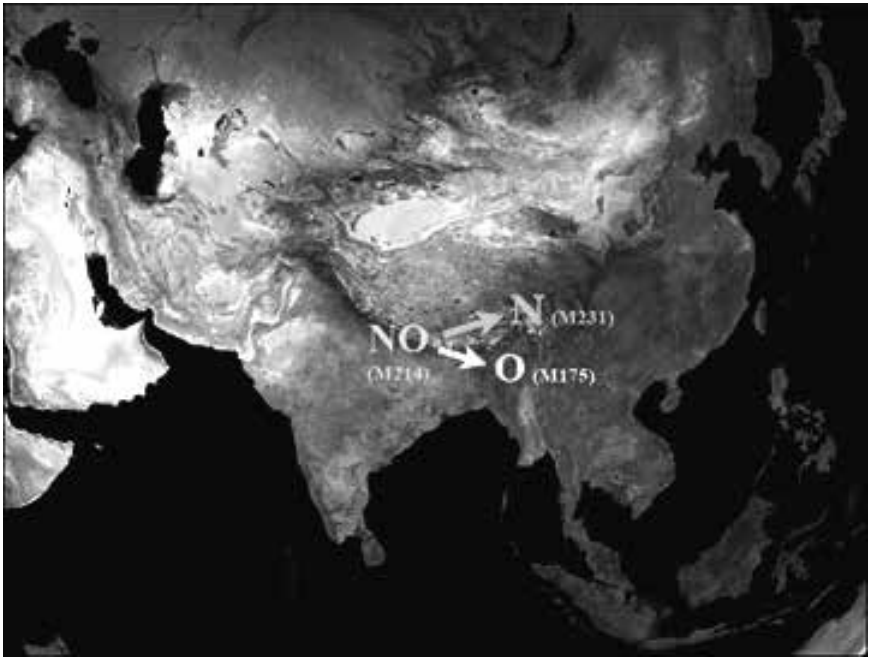

Figure 3.4: The split-up of paternal lineage NO (M214) into the haplogroups N (M231) and O (M175)

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 Lappland **(Figure 5)**. 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 northwards after the paternal founder lineages had already established themselves in the Americas. The Greater Yenisseian 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.

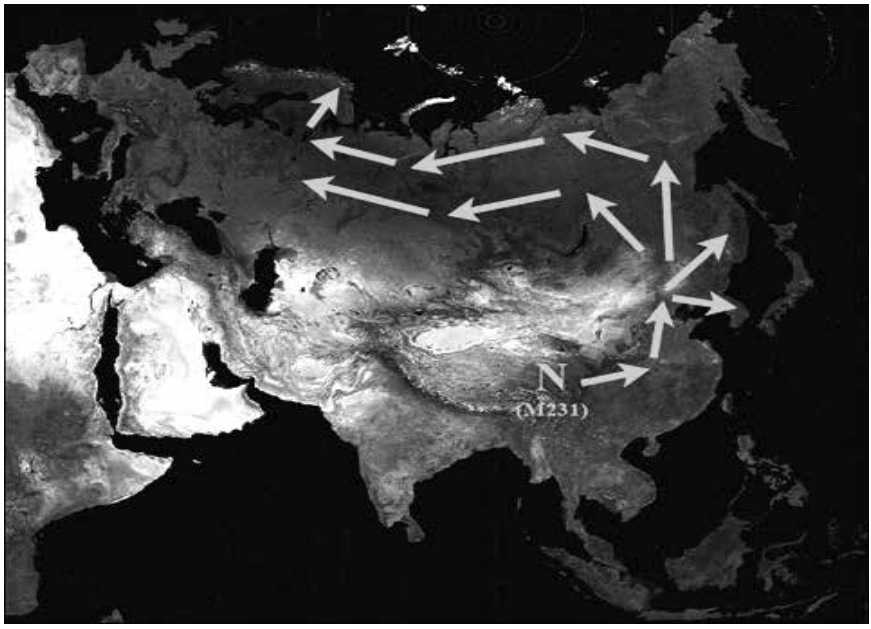

Figure 3.5: The counterclockwise spread of the paternal lineage N (M231), based on Rootsi et al. (2007)

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. 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 Tamyr peninsulas in northern Siberia.

The East Asian Linguistic Phylum

Julius von Klaproth (1783–1835) 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, Kradai, Austronesian and Austroasiatic (**Figures** 1, 6, 7, 8, 9). Later generations of linguists began to discern possible long-distance relationships between the recognised families. In 1901, Gustave Schlegel argued that Kradai 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.

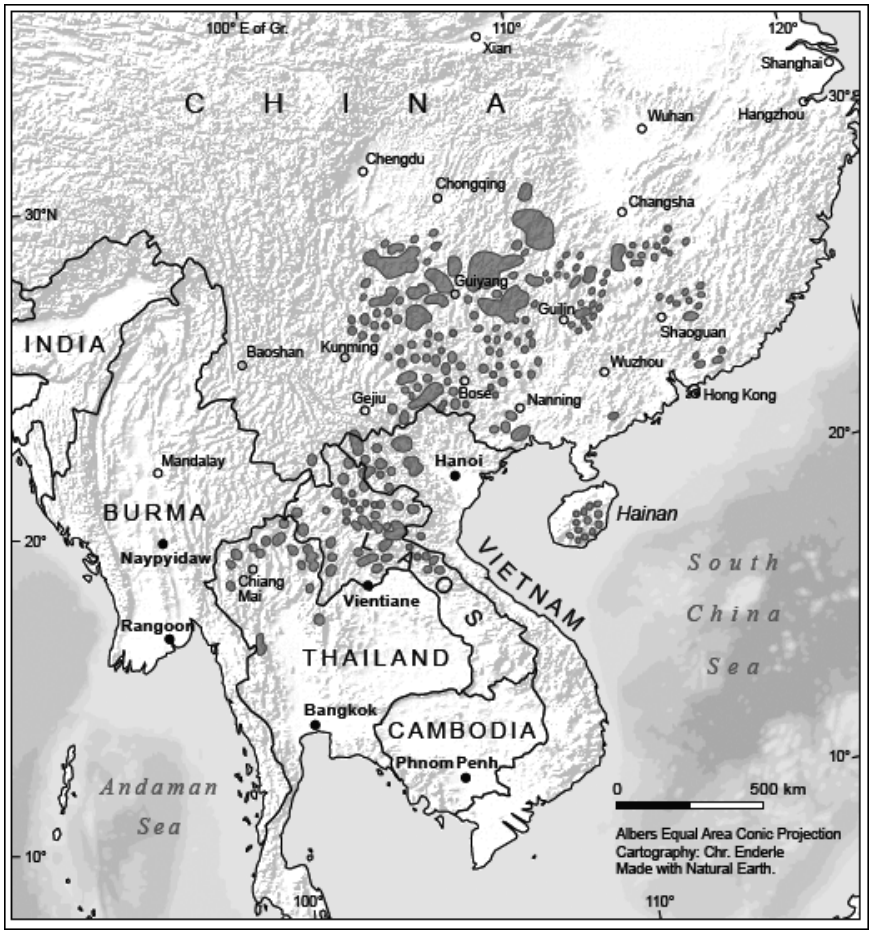

Figure 3.6: Geographical Distribution of Hmong-Mien

In 2005, Weera Ostapirat became the first to present a methodologically sound linguistic evidence that Kradai 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 Kradai 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 Kradai in an Austro-Tai family, Ostapirat has effectively reduced the number of East Asian language families from five to four.

Figure 3.7: Geographical Distribution of Kradai

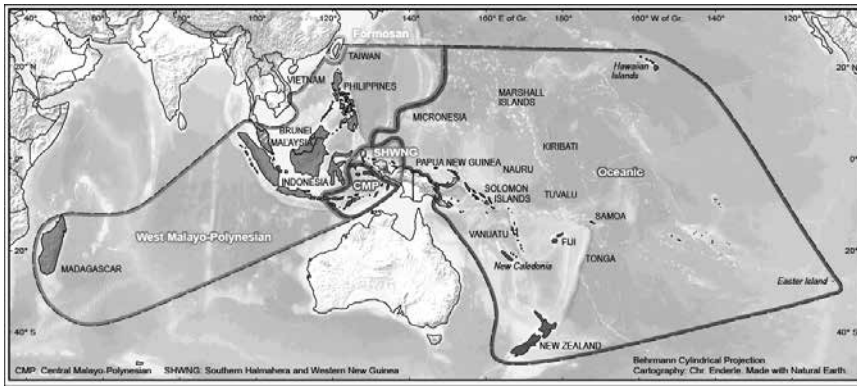

Figure 3.8: Geographical Distribution of Austronesian

Since the beginning of the twentieth 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".

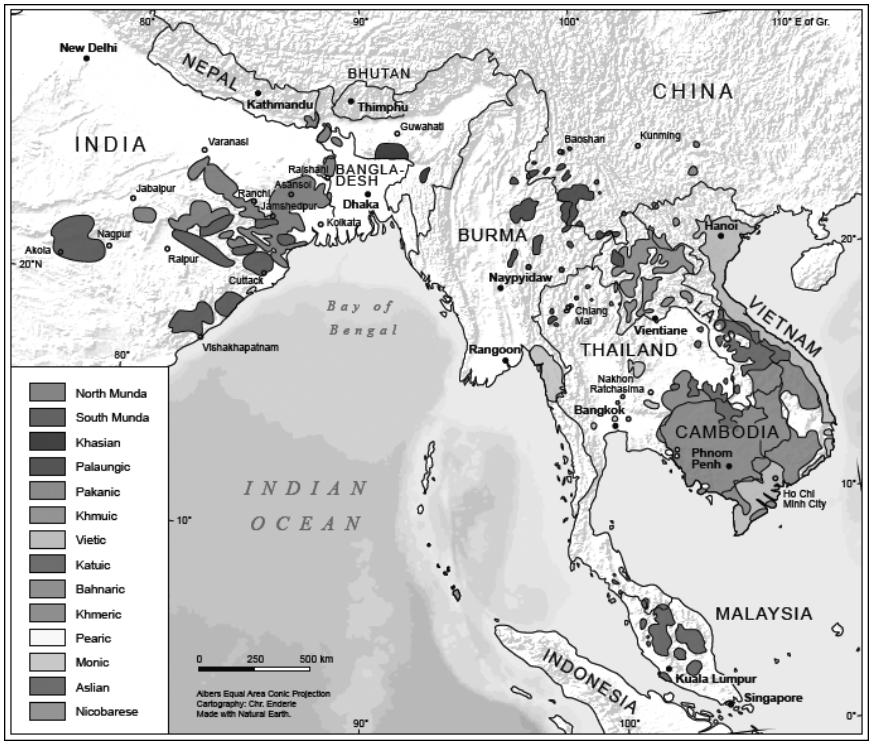

Figure 3.9: Geographical Distribution of Austroasiatic

August Conrady in 1916 and Kurt Wulff in 1934 each proposed a superfamily consisting of Austroasiatic, Austronesian, Kradai and Tibeto-Burman. Subsequently, Robert Blust in 1996 and Ilia Peiros in 1998 proposed an "Austric" superfamily comprising Austroasiatic, Austronesian, Kradai and Hmong-Mien. In 2001, a year before he died of congestive heart failure, Stanley Starosta had proposed the East Asian linguistic phylum encompassing Kradai, 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 disyllablic 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. At Benares in 2012, I presented the tweaked East Asian family tree depicted in **Figure 10.**

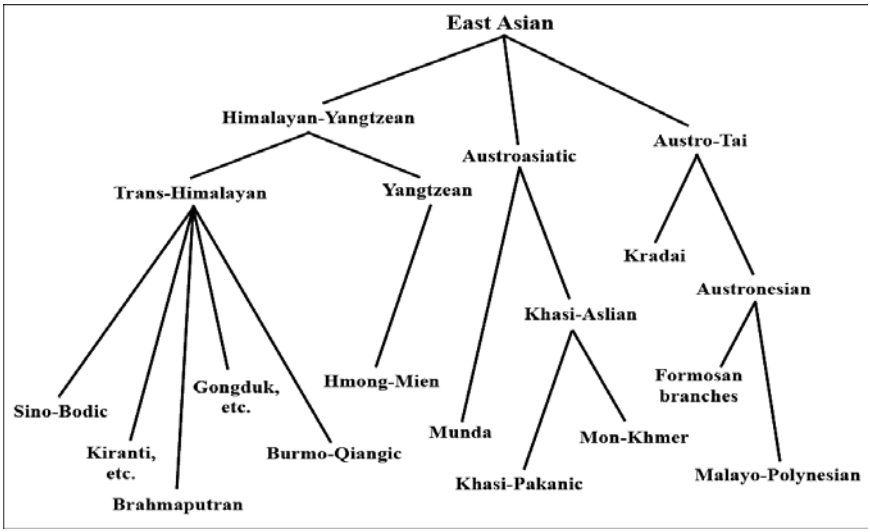

Figure 3.10: The 2012 Benares Recension: Revised East Asian Phylogeny

Eastern Himalayan Homeland

The East Asian linguistic phylum consists of the four language families: Trans-Himalayan, Hmong-Mien, Austroasiatic and Austro-Tai. The populations speaking these languages today are not characterised by just 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 **(Figure 11)**. 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 **(Figure 12).**

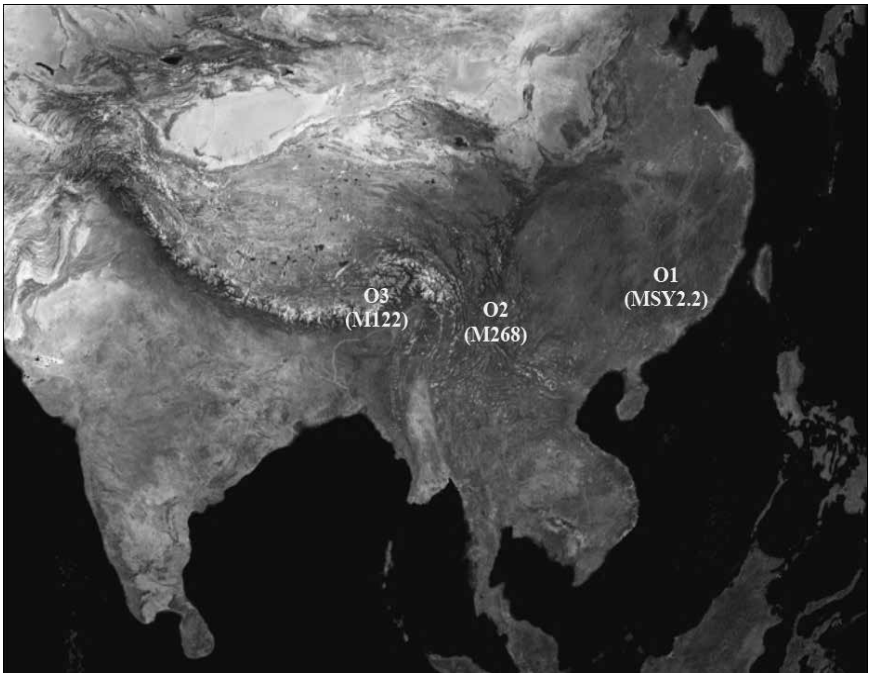

Figure 3.11: After the Last Glacial Maximum, the Y-chromosomal haplogroup O (M175) split into the subclades O1 (M119), O2 (M268) and O3 (M122). Bearers of the O2 (M268) paternal lineage domesticated Asian rice.

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.

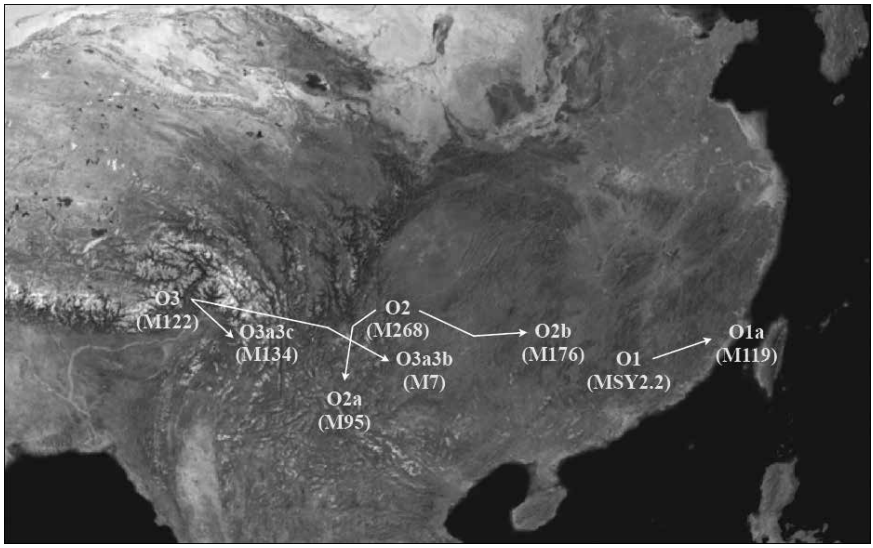

Figure 3.12: Paternal lineages branching into new subclades. Each event involved a linguistic bottleneck leading to language families that today are reconstructible as distinct linguistic phyla. The O1 (MSY2.2) lineage gave rise to the O1a (M119) subclade, which moved eastward to the Fujian hill tracts and across the straits to Formosa, which therefore became the homeland of the Austronesians. Bearers of O3a3b (M7) became the Proto-Hmong-Mien. In the Eastern Himalaya, the bearers of haplogroup O3a3c (M134) expanded and became the Trans-Himalayans. Haplogroup O2a (M95) is the Proto-Austroasiatic paternal lineage. The para-Austroasiatic fraternal clade O2b (M176) spread eastwards, sowing seeds along the way.

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

We can identify the O2b (M176) lineage with the Yayoi people, who introduced rice agriculture to Japan, perhaps as early as the second millennium BC, during the final phase of the Jōmon period (Tanaka *et al.* 2004; Hammer *et al.* 2006). The Yayoi appear to have introduced other crops of continental inspiration to the Japanese archipelago such as millet, wheat and melons. The gracile Yayoi immigrants soon outnumbered the more robust and less populous Jōmon, who had been the first anatomically modern humans to populate Japan. The Y-chromosomal haplogroup O2b and other O haplogroups in Japan are later arrivals but account for more

than half of all Japanese paternal lineages, with their highest frequencies in Kyūshū.

A *Father Tongue* theory for Altaic which assumes no close affinity between Altaic and Uralo-Siberian entails that an antique C haplogroup, perhaps C3, represents an early trace of a paternally disseminated linguistic phylum at a great time depth. Factors such as changes in the ambient material world, social upheaval and cultural transformation, which are known to accelerate the pace of language change, have played an ever more salient role in human life since the Neolithic revolution. It is conceivable, therefore, that language may have changed at a slower tempo in Paleolithic times. Scholars of the Altaic language family have reconstructed a very ancient linguistic relationship. No doubt, much of this old linguistic stratum was lost long ago. The remnants of this *Father Tongue* survive in Japan as Japanese and elsewhere in Asia as the other languages of the Altaic language family, i.e. Korean, Tungusic, Mongolic and Turkic. Another ancient *Father Tongue*, entirely distinct from Altaic and anciently introduced to Japan by the bearers of the Y-chromosomal haplogroup D2 (M55), also survives today in the Japanese archipelago as Ainu.

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 eastwards, 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 (**Figure 2**). Initially Trans-Himalayans expanded through Sichuan and Yunnan, north and northwest across the Tibetan plateau, westwards across the Himalaya and southwards into the Indo-Burmese borderlands.

On the Brahmaputran plain, the early Tibeto-Burmans encountered the Austroasiatics who had preceded them. The ancestral Trans-Himalayan paternal lineage O3a3c also spread from the Eastern Himalaya in a northeasterly direction across East Asia to the North China plains. Subsequently, at a shallower time depth, the Tibeto-Burman paternal lineage O3a3c spread from the Yellow River basin south into southern China, beginning with the Hàn expansion during the Qín 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 Lappland, many of these ancestors first passed through the Eastern Himalaya and crossed the Brahmaputra.

Notes

1. *Die Gesichts- und Körperbildung der Mongolen steht von der gewöhnlichen Form eben so sehr ab, als die der Neger. Und wenn irgend eine Nation verdient, als uraltes Stammvolk betrachtet zu werden; so kommt dieser Nahme mit recht den von allen anderen Asiatischen Völkern, der körperlichen und moralischen Beschaffenheit nach so sehr verschiedenen Mongolen zu.*
2. October 29, 2009 at the 4th International Conference on Austroasiatic Linguistics at Mahidol University in Bangkok, February 24, 2012 in a talk entitled "*The Present State of Sino-Tibetan Studies: Progress and Outstanding Issues*" at a special seminar for the Hakubi Project and Centre for Southeast Asian Studies at Kyōto University, and and October 26, 2012 at the Conference for Sino-Tibetan Languages and Linguistics at Nanyang Technological University in Singapore.
3. Haplogroup labels (O2a, O2b, etc.) are updated regularly by the Y-Chromosome Consortium. Mutation numbers (M95, M176, etc.) remain unchanged.

References

1. Beames, John (1868). *Outlines of Indian Philology.* London: Trübner and Company.
2. Benedict, Paul King (1975). *Austro-Thai: Language and Culture, with a Glossary of Roots.* New Haven: Human Relations Area Files.
3. Blust, Robert (1996). 'Beyond the Austronesian Homeland: The Austric Hypothesis and its Implications for Archaeology', pp. 117-160 in Ward H. Goodenough (ed), *Prehistoric Settlement of the Pacific*. Philadelphia: American Philosophical Society.
4. Bodman, Nicholas Cleaveland (1973). 'Some Chinese Reflexes of Sino-Tibetan s-Clusters', *Journal of Chinese Linguistics,* 1(3): 383-396.
5. Bouda, Karl (1950). 'Die Sprache der Burusch̃o', *Eusko-Jakintza, Revue d'Études Basques*, IV: 37-50.
6. Byrne, James (1885). *General Principles of the Structure of Language* (2 volumes). London: Trübner and Company.
7. Conrady, August (1916). 'Eine merkwurdige Beziehung zwischen den austrischen und den indochinesischen Sprachen', pp. 475-504 in *Aufsätze zur Kultur- und Sprachgeschichte vornehmlich des Orients: Ernst Kuhn zum 70. Geburtstage am 7. Februar 1916 gewidmet von Freunden und Schulern.* München: Verlag von M. & H. Marcus.
8. Cust, Robert N. (1878). *A Sketch of the Modern Languages of East India.* London: Trübner and Company.
9. Driem, George van (2013). East Asian Ethnolinguistic Phylogeography, *Bulletin of Chinese Linguistics,* 7(1):1-55.
10. Forbes, Charles James F.S. (1878). 'On Tibeto-Burman Languages', *Journal of the Royal Asiatic Society of Great Britain and Ireland,* X: 210-227.
11. Gāo Jīngyī (2008). *Hányǔ yú Běi'oū Yǔyán: Hányǔ yú Wūlāěr Yǔyán jí Yìnǒu Yǔyán Tóngyuán Tànjiū* ['Chinese and Nordic Languages: Investigating Correspondences between Chinese and Indo-European and Uralic']. Peking: Zhōngguó Shèhuì Kēxué Chūbǎnshè.

12. de Gobineau, Joseph Arthur (1854, 1855). *Essai sur l'inégalité des races humaines* (4 volumes). Paris: Firmin-Didiot et Compagnie.
13. Hammer, Michael F., Tatiana M. Karafet, Hwayong Park, Keiichi Omoto, Shinji Harihara, Mark Stoneking and Satoshi Horai (2006). 'Dual Origins of the Japanese: Common Ground for Hunter-Gatherer and Farmer Y Chromosomes', *Journal of Human Genetics*, 51: 47-58.
14. Hodgson, Brian Houghton (1857). 'Comparative Vocabulary of the Languages of the Broken Tribes of Népál', *Journal of the Asiatic Society of Bengal,* XXVI: 317-371.
15. Houghton, Bernard (1896). 'Outlines of Tibeto-Burman Linguistic Palæontology', *Journal of the Royal Asiatic Society,* 23-55.
16. Jones, Robert Reese (1995). 'Two-Hundred CALMSEA ['Culturally Appropriate Lexicostatistical Model for Southeast Asia'] Comparisons in Sino-Tibeto-Mayan', paper presented at the XXVIII International Conference on Sino-Tibetan Languages and Linguistics, University of Virginia at Charlottesville, October 1995.
17. Karafet, Tatiana M., Fernando L. Mendez, Monica B. Meilerman, Peter A. Underhill, Stephen L. Zegura and Michael F. Hammer (2008). 'New Binary Polymorphisms Reshape and Increase Resolution of the Human Y-Chromosomal Haplogroup Tree', *Genome Research,* 18: 830-838.
18. von Klaproth, Julius Heinrich (1823). *Asia Polyglotta.* Paris: A. Schubart.
19. Meiners, Christoph (1813). [posthumous]. *Untersuchungen uber die Verschiedenheiten der Menschennaturen (die verschiedenen Menschenarten) in Asien und in den Südländern, in den Ostindischen und Südseeinseln, nebst einer historischen Vergleichung der vormahligen und gegenwärtigen Bewohner dieser Continente und Eylande.* Tübingen: in der J.G. Cotta'schen Buchhandlung.
20. Ostapirat, Weera (2005). 'Kra-Dai and Austronesian: Notes on Phonological Correspondences and Vocabulary Distribution', pp. 107-131 in Laurent Sagart, Roger Blench and Alicia Sanchez-Mazas (eds.), *The Peopling of East Asia: Putting Together Archaeology, Linguistics and Genetics*. London: Routledge Curzon.
21. Ostapirat, Weera (2013). '*Austro-Tai Revisited*', 23rd Annual Meeting of the Southeast Asian Linguistic Society, Chulalongkorn University, 29 May 2013.
22. Peiros, Ilia (1998). *Comparative Linguistics in Southeast Asia,* Vol.142. Canberra: Pacific Linguistics.
23. Poloni, Estella Simone, Ornella Semino, Giuseppe Passarino, A.S. Santachiara-Benerecetti, I. Dupanloup, André Langaney and Laurent Excoffier (1997). 'Human Genetic Affinities for Y chromosome P49a,f/*Taq*I Haploptypes Show Strong Correspondence with Linguistics', *American Journal of Human Genetics,* 61: 1015- 1035 (cf. the erratum published in 1998 in the *American Journal of Human Genetics,* 62: 1267).
24. Poloni, Estella Simone, Nicolas Ray, Stefan Schneider and André Langaney. (2000). 'Languages and Genes: Modes of Transmission Observed through the Analysis of Male-Specific and Female-Specific Genes', pp. 185-186 in Jean-Louis Dessalles and Laleh Ghadakpour (eds.), *Proceedings: Evolution of Language, 3rd International Conference 3-6 April 2000*. Paris: École Nationale Supérieure des Télécommunications.

25. Przyluski, Jean (1924). 'Le sino-tibétain', pp. 361-384 in Antoine Meillet and Marcel Cohen, (eds.), *Les Langues du Monde.* Paris: Librairie Ancienne Édouard Champion.

26. Pulleyblank, Edwin George (1966). 'Chinese and Indo-Europeans', *Journal of the Royal Asiatic Society of Great Britain & Ireland* (New Series), 98 (1): 9-39.

27. Ramstedt, Gustaf John (1957). [posthumous]. *Einfuhrung in die altaische Sprachwissenschaft: Lautlehre* (Suomalais-Ugrilaisen Seuran Toimituksia, Mémoires de la Société Finno-Ougrienne, 104, 1). Helsinki: Suomalais-Ugrilaisen Seura.

28. Rasmussen, Morten, and Xiaosen Guo, Yong Wang, Kirk E. Lohmueller, Simon Rasmussen, Anders Albrechtsen, Line Skotte, Stinus Lindgreen, Mait Metspalu, Thibaut Jombart, Toomas Kivisild, Weiwei Zhai, Anders Eriksson, Andrea Manica, Ludovic Orlando, Francisco de la Vega, Silvano Tridico, Ene Metspalu, Kasper Nielsen, María C. Ávila-Arcos, J. Víctor Moreno-Mayar, Craig Muller, Joe Dortch, M. Thomas P. Gilbert, Ole Lund, Agata Wesolowska, Monika Karmin, Lucy A. Weinert, Bo Wang, Jun Li, Shuaishuai Tai, Fei Xiao, Tsunehiko Hanihara, George van Driem, Aashish R. Jha, François-Xavier Ricaut, Peter de Knijff, Andrea B. Migliano, Irene Gallego-Romero, Karsten Kristiansen, David Lambert, Søren Brunak, Peter Forster, Bernd Brinkmann, Olaf Nehlich, Michael Bunce, Michael Richards, Ramneek Gupta, Carlos Bustamante, Anders Krogh, Robert A. Foley, Marta Mirazón Lahr, François Balloux, Thomas Sicheritz-Pontén, Richard Villems, Rasmus Nielsen, Wang Jun, Eske Willerslev (2011). 'An Aboriginal Australian Genome Reveals Separate Human Dispersals into Asia', *Science,* 334 (6052): 94-98.

29. Reid, Lawrence A. (2005). 'The Current Status of Austric: A Review Footnote numbers 37, 38, 39 are in a different font and Evaluation of the Lexical and Morphosyntactic Evidence', pp. 132-160 in Laurent Sagart, Roger Blench and Alicia Sanchez-Mazas (eds.), *The Peopling of East Asia: Putting Together Archaeology, Linguistics and Genetics.* London: Routledge Curzon.

30. Renan, Ernest (1858). *De l'origine du langage* (deuxième édition, revue et considérablement augmentée). Paris: Michel Lévy, Frères, Librairies-Éditeurs.

31. Sagart, Laurent (1993). 'Chinese and Austronesian: Evidence for a Genetic Relationship', *Journal of Chinese Linguistics* 21 (1): 1-62.

32. Schlegel, Gustave (1901). Review: '*Elements of Siamese Grammar*' by O. Frankfurter, Ph.D., Bangkok: Printed at the American Presbyterian Mission Press, Leipzig, Karl W. Hiersemann, (1900), *T'oung Pao* (Série II), II: 76-87.

33. Schmidt, Wilhelm (1926). *Die Sprachfamilien und Sprachenkreise der Erde* (2 vols., the second of which is an atlas). Heidelberg: Carl Winter.

34. Starostin, Sergej Anatol'evič (1994). 'The Reconstruction of Proto-Kiranti', paper presented at the 27ème Congrès International sur les Langues et la Linguistique Sino-Tibétaines, Centre International d'Études Pédagogiques à Sèvres, 14 October 1994.

35. Steinthal, Heymann (1850). *Die Classification der Sprachen dargestellt als die Entwickelung der Sprachidee.* Berlin: Ferdinand Dümmler's Buchhandlung.

36. Steinthal, Heymann (1860). *Charakteristik der hauptsächlichsten Typen des Sprachbaues.* Berlin: Ferdinand Dümmler's Verlagsbuchhandlung.

37. Tanaka, Masashi, Vicente M. Cabrera, Ana M. González, José M. Larruga, Takeshi Takeyasu, Noriyuki Fuku, Li-Jun Guo, Raita Hirose, Yasunori Fujita, Miyuki Kurata,

Ken-ichi Shinoda, Kazuo Umetso, Yoshiji Yamada, Yoshiharu, Oshida, Yuzo Sato, Nobutaka Hattori, Yoshikuni Mizuno, Yasumichi Arai, Nobuyoshi Hirose, Shigeo Ohta, Osamu Ogawa, Yasushi Tanaka, Ryuzo Kawamori, Masayo Shamoto-Nagai, Wakako Maruyama, Hiroshi Shimokata, Ryota Suzuki and Hidetoshi Shimodaira (2004). 'Mitochondrial Genome Variation in Eastern Asia and the Peopling of Japan', *Genome Research,* 10A: 1832-1850.

38. Underhill, Peter A., Peidong Shen, Alice A. Lin, Jīn Lì, Giuseppe Passarino, Wei H. Yang, Erin Kauffman, Batsheva Bonné-Tamir, Jaume Bertranpetit, Paolo Francalacci, Muntaser Ibrahim, Trefor Jenkins, Judith R. Kidd, S. Qasim Mehdi, Mark T. Seielstad, R. Spencer Wells, Alberto Piazza, Ronald W. Davis, Marcus W. Feldman, Luigi Luca Cavalli-Sforza and Peter J. Oefner (2000). 'Y Chromosome Sequence Variation and the History of Human Populations', *Nature Genetics*, 26: 358-361.
39. Underhill, Peter A., G. Passarino, A.A. Lin, P. Shen, Marta Mirazón-Lahr, Robert A. Foley, P.J. Oefner and Luigi Luca Cavalli-Sforza (2001). 'The Phylogeography of Y Chromosome Binary Haplotypes and the Origins of Modern Human Populations', *Annals of Human Genetics,* 65: 4-62.
40. Wulff, Kurt (1934). *Chinesisch und Tai: Sprachvergleichende Untersuchungen*. Copenhagen: Levin & Munksgaard.