Austroasiatic phylogeny and the Austroasiatic homeland in light of recent population genetic studies

George VAN DRIEM
Leiden University, The Netherlands

Abstract

Currently, the most informed and authoritative Austroasiatic *Stammbaum* is the language family tree presented by Diffloth (2005). Diffloth has also identified a number of reconstructible pan-Austroasiatic lexical roots with linguistic palaeontological significance for the probable location of the Austroasiatic homeland or, at least, for minimal climatological conditions which must have prevailed in such a homeland. What can population genetic studies of Tibeto-Burman and Indo-European language communities tell us about how the emerging Austroasiatic picture could be interpreted? Does the Father Tongue hypothesis apply to Austroasiatic or to portions thereof? The three versions of prehistory afforded by archaeology, population genetics and linguistics can and should be kept distinct in the argumentation used in correlative studies. The Centripetal Migration model is proposed and juxtaposed to the Farming-Language Dispersal model.[1]

In 2001, Gérard Diffloth presented a new phylogeny of the Austroasiatic family of languages based on years of meticulous field work and linguistic comparison. This new *Stammbaum* is reproduced in modified form in Diagram 1. In contrast to earlier family trees, Diffloth's Austroasiatic family splits up into three major nodes, i.e. Munda, Khasi-Khmuic and a new 'Mon-Khmer'. In this new tripartite division, Munda is still one of the primary branches of Austroasiatic, representing the native heart of the Indian subcontinent. The Khasi-Khmuic branch represents what might be thought of as 'Inland Austroasiatic', and a more precisely delineated Mon-Khmer represents 'Littoral Austroasiatic'. The new Mon-Khmer comprises Khmero-Vietic and Nico-Monic. Each of the two sub-branches of Mon-Khmer is further subdivided, with Nico-Monic consisting of Asli-Monic and Nicobarese, and Khmero-Vietic breaking up into Vieto-Katuic and Khmero-Bahnaric. Conspicuously, Diffloth had initally left out Pearic on purpose because its genetic affinity was still, as he put it, *en chantier,* but it is at least safe to say that its greatest genetic affinity is not with the Munda or Khasi-Khmuic branches, but with Mon-Khmer. Many more phylogenetic insights are

---

[1] This article is a revised draft of a paper presented on 29 June 2006 at the pre-ICAAL3 pilot seminar held at l'École Française d'Extrême-Orient in Siem Reap, a meeting of scholars whose aim it was to resurrect the International Conference on Austroasiatic Linguistics. See <www.iias.nl/icaal>. Many thanks to my team-mates in the project Languages and Genes of the Greater Himalayan Region, particularly Peter de Knijff, Mark Jobling, Thirsa Kraaijenbrink and Emma Parkin. Theories of remote linguistic relationship such as Austric or Sino-Austronesian are not discussed in this brief paper, but have been discussed elsewhere, e.g. van Driem (2001, 2005).

contained in Diffloth's burgeoning, highly detailed but as yet unpublished Austroasiatic comparative database.

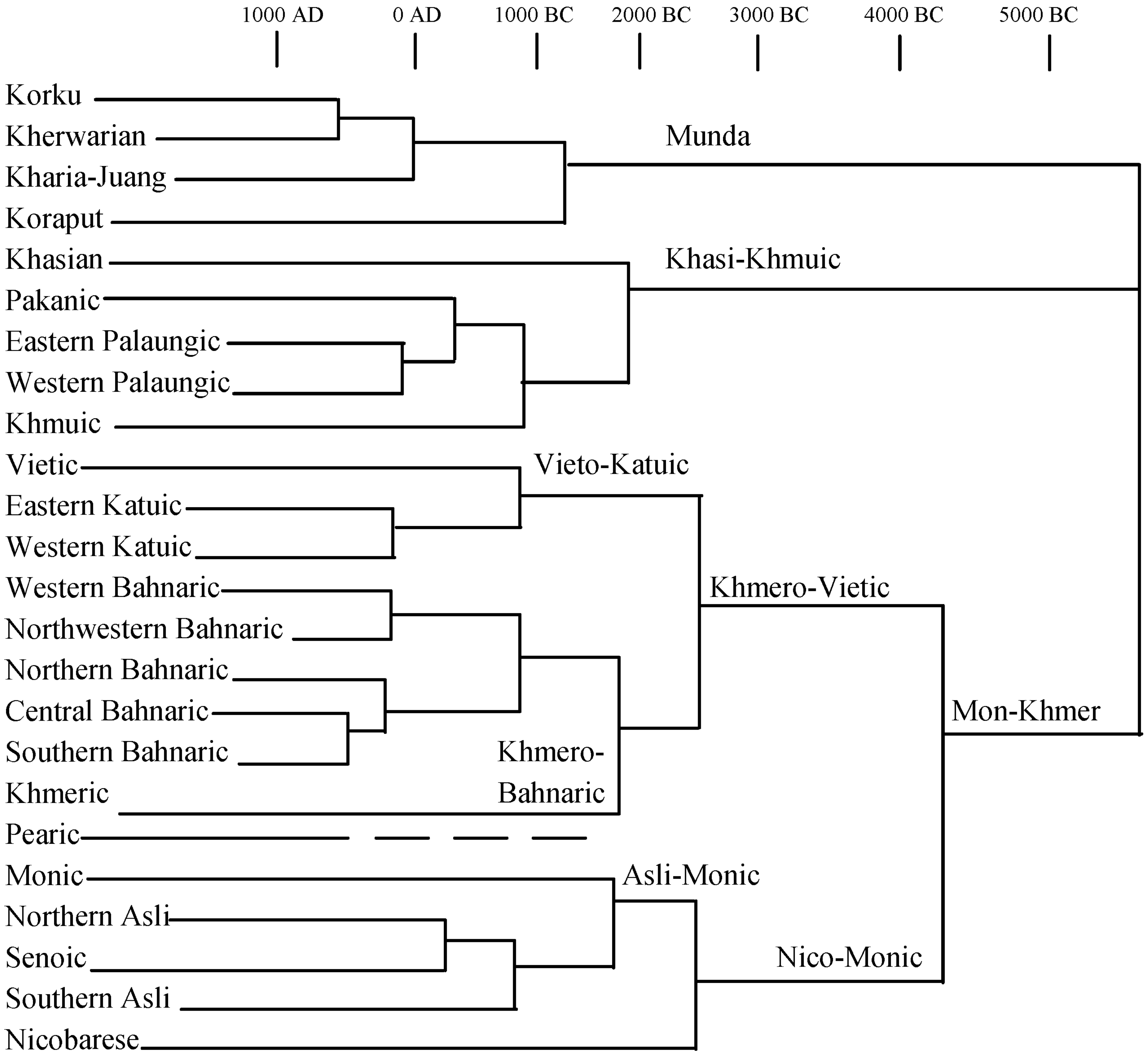

Diagram 1. Austroasiatic with Gérard Diffloth's tentative calibration of time depths for the various branches of the language family (modified from Diffloth 2001, 2005). The precise phylogenetic propinquity of Pearic, after Khmeric loan layers have been stripped off, remains uncertain except that Diffloth 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.

Any reconstruction of Austroasiatic population prehistory must start out from the present and historically attested geographical distribution of Austroasiatic subgroups. Diagram 2 shows the geographical distribution of Austroasiatic subgroups with the exception of the recently documented enclaves of Pakanic in southern China. We can all look forward to Gérard Diffloth's new detailed Austroasiatic map which is currently in production. When we compare the new phylogenetic model for Austroasiatic with the geographical distribution of Austroasiatic subgroups, a number of hypotheses concerning the possible location of an Austroasiatic homeland suggest themselves. In fact, in the past the most diverse homeland sites have been proposed for Austroasiatic, and most of these are discussed in my handbook (van Driem 2001:289-332).

On the basis of linguistic palaeontology Diffloth has argued that the reconstructibility at the Proto-Austroasiatic level of words for tree monitor, ant eater, buffalo, mountain goat, bear cat, elephant, peacock, rhinoceros and bamboo rat as well as the rich reconstructible rice cultivation vocabulary imply that the Austroasiatic homeland was located in the tropics. The Hémǔdù culture at the mouth of the Yangtze (5000-4500 BC) provides the best unambiguous evidence for a population for whom rice is the staple. The oldest direct evidence for domesticated rice, however, dates from 6500 BC and is from the Bāshídàng and Péngtóushān sites belonging to the Péngtóushān culture (7500-6100 BC) on the middle Yangtze in what today is Húnán and from the Jiǎhú culture (6000-7000 BC) on the Huái river further north in what today is Hénán.[2]

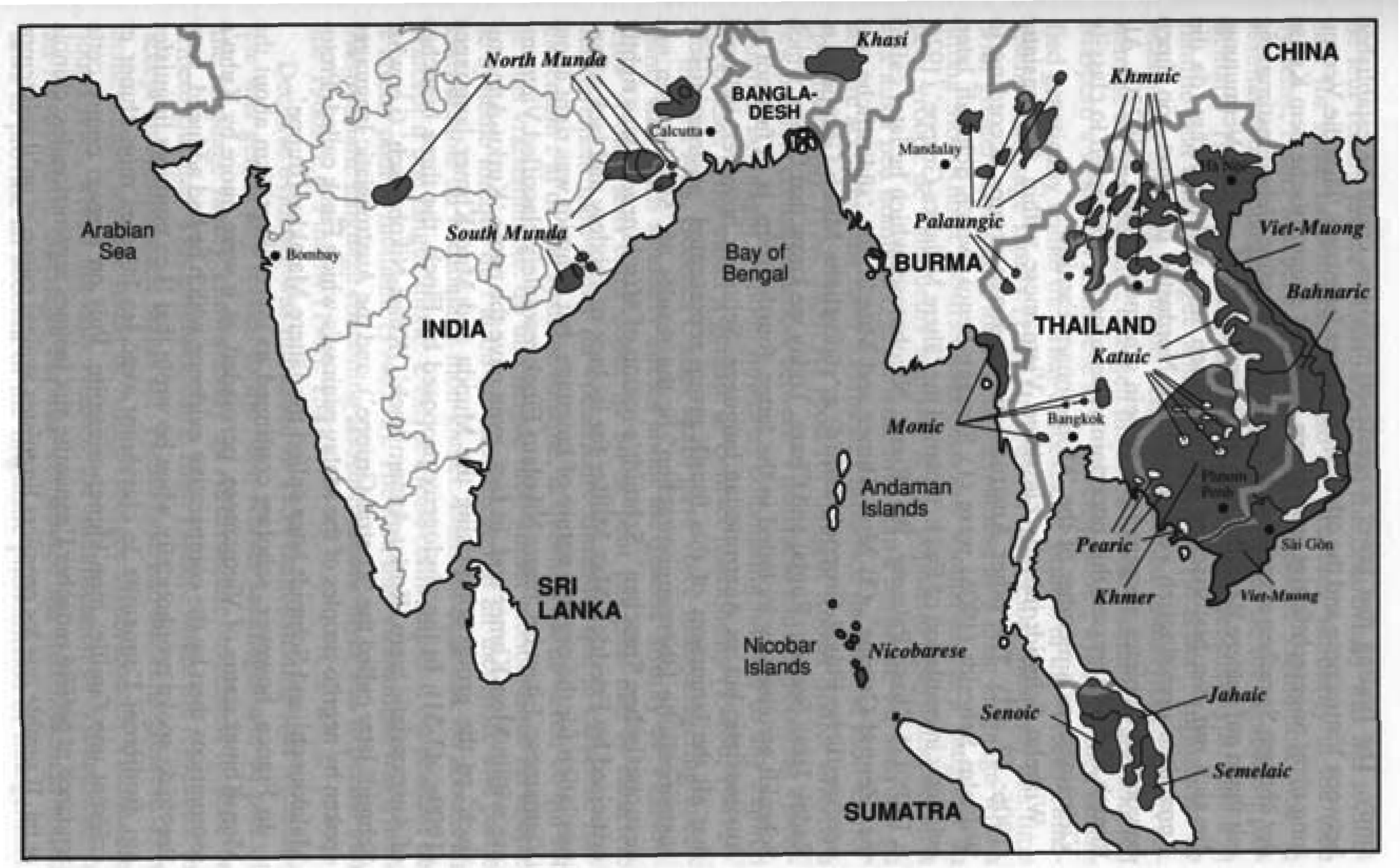

Diagram 2. Geographical distribution of Austroasiatic subgroups (van Driem 2001:267). Recently documented Pakanic enclaves in southern China are not yet shown.

Since the archaeological sites reflecting the oldest known rice cultivators are located along the middle Yangtze, Diffloth logically raised the palaeoclimatological question whether the faunal landscape which existed in this area at the putative time depth of Proto-Austroasiatic would be compatible with the environment suggested by linguistic palaeontology. Clearly, by the

---

[2] More recently, cultivated rice has been recovered from Nánguānlǐ in southeastern Taiwan dating from ca. 3000 BC (Tsang 2004).

faunal criterion large tracts of the Indian Subcontinent and Southeast Asia also remain homeland candidates.

Any successful correlative study of the historical linguistic picture and the population genetics of the the modern language communities will have to provide an account for the manifest somatological or phenotypical difference between Munda speakers on one hand and the speakers of Khasi-Khmuic and Mon-Khmer languages on the other, as well as comparably great differences between Aslian negrito groups and the linguistically closely related Nicobarese. The meaningfulness of any conjectures that we base on such correlative studies depends on the reliability of the linguistic reconstructions and language family tree as well as on the degree of resolution, refinement and thoroughness of sampling of our genetic assays.

Operating on the assumption that frequency gradients of Y haplogroups, mtDNA polymorphisms or autosomal haplotypes may correlate precisely or partially with the distribution of Austroasiatic language communities, we may still wonder whether such gradients necessarily reflect the people who introduced and disseminated any putative proto-language. The linguistic ancestors of a language community were not necessarily the same people as the biological ancestors of that community. 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.

The genetic picture also shows a certain sexual dimorphism in linguistic prehistory. In Baltistan, located in what today is northern Pakistan, the local Tibetan dialects are the most conservative of all Tibetan languages, preserving consonant clusters retained in Classical Tibetan orthography but wholly lost in most other Tibetan dialects. Yet the Balti abandoned the Tibetan script after they were converted to Islam in the fifteenth century, although native activists have in recent years begun reintroducing the Tibetan script, e.g. on shop signs, to the displeasure of central government authorities. Paradoxically, the old consonant clusters ceased to be pronounced as such in most areas throughout Tibet where the conservative 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 appears 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 *puras*, and the Aryan god Indra is a *puramïdara* 'destroyer of aboriginal forts', who shattered ninety such strongholds (Wheeler 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.* 2004a, 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 by 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. 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 Bronze Age Andronovo 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 observation made by Sahoo *et al.* (2006) 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.

In the East, a less controversial instance of the Father Tongue hypothesis is found. A population genetic study of 23 Hàn populations (Wen *et al.* 2004) has corroborated the picture which linguists and historians had developed and have long entertained, of a martial and therefore male-biased

Hàn expansion southward during the sinification of what today is southern China. The southern Hàn paternal lineage shows preponderant northern Hàn penetration alongside a faint pre-Sinitic signature. Males from the north were the primary contributor to the paternal gene pool of southern Hàn populations, whereas the mitochondrial DNA of southern Hàn populations contains roughly equal contributions from pre-Sinitic and Hàn maternal ancestors.

The population genetic work is ongoing, and some preliminary findings are prone to being interpreted prematurely in terms of their potential significance for population prehistory. Just two years ago, an article by Langstieh *et al.* (2004) created a stir amongst scholars of Khasi because the study addressed the provenance of the Garos and Khasis of the Meghalaya. This valuable contribution raised more questions than it answered. The researchers claimed that the population of the Meghalaya is homogeneous, whereas the Garos and Khasi tribes are linguistically unrelated. Judging from their median joining network, the Garos would appear to be an ethnic subset of the Khasis, something which suggests that the Garos are more homogeneous as a group than the Khasi tribes. Moreover, the purported homogeneity of the populations of the Meghalaya was based on comparsion with the Chinese and North American Indians! Obviously it would be more meaningful to conduct fine-mesh genetic comparison of the Khasi with Pakanic, Palaungic and Khmuic language communities, who are their closest linguistic relatives. By the same token, fine-mesh studies should be undertaken to compare the Garos with the Bodos, Rabhas and Dimasas, who are their closest linguistic relatives, as well as other linguistically less related population groups of northeastern India.

Short tandem repeats (STR) are highly polymorphous, but the short tandem repeats chosen by Langstieh *et al.* were not necessarily the optimal choice as genetic markers for gauging differences between closely related populations. In all of the Himalayan groups which we have been testing — and our sampling represents a highly varied and heterogeneous collection of peoples and language communities — we do not always see that much variation in the short tandem repeats as these researchers have found in the Meghalaya (Kraaijenbrink *et al.* 2006, 2007, Parkin *et al.* 2006a, 2006b). So, the peoples of the Meghalaya show up as a highly heterogeneous population, but the researchers cannot yet know this for sure, for they have not been able to compare their findings with data on other relevant groups. Further studies will have to corroborate the impression that the Meghalaya may be an area where the antiquity and genetic heterogeneity of the populations is relatively great.

In addition to the studies already mentioned, relevant population genetic studies have begun to chart the autosomal lineages, the mitochondrial or maternal lineages and the Y chromosome haplogroups representing the paternal lineages of Austroasiatic language communities and neighbouring population groups, e.g., Ashma *et al.* (2002), Banerjee *et al.* (2005a, 2005b), Debnath and Chaudhuri (2005), Ding *et al.* (2000), Kashyap *et al.* (2002a, 2002b), Krithika *et al.* (2005), Kumar *et al.* (2004), Maity *et al.* (2003), Sahoo *et al.* (2002), Shi *et al.* (2005, 2006), Singh *et al.* (2006), Sù *et al.* (1999, 2000), Thomas *et al.* (2004), Watkins *et al.* (2005).

Much progress has been made in Y chromosome phylogeny since the seminal contribution by Underhill *et al.* (2001). A number of research teams have mooted a possible link between the distribution of Austroasiatic language communities and the M95 mutation, i.e. Y chromosomal haplogroup O2a, e.g. Sù *et al.* (2000), Kayser *et al.* (2003), Kivisild *et al.* (2003), Cordaux *et al.* (2004b). Frequency gradients for Y haplogroup O2a are mapped for the Indian Subcontinent and Southeast Asia by Sahoo *et al.* (2006) and are shown here in Diagram 3. At the same time, maternal lineages of Munda groups appear to be old and indigenous to the Subcontinent, as indeed can be said of many Indian mitochondrial lineages (Kivisild *et al.* 1999a, 1999b, 2003). So, the mitochondrial picture indicates that the Munda maternal lineage derives from the earliest human settlers on the Subcontinent, whilst the predominant Y chromosome haplogroup argues for a Southeast Asian paternal homeland for Austroasiatic language communities in India. In this context, Sahoo *et al.* (2006: 847) rightly caution 'against simplistic interpretations of either linguistic or genetic correlations'.

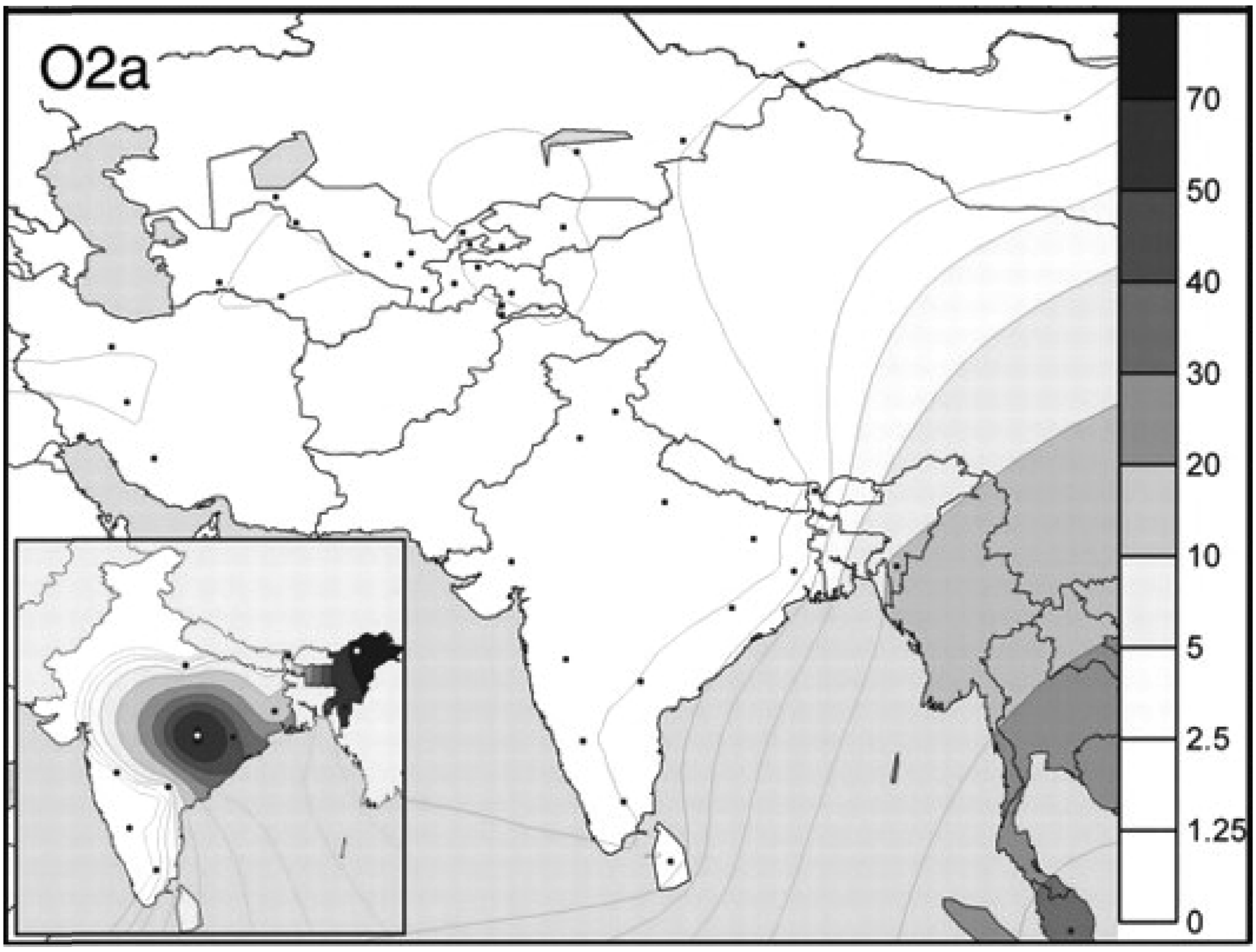

Diagram 3. Spatial frequency distribution of Y haplogroup O2a for caste populations [big map] and tribal populations [inset map], reproduced here from Sahoo *et al.* (2006: 846).

At the same time, some of the formulations in Sahoo *et al.* (2006) provide grounds for cautioning against the use of just a single explanatory model in our interpretation of the genetic, archaeological and linguistic data. Portions of the article reflect a *Hineininterpretation* of the Farming-Language

Dispersal theory into the genetic findings. This slant in no way diminishes the value of the proposed correlation of the Y chromosomal haplogroup O2a with the geographical distribution of Austroasiatic language communities. Yet this interpretation of genetic findings raises a more general issue which is of central relevance to the ways that we think about the prehistory of language families such as Austroasiatic.

It is tempting to assume that genes, languages and archaeological horizons have always tended to move in tandem with the incremental spread of Neolithic agriculture and to convince ourselves that this model generates the most parsimonious explanations. In fact, realities on the ground were often more complex. This complexity is not only suggested by the dissonance between the different pictures of prehistory reconstructible through the three disciplines, but more so by the multi-layered nature of the distinct pictures which emerge from linguistics, population genetics and archaeology. For example, Ossetian, an East Iranian language is spoken in an area which lies decidedly to the west of most West Iranian language communities, attesting to the ancient migration of the Alans and Sarmatians to the north central Caucasus.

The geographical distribution of gene frequencies not only reveals distinct migrations, sometimes in opposing directions at different time depths, but detailed future studies may also enable us to ascertain the relative chronology of the distinct layers of genetic diffusion at different times across the same areas. Archaeology defines specific cultural assemblages with definable horizons and identifiable colonial exponents. The farming-language dispersal model necessarily works in the case of Austronesian, where the geographical spread of the language family has to a major extent resulted from the colonisation of previously uninhabited insular environments emanating from Formosa, or perhaps from Hémŭdù via Formosa. Yet we must question whether the latter theory has the same explanatory power to account for the spread of language families under the circumstances which prevailed on the land masses where most of prehistory unfolded. For an archaeologist contemplating language families, the urge is inevitably irresistible to associate the geographical spread of technologically advanced Neolithic civilisations into more backward areas with the spread of peoples and language families.

More fundamentally, the premises of the farming-language spread theory ought to be questioned. The surplus generated by an agricultural economy and the stratified social and command structure enabled by a Neolithic lifestyle are held to have driven demographic spread into many areas. This argument is plausible, but this argument is not the crux of the farming-language dispersal theory. Crucial to the model is the tenet that the incremental spread of the Neolithic as such is associated with ‘the foundation dispersals’ of language families. This theory therefore presumes that the ancient spread of language families unfolded in the same direction as the demographic spread driven by Neolithic agriculture.

The very opposite may be what actually happened in many cases. Across the Fertile Crescent, agriculture was adopted by ethnolinguistically unrelated populations, and agriculture spread effortlessly across ethnolinguistic boundaries without disrupting them in any significant way. Sumerian pictographic script, developed ca. 3200 BC, appeared millennia after the invention of agriculture. Sumerian, Elamite, Akkadian,[3] Hurrian, Hattic and other contemporaneous agricultural civilisations were in all likelihood not the first cultivators of the region. Yet even these antique agricultural language communities have left no surviving linguistic descendants. The earliest recorded and reconstructible history of the Near East bears witness to the permeability of linguistic boundaries for the dissemination of agriculture and crops.

The Bronze Age of Asia Minor and Mesopotamia is characterised by a long period of incursive population movements into, rather than out of, Anatolia and the Fertile Crescent, lured by the relative affluence of urban centres supported by agricultural surplus. Gutæans, Amorites, Kassites and other peoples were drawn in by the promise of the good life. Most linguistic reconstructions presume that Indo-European groups such as the Hittites and Mitanni likewise came to settle in Asia Minor and the Fertile Crescent from elsewhere. Toponymical evidence and details about the cults of certain deities have been used to argue that even the Sumerians originally migrated from an earlier northern homeland to lower Mesopotamia. Were the motivations of migrating peoples in agricultural and pre-agricultural societies genuinely different at the Neolithic horizon than at later times?

Tidings of technologically advanced urban societies may in the course of prehistory have provided ample motivation for migration, with enticing prospects of plunder and material advancement. We must consider such alternatives especially in those cases where the linguistic picture suggests a radically different view of prehistory than does the spread of material culture as reflected in the known archaeological record. The introduction of Proto-Sinitic, a branch of Tibeto-Burman, into the Yellow River basin is a case in point. This theory, which I shall call the Centripetal Migration model, is diametrically opposed to the centrifugal Farming-Language Dispersal theory. The Centripetal Migration model may also apply to portions of Austroasiatic prehistory.

More crucially, an essential trait of the Centripetal Migration theory is that this model assumes that migrations in prehistory could have unfolded both in centrifugal and centripetal directions with respect to centres of technologically advanced and later urban civilisations. The motives for migrations were no doubt diverse, and no single model, such as the Farming-Language Dispersal theory, can account for all demographic developments and linguistic intrusions, even across the Neolithic horizon. Even the chief proponents of the Farming-Language Dispersal theory do not entertain the idea that all languages were spread by early farmers, e.g. Bellwood (2005). At the same time, we must also not lose sight of the fact that vast tracts of the

---

[3] Today Afroasiatic languages are spoken throughout this area, but none are descended directly from the extinct branch which Akkadian represents.

Himalayas, Burma, northeastern India and neighbouring southwestern China remain archaeologically under-explored or unexplored.

In conclusion, groundbreaking research in population genetics has begun to suggest that the geographical distribution of Austroasiatic may be connected to a well-defined Y chromosomal haplogroup. What I called the 'Father Tongue hypothesis' at the Indo-Pacific Prehistory Association conference in Taipei in 2002, based largely on the work of Poloni *et al.* (1997, 2000), may also apply to Austroasiatic, either wholly or in part, on the basis of the population genetic studies completed to date. The veracity of the Father Tongue hypothesis is the inherent underlying assumption when geneticists propose that a particular Y haplogroup, say O2a, corresponds to the geographical spread of a particular language family, such as Austroasiatic. Diagram 4 illustrates the portion of the Y chromosomal phylogeny thought to be relevant to Austroasiatic.

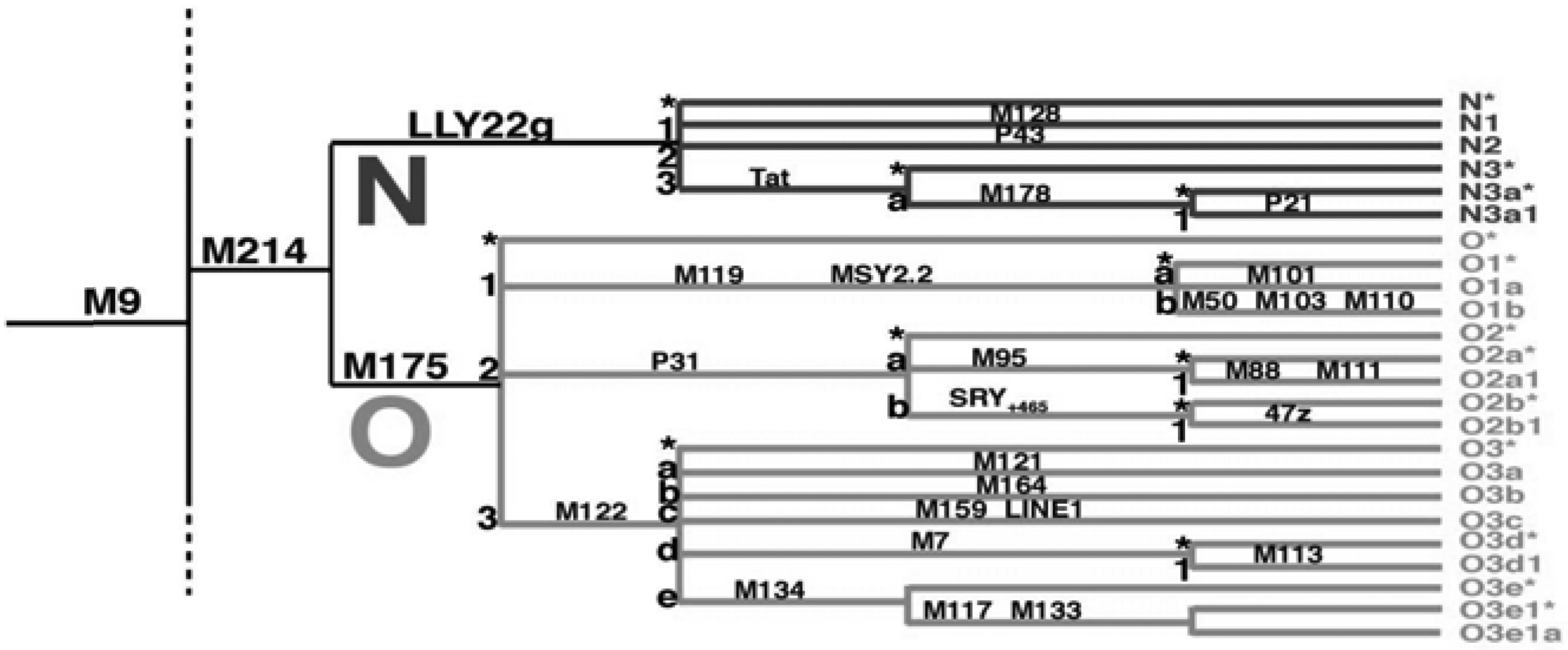

Diagram 4. The portion of the Y chromosome phylogenetic tree relevant to the Father Tongue hypothesis with regard to Austroasiatic, provided by Mark Jobling and Emma Parkin.

However, 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 Tibeto-Burman, if not more so. Careful correlation of linguistic and population genetic findings may enable us to reconstruct early language contact situations and ancient cases of language shift and linguistic intrusions that might account for the phenotypical difference between Munda speakers and Khasi-Khmuic and Mon-Khmer language communities as well as between the Aslian negrito populations, Aslian non-negrito populations and the Nicobarese. The Father Tongue hypothesis may not apply in all cases for the biological ancestry of all 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 Hàn.

Although Sahoo *et al.* (2006) clearly favour a Southeast Asian homeland for Austroasiatic, their findings cannot yet conclusively establish that Southeast Asia is the point of origin for the O2a haplogroup. The exciting hypothesis that the O2a haplogroup may correlate with the linguistic spread of

Austroasiatic also remains to be demonstrated in convincing detail. A fine-mesh genetic sampling of all Austroasiatic populations — not just the most populous, national majority or prestige groups — will be required in order to determine which precise area could be the probable point of origin of polymorphic genomic markers which could be correlated with the linguistic spread of Austroasiatic. Furthermore, the detailed geography of the entire O branch of Y chromosomal haplogroups has yet to be reconstructed at a satisfactorily high resolution. I call upon all interested parties to join forces and help us in this endeavour.

REFERENCES

Ashma, R., and V.K. Kashyap. 2002. "Genetic polymorphism at 15 STR loci among three important subpopulations of Bihar, India." *FSI* 130: 58-62.

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.

Banerjee, Jheelam, R. Trivedi and V.K. Kashyap. 2005a. "Y-STR haplotypes in autochthonous tribal population of the Chotanagpur Plateau, India." *JFS* 50(3):paper JFS2004531.

Banerjee, Jheelam, R. Trivedi and V.K. Kashyap. 2005b. "Polymorphism at 15 short tandem repeat AmpFℓSTR® Identifier™ loci in three aboriginal populations in India: An assessment in human identification." *JFS* 50(5):paper JFS2005151.

Bellwood, Peter. 2005. *First Farmers: The Origins of Agricultural Societies.* Oxford: Blackwell.

Bellwood, Peter. 2006. "Asian farming diasporas? Agriculture, languages and genes in China and Southeast Asia." In Miriam T. Stark, ed., *Archaeology of Asia*, pp. 96-118. Oxford: Blackwell.

Cordaux, Richard, et al. 2003. "Mitochondrial DNA analysis reveals diverse histories of tribal populations from India." *European Journal of Human Genetics* 11:253-264.

Cordaux, Richard, et al. 2004a. "The northeast Indian passageway: A barrier or corridor for human migrations?" *MBE* 21:1525-1533.

Cordaux, Richard, et al. 2004b. "Independent origins of Indian caste and tribal paternal lineages." *MBE* 14:231-235.

Debnath, M., and T.K. Chaudhuri. 2005. "HLA-A and HLA-B distribution in the Toto — a vanishing sub-Himalayan tribe of India." *Tissue Antigens* 67:64-65.

Diffloth, Gérard. 2001. "Tentative calibration of time depths in Austroasiatic branches." *Paper presented at the Colloque Perspectives sur la Phylogénie des Langues d'Asie Orientales at Périgueux, 30 August 2001.*

Diffloth, Gérard. 2005. "The contribution of linguistic palaeontology to the homeland of Austroasiatic." In Laurent Sagart, Roger Blench and Alicia Sanchez-Mazas, eds., *The Peopling of East Asia: Putting Together the Archaeology, Linguistics and Genetics,* pp. 77-80. London: Routledge Curzon.

Ding Yuan-Chun, et al. 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. 2001. *Languages of the Himalayas: An Ethnolinguistic Handbook of the Greater Himalayan Region with an Introduction to the Symbiotic Theory of Language* (2 vols.). Leiden: Brill.

van Driem, George. 2005. "Sino-Austronesian vs. Sino-Caucasian, Sino-Bodic vs. Sino-Tibetan, and Tibeto-Burman as default theory", pp. 285-338. In Yogendra Prasada Yadava, Govinda Bhattarai, Ram Raj Lohani, Balaram Prasain and Krishna Parajuli, eds., *Contemporary Issues in Nepalese Linguistics*. Kathmandu: Linguistic Society of Nepal.

van Driem, George. 2007. "The diversity of the Tibeto-Burman language family and the linguistic ancestry of Chinese." *Bulletin of Chinese Linguistics* 1(2):211-270.

Kashyap, V.K., N. Sarkar and R. Trivedi. 2002a. "Allele frequencies for STR loci of the PowerPlex™ 16 multiplex system in five endogamous populations of India." *FSI* 126:178-186.

Kashyap, V.K., Saurav Guha and R. Trivedi. 2002b. "Concordance study on 15 STR loci in three major populations of the Himalayan state Sikkim." *JFS* 47(5):paper JFS2001333_475.

Kayser, Manfred, et al. 2003. "Reduced Y-chromosome, but not mitochondrial DNA, diversity in human populations from west New Guinea." *American Journal of Human Genetics* 72:281-302.

Kivisild, Toomas, et al. 1999a. "The place of the Indian mitochondrial DNA variants in the global network of maternal lineages and the peopling of the old world." In Surinder S. Papiha, R. Deka and R. Chakraborty, eds., *Genomic Diversity: Applications in Human Population Genetics,* pp. 135-152. New York: Kluwer Academic, Plenum Publishers.

Kivisild, Toomas, et al. 1999b. "Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages." *Current Biology* 9(22):1331-1334.

Kivisild, Toomas, et al. 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, et al. 2006. "Allele frequency distribution for 21 autosomal STR loci in Nepal." *FSI* 168(2-3):227-231.

Kraaijenbrink, Thirsa, et al. 2007. "Allele frequency distribution for 21 autosomal STR loci in Bhutan." *FSI* 170:68-72.

Krithika, S., R. et al. 2005. "Genetic diversity at 15 microsattelite loci among the Adi Pasi population of Adi tribal cluster in Arunachal Pradesh, India." *Legal Medicine* 7:306-310.

Kumar, Vikrant, Debashis Basu and B. Mohan Reddy. 2004. "Genetic heterogeneity in northeastern India: Reflection of tribe-caste continuum in the genetic structure." *American Journal of Human Biology* 16:334-345.

Langstieh B.T., et al. 2004. "Genetic diversity and relationships among the tribes of Meghalaya compared to other Indian and continental populations." *Human Biology* 76(4):569-590.

Maity, B., S.C. Nunga and V.K. Kashyap. 2003. "Genetic polymorphism revealed by 13 tetrameric and 2 pentameric STR loci in four Mongoloid tribal populations." *FSI* 132:216-222.

Metspalu, Mait, et al. 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, doi:10.1186/1471-2156-5-26].

Parkin, Emma J., et al. 2006a. "26-locus Y-STR typing in a Bhutanese population sample." *FSI* 161(1):1-7.

Parkin, Emma J., et al. 2006b. "High-resolution Y-STR haplotypes in a Nepalese population sample: Isolation and drift in the Himalayas." *FSI* 166(2-3):176-181.

Poloni, Estella Simone, et al. 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 *American Journal of Human Genetics* 62:1267).

Poloni, Estella Simone, et al. 2000. "Languages and genes: modes of transmission observed through the analysis of male-specific and female-specific genes." In Jean-Louis Dessalles and Laleh Ghadakpour, eds., *Proceedings: Evolution of Language, 3rd International Conference 3-6 April 2000*, pp. 185-186. Paris: École Nationale Supérieure des Télécommunications.

Qamar, Raheel, et al. 2002. "Y-chromosomal variation in Pakistan." *American Journal of Human Genetics* 70:1107-1124.

Quintana-Murci, Lluís, et al. 2001. "Y-chromosome lineages trace diffusion of people and languages in southwestern Asia." *American Journal of Human Genetics* 68:537-542.

Quintana-Murci, Lluís, et al. 2004. "Where West meets East: The complex mtDNA landscape of the southwest and central Asian corridor." *American Journal of Human Genetics* 74:827-845.

Sahoo, Sanghamitra, and V.K. Kashyap. 2002. "Genetic variation at 15 autosomal microsatellite loci in the three highly endogamous tribal populations of Orissa, India." *FSI* 130:189-193.

Sahoo, Sanghamitra, et al. 2006. "A prehistory of Indian Y chromosomes: evaluating demic diffusion scenarios." *Proceedings of the National Academy of Sciences* 103(4): 843-848.

Shi Hong, et al. 2005. "Y-chromosome evidence of southern origin of the east Asian-specific haplogroup O3-M122." *American Journal of Human Genetics* 77:408-419.

Shi, L., S.B. Xu, et al. 2006. "HLA-A, HLA-B and HLA-DRB1 alleles and haplotypes in Naxi and Han populations in southwestern China (Yunnan province)." *Tissue Antigens* 67:38-44.

Singh, Anamika, R. Trivedi and V.K. Kashyap. 2006. “Genetic polymorphism at 15 tetrameric short tandem repeat loci in four aboriginal tribal populations of Bengal.” *JFS* 51(1):183-187.

Sù Bīng, et al. 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, et al. 2000. “Y chromosome haplotypes reveal prehistorical migrations to the Himalayas.” *Human Genetics* 107(6):582-590.

Thangaraj, Kumarasamy, et al. 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.

Thomas, R, S.B. Nair and M. Banerjee. 2004. “HLA-B and HLA-C alleles and haplotypes in the Dravidian tribal populations of southern India.” *Tissue Antigens* 64:58-65.

Tsang, Cheng-hwa. 2004. “Recent discoveries of the Tapenkeng culture in Taiwan.” In Laurent Sagart, Roger Blench and Alicia Sanchez-Mazas, eds., *The Peopling of East Asia: Putting Together Archaeology, Linguistics and Genetics,* pp. 63-74. London: Routledge Curzon.

Underhill, Peter A., et al. 2001. “The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations.” *Annals of Human Genetics* 65:43-62.

Watkins, W.S., et al. 2005. “Diversity and divergence among tribal populations of India.” *Annals of Human Genetics* 69:680-692.

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) (3rd edition). Cambridge: Cambridge University Press.

Wells, R. Spencer, et al. 2001. “The Eurasian heartland: A continental perspective on Y-chromosome diversity.” *Proceedings of the National Academy of Sciences* 98:10244-10249.

Wen Bo, et al. 2004. “Genetic evidence supports demic diffusion of Han culture.” *Nature* 431:302-305.

Zerjal, Tatjana, et al. 1997. “Genetic relationships of Asian and northern Europeans, revealed by Y chromosomal DNA analysis.” *American Journal of Human Genetics* 62:1174-1183.

Received: 15 August 2006

Himalayan Languages, Leiden University
Post Office Box 9515, 2300 RA Leiden
THE NETHERLANDS
<george.van.driem@gmail.com>