

Stratification in the peopling of China: how far does the linguistic evidence match genetics and archaeology?

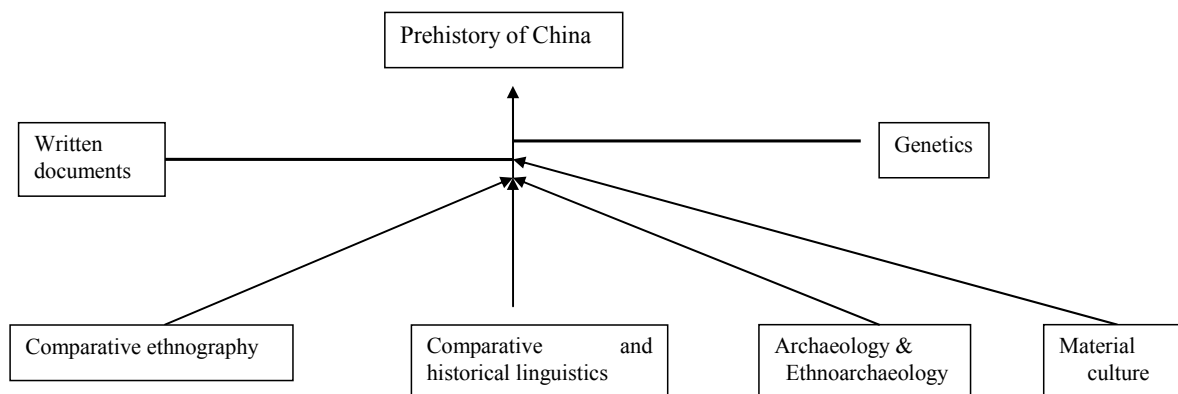
Roger Blench
Mallam Dendo
8, Guest Road
Cambridge CB1 2AL
United Kingdom
Voice/ Fax. 0044-(0)1223-560687
Mobile worldwide (00-44)-(0)7967-696804
E-mail R.Blench@odi.org.uk
<http://www.rogerblench.info/RBOP.htm>

1. Introduction

1.1 The problem: synthesising linguistics, archaeology and genetics

The concept of synthesising linguistics, archaeology and genetics in the reconstruction of the past is becoming a commonplace; but the reality is that each discipline largely pursues its own methods and what little interaction there is remains marginal. Hence many of the questions asked are internal to the discipline, addressed to colleagues, not the larger sphere of understanding the past. China and East Asia in general are a particularly difficult case because so much of the linguistics and archaeology is driven by an emphasis on high culture. Major archaeological texts refer neither to linguistics nor genetics and speculation about the identity of non-Chinese groups mentioned in the texts tends to be unanchored. In addition, ideology surrounding the definition of minorities in China has confused the analysis in genetics papers. This situation has begun to change and a review of the current situation may be useful¹.

Figure 1. Elements in reconstructing China's prehistory



A preliminary outline of an agenda for inter-disciplinary study is set out in Wang (1998) who characterised linguistics, archaeology and genetics as ‘three windows on the past’. This model should be expanded by separating out the potential information in written documents from the results of comparative linguistics based on spoken language. In addition, comparative ethnography has so far only featured in the archaeology window. But, especially in Taiwan, information on the distribution of

¹ Paper presented at the Symposium *Human migrations in continental East Asia and Taiwan: genetic, linguistic and archaeological evidence*. Geneva June 10-13, 2004. Université de Genève. The revision of April 2006 has benefited from two subsequent field trips to China, and I would particularly like to thank my colleagues at the Institute of Zoology in Kunming, especially Qi Xuebin and Su Bing, for extensive discussion on some of these issues. I would also like to acknowledge the input of Laurent Sagart and George van Driem, who have strongly influenced my views on some issues set out here.

material and cultural traits is rich and can potentially be incorporated into larger models. Figure 1 represents a potential multi-disciplinary framework for reconstructing China's past.

2. The linguistic pattern of present-day China

2.1 General

Although dominated numerically by languages of the Sino-Tibetan phylum, China is highly diverse linguistically (Ramsey 1987). Table 1 shows the main language phyla represented;

Table 1. Language Phyla of China

Phylum	Examples
Sino-Tibetan/ Tibeto-Burman	Chinese, Yi, Pumi, Naxi, Bai, Tujia
Hmong-Mien = Miao-Yao	Ho Te, Hmong, Pa Hng
Altaic (Turkic, Mongolic, Tungusic)	Ili Turko, Monguor, Evenki
Daic= Tai-Kadai = Kra-Dai	Zhuang, E
Austroasiatic	Blang, Mang, U
Austronesian (Chamic)	Tsat
Indo-European	Tajik, Wakhi, Russian, Macanese
Korean	
Unclassified	Waxianghua, Wutunhua

Ethnologue 2005 estimates lists 236 languages for China, but this includes 13 dialects of Chinese and Chinese sign language. However, new languages are regularly being recorded, such as the Austroasiatic Bagan, yet to be classified within Mon-Khmer. This figure contrasts with the official count of 56 [55 + Han] which includes all the minorities of Taiwan under a single name, Gaoshan (Yin 1989). Despite the unlikely nature of the official figure, it continues to be propagated on websites and official documents. According to 1990 data, minorities constitute some 91 million or 6.5% of the population, so they are relatively numerous compared with other countries in the region (MacKerras 1997). Despite this, a number of Chinese linguistic sources are available that provide data on the highly diverse languages subsumed under the official minorities, and survey programmes continue to record new languages. Even so, this is probably a fraction of the number of languages that used to exist; the spread of the Han over the last 3000 years has probably eliminated considerably more diversity.

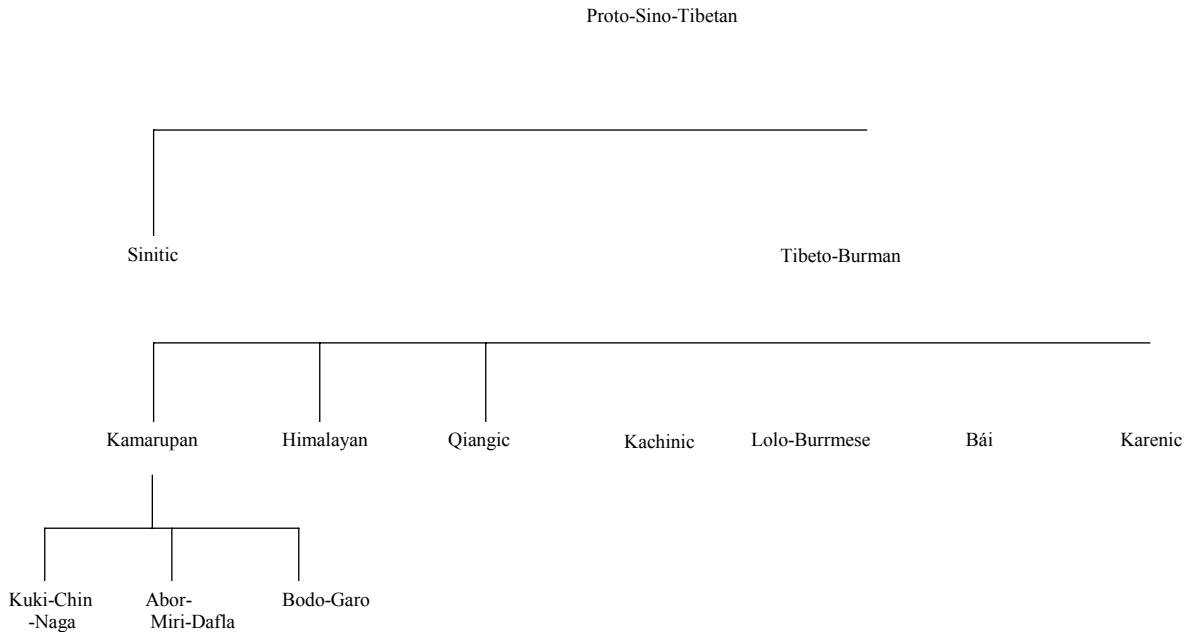
2.2 Sino-Tibetan

Sino-Tibetan is the phylum with the second largest number of speakers after Indo-European, largely because of the size of the Chinese population. Current estimates put their number at ca. 1.3 billion (Ethnologue 2005). Apart from Burmese and Tibetan, most other languages in the phylum are small and remain little-known, partly because of their inaccessibility. The name Sino-Tibetan carries with it heavy historical baggage. Originating as 'Indo-Chinese' in the middle of the nineteenth century, it originally carried racial connotations (Driem 2002). The first recognition of the phylum probably dates to Julius von Klaproth (1823) who recognised three parallel branches, Chinese, Burmese and Tibetan (see Van Driem, this volume). Von Klaproth also explicitly excluded Austroasiatic and Daic, unlike many later classifiers, who sequentially included almost all the phyla in this region, in a series of now discarded phylogenies. Strangely, the notion that the Sinitic branch (i.e. the varieties of Chinese) is not related to the rest of Tibeto-Burman, is still held in some quarters and even recent conspectuses (e.g. Thurgood & LaPolla 2003) feel the obligation to refer to this view.

Considering the importance of Sino-Tibetan and its history of scholarship, there is a striking lack of agreement as to its internal classification. Broadly speaking, the opposing camps are those who

consider Sinitic as a primary subgroup of Sino-Tibetan (Benedict 1972, 1976, Bodman 1980, Matisoff 2003, Bradley 1997, Thurgood and La Polla 2003) and those who would place it at some lower node within the remaining languages, thereby applying the name Tibeto-Burman to the whole phylum (Shafer 1974, Van Driem 1997). Under the second proposal, Sinitic would thus be incorporated within the group conventionally defined in opposition to Sinitic. These two views are reflected in **Figure 2** and **Figure 3**. **Figure 2** shows the internal structure of Sino-Tibetan according to Matisoff (2001:297), which can be taken to represent the mainstream.

Figure 2. Sino-Tibetan according to Matisoff (2001)

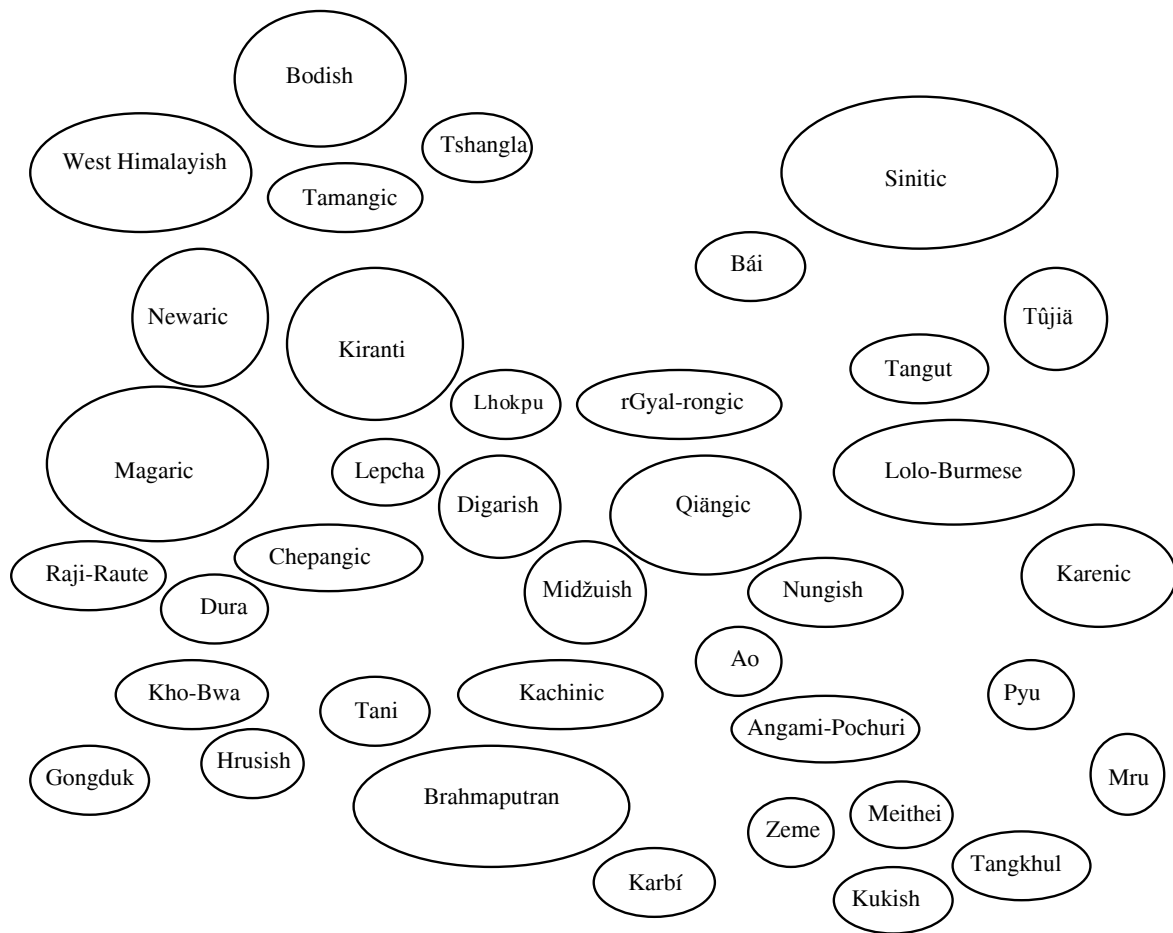


The groups represented here are by and large ‘geographic’ categories; Kamarupan and Himalayan have no genetic status. Even this view has never been justified in print, despite the space afforded by the 800 pages of Matisoff (2003). Moreover, the exclusion by Matisoff of many small branches of Sino-Tibetan and the branching of others from a single node does not suggest this is a fully worked-out theory. The equally agnostic alternative is represented in the view of Van Driem (2003), in his ‘fallen leaves’ schema (**Figure 3**);

Van Driem’s model presents no assumptions at all about subgrouping except to map already well-recognised groups. This is an entirely geographical model, which places generally agreed subgroups in proximity, with area of the ellipse representing their size, but advances no hypothesis about their ultimate relationships. Whether this represents progress is debatable, but the ‘fallen leaves’ model has the virtue of treating all branches of Sino-Tibetan as of equal status and requiring that their position be ultimately defined. Van Driem would argue that this is a fair representation of the current state of our knowledge.

It is hard not to gain the impression that a state of academic warfare exists between the various camps. Languages given prominence by one side are ignored by the other. For example, Gongduk, first drawn attention to in Driem (2001), appears to be a highly divergent Himalayan language that may be of great significance for the reconstruction of the phylum. However, no mention of this language is made in either Matisoff (2003) or Thurgood & LaPolla (2003).

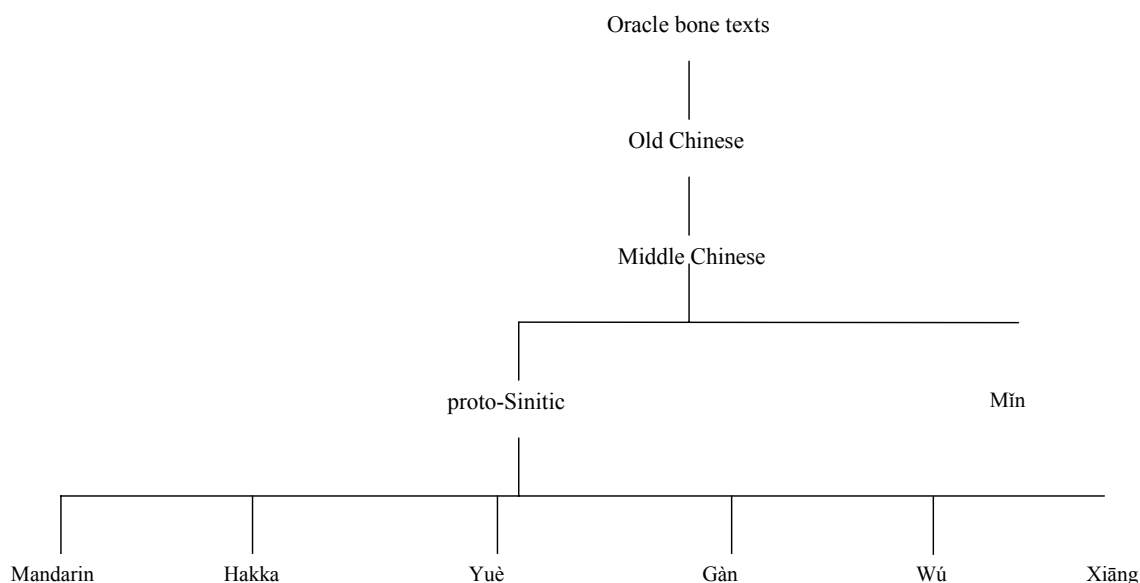
Figure 3. ‘Fallen leaves’ model of Sino-Tibetan according to Driem (2005)



Sinitic

Sinitic is a general term for all the languages deemed to be part of the Chinese subgroup. This label is not uncontroversial, as the Bái language has been analysed as a remote relative of Chinese, or as a distinct Sino-Tibetan language heavily influenced by Chinese (see summary in Norman 2003:73). Sinitic is generally divided into some seven recognised dialect groups, Mandarin, Mǐn, Hakka, Yuè, Gà, Wú and Xiāng, but Norman (2003:72) argues that these are ill-defined and that the unity of groups such as Wú is far from proven. Moreover, the Mǐn dialects are distinct from the other branches of Sinitic, and presumably represent a primary branching. Modern-day Sinitic lects can be reconstructed to a proto-language, referred to as ‘Common Chinese’. However, Chinese is exceptional in that there are decipherable records going back at least to the Shang oracle bones, roughly the 13th century BC (Herforth 2003:59). By the Warring States period (475-221 BC) a corpus of prose texts allows us to undertake a reasonable analysis of phonology and grammar of what is usually called Zhou Chinese (Baxter 1992, Sagart 1999). It appears that the morphology and grammar of this language are strikingly different from Modern Chinese and rather resemble other branches of Sino-Tibetan. Figure 4 shows the historical relation between these groups as presently understood;

Figure 4. Family tree of Sinitic languages



A figure such as this suggests a unidirectional evolution of Chinese, but presumably the early written forms were one lect among many spoken forms of which no record remains.

China also has intriguing ‘remnant’ languages such as Tujia (Braslett & Braslett 2005), hard to classify because they have been so heavily Sinicised. The Tujia² are likely to be the modern descendants of the Ba people, whose kingdom (near modern-day Chongqing) flourished between 600 BC and 400 BC but fell to the Qin in 316 BC. The Ba appear in historical records as the Tujia from about 1300 AD onwards. Written Chinese texts also contain material on other Sino-Tibetan languages that can provide rather fragmentary insights into language diversity in the past (Wang 1998). Bái words are recorded in the *Manshu*, a work of the Tang Dynasty, while the Han dynasty *Bailangge* [=Pai-Lang] is written in a Tibeto-Burman language, probably related to Yi. The Bái language is spoken around Dali in northwestern Yunnan by some 1,250,000 people (Ethnologue 2005). Although officially classified as Tibeto-Burman, evidence for this is problematic because of the complex layers of ancient loans from Chinese and other languages (Wiersma 2003). These two languages are now generally considered to be Sino-Tibetan, although the problem is usually that they have many layers of Sinitic loanwords and hence it is difficult to sort out their core vocabulary. It may be that these are traces of a much more diverse earlier Sino-Tibetan population largely eliminated by the spread of Sinitic. The notion that Bái is co-ordinate with Sinitic would have to be squared with the new understanding of the place of Chinese in the Tibeto-Burman ‘tree’.

A language that is still puzzling is Waxianghua, spoken by 300,000 people (in 1995) in a 6000 km² area in western Hunan Province, Wuling Mountains, including Yuanling, Chunxi, Jishou, Guzhang, and Dayong counties. It differs greatly from both Southwestern Mandarin (Xinan Guanhua) and Xiang Chinese (Hunanese), but is relatively uniform within itself. It has so far remained unclassified. Similarly, the Wutun language with some 2000 speakers in Eastern Qinghai Province, Huangnan Tibetan Autonomous Prefecture mixes elements of Chinese, Tibetan and Mongolian, but its genetic affiliation remains unknown.

External affiliations of Sino-Tibetan

The external affiliations of Sino-Tibetan are also controversial. Sino-Tibetan has been linked with almost every phylum in East Asia (and the New World) and it is hard to make a judgment on this

² A valuable website on Tujia language and culture is <http://www.braslett.org.uk/tujia/ehome.html>

potential for promiscuous cohabitation. Most scholars consider the similarities in lexicon and phonology between Sino-Tibetan and the other phyla with which it is in geographical proximity to be the result of contact. Two macrophylic models have been argued for in recent times; Sino-Caucasian and Sino-Austronesian. Sino-Caucasian has been principally promoted by the late Sergei Starostin (1991) and Sino-Austronesian by Laurent Sagart (1994, 2005). Van Driem (this volume) has reviewed these models and finds the case for Sino-Caucasian flawed by irregular correspondences and wayward semantics. Sino-Austronesian is altogether more promising and most of those who have considered the evidence conclude that Sagart has made a case for links between the two phyla. The issue is whether it can be demonstrated that this is the result of genetic affiliation as opposed to contact.

Archaeological and genetic correlates of Sino-Tibetan expansion

Is it therefore worth trying to make proposals for the pattern of Sino-Tibetan expansion within this mosaic of uncertainty? Probably only generalisations of a very broad kind are useful. The first is that Sino-Tibetan may well be substantially older than is usually thought. The pattern seems to be a number of well-defined groups that have expanded in the last few thousand years and a scatter of archaic languages with unusual features that are very different one from another. This suggests that the source populations were fragmented hunter-gatherer groups spread over a wide area between the Himalayan Plateau and North China, at least 10-12,000 years ago. This period is very poorly known in the archaeology of mainland China but perhaps can be identified with the Shengwen (= 'cord-marked') pottery found between the Yangzi and Yellow rivers. Better known is the Chulmun pottery of the Korean peninsula, which is clearly associated with an alternation between land-mammal hunting and exploitation of marine resources. If this is so, a model that has populations spreading down river valleys, popular among models of phyllic expansion in this region, is inappropriate here; these were probably hunters spreading across open terrain. Once agriculture began, the early adopters gained a massive advantage and some groups spread preferentially, most notably the Sinitic-speakers. The topography allowed the survival of archaic groups in montane areas; hence the pattern of fragmentation of Sino-Tibetan.

A quite different view is canvassed by Matisoff on the STEDT website³. He says;

The Proto-Sino-Tibetan (PST) homeland seems to have been somewhere on the Himalayan plateau, where the great rivers of East and Southeast Asia (including the Yellow, Yangtze, Mekong, Brahmaputra, Salween, and Irrawaddy) have their source. The time of hypothetical ST unity, when the Proto-Han (= Proto-Chinese) and Proto-Tibeto-Burman (PTB) peoples formed a relatively undifferentiated linguistic community, must have been at least as remote as the Proto-Indo-European period, perhaps around 4000 BC. The TB peoples slowly fanned outward along these river valleys, but only in the middle of the first millennium A.D. did they penetrate into peninsular Southeast Asia, where speakers of Austronesian (= Malayo-Polynesian) and Mon-Khmer (Austroasiatic) languages had already established themselves by prehistoric times. The Tai peoples began filtering down from the north at about the same time as the TB's. The most recent arrivals to the area south of China have been the Hmong-Mien (Miao-Yao), most of whom still live in China itself.

This model does not seem to account for the internal diversity of Sino-Tibetan, nor the relative internal diversity of individual branches. If Sinitic and Tibeto-Burman are a primary split, why is Tibeto-Burman so much more internally divided? Six thousand years seems a short period to arrive at the present diversity when compared to Austronesian, which should be of comparable antiquity.

Wherever Sinitic originates within Sino-Tibetan, there is a broad consensus that its main spread has been north-south from the millet-growing to the rice-growing areas and that it has assimilated or overwhelmed a diverse *in situ* population (e.g. Fitzgerald 1972; Lee 1978; LaPolla 2001). It is

³ At <http://stedt.berkeley.edu/>

therefore *unlikely* that Sinitic can be identified with the earliest Neolithic communities in North China such as the Péilǐgǎng or Císhān (6500 BP onwards) and it is more helpful to think of Sinitic as one of Barnes' (1993:108) 'Late Neolithic Elites' emerging between 3500-2000 BC. The notable feature of the end of this period is the appearance of bronze vessels in the archaeological record and it easy to imagine the inception of the Shang as marking the take-off of Sinitic. Presumably, a major element in the *in situ* population was Hmong-Mien-speaking, but unless these groups were considerably north of their present location, the agriculturists of Císhān were not Hmong-Mien either. Van Driem (1998) has canvassed Sichuān as the likely original homeland of Sino-Tibetan (Tibeto-Burman in his terms). A view something like this is supported in a study of Y chromosome haplotypes reported in Bing Su et al. (2000) who argue that proto-Sino-Tibetan was spoken in northern Sichuān and dispersed westwards to the Himalayas and east and south to create the Chinese dialects. However, they also argue that this nucleus was the lineal descendant of early Neolithic millet-growers, which seems highly unlikely. There is no obvious candidate for the ethnolinguistic identity of the millet-growers of Péilǐgǎng and it may be they have no linguistic descendants.

An interesting example of the politicisation of archaeological narratives is the description by Da-Shun (1995) of the Hongshan culture of Liaoning Province, northeast of Beijing. This is usually dated to 4-3000 BC, i.e. roughly contemporaneous with the Yangshao. Despite being well outside the imperial boundaries, Da-Shun sees this as 'the dawn of Chinese civilization' and attempts to link it with that civilization through a series of typological indicators, a writing system, bronze metallurgy etc. A particular type of altar, also found elsewhere in China is part of the thread that links this region with the later Ming dynasty. The reality is that there is no evidence that this region would have been Sinitic-speaking at this period; it is much more credible that the inhabitants would have been Altaic speakers, either speaking pre-Mongolic or Koreanic languages.

2.3 Hmong-Mien

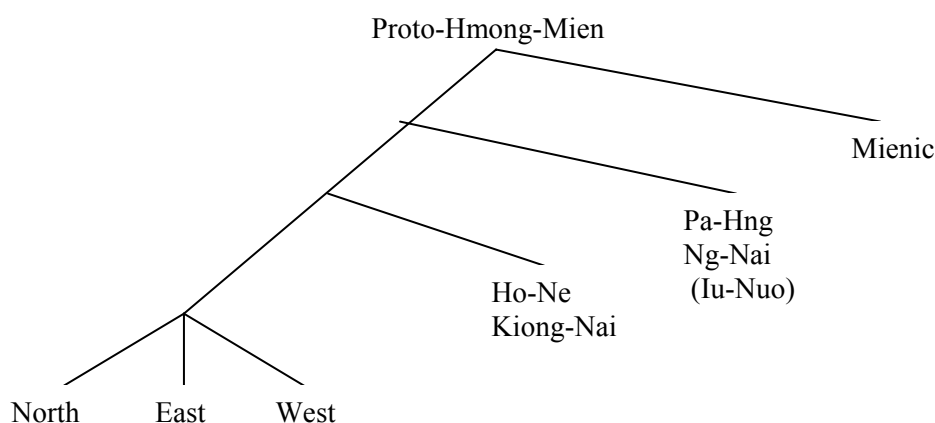
The Hmong-Mien⁴ [=Miao-Yao] languages are spoken mostly in China with some groups also in Laos, Việt Nam and Thailand. Their centre of gravity is between the Yangzi and the Mekong rivers. Hmong-Mien languages are quite close to one another, and although Ethnologue (2005) lists some 35 languages, many of these are mutually intelligible lects. There have been various comparative overviews of the group, starting with Purnell (1970), Wang (1994), Wang & Mao (1995) and Niederer (1998). Tapp et al. (2004) have edited a comprehensive overview of recent scholarship that includes much valuable bibliography.

The internal structure of the family is still very much in flux as new information becomes available. Of particular interest is the Pa-Hng language which has evidence of features of both main branches, although it is apparently closer the Hmongic (Niederer 2004). Still debated is the place Ho Te, or She, a language spoken in Southeastern Guangdong Province by less than 1000 speakers. Figure 5 shows the internal classification of Hmong-Mien according to Niederer (2004:141).

The Hmong-Mien languages have been linked with almost all the East Asian language phyla, but never conclusively. Despite sharing much common vocabulary with their neighbours, evidence for a genetic link is lacking and there is every reason to consider these language constitute a small but independent phylum. All the Hmong-Mien languages are relatively close to one another and the date of their overall dispersal is quite recent. This suggests that one bottleneck brought the phylum into existence, and Mienic at least went through another bottleneck, as it appears to be far more uniform than Hmongic.

⁴ This formulation is also not uncontroversial since it privileges the names of two particular dominant groups influential in the United States (Tapp 2004 fn. 11)

Figure 5. Hmong-Mien languages according to Niederer (2004)



The linguistic geography of Hmong-Mien suggests very strongly that these people were scattered by the incoming Han and probably forced southwards into modern Laos and Thailand, probably in the last 3-4000 years. This has sparked a number of debates on the relative antiquity of these groups; if Hmong-Mien preceded Chinese, should it not be more diverse? Sagart (p.c.) has put forward the suggestion that pre-Mien was adopted by the Chu state (500 BC onwards) which would have had a Sinicised bureaucracy. The northern distribution of Mien probably represents the boundary of this state. Yao, the more southerly languages, must have escaped this state at some stage and were perhaps within another state, as Yao languages have a unique set of Sinitic loans.

Ratliff (2004) has made use of the reconstructions of proto-Hmong-Mien by Wang and Mao (1995) to explore the likely environment of the putative homeland of these people. Two valuable conclusions can be drawn from this; the Hmong-Mien were already established farmers prior to their dispersal and animals and plants reconstructible to proto-Hmong-Mien point to a homeland south of the Yangtze River⁵.

Despite the lack of internal diversity in Hmong-Mien, it seems difficult to imagine that pre-Hmong-Mien are not ancient inhabitants of the East Asian area. It seems as if the other more diverse relatives of Hmong-Mien must have been eliminated by the Han expansion and the languages still in existence are the result of a secondary expansion. The pre-Hmong-Mien may therefore be identified with one of the Neolithic pottery horizons, but it seems unlikely that the present Hmong-Mien diaspora would have any direct correlate, since their dispersal is based on a pattern of refuge rather than positive expansion.

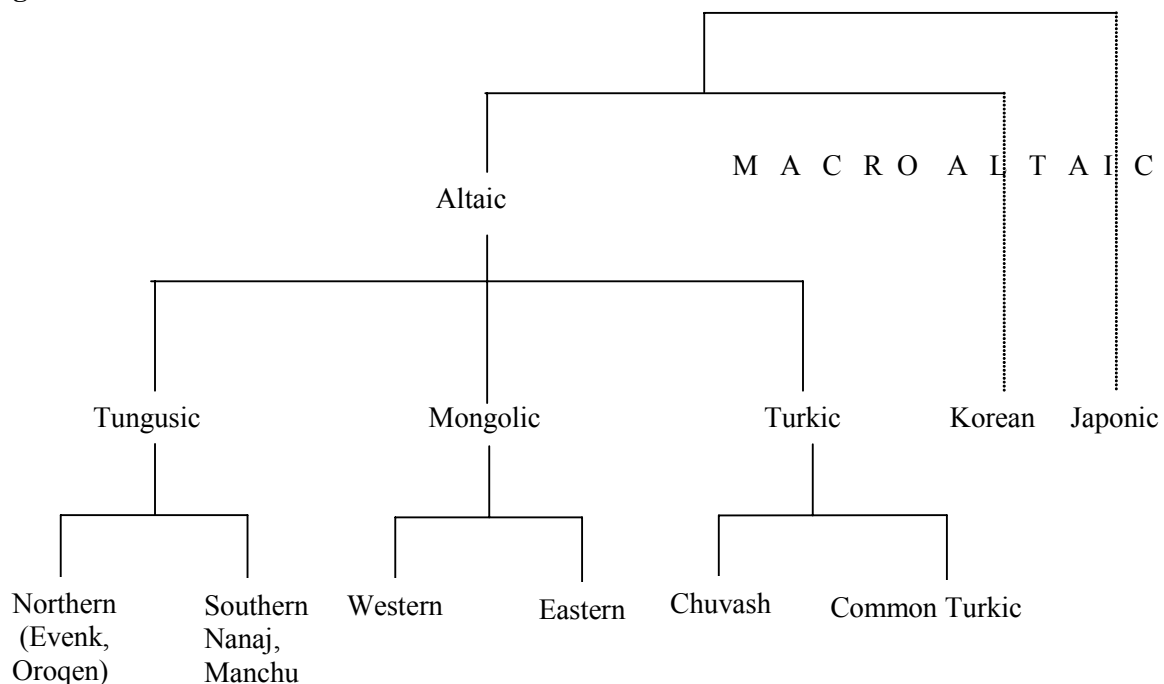
2.4 Altaic

The minimal set of Altaic languages consists of Turkic, Mongolic and Tungusic, spread from Turkey to Siberia, and all attested in China. The macro-phylum, Macro-Altaic, remains controversial. Most scholars accept the membership of Korean, fewer Japanese (see Martin 1991, Miller 1996, this volume). Figure 6 shows a tree representing the Altaic and Macro-Altaic groupings. Starostin et al. (2002) have published a major etymological dictionary of comparative Altaic, which provides rich material for interpreting its prehistory. Compared with the other phyla discussed here, Altaic has a very unusual substructure. Each of its branches are internally very close-knit, but very different from one another. Indeed the dispersal of Turkic has largely taken place in historical time (Golden 1998). The pattern of the phylum points to the likely loss of other branches of Altaic intermediate between

⁵ This might seem evident, given their present location, but Hmong oral traditions have been widely held to point to environments much further north, in Siberia (see Tapp et al. 2004).

those still in existence. The differences between branches have led some scholars to claim that Altaic is not a phylum but a bundling of languages that have interacted (Janhunen 1994).

Figure 6. Altaic and Macro-Altaic



Today the Turkic languages are spread across Central Asia from Sakha (Yakutia) to the Turkish republic, with their centre of gravity in Asian Russia. They are represented in China by Salar, related to Crimean Turkish, and the Uyghur languages and are probably a relatively recent intrusion. The principal sources on the languages and history of this group are Menges (1995) and Johanson and Csató (1998). Generally speaking, the Turkic languages are very closely related and are consistent with a pattern of expansion from the present-day region of modern Mongolia, both westwards to Turkey and north to Sakha.

Mongolic languages are spoken throughout much of modern Mongolia, with outlying Mongolic languages spoken in China and Afghanistan (Janhunen 2003). Much of the region today is dominated by Khalkh Mongol, but the relative uniformity of Mongolic can be attributed to the empire founded by Chinggis Khan (ca. 1200 to 1400 AD) which grew to control the largest land empire ever recorded and probably eliminated earlier ethnic and linguistic diversity. Janhunen (1993) has analysed lexical elements borrowed from Mongolic into Manchurian Tungusic to argue that the family formerly exhibited much greater diversity. Some of the isolated Mongolic languages in China may well be remnants of military expeditions rather than traces of earlier expansions. Kolman *et al.* (1996) sampled the mtDNA of Mongolian populations within Mongolia extensively, and found a high degree of genetic homogeneity, as well as a close link to New World populations. Whether such homogeneity would be reproduced if the sample were extended to Mongolic populations outside Mongolia is unclear, since this may simply reflect the recent dominance of the Khalkh.

The region of Mongolia had a much warmer climate in the early Holocene, and much of the high plateau was heavily forested. As a consequence, subsistence strategies were quite diverse and it is assumed there was agriculture in this period, although this is an inference from Yangshao pottery finds rather than direct evidence (Barnes 1993:154). Reconstructions in Starostin (2002) also point to an important agricultural component in early Altaic although the glottochronological dates they attribute to the phylum make no sense in terms of known archaeology. When the climate became more arid in the third millennium BC, there was a development of nomadic pastoralism. At the same time, rock-engravings show horse-drawn chariots and these are presumably ancestral to the carts

essential to transhumance in Mongolia today. It would not be unreasonable to link this evolution of pastoralism with the expansion of the Mongolic languages. Although today Mongolic is quite undiverse, this is the result of the spread of Khalkh Mongol following the establishment of the Khanates in the medieval period. But there is every reason to think that pastoral peoples, herding horses and other species, have been on the northern borders of China for a long period. Janhunen (1998) has explored the vocabulary of the horse in Central Asia and points out that the terms are all related in almost all the phyla of this region (Table 2);

Table 2. Horse terms in East Asia

Language group	proto-form
Mongolic	morin
Tungusic	murin
Korean	mar
Japanese	uma
Chinese	ma

Source: Janhunen (1998)

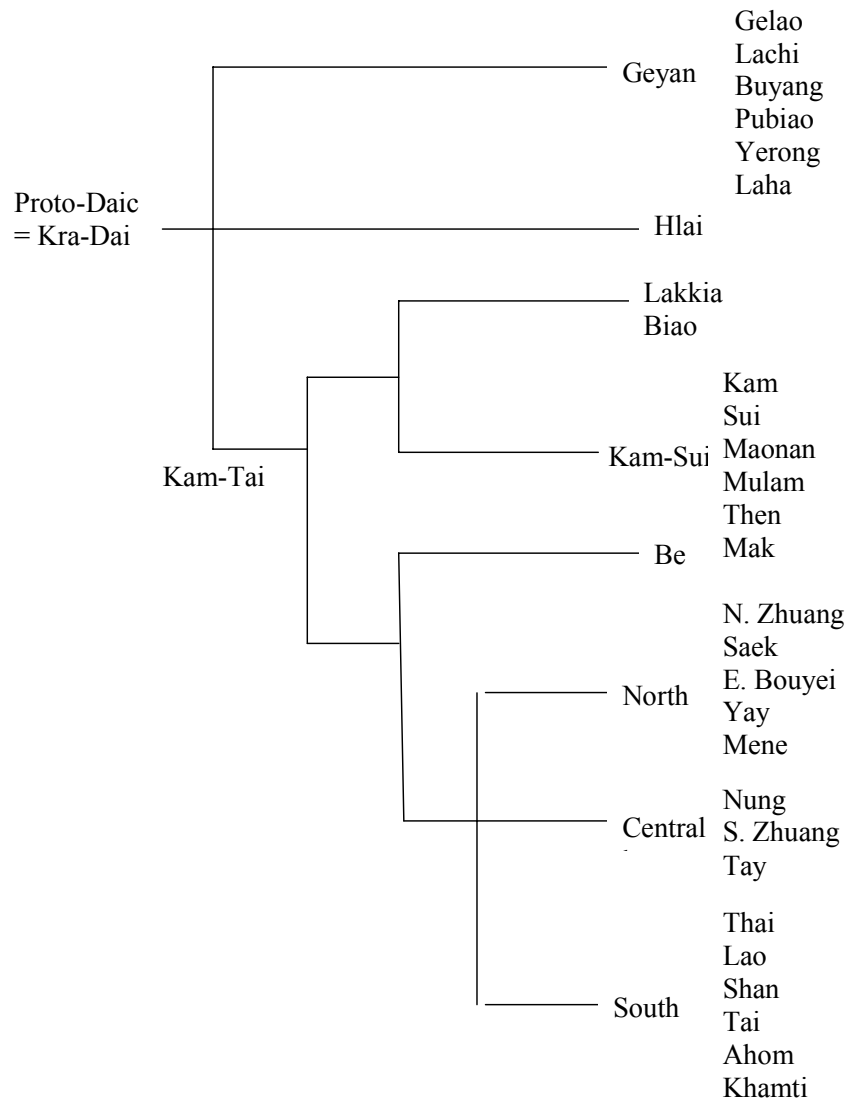
This suggests that horse culture was spread rapidly by a single group; linguistic geography points strongly to Mongolic speakers. As Janhunen (1998) points out, its absence in Turkic suggests that it is not an Altaic root, but a series of ancient loanwords. Norman (1988:18) identifies some loanwords from Altaic into Old Chinese, for example OC **duk*, ‘calf’ and Starostin (this volume) proposes others such as 𪛗 *rəp* &bdq; k+fq`m `r ɛnc&-

Apart from Manchu, the Tungusic languages all have a small number of speakers whose populations were until recently hunter-gatherers (Doerfer 1978). Starostin (2002) points out that speakers may have undergone a reversion to foraging as Tungusic shares names for crops with other Altaic languages. The Tungusic languages are not highly diverse compared with other Siberian populations, suggesting that their expansion is probably quite recent. However, what remains of Tungusic today does not reflect its previous importance. Manchu was the language of the ruling class in China until recently although it has now almost disappeared (Svanberg 1988). Tungusic groups were probably spread more widely across northern Heilongjiang Province and the adjacent Mongol-speaking area, and were in early interaction with Koreanic speakers. Tungusic speakers may have introduced the Northern Bronze Complex into the Korean peninsula during the 1st millennium BC, and also that the Rong people, associated with the Upper Xiajiadian in south-eastern Mongolia, represent a southern intrusion of Tungusic (Barnes 1993:165).

2.5 Daic

The Daic or Tai-Kadai languages, of which Thai is the most well-known and widespread representative, are spoken from southern Thailand into Burma, Laos, Vietnam and China. Up-to-date maps of their distribution are given in Edmondson & Solnit (1997a) who estimate the number of speakers of these languages as at least 80 million. Overviews of the phylum are given in Edmondson & Solnit (1988, 1997a). Figure 7 shows the view of the internal relationships of Daic given by Edmondson & Solnit (1997b);

Figure 7. Daic according to Edmondson & Solnit (1997b)



All the diversity of Daic languages is in China; despite the southward extension of Thai today, the likely origin of Daic is in Kweichow. The external affiliations of Daic have remained highly controversial, sharing as it does many features with surrounding language phyla, notably Austroasiatic, Hmong-Mien and Sino-Tibetan. These were used by Benedict (1975, 1990a) to erect ‘Austro-Tai’, a macrophylum that would unite Austroasiatic, Hmong-Mien, Daic and Austronesian. The general trend however, has been in the opposite direction; to regard each of these phyla as distinct and unrelated. Thurgood (1994) has demonstrated that much of the evidence for hypotheses that linking together the major language phyla of SE Asia, such as Benedict’s Austro-Tai, derive from ancient loanwords rather than genuine cognacy.

Ostapirat (2005) has recently proposed a series of regular correspondences linking Daic with Austronesian. Ostapirat assumes a simple model of a primary split with Daic being the Austronesians who stayed at home. But this seems unlikely as Daic looks more like a branch of proto-Philippines and does not share in the complexities of Formosan. Sagart (2005) has fleshed out a proposal which has proto-Daic speakers migrating back across from the northern Philippines to the region of Hainan island; hence the distinctiveness of Hlai, Be and Daic, resulting from radical restructuring following contact with Hmong-Mien and Sinitic. If so, such a migration would be around 4000 BP, in conformity with current dates for the first incursions in the Northern Philippines.

Daic languages are not all that diverse and almost certainly a candidate for a major agricultural expansion. Despite this, there is no obvious archaeological correlate. Blench (2005) has presented some evidence for thinking that speakers of proto-Daic were not originally rice-cultivators, but borrowed these techniques from Austroasiatic speakers. Reconstruction has yet to produce evidence for their subsistence strategies, and it may be that they were originally cultivators of tubers such as taro, which would fit with the links with the islands. But without a deeper knowledge of the pattern of Daic dispersal it is hard to link them directly with any of the known archaeological horizons of south China.

2.6 Austroasiatic

Austroasiatic languages are the most poorly researched of all those under discussion. Many are not documented at all and some recently discovered in China are effectively not classified. The genetics of Austroasiatic speakers are almost unresearched. Austroasiatic is conventionally divided into two families, Mon-Khmer (in SE Asia) and Muṅḍā (in India). Diffloth (2005:79) now considers Austroasiatic to have three primary branches but no evidence for these realignments has been published. Indeed Austroasiatic classification has been dogged by a failure to publish data, making any evaluation of competing hypotheses by outsiders a merely speculative exercise. With these reservations, therefore, Figure 8 shows this most recent ‘tree’ of Austro-Asiatic;

The main branch of Austroasiatic in China is Western Palaungic, a cluster of languages such as Hu and Kon Keu and some of the Waic languages close to the Burmese border. Palyu, a branch of Austroasiatic, consists of two languages, Bogan and Bolyu. Bolyu, also known as Lai, is found isolated from the remainder of Austroasiatic in Guangxi and is probably a migrant group from further south (Benedict 1990b). There are also four unclassified Austroasiatic languages in China listed above as Pakanic; Bungan, Buxinhua, Kemiehua and Kuanhua spoken by very small populations in southwestern Yunnan.

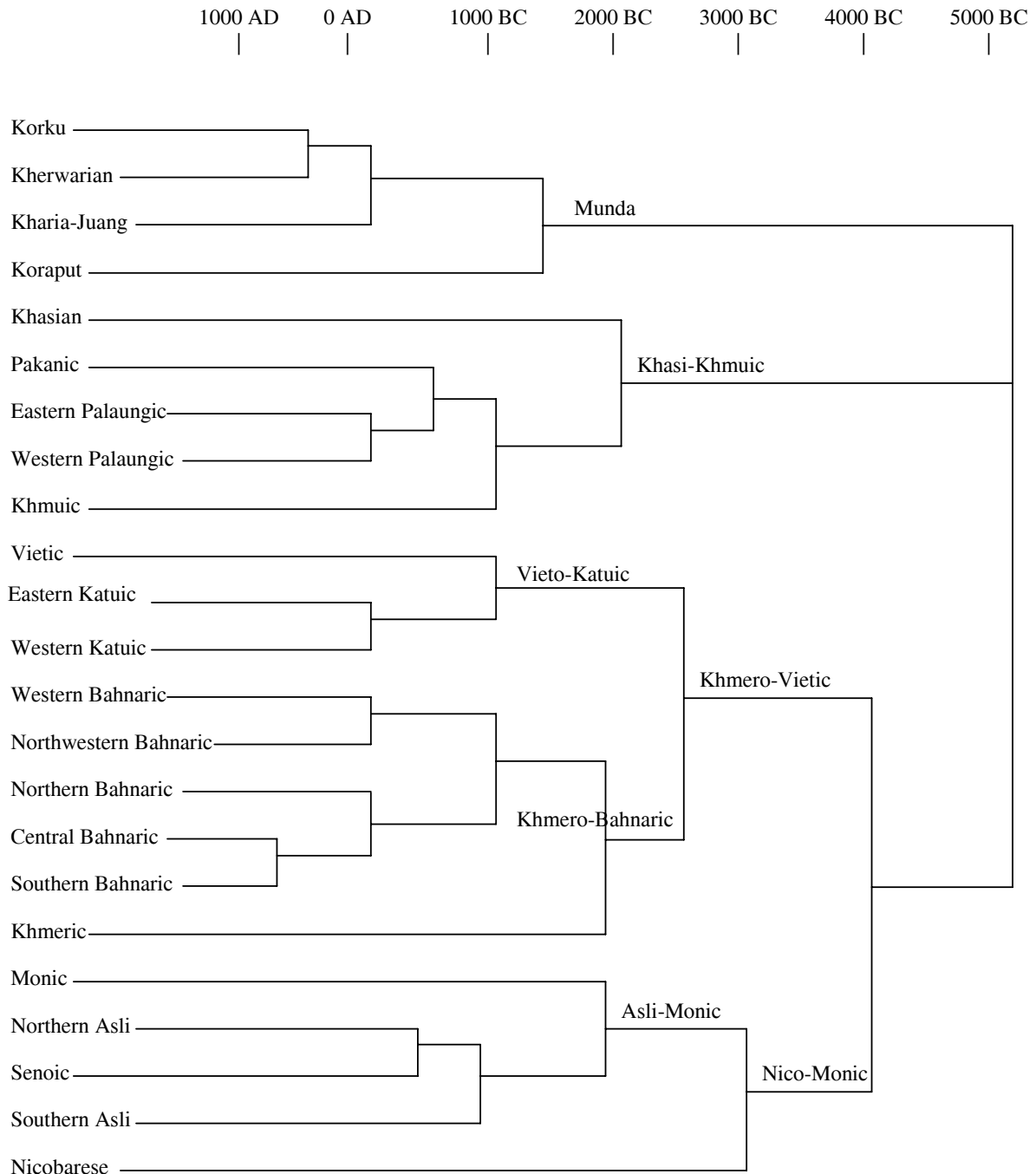
It has long been argued that Austroasiatic was once much more widespread in China and was driven south by the expansion of the Han (Norman & Mei 1976). Some names of zodiacal animals, the Old Chinese word for ‘river’ and ‘tiger’ appear to be borrowings (Norman 1988:18). This has been related to a more general identification of northern regions as the homeland of Austroasiatic. Van Driem (2001) describes a number of theories including the ‘northern shores of the Bay of Bengal’. Blust (1996) put forward the idea that the homeland of proto-Austric (a hypothetical macrophylum uniting Austronesian and Austroasiatic) was in Leaping Tiger Gorge in Yunnan but this has been left adrift by doubts about the validity of Austric. Diffloth’s (2005) claim that Austroasiatic speakers typically spread along river valleys seems to be justified, although they obviously became seagoing at some point. Austroasiatic languages are very fragmented; the spread of Austronesian, Sino-Tibetan and Daic in more recent times has isolated populations among other phyla.

Although there have been many promises, there are *no* justified proto-Austroasiatic reconstructions. It is impossible to see whether faunal or crop names are really supported by a reconstructed proto-language. Diffloth (2005:78) has claimed that faunal reconstructions support a tropical origin and that an elaborated rice terminology indicates an already agricultural society, but the evidence for this remains unpublished and without a date, so it is difficult to relate to a dated palaeoenvironment. The South China/Myanmar/Laos is an important area of biodiversity and there is evidence that Austroasiatic languages were once more widespread in China. Is it likely that southern China was the homeland area of Austroasiatic?

A possible archaeological correlation is the geometric cord-marked pottery that is found in South China prior to 5000 BC (Chang 1986:95). Pottery has been recovered from sites such Hsien-jen-tung and Tseng-p’i-yen dated by TL to >7000 BP, which makes it the earliest pottery in China. This was originally assumed to be similar to the ‘Neolithic’ represented by Spirit Cave in NE Thailand, but the notion that this represented early farmers has now been discredited (Higham & Thosarat 1998).

Nonetheless, similarities between the artefacts do suggest they represent a related culture unless the pottery in Spirit Cave is intrusive. This distribution area also correlates with Daic speakers but if our sense of the coherence of Daic is correct, too early to represent their expansion. This date does approximately correlate with those advanced by Diffloth although he canvases quite a different area of origin for Austroasiatic.

Figure 8. Austroasiatic with calibrated time-depths according to Diffloth (2005)



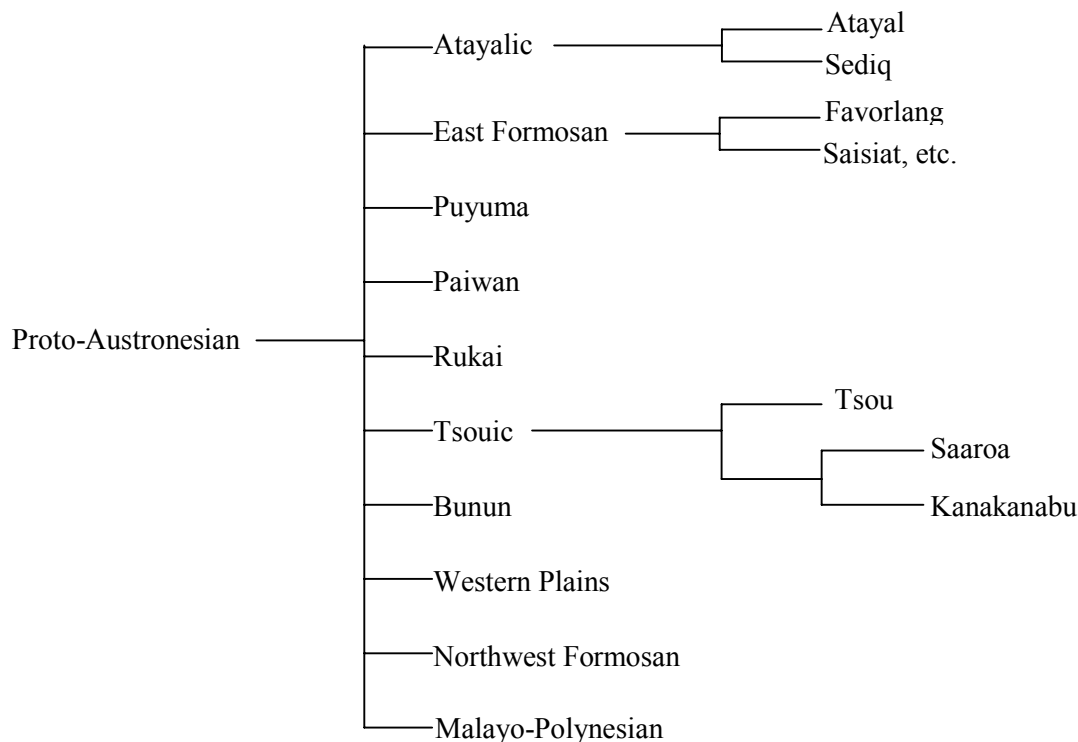
2.7 Austronesian

Austronesian is the second-largest language phylum in the world after Niger-Congo and certainly one of the most widespread, stretching from Easter Island to Madagascar (Bellwood et al. 1995). Compared with many of the other phyla in this region, its internal structure is relatively transparent

and there are few doubts about the languages it includes (with the exception of Daic, see above). Its possible external affiliations are numerous and almost all language phyla of the adjacent mainland have been canvassed. China is presently on the very edge of its distribution and the sole Austronesian language, Tsat, spoken in China today, is not a representative of an older stratum of Austronesian connected to Formosan, but a later migration from insular SE Asia (Thurgood 1999). Tsat is a close relative of Ronglai, a Chamic language found in Vietnam and the founders of the Utsat community probably fled to Hainan after break-up of the Cham Empire.

The usual view of Austronesian is that Formosan forms one branch opposed to the remainder, Malayo-Polynesian. Blust (1999) has challenged this by suggesting that Formosan languages are so diverse as to form a series of high-level primary branches. Figure 9 shows the top-level structure of Austronesian according to Blust (1999);

Figure 9. Austronesian according to Blust (1999)

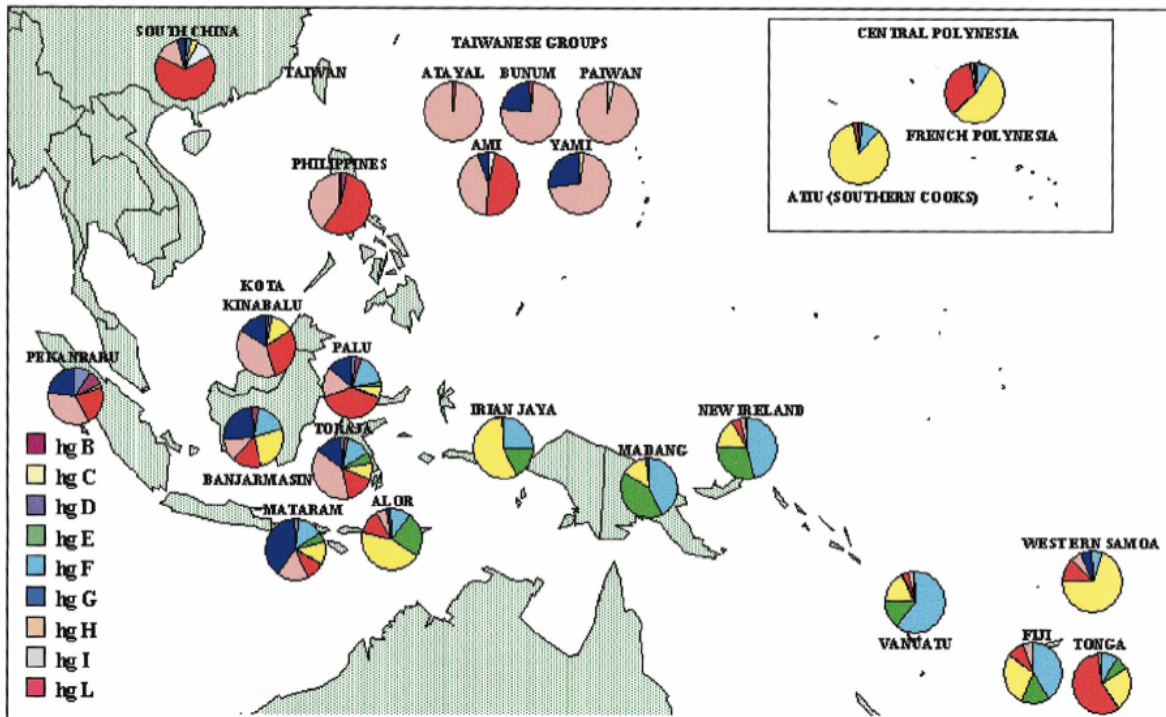


Sagart (2004, this volume) has been active in arguing for a ‘Sino-Tibetan-Austronesian’ and has argued that the Formosan languages diversified in a chain around the island before expanding southwards, which would explain why there is a significant chronological gap between settlement from the mainland and further expansion towards the Philippines.

Although there are no Formosan-type languages spoken in China today, it is widely accepted that the ancestors of the Austronesian peoples crossed from the mainland. At that period, the population would have consisted of Pleistocene hunter-gatherers represented by the cave site at Ch'ang-pin on the eastern coast and the sites of O-luan-pi II and Lung-K'eng on the southern coast. A link with the Ta Peng Keng, or Corded Ware culture, was first proposed in Ferrell (1966:124) and was later taken up by a variety of authors, most recently Tsang (2005). The Hemudu site in Zhejiang, south of Shanghai, north of Taiwan is usually identified as a typical source area (Chang 1981). The inhabitants of Hemudu were rice-growers, with advanced woodworking and maritime technology. The pottery at Hemudu is black, cord-marked ware that shares designs with the Ta Peng Keng, but is obviously at the extreme margin of its distribution. Tsang (2005:71) argues that the most likely source area is the Pearl River Delta and that recent finds show close affinities with the Neolithic of Hongkong.

Genetics broadly supports these conclusions; Melton *et al.* (1998) argued from an analysis of Taiwanese DNA for an Austronesian homeland on the mainland. Capelli *et al.* (2001) explored the patterns of paternal DNA, using 10 haplogroups, in Austronesian, Papuan and South China populations. Although the authors seem more interested in demonstrating the absence of a contribution from *Homo erectus*, the distribution of their haplogroups H and L have some interesting stories to tell about the Austronesian expansion. L is dominant in South China populations, common in Amis, the Philippines and parts of Indonesia, virtually disappearing in Melanesia and re-appearing markedly in Polynesia. Haplogroup H is present in South China but becomes dominant in most of the Formosan groups, and is present throughout Indonesia. Figure 10 shows the geographic locations of populations analysed by Capelli *et al.* (2001) with the proportions of each haplogroup;

Figure 10. Geographic locations of populations analysed by Capelli *et al.* (2001)



Source: Capelli *et al.* (2001:435) [including mis-spellings]

Haplogroup C looks rather as if it represents the Papuan-related Pleistocene hunter-gatherers of Indonesia, although one would expect these to be also present in the Philippines.

2.8 Indo-European

Northwest China also has Indo-European outliers, notably Tajik (Sarikoli) and the Wakhi, Iranian languages of the Pamir branch spoken around Xinjiang, relatively recent intrusions, relics of the Silk Route trade. China is the home of Tocharian, a language attested in manuscripts found in the Tāklimakan desert. The linguistic features of Tocharian link it to Celtic and Italic, rather than the Indo-Iranian languages that would seem immediately more likely. A further impetus to these discoveries has been given by the mummies first uncovered in Xinjiang in 1988, which have been recorded at various sites, all representing linked but distinct historical layers, dating back to 4000 bp (Mair 1998; Barber 1999; Mallory & Mair 2000). The features of the mummies are surprising by any standards, since the figures are up to 2m. in height, with European features including marked beards, wearing cloths apparently woven in plaid patterns and with women wearing tall ‘Welsh’ conical hats. Needless to say, this hardly squares with nationalist ideologies about Chinese origins, but these images have also sparked a bout of speculation from the European side, with wandering tribes of Celts setting up camp in northwest China and bringing all good things to inner Asia.

Tocharian documents date from the 7-8th centuries; the Tarim Basin mummies from 2000 BC. So the question has been, did the mummies ‘speak’ an Indo-European language? Assuming we are not dealing with stray Celtic supporters, it is reasonable to assume that at least some were Indo-European speakers and that they were hunter-gatherers who somehow wandered this long distance in pursuit of animals. But we cannot prove this and indeed various claims have been made for other affiliations, including Uyghur etc. But thinking of these people as the ancestors of the Tocharians and possibly the people who transmitted some early Indo-European loans in Sinitic⁶ would be the simplest solution.

Mallory & Mair (2000: 302) consider the problems at some length and conclude there is probably no unitary solution. Without unwinding the whole argument, they conclude the mummies probably fall into four different groups in terms of physical type and that these are partly correlated with locations and dates. Table 3 shows their assignments;

Table 3. Physical types and linguistic affiliation of Tarim Basin mummies

Location	Hypothetical language	Physical type	Date
Chārchān	Prākṛit, Korānian	?	1000 BC
Lopnur	Prākṛit, Korānian	Proto-Europoid, Indo-Afghan	1800 BC
Qumul	? Tocharian A	Proto-Europoid	1000 BC
Turpan	? Tocharian A	Proto-Europoid	4-5 th centuries BC
	Tocharian B	Indo-Afghan, Pamir-Ferghana	

Source: Mallory & Mair (2000: 302)

The general conclusion is that there are two distinct layers of Europoid populations represented among the Tarim mummies, one representing Tocharian and thus affiliated to far western populations, the other more closely relating to the Indo-Iranian languages and the peoples of the Hindu Kush.

2.9 Korean

China is on the very edge of the Korean-speaking area, in Jilin Province, adjacent to the North Korean border. Korean today is an isolated language, linked to Altaic, but not closely. However, in an earlier period there must have been a linguistic family, Koreanic, with more diversity than is apparent today, and probably spread over a broader area of NE China. Accounts of the ‘Neolithic’ in Jilin (Zhen-hua 1995) and Heilongjiang Provinces (Ying-jie 1995) suggest they a similar culture with strong links to the Korean peninsular, dating to >4000 to >2000 BC. Fish and aquatic resources were apparently of

⁶ The much cited example of ‘honey’, Old Chinese **mit* < Tocharian B *mit*. See Pulleyblank (1983). Sinitic also borrows from Indo-Aryan, for example the words for ‘grape’ and ‘jasmine’ (Norman 1988:19).

major importance in their diet and are characterised by incised and impressed pottery with geometric markings. It is possible that these regions were originally populated by Koreanic speakers.

3. Contributions from genetics

Recent years have seen an explosion of publications on molecular biology in relation to East Asian populations. Although some of these address the question of the peopling of China, it is often difficult to match their conclusions with archaeology and linguistics. Chu *et al.* (1998) exemplify the rather worrying tendencies of 'official' genetics. They start with the figure of 55 minorities, which includes Taiwanese populations and conclude, 'Genetic profiles of 28 populations sampled in China supported the distinction between southern and northern populations, while the latter are biphyletic. Linguistic boundaries are often transgressed across language families studied, reflecting substantial gene flow between populations.' From this they jump to an amazing map of the peopling of China which certainly makes no sense in relation to any archaeological or linguistic data. Ding *et al.* (2000) then directly contradict this. They say, 'Archaeological, anatomical, linguistic, and genetic data have suggested that there is an old and significant boundary between the populations of north and south China. We use three human genetic marker systems and one human-carried virus to examine the North-south distinction. We find no support for a major north-south division in these markers; rather, the marker patterns suggest simple isolation by distance.'

By contrast, Guo *et al.* (1998) who looked at the types of JC polyomavirus, found that it subdivides into four major types in China, and that there is a very distinctive Mongolian type, B1-b, not found in Han populations, whereas other minor Mongolian types are. In China itself, the CY type was characteristic of North China and SC of the South. These differences are attributed to extensive mixing with minorities in the south and rather less with Altaic populations further north.

The approach taken by Mountain *et al.* (1992) to the evolution of Sinitic is quite innovative. Because Chinese surnames are extremely conservative they were used as a proxy for genetic affiliation. The linguistic traits of seven main dialect groups of Sinitic were compared with the patterning of surnames in the same geographic areas. Interestingly, the correspondence with lexical features was much greater than with phonological features.

Kivisild *et al.* (2003) confirm the geographical rather than ethnolinguistic specificity of East Asian DNA, although the distribution of the M7 haplogroup 'branch' and its 'twigs' suggests specificity in the case of isolated or island populations, such as Korea, Japan and insular SE Asia. This strongly suggests that, in a sense, as with languages, that large, contiguous mainland areas lead to massive interchange, whether genetically or linguistically.

Oota *et al.* (2002) compared mtDNA variation in continental Asia. They studied, 'mtDNA HV1 sequences for 84 Xi'an and 82 Changsha Han Chinese, 89 Honshu Japanese, and 35 Vietnamese. Comparison of these sequences with other Asian mtDNA sequences reveals high variability within populations, but extremely low differentiation among Asian populations. Correlations between genetic distance and geographic distance, based on mtDNA and Y chromosome variation, indicate a higher migration rate in females than in males. This may reflect patrilocality, as suggested previously, but another plausible hypothesis is that the demographic expansion associated with the spread of agriculture in Asia may be responsible for the extreme genetic homogeneity in Asia.' This seems highly unlikely. Sampling large urbanised groups will probably show evidence of large-scale genetic interchange; to be convincing, the sample would have to include a wide scatter of minorities.

A more assimilable scenario is that exemplified in Bo Wen *et al.* (2004) which looks at sex-biased admixture in 'Southern Tibeto-Burmans' (Bái, Lolo-Burmese, Tùjiā etc.). Haplotype group distributions of Y-chromosome and mtDNA markers indicate that the genetic structure of these populations were 'primarily formed by two parental groups: northern immigrants and native southerners'. The implication is that a key element of ethnolinguistic group formation may have been

the migration of males, who took wives among *in situ* populations. This may be a useful model for the process of Sinicisation and in particular it can be mapped against the deep influence of Sinitic on Bái and Tùjiǎ. Nonetheless, it is unclear what social and migratory process this reflects; perhaps the movement of soldiers or seasonal hunters or cultivators. Zhang et al. (2006) compared CCR2 allele polymorphisms in 15 Chinese ethnic populations distributed widely across the country and taking in all the main language phyla. They found significant allelic variation principally between the Tungusic and Mongolic populations in the northeast and the southern groups, but much less variation between those in the northwest (Xinjiang etc.). They attribute this to a significant founder effect from this region, which would certainly support the linguistic evidence.

4. The peopling of China

Many questions about the dating the spreads of the language phyla of China remain in doubt and therefore answers are highly tentative. But it may be useful to clarify the useful questions.

What populations underlay the Sinitic/ Han Chinese?

The underlying population was probably ethnolinguistically highly diverse but would have consisted of Tungusic-Koreanic speakers in the North, Hmong-Mien in the centre, intertwined with other Sino-Tibetan groups, and Austroasiatic and Austronesian speakers in the south. There may well have been more language isolates, especially in coastal areas representing the type of phyltic diversity seen in Siberia. In the far northwest, where Chinese expansion is more recent, there would have been at least two different resident Indo-European groups.

When and from which direction was the Sinitic expansion?

This expansion was from north to south, from millet cultivating to the humid areas where irrigated rice was possible. However, Sinitic languages underwent a significant bottleneck some >2000 years ago and records of the language prior to that are highly idiosyncratic. We can therefore know much less about the expansion of Sinitic prior to this event.

What populations came after the Chinese?

The Turkic speakers in the Xinjiang region represent a late incursion. Pre-Mongolic speakers would have made incursions on the settled villages in northern China as nomadic pastoralism developed. It is also possible that Tungusic is a recent expansion following a bottleneck. The expansion of Daic would have roughly coincided with the expansion of Sinitic and may represent a remigration of Austronesian from insular SE Asia.

What drove the expansion of different phyla or groupings?

Table 4 shows some very speculative motives for the expansion of East Asian language phyla with even more speculative dates.

Table 4. What drove the expansion of East Asian language phyla?

Phylum	Comment	Date BP
Early Sino-Tibetan	? dialect diversification typical of hunter-gatherers	>6000
Sinitic	Neolithic agriculture	3500
Mongolic	development of horse culture and nomadic pastoralism	4500
Tungusic	dialect diversification typical of hunter-gatherers [excepting Manchu]	?
Hmong-Mien	not relevant since present distribution is a late artefact of Sinitic expansion	3000
Daic	agriculture [but of what type?]	3500
Austroasiatic	? tuber and fruit-based agriculture in river valleys	7000
Austronesian	Fishing and rice based subsistence [although to what extent there was an expansion in China is unknown]	7000

An unsatisfactory aspect of this is that we would want to attribute approximate dates to some groups on the basis of the synchronic diversity of their languages. But the present form of Hmong-Mien cannot be very old because existing languages are so tightly knit. This is probably an artefact of the assimilation of much of its prior diversity by Sinitic and its roots will lie much deeper. Similarly, with Sino-Tibetan, the languages that reflect an earlier diversity have become not only isolated but heavily Sinicised, making it difficult to analyse the extent to which they reflect an older stratum of dispersal.

5. Where next?

The paper sets out recent evidence for the distribution of the different language phyla in China and their possible archaeological and genetic correlates. But:

- The linkage between archaeological cultures and ethnolinguistic groupings remains sketchy.
- The antiquity of these groupings is highly controversial.
- The internal classification of Sino-Tibetan is very unsettled, although this is essential to making a rational model.
- Genetics input has been more effective a higher levels in establishing the overall affinities of the mainland populations and less in terms of particular language phyla. Indeed the evidence is that genetic variation is determined more by geography than by linguistic affiliation. This is probably to be expected, given the high levels of interaction between languages.

The reasons for this are;

- Historical linguistics has a very long way to go, especially in reconstructing lexical items that could be linked to subsistence and thence to archaeology. Some phyla remain very poorly served [and it is common for unsubstantiated proto-forms to be published].
- Archaeology remains very patchy with some areas well-known, others not.
- Genetics seems to be solving some large-scale problems about human settlement of the region. But it is difficult to know whether it can contribute to the problems of the interface of linguistics and archaeology. More reliable sampling frames would help.

Genetics and archaeology are proceeding apace in China and the coming years are likely to generate significant new data which will certainly clarify some of the issues raised in this chapter.

References

- Barber, E.W. 1999. *The Mummies of Urumqi*. London: Macmillan.
- Barnes, Gina L. 1993. *China, Korea and Japan: the rise of civilization in East Asia*. London: Thames and Hudson.
- Baxter III, William Hubbard 1992. *A Handbook of Old Chinese Phonology*, Berlin: Mouton de Gruyter.
- Bellwood, P., J.J. Fox and D. Tryon (eds) 1995. *The Austronesians*. Canberra: Australian National University.
- Benedict, Paul King 1972. *Sino-Tibetan: A Conspectus*. Cambridge University Press, Cambridge.
- Benedict, Paul King 1975. *Austro-Thai: Language and Culture, with a Glossary of Roots*. Human Relations Areas Files Press, New Haven.
- Benedict, Paul King 1976. Sino-Tibetan: another look. *Journal of the American Oriental Society*, 96, 2: 167-197.
- Benedict, Paul King 1990a. *Japanese/Austro-Tai* (Linguistica Extranea Studia 20). Ann Arbor: Karoma Publishers.
- Benedict, Paul King 1990b. How to tell Lai: an exercise in classification. *Linguistics of the Tibeto-Burman Area*, 13(2):1-26.
- Blench, R.M. 2005. From the mountains to the valleys: understanding ethnolinguistic geography in SE Asia. In: *The peopling of East Asia*. Sagart, L. Blench, R.M. & A. Sanchez-Mazas (eds.) 31-50. London: Routledge.
- Blust, R.A. 1996. Beyond the Austronesian homeland: The Austric hypothesis and its implications for archaeology', in W.H. Goodenough (ed.) *Prehistoric Settlement of the Pacific* (Transactions of the American Philosophical Society, vol. 86, pt. 5, Philadelphia: American Philosophical Society.
- Blust, Robert 1999. Subgrouping, circularity and extinction: some issues in Austronesian comparative linguistics. In: *Selected papers from Eighth International Conference on Austronesian linguistics*. E. Zeitoun & P. Jen-Kuei Li eds. 31-94. Taipei: Academica Sinica.
- Bo Wen, Xuanhua Xie, Song Gao, Hui Li, Hong Shi, Xiufeng Song, Tingzhi Qian, Chunjie Xiao, Jianzhong Jin, Bing Su, Daru Lu, Ranajit Chakraborty and Li Jin. 2004. Analyses of Genetic Structure of Tibeto-Burman Populations Reveals Sex-Biased Admixture in Southern Tibeto-Burmans. *Am. J. Hum. Genet.* 74:856-865.
- Bodman, Nicholas Cleaveland 1980. Proto-Chinese and Sino-Tibetan: data towards establishing the nature of the relationship. In *Contributions to Historical Linguistics: Issues and Materials*. Frans van Coetsem and Linda R. Waugh eds. 34-199. Leiden: E.J. Brill.
- Bradley, D. 1997. Tibeto-Burman languages and classification. In: *Tibeto-Burman languages of the Himalayas*. D. Bradley (ed.) 1-72. Pacific Linguistics A-86. Canberra: Australian National University.
- Brassett, Philip & Cecilia 2005. *Imperial Tiger Hunters: An Introduction to the Tujia People of China*. Publisher: Antony Rowe Publishing Services.
- Capelli, Cristian et al. 2001. A Predominantly Indigenous Paternal Heritage for the Austronesian-Speaking Peoples of Insular Southeast Asia and Oceania *Am. J. Hum. Genetics*, 68:432-443.
- Chang, Kwang-Chih 1981. The affluent foragers in the coast areas of China: extrapolation from evidence on transition to agriculture. In: *SENRI Ethnological Studies*, 9:177-186. Osaka; National Museum of Ethnology.
- Chang, Kwang-Chih (= Zhāng Guāngzhí) 1986. *The Archaeology of Ancient China* (4th edition, revised and enlarged). New Haven: Yale University Press.
- Chu, J.Y., Huang, W., Kuang, S.Q., Wang, J.M., Xu, J.J., Chu, Z.T., Yang, Z.Q., Lin, K.Q., Li, P., Wu, M., Geng, Z.C., Tan, C.C., Du, R.F. and Jin, L. 1998. Genetic relationship of populations in China. *Proceedings of the National Academy of Sciences of the United States of America*, 95: 11763-11768.
- Da-Shun, Guo 1995. Hongshan and related cultures. In: *The archaeology of Northeast China: beyond the Great Wall*. Nelson, Sarah M. ed. 21-64. London and New York: Routledge.
- Diffloth, Gérard. 2005. The contribution of linguistic palaeontology and Austroasiatic. pp. 77-80 in Laurent Sagart, Roger Blench and Alicia Sanchez-Mazas, eds., *The Peopling of East Asia: Putting Together Archaeology, Linguistics and Genetics*. London: Routledge Curzon.

- Ding YC, Wooding S, Harpending HC, Chi HC, Li HP, Fu YX, Pang JF, Yao YG, Yu JG, Moyzis R, Zhang Y 2000. Population structure and history in East Asia. *PNAS, USA*, 97:14003–14006
- Doerfer, G. 1978. Classification problems of Tungus. *Beiträge zur nordasiatischen Kulturgeschichte. Tungusica* 1:1-26. Wiesbaden: Otto Harrassowitz.
- Driem, G. van 1997. Sino-Bodic. *Bulletin of the School of Oriental and African Studies*, 60 (3): 455-488.
- Driem, G. van 1998. Neolithic correlates of ancient Tibeto-Burman migrations. in Roger Blench and Matthew Spriggs (eds) *Archaeology and Language II*, London: Routledge.
- Driem, G. van 2001. *Languages of the Himalayas: An Ethnolinguistic Handbook of the Greater Himalayan Region containing an Introduction to the Symbiotic Theory of Language*. 2 vols. Leiden: Brill.
- Driem, G. van 2002. Tibeto-Burman replaces Indo-Chinese in the 1990s: Review of a decade of scholarship. *Lingua*, 111: 79-102.
- Driem, G. van 2005. The contribution of linguistic palaeontology and Austroasiatic. pp. 77-80 in Laurent Sagart, Roger Blench and Alicia Sanchez-Mazas, eds., *The Peopling of East Asia: Putting Together Archaeology, Linguistics and Genetics*. London: Routledge Curzon.
- Edmondson, J.A. and D.B. Solnit eds. 1988. *Comparative Kadai: linguistic studies beyond Tai*. Dallas: Summer Institute of Linguistics and the University of Texas at Arlington.
- Edmondson, J.A. and D.B. Solnit eds. 1997a. *Comparative Kadai: the Tai branch*. Dallas: Summer Institute of Linguistics and the University of Texas at Arlington.
- Edmondson, J.A. and D.B. Solnit eds. 1997b. Introduction. In: *Comparative Kadai: the Tai branch*. 1-32. Dallas: Summer Institute of Linguistics and the University of Texas at Arlington.
- Ferrell, R. 1966. The Formosan tribes: a preliminary linguistic, archaeological and cultural synthesis. *Bulletin of Institute of Ethnology*, 21: 97-130.
- Fitzgerald, C.P. 1972. *The Southern expansion of the Chinese people: southern fields and southern oceans*. London: Barrie & Jenkins.
- Golden, P.B. 1998. The Turkic peoples, a historical sketch. In: Johanson, L. and E. A. Csató eds. *The Turkic Languages*. 16-29. London: Routledge.
- Guo J, Sugimoto C, Kitamura T, Ebihara H, Kato A, Guo Z, Liu J, Zheng SP, Wang YL, Na YQ, Suzuki M, Taguchi F, Yogo Y. 1998. Four geographically distinct genotypes of JC virus are prevalent in China and Mongolia: implications for the racial composition of modern China. *J Gen Virol* 79:2499–2505.
- Herforth, D. 2003. A sketch of late Zhou Chinese grammar. In Thurgood, Graham and Randy J. LaPolla (eds.). *The Sino-Tibetan languages*. Routledge Language Family Series. 57-71. London and New York: Routledge.
- Higham, Charles & Rachanie Thosarat 1998. *Prehistoric Thailand: from early Settlement to Sukhothai*. Bangkok: River Books.
- Janhunen, Juha 1993. The teens in Jurchen and Manchu revisited. *Mémoires de la Société Finno-Ougrienne* 215:169-184. [Festschrift für Raija Bartens].
- Janhunen, Juha 1994. Additional notes on Japanese and Altaic [1 & 2]. *Journal de la Société Finno-Ougrienne*, 85:236-240 & 256-260. Helsinki.
- Janhunen, Juha 1998. Ethnicity and Language in Prehistoric Northeast Asia. In: *Archaeology and Language II*, R.M. Blench and M. Spriggs eds. 195-208. London: Routledge.
- Janhunen, J. 2003. *The Mongolic Languages*. London: Routledge.
- Johanson, L. and E. A. Csató eds. 1998. *The Turkic Languages*. London: Routledge.
- Kivisild T, Tolk H-V, Parik J, Wang Y, Papiha SS, Bandelt H-J, Villems R. 2002. The emerging limbs and twigs of the East Asian mtDNA tree. *Molecular and Biological Evolution*, 19:1737–1751.
- Klaproth, Julius Heinrich von 1823. *Asia Polyglotta*. Paris: A. Schubart.
- Kolman, Connie J., Nyamkhisig Sambuughin & Eldredge Bermingham 1996. Mitochondrial DNA Analysis of Mongolian Populations and Implications for the Origin of New World Founders. *Genetics*, 142: 1321-1334.
- LaPolla, Randy J. 2001. The role of migration and language contact in the development of the Sino-Tibetan language family. In: *Areal Diffusion and Genetic Inheritance*. A. Y. Aikhenvald & R.M.W. Dixon (eds.) 225-254. Oxford: Oxford University Press.

- Lee, J. 1978. Migration and expansion in Chinese history. In: *Human migration: patterns and policies*. W.H. McNeill & R.S. Adams eds. 20-47. Bloomington & London: Indiana University Press.
- MacKerras, Colin 1997. *China's Minorities: Integration and Modernization in the Twentieth Century*. Oxford: Oxford University Press
- Mair, V. ed. 1998. *The Bronze Age and Early Iron Age Peoples of Eastern Central Asia*. [2 vols.] Washington, DC/ Philadelphia: Institute for the Study of Man/ University of Pennsylvania.
- Mallory, J.P. & Victor H. Mair 2000. *The Tarim mummies*. London: Thames & Hudson.
- Martin, S.E. 1991. Recent research on the relationships of Japanese and Korean. In: *Sprung from some common source*. S. Lamb ed. 269-292. Stanford: Stanford University Press.
- Matisoff, J.A. 2001. Prosodic diffusibility in South-East Asia. In: *Areal Diffusion and Genetic Inheritance*. A. Y. Aikhenvald & R.M.W. Dixon (eds.) 291-327. Oxford: Oxford University Press.
- Matisoff, J.A. 2003. *Handbook of proto-Tibeto-Burman*. Berkeley: University of California Press.
- Melton T, Clifford S, Martinson J, Batzer M, Stoneking M. 1998. Genetic evidence for the proto-Austronesian homeland in Asia: mtDNA and nuclear DNA variation in Taiwanese aboriginal tribes. *American Journal of Human Genetics*, 63:1807–1823
- Menges, K.H. 1995. [2nd ed.] *The Turkic peoples and languages*. Wiesbaden: Harrassowitz.
- Miller, R.A. 1996. *Languages and history: Japanese, Korean and Altaic*. Bangkok: White Orchid Press.
- Mountain, Joanna K. et al. 1992. Congruence of genetic and linguistic evolution in China. *Journal of Chinese Linguistics*, 20(2):315-332.
- Niederer, Barbara 1998. *Les langues Hmong-Mjen (Miáo-Yáo)—Phonologie Historique*. Lincom Studies in Asian Linguistics. 07. München: Lincom Europa.
- Niederer, Barbara 2004. Pa-hng and the classification of the Hmong-Mien languages. In: Tapp, N. Michaud, J., Culas, C. & G. Yai Lee eds. *Hmong/Miao in Asia*. 129-146. Bangkok: Silk Worm Books.
- Oota, H. et al. 2002. Extreme mtDNA homogeneity in continental Asian populations. *Am J Phys Anthropol*, 118:146-153.
- Ostapirat, Weera. 2005. Kra-Dai and Austronesian: Notes on phonological correspondences and vocabulary distribution, pp. 107-131 in Laurent Sagart, Roger Blench and Alicia Sanchez-Mazas, eds., *The Peopling of East Asia: Putting Together Archaeology, Linguistics and Genetics*. London: Routledge Curzon.
- Norman, J. 1988. *Chinese*. Cambridge Language Surveys. Cambridge and New York: Cambridge University Press.
- Norman, J. 2003. The Chinese dialects: phonology. In Thurgood, Graham and Randy J. LaPolla (eds.). *The Sino-Tibetan languages*. Routledge Language Family Series. 72-83. London and New York: Routledge.
- Norman, J. & Mei T. 1976. The Austroasiatics in ancient south China: some lexical evidence. *Monumenta Serica*, 32:274-301.
- Pulleyblank, Edwin G. 1983. The Chinese and their neighbours in prehistoric and early historic times. In: *The Origins of Chinese Civilization*. David N. Keightley ed. 411-466. Berkeley: University of California Press.
- Purnell, H.C. Jr. 1970. *Towards a reconstruction of Proto-Miao-Yao*. Ph.D. Cornell University.
- Ramsey, S.R. 1987. *The Languages of China*. Princeton, NJ: Princeton University Press.
- Ratliff, Martha 2004. Vocabulary of environment and subsistence in the Hmong-Mien proto-language. In: Tapp, N. Michaud, J., Culas, C. & G. Yai Lee eds. *Hmong/Miao in Asia*. 147-166. Bangkok: Silk Worm Books.
- Sagart, Laurent 1994. Proto-Austronesian and Old Chinese evidence for Sino-Austronesian. *Oceanic Linguistics*, 33(2):271-308.
- Sagart, Laurent 1999. *The Roots of Old Chinese*. Amsterdam: John Benjamins.
- Sagart, Laurent 2004. The higher phylogeny of Austronesian and the position of Tai-Kadai. *Oceanic Linguistics*, 43 (2): 411-444.
- Sagart, Laurent 2005. Sino-Tibetan-Austronesian: an updated and improved argument, pp. 161-176 in Laurent Sagart, Roger Blench and Alicia Sanchez-Mazas, eds., *The Peopling of East Asia: Putting Together Archaeology, Linguistics and Genetics*. London: Routledge Curzon.

- Shafer, Robert 1974. *Introduction to Sino-Tibetan*. Wiesbaden: Otto Harrassowitz.
- Starostin, S.A. 1991. On the hypothesis of a genetic connection between the Sino-Tibetan languages and the Yeniseian and North Caucasian languages. In Vitaly Sheveroshkin, ed. *Dene-Sino-Caucasian*. 12-42. Bochum: Universitätsverlag Dr. N. Brockmeyer.
- Starostin, S.A., Anna Dybo & Oleg Mudrak 2003. *Etymological dictionary of the Altaic languages*. Leiden/ Boston: Brill.
- Su Bing, Xiao C., Deka R., Seielstad M.T., Kangwanpong D., Xiao J., Lu D., Underhill, P., Cavalli-Sforza L.L., Chakraborty R., Jin L. 2000. Y chromosome haplotypes reveal prehistorical migrations to the Himalayas. *Human Genetics*, 107:582–590.
- Svanberg, Ingvar 1988. *The Altaic speakers of China: numbers and distribution*. Uppsala: Uppsala University.
- Tapp, N. Michaud, J., Culas, C. & G. Yai Lee eds. 2004. *Hmong/Miao in Asia*. Bangkok: Silkworm Books.
- Thurgood, G. 1994. Tai-Kadai and Austronesian: the nature of the relationship. *Oceanic Linguistics*, 33(2):345-368.
- Thurgood, G. 1999. *From Ancient Cham to modern dialects: two thousand years of language contact and change*. Honolulu: University of Hawai'i Press.
- Thurgood, Graham and Randy J. LaPolla (eds.). 2003. *The Sino-Tibetan languages*. Routledge Language Family Series. London and New York: Routledge.
- Tsang, C.H. 2005. Recent discoveries of the Tap'enkeng culture in Taiwan: Implications for the problem of Austronesian origins. In: Sagart, L. Blench, R.M. & A. Sanchez-Mazas, (eds.) *Perspectives on the Phylogeny of East Asian Languages*. 63-73. London: Curzon Press.
- Wang Fushi 1994. *Reconstruction of the Proto-Miao Language*. [In Chinese and English] Tokyo: Institute for the Study of the Languages and Cultures of Asia and Africa.
- Wang Fushi and Mao Zongwu 1995. *Miao-yao yu guyin gouni*, Beijing: Zhongguo Shehui Kexue.
- Wang, William, S-Y. 1998. Three windows on the past. In: Mair, V. ed. *The Bronze Age and Early Iron Age Peoples of Eastern Central Asia*. 508-534. [2 vols.] Washington, DC/ Philadelphia: Institute for the Study of Man/ University of Pennsylvania.
- Wiersma, G. 2003. Yunnan Bai. In Thurgood, Graham and Randy J. LaPolla (eds.). *The Sino-Tibetan languages*. Routledge Language Family Series. 651-673. London and New York: Routledge.
- Yin, Ma ed. 1989. *China's minority nationalities*. Beijing: Foreign languages press.
- Ying-Jie, Tan *et al.* 1995. The Neolithic in Heilongjiang Province. In: *The archaeology of Northeast China: beyond the Great Wall*. Nelson, Sarah M. ed. 118-144. London and New York: Routledge.
- Zhang, X. L., C. Y. Zhang, Y. Yu, F. Chen, P. Li & S. B. Fu 2006. CCR2 allele polymorphisms in 15 Chinese ethnic populations. *International Journal of Immunogenetics* 33, 45–48.
- Zhen Hua 1995. Recent Neolithic discoveries in Jilin Province. In: *The archaeology of Northeast China: beyond the Great Wall*. Nelson, Sarah M. ed. 89-117. London and New York: Routledge.