

# 16 *Rice and the Austroasiatic and Hmong-Mien homelands*

---

GEORGE VAN DRIEM

## 1 Introduction

The locations of the Austroasiatic and Hmong-Mien homelands have been tied to the origins of Asian rice cultivation. The lexical evidence supports the view that ancient Austroasiatics and ancient Hmong-Mien were the first rice cultivators. Historical linguistic insights shed light on the possible geographical location of the two respective homelands. Controversy regarding the origins of rice cultivation that has been waged since 1883 has, by and large, been resolved, though the story is complex. The palaeobotanical evidence is evaluated in light of molecular genetic findings on Asian rice. The overall picture is enhanced by insights from the genetics of Asian human populations, including modern Austroasiatic and Hmong-Mien language communities.

## 2 Phylogeny and linguistic palaeontology

On the basis of linguistic and ethnographic arguments, the presence of Austroasiatic populations in South and Southeast Asia has long been held to antedate the advent of Indo-European and Kradai, alias Daic, and perhaps even Tibeto-Burman and Dravidian in this region. Most conceivable theories about the ancestral homeland of Austroasiatic have already been put forward. 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. Today the principal contenders for the Austroasiatic homeland are the Indian subcontinent, mainland Southeast Asia and the middle Yangtze.

From a purely linguistic point of view, the location of the Austroasiatic ancestral homeland can be argued principally on several grounds. These include the geographical centre of gravity of the family based on the distribution of modern Austroasiatic language communities and the deepest phylogenetic divisions in the family tree, as well as historical grammar and linguistic palaeontology. The phylogeny of Austroasiatic presented by Gérard Diffloth is shown in Figure 1. Paul Sidwell has recently presented a rival model (see Sidwell and Blench, this volume), but a comparative appraisal of the two competing phylogenetic models would fall well beyond the scope of this article.

In view of the deepest historical division in the family's linguistic phylogeny, between Munda in the west and Khasi-Aslian in the east, the geographical distribution of the modern language communities would put the geographical centre of the family on either

side of the Ganges and Brahmaputra delta. Even the deepest division within the eastern or Khasi-Aslian trunk, that is the split into Khasi-Pakanic and Mon-Khmer, would suggest a point of dispersal for Khasi-Aslian between South Asia proper and Southeast Asia proper, somewhere in the northern littoral of the Bay of Bengal.

Historical grammarians have long observed typological differences between the Munda and the Khasi-Aslian branches of Austroasiatic. These structural differences raise the historical linguistic question of which grammatical features represent the original state and which structural traits are innovations. Donegan & Stampe (1983, 2004) propose that Austroasiatic spread from the Indian subcontinent to Southeast Asia, but argue, paradoxically, that 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-Aslian language communities migrated towards Southeast Asia. In their view, the typological changes in Munda were triggered by a prosodic shift to a falling rhythm, whereas the analytic head-initial typology observed in Khasi-Aslian languages reflects the more original Austroasiatic state of affairs. By contrast, Zide & 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-Aslian languages of Southeast Asia. It is interesting to note that these opposing views of Austroasiatic historical grammar both presume a similar view with regard to the whereabouts of an original Austroasiatic homeland.

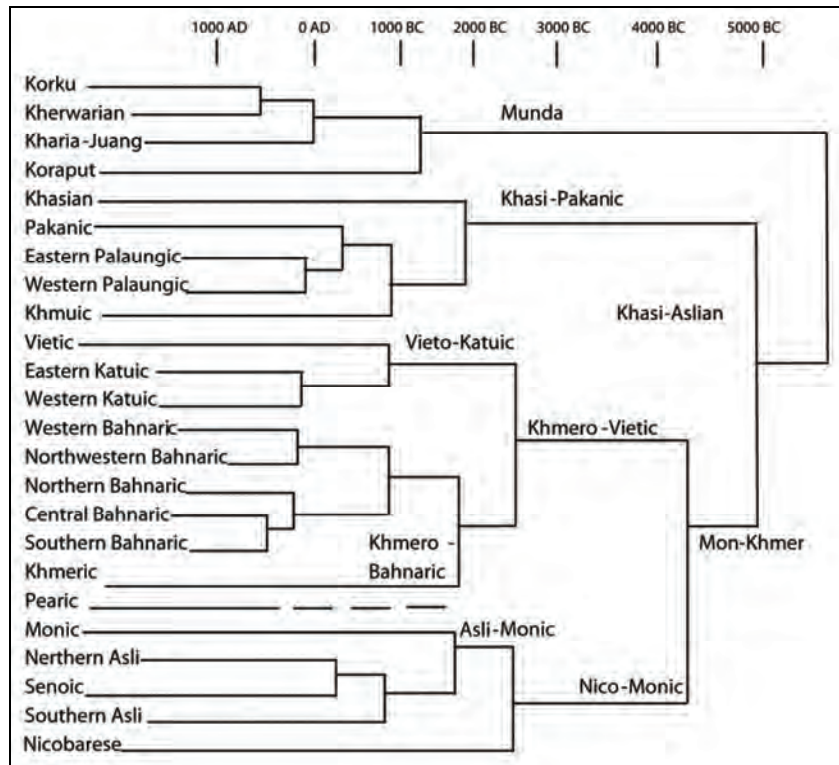
Alongside linguistic phylogeny and historical grammar, linguistic palaeontology offers a window onto the Austroasiatic past. Linguistic palaeontology, a term introduced by Adolphe Pictet in 1859, is an attempt to understand the ancient material culture of a language family on the basis of the lexical items which can be reliably reconstructed for the common ancestral language. The nature of the reconstructible Austroasiatic lexicon can be assessed in relation to the findings of palaeoecology, palaeobotany and archaeology.

Diffloth has shown that the reconstructible Austroasiatic lexicon paints the picture of a fauna, flora and ecology of a tropical humid homeland environment. The reconstructible lexicon possesses three salient isoglosses diagnostic for the faunal ecology of the Proto-Austroasiatic homeland reconstructible all the way to the Austroasiatic level and reflected in all branches of the family. The etyma \*mra:k ‘peacock *Pavo muticus*’, \*tərkuət ‘tree monitor lizard *Varanus nebulosus* or *bengalensis*’ and \*tənyu:ʔ ‘binturong’ or the ‘bear cat *Arctitis binturong*’, a black tropical mammal that is the largest of the civet cats (Diffloth 2005:78). All of these species are not native to areas that currently lie within China, and, to our present knowledge, these species were never native to the area that is today China. Such linguistic palaeontological evidence therefore appears to render the middle Yangtze homeland hypothesis less likely.

Diffloth also adduced additional Proto-Austroasiatic roots indicative of a tropical or subtropical climate, viz. \*(bən)jo:l ~ \*j(ərm)o:l ‘pangolin, *Manis javanica*’, \*dəkan ‘bamboo rat, *Rhizomys sumatrensis*’ (an Austroasiatic root which has found its way into Malay as a loan), \*kaciaŋ ‘the Asian elephant, *Elephas maximus*’, \*kiaē ‘mountain goat, *Capricornis sumatrensis*’, \*rəma:s ‘rhinoceros, *Dicerorhinus sumatrensis*’ and \*tənriak ‘buffalo, *Bubalus bubalus*’.

Osada (1995) and others have argued that the reconstructible lexicon strongly qualifies the ancient Austroasiatics as the most likely candidates for the first cultivators of rice. Diffloth (2005:78) adduces a rich repertoire of reconstructible roots representing rice, robustly reflected in all branches of Austroasiatic, viz. \*(kə)ba:ʔ ‘rice plant’, \*rəŋko:ʔ ‘rice grain’, \*cəŋka:m ‘rice outer husk’, \*kəndək ‘rice inner husk’, \*phe:ʔ ‘rice bran’, \*təmpal ‘mortar’, \*jənreʔ ‘pestle’, \*jəmpiər ‘winnowing tray’, \*gu:m ‘to winnow’, \*jərmuəl

‘dibbling stick’ and \*kəntu:ʔ ‘rice complement’, that is accompanying cooked food other than rice.



**Figure 1:** Gérard Diffloth’s (2009) Austroasiatic phylogeny with his tentative calibration of time depths (Diffloth 2001, 2005) for the various branches of the family. The two main trunks of the Austroasiatic tree are Munda in eastern, northeastern and central India and Khasi-Aslian, which stretches from the Meghālaya in the northeast of the subcontinent to the Nicobars, Malay peninsula and Mekong delta in Southeast Asia. The precise phylogenetic position of Pearic remains uncertain.

Nicole Revel (1988) contributed one of the most elaborate ethnobotanical studies on rice, rice cultivation practices and rice terminology in various Asian language communities. Her copious data already suggested the ancestral Hmong-Mien as another possible candidate for the early cultivators of rice. Progress has been made in our understanding of Hmong-Mien historical phonology, viz. Haudricourt (1954), Purnell (1970), Wáng & Máo (1995), Niederer (1998) and Ratliff (2010). The reconstructible lexicon specific to rice cultivation in Hmong-Mien is just as impressive as the Austroasiatic repertoire.

Martha Ratliff points out that, in Hmong-Mien, ‘[u]nlike the terms for the major non-rice crops..., terms for rice are less likely to show similarities with forms with similar meaning in neighboring families. Quite remarkable is the poor correlation with terms from Kra-dai and Austronesian, the languages of major rice growing peoples’ (2004:158–159). She identifies three rice cultivation etyma as being unambiguously original to the Hmong-Mien phylum: Proto-Hmongic \*ntsuw<sup>C</sup> ‘husked rice’, Proto-Hmong-Mien \*hnrəŋH ‘cooked rice’ and \*hnən ‘rice head, head of grain’. For the latter two items, Ratliff (2010) provides Baxter and Sagart’s Old Chinese reconstructions 饌 \*ŋaŋ(?)<sup>s</sup> and 囊 \*nʃaŋ.<sup>1</sup>

1 Hmong-Mien forms are given here in Ratliff’s new phonological reconstruction.

Ratliff (2004) suggests that the reconstructible Proto-Hmong-Mien rice cultivation etyma \*mblut ‘glutinous’, \*ljiŋ ‘paddy field’, \*ljim ‘sickle’, \*ŋkjuəX ‘rice cake’, \*tuX ‘husk/pound rice’ and \*tsjeŋH ‘rice steamer’ might have a Sinitic origin. In her 2010 handbook, she relates these roots to the Old Chinese forms 秣 \*m.lut, 田 \*lʃiŋ, 鎌 \*[r]em and 粃 \*[g](r)aʔ, 擣 \*tʃuʔ and 甑 \*s-təŋ-s respectively. Yet none of these Sinitic terms are well reflected across Tibeto-Burman, and none are currently reconstructible to the ancestral Tibeto-Burman proto-language. Rather, these etyma may all very well represent loans from Hmong-Mien into Sinitic, the one branch of Tibeto-Burman with the longest history of contact with Hmong-Mien.

Additional rice cultivation etyma reconstructible to Proto-Hmong-Mien include \*mbləu ‘rice plant, paddy’ and \*mphiek ‘chaff’. Reconstructible to the Proto-Hmongic level are the etyma \*S-phjæ<sup>C</sup> ‘chaff’, \*mbljæ<sup>C</sup> ‘have food with rice’, \*ʔrin<sup>A</sup> ‘dry (rice) in sun’ and \*tshɛŋ<sup>B</sup> ‘husked rice or millet’, and the rice measure etymon \*hrau<sup>A</sup> is reconstructible to the Proto-Mienic level.

The direction of some borrowings may have been from Sinitic into Hmong-Mien. Ratliff relates Proto-Hmongic \*ʔjeŋ<sup>A</sup> and Proto-Mienic \*ʔjaŋ<sup>A</sup> ‘seedling’ to Middle Chinese 秧 ʔjang and Old Chinese \*ʔaŋ, Proto-Hmongic \*ljeŋ<sup>A</sup> ‘rice measure’ to Old Chinese 量 \*[r]aŋ, and Proto-Mienic \*hmei<sup>B</sup> ‘husked rice’ to Old Chinese 米 \*[m]ʃijʔ. Yet of these, only the latter etymon appears to be reflected in Tibeto-Burman outside of Sinitic, viz. in Bodo-Koch languages.

Rice cultivation terminology could have been borrowed into Sinitic from ancient Hmong-Mien rice cultivators at a time when Proto-Sinitic millet growers came into increasing contact with their southern neighbours. The main split in the Hmong-Mien family is between Hmong and Mien, and the scattered distributions of modern communities of either branch have roughly the same geographical range, which is roughly bisected by the Pearl River. On the basis of the historical sources, it has long been mooted that the geographical centre of gravity of the family would originally have lain further north along the middle Yangtze (Cushman 1970). There is currently no palaeobotanical evidence for the co-cultivation of rice and foxtail millet along the middle Yangtze until around 3800 BC (Nasu et al. 2006).

### 3 Rice domestications and the archaeological record

The rice story is complex, and the plot of the story has changed more than once in the course of time. Alphonse-Louis-Pierre Pyrame de Candolle, director of the botanical garden in Geneva, argued that the origin of cultivated rice lay in China and that rice was introduced to India from China (1883:285, 309–311). Nikolai Ivanovič Vavilov (1926) paid tribute to Pyrame de Candolle, but argued against his claim of a Chinese origin for rice. In his elaborate model of the centres of plant domestication,<sup>2</sup> Vavilov contended instead that the origin of Asian rice lay in India, whence the crop had spread to China and Japan. The controversy has continued until recently, when molecular genetics was applied to the study of rice.

At times, South Asia would have the upper hand, as when Haudricourt & Hédin (1987:159–161, 176) proclaimed that the origin of rice cultivation lay ‘incontestablement’ in the Indian subcontinent. Afterwards, the consensus of scholarship moved the homeland

2 Themes repeat themselves in history, and Purugganan and Fuller’s map ‘Centres of plant domestication’ (2009:844) harks back to Vavilov’s 1926 map.

of rice agriculture from the Ganges to the Yangtze, and for a number of years conventional wisdom dictated that rice was domesticated in the Middle Yangtze, perhaps as early as the sixth millennium BC. Yet this idea was challenged by champions of an Indian homeland for rice. Palaeobotanical remains recovered at archaeological sites were adduced as evidence. I have discussed these shifting stances in the older rice literature elsewhere (van Driem 2001:324–327).

Archaeologists favouring a homeland for rice cultivation either in India or in China have been prone to exaggerate the antiquity and nature of the evidence. Morphometric criteria for distinguishing between actual rice domestication as opposed to mere rice cultivation were also not strictly observed. Both tendencies prompted Fuller, Harvey and Qin to exhort archaeologists ‘to cease and desist in presuming that all finds are domesticated and equate to agriculture’ (2007:328). Actual rice domestication is evinced by phenotypically observable genetic modification. Features such as long vs. short awns, shattering vs. non-shattering panicles, open vs. densely packed panicles and high dormancy vs. uniform germination are not always easy to establish by studying the microanatomy of archaeologically attested rice remains.

The chilly Younger Dryas, which lasted from ca. 10,700 to 9500 BC, is sometimes thought to have compelled the people behind the early Natufian culture in the Levant to turn to agriculture and cereal cultivation. Further east at this time, on the semi-arid Gangetic plain at the end of the mid-Holocene wet period, habitats for wild rices increasingly shifted to oxbow ponds as palaeochannels dried up and turned into oxbows. This shift favoured monsoonal rather than marshland rice species, including *Oryza nivara*, the wild progenitor of *Oryza indica* (Fuller 2006a).

The proponents of an original homeland for rice cultivation on the Indian subcontinent pointed at the presence of domesticated rice and ceramic culture from the Gangetic basin and at Doab sites such as Koldihawa and Mahagarha, dating from the seventh millennium BC (Sharma et al. 1980; Pal 1990; Agrawal 2002). Later, newer sites with rice remains and more reliable dates were reported at Lahuradewa (Lahurādevā), Tokwa (Ṭokuvā) and Sarāī Nahar Rāī.

At Lahuradewa, the early farming phase of period 1A in the site’s stratigraphy has radiocarbon dates ranging from ca. 5300 to 4300 BC. Carbonised material from this period was collected by the flotation method, yielding *Setaria pumila* and *Oryza rufipogon* as well as a morphologically distinct, fully domesticated form of rice ‘comparable to cultivated *Oryza sativa*’ (Tewari et al. 2003). More recently, accelerator mass spectroscopy dates were obtained on the rice grains themselves, corroborating the antiquity of rice agriculture at the site (Tewari et al. 2006, 2009). However, the preserved awn bases and attached rachillae in the Lahuradewa material suggest wild rice, probably *Oryza nivara*. Moreover, husk patterns are not a good diagnostic for distinguishing between *rufipogon*, *nivara* and *sativa* rice (Fuller 2006a, 2008; Fuller & Qin 2009).

Rice was certainly cultivated in the Gangetic basin by 7000 BC, but the current evidence for the actual domestication of *nivara* rice in the middle Ganges dates from no earlier than the third millennium BC. Rice agriculture is reported by Pokharia (2008) from the third millennium BC at the Ganges site of Tokwa near Allahabad. Fuller stresses that meticulous study will have to verify whether Tokwa is of greater antiquity than the second millennium BC, and problems arise from the allegedly disturbed stratigraphy of a few middle Ganges sites. At the same time, a reassessment of radiocarbon datings has tended to move crop domestication in the Indian subcontinent and Southeast Asia from a hoary past into the fifth millennium BC and may have chronological implications for the dawning of

the Neolithic horizon in some areas, for example Higham (2002, 2004), Fuller, Boivin & Korisettar (2007), Fuller, Harvey & Qin (2007).

Moreover, a fine distinction is being made between the cultivation and the actual domestication of crops, based on precise morphometric studies of phenotypical evidence in the form of grain size and anatomical changes accompanying the reduction of seed shattering. Yet the transition manifests itself as a longer and more gradual process in the archaeological record for Asian rice than seen for comparable morphological changes in barley and einkorn wheat in the Levant (Purugganan & Fuller 2009). Indeed, the genetic pathways which led to reduced seed shattering in different domesticated grasses such as wheat, barley, rice, etc. are many and species-specific (Li & Gill 2006). On the Gangetic plain, therefore, groups were using rice by 7000 BC, but between this time and the appearance of better evidence for agriculture, dating from 2500 to 2000 BC, the current archaeological record in the Indian subcontinent offers little evidence from which to infer cultivation practices or the evolution of rice domestication traits.

Similarly in East Asia, claims have been published dating the earliest rice cultivation in the Yangtze basin to as long ago as 10,000 BC, cf. Chén (1999). Certainly, sedentary settlements of the 彭頭山 Péngtóushān culture flourish along the middle Yangtze between 8000 and 6000 BC. However, the currently available evidence indicates that immature morphologically wild rice was cultivated by foragers before the actual domestication of the crop. This state of affairs is in evidence along the middle Yangtze at the Péngtóushān culture site 八十擋 Bāshídàng (7000–6000 BC) as well as in the Hángzhōu Bay littoral around the Yangtze delta at sites such as 跨湖橋 Kuàhúqiáo, 馬家浜 Mǎjiābāng (5000–3000 BC) and 河姆渡 Hémǔdù (5000–4500 BC).

The appearance of a minority of domesticated-type spikelet bases alongside the palynological and micro-charcoal data from Kuàhúqiáo suggest clearance by burning and water management already between 6000 and 5700 BC (Fuller, Harvey & Qin 2007; Fuller & Qin 2009). Yet the actual domestication of *japonica* type rice through genetic modification by selective breeding was probably effectuated along the lower Yangtze in the fifth millennium BC by people, who previously relied far more heavily on the collecting of acorns, water chestnuts and foxnuts before becoming reliant on rice cultivation. Morphological studies of rachides (spikelet bases) found at 田螺山 Tiánluóshān, however, show that cultivated rice first underwent domestication in the Lower Yangtze between 4900 and 4600 BC. This transition is evident in the increased proportion of rice versus other plant remains and the increased proportion of rice with non-shattering panicles. Dates were obtained by direct accelerator mass spectrometry radiocarbon dating on nuts and rice grains (Fuller et al. 2009; Nakamura 2010; Zhao 2010).

The picture which emerges is that the beginnings of rice agriculture were gradual, and the domestication process was initially slow and finished significantly later than often assumed. Over the course of the Chinese Neolithic, nut-gatherers switched to being farmers. The wild progenitor *Oryza rufipogon* was not fully domesticated in the lower Yangtze to yield early *Oryza japonica* until ca. 4000 BC (Fuller 2005a, 2005b, 2005c, 2006a, 2006b, 2006c, 2007a, 2007b; Yasuda 2002; Zong et al. 2007). Unambiguous morphological evidence of domesticated rice appears in the middle Yangtze ca. 4500 BC and in the lower Yangtze ca. 4000 BC (Fuller, Harvey & Qin 2007).

The cultivation of *japonica* rice spread from the Yangtze basin core area only after 4600 BC (Fuller & Qin 2009; Ruddiman et al. 2008). Rice reached the Yellow River basin during the third millennium BC (Crawford & Shen 1998). If the conjectural dates from the Chulmun site are tentatively accepted, then rice cultivation could also have entered the

Korean peninsula in the course of the third millennium BC (Ahn 2010), whereas paddy agriculture appears in the Japanese archaeological record in the course of the first millennium BC (Tanaka et al. 2010). Rice cultivation reaches Formosa and Vietnam between 2500 and 2000 BC (Higham & Lu 1998), but only spreads throughout the Indochinese peninsula between 1500 and 500 BC (Weber et al. 2010). Yet, whereas palaeobotanical investigations yield highly valuable and insightful evidence, molecular genetic findings have provided even more crucial insights to the issue of rice domestication.

#### 4 From morphometrics to rice genetics

By the 1990s, it had been understood that there were two principal domesticated varieties of rice, *Oryza sativa*, var. *indica*, and *Oryza sativa*, var. *japonica*. Some held that the *javanica* cultivar represented yet a third type of cultivated Asian rice.<sup>3</sup> The two cultivars *indica* and *japonica* were held to be phylogenetically distinct and to have distinct wild progenitors, which were identified as being *Oryza nivara* and *rufipogon* respectively. *Oryza rufipogon* is a perennial swamp species ranging throughout South and Southeast Asia and southern China. *Oryza nivara* is an annual which grows in moist soil or shallow water and ranges throughout South Asia and mainland Southeast Asia. Both species grow sympatrically and naturally hybridise with each other as well as with cultivated rice. Early genetic studies corroborated this view (Vaughan 1994).

The old controversy between an Indian versus a Chinese homeland for the origin of cultivated rice was rendered obsolete, although the new view did not dampen the fervour of those who strove to turn up the earliest evidence for rice domestication in India and in China. In fact, this archaeological contest was given a new impetus by the realisation amongst rice specialists that *Oryza rufipogon* actually constitutes a single diverse species encompassing both the annual self-pollinating *Oryza nivara*, adapted to disturbed shallow-water environments, and the perennial *Oryza rufipogon* in the strict sense, which can out cross and is adapted to stable deep-water environments. Some researchers still prefer to treat the two as distinct species, whereas others consider them to be distinct ecotypes under a single *rufipogon* species complex.

Genetically, there is no significant barrier to reproduction, and the two varieties essentially share a common genome. The different ecological niches aid and abet isolation and diversification within the species, but hybridisation is common. On this issue, Chen et al. (2008) have identified the S5 gene, which encodes for an aspartic protease which conditions fertile embryo sacs, as a major locus enforcing hybrid sterility between *indica* and *japonica* rice, and thereby contributing to the observed high rate of post-zygotic reproductive isolation of the two cultivars. Yet gene flow is facilitated by a third neutral allele which, when it occurs, yields fertile hybrids with either species.

In the 1970s, rice dwarf virus devastated many high-yield paddy fields in India and Indonesia. When resistance genes were found in *Oryza nivara*, the wild *nivara* was crossed with the most widely used of the afflicted *sativa* strains, yielding a hybrid which now

---

3 In contrast to Asian rice, African cultivated rice (*Oryza glaberrima*) is genetically related most intimately to the wild species *Oryza barthii* and, to a lesser extent, to *Oryza longistaminata*. African rice represents an independent domestication of a cultivated rice species geographically and genetically distinct from Asian rice, as shown by analyses of mitochondrial, chloroplast and nuclear DNA markers (Duan et al. 2007). Linguistic and genetic evidence both indicate that African rice was domesticated independently of Asian rice, probably in the upper Niger river delta (Blench 2006; Sweeney & McCouch 2007).

covers over 100,000 km<sup>2</sup> of Asian paddy fields. By the same token, so-called ‘weedy rice’ arises from the natural hybridisation of cultivated rice with wild *Oryza rufipogon* or *nivara*. Such de-domesticated rice was first detected in Sri Lanka in 1992, and today much yield loss in Ampara and Puttalam districts is attributed by farmers to the prevalence of weedy rice. Thus, it is defensible to view *Oryza rufipogon*, including *nivara*, as a single species.

Notwithstanding arguments for the single species status of *rufipogon* rice *sensu lato*, studies of the rice genome indicate independent domestications for the *indica* and *japonica* varieties of rice. This conclusion is based both on the time depth of the calculated divergence of their respective wild progenitors, viz. somewhere between 200,000 and 400,000 years ago (Ma & Bennetzen 2004; Vitte et al. 2004; Zhu & Ge 2005), and the independent adaptive selection acting on the *Hd6* photosensitivity gene, which enabled rice to spread beyond the confines of the tropics and subtropics to be cultivated globally from 53°N and 40°S (Yamane et al. 2009).

Londo et al. (2006) conducted a phylogeographic analysis of 203 cultivars of domesticated rice and over 129 populations of wild rice, that is *Oryza rufipogon* and *nivara*, spanning the entire natural geographical range of wild rice. The three distinct gene regions investigated (that is the maternally inherited chloroplast marker *atpB-rbcL*, the neutral nuclear pseudogene *p-VATPase*, presumed to be under little selective constraint, and the functional nuclear gene *S*-adenosyl methionine synthetase) yielded a consistent signal of domestication from distinct wild ancestral gene pools in two distinct geographical regions. Whilst the origin of the *Oryza sativa indica* lay in the region encircling the Bay of Bengal, which stretches from eastern and northeastern India to modern Burma and Thailand, the origin of the *japonica* cultivar lay in the Middle Yangtze basin in what today is southern China. Moreover, Londo et al. (2006) opine that *ghaiyā* or upland rice could represent a third distinct domestication event in or near the Indian subcontinent.

A major quantitative trait locus on the rice genome, *sh4*, is a single nucleotide substitution (SNP) encoding for a transcription factor. The consequent substitution of the amino acid lysine by asparagine in the Myb3 DNA-binding domain led to only partial development of the abscission zone where the mature grain detaches from the pedicle. Diminished gene function, expressed in the reduced brittleness of the rachides, is responsible for the reduction of grain shattering. The wild progenitors of domesticated rice exhibiting the latter trait were selected by early farmers to yield the non-shattering rice cultivars which could be efficiently harvested without the loss of grain. Yet it appears that the gene *sh4* had a single origin (Li et al. 2006a, 2006b; Lin et al. 2007; Onishi et al. 2007), whereas the phylogenetic evidence indicates separate domestications for *japonica* and *indica* rice. Therefore, the trait *sh4* is believed to have spread by introgression from its initial source, probably an early variety of domesticated rice, to other rice cultivars and also to the wild progenitors (Sweeney & McCouch 2007; Sang 2009).

Current genetic data therefore indicate the occurrence of two independent domestications, which were not entirely independent of each other. The available genetic data indicate overall gene flow from *japonica* to *indica*, but the set of domestication alleles studied to date is still limited, and this impression could change in the future (Kovach et al. 2007). Assuming that future research does not undermine this impression,<sup>4</sup> the most parsimonious model would assume that the *sh4* single nucleotide mutation first arose in

---

4 Since the synthesis of the genetics literature presented here was first circulated in 2009, portions of this reconstruction have been echoed in two subsequent discussions of the rice genetics literature (Purugganan 2010; Fuller et al. 2010).



*japonica* and then later spread back into the gene pool by introgression. It turns out that this rather complex genetic picture makes fairly good sense of the currently available palaeobotanical evidence.

The available morphometric evidence for Neolithic Gangetic rice, as collected and assessed by Harvey et al. (2006), indicates that grains are on the small side, congruent with morphologically wild rice, and perhaps containing immature grains. The cumulative evidence could be seen as indicating a long history of cultivation in the Ganges. This process initially involved the gathering of morphologically wild rice, with the evolution of non-shattering forms occurring quite late after the introduction of domesticated *japonica* rice carrying the key mutations, especially the recessive *sh4* allele, and perhaps the white pericarp *Rc* mutation (Kovach et al. 2007; Sweeney & McCouch 2007).

## 5 Rice domestications and geography

The reconstructible lexical evidence warrants us entertaining the hypothesis that the Proto-Hmong-Mien were the initial cultivators of *japonica* rice, and that ancient Austroasiatics were the initial cultivators of *indica* rice. The Austroasiatic and Hmong-Mien homelands need not have been coeval. A homeland is just a geographical region where a proto-language is thought to have been spoken at one point in time. The homelands of two language phyla may not necessarily have co-existed in time, as the homelands of different language families could be attributed to distinct time depths in the past. On a grander scale, no putative linguistic homeland was the original cradle of mankind.

However, for the first time since 1883, a coherent story about the domestication of Asian rice has emerged that could tell us where these homelands may have lain. Yet the plot of this tale may change again, and we can anticipate whence possible changes in the story line might come. Today both the palaeobotanical picture and the available genetic evidence remain incomplete.

Palaeobotanical evidence for an early domestication of Asian rice might as yet turn up in India. More crucially, despite the archaeological work conducted in the Ganges and Yangtze basins, much of the archaeology of ancient rice agriculture remains simply unknown because no substantive archaeological work has been done on the Neolithic in the most relevant areas, for example northeastern India, Bangladesh, the Indo-Chinese borderlands and Burma. The absence of evidence does not constitute evidence of absence, and the sheer dearth of archaeological research in these areas leaves entirely open the possibility that rice cultivation may have originated in this region.

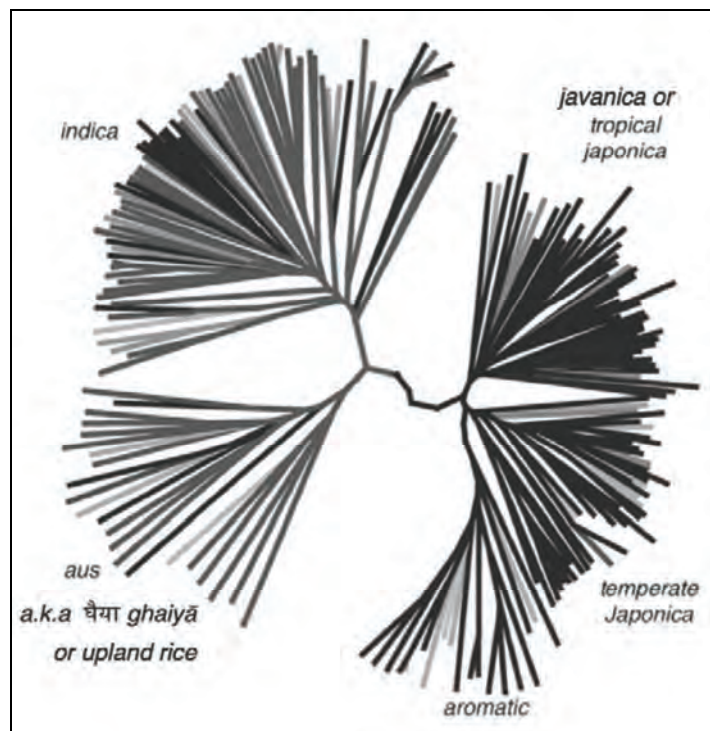
Epistemologically, therefore, the situations in the Indochinese peninsula versus northeastern India and Burma are essentially contrasting. Whereas rice agriculture currently appears in the archaeological record of the Indochinese peninsula between 1500 and 500 BC (Weber et al. 2010), the absence of Neolithic sites is not the crucial empirical issue in Burma and northeastern India. Rather, the pivotal issue is the absence of archaeological research to the present day in this vast region. Whilst a Lower Mekong homeland hypothesis for Austroasiatic is, in view of the linguistic palaeontological evidence, severely challenged by the lack of early sites for rice agriculture in mainland Southeast Asia, the sheer archaeological neglect of the Salween, Irrawaddy and Lower Brahmaputran basins simply does not permit any inferences one way or the other.

A related but distinct empirical issue is the archaeological recoverability of rice agriculture sites. The traces of ancient farming communities would tend to have been

better preserved in the hill tracts surrounding the Brahmaputran flood plains than on the fertile fields themselves.<sup>5</sup> Yet 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 silty sediments of the sinuous lower Brahmaputran basin. Maybe the palaeobotanical evidence for the earliest domestication of rice was washed out by the Brahmaputra long ago and now lies submerged in the depths of the Bay of Bengal.

In future, genetic studies may likewise recast the rice story in a new light as more high-resolution genotyping and resequencing enables the identification of relevant markers of selection by domestication. Already the available genetic data on Asian rice yields information directly relevant to the hypothesis of ancient Austroasiatics and Hmong-Mien as the earliest rice cultivators of Asian rice. These considerations emerge from the diversity of the rice genome itself.

Genetic studies have provided a new view of Asian rice cultivars. Both *javanica*, now often called ‘tropical *japonica*’, as well as ‘temperate *japonica*’ issued from the same domestication process (Figure 2).



**Figure 2:** An unrooted phylogenetic tree of the diversity of *Oryza sativa* based on 169 nuclear simple sequence repeats (SSRs) and two chloroplast markers in 234 cultivars of *Oryza sativa*, adapted from Garris et al. (2005) and Kovach et al. (2007).

Despite the length of the grains and the Indian associations evoked by some aromatic cultivars in the minds of rice lovers today, the aromatics such as *bāsmati* rice likewise ultimately originate from the *japonica* domestication in the Yangtze basin (Parsons et al. 1999; Garris et al. 2005). By contrast, *ghaiyā* or upland rice is most closely affiliated with *indica* rice. However, some geneticists have mooted the possibility that upland rice or

5 Even in East Asia, many salvageable rice agriculture sites are in the foothills or at the base of the foothills (Nakamura 2010).

‘haplotype D’ rice represents a third distinct domestication event which unfolded in or near the Indian subcontinent (Londo et al. 2006).

In the rice genetics literature, upland or dry-cultivated rice sometimes currently goes by the Bengali name আউশ *āuš*, anglicised as ‘aus’, an orthography which belies the pronunciation.<sup>6</sup> The Dzongkha name for upland rice in Bhutan is simply *kambjâ* ‘dry paddy’, whereas in Nepal upland rice goes by the proper name of घैया *ghaiyā* (Figure 3). Whether or not *ghaiyā* represents a third domestication event, upland rice is genetically more closely affiliated to *indica* rice. Upland rice has a short growing season, maturing early in the Nepali month of Āṣādh (Asār), that is June-July. Upland rice is tolerant to drought and requires little or no irrigation. These adaptive traits have rendered *ghaiyā* ideally suited to hillside niche environments in Nepal, Bhutan, the Himalayan foothills, Bangladesh and northeastern India, where it is still commonly grown.

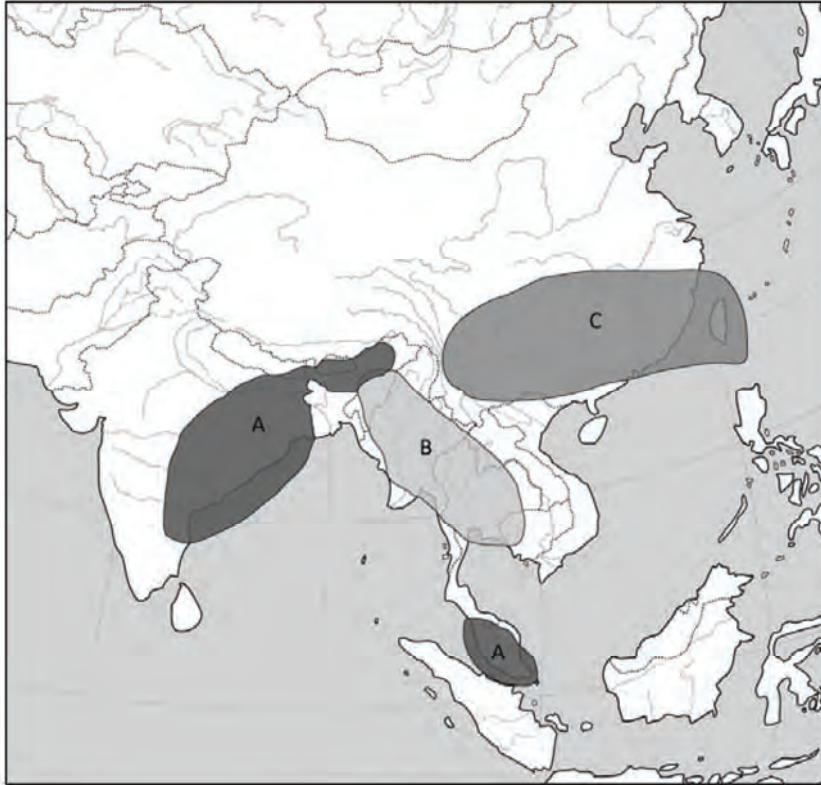
Based on the genetic markers in the wild precursor *Oryza rufipogon*, geneticists have identified the likely geographical ranges for the domestication of (A) *ghaiyā* or upland rice, (B) wet *indica* rice and (C) the *japonica* cultivar, shown in Figure 4. Whereas *japonica* rice is likely to have been domesticated somewhere in the Yangtze basin and its periphery, the most likely region for the domestication of wet *indica* rice stretches from the Brahmaputra to the Mekong. The likely geographical range for an independent domestication of upland rice, however, stretches from the Brahmaputra basin in the northeast to Pulicat Lake on the Coromandel Coast in the southwest and includes a discontinuous patch at the end of the Malay peninsula.



**Figure 3:** घैया *Ghaiyā* or upland rice (courtesy of Hem Bahādur Thāpā, alias Himāl bhānjā).

If Asian rice was indeed domesticated thrice instead of twice, then this finding would raise a whole set of new questions. However, if wet *indica* and upland rice diverged after a single initial domestication process, then the primordial *indica* cultivar may have been domesticated in a complex habitat which included both wet lowlands and dryer hills. Northeastern India and the Indo-Burmese borderlands straddle the area where the geographical range for the first domestication of wet *indica* rice and dry upland rice overlap, that is areas A and B in Figure 4.

6 An English speaker could come close to the Bengali pronunciation by pronouncing the name as ‘ouch’, but substituting an *sh* sound for the *ch* at the end of the word. Alternatively, the Nepali name *ghaiyā* might be a better alternative.



**Figure 4:** The geographical ranges for the domestication of (A) *ghaiyā* or upland rice, (B) wet *indica* rice and (C) the *japonica* cultivar, based on the genetic markers in the wild precursor *Oryza rufipogon* (adapted from Londo et al. 2006).

What else can be said about this area of overlap? First, northeastern India falls within the area whence linguistic phylogeny suggests that the ancient Austroasiatics could have originated. Second, this area forms part of a vast archaeologically largely unexplored region. Third, northeastern India and the Indo-Burmese borderlands have maintained highly diverse rice cultures to the present day.

At least five species of wild rice are native to northeastern India, viz. *Oryza nivara*, *Oryza officianalis* (*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*, that is mid February to early April. The seedlings are not transplanted but ripen in just four months in fields which must be constantly weeded. In *bāu 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 *śāli kheti*, the rice is sown from mid May to mid June, and the seedlings are transplanted. *Śāli kheti* rice varieties are suspected to derive from the wild *officianalis* 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. Instead, wild *rufipogon* rice is used in Assam and other parts of northeastern India as cattle feed (Hazarika 2006b).

Beyond linguistic phylogeny, historical grammar, linguistic palaeontology, palaeobotanical findings and rice genetics, the homeland question can be argued on the basis of which genetic gradient or haplogroups in modern human populations might be

credibly related to either an ancient linguistic intrusion or to a demic spread of the language family in prehistory.

## **6 Father tongues and paternal homelands**

Often languages appear to show a correlation with the geographical distribution of genetic markers in the populations speaking these languages. Yet often enough there is a disconnect between the genetic affiliation of a population for a certain marker and the linguistic affinity of their language. The linguistic ancestors of a language community were not necessarily the same people as the biological ancestors of that community. Such discrepancies can often be as informative for our understanding of the sociolinguistic past as are the grand correlations. Moreover, the time depth accessible to geneticists stretches far beyond the ken of the linguistically reconstructible past based on systematic correspondences between related languages studied by means of the comparative method.

Human population genetics has made vast strides forward in recent decades. Classical population genetics developed mathematical models such as the Hardy-Weinberg equilibrium, quantified genetic diversity and distance, measured admixture and the effects of population size on selection and drift. In an attempt to reconstruct prehistoric migrations, Cavalli-Sforza, Menozzi & Piazza (1994) famously developed principal component analysis. However, since then, such summary statistical methods have been superseded by the more refined instrumentarium of modern human genetic phylogeography, which is not constrained to dealing with hypothetical prehistoric populations as discrete entities.

Human genetic phylogeography is largely based on the study of STRs and SNPs. Short tandem repeats (STRs) or ‘microsatellites’ are repeats of short segments of DNA generally less than five nucleotides in length. Microsatellites have a high rate of mutation and a large number of alleles. Single nucleotide polymorphisms (SNPs) are generally biallelic and tend to have a lower rate of mutation. Paternal ancestry of a population can be traced through polymorphisms on the paternally inherited Y chromosome, whereas the maternal ancestry of a population can be reconstructed on the polymorphisms in the maternally inherited mitochondrial DNA.<sup>7</sup> Ever more salient markers are being identified on the more numerous autosomes that will be as informative for reconstructing population prehistory as the polymorphisms on the non-recombinant portion of the genome. Human genetic phylogeography employs laborious formal methods, such as founder analysis, in order to date and measure prehistoric human migrations.

The linguistic affinity of a language often correlates better with the paternally inherited polymorphisms of the populations which speak that language than do other markers. I have called this tendency, recognised by Poloni et al. (1997, 2000), the father tongue hypothesis. Our genetic ancestry shows what I have described with an anatomical metaphor as sexual dimorphism, and so an uneven correlation with our linguistic prehistory. Some languages appear to be mother tongues, whereas others show up as father tongues. At many times and in many places in prehistory, the father tongue appears to have been the guiding mechanism in language shift (van Driem 2005, 2006, 2007). The dynamics of a process whereby mothers passed on the language of their spouses to their offspring has major

---

7 Only rarely does paternal mitochondrial DNA from the base of the sperm flagellum survive the process of selective destruction, dilution and inactivation during early embryogenesis by the vast surplus of oocyte mitochondria (Schwartz & Vissing 2002; Williams 2002). The reported case remains an anomaly in humans.

implications for our understanding of language change and so for historical linguistics and historical sociolinguistics in general.

Genetic studies tend often to corroborate models of ethnolinguistic prehistory previously mooted by linguists and ethnographers. For example, the spread of Indo-Aryan languages unambiguously attests to an ancient linguistic intrusion into the Subcontinent from the northwest. The Indo-Iranian intrusion onto the Iranian plateau and into the Subcontinent from the northwest appears also to have left a corresponding genetic legacy in the form of Y haplogroups M11-L, M17-R1a and M124-R2 spreading from the northwest 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; Sahoo et al. 2006).<sup>8</sup>

Similarly, a population genetic study of 23 Hân populations (Wen et al. 2004) has corroborated the picture which linguists and historians have long entertained, of a martial and therefore male-biased Hân expansion southward during the sinification of what today is southern China. The southern Hân paternal lineages, identified by Wen et al. (2004) with haplogroups M122-O3 and M134-O3e (O3a3c), show preponderant northern Hân penetration alongside a faint pre-Sinitic signature. Males from the north were the primary contributor to the paternal gene pool of southern Hân populations, whereas the mitochondrial DNA of southern Hân populations contains roughly equal contributions from pre-Sinitic and Hân maternal ancestors.

Yet certain languages are clearly mother tongues in that the mitochondrial markers of their speakers correlate with those of populations speaking related languages, whereas their Y chromosomal markers show no such correlation. The conservative Tibetan languages or mother tongues of Baltistan are a case in point (Poloni et al. 1997, 2000; Zerjal et al. 1997; Quintana-Murci et al. 2001; Qamar et al. 2002). Sometimes the genetic markers of a population studied to date show no clear correlation with the language spoken by a population. The evident lack amongst Hungarians of the TatC deletion defining the Y chromosomal haplogroup Tat-N3 (N1c),<sup>9</sup> despite its prevalence amongst other Uralic language communities, sheds some additional light on the complex sociolinguistic history of the migrations which have passed through Pannonia and the Magyar linguistic assimilation. Modern language communities have diverse ancestries, resulting from different initial waves of peopling as well as subsequent migrations (Li et al. 1999).

Any first successful human colonisation of a new land mass necessarily involves women. If an island were first settled exclusively by men, then these colonists would leave no progeny. Once women are in place, however, migrations made up mainly or even exclusively of men can follow and successfully introduce new genetic lineages into resident populations if there are any present. Linguistic intrusions need not involve women. The introduction of Indo-Iranian languages into the Indian subcontinent has already been discussed in this regard. In a similar vein, studies of mitochondrial DNA

---

8 Claims that Y chromosomal lineages, particularly R haplogroups, of Indo-Aryan high caste populations are instead autochthonous to India (e.g. Sengupta et al. 2006; Sharma et al. 2009; Reich et al. 2009; Chakravarti 2009) appear to be based on interpretations of modern frequency gradients which fail to take into account the genetic aftermath of known migrations which altered the genetic landscape of Central Asia subsequent to the Aryan ethnolinguistic intrusion. The topic is still under debate.

9 Alongside the familiar Y chromosome haplogroup labels, I provide between parentheses the newer haplogroup labels introduced in 2008 by the Y Chromosome Consortium (Karafet et al. 2008).

lineages show that any involvement by women in the linguistic and cultural colonisation of insular Southeast Asia by Austronesians from Formosa must likewise have been demographically minor (Ballinger et al. 1992; Macauley et al. 2005; Hill et al. 2007).

Fine-grain genetic studies of human populations often show that the maternal ancestry of groups is more complex. The maternal lineages of *orang asli* groups, based on studies of control-region and coding-region markers in the mitochondrial DNA, show that *orang asli* groups represent a microcosm of demographic processes that no doubt took place throughout Asia and must often have had a comparable degree of complexity. (cf. Oppenheimer, Fix, Bulbeck, this volume.) The maternal ancestry of the Semang dates back to the initial settlement of the Malay Peninsula from Africa 50,000 years ago. The Senoi on the other hand trace half of their maternal ancestry back to the same initial settlement, whereas the other half derives from women who came to the southern part of the peninsula as part of an early Austroasiatic speaking agriculturalist settlement 4000 years ago. The maternal ancestry of the Malays is more diverse and shows affinities with both insular and mainland Southeast Asia at time depths suggesting Late Glacial, early Holocene and Neolithic dispersals (Hill et al. 2006).

The mitochondrial ancestry of language communities in what today is China is likewise complex. However, on the whole the maternal ancestry appears older and more sedentary than the more dynamic episodes in the spread of Y chromosomal lineages against this mitochondrial background (for example Yao et al. 2002; Kong et al. 2003). The mitochondrial DNA pool in East Asia tends to be locally region-specific and largely derives from two super-haplogroups M and N. Mitochondrial lineages reflect early settlement patterns in East Asia and support an early phylogenetic partitioning between northern and southern populations (Kivisild et al. 2002). The majority of Hmong-Mien populations belong to haplogroups that tend to be prevalent in southern China, for example B, R9, N9a, and M7. Historically, the Hmong have had more contact with Chinese than have the Mien, and this contact has left genetic traces in the Hmong heritage (Wen et al. 2005). Mitochondrial haplogroup M7 characteristically has a southern distribution in East Asia, especially in the Yellow Sea littoral.<sup>10</sup>

Similarly, the maternal lineages in the Indian subcontinent generally show that there was a relatively rapid dispersal out of eastern Africa, at which time the mitochondrial heritage of the Subcontinent was first firmly established. The Subcontinent shows the effects of having been at a crossroads, but mitochondrial lineages often tend to be old and to antedate the last glacial maximum (Endicott et al. 2007). On the Indian subcontinent, a study of mitochondrial lineages indicates that Austroasiatic tribal groups show a far higher diversity of maternal lineages than Dravidian tribal populations, and that their maternal lineages more closely represent those of earlier settlement. The calculated expansion times

---

10 The M7 mitochondrial daughter groups M7a and M7b2, specific to Japanese and Korean populations, attest to a pre-Jōmon contribution to the modern Japanese mitochondrial DNA pool. The estimated coalescence times for the subclades M7a, M7b, and M7c range between 6000 and 18,000 years ago. This date suggests either that these star-like clades reflect a resettlement process around the Sea of Japan from the south after the Last Glacial Maximum, contemporary with the spread of Suyanggae microblades before the onset of the Jōmon culture, or that M7a and M7b entered Japan during initial settlement over 30,000 years ago and underwent a genetic bottleneck during the Last Glacial Maximum. By contrast, the mitochondrial haplogroups A5, B5, C, F1a, N9a, and Z, which are shared between Koreans and Japanese and are virtually absent in Ryūkyūans and in the Ainu, testify to later migrations through the Korean peninsula to Japan, probably during the Yayoi agricultural intrusion 2300 years ago. The presence of the mitochondrial lineage Y1 amongst the Ainu testifies to the migration of Siberian populations to the Japanese archipelago from the north (Kivisild et al. 2002; cf. Igawa et al. 2009).

for Austroasiatic mitochondrial lineages is 17,000 years older than those of Dravidian tribal groups in the Subcontinent and 22,000 older than Indian Tibeto-Burman maternal lineages (Roychoudhury et al. 2001). Whether and how long this picture will be upheld by emergent genetic data remains to be seen.

In contrast to this prehistory of maternal lineages, the paternal ancestry of modern language communities generally shows a far more dynamic history subsequent to the first settlement of Asia by anatomically modern humans out of Africa. Single nucleotide polymorphisms (SNPs) and short tandem repeats (STRs) on the Y chromosome support two distinct northern and southern migration routes into East Asia. The northern populations expanded before the Last Glacial Maximum, presumably exploiting Siberian megafauna between 34,000 and 22,000 years ago, whereas southern populations expanded after the Last Glacial Maximum, between 18,000 and 12,000 years ago, but did so more vigorously than the northern populations. The southern expansion was more reliant on the exploitation of bountiful plant resources in the warmer more stable climates (for example Zerjal et al. 1997; Chũ et al. 1998; Sù et al. 1999, 2000; Ding et al. 2000; Shi et al. 2005; Xue et al. 2006; Karafet et al. 2008). In East Asia, the Y haplogroup M231-N has been interpreted by some geneticists as representing the legacy of ancient settlement by way of a boreal route, whereas the haplogroup M175-O and its many branches represent the more robust southern settlement via a littoral expansion. The arguments with respect to the spread of the Y haplogroup RPS4Y-C are undecided.

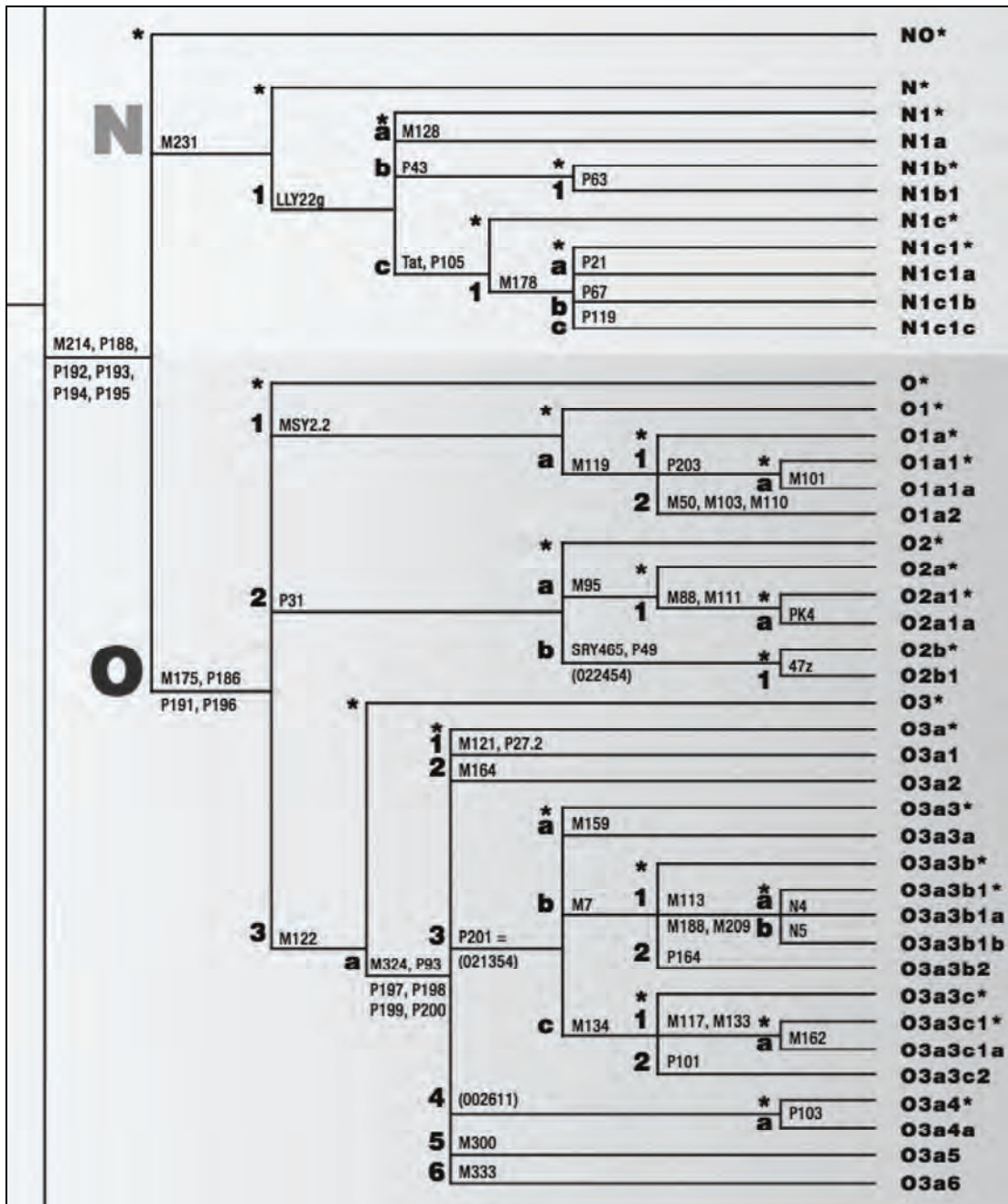
The correlation of the Y chromosomal haplogroup M134-O3e (O3a3c) with the spread of Tibeto-Burman language family has already been hypothesised (van Driem 2006). On the assumption of the veracity of the father tongue hypothesis for Austroasiatic, I identified the Y chromosomal haplogroup M95-O2a as the marker for the spread of Austroasiatic on the basis of the then available genetic data (van Driem 2007). This view has been corroborated by subsequent genetic studies, for example Kumar et al. (2007). The available genetic data also enable us now to propose a correlation of the Y chromosomal haplogroup M7-O3d (O3a3b) with the spread of Hmong-Mien. The relevant portion of the Y chromosomal tree is shown in Figure 5.

The undifferentiated Y chromosomal haplogroup M122-O3 is ancestral to both haplogroups M134-O3e (O3a3c) and M7-O3d (O3a3b), whereas the undifferentiated Y chromosomal haplogroup M175-O is ancestral to both haplogroups M122-O3 and M95-O2a. The undifferentiated haplogroups M175-O and M122-O3 currently appear at the highest frequency in Southeast Asian populations. At the same time, the high frequency of undifferentiated M122-O3 in assayed Southeast Asian Austroasiatic language communities appears to corroborate the view that the linguistic differentiation into the known linguistic phyla of Austroasiatic, Hmong-Mien and Tibeto-Burman was posterior in time to the differentiation of the haplogroups M95-O2a, M7-O3d (O3a3b) and M134-O3e (O3a3c) respectively. The precise geographical areas where these mutations took place cannot be pinpointed, but these areas no doubt lay between the Indus and the Yangtze delta, and more probably between the Brahmaputra and the Mekong delta, where the incidence of the undifferentiated haplogroups M175-O and M122-O3 is still relatively high.

Whereas the mitochondrial lineages in Austroasiatic groups are locally region-specific, and some of these lineages date back roughly to the advent of the first anatomically modern humans (Reddy et al. 2007), all Nicobarese assayed to date show the Y chromosomal haplogroup M95-O2a. Sahoo et al. (2006) find haplogroup M95-O2a to occur at a frequency of 77% in Austroasiatic groups in India, and Reddy et al. (2007) show a decreasing frequency gradient of this signature haplogroup from Mundari (53%) to Khasi (30%) to Southeast Asian Austroasiatics (ca. 23%). Assuming the veracity of the father



tongue hypothesis for the spread of Austroasiatic, the available data could be interpreted as pointing towards the Brahmaputra basin as the point of origin for this language family.



**Figure 5:** The portion of the Y chromosome phylogenetic tree relevant to the Father Tongue hypothesis with regard to Austroasiatic, Hmong-Mien, Tibeto-Burman and Austronesian and the theory regarding the peopling of eastern Asia via a boreal and a meridional route, reproduced from Karafet et al. (2008) with the kind permission of the Cold Spring Harbor Laboratory Press.

Reddy et al. (2007) measure an overall frequency of 7% for the Y haplogroup M95-O2a in Tibeto-Burman populations on the Indian subcontinent. The highest frequency of the M95-O2a haplogroup is found in tribal populations in Orissa, Chattisgarh and Jharkhand (Sengupta et al. 2006). However, the 47% frequency of the Y haplogroup M95-O2a found

by Sahoo et al. (2006) in Tibeto-Burman groups of northeastern India suggests that a subset of the paternal ancestors of particular Tibeto-Burman populations in northeastern India, for example certain Bodo-Koch communities, may originally have been Austroasiatic speakers who either married into Tibeto-Burman communities or who were linguistically assimilated by ancient Tibeto-Burmans. Indeed, median-joining network analysis of haplogroup M95-O2a microsatellite haplotypes suggested a division in the Indian subcontinent between Tibeto-Burmans vs. Austroasiatic and Dravidian language communities. The Austroasiatics and Dravidians also show greater Y-chromosomal microsatellite (that is short tandem repeat) diversification than is found in the Tibeto-Burman language communities.

Continuing to operate on the same set of hypotheses, the high incidence of the Y chromosomal haplogroup M7-O3d (O3a3b) in Austroasiatic language communities of Southeast Asia could indicate that some of the people who were assimilated by the spread of Austroasiatic into Southeast Asia may have been ancient Hmong-Mien or some related group long forgotten. The low incidence of haplogroup M95-O2a amongst Hmong-Mien speakers could indicate an Austroasiatic contribution to ancient Hmong-Mien populations, and would certainly have implied contact in the past. The incidence of the Y haplogroup M7-O3d (O3a3b) in Austroasiatic communities of the Indian subcontinent, by contrast, is undetectably low.

Perhaps the M175-O haplogroup first split up in northeastern India, with P31-O2 initially staying behind in the eastern part of the Indian subcontinent, and M122-O3 moving into the Indo-Burmese borderlands. Thence the derivative clade M134-O3e (O3a3c) headed north, and M7-O3d (O3a3b) headed east. Subsequently, Austroasiatic bearers of the M95-O2a haplogroup also continued to the east, where they came into contact with the Hmong-Mien whose ancestors had preceded them, and to the southeast, where their numbers expanded and they went on to colonise much of insular Southeast Asia. The significant incidence of the M95-O2a paternal haplogroup in Hmong-Mien populations and the M7-O3d (O3a3b) haplogroup in the paternal ancestry of eastern Austroasiatic populations of Southeast Asia suggests an intimate contact relationship between ancient Austroasiatics and ancient Hmong-Mien.

In his posthumously published theory of an East Asian superphylum comprising distant linguistic relationships, Starosta proposed a Proto-Yangtzean supergroup consisting of Austroasiatic and Hmong-Mien (van Driem 2008). Could Proto-Yangtzean correspond to a real linguistic taxon? If we assume the father tongue hypothesis for the spread of language, then the phylogeny of East Asian Y chromosomal haplogroups gives no reason to assume a closer relationship between these two language families. Instead a distant relationship between Tibeto-Burman and Hmong-Mien might be mooted to have existed at a hoary time depth beyond the linguistically reconstructible past.

Between the ancient Austroasiatics and the ancient Hmong-Mien, on the other hand, both human genes and knowledge about rice domestication were shared in the eastern half of a long and complex interaction sphere which extended from the Ganges in the west to the Yangtze and Mekong in the east. In fact, this more complex view is at least partially corroborated by our own findings in the realm of human genetic phylogeography, relating to how sex-specific admixture in Austroasiatic populations of the Indian subcontinent may have been correlated with a spread from east to west (Chaubey et al. 2010).

When we turn to the Kra-dai groups that now also inhabit this region, linguistically the Kra-dai languages of mainland Southeast Asia and southern China appear to be an old offshoot of Austronesian, dating from just after the Formosan exodus (Ostapirat 2005; Sagart 2004). In view of the continuing refinements to the 2002 and 2005 models of the Y

chromosomal haplogroup tree (Karafet et al. 2008), the currently emerging Y chromosomal picture based on single nucleotide polymorphisms suggests that Kradai peoples could descend mainly from ancient Hmong-Mien and Austroasiatic language communities which were linguistically assimilated by ancient Austronesian remigrants to the East Asian mainland. The Y chromosomal haplogroup M119-O1 (O1a) occurs at a high frequency amongst the Austronesian aboriginal peoples of Formosa and also, albeit in a much lower frequency, in the Philippines and southeastern China, especially in Kradai language communities (Abdulla et al. 2009). The paternal genetic imprint of the ancient Austronesians is but faint in comparison to their linguistic and cultural impact, which has extended from Formosa across half the planet.<sup>11</sup>

The slight disconnect between the immediate linguistic affinity of Kradai peoples and their predominant Y chromosomal signature indicates that the Kradai languages probably spread across populations which by and large retained their original genetic make-up. The ancient mtDNA recovered in northeastern Thailand from the Bronze Age site Noen U-loke, dating from 1500 BC to 500 AD, and from the Iron Age site Ban Lum-Khao, dating from 1200 to 400 BC, most closely resembles the mitochondrial lineages of a local Austroasiatic group, i.e. the Chaobon who speak the Monic language Nyahkur (Lertrit et al. 2008).<sup>12</sup> In a similar vein, the numerous insular Southeast Asian Austronesian language communities, where the M95-O2a haplogroup is also prevalent, may represent ancient Austroasiatics who were linguistically assimilated after the Formosan exodus.

Both linguists and ethnographers have for various reasons presumed that Austroasiatic is a particularly old language family. The prehistoric range of the language phylum could have been vast. If we assume the father tongue hypothesis and a correlation of the Y chromosomal haplogroup M95-O2a with this phylum, then the ancient range of Austroasiatic could have stretched from the Godāvārī to the Halmahera Sea. The theory of an Austroasiatic substrate in insular Southeast Asia dates back at least to Skeat and Blagden (1906). The view was explicitly or implicitly entertained in much of the anthropological literature of the period. However, actual linguistic evidence has seldom been adduced, with the notable exception of a study by Adelaar (1995). The culture of Enggano off the west coast of Sumatra has recently been studied in the light of an Austroasiatic substrate theory (Blench 2009). The strongest evidence for this old theory, however, appears to be prevalence of the Y chromosomal signature haplogroup that could have been associated with the ancient Austroasiatics.

Assuming the father tongue hypothesis, it appears that we can at least paint the linguistic map of much of eastern Asia with a broad brush, based on our understanding of the phylogeny of Y chromosomal haplogroups. Detailed, careful correlations of linguistic and population genetic findings based on a far more fine-meshed and a far more ethnolinguistically well-informed population genetic sampling throughout Asia, in light of the higher-resolution Y haplogroup phylogeny currently available, will hopefully enable us

---

11 A number of the roughly one hundred alleged correspondences adduced as evidence for a genetic relationship between Austronesian and the southernmost Andamanese languages Ōnge and Jarawa (Blevins 2007) are not compelling. Yet it cannot be excluded that some of the forms might reflect the residue of contact influence.

12 Apparently the Thai research group observed numerous caveats which have appeared in the literature regarding the processing and treatment of ancient mtDNA. For example, Yao et al. (2003) stress that unambiguous conclusions regarding mitochondrial prehistory must be based not just on hyper-variable segment 1 (HVS 1) sequences which are held to be indicative for geographical or ethnolinguistic affinity, but also on coding region polymorphisms diagnostic for mtDNA haplogroup membership.

to reconstruct early language contact situations and ancient cases of language shift and linguistic intrusions with far greater detail. In a careful interdisciplinary approach lies the promise of sociolinguistic reconstructions of episodes of our past which could account for the phenotypical differences readily observable and long noted by ethnographers between Munda speakers and Khasi-Aslian language communities or between Aslian negrito populations, Aslian non-negrito populations and the Nicobarese. An understanding of prehistory from the different vantage points of linguistics, archaeology and palaeobotany, and cultural and physical anthropology will provide a richer picture of our shared past.

## Acknowledgements

This chapter was first presented at the 4th International Conference on Austroasiatic Linguistics at Mahidol University on 29 October 2009. I should like to thank Martha Ratliff for sharing an electronic copy of the index of her then forthcoming handbook on Hmong-Mien historical phonology (2010). Figure 4 was prepared by Peter Johan Nijland.

## References

- Abdulla, Mahmood Ameen, Ikhlaq Ahmed, Anunchai Assawamakin, Jong Bhak, Samir K. Brahmachari, Gayvelline C. Calacal, Amit Chaurasia, Chien-Hsiun Chen, Jieming Chen, Yuan-Tsong Chen, Jiayou Chu, Eva Maria C. Cutiongco-de la Paz, Maria Corazon A. de Ungria, Frederick C. Delfin, Juli Edo, Suthat Fuchareon, Ho Ghang, Takashi Gojobori, Junsong Han, Sheng-Feng Ho, Boon Peng Hoh, Wei Huang, Hidetoshi Inoko, Pankaj Jha, Timothy A. Jinam, Li Jin, Jongsun Jung, Daoroon Kangwanpong, Jatupol Kampuansai, Giulia C. Kennedy, Preeti Khurana, Hyung-Lae Kim, Kwangjoong Kim, Sangsoo Kim, Woo-Yeon Kim, Kuchan Kimm, Ryosuke Kimura, Tomohiro Koike, Supasak Kulawonganuchai, Vikrant Kumar, Poh San Lai, Jong-Young Lee, Sunghoon Lee, Edison T. Liu, Partha P. Majumder, Kiran Kumar Mandapati, Sangkot Marzuki, Wayne Mitchell, Mitali Mukerji, Kenji Naritomi, Chumpol Ngamphiw, Norio Niikawa, Nao Nishida, Bermseok Oh, Sangho Oh, Jun Ohashi, Akira Oka, Rick Ong, Carmencita D. Padilla, Prasit Palittapongarnpim, Henry B. Perdigon, Maude Elvira Phipps, Eileen Png et al. (30 co-authors). 2009. Mapping human genetic diversity in Asia. *Science* 326. 1541–1545.
- Adelaar, Karl Alexander. 1995. Borneo as a crossroads for comparative Austronesian linguistics. In Peter Bellwood, James J. Fox & Darrell Tryon (eds) *The Austronesians: Historical and Comparative Perspectives*, 75–95. Canberra: Australian National University.
- Agrawal, D. P. 2002. The earliest pottery and agriculture in South Asia. In Yoshinori Yasuda (ed.) *The Origins of Pottery and Agriculture*, 81–88. New Delhi: Lusre Press and Roli Books, for the International Center for Japanese Studies.
- Ahn, Sung-Mo. 2010. The emergence of rice agriculture in Korea: archaeobotanical perspectives. *Archaeological and Anthropological Sciences* 2(2). 89–98.
- Baig, M. M., A. A. Khan & K. M. Kulkarni. 2004. Mitochondrial DNA diversity in tribal and caste groups of Maharashtra (India) and its implication on their genetic origins. *Annals of Human Genetics* 68. 453–460.
- Ballinger, S. W., T. G. Schurr, Antonio Torroni, Y. Y. Gan, J.A. Hodge, K. Hassan, K. H. Chen & D. G. Wallace. 1992. Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient Mongoloid migrations. *Genetics* 130. 139–152.
- Blench, Roger. 2006. *Archaeology, Language and the African Past*. Lanham: Alta Mira Press.

- . 2009. *The Enggano: archaic foragers and their interactions with the Austronesian world*. Manuscript 9 September 2009.
- Blevins, Juliette. 2007. A long lost sister of Proto-Austronesian? Proto-Ongan, mother of Jarawa and Onge of the Andaman Islands. *Oceanic Linguistics* 46(1). 154–198.
- de Candolle, Alphonse-Louis-Pierre Pyrame. 1883. *Origine des plantes cultivées*. Paris: Librairie Germer Ballière et Compagnie.
- Cavalli-Sforza, Luigi Luca, Paolo Menozzi & Alberto Piazza. 1994. *The History and Geography of Human Genes*. Princeton, New Jersey: Princeton University Press.
- Chakravarti, Aravinda. 2009. Tracing India's invisible threads. *Nature* 461. 487–488.
- Chaubey, Gyaneshwer, Mait Metspalu, Ying Choi, Reedik Mägi, Irene Gallego Romero, Siiri Rootsi, Pedro Soares, Mannis van Oven, Doron M. Behar, Siiri Rootsi, Georgi Hudjashov, Chandana Basu Mallick, Monika Karmin, Mari Nelis, Jüri Parik, Alla Goverdhana Reddy, Ene Metspalu, George van Driem, Yali Xue, Chris Tyler-Smith, Kumarasamy Thangaraj, Lalji Singh, Mait Metspalu, Martin B. Richards, Marta Mirazon Lahr, Manfred Kayser, Richard Villems and Toomas Kivisild. 2010. Population genetic structure in Indian Austroasiatic speakers: The role of landscape barriers and sex-specific admixture. *Molecular Biology and Evolution*, doi: 10.1093/molbev/msq28.
- Chen, Jiongjiong, Jihua Ding, Yidan Ouyang, Hongyi Du, Jiangyi Yang, Ke Cheng, Jie Zhao, Shuqing Qiu, Xuelian Zhang, Jialing Yao, Kede Liu, Lei Wang, Caiguo Xu, Xianghua Li, Yongbiao Xue, Mian Xia, Qing Ji, Jufei Lu, Mingliang Xu & Qifa Zhang. 2008. A triallelic system of *S5* is a major regulator of the reproductive barrier and compatibility of *indica-japonica* hybrids in rice. *Proceedings of the National Academy of Sciences of the United States of America* 105(32). 11436–11441.
- Chén Xīngcàn. 1999. On the earliest evidence of rice cultivation in China. In Peter Bellwood & Ian Lilley (eds) *Bulletin of the Indo-Pacific Prehistory Association Bulletin, 18 (Indo-Pacific Prehistory: The Melaka Papers, Vol. 2)*, 81–94. Canberra: Australian National University.
- Chǔ Jiāyòu, 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 & 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.
- Cordaux, Richard, Nilmani Saha, Gillian R. Bentley, Robert Aunger, S.M. Sirajuddin & Mark Stoneking. 2003. Mitochondrial DNA analysis reveals diverse histories of tribal populations from India. *European Journal of Human Genetics* 11. 253–264.
- Cordaux, Richard, Gunter Weiss, Nilmani Saha & Mark Stoneking. 2004. The northeast Indian passageway: A barrier or corridor for human migrations? *Molecular Biology and Evolution* 21. 1525–1533.
- Crawford, Gary W. & Chen Shen 1998. The origins of rice agriculture: Recent progress in East Asia. *Antiquity* 72. 858–866.
- Cushman, Richard D. 1970. *Rebel Haunts and Lotus Huts: Problems in the Ethnohistory of the Yao*. Ithaca, New York: Cornell University doctoral dissertation.
- 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.
- . 2005. The contribution of linguistic palaeontology to the homeland of Austroasiatic. In Laurent Sagart, Roger Blench & Alicia Sanchez-Mazas (eds) *The Peopling of East Asia:*

- Putting Together the Archaeology, Linguistics and Genetics*, 77–80. London: Routledge Curzon.
- 2009. *More on Dvaravati Old Mon*. Paper presented at the Fourth International Conference on Austroasiatic Linguistics, Mahidol University at Salaya, 29 October 2009.
- 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 & 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.
- Donegan, Patricia J. & David Stampe. 1983. Rhythm and holistic organization of language structure. In John F. Richardson, Mitchel Marks & Amy Chukerman (eds) *Papers from the Parasession of the Interplay of Phonology, Morphology and Syntax*, 337–363/ Chicago: Chicago Linguistics Society.
- 2004. Rhythm and the synthetic drift of Munda. In Rajendra Singh (ed.) *The Yearbook of South Asian Languages and Linguistics 2004*, 3–36. Berlin: Mouton de Gruyter.
- Driem, George van. 2001. *Languages of The Himalayas: An Ethnolinguistic Handbook of the Greater Himalayan Region containing an Introduction to the Symbiotic Theory of Language*, vol. 1, Brill, Leiden.
- 2005. Tibeto-Burman vs. Indo-Chinese: Implications for population geneticists, archaeologists and prehistorians. In Laurent Sagart, Roger Blench & Alicia Sanchez-Mazas (eds) *The Peopling of East Asia: Putting Together the Archaeology, Linguistics and Genetics*, 81–106. London: Routledge Curzon.
- 2006. The diversity of the Tibeto-Burman language family and the linguistic ancestry of Chinese. *Bulletin of Chinese Linguistics* 1(2). 211–270.
- 2007. Austroasiatic phylogeny and the Austroasiatic homeland in light of recent population genetic studies. *Mon-Khmer Studies* 37. 1–14.
- 2008. To which language family does Chinese belong, or what's in a name. In Alicia Sanchez-Mazas, Roger Blench, Malcolm D. Ross, Ilia Peiros & Marie Lin (eds) *Past Human Migrations in East Asia: Matching Archaeology, Linguistics and Genetics*, 219–253. London and New York: Routledge.
- Duan Shihua, Baorong Lu, Zhong Li, Jiping Tong, Jin Kong, Wen Yao, Shaoqing Li & Yingguo Zhu. 2007. Phylogenetic analysis of AA-genome *Oryza* species (Poaceae) based on chloroplast, mitochondrial and nuclear DNA sequences. *Biochemical Genetics* 45(1/2). 113–129.
- Endicott, Phillip, Mait Metspalu & Toomas Kivisild. 2007. Genetic evidence on modern human dispersals in South Asia: Y chromosome and mitochondrial DNA perspectives: The world through the eyes of two haploid genomes. In Michael D. Petraglia & Bridget Allchin (eds) *The Evolution and History of Human Populations in South Asia: Interdisciplinary Studies in Archaeology, Biological Anthropology, Linguistics and Genetics*, 341–244. Dordrecht: Springer.
- Fuller, Dorian Q. 2005a. Ceramics, seeds and culinary change in prehistoric India. *Antiquity* 79. 761–777.
- 2005b. Formation processes and palaeoethnobotanical interpretation in South Asia. *Journal of Interdisciplinary Studies* 2(1). 93–115.
- 2005c. The Ganges on the world Neolithic map: The significance of recent research on agricultural origins in northern India. *Prāgdhārā* 16. 187–206.

- 2006a. Agricultural origins and frontiers in South Asia: A working synthesis. *Journal of World Prehistory* 20. 1–86.
- 2006b. Dung mounds and domesticators: Early cultivation and pastoralism in Karnataka. In Catherine Jarrige & Vincent Lefèvre (eds) *South Asian Archaeology 2001: Proceedings of the Sixteenth International Conference on South Asian Archaeology, European Association of South Asian Archaeologists, Paris, 2-6 July 2001*, 117–127. Paris: Éditions Recherche sur les Civilisations.
- 2006c. Silence before sedentism and the advent of cash crops: A status report on early agriculture in South Asia from plant domestication to the development of political economies (with an excursus on the problem of semantic shift among millets and rice). In Osada Toshiki (ed.) *Proceedings of the pre-Symposium of the Research Institute for Humanity and Nature and 7th Ethnogenesis of South and Central Asia Round Table Harvard-Kyoto Roundtable*, 175–213. Kyoto: Research Institute for Humanity and Nature.
- 2007a. Contrasting patterns in crop domestication and domestication rates: Recent archaeobotanical insights from the Old World. *Annals of Botany*, 1–22.
- 2007b. Non-human genetics, agricultural origins and historical linguistics in South Asia. In Michael D. Petraglia & Bridget Allchin (eds) *The Evolution and History of Human Populations in South Asia: Interdisciplinary Studies in Archaeology, Biological Anthropology, Linguistics and Genetics*, 393–443. Dordrecht: Springer.
- Fuller, Dorian Q, Nicole Boivin & Ravi Korisettar. 2007. Dating the Neolithic of South India: New radiometric evidence for key economic, social and ritual transformations. *Antiquity* 81. 755–778.
- Fuller, Dorian Q., Emma Harvey & Ling Qin. 2007. Presumed domestication? Evidence for wild rice cultivation and domestication in the fifth millennium BC of the Lower Yangtze region. *Antiquity* 81. 316–331.
- Fuller, Dorian Q. 2008. Neolithic Cultures. In Deborah M. Pearsall (ed.) *Encyclopaedia of Archaeology, Vol. 1*, 756–768. New York: Academic Press.
- Fuller, Dorian Q. & Ling Qin. 2009. Water management and labour in the origins and dispersal of Asian rice. *World Archaeology* 41(1). 88–111.
- Fuller, Dorian Q, Ling Qin, Yunfei Zheng, Zhijun Zhao, Xugao Chen, Leo Aoi Hosoya & Guo-Ping Sun. 2009. The domestication process and domestication rate in rice: Spikelet bases from the Lower Yangtze. *Science* 323. 1607–1610.
- Fuller, Dorian Q, Yo-ichiro Sato, Cristina Castillo, Ling Qin, Alison R. Weisskopf, Eleanor J. Kingwell-Banham, Jixiang Song, Sung-Mo Ahn & Jacob van Etten. 2010. Consilience of genetics and archaeobotany in the entangled history of rice. *Archaeological and Anthropological Sciences* 2(2). 115–131.
- Garris, Amanda J., Thomas H. Tai, Jason Coburn, Steve Kresovich & Susan McCouch. 2005. Genetic structure and diversity in *Oryza sativa* L. *Genetics* 169. 1631–1638.
- Harvey, Emma L., Dorian Q. Fuller, R.K. Mohanty & Basanta Mohanta. 2006. Early agriculture in Orissa: Some archaeobotanical results and field observations on the Neolithic. *Man and Environment*, XXXI(2). 21–32.
- Haudricourt, Andre-Georges. 1954. Introduction à la phonologie historique des langues miao-yao. *Bulletin de l'École Française d'Extrême-Orient* 44. 555–574.
- Haudricourt, André-Georges & Louis Hédin. 1987. *L'homme et les plantes cultivées*. Paris: Éditions A.-M. Métailié.

- Hazarika, Manjil. 2005. *Neolithic culture of Northeast India with Special Reference to the Origins of Agriculture and Pottery*. Pune (Poona): Unpublished Master's thesis, Deccan College.
- . 2006a. Neolithic culture of northeast India: A recent perspective on the origins of pottery and agriculture. *Ancient Asia* 1. 25–43.
- . 2006b. Understanding the process of plant and animal domestication in northeast India: A hypothetical approach. *Asian Agri-History* 10(3). 203–212.
- Higham, Charles 2002. *Early Cultures of mainland Southeast Asia*. Bangkok: River Books.
- . 2004. Mainland Southeast Asia from the Neolithic to the Iron age. In Ian Glover & Peter Bellwood (eds) *Southeast Asia: From Prehistory to History*, 41–67. Abingdon: Routledge Curzon.
- Higham, Charles & Tracey Lu. 1998. The origins and dispersal of rice cultivation. *Antiquity* 72. 867–877.
- Hill, Catherine, Pedro Soares, Maru Mormina, Vincent Macaulay, William Meehan, James Blackburn, Douglas Clarke, Joseph Maripa Raja, Patimah Ismail, David Bulbeck, Stephen Oppenheimer & Martin Richards. 2006. Phylogeography and Ethnogenesis of Aboriginal Southeast Asians. *Molecular Biology and Evolution*, 23(12). 2480–2491.
- Hill, Catherine, Pedro Soares, Maru Mormina, Vincent Macauley, Dougie Clarke, Petya B. Blumbach, Matthieu Vizuete-Forster, Peter Forster, David Bulbeck, Stephen Oppenheimer & Martin Richards. 2007. A mitochondrial stratigraphy for Island Southeast Asia. *American Journal of Human Genetics* 80. 29–43.
- Igawa, Kazunari, Yoshitaka Manabe, Joichi Oyamada, Yoshikazu Kitagawa, Katsutomo Kato, Kazuya Ikematsu, Ichiro Nakasono, Takayuki Matsushita & Atsushi Rokutanda. 2009. Mitochondrial DNA analysis of Yayoi period human skeletal remains from the Doigahama site. *Journal of Human Genetics* 54. 581–588.
- Karafet, Tatiana M., Fernando L. Mendez, Monica B. Meilerman, Peter A. Underhill, Stephen L. Zegura & Michael F. Hammer. 2008. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Research* 18. 830–838.
- Kivisild, Toomas, Katrin Kaldma, Mait Metspalu, Jüri Parik, Surinder S. Papiha & 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. In Surinder S. Papiha, R. Deka & R. Chakraborty (eds) *Genomic Diversity: Applications in Human Population Genetics*, 135–152. 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 & Richard Villems. 1999b. Deep common ancestry of Indian and Western-Eurasian mitochondrial DNA lineages. *Current Biology* 9(22). 1331–1334.
- Kivisild, Toomas, Helle-Viivi Tolk, Jüri Parik, Yiming Wang, Surinder S. Papiha, Hans-Jürgen Bandelt & Richard Villems. 2002. The emerging limbs and twigs of the East Asian mtDNA tree. *Molecular Biology and Evolution* 19(10). 1737–1751.
- 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 & 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.



- Kong Qing-Peng, Yong-Gang Yao, Chang Sun, Hans-Jürgen Bandelt, Chun-Ling Zhu & Ya-Ping Zhang. 2003. Phylogeny of East Asian mitochondrial DNA lineages inferred from complete sequences. *American Journal of Human Genetics* 73. 671–676.
- Kovach, Michael J., Megan T. Sweeney & Susan McCouch. 2007. New insights into the history of rice domestication. *Trends in Genetics* 23(11). 578–587.
- Kumar, Vikrant, Arimanda N. S. Reddy, Jagedeesh P. Babu, Tiprisetti N. Rao, Banrida T. Langstieh, Kumarasamy Thangaraj, Alla G. Reddy, Lalji Singh & Battini M. Reddy. 2007. Y-chromosome evidence suggests a common paternal heritage of Austro-Asiatic populations. *BioMed Central Evolutionary Biology* 7: 47. 14 pp. <http://www.biomedcentral.com/1471-2148/7/47>.
- Lertrit, Patcharee, Samerchai Poolsuwan, Rachanie Thosarat, Thitima Sanpachudayan, Hathaichanoke Boonyarit, Chatchai Chinpaisal & Bhoom Suktitipat. 2008. Genetic history of Southeast Asian populations as revealed by ancient and modern human mitochondrial DNA analysis. *American Journal of Physical Anthropology* 137(4). 425–440.
- Li Changbao, Ailing Zhou & Tao Sang. 2006a. Genetic analysis of rice domestication syndrome with wild annual species *Oryza nivara*. *New Phytologist* 170. 185–194.
- . 2006b. Rice domestication by reducing shattering. *Science* 311. 1936–1939.
- Li Wanglong & Bikram S. Gill. 2006. Multiple genetic pathways for seed shattering in the grasses. *Functional Integrated Genomics* 6. 300–309.
- Lì Jīn, Peter A. Underhill, Vishal Doctor, Ronald W. Davis, Peidong Shen, L. Luca Cavalli-Sforza & Peter J. Oefner. 1999. Distribution of haplotypes from a chromosome 21 region distinguishes multiple prehistoric human migrations. *Proceedings of the National Academy of Sciences of the United States of America* 96. 3796–3800.
- Lin Zhongwei, Megan E. Griffith, Xianran Li, Zuofeng Zhu, Lubing Tan, Yongcai Fu, Wenxu Zhang, Xiangkun Wang, Dauxin Xie & Chuanqing Sun. 2007. Origin of seed shattering in rice (*Oryza sativa* L.). *Planta* 266. 11–20.
- Londo, Jason P., Yu-chung Chiang, Kuo-Hsiang Hung, Tzen-Yuh Chiang & Barbara A. Schaal. 2006. Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, *Oryza sativa*. *Proceedings of the National Academy of Sciences of the United States of America* 103(25). 9578–9583.
- Ma Jianxin & Jeffrey L. Bennetzen. 2004. Rapid recent growth and divergence of rice nuclear genomes. *Proceedings of the National Academy of Sciences of the United States of America* 101(34). 12404–12410.
- Macauley, Vincent, Catherine Hill, Alessandro Achilli, Chiara Rengo, Douglas Clarke, William Meehan, James Blackburn, Ornella Semino, Rosario Scozzari, Fulvio Cruciani, Adi Taham, Norazila Kassim Shaari, Joseph Maripa Raja, Patimah Ismail, Zafarina Zainuddin, William Goodwin, David Bulbeck, Hans-Jürgen Bandelt, Stephen Oppenheimer, Antonio Torroni & Martin Richards. 2005. Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science* 308. 1034–1036.
- 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 & 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* 5: 26. 25 pp. doi:10.1186/1471-2156-5-26.

- Nakamura Shin-ichi. 2010. The origin of rice cultivation in the Lower Yangtze region. *China. Archaeological and Anthropological Sciences* 2(2). 107–113.
- Nasu Hirō, Arata Momohara, Yoshinori Yasuda & Jiejun He. 2006. The occurrence and identification of *Setaria italica* (L.) P. Beauv. (foxtail millet) grains from the Chengtoushan site (ca. 5800 cal B.P.) in central China, with reference to the domestication centre in Asia. *Vegetation History and Archaeobotany* 16(6). 481–494.
- Niederer, Barbara. 1998. *Les langues hmong-mjen (miao-yao): Phonologie historique*. München: Lincom Europa.
- Onishi Kazumitsu, Kyoko Takagi, Manabu Kontani, Takumi Tanaka & Yoshio Sano. 2007. Different patterns of genealogical relationship found in the two major QTLs [i.e. quantitative trait loci] causing reduction of seed shattering during rice domestication. *Genome* 50. 757–766.
- Osada Toshiki. 1995. *Mundajin no Nōkō Bunka to Shokuji Bunka: Minzoku Gengogakuteki Kōsatsu*. [‘The Rice and Food Culture of Munda in Eastern India: An Ethnolinguistic Study’]. Kyōto: Kokusai Nihon Bunka Kenkyū Sentā.
- Ostapirat, Weera 2005. Kra-Dai and Austronesian: Notes on phonological correspondences and vocabulary distribution. In Laurent Sagart, Roger Blench & Alicia Sanchez-Mazas (eds) *The Peopling of East Asia: Putting Together Archaeology, Linguistics and Genetics*, 107–131. London: Routledge Curzon.
- Pal, J.N. 1990. The early farming culture of northern India. *Bulletin of Deccan College Post-Graduate and Research Institute* 49. 297–304.
- Parsons, B. J., H. J. Newbury & M. T. Jackson. 1999. The genetic structure and conservation of aus, aman and boro rices from Bangladesh. *Genetic Resources and Crop Evolution* 46(6). 587–598.
- Pictet, Adolphe. 1859, 1863. *Les origines indo-européennes ou les aryas primitifs: Essai de paléontologie linguistique (première partie, seconde partie)*. Paris: Joël Cherbuliez.
- Pokharia, Anil. 2008. Palaeoethnobotanical record of cultivated crops and associated weeds and wild taxa from Neolithic site, Tokwa, Uttar Pradesh, India. *Current Science* 94(2). 248–255.
- Poloni, Estella Simone, Ornella Semino, Giuseppe Passarino, A.S. Santachiara-Benerecetti, I. Dupanloup, André Langaney & 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 the *American Journal of Human Genetics* 62. 1267).
- Poloni, Estella Simone, Nicolas Ray, Stefan Schneider & André Langaney. 2000. Languages and genes: Modes of transmission observed through the analysis of male-specific and female-specific genes. In Jean-Louis Dessalles & Laleh Ghadakpour (eds) *Proceedings: Evolution of Language, 3rd International Conference 3-6 April 2000*, 185–186. Paris: École Nationale Supérieure des Télécommunications.
- Purnell, Herbert C., Jr. 1970. *Toward a Reconstruction of Proto Miao-Yao*. Ithaca: Cornell University Ph.D. dissertation.
- Purugganan, Michael D. & Dorian Q Fuller. 2009. The nature of selection during plant domestication. *Nature* 457. 843–848.
- Purugganan, Michael D. 2010. The evolution of rice: Molecular vignettes of its origin and spread. *Archaeological and Anthropological Sciences* 2(2). 61–68.

- Qamar, Raheel, Qasim Ayub, Aisha Mohyuddin, Agnar Helgason, Kehkashan Mazhar, Atika Manoor, Tatiana Zerjal, Chris Tyler-Smith & 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 & 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 & Ken McElreavey. 2004. Where West meets East: The complex mtDNA landscape of the Southwest and Central Asian corridor. *American Journal of Human Genetics* 74. 827–845.
- Ratliff, Martha. 2004. Vocabulary of environment and subsistence and in the Hmong-Mien protolanguage. In Nicholas Tapp, Jean Michaud, Christian Culas & Gary Yia Lee (eds) *Hmong/Miao in Asia*, 147–165. Chiang Mai: Silkworm Books.
- . 2010. *Hmong-Mien Language History*. Canberra: Pacific Linguistics.
- Reddy, B. Mohan, B.T. Langstieh, Vikrant Kumar, T. Nagaraja, A. N. S. Reddy, M. Aruna, K. Thangaraj, A.G. Reddy & Lalji Singh. 2007. Austro-Asiatic tribes of Northeast India provide hitherto missing genetic link between South and Southeast Asia. *Public Library of Science One* 2(11). e1141. doi:10.1371/journal.pone.0001141.
- Reich, David, Kumarasamy Thangaraj, Nick Patterson, Alkes L. Price & Lalji Singh. 2009. Reconstructing Indian population prehistory. *Nature* 461. 489–495.
- Revel, Nicole. 1988. *Le riz en Asie du sud-est* (3 vols.). Paris: Éditions de l'École des Hautes Études en Sciences Sociales.
- Roychoudhury, Susanta, Sangita Roy, Analabha Basu, Rajat Banerjee, H. Vishwanathan, M.V. Usha Rani, Samir K. Sil, Mitashree Mitra & Partha P. Majumder. 2001. Genomic structures and population histories of linguistically distinct tribal groups of India. *Human Genetics* 109. 339–350.
- Ruddiman, W. F., Z. Guo, X. Zhou, H. Wu & Y. Yu. 2008. Early rice farming and anomalous methane trends. *Quaternary Science Reviews* 27. 1291–1295.
- Sagart, Laurent. 2004. The higher phylogeny of Austronesian and the position of Tai-Kadai. *Oceanic Linguistics* 43(2). 411–444.
- Sahoo, Sanghamitra, Anamika Singh, G. Himabindu, Jheeman Banerjee, T. Sitalaximi, Sonali Gaikwad, R. Trivedi, Phillip Endicott, Toomas Kivisild, Mait Metspalu, Richard Villems & V. K. Kashyap. 2006. A prehistory of Indian Y chromosomes: Evaluating demic diffusion scenarios. *Proceedings of the National Academy of Sciences of the United States of America* 103(4). 843–848.
- Sang, Tao. 2009. Genes and mutations underlying domestication transitions in grasses. *Plant Physiology* 149. 63–70.
- Schwartz, Marianne & John Vissing. 2002. Paternal inheritance of mitochondrial DNA. *New England Journal of Medicine* 347(8). 576–580.
- Sengupta, Sanghamitra, Lev A. Zhivotovsky, Roy King, S.Q. Mehdi, Christopher A. Edmonds, Cheryl-Emiliane T. Chow, Alice A. Lin, Mitashree Mitra, Samir K. Sil, A. Ramesh, M. V. Usha Rani, Chitra M. Thakur, L. Luca Cavalli-Sforza, Partha P. Majumder & Peter A.

- Underhill. 2006. Polarity and temporality of high-resolution Y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of Central Asian pastoralists. *The American Journal of Human Genetics* 78. 202–221.
- Sharma, G.R., V. D. Misra, D. Mandal, B. B. Misra & J. N. Pal. 1980. *Beginnings of Agriculture: From Hunting and Food Gathering to Domestication of Plants and Animals*. Allahabad: Abhinav Prakashan.
- Sharma, Swarkar, Ekta Rai, Prithviraj Sharma, Mamata Jena, Shweta Singh, Katayoon Darvishi, Audesh K. Bhat, A.J.S. Bhanwer, Pramod Kumar Towari & Rameshwar N.K. Bamezai. 2009. The Indian origin of paternal haplogroup R1a1\* substantiates the autochthonous origin of Brahmins and the caste system. *Journal of Human Genetics* 54. 47–55.
- Shi Hong, Dong Yong-li, Wen Bo, Xiao Chun-Jie, Peter A. Underhill, Shen Peidong, Ranajit Chakraborty, Jīn Lì & 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 & 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 & 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 & Jīn Lì. 2000. Y chromosome haplotypes reveal prehistorical migrations to the Himalayas. *Human Genetics* 107(6). 582–590.
- Sweeney, Megan T. & Susan R. McCouch. 2007. The complex history of the domestication of rice. *Annals of Botany* 100. 951–957.
- Tanaka Katsunori, Takeshi Honda & Ryuji Ishikawa. 2010. Rice archaeological remains and the possibility of DNA archaeology: Examples from Yayoi and Heian periods of northern Japan. *Archaeological and Anthropological Sciences* 2(2). 69–78.
- Tewari, Rakesh, R. K. Srivastava, K. K. Singh, K. S. Saraswat & I. B. Singh. 2003. Preliminary report of the excavation at Lahuradewa, District Sant Kabir Nagar, U.P. 2001–2002: Wider archaeological implications. *Prāgdhāra, Journal of the Uttar Pradesh Archaeological Department* 13. 37–76.
- Tewari, Rakesh, R. K. Srivastava, K. K. Singh, K. S. Saraswat, I. B. Singh, M. S. Chauhan, A. K. Pokharia, A. Saxena, V. Prasad & M. Sharma. 2006. Second Preliminary Report of the excavations at Lahuradewa. District Sant Kabir Nagar, U. P. 2002–2003–2004 & 2005–2006. *Prāgdhāra, Journal of the Uttar Pradesh Archaeological Department* 16. 35–68.
- Tewari, Rakesh, R. K. Srivastava, K. K. Singh, K. S. Saraswat, I. B. Singh & K.K. Singh. 2009. Early farming at Lahuradewa, Ganga Basin. *Prāgdhāra, Journal of the Uttar Pradesh Archaeological Department* 18. 347–374.
- Thangaraj, Kumarasamy, Vempati Sridhar, Toomas Kivisild, Alla G. Reddy, Gyaneshwer Chaubey, Vijay Kumar Singh, Suminder Kaur, Pooja Agarawal, Amit Rai, Jalaj Gupta, Chandana Basu Mallick, Niraj Kumar, Thrimulaisamy P. Velavan, Rajanbabu Suganthan, Divi Udaykumar, Rashmi Kumar, Rachana Mishra, Arif Khan, Chitikineni Annapurna & 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.
- Vaughan, Duncan A. 1994. *The Wild Relatives of Rice: A Genetic Resources Handbook*. Manila: International Rice Research Institute.
- Vavilov, Nikolai Ivanovič. 1926. Studies on the origin of cultivated plants. *Trudy po Prikladnoi Botanike, Genetike i Selekcii (Bulletin of Applied Botany, Genetics and Plant Breeding)* 16. 1–248.
- Vitte, C., T. Ishii, F. Lamy, D. Brar & O. Panaud. 2004. Genomic paleontology provides evidence for two distinct origins of Asian rice (*Oryza sativa* L.). *Molecular Genetic Genomics* 272. 504–511.
- Wáng Fūshì & Máo Zōngwǔ. 1995. *Miáo-Yáo Yǔ Gǔyīn de Gòunǐ* [‘Phonological Reconstruction of Hmong-Mien’]. Peking: Zhōngguó Shèhuì Kēxué Chūbǎnshè.
- Weber, Steve, Heather Lehman, Timothy Barela, Sean Hawks & David Harriman. 2010. Rice or millets: Early farming strategies in prehistoric central Thailand. *Archaeological and Anthropological Sciences* 2(2). 79–88.
- Wells, R. Spencer, Nadira Yuldasheva, Ruslan Ruzibakiev, Peter A. Underhill, Irina Evseeva, Jason Blue-Smith, Jīn Lì, Sù Bīng, Ramasamy Pitchappan, Sadagopal Shanmugalakshmi Karupiah 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 Mirrahimov, Ashir Chariev & 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 & 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.
- Wen Bo, Hui Li, Song Gao, Xianyun Mao, Yang Gao, Feng Li, Feng Zhang, Yungang He, Yongli Dong, Youjun Zhang, Wenju Huang, Jianzhong Jin, Chunjie Xiao, Daru Lu, Ranajit Chakraborty, Sù Bīng, Ranjan Deka & Jīn Lì. 2005. Genetic structure of Hmong-Mien speaking populations in East Asia as revealed by mtDNA lineages. *Molecular Biology and Evolution* 22(3). 725–734.
- Williams, R. Sander. 2002. Another surprise from the mitochondrial genome. *New England Journal of Medicine* 347(8). 609–612.
- Xue Yali, Tatiana Zerjal, Weidong Bao, Suling Zhu, Qunfang Shu, Jiujiu Xu, Ruofu Du, Songbin Fu, Pu Li, Matthew E. Hurles, Huanming Yang & Chris Tyler-Smith. 2006. Male demography in East Asia: A north-south contrast in human population expansion times. *Genetics* 172. 2431–2439.
- Yamane, Hiroko, Tomoko Ito, Harumi Ishikubo, Masaki Fujisawa, Harumi Yamagata, Kozue Kamiya, Yukiya Ito, Masao Hamada, Hiroyuki Kanamori, Hiroshi Ikawa, Yuichi Katayose, Jianzhong Wu, Takuji Sasaki & Takashi Matusmoto. 2009. Molecular and evolutionary analysis of the *Hd6* photoperiod sensitivity gene within genus *Oryza*. *Rice* 2. 56–66.
- Yao Yong-Gang, Long Nie, Henry Harpending, Yun-Xin Fu, Zhi-Gang Yuan & Ya-Ping Zhang. 2002. Genetic relationship of Chinese ethnic populations revealed by mtDNA sequence diversity. *American Journal of Physical Anthropology* 118. 63–76.