


GLIMPSES OF THE ETHNOLINGUISTIC PREHISTORY OF NORTHEASTERN INDIA

GEORGE VAN DRIEM

OLD AND NEW LINGUISTIC PHYLA IN THE NORTHEAST OF THE SUBCONTINENT

The world’s two most populous families of languages meet in the Himalayas. These are Tibeto-Burman, which includes Cantonese and Mandarin, and Indo-European, to which languages such as English and Bengali belong. In addition to these two great linguistic phyla, Kra-Dai alias Daic, Austroasiatic and Dravidian language communities skirt the eastern Himalayan region. For example, the Austroasiatic language Khasi is spoken in the Indian state of Meghalaya. Ahom, a now extinct Kra-Dai tongue, was once a prominent cultural language in northeastern India, where scattered Daic language communities are still settled today. The Dravidian tongues Dhangar and Jhangar, which are basically dialects of Kurukh or Uraon, are spoken in Nepal’s eastern Terai. Kurukh is also spoken in scattered communities throughout northeastern India by people displaced by British colonial policies in India involving resettlement, not unlike the policy of transmigrasi in the former Dutch East Indies. Yet the Himalayas would appear to be peripheral to our understanding of the prehistory of Dravidian. The crux of the ethnolinguistic prehistory of the eastern Himalayan region are the language families Tibeto-Burman, Austroasiatic, Indo-European and, to a lesser extent, Kra-Dai.

The advent of Indo-European and Kra-Dai languages to the northeastern portion of the Indian subcontinent is relatively recent and to some extent historically attested. Yet whilst the impact of Kra-Dai today is marginal, the intrusion of Indo-European into the region has been more robust and is represented by expansive languages such as Assamese, Hindi, Bengali and English. At the same time, historical, ethnographic, archaeological, anthropological and linguistic data all point to an Austroasiatic and Tibeto-Burman presence in northeastern India which must date back to some hoary period of ethnolinguistic prehistory. Austroasiatic and Tibeto-Burman, therefore, hold the key to understanding the population prehistory of northeastern India and the Indo-Burmese borderlands.

The deeper phylogeny of the Indo-European language family was
once conceived by August Schleicher as a branching oak tree, but the
phylum has increasingly assumed a more rake-like appearance in more
recent literature. The currently best-informed family tree structure of the
Tibeto-Burman phylum is likewise essentially rake-like, a situation for
which I introduced the metaphor of fallen leaves (figure 9.1).

Figure 9.1. The fallen leaves diagram for Tibeto-Burman. Some subgroups
are well-established, whilst others are less so. Brahmaputran may include
Kachinic and Dhimalish. For the sake of argument, this diagram breaks up the
traditional catch-all ‘Qiängic’ into the Ėrsū cluster and a truncated ‘Qiäng’,
not to posit a robust phylogenetic hypothesis but to challenge, thus empha­sising that crucial work in this area has been left undone. The precise phylo­genetic relationships between the diverse rGyal-rong languages, Ėrgong, Qiäng, Mi-räng (Mûyä), Tangut, Ėrsū, Lûsû, Tōsu (Duôxû), Nâmûyï, Shihîng, Gṳiíqîng, Chîyo (Quêyû), Zhäuser and Prînnî (Pâmî) have yet to be dem­onstrated. What is sometimes called ‘Northern Qiängic’ is supposed to include
the rGyal-rongic group recognised by Jackson Sun (Sûn Tiänxîn) and Huang
Bûfán. Hopefully scholars working on the Tibeto-Burman languages of Sichuán
and Yûnnán provinces will in the coming years shed light on the structure
of this portion of the Tibeto-Burman family tree (van Driem 2001, 2006).
The geographical distribution of the major Tibeto-Burman subgroups saliently shows the densest concentration in the northeastern portion of the Indian subcontinent, with most major subgroups being exclusive to this area (figure 9.2). The most authoritative phylogenetic tree for Austroasiatic at present is the model developed by Gérard Diffloth (figure 9.3).

![Phylogenetic Tree of Austroasiatic Languages]

Figure 9.3. Austroasiatic with Gérard Diffloth's tentative calibration of time depths for the various branches of the language family (modified from Diffloth 2001, 2005). The precise phylogenetic propinquity of Pearic, after Khmeric loan layers have been stripped off, remains uncertain except that Diffloth observes that Pearic is Mon-Khmer and not ‘une espèce de vieux khmèr’, as some scholars once maintained. This diagram arranges in a tree-shaped phylogeny the fourteen recognised branches of Austroasiatic, i.e. North Munda, South Munda, Khasian, Pakanic, Palaungic, Khmuic, Vietic, Katuic, Bahnaric, Khmeric, Pearic, Monic, Aslian and Nicobarese.

Some novel insights into the phylogeny of Kra-Dai have been put forward by Edmondson and Solnit (1988, 1997) and Ostapirat (2005).

The location of the Austroasiatic ancestral homeland can be argued from a purely linguistic point of view principally on the basis of linguistic paleontology and the geographical centre of gravity of the family. The latter is ascertained from the distribution of modern Austroasiatic language communities and the deep phylogenetic divisions in the family. The distribution of the modern language communities and the geography of the deepest divisions in linguistic phylogeny would put the geographical centre of the family somewhere between South Asia and Southeast Asia, in the area around the northern coast of the Bay of Bengal. Whether we assume that the deepest division in the family lies between Munda and the rest, as an older generation of scholars used to suspect, or assume the veracity of Diffloth’s new tripartite division, then the geography of deep historical divisions in linguistic phylogeny would compel us to look for a homeland on either side of the Ganges and Brahmaputra delta, although we would be unable to say whether this homeland would have to have lain to the east or to the west of the delta.¹

When linguists look beyond what linguistic phylogeny can tell them, they must ask which archaeological transition or modern genetic gradient can be related with confidence to an ancient linguistic intrusion or to the prehistorical spread of a language family. When linguists resort to linguistic paleontology, they must have recourse to the findings of palaeoclimatologists and, more particularly, palaeobotanists. Most conceivable theories about the homelands of Tibeto-Burman and Austroasiatic have already been put forward. The idea of a Tibeto-Burman homeland situated in or nearby present-day Sichuán has been entertained since the nineteenth century, especially by British scholars in India. Sinocentrists favour a northern Tibeto-Burman homeland in the lower Yellow River basin on the plains of northern China, whereas some have proposed a provenance within the Himalayan region itself. Scholars have sought to situate the Austroasiatic Urheimat as far west as the Indus valley and as far east as the Yangtze delta or insular Southeast Asia. However, the main contenders today for the Austroasiatic homeland are the Indian subcontinent, mainland Southeast Asia and the middle Yangtze.¹

¹The presence of many speakers of Munda languages in northeastern India is a legacy of resettlement to Assam orchestrated by the East India Company, an economic policy mentioned above in connexion with speakers of Northern Dravidian languages.
ARCHAEOLOGY AND LINGUISTIC PALAEONTOLOGY

The fundamental epistemological question that will continue to haunt us is whether the spread of a recognisable Neolithic and Bronze Age assemblage can actually ever be taken with certainty to reflect the spread of a language and so of a language family. Archaeology reflects what we have been able to glean about the material culture of past communities. In fact, how often can we be certain which language was spoken by ancient stone knappers or by the potters behind a particular ceramic culture in some archaeologically attested pre-literate society? Indeed, we must ask whether the modern geographical distribution of the Tibeto-Burman language family correlates with the mute testimony of any single portion of the archaeological record that happens to have been preserved, discovered and studied by archaeologists. A more general issue is time depth. Archaeological transitions are reconstructed at very different times in the past, e.g., the palaeontologically attested spread of anatomically modern humans, the spread of agriculture, and the sometimes well-defined patterns of dispersal of identifiable cultural assemblages in the Neolithic and Bronze Age. At the same time, many known historical transitions and conquests with linguistic consequences have left little or no clear-cut traces whatsoever in the archaeological record.

Therefore, we are free to speculate, for example, whether the expansion of early Bodic language communities into the Himalayas was associated with the sudden appearance of exponents of the Mäjiāyáo Neolithic in eastern Tibet at mKhar-ro and in Kashmir at Burzahom at phase (2700-2300 BC) and the phase (2200-1900 BC) of the Mäjiāyáo cultural assemblage in Gănsū shrank during a period of climate change between the Mäjiāyáo phase (2700-2300 BC) and the Bănsān phase (2200-1900 BC) of the Mäjiāyáo sequence. This at least is a scenario which I argued in several earlier publications (van Driem 1998, 2001, 2002). Recent palaeoecological evidence indicates that the vast and once heavily forested Tibetan Plateau underwent large-scale deforestation precisely during this period of projected Bodic expansion. The palaeobotanical evidence indicates that this deforestation occurred at the hands of human settlers (Kaiser et al. 2006, Miehe et al. 2006, Wu et al. 2006, Kaiser et al. 2007). Were these people perhaps Bodic colonists from the Mäjiāyáo Neolithic core area who introduced to the Tibetan Plateau a new lifestyle with deleterious ecological ramifications? Recently, I have presented several alternative scenarios which differently relate the traceable patterns of dispersal of cultural assemblages in the archaeological record with the present geographical distribution of Tibeto-Burman language communities (van Driem 2006).

Linguistic palaeontology, a term introduced by Adolphe Pictet in 1859, is an attempt to understand the ancient material culture of a language family on the basis of the lexical items which can be reliably reconstructed for the common ancestral language. Linguistic palaeontology, like any attempt to give an ethnolinguistic interpretation to the archaeological record, invariably raises complex issues. Elsewhere, I have discussed at some length the arguments relating to what we can glean about ancient Tibeto-Burman culture and the role played by broomcorn millet Panicum miliaceum and by foxtail millet Setaria italica, the latter reflected in languages as far flung as Old Chinese 楠 tsisik in the Yellow River basin and Lhokpu2 cákt’o foxtail millet’ in modern southwestern Bhutan (van Driem 2006). At the present time, the earliest archaeologically attested domestic millet dates from before 6000 BC at Xinglonggou near 赤峰 Chifeng, where a Neolithic culture without sickles once flourished (Zhào 2005).

Linguistic palaeontology strongly qualifies the ancient Austroasiatics as the most likely candidates for the first cultivators of rice. Moreover, Diffloth has shown that the reconstructible Austroasiatic lexicon paints the picture of a fauna, flora and ecology of a tropical humid homeland environment. He adduces three salient isoglosses diagnostic for the faunal ecology of the Proto-Austroasiatic homeland that are reconstructible all the way back to common Austroasiatic and are reflected in all branches of the family, i.e., *mrak ‘peacock Pavo muticus’, *tarkuat ‘tree monitor lizard Varanus nebulosus or bengalensis’ and *tonyu ‘binturong’ or the ‘bear cat Arctitis binturong’, a black tropical mammal that is the largest of the civet cats (Diffloth 2005:78). None of these species are native to areas that currently lie within China, and, to our present knowledge, none were ever native to the area that is today China. Such linguistic palaeontological evidence therefore appears to render the middle Yangtze homeland hypothesis less likely.

More reconstructible Proto-Austroasiatic roots indicative of a tropical or subtropical climate are adduced by Diffloth (2005:78), i.e., *(baw)jol ~

2 The Lhokpu are an inbred and genetically highly distinct group within the Himalayan region as a whole; Kraaijenbrink et al. 2006a, Parkin et al. 2006a. The impact of matrilocal and cross-cousin endogamy is clearly discernible in the genetic signature of this language community. Many of the ancient Tibeto-Burman groups may have been matrilineal, matrilocal societies with uxorial local marriage such as the modern Lhokpa and Gongdruk of Bhutan.
For years conventional wisdom in archaeological circles dictated that the origin of rice cultivation back to the Indian subcontinent. Against the background of older datings of domesticated rice and ceramic culture from Gangetic basin and Doab sites such as Koldihawa and Mahagarha, reportedly dating from the seventh millennium BC (Sharma et al. 1980, Pal 1990, Agrawal 2002), there are now newer sites with more reliable dates at Lahuradewa (Lahurādevā), Tōkuvā and Sarāī Nahār Rāī.

At the Lahuradewa site (26°46' N, 82°57' E), the early farming phase, corresponding to period 1A in the site’s clear-cut stratigraphy, has radiocarbon dates ranging from ca. 5300 to 4300 BC.Carbonised material from period 1A was collected by the flotation method, yielding *Setaria glauca* and *Oryza rufipogon* as well as a morphologically distinct, fully domesticated form of rice ‘comparable to cultivated *Oryza sativa*’ (Tewari et al. 2002). More recently, accelerator mass spectrometry dates were obtained on the rice grains themselves, corroborating the antiquity of rice agriculture at the site.

Most recently, new radiocarbon dates for rice agriculture have been coming from the Ganges basin, with the Tōkuvā site near Allahabad now yielding dates comparable in antiquity to those of the Lahuradewa site (Vasant Shinde [Vasant Śivarām Śinde], personal communication 27 November 2007), and exciting new dates for ancient rice agriculture are also emerging from Sarāī Nahār Rāī (Manjil Hazarika, personal communication 7 March 2008). Of course, we are living at a time when a more reliable calibration of radiocarbon dates in general has become a matter of great urgency.

Further east, at least five species of wild rice are native to northeastern India, viz. *Oryza nivara*, *Oryza officinalis* (*O. latifolia*), *Oryza perennis* (*O. longistaminata*), *Oryza meyeriana* (*O. granulata*) and *Oryza rufipogon*, and reportedly over a thousand varieties of domesticated rice are currently in use in the region (Hazarika 2005, 2006a). The different varieties of rice in northeastern India are cultivated in three periods by distinct cultivation processes. In the process of *āhu kheti*, the rice is sown in the months of *Phāgun* and *Sot*, i.e., mid February to early April. The seedlings are not transplanted but ripen in just four months in fields that must be constantly weeded. In *bāhu kheti*, the rice seedlings are sown from mid March to mid April in ploughed wet fields and likewise do not need to be transplanted. In *sāli kheti*, the rice is sown from mid May to mid June, and the seedlings are transplanted. *Sāli kheti* rice varieties are suspected to derive from the wild *officinalis* rice still widely found in swampy village areas. The wild *rufipogon* rice cannot be used for human
consumption because the plants shed their seeds before they ripen, so it is used in Assam and other parts of northeastern India as cattle feed (Hazarika 2006b).

Whilst claims have been published dating the earliest rice cultivation in East Asia to as long ago as 10,000 BC, the currently available evidence indicates that immature morphologically wild rice may have been used by foragers before actual domestication of the crop, e.g., at the 八十擓 Bāshídàng site (7000-6000 BC) belonging to the 彭頭山 Pengtoushan culture in the Middle Yangtze and at sites in the Yangtze delta area such as 跨湖橋 Kuăihuáijiāo, 馬家浜 Mǎjiàbīng (5000-3000 BC) and 河姆渡 Hémǔdù (5000-4500 BC). However, only ca. 5000 BC was the actual cultivation of rice probably first undertaken by people in the Lower Yangtze, who at the time relied far more heavily on the collecting of acorns and water chestnuts (Yasuda 2002; Fuller 2005a, 2005b, 2005c, 2006a, 2006b, 2006c, 2007a, 2007b; Fuller et al. 2007; Zong et al. 2007). There is also currently no evidence for the co-cultivation of rice and foxtail millet along the middle Yangtze until around 3800 BC (Nasu et al. 2006).

Today, our understanding of the palaeoethnobotanical picture is more complex. The two main domesticated varieties of rice, Oryza indica and Oryza japonica, are phylogenetically distinct and would appear to have been domesticated separately. Oryza indica derives from the wild progenitor Oryza nivara and was first cultivated in South Asia or western Southeast Asia, perhaps in two separate domestication events. On the semi-arid Gangetic plain at the end of the mid-Holocene wet period, habitats for wild rices increasingly shifted to oxbows as palaeochannels dried up and turned into oxbow ponds. This shift favoured monsoonal rather than marshland rice species, including Oryza nivara (Fuller 2006a).

Oryza japonica derives from the wild progenitor Oryza rufipogon, and it is currently believed that the rufipogon variety was first cultivated to yield early Oryza japonica along the Middle Yangtze. Harvey et al. (2006) have critically reassessed the morphometrics of rice finds associated with various Neolithic sites throughout the Yangtze basin in light of recent genetic findings. It appears that the wild progenitor Oryza rufipogon was not fully domesticated in the Lower Yangtze to yield early Oryza japonica until ca. 4000 BC. Generally, the archaeological record shows a delay of one to two millennia between the beginning of cultivation and the first clear evidence of domestication sensu stricto, i.e., genetic modification by selective breeding.

Twelve wild forest-margin rice species are known and found mostly in Southeast Asia as well as at old sites of human habitation, e.g., Jiāhú on the Huáí River in Hénán in the seventh millennium BC or Hémūdù in the Yangtze delta in the first half of the fifth millennium BC. Extinct wild varieties of rice also appear to be preserved in the modern japonica genome. Based on the genetics of the officinalis variety, the seasonally wet, puddle-adapted Oryza nivara, and the always wet perennial Oryza rufipogon, there may be evidence for multiple rice domestications in South, Southeast and East Asia. So, maybe the domesticators of Oryza nivara were ancient Austroasiatics, and maybe the domesticators of ancient Oryza rufipogon were ancient Hmong-Mien.

O’Connor (1995) and Blench (2001) have argued that irrigated rice agriculture enabled people to seize control of lowlands and flood plains. People were able to move down from upland areas that had hitherto been more favourable habitats after wet cultivation had transformed lowlands from epidemiologically undesirable places into bountiful habitats. But what if the first cultivators and domesticators of rice already inhabited lowland river basins and flood plains, such as the Ganges or Yangtze basins or even the Brahmaputran flood plains?

Notwithstanding the excellent archaeological work conducted in the Ganges and Yangtze River basins, much of the archaeology of ancient rice agriculture is simply not known because no substantive archaeological work has been done on the Neolithic in the most relevant areas, e.g., northeastern India, Bangladesh, the Indo-Burmese borderlands and Burma. The sheer dearth of archaeological research in these areas leaves entirely open the possibility that rice cultivation may have originated in this region. We might expect to find traces of ancient farming communities better preserved in the hill tracts surrounding the Brahmaputran flood plains than on the fertile fields themselves, although the earliest rice-based cultures may first have developed on those very flood plains. Perhaps the remains of the first rice cultivating cultural assemblages lie buried forever in the stiły sediments of the sinuous lower Brahmaputra basin or were washed out by the Brahmaputra long ago into the depths of the Bay of Bengal.

Population Genetics and Two Old Linguistic Phyla

Very often language seems to be less ambiguously correlated with the geographical distribution of genetic markers in the populations speaking the languages in question than either genes or languages can be
correlated and contrasted with the fragments of material culture that happen to have resurfaced unscathed from the sands of time. So, can genes and languages generally be correlated and contrasted with each other in a more meaningful way? On the one hand, the linguistic ancestors of a language community were not necessarily the same people as the biological ancestors of that community. On the other hand, the genetic picture often shows sexual dimorphism in linguistic prehistory. Some languages appear to be mother tongues, whereas others show up as father tongues.

In Baltistan, in what today is northern Pakistan, the phonologically highly conservative local Tibetan dialects appear to correlate with the predominantly Tibeto-Burman mitochondrial DNA, which reflects the Balti community’s maternal ancestry (Poloni et al. 1997, 2000; Zerjal et al. 1997; Quintana-Murci et al. 2001; Qamar et al. 2002), whilst the intrusive paternal Y haplogroups from the Near East appear to correlate with the forcible conversion of the area to Islam in the fifteenth century. By contrast, the ‘Father Tongue hypothesis’ may apply to the spread of Indo-Aryan into the Indian subcontinent and, further east, to the spread of Sinitic during the Han demographic expansion. In fact, a likely correlation between the linguistic affinity and the Y haplogroup of a population appears to be a more widespread phenomenon.

At many times and in many places in prehistory, the father tongue may have been the guiding mechanism in language shift. The dynamics of a process whereby mothers passed on the language of their spouses to their offspring also has major implications for our understanding of language change. If the language shift giving rise to the Sinitic languages and perhaps also the eastward spread of Indo-Aryan speech across northern India took place in this way, then such languages may have begun as languages belonging to another phylum until they reached the stage currently attained by Michif. In origin at least, Michif is genetically an Alqonquian language that was spoken by women who relexified the language with the French spoken by their husbands to such an extent that the genetic affinity has nearly been obscured (Bakker 1992, 1994; van Driem 2001:169-73). If the process of relexification were to continue beyond the stage attained by Michif, then a language could conceivably change its genetic affinity even though the dynamics of the process would introduce a discontinuity with its past. If such a process took place, could it ever be reconstructed linguistically?

Additionally, though these findings have not yet reached a wider public, geneticists have increasingly been identifying single nucleotide polymorphisms on the autosomes that are diagnostic for geographical and racial affinity and probably reflect salient episodes of our collective population prehistory. When we turn now to Tibeto-Burman and Austroasiatic, however, to the present state of our knowledge the predominant Y haplogroups in the populations tested to date correlate best with the linguistic affinity of those communities. In other words, both Austroasiatic and Tibeto-Burman may, on the grander scale over time, have preponderantly, albeit not exclusively, spread by the Father Tongue mechanism. By this interpretation, the shared Y haplogroup is assumed to reflect the founding dispersal of the language family.

A few of the pioneering genetic assays of Tibeto-Burman populations inside China ventured some plausible claims. The reduced polymorphism of northern populations of East Asia, which represent a subset of the haplotypes found in southern populations, was taken to reflect the peopling of the north after the Ice Age (Su et al. 1999). By contrast, the high frequency of H8, a haplotype derived from M122C, was seen as reflecting a genetic bottleneck effect that occurred during an ancient southwesterly migration about 10,000 years ago, suggesting a demic diffusion at the outset of the Neolithic (Su et al. 2000, Ding et al. 2000, Shi et al. 2005). Another study suggested that Han Chinese did not originate in the Yellow River basin but had more recently migrated to this area from southwestern China (Chü et al. 1998).

Comparison of various haplogroup frequencies exhibited by Tibetans vs. Tujia, Bai and Lolo-Burmese groups showed all these Tibeto-Burman groups to have a high frequency of the Y-chromosomal haplogroups O3* and O3*, with the average hovering approximately around 40%. These findings were interpreted as supporting a male-biased infiltration from the Bodish area in Amdo into Yunnan and Hunan about two and a half millennia ago. However, ‘the less drastic bias between male and female lineages’ suggested that these putative southward migrations ‘likely occurred with the involvement of both sexes rather than as conquests involving expedition forces primarily consisting of male soldiers’ (Wen et al. 2004).

Interestingly, genetic studies often appear largely to corroborate the long-standing intuitions of linguists, historians and ethnographers working in the region and even to support their previously published models of ethnonlinguistic prehistory. Yet these Chinese studies are limited by the fact that most Tibeto-Burman language communities and even most branches of the language family are exclusively represented outside of China. The picture of the Tibeto-Burman past has been rendered far
Figure 9.4. The portion of the Y chromosome phylogenetic tree relevant to the Father Tongue hypothesis, with regard to the Austroasiatic and Tibeto-Burman linguistic phyla, provided by Mark Jobling and Emma Parkin.

more complete by findings of our own research team, which has conducted the most extensive sampling of Tibeto-Burman populations in the Himalayan region (Kraaijenbrink et al. 2006a, 2006b; Parkin et al. 2006a, 2006b). These findings have allowed us to make novel inferences about the population prehistory of Tibeto-Burman, and has also uncovered some wholly unexpected finds, such as the genetic affinities of the Black Mountain Monpa. Additional findings from northeastern India, Tibet and Burma will enable us to identify the possible molecular correlates corresponding to more episodes in the spread in Asia of ancient Tibeto-Burman language communities.

The Y-chromosomal haplogroup O3e (M134), which seems to tag Tibeto-Burman language communities within and outside of the Himalayan region, may very well have a time depth of at least 10,000 years (figure 9.4). If so, this would put us in a time frame that compels us to consider in which localities the ancient Tibeto-Burmans may have dwelt during the last glacial maximum. Could the southern flank of the Himalayas have served as a vast refuge area during the last Ice Age, or did the early Tibeto-Burmans at this time dwell in some area to the northeast or to the east of the Himalayas? Palaeoclimatologists remain divided into rival camps on the question of the extent of the glaciation in the Himalayas in recent prehistory, e.g., Kuhle (1985, 1986, 1987, 1988a, 1988b, 1990a, 1990b, 1991, 1995, 1997, 1999, 2005), Thompson et al. (1989), Lehmkuhl (1995), Schäfer et al. (2002), Owen et al. (2002), Owen et al. (2003), Jin et al. (2005), Lehmkuhl and Owen (2005), Vandenberghe (2007). By virtue of the sheer scale and diversity of the topography, the Himalayas harbour a panoply of climatological enclaves and sheltered areas with their own specific microclimate. Did the Himalayas offer hospitable ice-age refugia to the ancestral Tibeto-Burmans?

In seeking an answer to this question, northeastern India still remains scantily documented and poorly understood in archaeological terms. Much of the major work in this region was carried out over a generation ago and has been discussed and referenced in my handbook (van Driem 2001). Yet most such work pertains to the Neolithic, a later period which does not help shed light on the issue of possible ice-age refuge habitats. By contrast, a partial answer emerges from the far more complete picture that we have for Nepal thanks to the posthumously published work of Gudrun Corvinus, who, tragically, was murdered in her own flat in Pūne by the henchman of a local real estate criminal. Ironically, she had chosen to live in Pūne to avoid the mayhem, extortion and murder that Maoists were inflicting indiscriminately on their fellow countrymen in the regions where she conducted her work.

On the basis of a lifetime of palaeontological and archaeological research in the sub-Himalayan tracts of Nepal, Gudrun Corvinus (2007) developed the theory that the Early and Middle Palaeolithic and microlithic traditions in the western Terai and Śivālik derive from contemporaneous cultures further south on the Indian subcontinent in a period still characterised by low population density. However, the Pāṭu culture in the eastern Śivālik and the Brākhuṭī culture in the western Terai show a later influence emanating westward across the Himalayan foothills from Southeast Asia in the Late Pleistocene and Holocene, with the archaeological record suggesting an increase in population density at the end of the last glacial maximum roughly coincident with the flourishing of the Brākhuṭī culture.

Corvinus’ comprehensive pioneering work in Nepal has yielded good stratigraphies and optical and infra-red stimulated luminescence datings of key sediment layers based on work at sites along the Rāto Kholā south of Sīndhuli in Mahottarī district in the eastern Śivālik, sites at and near Sāṭpāṭi hill east of Bhaīrahaṇa in Lumbinī district and various sites in the Dāṅg and Deukhuri basin. The archaeological record suggests a long prehistory of human habitation. For example, the alluvial and colluvial hill-wash deposits composing the Gidhiniya and Babai formations in the Tuī basin in Dāṅg district have yielded abundant lithic material from the Early, Middle and Late Palaeolithic periods as well as from the Mesolithic and Neolithic periods. An Early Palaeolithic presence of the Acheulian tradition in South Asia in the early Middle Pleistocene
is indicated by Acheulian bifacials and flake tools industry at Gadarī in Dāng in the western Terai and at Sātpā in central Nepal just west of the Nārāyāni River. Alluvial terrace deposits at the Arjun site in the badlands on the left bank of the Arjun River show Levallois technology appearing in the Middle Palaeolithic.

At the same time the foothills of the central Himalayas show a complex cultural prehistory. Nepal straddles the so-called Movius line and represents a transition zone between two traditions of lithic technology. More Occidental technocomplexes as represented by artefacts of the Acheulian and Levallois traditions culminating in late Palaeolithic flaking and microlithic industries. The Pātu culture in Mahottarī district in eastern Nepal, however, stands outside of the main subcontinental context and shows clear connections with coeval technocomplexes in the forested habitats of Southeast Asia. Pātu technology is characterised by cobbled-tools and less distinctly retouched small flake tools and, in the Mesolithic context, by macrooliths such as adzes and cobbled tools as well as the unifacial, flat-based and steep-edged tools called sumatraliths, which clearly suggest an affinity with the Hoabinhian.

Yet in later strata, the Brāhkhūṭ culture in the Tūi valley in the Dāng and Deukhurī basin preserves core scrapers and unifacial choppers manufactured by stone-knapping techniques that would appear to have been very particular to that area of present-day Nepal and quite distinct from more Occidental technocomplexes of the period. At the same time, the large flake core industry at Brāhkhūṭ comprises sumatraliths, some high-crested and some in the shape of a horseshoe, whilst adzes of the Oriental type found at Pātu are lacking at Brāhkhūṭ. The precise chronological relationship between the Pātu culture in the eastern Sīvālik and the Brāhkhūṭ culture in the western Terai has not yet been clarified, but heavy-duty stone tools would appear to have been de rigueur in the ancient forests of the Terai and Sīvālik hills.

In turning from the archaeological record to the molecular biological legacy, the population genetic data in the Himalayan region correspond with the linguistic divide more sharply than in most other parts in the world. Whilst gradients of biological markers often flow fuzzily across deep linguistic boundaries, in the Himalayas both the genetic and linguistic divides between Tibeto-Burman and Indo-European remain sharp. Population geneticists also corroborate what linguists and ethnographers have long known, namely that the Himalayas themselves do not constitute the real geographical divide between Tibeto-Burman and Indo-European. Rather the divide runs roughly through the sub-Himalayas or the Terai. Similarly, there is a marked discontinuity between Neolithic and Bronze Age traditions up in the hills and those down on the plains. Yet such later archaeological assemblages appear, to our current state of knowledge, to be younger than the population genetic divide, and perhaps also to the linguistic one.

The genetic divide between Tibeto-Burman and Austroasiatic in the region shows a far more complex structure than the clear line demarcating Tibet-Burman from Indo-European. To the south, in the Brahmaputran basin and the Indo-Burmese borderlands, some of the spread of Tibet-Burman may have been at the expense of indigenous Austroasiatic populations whom the Tibet-Burmans assimilated linguistically. The Y haplogroup O2a (M95) is represented at a frequency of 77% in Austroasiatic groups in India and 47% in Tibet-Burman groups of northeastern India (Sahoo et al. 2006). This pattern could suggest that Tibet-Burman paternal lineages partially replaced indigenous Austroasiatic lineages in the northeast of the Indian subcontinent in the distant past, and that Austroasiatic populations preceded the Tibet-Burmans in this region, as linguists and ethnographers have speculated for over a century and a half (van Driem 2001). The geographical extent of the Bodo-Kok languages and the shallow time depth of this sub-branch of Tibet-Burman might represent the linguistic corollaries of such partial genetic replacement. My earlier arguments regarding the locations and geographical proximity of a more littoral Austroasiatic homeland and a more montane Tibet-Burman homeland, both within or near the northeastern portion of the Indian subcontinent (van Driem 2006, 2007), continue to be supported by more recent genetic studies involving the Y-chromosomal haplogroups O2a and O3, e.g., Kumar et al. (2007).

Austroasiatic is an old language family, and we would expect the population history of this family to be at least as complex as that of Tibet-Burman, if not more so. In future, more detailed and careful correlation of linguistic and population genetic findings based on more fine-mesh population genetic sampling may enable us to reconstruct early language contact situations and ancient cases of language shift and linguistic intrusions that might, for example, account for the phenotypical differences readily observable between Munda speakers as opposed to Khasi-Khmuic and Mon-Khmer language communities as well as between Aslian negrito populations, Aslian non-negrito populations and the Nicobarese. Somewhat in parallel with such somatological
observations by ethnographers, linguists have long observed corresponding typological differences between various branches of Austroasiatic.

Donegan and Stampe (1993, 2004) suggest that Austroasiatic spread from the Indian subcontinent to Southeast Asia but argue that paradoxically the synthetic head-final typology of Munda languages resulted from an innovative process of drift which unfolded within South Asia after the linguistic ancestors of modern Khasi-Khmuic and Mon-Khmer language communities migrated towards Southeast Asia. In their view, the typological change in Munda was triggered by a prosodic shift to a falling rhythm, whereas the analytic head-initial typology observed in Khasi-Khmuic and Mon-Khmer languages reflects the more original Austroasiatic state of affairs. By contrast, Zide and Anderson (1999, 2003) have argued that Munda verbal morphology is a conservative retention, and that older Austroasiatic grammatical systems were secondarily lost in the Khasi-Khmuic and Mon-Khmer languages of Southeast Asia.

Yet if the Father Tongue hypothesis holds true for the spread of Austroasiatic into South Asia (van Driem 2007), then this outcome would vindicate Robert von Heine-Geldern’s view of the Munda as the result of the ‘Einwanderung mongolider austrasiatischer Stämme in Vorderindien’ and of their ‘Mischung mit Dravida und Urbevölkerungselementen’ (1928, 1932). In that case, the typological divergence between Munda as opposed to Khasi-Khmuic and Mon-Khmer, lucidly discussed by Donegan and Stampe, may be the result of the adoption of an intrusive paternal tongue by indigenous pre-Austroasiatic populations of the Indian subcontinent. The apparent Munda penchant for a falling prosodic rhythm might then be just one residue of a far-reaching action de substrat.

If, however, the Father Tongue hypothesis holds true for Austroasiatic, and Zide and Anderson are correct, then the Munda descendants of the linguistically assimilated South Asian indigenes may have more faithfully preserved the original Austroasiatic morphology and grammar of the early bearers of the Y-chromosomal haplogroup O2a than is now seen reflected in modern Khasi-Khmuic and Mon-Khmer languages, which underwent divergent areal developments specific to Southeast Asia.

On the other hand, the Father Tongue hypothesis may very well not apply in all cases for the biological ancestry of Austroasiatic language communities, just as language spreading solely via the paternal line cannot account for the linguistic identity of all Tibeto-Burman populations, e.g., maternal Balti vs. paternal Hân. More decisive insights into both the historical linguistics and historical typology of Austroasiatic and Tibeto-Burman, newly identified single nucleotide polymorphisms on the autosomes, ethnolinguistically informed fine-mesh genetic assays of Aslian negrito populations, Aslian non-negrito populations, the Nicobarese and peoples of Tibet and Burma, and insights from other disciplines may help us to retrieve more undiscovered bits and pieces of prehistory that may not have been irretrievably lost.
BIBLIOGRAPHY


— 1997. New findings concerning the Ice Age (Last Glacial Maximum); Glacier cover of the East-Parim, of the Nanga Parbat up to the Central Himalaya and of Tibet, as well as the age of the Tibetan inland ice. *GeoJournal* 42(2-3): 87-257.


