

# Reflections on the Ethnolinguistic Prehistory of the Greater Himalayan Region

GEORGE VAN DRIEM (Leiden)

's Werelds twee grootste taalfamilies komen elkaar tegen in de Himalaya. In de omstreken bevinden zich ook verscheidene taalgemeenschappen die nog eens vier grote taalfamilies vertegenwoordigen. Verder herbergt het gebied drie geïsoleerde talen. De taalwetenschap, de archeologie en het erfelijkheidsleerkundige onderzoek naar menselijke bevolkingsgroepen ontsluiten uiteenlopende hoofdstukken uit de voorgeschiedenis van de mensheid. Hoe kunnen de reconstructies van het verleden die deze drie disciplines opleveren op zinvolle wijze in elkaar gepast worden om tot een samenhangend en geloofwaardig beeld van ons verleden te komen?

## 1 Language phyla of the Himalayan region

The world's two most populous families of languages meet in the Himalayas. These are the Indo-European phylum, to which languages such as English and Bengali belong, and Tibeto-Burman, which includes Cantonese and Mandarin Chinese. In addition to these two great language families, Daic *alias* Kra-Dai, Austroasiatic, Altaic and Dravidian language communities skirt the periphery of the Himalayan region. For example, Kyrgyz is spoken in the Tiānshān and Khasi in the Meghālaya. The Ahom language was once spoken in northeastern India, where scattered Daic language communities are still settled today. Dhangar and Jhangar, dialects of Kurukh or Uraon, are spoken in Nepal's eastern Terai. Yet the Himalayas would appear to be as peripheral to our understanding of the prehistory of Kra-Dai, Austroasiatic, Altaic and Dravidian as these four language families are peripheral to the Himalayas.

The real crux to the ethnolinguistic prehistory of the greater Himalayan region are the two great language families Tibeto-Burman and Indo-European and, even more so, the several language isolates of the region. 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 family tree structure of the Tibeto-Burman phylum is likewise essentially rake-like, a situation for which I introduced the metaphor of fallen leaves.

Two magnificent language isolates are spoken, far apart from each other, in the heart of the Himalayas. Burushaski is spoken by some 80,000 people in the high alpine valleys of Hunza-Nager and Yasin near Gilgit in Pakistan. The nearly extinct Kusunda lingers on in the memories of perhaps no more than two dozen elderly speakers. The remaining ethnic Kusunda reside in Pyūṭhāñ, Dāng and Tanahū districts in western Nepal. A third language isolate, Nahali, survives out in the far periphery of the Himalayan region, in the Gavilgaḍh Hills in the Indian state of Chattīsgaḍh.

For Burushaski an ancient relationship with the Yenisseian languages has been argued, i. e. Hyde Clark (1870), Toporov (1969, 1971), van Driem (2001). I named the hypothetical language family ‘Karasuk’ after the archaeological complex in western Siberia which Členova (1972) identified with the ancient Yenisseians. The Karasuk cultural assemblage broke up, the main movement following the Yenissei northwards downstream to the area around today’s Krasnojarsk, and the other movement proceeding west onto the eastern steppe into territory inhabited by bearers of the Andronovo culture. I also proposed the alternative, more neutral name ‘Greater Yenisseian’ for this hypothetical family, comprising Burushaski and Yenisseian.

Assuming the veracity of Členova’s theory, I suggested that the main northward movement gave rise to the historically attested Yenisseian language communities. The other Karasuk culture bearers, who inhabited the northwestern Minusinsk basin, moved southwest in the seventh and sixth centuries BC and gradually merged with the Andronovo Bronze Age culture of Central Asia. Many scholars associate the Andronovo complex with the ancient Indo-Iranians before their southward migration east into the Subcontinent and westward across the Iranian plateau. This hypothetical scenario would explain the location of the Burusho in Hunza-Nager and Yasin.

Kusunda has recently become a fresh recruit for an expanded version of the Indo-Pacific theory (Whitehouse *et al.* 2004). Since the late eighteenth century, scholars have entertained the idea that all Asian negrito peoples or *ozeanischen Neger* shared a deep common ethnolinguistic origin. The linguistic supposition underlying this old hypothesis was ‘disproved’ in 1831 by Samuel Rafinesque in the first ever application of lexicostatistics, which Rafinesque himself invented (Dumont d’Urville 1834). None the less many linguists and anthropologists continued to entertain the hypothesis, e.g. Finck (1909), Skeat and Blagden (1906). Some called this assumption the ‘Pan-Negrito Theory’ or ‘Indo-Pacific’. The late Joseph Greenberg (1971) revitalised the idea, and his pupil Merritt Ruhlen and three colleagues have recently attempted to include the Kusunda language into Indo-Pacific, viz. Whitehouse *et al.* (2004).

The recent comparative essay is of questionable linguistic merit, however. Data are presented not as in the sources, but in newer international phonetic notation. In the process, *ad hoc* interpretations and choices had to be made, and it is my impression that the choices have had the effect of making forms look more similar than they might be in reality. This unnecessary standardisation can be especially unreliable when retranscribing Papuan forms with no standardised orthography. The sources for forms adduced by the authors from little-known languages are provided only in the ‘supporting data’ downloadable from the website. References should have been given a more prominent place, as is customary in the linguistic literature, since recourse to the original sources is required for any detailed critical appraisal of the study in question, particularly because the comparison deals exclusively with poorly documented languages.

More fundamentally, the methodological issue of comparing look-alikes in modern languages without any understanding of the historical phonology of the languages in question has been dealt with in numerous critiques of long-range linguistic comparison. It will always be tempting to relate forms like Greek θεός ‘god’ and Latin *deus* ‘god’, first adduced as Indo-European cognate etyma by Marcus van Boxhorn in 1647, but the two are not cognate at all, as Müller argued in 1865.

I need not belabour the methodological inadequacies and pitfalls of false cognates and long-range comparisons in general, since so many others have detailed the arguments eloquently in the literature. In my assessment, the

case for Indo-Pacific remains unconvincing, and the same applies *a fortiori* to the proposed inclusion of Kusunda in this hypothetical linguistic construct. However, I am pleased that the authors were able to state their case in a prominent non-specialist journal so that their arguments are publicly available, since the article would not have passed muster for publication in a standard linguistic journal.

Various theories have been proposed to associate Nahali with some recognised linguistic phylum, notably Austroasiatic. The arguments are discussed at some length in my handbook (van Driem 2001). None of these hypotheses has been convincingly demonstrated. Keralapura Shreevinasaiah Nagaraja (2006) has since 1998 gathered linguistic data in the Nahali language community, where he spends about a fortnight each year. His data and the material collected by earlier scholars corroborate the conservative view that Nahali has no demonstrable genetic affinity with any known linguistic phylum. In short, Nahali does not appear to be a member, peripheral or otherwise, of the Austroasiatic phylum. Once Korku loan words and other obvious borrowings are removed, Nahali looks very much like a genuine language isolate representing the sole vestige of a distinct pre-Austroasiatic, pre-Dravidian and pre-Indo-European ethno-linguistic stratum.

## 2 Archaeology, linguistic intrusions and biological ancestors

The fundamental issue is whether the spread of a recognisable Neolithic and Bronze Age assemblage can actually ever be taken to reflect the spread of a language and so of a language family. For example, was the expansion of Bodic into the Himalayas associated with the sudden appearance of colonial exponents of the Mǎjiāyáo Neolithic in eastern Tibet at mKharro and in Kashmir at Burzahom at the same time that the core area in Gānsù shrank during a period of climate change between the Mǎjiāyáo phase (2700–2300 BC) and the Bānshān phase (2200–1900 BC) of the Mǎjiāyáo sequence? This is a scenario which I outlined and argued in several earlier publications (van Driem 1998, 2001, 2002).

More recently, I have outlined a number of 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). 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. Linguistic palaeontology invariably raises complex issues, and elsewhere I have discussed some of 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 稷 *btsik* in the Yellow River basin and Lhokpu<sup>1</sup> *cǎ'kto* 'foxtail millet' in modern southwestern Bhutan (van Driem 2006).

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 (Sù *et al.* 1999), whereas 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 (Sù *et al.* 2000, Ding *et al.* 2000, Shi *et al.* 2005). Another study suggested that Hàn Chinese did not originate in the Yellow River basin but had more recently migrated to this area from southwestern China (Chǔ *et al.* 1998).

Comparison of various haplogroup frequencies exhibited by Tibetans vs. Tǔjiā, Bái and Lolo-Burmese groups showed all Tibeto-Burman groups to have a high frequency of the Y-chromosomal haplogroups O3e and O3\*, with the average hovering approximately around 40%. The findings were interpreted as supporting a slightly male-biased infiltration from the Bodish area in Amdo into Yúnnán and Húnán about two and a half millennia ago, though '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).

---

<sup>1</sup> 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 matrilocality 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 uxori-local marriage such as the modern Lhokpu and Gongduk of Bhutan.

However useful these genetic studies are, they were 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 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). Our team has identified markers which we believe to be specifically correlated to the spread in Asia of Tibeto-Burman language communities. Our results will be published in due course in an appropriate population genetics journal, and I am not at liberty to detail the findings here.

Far away to the south, in the Brahmaputran basin and the Indo-Burmesse borderlands, however, some of the spread of Tibeto-Burman may have been at the expense of indigenous Austroasiatic populations who were assimilated linguistically. The Y haplogroup O2a is represented at a frequency of 77% in Austroasiatic groups in India and 47% in Tibeto-Burman groups of northeastern India (Sahoo *et al.* 2006). This patterning could suggest that Tibeto-Burman paternal lineages may have partially replaced indigenous Austroasiatic lineages in the northeast of the Indian Subcontinent and that Austroasiatic populations preceded the Tibeto-Burmans in this area, as linguists and ethnographers have speculated for over a century and a half.

For Indo-European, Mallory (1989) has provided a good recapitulation of the many competing proposed correlations between archaeology and linguistics, and since then many alternative proposals and speculations have been elaborated on this broad interdisciplinary theme. Archaeology reflects what we have been able to glean about the material culture of past communities. In fact, how often can we be quite 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?

Very often language seems to be less ambiguously correlated with the geographical distribution of genetic markers in the populations speaking the languages in question. Even so, the linguistic ancestors of a language community were not necessarily the same people as the biological ancestors of that community. For example, geneticists have until now looked in vain for markers which identify Hungarians as a Uralic language community.

Even the Y-chromosomal haplogroup N-TatC (N43), which is found at a high frequency throughout Uralic language communities, does not seem to be prevalent in Hungary.<sup>2</sup> Rather, Hungarians look genetically quite a lot like a Western Slavic language community, and there is little trace at the moment of a Uralic genetic signature (Tambets *et al.* 2001). Perhaps the Magyars who penetrated into Pannonia introduced a Uralic language but not much else. Perhaps Uralic Y chromosome lineages simply died out in Hungary for whatever reason.

We invariably get all of our DNA from our biological parents, but only in most cases is our native language also that of our parents. So, notwithstanding the probabilistic correlation between languages and genes, the discrepancies between the two versions of prehistory can tell us at least as much about what went on in the past as the grand correlations. Population genetics tells us about the spread of genotypes, whether this is caused by circumstances of origin, migration or natural selection. Whatever the case may be, the Hungarian language constitutes incontrovertible linguistic evidence that the Magyars came to Pannonia. The historically attested Magyar linguistic intrusion may now be genetically invisible, but the Hungarian language is linguistically very much in evidence. Given the extremely low population numbers which characterised prehistoric human demography, evidently no colossal throng of people was needed to effectuate a linguistic incursion.

The genetic picture, moreover, sometimes 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 local Tibetan dialects are perhaps the most conservative of all Tibetan languages, preserving consonant clusters retained in Classical Tibetan orthography but which have been lost or have succumbed to mergers in most other Tibetan dialects. Yet the Balti abandoned the Tibetan script after they were converted to Islam in the fifteenth century.<sup>3</sup> Paradoxically, the old consonant clusters ceased to be pronounced as such in most areas throughout Tibet where the conser-

---

<sup>2</sup> Outside of Uralic speech communities, the haplogroup is also found at a high frequency amongst the Yakut, Even and Tuva.

<sup>3</sup> In recent years, native activists have begun reintroducing the Tibetan script, e. g. on shop signs, somewhat to the displeasure of central government authorities.

vative indigenous orthography representing these phonological segments remained in use. Genetic studies of the Balti populations show intrusive Y haplogroups from the Near East, whereas the mitochondrial DNA of the Baltis is predominantly Tibetan mtDNA (Poloni *et al.* 1997, 2000, Zerjal *et al.* 1997, Quintana-Murci *et al.* 2001, Qamar *et al.* 2002). So, the religion of the Balti would appear to be a paternal heritage, whilst the languages that they speak are literally mother tongues.

Genetic studies have suggested that the distribution of Indo-Aryan language communities in northern India patterns well with intrusive Y haplogroup frequencies emanating from the northwest, reflecting what many linguists and archaeologists had long thought about Indian prehistory. The picture of an Aryan invasion emerging from the R̥gveda, in the words of Mortimer Wheeler, ‘constantly assumes the form of an onslaught upon the walled cities of the aborigines’, i.e. the *purās*, and the Aryan god Indra is a *puram̐dara* ‘destroyer of aboriginal forts’, who shattered ninety such strongholds (1966, 1968). Many scholars have connected this destruction of aboriginal fortresses and the conquest of subjugated Dasyus recounted in the Aryan hymns to the extinguishing of the Indus Valley civilisation. At any rate, the activities depicted were a predominantly male occupation.

Genetic studies have suggested that the Y haplogroups L, R1a and R2 spread from the northwest along with Indo-Aryan language across northern India and to Ceylon, whereas mitochondrial lineages prevalent in India are overwhelmingly indigenous to the Subcontinent (Kivisild *et al.* 1999a, 1999b, Wells *et al.* 2001, Cordaux *et al.* 2003, Kivisild *et al.* 2003, Baig *et al.* 2004, Cordaux *et al.* 2004, Metspalu *et al.* 2004, Quintana-Murci *et al.* 2004, Thangaraj *et al.* 2005). At the same time, the spread of Indo-Aryan languages unambiguously attests to an ancient linguistic intrusion into the Subcontinent from the northwest. So, were Vedic and Avestan introduced as father tongues?

A recent study (Sahoo *et al.* 2006) attempts to challenge the Y chromosome picture. The study is a major leap forward, but the sampling is still coarse, and the survey neglects to systematically distinguish between Turks, Kurds and other language communities in the Near East and between Indo-Iranian and Turkic language communities in Central Asia. A fine-mesh and more ethnolinguistically informed sampling remains a realisable goal. More crucially, the reasoning in Sahoo *et al.* (2006), edited by

Colin Renfrew, omits to take note that Central Asia saw major incursions of Altaic populations in historical times, and that an ethnolinguistically low-resolution survey of present Central Asia Y chromosomal genography cannot be presumed to reflect the genography of the region during, say, the Andronovo Bronze Age culture and the Bactria Margiana archaeological complex.

In fact, the probable replacement of Y chromosomal lineages during the Altaicisation of Central Asia is consonant with the team's observation that the Y haplogroups E, I, G, J\* and R1\*, which have a combined frequency of 53% in Turks of Asia Minor and 24% in Central Asia, are virtually absent in India, except for a trickling of R1\*. Also absent in India are haplogroups C3, D, N and O, which are 'specific to Central Asia', where they have a combined frequency of 36%. Likewise, the complete absence in India of the derived C3 lineages, which account for over 95% of the C haplogroup variation in Central Asia, 'cannot be ascribed to a recent admixture from the north' (op. cit. 845). At the same time, the J2 haplogroup, which appears to emanate from the Arabian Peninsula and, unlike haplogroups N and R1a, attains no high frequency in Ceylon, 'indicates an unambiguous recent external contribution, from West Asia rather than Central Asia' (op. cit. 87), and indeed this gradient probably reflects the historically attested male-borne eastward spread of Islam.

Therefore, what I first called the 'Father Tongue hypothesis' in Taipei in 2002, based largely on the work of Poloni *et al.* (1997, 2000), may apply to the spread of Indo-Aryan into the Indian subcontinent and, as I have argued elsewhere (van Driem 2006), may perhaps also hold for the spread of Sinitic during the Hân demic expansion. Though there are numerous contrary cases such as the Tibetan mother tongues of Baltistan, as a general principle the Father Tongue hypothesis may at many times and in many places in prehistory have been an important mechanism in language shift.

The dynamics of a process whereby mothers passed on the language of their spouses to their offspring 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 Algonquian 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: 169–173). 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. Can such a process ever be reconstructed linguistically?

A general issue is that of 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. So, can genes and languages generally be correlated and contrasted with each other in a more meaningful way than either can be with the fragments of material culture that happen to have resurfaced unscathed from the sands of time?

For example, the Y chromosome haplogroup which seems to tag virtually all Tibeto-Burman language communities in and outside of the Himalayan region may very well have a time depth of at least 10,000 years. This finding suggests that the southern flank of the Himalayas could have been a vast refuge area during the last Ice Age. If so, does this event correlate with any transition that has left visible traces in the archaeological record? Which archaeological transition or modern genetic gradient can we relate with confidence to an ancient linguistic intrusion or to the prehistorical spread of a language family?

In the Himalayan region, the population genetic data correspond with the linguistic divide more sharply than in most other parts in the world. Often gradients of biological markers flow fuzzily across deep linguistic boundaries. In the Himalayas, the genetic and linguistic divide match up well. Population genetics also corroborates what linguists and ethnographers have long known, namely that the Himalayas themselves do not constitute the geographical divide between Tibeto-Burman and Indo-

European. Rather the divide runs approximately 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 these archaeological assemblages appear, to our current state of knowledge, to be much younger than – and therefore posterior in time to – the population genetic divide, and perhaps also to the linguistic one.

### **3 The population genetics of communities speaking language isolates**

Nahali, Burushaski and Kusunda are linguistically as singular in the Asian context as Basque is in the European context. For centuries now it has been mooted that the Basques may represent the linguistic vestige of an ancient population layer in Europe, whether their linguistic ancestors were already on the Iberian peninsula before the last Ice Age or were the first to colonise Europe after the retreat of the glaciers. Similarly, the genographic study of communities speaking language isolates in the Asian heartland may identify vestiges of earlier population strata.

Recently, population genetic research has begun on the Nahali language community conducted by friends at the Centre for Cellular and Molecular Biology in Hyderabad, who have collected 32 samples in the field. The mtDNA data are already in, and they look exciting. The Y chromosome and autosomal data are still being processed, and new insights are trickling in as I write this piece, i.e. March 2007. The findings of the Indian genetics team will be published in the foreseeable future, and the importance of their Nahali findings to our understanding of the population prehistory of the Subcontinent is great.

As for the Burushaski language community, Qamar *et al.* (2002) tell us that the short tandem repeats (STRs) on the Y chromosome ‘do not support’ the utterly incredible hypothesis that the Burusho are the descendants of Alexander the Great’s army in Pakistan, an outlandish idea that was evidently once put forward by John Biddulph in 1882 but has never been entertained seriously by anyone since. The predominant Y haplogroups found in the Burusho are R1a and R1b, neither of which is remarkable in the South Asian context.

In a study of the paternal lineages of native Siberian populations, the sole surviving Yenisseian language community, Ket, was shown to be the

furthest outlier in a multi-dimensional scaling (MDS) plot of native ethnic groups of Siberia based on genetic  $\Phi_{ST}$  distances. Closest to the Ket are the Selkup. Once known as the Selkup speaking Ostiak Samoyeds, this Uralic language community of the Southern Samoyedic branch used to live with the Ket in a commensal relationship that has been described in the ethnographic literature since the eighteenth century. The Ket also exhibit the lowest heterozygosity of any native Siberian ethnic group. The predominant Y-chromosomal haplogroup amongst the Ket and Selkup is Q, which was found at a frequency of 93.8 % and 66.4 % respectively in the two populations. The extremely high frequency of this haplogroup is attributed ‘to intergenerational genetic drift coupled with founder effects’, a supposition supported by the extremely low levels of Y-chromosomal short tandem repeat (STR) diversity associated with haplogroup Q (Karafet *et al.* 2002: 784). Haplogroup Q is found only in low frequencies in other northern Siberian populations.

Therefore, there is no evidence of a specific genetic link between the paternal lineages of the Burusho and Ket, unless we venture to push the date for the hypothetical Greater Yenisseian linguistic phylum back to the rather distant time depth of haplogroup P, which was ancestral to both haplogroups Q and R. A father tongue hypothesis for Greater Yenisseian involving haplogroup P would bring us to a time depth much before the last Ice Age but still long after anatomically modern humans had first begun to people the Subcontinent. What empirically supported speculations would comparative historical linguists be able to contribute about such time depths? What types of linguistic evidence would linger on after such a vast stretch of time, and what shape would this evidence take?

Could a great time depth explain why the surviving correspondences between Yenisseian and Burushaski are limited to structural particulars of their flamboyant verbal morphology and a few conjugational specifics? The two Ket verbal agreement systems, named D and B, correspond to the distinction in agreement patterns between Burushaski ‘*d*-Verben’ and other verbs. Yenisseian and Burushaski biactantial verbal agreement marking are both gender-dependent in the third person, and the peculiar gender systems of Yenisseian and Burushaski too are similar. Possible vestiges of cognate morphology include the Burushaski second person singular patient-subject prefix <gu- ~ gó-> vs. Ket second person agent-subject

prefix <ku- ~ k- ~ gu- ~ γu->, Burushaski first person singular patient-subject prefix <a- ~ á-> vs. Ket first person singular prefix <ba- ~ bə- ~ va- ~ və->, Burushaski non-feminine third person singular patient-subject prefix <i- ~ é-> vs. Ket inanimate third person singular prefix <i- ~ u->, Burushaski third person plural patient-subject prefix <u- ~ ó-> vs. Ket third person plural prefix <bu- ~ vu->, Burushaski plural agent-subject suffix <-en> vs. Ket plural subject-agent suffix or infix <-(V)n ~ -(V)ŋ ~ -(V)ŋ>.

So, should these correspondences convince us of the reality of Greater Yenissonian as a linguistic phylum? Might Greater Yenissonian be analogous to the case of Chukotko-Kamchatkan, where the evidence for a widely accepted ancient genetic relationship is restricted to vestiges of a common morphological system with little or no compelling lexical correspondences? If the Greater Yenissonian linguistic phylum is accepted as real, then are Burushaski and Yenissonian its only members? Assuming the father tongue hypothesis for the sake of argument, then the global pattern for haplogroups P, R and Q would have to be accounted for in the broader Eurasian ethnolinguistic context, not just for the pieces of the puzzle represented by Burushaski and Yenissonian.

For example, an explanation would have to be found for the prevalence of R haplogroups outside of the Burusho language community. Could it be plausibly argued, for example, that the R1a haplogroup frequency in the northwestern portion of the Subcontinent represents a vestige of an ancient Greater Yenissonian paternal lineage persisting in South Asian communities which later were linguistically assimilated by incursive Indo-Aryans? Could Burushaski and Ket be the surviving linguistic vestiges of this antique stratum? In this haplogroup P scenario, the hypothetical 'Karasuk' linguistic phylum would have to date back to a very long time before there ever was such a thing as the Karasuk archaeological complex in western Siberia, so that 'Greater Yenissonian' would be a better name for the hypothetical phylum than the archaeologically inspired 'Karasuk'.

Yet there is still another way of looking at the genetic picture. Just as no Uralic paternal lineage has yet been identified in the Hungarians, an alternative possibility is that, if the ancient Burusho did branch off from the Yenisnians at the time of the Karasuk cultural assemblage and if the Burusho originally once did have a high incidence of the Y-chromosomal

haplogroup Q, then this heritage could subsequently have been obliterated by paternal input from neighbouring Indo-Aryan language communities. Currently, no published studies tell us much about the mitochondrial DNA of the Burusho or Yenisseians.

Turning now to the third language isolate, it has already been mentioned that the Indo-Pacific ethnolinguistic theory is rooted in the ‘Pan-Negrito Theory’ of physical anthropologists of yore. The theory essentially represents what Roger Blench has called ‘a crinkly hair hypothesis’. Phenotypically, however, the Kusunda exhibit no traits that could credibly be qualified as ‘Negroid’ or ‘negrito’. On the genetic front, I am pleased to be able to report that DNA samples were collected from the Kusunda on a recent field campaign. In addition to mtDNA, the Y chromosome was collected from three of the four patrilinear Kusunda clans known from the ethnographic literature, i. e. Śāhī, Siṃha and Sen, but not yet Khān. It is known that the Kusunda sometimes claim Ṭhakurī status, which raises a few questions in view of the role which this subcaste of the warrior caste has played in Nepal’s history. There is much that we might yet come to know about the successive waves of peopling that have washed across the Himalayas.



MAP: Geographical distribution of *Kusunda* and *Burushaski*, two language isolates in the Himalayas, and of *Nahali*, a language isolate in the heart of the Subcontinent

## References

- Baig, M. M., A. A. Khan and K. M. Kulkarni. 2004. 'Mitochondrial DNA diversity in tribal and caste groups of Maharashtra (India) and its implication on the their genetic origins', *Annals of Human Genetics*, 68: 453–460.
- Bakker, Peter (i. e. Pieter Jan Bakker). 1992. '*A Language of Our Own*', the *Genesis of Michif, the Mixed Cree-French Language of the Canadian Métis*. Universiteit van Amsterdam: doctoral dissertation.
- Bakker, Peter (i. e. Pieter Jan Bakker). 1994. 'Michif, the Cree-French mixed language of the Métis buffalo hunters in Canada', pp. 13–33 in Peter Bakker

- and Maarten Mous, eds., *Mixed Languages, 15 Case Studies in Language Intertwining* (Studies in Language and Language Use, 13). Amsterdam: Instituut voor Fundamenteel Onderzoek naar Taal en Taalgebruik.
- Belyi, Vilen V. 1997. 'Rafinesque's linguistic activity', *Anthropological Linguistics*, 39 (1): 60–73.
- Biddulph, John. 1882. *Tribes of the Hindoo Koosh*. Calcutta: Office of the Superintendent of Government Printing.
- van Boxhorn, Marcus Zuerius. 1647. *Antwoord van Marcus Zuerius van Boxhorn op de Vraaghen, hem voorgesteld over de Bediedinge van de tot noch toe onbekende Afgodinne Nehalennia, onlancx uytgegeven. In welke de ghe-meine herkomste van der Griecken, Romeinen, ende Duytschen Tale uyt den Scythen duydelyck bewesen, ende verscheiden Oudheden van dese Volckeren grondelijck ontdeckt ende verklaert*. Leyden: Willem Christiaens van der Boxe.
- Chū Jiāyòu and W. Huang, S. Q. Kuang, J. M. Wang, J. J. Xu, Z. T. Chu, Z. Q. Yang, K. Q. Lin, P. Li, M. Wu, Z. C. Geng, C. C. Tan, R. F. Du and Jīn Lì. 1998.] 'Genetic relationship of populations in China', *Proceedings of the National Academy of Sciences of the United States of America*, 95: 11763–11768.
- Členova, Natalija L'vovna. 1972. *Xronologija pamjatnikov karasukskoj èpoxi* (Materialy i issledovanija po arxeologii, SSSR, N° 182). Moscow: Izdatel'stvo 'Nauka'.
- Cordaux, Richard, Nilmani Saha, Gillian R. Bentley, Robert Aunger, S. M. Sirajuddin and Mark Stoneking. 2003. 'Mitochondrial DNA analysis reveals diverse histories of tribal populations from India', *European Journal of Human Genetics*, 11: 253–264.
- Cordaux, Richard, G. Weiss, Nilmani Saha and Mark Stoneking. 2004. 'The northeast Indian passageway: A barrier or corridor for human migrations?', *Molecular Biology and Evolution*, 21: 1525–1533.
- Crawford, Gary W. 2006. 'East Asian plant domestication', pp. 77–95 in Miriam T. Stark, ed., *Archaeology of Asia*. Oxford: Blackwell. [actually appeared in early 2005]
- Ding Yuan-Chun, Stephen Wooding, Henry C. Harpending, Chi Han-Chang, Li Hai-Peng, Fu Yun-Xin, Pang Jun-Feng, Yao Yong-Gang, Xiang Yu Jing-Gong, Robert Moyzis and Zhang Ya-ping. 2000. 'Population structure and history in East Asia', *Proceedings of the National Academy of Sciences of the United States of America*, 97: 14003–14006.
- van Driem, George. 1998. 'Neolithic correlates of ancient Tibeto-Burman migrations', pp. 67–102 in Roger Blench and Matthew Spriggs, eds., *Archaeology and Language II*. London: Routledge.

- van Driem, George. 2001. *Languages of the Himalayas: An Ethnolinguistic Handbook of the Greater Himalayan Region, containing an Introduction to the Symbiotic Theory of Language* (2 vols.). Leiden: Brill.
- van Driem, George. 2002. 'Tibeto-Burman phylogeny and prehistory: Languages, material culture and genes', pp. 233–249 in Peter Bellwood and Colin Renfrew, eds., *Examining the Farming/Language Dispersal Hypothesis*. Cambridge: McDonald Institute for Archaeological Research.
- van Driem, George. 2006. 'The diversity of the Tibeto-Burman language family and the linguistic ancestry of Chinese', *Bulletin of Chinese Linguistics*, 1 (2).
- van Driem, George. [forthcoming]. 'Austroasiatic phylogeny and the Austroasiatic homeland in light of recent population genetic studies', *Mon-Khmer Studies*.
- Dumont d'Urville, Jules Sébastien César. 1834. *Voyages de découvertes de l'Astrolabe exécuté par ordre du Roi pendant les années 1826-1827-1828-1829 sous le commandement de M. J. Dumont d'Urville: Philologie*. Paris: le Ministère de la Marine [*Philologie* was one of eleven tomes to appear under the main title].
- Finck, Franz Nikolaus. 1909a. *Die Sprachstämme des Erdkreises*. Leipzig: B. G. Teubner.
- Greenberg, Joseph. 1971. 'The Indo-Pacific hypothesis', pp. 805–871 in Thomas A. Sebeok, ed., *Linguistics in Oceania* (Trends in Linguistics, Vol. 8). 's Gravenhage: Mouton.
- Hyde Clark. 1870. 'Response to Leitner's verbal presentation' (November 23<sup>rd</sup>, 1869), *Journal of the Ethnological Society of London*, New Series, II: 32–34.
- Karafet, Tatiana M., Ludmila P. Osipova, Marina A. Gubina, Olga L. Posukh, Stephen L. Zegura and Michael F. Hammer. 2002. 'High levels of Y-chromosome differentiation among native Siberian populations and the genetic signature of a Boreal hunter-gatherer way of life', *Human Biology*, 74 (6): 761–789.
- Kivisild, Toomas, Katrin Kaldma, Mait Metspalu, Jüri Parik, Surinder S. Papiha and Richard Villems. 1999a. 'The place of the Indian mitochondrial DNA variants in the global network of maternal lineages and the peopling of the Old World', pp. 135–152 in Surinder S. Papiha, R. Deka and R. Chakraborty, eds., *Genomic Diversity: Applications in Human Population Genetics*. New York: Kluwer Academic, Plenum Publishers.
- Kivisild, Toomas, Michael J. Bamshad, Katrin Kaldma, Mait Metspalu, E. Metspalu, M. Reidla, S. Laos, Jüri Parik, W. Scott Watkins, Mary E. Dixon,

- Surinder S. Papiha, S. S. Mastana, M. R. Mir, V. Ferak and Richard Villems. 1999b. 'Deep common ancestry of Indian and Western-Eurasian mitochondrial DNA lineages', *Current Biology*, 9 (22): 1331–1334.
- Kivisild, Toomas, S. Rootsi, Mait Metspalu, S. Mastana, Katrin Kaldma, Jüri Parik, E. Metspalu, M. Adojaan. H.-V. Tolk. V. Stepanov, M. Gölge, E. Usanga, Surinder S. Papiha, C. Cinnioglu, R. King. Luca Luigi Cavalli-Sforza, Peter A. Underhill and Richard Villems. 2003. 'The genetic heritage of the earliest settlers persist both in Indian tribal and caste populations', *American Journal of Human Genetics*, 72: 313–332.
- Kraaijenbrink, Thirsa, and George van Driem, Karma Tshering of Gaselô and Peter de Knijff. 2006a. 'Allele frequency distribution for 21 autosomal STR loci in Bhutan', *Forensic Science International*, [in press].
- Kraaijenbrink, Thirsa, and George van Driem, Jean Robert M.L. Ongenort, Nirmal Man Tuladhar and Peter de Knijff. 2006b. 'Allele frequency distribution for 21 autosomal STR loci in Nepal', *Forensic Science International*, [in press].
- Mallory, James Patrick. 1989. *In search of the Indo-Europeans*. London: Thames and Hudson.
- Metspalu, Mait, Toomas Kivisild, Ene Metspalu, Jüri Parik, Georgi Hudjashov, Katrin Kadma, Piia Serk, Monika Karmin, Doron M. Behar. M. Thomas P. Gilbert, Phillip Endicott, Sarabjit Mastana, Surinder S. Papiha, Karl Skorecki, Antonio Torroni and Richard Villems. 2004. 'Most of the extant mtDNA boundaries in South and Southeast Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans', *BioMed Central, Genetics*, 2004 (5): 26 [25 pages].
- Müller, Friedrich Max. 1875 [1865]. 'On θεός and Deus', pp. 239–242, Vol. IV in *Chips from a German Workshop* (4 vols.). London: Longmans, Green and Company.
- Nagaraja, Keralapura Shreevinasaiah. 2006. 'Nahali and Korcu: A comparative note', paper presented at the pilot meeting for the 3<sup>rd</sup> International Conference on Austroasiatic Linguistics held at l'École française d'Extrême-Orient at Siem Reap on 28 June 2006.
- Parkin, Emma J., Thirsa Kraayenbrink, Karma Tshering of Gaselô, George van Driem, Peter de Knijff and Mark Jobling. 2006a. '26-locus Y-STR typing in a Bhutanese population sample', *Forensic Science International*, 161 (1): 1–7.
- Parkin, Emma J., Thirsa Kraayenbrink, Jean Robert M. L. Ongenort, George van Driem, Nirmal Man Tuladhar, Peter de Knijff and Mark Jobling. 2006b. 'Diversity of 26 Y-STR haplotypes in a Nepalese population sample: Iso-

- lation and drift in the Himalayas', *Forensic Science International*, 166 (2-3): 176–181.
- 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/*TaqI* haplotypes show strong correspondence with linguistics', *American Journal of Human Genetics*, 61: 1015–1035 (cf. the erratum published in 1998 in *American Journal of Human Genetics*, 62: 1267).
- 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, 3<sup>rd</sup> International Conference 3–6 April 2000*. Paris: École Nationale Supérieure des Télécommunications.
- Qamar, Raheel, Qasim Ayub, Aisha Mohyuddin, Agnar Helgason, Kehkashan Mazhar, Atika Manoor, Tatiana Zerjal, Chris Tyler-Smith and S. Qasim Mehdi. 2002. 'Y-chromosomal variation in Pakistan', *American Journal of Human Genetics*, 70: 1107–1124.
- Quintana-Murci, Lluís, C. Krausz, Tatiana Zerjal, S. Hamid Sayar, Michael F. Hammer, S. Qasim Mehdi, Qasim Ayub, Raheel Qamar, Aisha Mohyuddin, U. Radhakrishna, Mark A. Jobling Chris Tyler-Smith and Ken McElreavey. 2001. 'Y-chromosome lineages trace diffusion of people and languages in southwestern Asia', *American Journal of Human Genetics*, 68: 537–542.
- Quintana-Murci, Lluís, Raphaëlle Chaix, R. Spencer Wells, Doron M. Behar, Hamid Sayar, Rosaria Scozzari, Chiara Rengo, Nadia al-Zahery, Ornella Semino, A. Silvana Santachiara-Benerecetti, Alfredo Coppa, Qasim Ayub, Aisha Mohyuddin, Chris Tyler-Smith, S. Qasim Mehdi, Antonio Torroni and Ken McElreavey. 2004. 'Where West meets East: The complex mtDNA landscape of the Southwest and Central Asian corridor', *American Journal of Human Genetics*, 74 (5): 827–845.
- Rafinesque, Constantine Samuel. 1831. *Mémoires sur l'origine des nations nègres, ou Introduction à l'histoire des nègres indigènes d'Asie, d'Afrique Polynésie, Amérique et Europe par C. S. Rafinesque, professeur des Sciences historiques et naturelles à Philadelphie*. Manuscript kept at the American Philosophical Society in Philadelphia [cited by Belyi 1997].
- Sahoo, Sanghamitra, Anamika Singh, G. Himabindu, Jheeman Banerjee, T. Sitalaximi, Sonali Gaikwad, R. Trivedi, Phillip Endicott, Toomas Kivisild, Mait Metspalu, Richard Villems and V. K. Kashyap. 2006. 'A prehistory of Indian Y chromosomes: Evaluating demic diffusion scenarios', *Proceedings of the National Academy of Sciences*, 103 (4): 843–848.

- Shi Hong, Dong Yong-li, Wen Bo, Xiao Chun-Jie, Peter A. Underhill, Shen Peidong, Ranajit Chakraborty, Jīn Lì and Sù Bīng. 2005. 'Y-chromosome evidence of southern origin of the East Asian-specific haplogroup O3-M122', *American Journal of Human Genetics*, 77: 408–419.
- Skeat, Walter William, and Charles Otto Blagden. 1906. *Pagan Races of the Malay Peninsula* (2 vols.). London: Macmillan and Company.
- Sù Bīng, Xiao Junhua, Peter Underhill, Ranjan Deka, Zhang Weiling, Joshua Akey, Huang Wei, Shen Di, Daru Lu, Luo Jingchun, Chu Jiayou, Tan Jiazhen, Shen Peidong, Ron Davis, Luca Cavalli-Sforza, Ranajit Chakraborty, Xiong Momiao, Du Ruofu, Peter Oefner, Chen Zhu and Jīn Lì. 1999. 'Y-chromosome evidence for a northward migration of modern humans into eastern Asia during the last Ice Age', *American Journal of Human Genetics*, 65: 1718–1724.
- Sù Bīng, Chunjie Xiao, Ranjan Deka, Mark T. Seielstad, Daoroong Kangwanpong, Junhua Xiao, Daru Lu, Peter Underhill, Luca Cavalli-Sforza, Ranajit Chakraborty and Jīn Lì. 2000. 'Y chromosome haplotypes reveal prehistorical migrations to the Himalayas', *Human Genetics*, 107 (6): 582–590.
- Tambets, Kristiina, Siiri Rootsi, Toomas Kivisild and Richard Villems. 2001. 'The concepts of Richard Indreko about the origins of the Finno-Ugric speakers and the population genetics of the extant northeast European populations', *Trames, Journal of the Humanities and Social Sciences*, 5 (55/50), (1): 59–74.
- Thangaraj, Kumarasamy, Vempati Sridhar, Toomas Kivisild, Alla G. Reddy, Gyaneshwer Chaubey, Vijay Kumar Singh, Suminder Kaur, Pooja Agarwal, Amit Rai, Jalaj Gupta, Chandana Basu Mallick, Niraj Kumar, Thirulaisamy P. Velavan, Rajanbabu Suganthan, Divi Udaykumar, Rashmi Kumar, Rachana Mishra, Arif Khan, Chitikineni Annapurna and Lalji Singh. 2005. 'Different population histories of the Mundari- and Mon-Khmer-speaking Austro-Asiatic tribes inferred from the mtDNA 9-bp deletion/insertion polymorphism in Indian populations', *Human Genetics*, 116: 507–517.
- Toporov, Vladimir Nikolaevič. 1969. 'K voprosu o tipologičeskoj blizosti enisejskix jazykov i burušaski', pp. 217–220 in *Proisxoždenie aborigenov Sibiri i ix jazykov (Materialy mežvuzovskoj konferencii 11–13 maja 1969 g.)*. Tomsk.
- Toporov, Vladimir Nikolaevič. 1971. 'Burushaski and Yenisean languages: Some parallels', *Travaux Linguistiques de Prague*, 4: 107–125.
- Underhill, Anne P., and Junk Habu. 2006. 'Early communities in East Asia: Economic and sociopolitical organisation at the local and regional levels', pp. 121–148 in Miriam T. Stark, ed., *Archaeology of Asia*. Oxford: Blackwell. [actually appeared in early 2005]

- Wells, R. Spencer, Nadira Yuldasheva, Ruslan Ruzibakiev, Peter A. Underhill, Irina Evseeva, Jason Blue-Smith, Jīn Lì, Sù Bīng, Ramasamy Pitchappan, Sadagopal Shanmugalakshmi Karuppiiah Balakrishnan, Mark Read, Nathaniel M. Pearson, Tatiana Zerjal, Matthew T. Webster, Irakli Zholoshvili, Elena Jamarjashvili, Spartak Gambarov, Behrouz Nikbin, Ashur Distiev, Oganazar Aknazarov, Pierre Zalloua, Igor Tsoy, Mikhail Kitaev, Mirsaid Mirrakhimov, Ashir Chariev and Walter F. Bodmer. 2001. 'The Eurasian heartland: A continental perspective on Y-chromosome diversity', *Proceedings of the National Academy of Sciences*, 98: 10244–10249.
- Wen Bo, Xie Xuanhua, Gao Song, Li Hui, Shi Hong, Song Xiufeng, Qian Tingzhi, Xiao Chunjie, Jin Jianzhong, Sù Bīng, Lu Daru, Ranajit Chakraborty and Jīn Lì. 2004. 'Analyses of genetic structure of Tibeto-Burman populations reveals sex-biased admixture in southern Tibeto-Burmans', *American Journal of Human Genetics*, 74 (5): 856–865.
- Wheeler, Robert Eric Mortimer. 1966. *Civilizations of the Indus Valley and Beyond*. London: Thames and Hudson.
- Wheeler, Robert Eric Mortimer. 1968. *The Indus Civilization* (Supplementary Volume to the Cambridge History of India) (3<sup>rd</sup> edition). Cambridge: Cambridge University Press.
- Whitehouse, Paul, Timothy Usher, Merritt Ruhlen and William Shìyuán Wáng. 2004. 'Kusunda: An Indo-Pacific language in Nepal', *Proceedings of the National Academy of Sciences of the United States of America*, 101 (15): 5692–5695.
- Zerjal, Tatjana, B. Dashnyam, A. Pandya, Manfred Kayser, Lutz Roewer, F. R. Santos, W. Schiefenhovel, N. Fretwell, M. A. Jobling, S. Harihara, K. Shimuzu, D. Semjidmaa, Antti Sajantila, P. Salo, M. H. Crawford, E. K. Ginter, Oleg V. Efgrafov and Chris Tyler-Smith. 1997. 'Genetic relationships of Asian and Northern Europeans, revealed by Y chromosomal DNA analysis', *American Journal of Human Genetics*, 62: 1174–1183.