

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

Paper for the Symposium :

## **Human migrations in continental East Asia and Taiwan: genetic, linguistic and archaeological evidence**

Geneva June 10-13, 2004.  
Université de Genève

**[DRAFT CIRCULATED FOR COMMENT]**

Roger Blench  
Mallam Dendo  
8, Guest Road  
Cambridge CB1 2AL  
United Kingdom  
Voice/Answerphone/Fax. 0044-(0)1223-560687  
E-mail [R.Blench@odi.org.uk](mailto:R.Blench@odi.org.uk)  
[http://homepage.ntlworld.com/roger\\_blench/RBOP.htm](http://homepage.ntlworld.com/roger_blench/RBOP.htm)

**Cambridge, Sunday, 06 June 2004**

## TABLE OF CONTENTS

<b>FIGURES.....</b>	<b>i</b>
<b>1. Introduction.....</b>	<b>3</b>
1.1 The problem: linking linguistics, archaeology and genetics.....	3
1.2 Methodological issues.....	3
<b>2. The linguistic pattern of present-day China.....</b>	<b>5</b>
2.1 General.....	5
2.2 Sino-Tibetan.....	5
2.3 Miao-Yao.....	9
2.4 Altaic.....	10
2.5 Daic.....	11
2.6 Austroasiatic.....	13
2.7 Austronesian.....	15
2.8 Indo-European.....	17
2.9 Korean.....	18
2.10 Isolates?.....	18
<b>3. Contributions from genetics.....</b>	<b>18</b>
<b>4. The peopling of China.....</b>	<b>20</b>
<b>5. Where next?.....</b>	<b>21</b>

## TABLES

Table 1. Language Phyla of China.....	5
Table 2. Horse terms in East Asia.....	11
Table 3. Physical types and linguistic affiliation of Tarim Basin mummies.....	18
Table 4. What drove the expansion of East Asian language phyla?.....	20

## FIGURES

Figure 1. Possible elements in reconstructing China's prehistory ?.....	3
Figure 2. Sino-Tibetan according to Matisoff (2001).....	6
Figure 3. 'Fallen leaves' model of Sino-Tibetan according to Van Driem (many places).....	7
Figure 4. Miao-Yao [=Hmong-Mien] according to Matisoff (2001).....	9
Figure 5. Altaic and Macro-Altaic.....	10
Figure 6. Daic according to Edmondson & Solnit (1997).....	12
Figure 7. Austroasiatic with calibrated time-depths according to Diffloth.....	14
Figure 8. Austronesian according to Blust (1999).....	16
Figure 9. Geographic locations of populations analysed by Capelli et al. (2001).....	17

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

Roger Blench  
Mallam Dendo Ltd.

Overviews of the prehistory of China have been dominated by a crypto-Marxist developmental schema, whereby archaeological horizons lead inexorably towards the historical, Han-dominated political structures. The substantial and complex minorities also present in China are rarely mentioned and little is known of the way the present ethnolinguistic pattern of China was established. The situation has not been helped by official stereotypes in China which established a standard number of 55 minorities into which all groups must be shoehorned (a total which includes the Gaoshan, a collective term for all the Austronesian peoples of Taiwan). Although it is now recognised that internal diversity is greater, this framework continues to be acknowledged in many places and indeed used in genetics articles, despite its inappropriateness. Ethnologue estimates there are around 200 languages in China, but this includes 13 dialects of Chinese. The following language phyla are represented in China;

- Sino-Tibetan/Tibeto-Burman
- Austroasiatic
- Daic= Tai-Kadai = Kra-Dai
- Hmong-Mien = Miao-Yao
- Altaic (Turkic, Mongolic, Tungusic)
- Austronesian (Chamic)
- Indo-European (Tajik, Wakhi, Russian etc.)
- Korean
- Unclassified (Waxianghua)

According to 1998 data, minorities constitute some 91,000,000, so they are relatively numerous compared with other countries in the region. However, 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. In particular, two groups are controversial; the putative mainland Austronesians and the Tocharians, who were probably related to the Celtic Tarim Basin mummies.

What are the issues in the peopling of China?

- What populations underlay the Han Chinese?
- When and from which direction was the Chinese expansion?
- What populations came after the Chinese?
- What drove the expansion of different phyla?
- What are the archaeological and genetic correlates of these phyllic expansions?

The paper sets 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

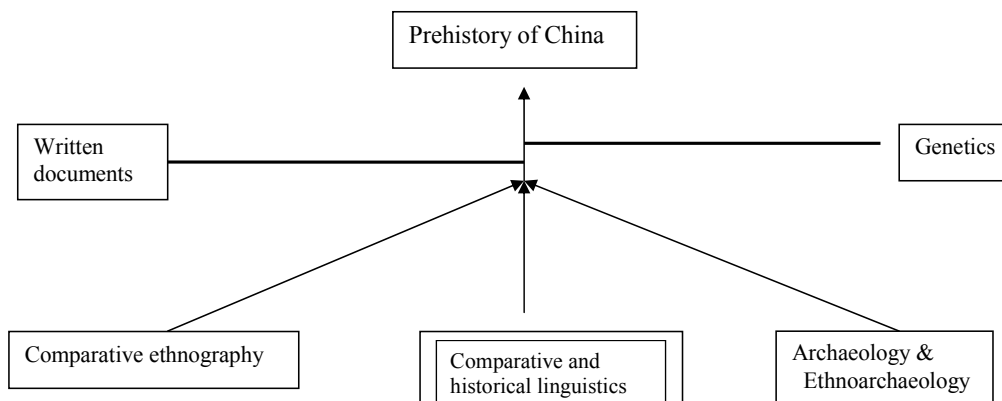
## 1. Introduction

### 1.1 The problem: linking linguistics, archaeology and genetics

The concept of linking linguistics, archaeology and genetics in the reconstruction of the past is becoming a commonplace at certain types of academic conference; but the reality is that each discipline largely pursues its own methods and what little interaction there is remains marginal. Generally, despite much talk of interdisciplinary work, individual disciplines are driven by their own methods. Hence many of the questions they ask are internal, addressed to colleagues, not the discipline.

China is a particularly bad case because so much of the linguistics and archaeology is driven by an obsession with high culture. Major archaeological texts refer neither to linguistics nor genetics and there is rather wild speculation about the identity of some non-Chinese groups mentioned in the texts. In addition, the 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 useful. A preliminary outline of an agenda for inter-disciplinary study is set out in Wang (1998) who characterised linguistics, archaeology and genetics and ‘three windows on the past’. Figure 1 represents a potential multi-disciplinary framework for reconstructing China’s past.

**Figure 1. Possible elements in reconstructing China’s prehistory ?**



### 1.2 Methodological issues

A key assumption of this type of trans-disciplinary enterprise is that results can be matched, that patterns of language distribution are, in principle, congruent with archaeology. But this is not accepted, particularly by many archaeologists, for whom linguistics is simply a separate discipline and for whom ‘the makers of the pots must remain silent’. The argument is that since both archaeology and linguistics are direct reflections of human activities, they must, in some way, be congruent. One good reason for thinking this is there is a clear congruence in the present; culture and language *are* clearly linked and divergences can be explained by relatively simple sociolinguistic processes<sup>1</sup>. The single biggest problem in linking various approaches is that within a discipline it is neither fashionable nor popular to frame questions in terms of the questions asked by another discipline. So archaeologists give almost no time to match the patterns of the cultures they delineate with historical linguistics and linguists are often uninterested in collecting reconstructing terms and concepts that could illuminate historical hypotheses.

This potential for congruence is *not* necessarily the case with genetics; genes are not people, and they have a distributional logic quite different from languages and cultures. They reflect extensive and complex patterns of human interactions with each other and the environment at a one-to-one and one-to-many level. It seems

---

<sup>1</sup> English is the most intensively studied language in the world, and recent explorations of its varieties make it perfectly possible to account for both variation and the congruence or otherwise of the cultures of those who speak it.

to be no quirk of the analytic process that maps of different marker systems and haplogroups seem to reflect geography more than ethnicity and often do not map against each other. One consequence of this is ‘Cavalli-Sforza syndrome’ where the geneticist sorts frantically through a series of highly diverse maps and eventually finds one that approximately corresponds to a known linguistic or archaeological grouping. By extracting different statistical components the match is improved and a triumph for congruence is announced. But this is frankly improbable; all other types of data suggest that on any large contiguous land mass, populations interact in such an intensive and complex fashion as to make congruence unlikely. It is only on islands or other isolated locales where interaction with external populations is more constrained that there is a conservation of genetic traits; hence there has been more success with integrated account in Polynesia than in Eurasia or Africa. There may also be an issue of scale; the conclusions of genetics seem better drawn over very large areas. Much of the recent successes of genetics have been in modelling of the expansion of modern humans out of Africa, the demonstration that there was no genetic interchange with *in situ* hominids and the demarcation of different migration routes. This is reflected in typical ‘geographical’ results, where patterns underlie the broad processes of human expansion, rather than the micro-movements associated with local cultural processes.

Another aspect of genetics that is difficult to match to the other disciplines are its dates. Geneticists often believe in a ‘clock’ that can measure the speed of genetic divergence and thus they assign dates to human dispersals (some egregious examples are given later in the paper). It seems difficult to know how such a clock can be calibrated against anything except archaeology. Apart from the remaining adherents of glottochronology, few linguists now believe that language dispersals can be dated except by correspondences with archaeology; the same should hold for genetics.

Related to this is an issue often raised by geneticists, that of language diversity. Genetics can often put a quantitative measure on diversity and wonder whether this can be mapped against linguistic diversity. This seems as if it ought to work, but it doesn’t, because languages diversify in different ways. Australian and Papuan are well-known for being highly diverse lexically and extremely uniform phonologically. Daic languages are quite uniform lexically but extremely diverse tonally. Khoesan and Nilo-Saharan languages are diverse in almost every conceivable way. Mountain *et al.* (1992) report on measures of diversity within Sinitic, but show that different categories of linguistic feature show different levels of diversity. This is not to say that diversity carries no information at all. The diversity within Australian and Papuan clearly indicate the long-term settlement of these regions; but whether anything more precise can be extracted from their variety is open to question.

Linguistics and archaeology are not of course driven by the spirit of pure enquiry; archaeology in particular is often prone to hijacking by nationalist agendas. This is not a new point, but the development of the nation state in the twentieth century has resulted in a bizarre framing of accounts of the past in terms of the boundaries of the present. It encourages archaeological accounts to view the horizons of the past as leading inexorably towards those of the present. Typically, in China, ancient cultures become precursors of the Han state, rather than perhaps, dead ends<sup>2</sup>. This is very persuasive but misleading; most of what we know about Sinitic suggests that the Han expansion is quite recent and therefore almost any older archaeological culture is *not* likely to be associated with Sinitic speakers.

A key issue in linguistics that can be very perplexing for outsiders in the East Asian region are macrophylic schemas. A number of scholars consider that many of the language phyla of East Asia are related to one another. Unfortunately, their maps of these relationships is highly variable. The affiliation of Sino-Tibetan is a particular problem, with a more ‘conventional’ view linking it with Miao-Yao or Daic and wider hypotheses that bring in Caucasian or Austronesian. Similarly, Austronesian, Austroasiatic and Daic are often linked. Indeed, some authors seem to think that all these phyla will ultimately prove to be related. When working on the problem of correlation with other disciplines it is best to retain a minimalist view;

---

<sup>2</sup> It is interesting to compare these with Stephen J. Gould’s strictures on models of evolution that are structured so as the always finish with the evolution of modern humans, rather than being full of byways and forking paths that lead nowhere.

namely that while these views *may* reach a consensus among scholars in the future, at present we need to look for the correlates of agreed groupings.

The reconstruction of some parts of Sino-Tibetan has been confused by the existence of archaic written texts. Much historical scholarship has gone into the reconstruction of Old Chinese, a language that would consistently account for the system of ancient texts. But there is, and can be, no evidence that such a language was ever spoken, and no necessary link with proto-Sinitic, a language reconstructed from the wide range of modern dialects. Similar problems have arisen by confusing Sanskrit with proto-Indo-Aryan, as Turner does in his magisterial volumes. Probably if we had better proto-Sinitic, there would be less problem about its place within the larger Sino-Tibetan schema.

Historical linguists tend to work with ‘tree’ models, where languages split, usually in binary fashion, and this is evidently convenient when trying to fashion a correspondence with archaeology as a chronology can be developed. But some linguists are sceptical of these models and it is clear that languages do not always develop in such a convenient fashion. Indeed it seems likely that the common pattern of mainland East Asian languages with reduced morphology, complex tones and simplified word structures represents massive convergence between different language phyla. Nonetheless, it is also difficult to work with non-trees, ‘fallen leaves’, because these present no sense of chronology.

## 2. The linguistic pattern of present-day China

### 2.1 General

Although dominated numerically by Sino-Tibetan, China is highly diverse linguistically. Table 1 shows the main language phyla represented;

**Table 1. Language Phyla of China**

<b>Phylum</b>	<b>Examples</b>
Sino-Tibetan/ Tibeto-Burman	
Hmong-Mien = Miao-Yao	
Altaic (Turkic, Mongolic, Tungusic)	
Daic= Tai-Kadai = Kra-Dai	
Austroasiatic	
Austronesian (Chamic)	Tsat
Indo-European	(Tajik, Wakhi, Russian etc.)
Korean	
Unclassified	Waxianghua

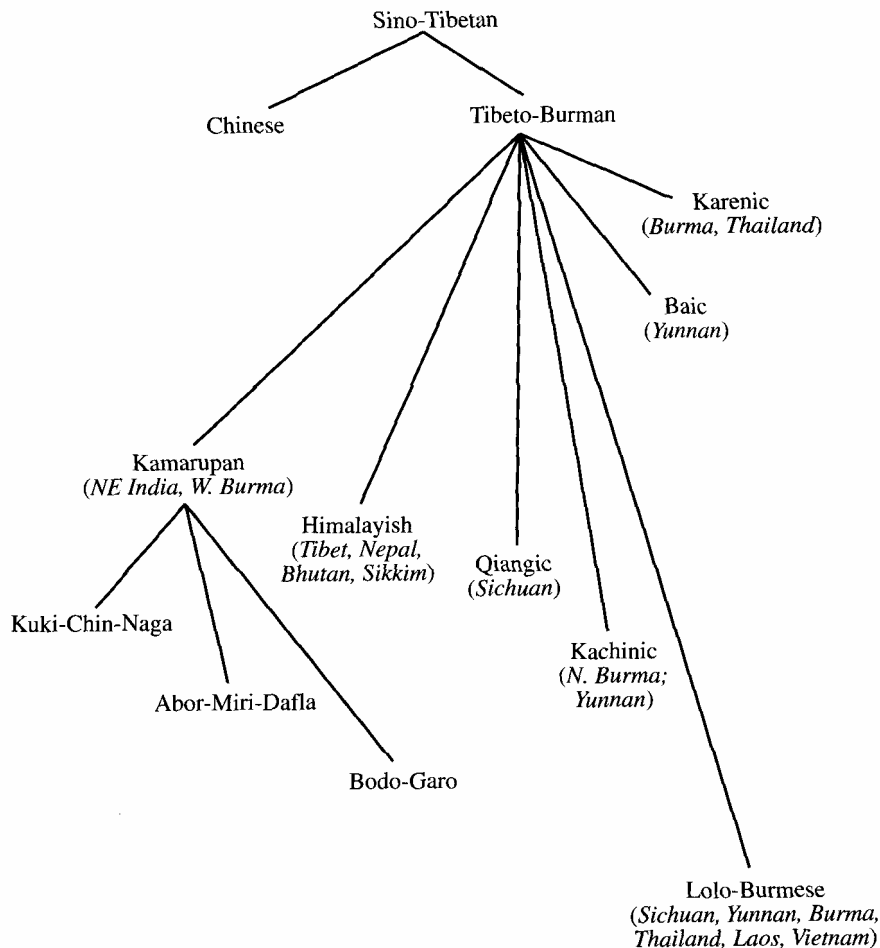
Ethnologue estimates there are around 200 languages, but this includes 13 dialects of Chinese. However, new languages are regularly being recorded, such as the Austroasiatic Bagan, yet to be classified within Mon-Khmer. This figure contrasts with the official count of 56 [55 + Han]. Despite the unlikely nature of the official figure, it continues to be propagated on websites and official documents. According to 1998 data, minorities constitute some 91,000,000, so they are relatively numerous compared with other countries in the region. However, 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. Two topics have generated considerable controversy; the putative mainland Austronesians and the Tocharians, who were probably related to the Celtic Tarim Basin mummies. These are discussed in the sections below which cover the phyla systematically.

### 2.2 Sino-Tibetan

Sino-Tibetan almost certainly has more speakers than any other language phylum, largely because of the size of Chinese. Most other languages are small and remain little-known, partly because of their inaccessibility. The

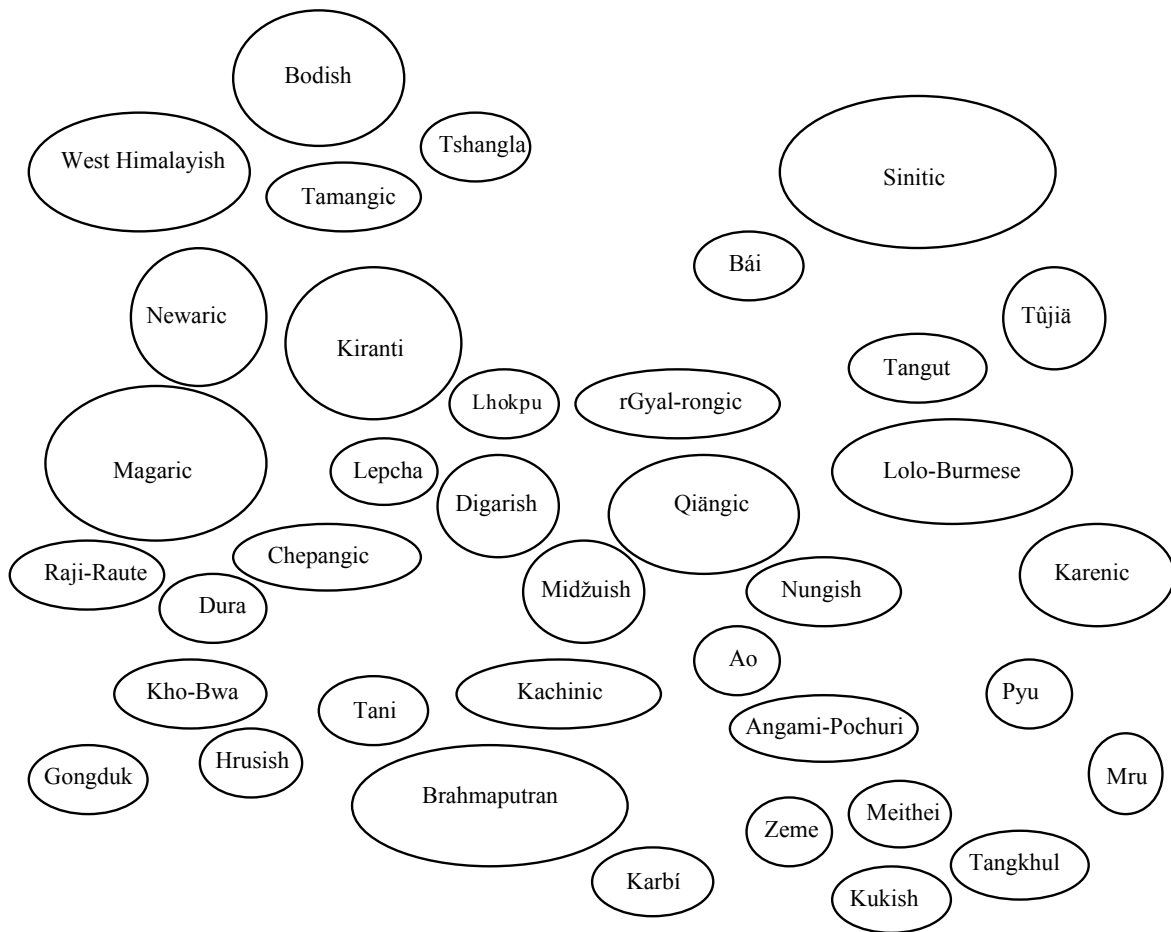
internal classification of Sino-Tibetan remains highly controversial, as is any external affiliation. Sinitic is not very diverse when compared with the rest of Sino-Tibetan and it is likely that it is a ‘recent’ expansion, although it could hardly be less than about 4000 years old. Figure 2 shows the internal structure of Sino-Tibetan according to Matisoff (2001:297) which is about as mainstream as any view. However, it should be emphasised that there is no agreed internal structure and very different views are held by other scholars.

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



This is a worryingly agnostic model as the output of STEDT, a long-running project on the comparative lexicon of Sino-Tibetan. Sinitic and Tibeto-Burman represent the primary split and the other branches are all primary divisions of Tibeto-Burman. I also suspect that this model air-brushes out any languages that don't ‘fit’, such as Gongduk and Magaric. A quite different model is the ‘fallen leaves’ schema of George van Driem (Figure 3);

**Figure 3. ‘Fallen leaves’ model of Sino-Tibetan according to Van Driem (many places)**



This is essentially a geographical model, which defines subgroups which are generally agreed and places them 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.

The key questions that emerge are; whether the primary branching is Sinitic (i.e. all Chinese languages) and the remainder (usually called Tibeto-Burman) or whether Sinitic is simply part of one branch, such as Bodic. Just as Hamito-Semitic privileged Semitic for entirely non-linguistic reasons, it is hard not to suspect that Chinese does not have the distinct status accorded it by the Matisoffian model, but whatever evidence exists for other schemas has failed to win significant assent from the scholarly community. The second major issue is the status of the problematic ‘remnant’ languages of the Himalaya, Gongduk, Magaric and others. Either these are early branchings from the Sino-Tibetan tree or they are ‘Kusundic’, remnants of earlier language phyla that have been Sino-Tibetanised. But having no model to account for them is like reconstructing Austronesian without paying attention to Formosan.

China also has intriguing ‘remnant’ languages such as Tujia, Bai and Waxianghua, hard to classify because they have been so heavily Sinicised. It may be that these are traces of a much more diverse earlier Sino-Tibetan population. Intriguingly, written Chinese texts contain material on other Sino-Tibetan languages that can provide rather fragmentary insights into language diversity in the past (Wang 1998). Bai 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 external affiliations of Sino-Tibetan are of course equally 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



potential for promiscuous cohabitation. Passionate advocacy of Sino-Caucasian and Sino-Austronesian undoubtedly conflict and they clearly cannot both be true. One important implication of this is that loanwords can be embedded to such a depth that it is difficult to distinguish them from fundamental vocabulary. The lessons of the interaction of Austronesian and Papuan are yet to be absorbed on the mainland.

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 from one another. This suggests that it was originally a scatter of 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, the model that has populations spreading down river valleys, popular in many models of phylic 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 website. 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 say, 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. It is therefore *unlikely* that Sinitic can be identified with the earliest Neolithic communities in North China such as the Péilǐgǎng or Cǐshān (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 is 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 Miao-Yao-speaking, but unless these groups were considerably north of their present location, the agriculturists of Cǐshān were not Miao-Yao either. 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

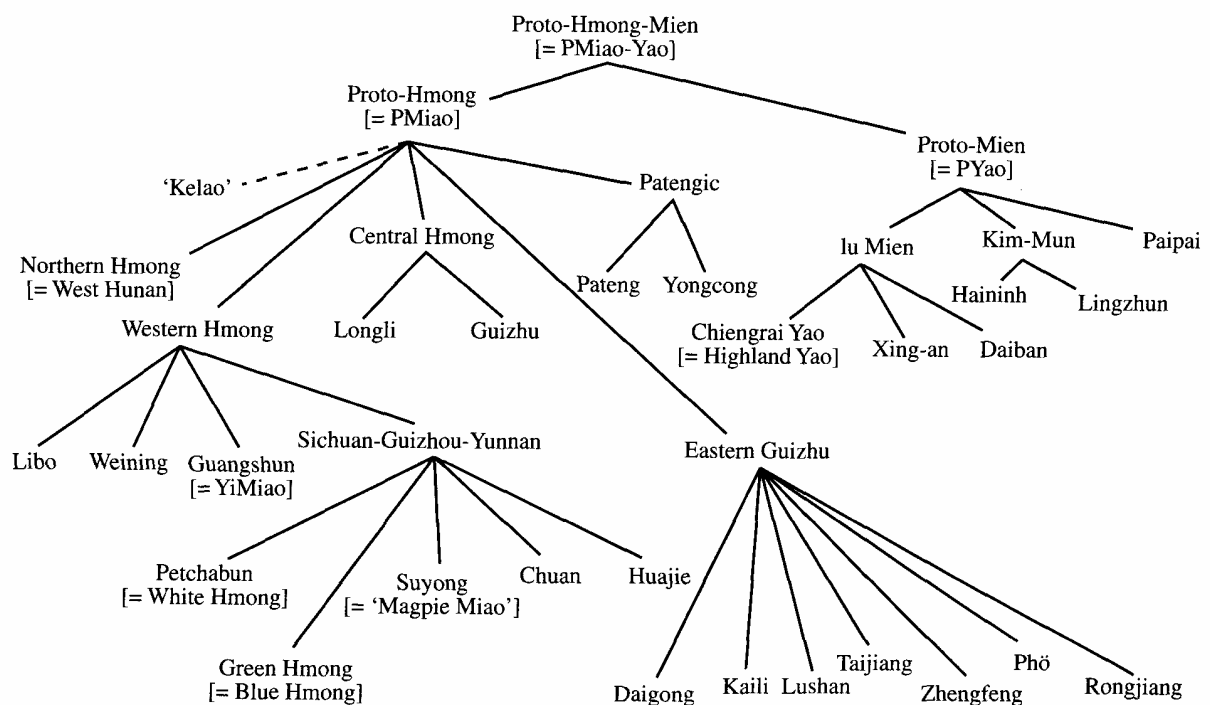
BC, i.e. roughly contemporaneous with the Yangshao. Despite being well outside the imperial boundaries, Da-Shun sees this as ‘the dawn of Chinese civilization’ and attempts to link it with that civilization through a series of typological indicators, a writing system, bronze metallurgy etc. A particular 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 Miao-Yao

The Miao-Yao [=Hmong-Mien] languages are spoken mostly in China with some groups also in Laos, Vietnam and Thailand. Their centre of gravity is between the Yangzi and the Mekong rivers. Miao-Yao languages are quite close to one another, and although the Ethnologue lists some 32 languages, many of these are mutually intelligible lects. There have been various comparative overviews of the group, starting with Purnell (1970), Wang Fushi (1994) and Niederer (1998). The linguistic geography of Miao-Yao 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 Miao-Yao preceded Chinese, should it not be more diverse? One long-running argument is whether the rice-terminology of Miao-Yao was borrowed by Sinitic speakers as they moved south (Blench 2004).

Figure 4 is one possible version of the internal structure of Miao-Yao, although it leaves the placing of She, Ho Nte uncertain.

**Figure 4. Miao-Yao [=Hmong-Mien] according to Matisoff (2001)**



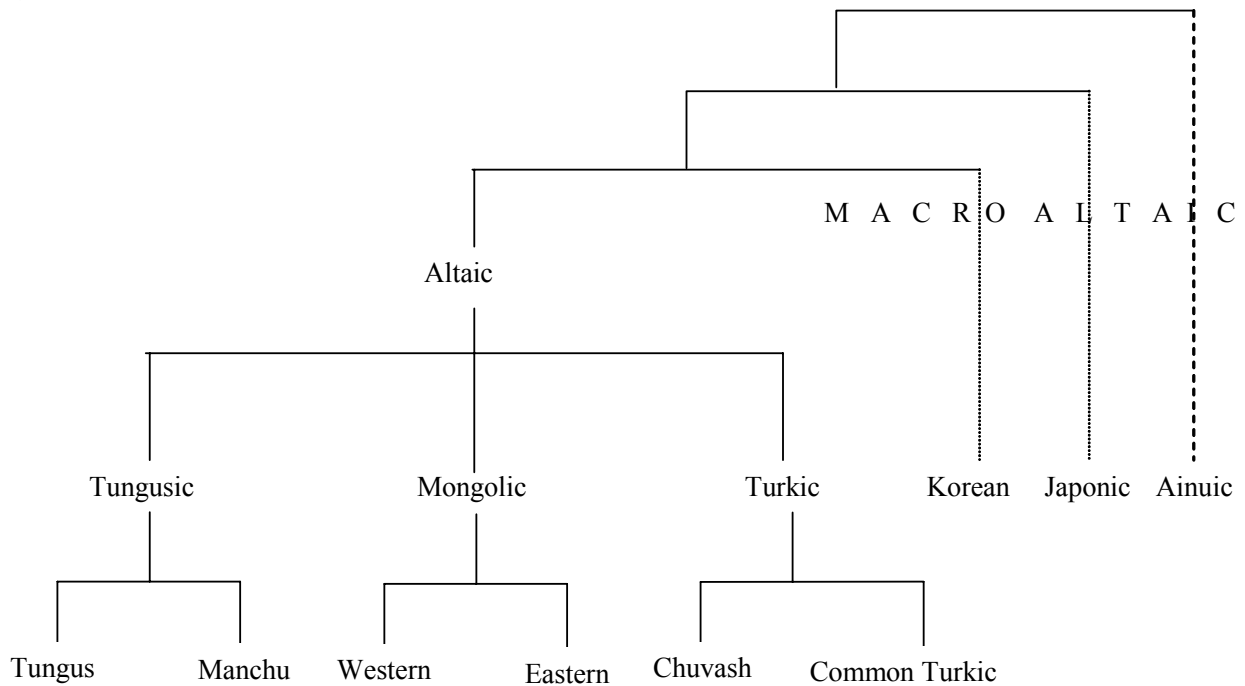
Despite the lack of internal diversity in Miao-Yao, it seems difficult to imagine these are not ancient inhabitants of the East Asian area. It seems as if the other more diverse relatives of Miao-Yao must have been eliminated by the Han expansion and the languages still in existence are the result of a secondary expansion. Japanese studies of the early cities such as Pengtoushan have argued that these are not Sinitic, but Miao-Yao. One of the planks of this argument is based on palynological work that shows the presence of *Liquidambar* sp. which is used for ‘spirit posts’ in Miao-Yao villages. The pre-Miao-Yao may therefore be

identified with one of the Neolithic pottery horizons, but it seems unlikely that the present Miao-Yao diaspora would have any direct correlate, since their dispersal is based on a pattern of refuge rather than positive expansion.

## 2.4 Altaic

The Altaic languages consist of Turkic, Mongolic and Tungusic, all of which are attested in China. They are highly internally divided, so much so that some scholars claim it is not a phylum but a bundling of languages that have interacted (Janhunen 1994). The proposed macro-phylum, Macro-Altaic, remains still more controversial although most scholars accept the membership of Korean, fewer Japanese. Surprisingly, the individual members of Altaic, Turkic etc., are very undiverse and the dispersal of Turkic has largely taken place in historical time. Figure 5 shows a tree representing the Altaic and Macro-Altaic groupings.

Figure 5. Altaic and Macro-Altaic



Today the Turkic languages 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 Csato (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.

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). But it is clear that 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 development of pastoralism with the expansion of the Mongolic languages. Although today these are quite undiverse, this may be 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**

<b>Language group</b>	<b>proto-form</b>
Mongolic	<b>morin</b>
Tungusic	<b>murin</b>
Korean	<b>mar</b>
Japanese	<b>uma</b>
Chinese	<b>ma</b>

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.

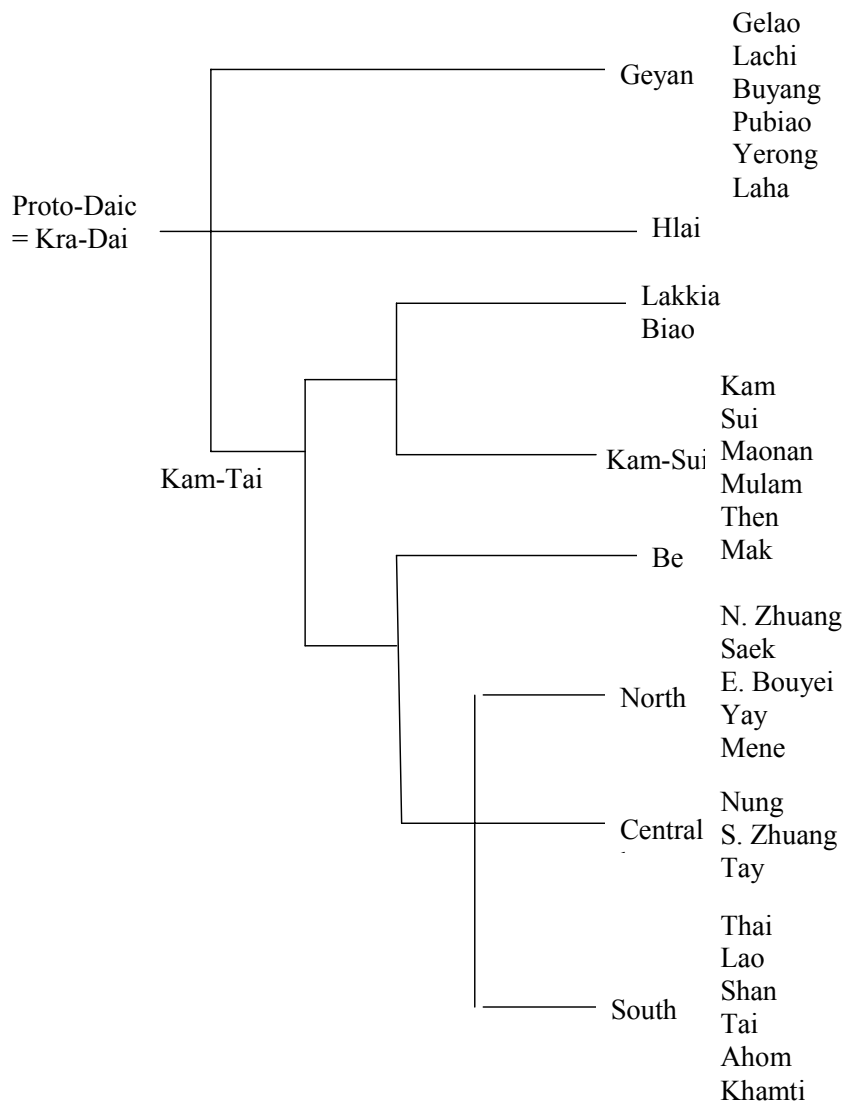
Mongolic languages today are dominated by Khalkh Mongol, spoken throughout much of modern Mongolia, with outlying Mongolic languages spoken in China and Afghanistan. The relative uniformity of Mongolic can be attributed to the empire of Chinggis Khan (ca. 1200 to 1400 AD) which grew to control the largest land empire ever recorded and probably eliminated earlier ethnic and linguistic diversity during this period. Janhunen (1993) has analysed lexical elements borrowed from Mongolic into Manchurian Tungusic to argue that the family formerly exhibited much greater diversity. Kolman *et al.* (1996) sampled 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.

Apart from Manchu, the Tungusic languages all have a small number of speakers whose populations were until recently hunter-gatherers. The Tungusic groups are quite likely the descendants of the LSA hunter-gatherers displaced by the rise of agriculture in North China. Surprisingly, however, the Tungusic languages are not highly diverse compared with other Siberian populations, suggesting that the Tungusic expansion is probably quite recent. However, what remains of Tungusic today may well not reflect its previous importance. Manchu was the language of the ruling class in China until recently but has almost disappeared (Svanberg 1988). It seems likely that Tungusic groups were spread more widely across northern Heilongjiang Province and the adjacent Mongol-speaking area, probably in interaction with Koreanic speakers. It is considered possible that Tungusic speakers were responsible for the introduction of the Northern Bronze Complex into the Korean peninsula during the 1<sup>st</sup> 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 Laos, Cambodia 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 6 shows the view of the internal relationships of Daic given by Edmondson & Solnit (1997b);

**Figure 6. Daic according to Edmondson & Solnit (1997)**



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, Miao-Yao and Sino-Tibetan. These were used by Benedict (1975) to erect 'Austro-Tai', a macrophylum that would unite Austroasiatic, Miao-Yao, 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 shown that the evidence for hypotheses, such as Benedict's Austro-Tai, linking together the major language phyla of SE Asia, derive from ancient loanwords. The recent proposal by Ostapirat (in press) to link Daic with Austronesian will no doubt be controversial, but it conforms better to the classic comparative method than any other on the table. Ostapirat assumes a simply model of a primary split with Daic being the Austronesians who stayed at home. But this seems unlikely. Daic looks like a branch of proto-Philippines and does not share in the complexities of Formosan. It may be better to think of proto-Daic speakers migrating back across from the northern Philippines to the region of Hainan island; hence the distinctiveness of Hlai and Be and Daic the result of radical restructuring following contact with Miao-Yao and Sinitic.

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 (2004) has presented some evidence for thinking that speakers of proto-Daic were not originally rice-cultivators, that they 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

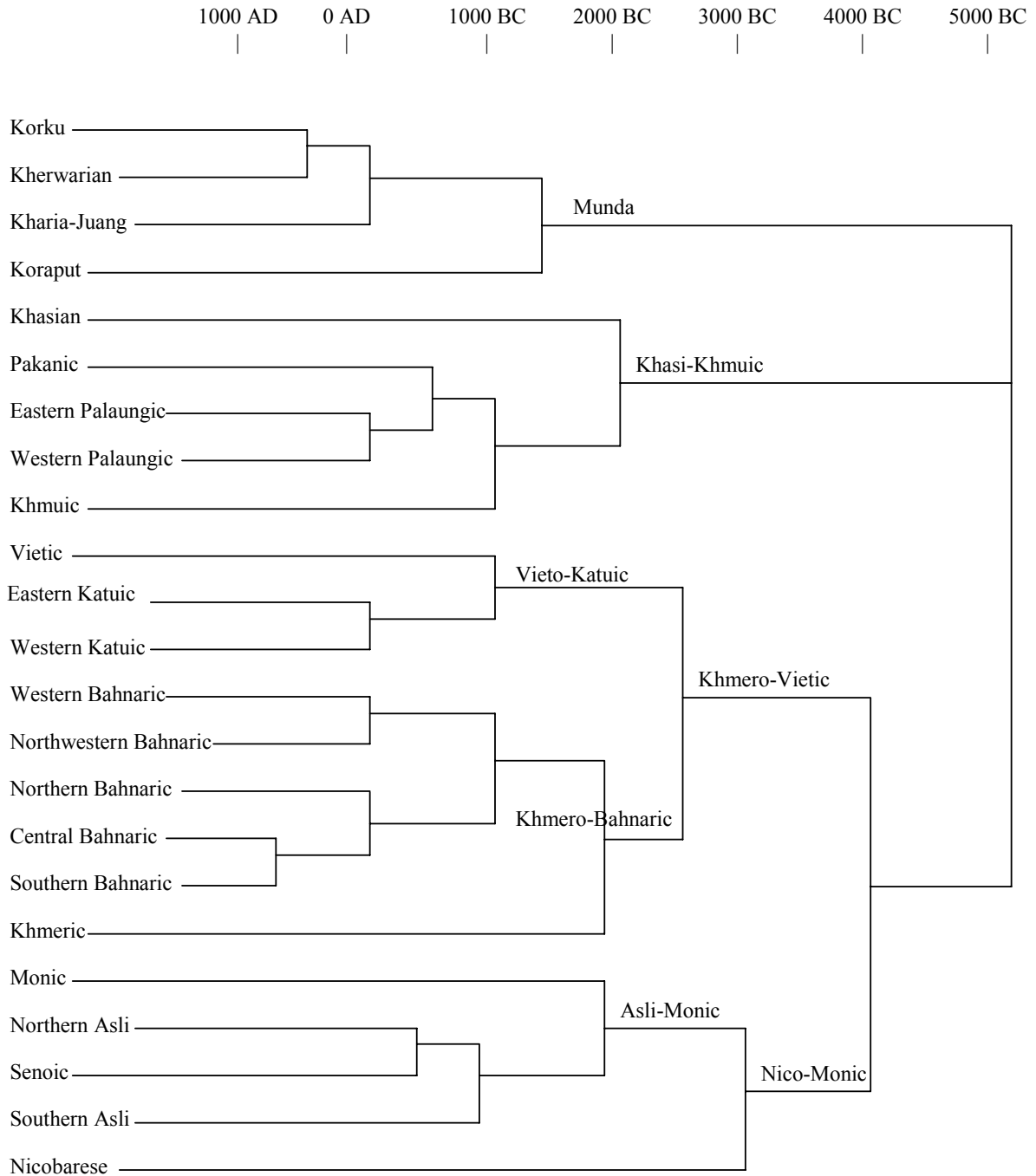
Austroasiatic languages are the most poorly researched of all those under discussion. Many are not documented at all and some recently discovered in China are effectively not classified. Parkin (1991) is a general anthropological overview of speakers of Austroasiatic languages as well as a comprehensive bibliography. The genetics of Austroasiatic speakers are almost unresearched. Austroasiatic is conventionally divided into two families, Mon-Khmer and Muṅḍā. Earlier classifications have elevated these last two to a primary branching of Austro-Asiatic, but no evidence for these realignments has been published. Indeed Austro-Asiatic 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 7 shows the most recent ‘tree’ of Austro-Asiatic;

The main branch of Austroasiatic in China is Palaungic, but there are other much less well-known languages. Apart from Palyu, a single branch language of Austroasiatic, there are also four unclassified Austroasiatic languages in China; Bagan, Buxinhua, Kemiehua and Kuanhua.

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’s 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, as the map shows; the spread of Austronesian, Sino-Tibetan and Daic in more recent times has isolated populations. The big question is where they began and how they spread outwards. Van Driem (2001) canvases a number of theories including the ‘northern shores of the Bay of Bengal’. Diffloth (in press) has claimed that faunal reconstructions support a tropical origin; 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 diversity and there is at least some evidence that Austroasiatic languages were once more widespread in China. What if this was the homeland area?

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 (Figure 7) although he canvases quite a different area of origin for Austroasiatic.

**Figure 7. Austroasiatic with calibrated time-depths according to Diffloth**



Contributions from genetics to the genesis of Austroasiatic are slight and not necessarily trustworthy. Roychoudhury *et al.* (2001) confidently assign improbably early dates to the Austroasiatic expansion based on a limited Muṅḍā sample;

Our data indicate Austro-Asiatic speakers underwent population expansion about 17,000 years prior to the Elamo-Dravidian speakers and about 5,000 years prior to the Tibeto-Burman speakers. The confidence intervals of the expansion times of Austro-Asiatic and Tibeto-Burman speakers are non-overlapping with those of the Dravidian speakers, while those of Austro-Asiatic and Tibeto-Burman speakers do overlap, indicating that the antiquity of expansion of the Austro-Asiatics is significantly greater than that of the Dravidians, but not of the Tibeto-Burman.

Roychoudhury *et al.* (2001)

‘Genomic structures and population histories of linguistically distinct tribal groups of India’

This appears to tell us that not only was the Austroasiatic expansion improbably early but was at about the same time as Sino-Tibetan.

## 2.7 Austronesian

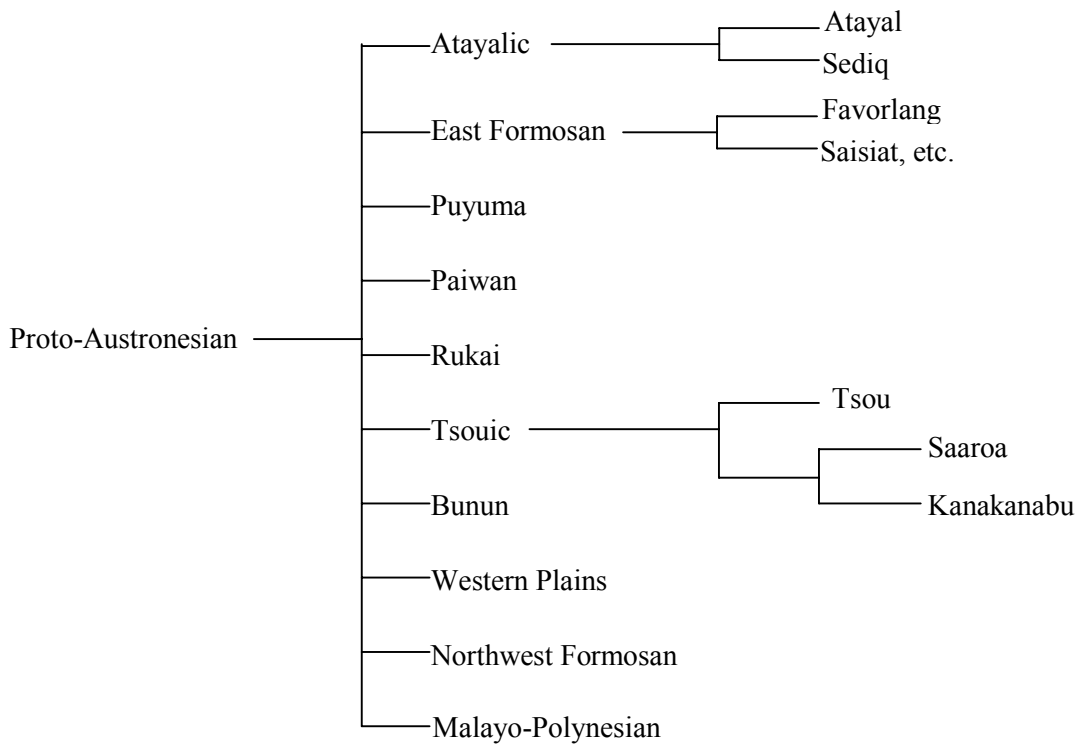
Austronesian is the second-largest language phylum in the world after Niger-Congo and certainly one of the most widespread, stretching from Easter Island to Madagascar. Compared with many of the other phyla in this region, its internal structure is relatively transparent and there are few doubts about 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 today on the very edge of its distribution and the sole Austronesian language, Tsat, spoken in China today is not 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 Roglai, a Chamic language found in Vietnam and the founders of the Utsat community probably fled to Hainan after break-up of the Cham Empire.

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. 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 (2004). 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. It does suggest a flow of migrants into Taiwan from the mainland across the strait from 7000 BP onwards. 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.

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 8 shows the top-level structure of Austronesian according to Blust (1999);

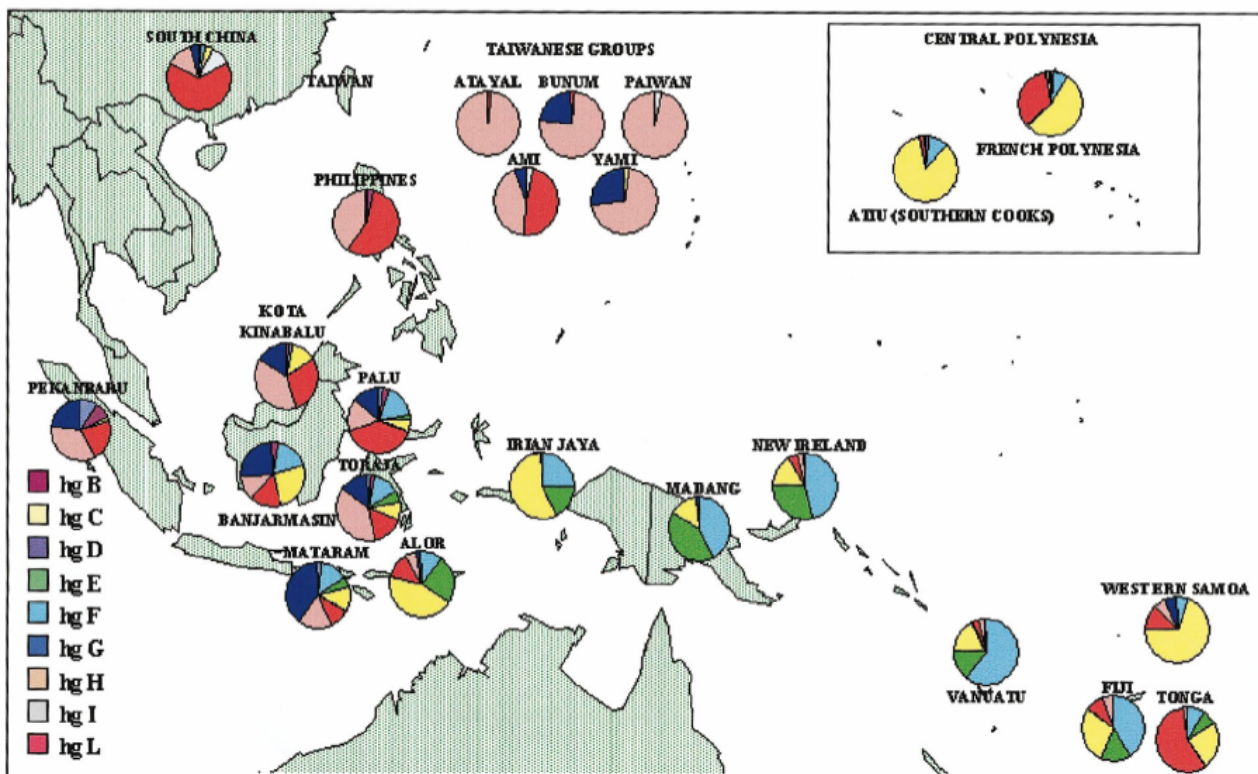


**Figure 8. Austronesian according to Blust (1999)**



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 Ami, 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 9 shows the geographic locations of populations analysed by Capelli *et al.* (2001) with the proportions of each haplogroup;

Figure 9. 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 you would expect these to be also present in the Philippines.

The tendency to mix ethnic groups, regions, countries and other incompatible units tends to make linguists highly uneasy. Nonetheless, the links between South China, Taiwan and the Philippines are quite striking as are variations between different Formosan groups. This strongly suggests that the settlement of Taiwan was a complex process, perhaps with inputs from different regions of the Chinese mainland, as well as back-settlement from 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 source of Tokharian, a language attested in manuscripts found in the T aklimakan desert. The linguistic features of Tokharian 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 can't 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 Sinitic<sup>3</sup> would be the simplest solution (Lubotsky 1998).

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
Chärchän	Präkrit, Koräanian	?	1000 BC
Lopnur	Präkrit, Koräanian	Proto-Europoid, Indo-Afghan	1800 BC
Qumul	? Tocharian A	Proto-Europoid	1000 BC
Turpan	? Tocharian A	Proto-Europoid	4-5 <sup>th</sup> 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 peninsula, dating to >4000 to >2000 BC. Fish and aquatic resources were apparently of 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.

## 2.10 Isolates?

Older accounts of the languages of China suggest that there are a number of isolates or languages difficult to classify, such as Bai and Tujia (Ramsey 1987). These 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. A language that is still puzzling is Waxianghua, spoken by 300,000 people in a 6000 km<sup>2</sup> 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. Neighbouring Han Chinese, Miao and Tujia people do not understand it. Some view it as a special variety of Chinese, others as a minority language, perhaps related to Miao.

## 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

<sup>3</sup> The much cited example of 'honey', Old Chinese \*mit < Tocharian B mit.

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 with 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.'

Kisivild *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.

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.

A more assimilable scenario is that exemplified in Bo Wen *et al.* (2004) which looks at sex-biased admixture in 'Southern Tibeto-Burmans' (Bai, Lolo-Burmese, Tujia 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 model for the process of Sinicisation and in particular it can be mapped against the deep influence of Sinitic on Bai and Tujia. Nonetheless, it is unclear what social and migratory process this reflects; perhaps the movement of soldiers or possible seasonal hunters of cultivators.

So far limited work on extraction of ancient DNA in China: Wang et al. (2000) an important exception. This paper compares mtDNA from ancient Chinese populations with those of today. It purports to show that the people of Linzi in Shandong province were closer to the Welsh than to modern-day Chinese and therefore that there has been significant population replacement. This has been comprehensively discredited (Yao et al. 2003).

#### 4. The peopling of China

As the paper suggests many questions about the dating and spreading the language phyla of China remain in doubt and therefore and 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 highly ethnolinguistically diverse but would have consisted of Tungusic-Koreanic speakers in the North, Miao-Yao in the centre, intertwined with a wide range of diverse 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 phyllic diversity seen in Siberia. In the far northwest, where Chinese expansion is more recent, there would have been at least two different resident IE groups.

##### **When and from which direction was the Sinitic expansion?**

This expansion would surely have been from north to south, from millet cultivating to the humid areas where irrigated rice was possible.

##### **What populations came after the Chinese?**

The Turkic speakers in the Xinjiang region represent a late incursion. Pre-Mongolic speakers would have perhaps made incursions on the settled villages in northern China as nomadic pastoralism developed. The expansion of Daic would have roughly coincided with the expansion of Sinitic.

##### **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?**

<b>Phylum</b>		<b>Date BP</b>
Early Tibetan	Sino- ? 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]	?
Miao-Yao	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. In its present form, Miao-Yao, for example, cannot be very old because the Miao-Yao languages are closely related. But 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 they reflect an older stratum of dispersal.

## 5. Where next?

The paper sets 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 at 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.

Part of this is an unfortunate fragmentation of scholarship; Austronesian studies is a model of collaboration that could well be emulated.

## 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.
- 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 1990. *Japanese/Austro-Tai* (Linguistica Extranea Studia 20). Ann Arbor: Karoma Publishers.
- 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.
- Bowcock, A., Ruiz-Linares, A., Tomfohrde, J., Minch, E., Kidd, J. R. & Cavalli-Sforza, L. L. (1994) *Nature (London)* **36**, 455–457.

- 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.
- 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.
- Cavalli-Sforza, L.L. 1998. *Trends Genet.* **14**, 60–65.
- 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, G. 1997. Austroasiatic. In: *Encyclopaedia Britannica*. 701-704. Chicago: Encyclopaedia Britannica.
- 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. *Proc Natl Acad Sci 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 1995. Black Mountain conjugational morphology, Proto-Tibeto-Burman morphosyntax, and the linguistic position of Chinese. In: *New Horizons in Tibeto-Burman Morphosyntax* (Senri Ethnological Studies 41). Yoshio Nishi, James Alan Matisoff and Yasuhiko Nagano eds. 229-259. Ōsaka: National Museum of Ethnology.
- 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 1999. A new theory on the origin of Chinese. in Peter Bellwood and Ian Lilley (eds) *Bulletin of the Indo-Pacific Prehistory Association Bulletin*, 18 (*Indo-Pacific Prehistory: The Melaka Papers, Vol. 2*), 43-58. Canberra: Australian National University.
- 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.
- Du, R., Yuan, Y., Hwang, J., Mountain, J. & Cavalli-Sforza, L. L. 1992. *J. Chinese Linguistics, Monograph Series No. 5*.
- 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.
- Ferrell, R. 1969. *Taiwan aboriginal groups: problems in cultural and linguistic classification*. Taipei: Institute of Ethnology, Academia Sinica, Monograph series, no. 17.
- Fitzgerald, C.P. 1972. *The Southern expansion of the Chinese people: southern fields and southern oceans*. London: Barrie & Jenkins.
- 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.
- Guo Z, Zheng SP, Sugimoto C, Wang YL, Zheng H-Y, Takasaka T, Kitamura T, Guo J, Yogo Y. 2001. JC Virus genotypes in northwestern China: implications for its population history. *Anthropological Science*, 109:203–212.

- 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.
- Johanson, L. and E. A. Csato eds. 1998. *The Turkic Languages*. London: Routledge.
- Jorde, L. B., Bamshad, M. J., Watkins, W. S., Zenger, R., Fraley, A. E., Krakowiak, P. A., Carpenter, K. D., Soodyall, H., Jenkins, T., Rogers, A. R., et al. (1995) *Am. J. Hum. Genet.* **57**, 523–538.
- 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. *Mol Biol Evol* 19:1737–1751.
- Kolman, Connie J., Nyamkhishig Sambuughin & Eldredge Bermingham 1996. Mitochondrial DNA Analysis of Mongolian Populations and Implications for the Origin of New World Founders. *Genetics*, 142: 1321-1334.
- Krings, M., Stone, A., Schmitz, R. W., Krainitzki, H., Stoneking, M. & Paabo, S. 1997. xxx *Cell* **90**, 19–30.
- 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.
- Macer, D., Fleming, J., Keyeux, G. & Knoppers, B. M. 1996. *Nature (London)* **379**, 11.
- 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. & D.Q. Adams eds. 1997. *Encyclopaedia of Indo-European culture*. London & Chicago: Fitzroy Dearborn.
- 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.
- Masica, C. 1990. *The Indo-Aryan languages*. Cambridge: Cambridge University Press.
- Matisoff, J.A., S.P. Barton & J.B. Lowe 1996. *Languages and dialects of Tibeto-Burman*. STEDT Monograph Series, 2. Berkeley: University of California 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.
- 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. *Am J Hum Genet*, 63:1807–1823
- Menges, K.H. 1995. [2nd ed.] *The Turkic peoples and languages*. Wiesbaden: Harrassowitz.
- Miller, R.A. 1994. Old loanwords in Japanese and ‘Omnicomparativismus’. *Journal de la Société Finno-Ougrienne*, 85:221-236. Helsinki.
- 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.
- Nelson, Sarah M. 1995. *The archaeology of Northeast China: beyond the Great Wall*. London and New York: Routledge.
- Niederer, Barbara 1998. *Les langues Hmong–Mjen (Miáo–Yáo)—Phonologie Historique*. Lincom Studies in Asian Linguistics. 07. München: Lincom Europa.
- Oota, H. et al. 2002. Extreme mtDNA homogeneity in continental Asian populations. *Am J Phys Anthropol*, 118:146-153.
- Parkin, R. 1991. *A guide to Austroasiatic speakers and their languages*. Honolulu: University of Hawaii Press.



- 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.
- Qian Y, Qian B, Su B, Yu J, Ke Y, Chu Z, Shi L, Lu D, Chu J, Jin L 2000. Multiple origins of Tibetan Y chromosomes. *Human Genetics*, 106:453–454
- Ramsey, S.R. 1987. *The Languages of China*. Princeton, NJ: Princeton University Press.
- Reid, Lawrence A. 1996. The current state of linguistic research on the relatedness of the language families of East and South-East Asia. *Bulletin of the Indo-Pacific Prehistory Association*, 15:87-91.
- Roberts, J.A.G. 1996. *A history of China. Volume I. Prehistory to c. 1800*. Phoenix Mill: Alan Sutton Publishing.
- Sagart, Laurent 1994. Proto-Austronesian and Old Chinese evidence for Sino-Austronesian. *Oceanic Linguistics*, 33(2):271-308.
- Schanfield, M.S. & Gershowitz, J. 1973. xxx *Am. J. Hum. Genet.* **25**, 567–574.
- Seielstad, M., Minch, E. & Cavalli-Sforza, L.L. (1998) xxx. *Nat. Genet.*, in press.
- Shafer, Robert 1974. *Introduction to Sino-Tibetan*. Wiesbaden: Otto Harrassowitz.
- Su B, Xiao C, Deka R, Seielstad MT, 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
- Su B, Xiao J, Underhill P, Deka R, Zhang W, Akey J, Huang W, Shen D, Lu D, Chu J, Tan J, Shen P, Davis R, Cavalli-Sforza LL, Chakraborty R, Xiong M, Du R, Oefner P, Chen Z, Jin L (1999) Y-chromosome evidence for a northward migration of modern humans into Eastern Asia during the last Ice Age. *Am J Hum Genet* 65:1718–1724.
- Svanberg, Ingvar 1987. The Loplyks: a vanishing fishing and gathering culture in Xinjiang. *Svenska Forskningsinstitutet i Istanbul*, 12:57-81.
- Svanberg, Ingvar 1988. *The Altaic speakers of China: numbers and distribution*. Uppsala: Uppsala University.
- Thurgood, Graham. 1985. Pronouns, verbal agreement systems, and the subgrouping of Tibeto-Burman. pp. 376-400 in *Linguistics of the Sino-Tibetan Area: the State of the Art — Papers presented to Paul K. Benedict for his 71st birthday*. eds. Graham Thurgood, James Alan Matisoff and David Bradley. Pacific Linguistics (Series C, No. 87), Canberra.
- 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.
- Tsang, C.H. 1992. *Archaeology of the Peng-hu Islands*. Special Publications, No. 95. Taipei: Institute of History and Philology, Academia Sinica,
- Tsang, C.H. 2004. 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*. 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, William, S-Y. ed. 1991. Languages and dialects of China. *Journal of Chinese Linguistics*, Monograph 3.
- 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.
- Whitfields, Roderick & Wang Tao. eds. 1999. *Exploring China's past*. London: Saffron Books.
- Wolpoff, M. H. 1989. xxx In *The Human Revolution*. eds. Mellars, P. & Stringer, C. 62–108. Edinburgh: Edinburgh Univ. Press.
- Yao Y-G, Kong Q-P, Man X-Y, Bandelt H-J, Zhang Y-P 2003. Reconstructing the evolutionary history of China: a caveat about inferences drawn from ancient DNA. *Mol Biol Evol* 20:214–219.
- Yao Y-G, Zhang Y-P. 2002. Phylogeographic analysis of mtDNA variation in four ethnic populations from Yunnan Province: new data and a reappraisal. *J Hum Genet* 47:311–318.

- Yao Y-G, Kong Q-P, Bandelt H-J, Kivisild T, Zhang Y-P 2002. Phylogeographic differentiation of mitochondrial DNA in Han Chinese. *Am J Hum Genet* 70:635–651
- Yao Y-G, Nie L, Harpending H, Fu Y-X, Yuan Z-G, Zhang Y-P. 2002. Genetic relationship of Chinese ethnic populations revealed by mtDNA sequence diversity. *Am J Phys Anthropology*, 118:63–76.
- 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, C. (1996) 'The rise of urbanism in the middle and lower Yangtze river valley', in Peter Bellwood and Diane Tillotson (eds) *Indo-Pacific Prehistory: The Chiang-Mai Papers* (Proceedings of the 15th Congress of the Indo-Pacific Prehistory Association, Chiang Mai, Thailand, 5 to 12 January 1994), pp. 63-68, Vol. 3 Canberra: Australian National University.
- Zhao TM, Lee TD (1989) *Gm* and *Km* allotypes in 74 Chinese populations: a hypothesis of the origin of the Chinese nation. *Hum Genet* 83:101–110
- 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.