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

- Roger Blench
- Mallam Dendo Ltd.

Linguistics, genetics and archaeology I

- 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 and crypto-Marxist schemas.

Linguistics, genetics and archaeology II

- **In addition, the ideology surrounding the definition of minorities in China has confused the analysis in genetics papers**
- **Major archaeological texts refer neither to linguistics nor genetics**
- **Situation has begun to change: a review of the current situation is useful**

Linguistics, genetics and archaeology III

- **So far limited work on extraction of ancient DNA in China: Wang et al. (2000) an important exception and even this has been queried**

- **Dates in genetics papers best described as ‘wild’; limited effort to reconcile these with linguistic or archaeological data**

An appreciation of the skills of James Matisoff

For one who has abandoned craving and is free from grasping,
who is skilled in etymology and terms, knowing the groupings and sequences of letters,
this is the final birth.
This one is called
the Great Being, the Great Sage.

Dhammapada (24.19)

Which 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 etc.)
- + Korean, Russian etc.
- Unclassified

WAXIANGHUA


INDO-EUROPEAN
TURKIC
MONGOLIC
AINU
TUNGUSIC
KOREANIC
JAPONIC
SINO-TIBETAN
(=Tibeto-Burman)
Tibeto-Burman
INDO-EUROPEAN
HMONG-MIEN
=Miao-Yao
AUSTROASIATIC
KRA-DAI
DAI
AUSTRONESIAN
AUSTROASIATIC
AUSTRONESIAN
DRAVIDIAN
AUSTRONESIAN
PAPUAN

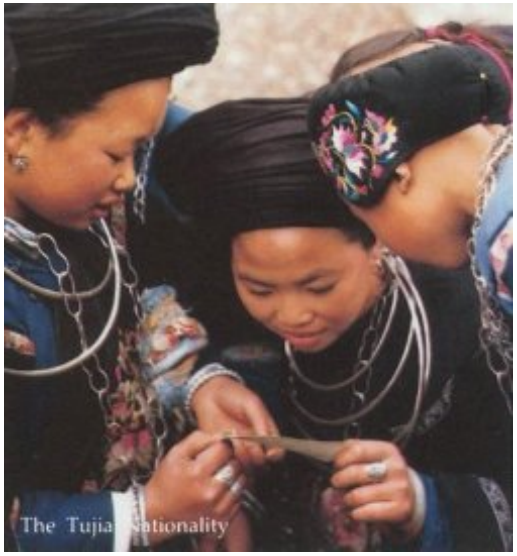

Language phyla of China
SINO-TIBETAN
Mandarin
1. Northern
2. Eastern
3. Southwestern
Southern
1. Wu
2. Gan
3. Xiang
4. Min
5. Hakka
6. Yue
Tibetan
1. Amdo
2. Khams
3. Dbusgtsang
Kam-Tai
Miao-Yao
INDO-EUROPEAN
Tajik
AUSTRO-ASIATIC
Mon-Khmer
ALTAI
Turkic
1. Kazakh
2. Uygur
3. Kirghiz
Mongolian
Manchu-Tungus
Korean
Heilongjiang
Jilin
Liaoning
Nei Mongol
Xinjiang
Beijing
BEIJING
Tianjin
Hebei
Ningxia
Shanxi
Shandong
Gansu
Shaanxi
Henan
Jiangsu
Shanghai
Anhui
Hubei
Zhejiang
Sichuan
Jiangxi
Hunan
Fujian
Guizhou
Taiwan
Yunnan
Guangxi
Guangdong
Hong Kong (U.K.)
Macau (Port.)
Hainan
Uninhabited
Qinghai
Xizang
Indian claim
Province-level boundary
Boundary representation is not necessarily authoritative.
0 500 Kilometers
0 500 Miles
WAXIANGHUA

How many languages are there in China?

- Ethnologue estimates around 200, but this includes 13 dialects of Chinese. However, new languages are regularly being recorded, such as the Austroasiatic Bugan, yet to be classified within Mon-Khmer. This contrasts with the official count of 56 [55 + Han].
- Despite the unlikely nature of this figure in continues to be propagated on websites and official documents

Chinese minorities, tourist version I

Chinese minorities, tourist version II



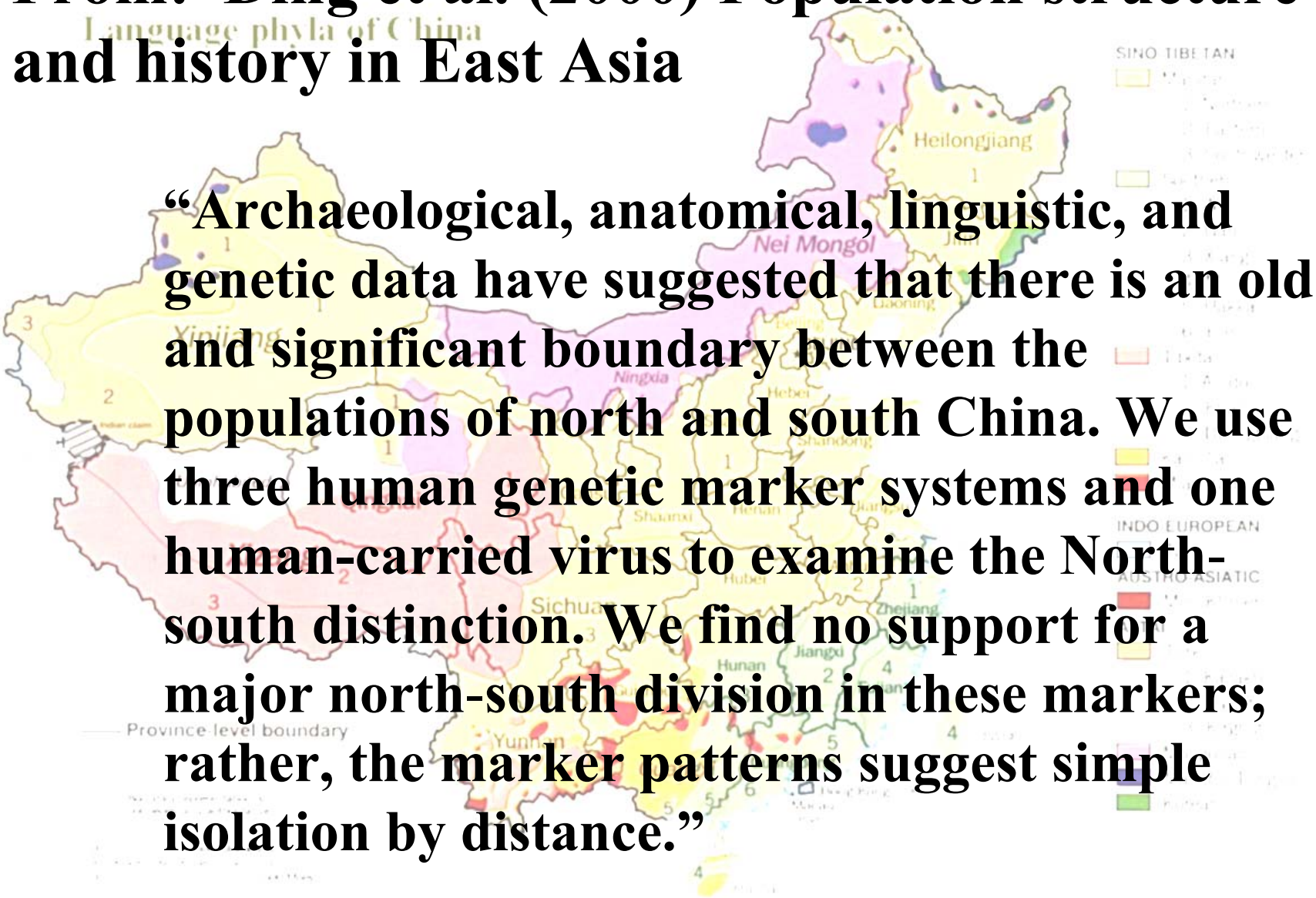
How many languages *were* there in China?

- 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?

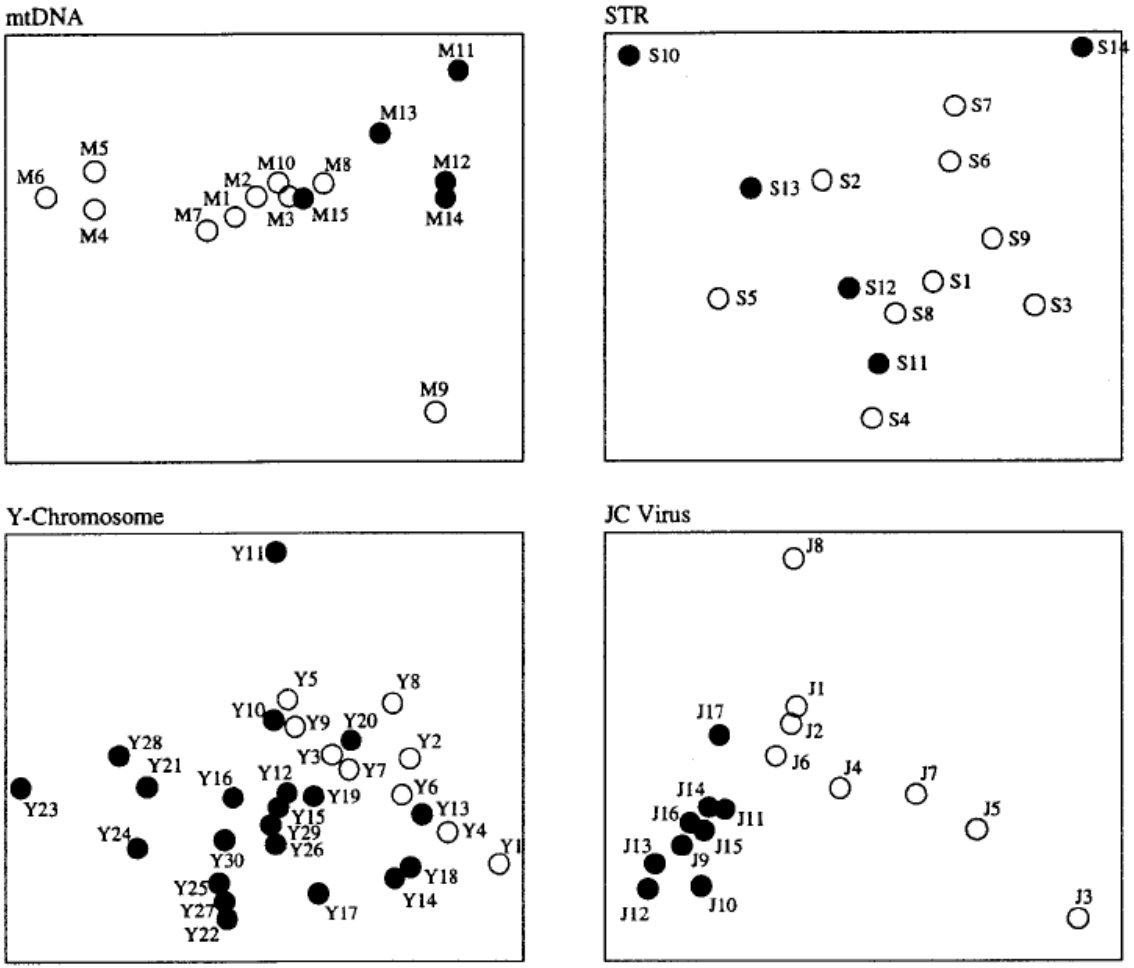
- What populations underlay the Han Chinese?
- When and from which direction was the Chinese expansion?
- What populations came after the Chinese?
- What drove the expansion of different phyla?
- What are the archaeological and genetic correlates of these phylic expansions?

From: Ding et al. (2000) Population structure and history in East Asia



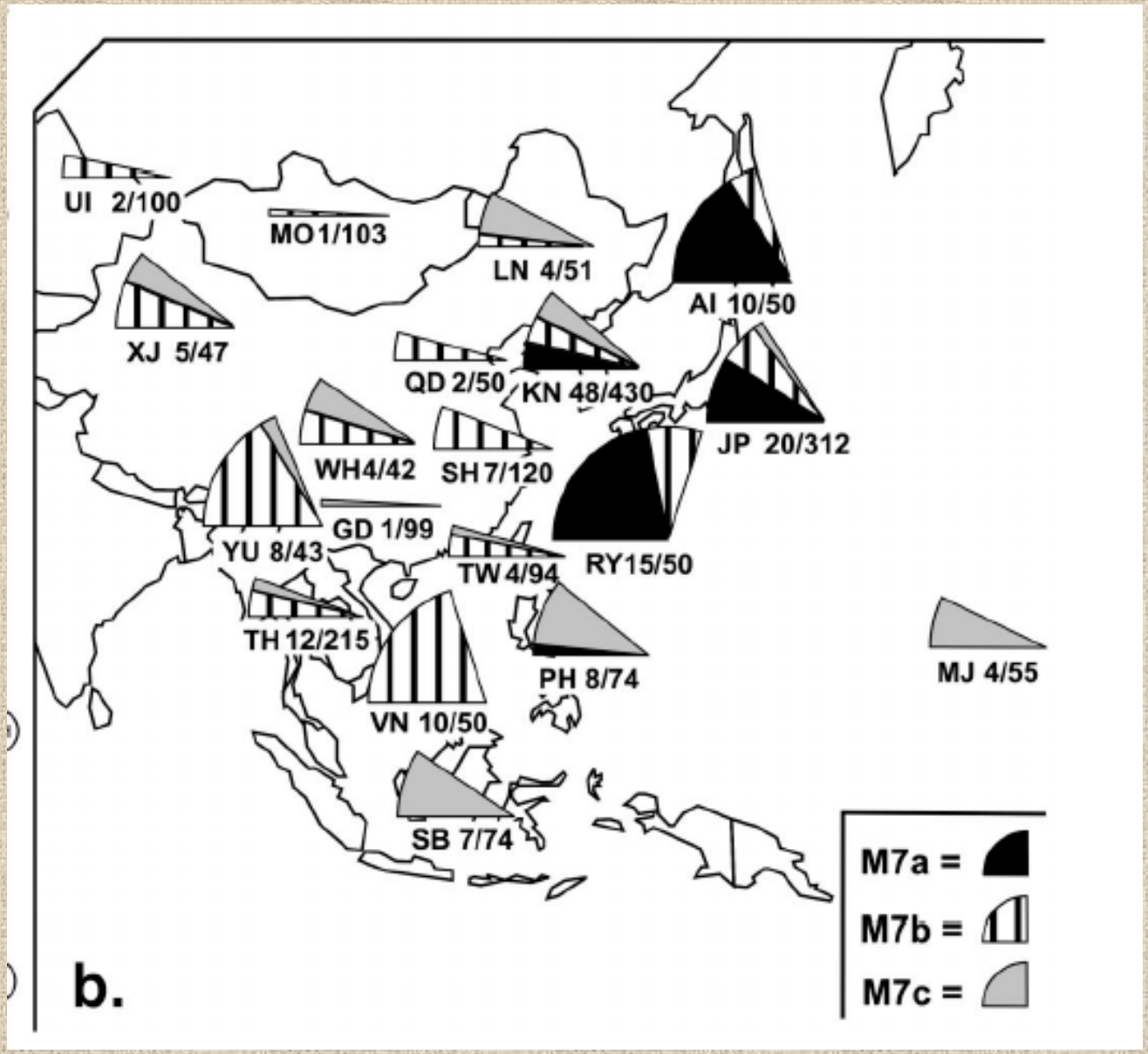
"Archaeological, anatomical, linguistic, and genetic data have suggested that there is an old and significant boundary between the populations of north and south China. We use three human genetic marker systems and one human-carried virus to examine the North-south distinction. We find no support for a major north-south division in these markers; rather, the marker patterns suggest simple isolation by distance."

From: Ding et al. (2000) Population structure and history in East Asia



Principal components maps. For each map, the *x* axis is the first and the *y* axis the second principal component. Northern populations are indicated by open and southern populations by closed circles.

Kivisild et al. 2003 'The Emerging Limbs and Twigs of the East Asian mtDNA Tree': distribution of the M7 haplogroup



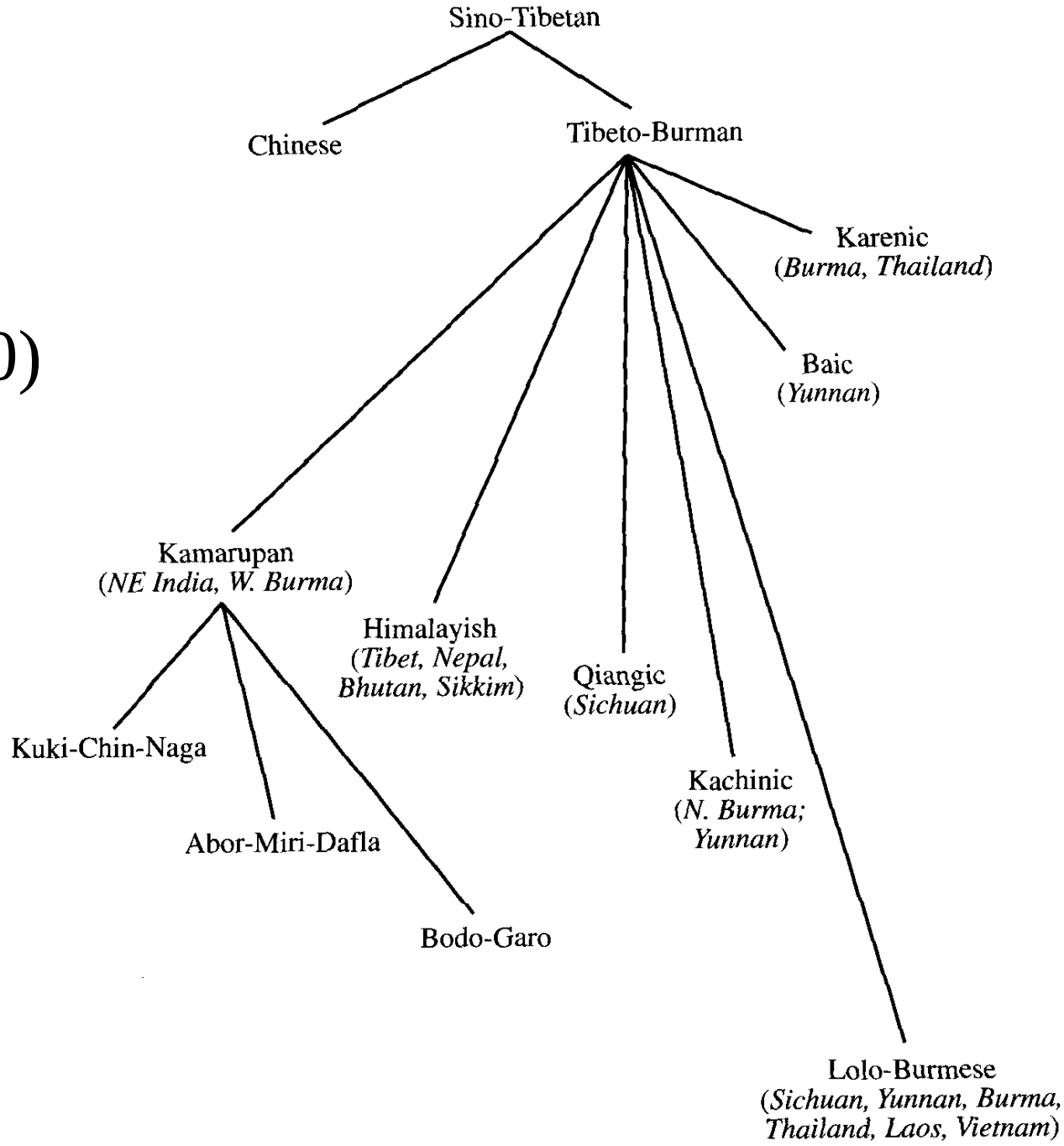
Kivisild et al. 2003 'The Emerging Limbs and Twigs of the East Asian mtDNA Tree': distribution of the M7 haplogroup

Kisivild confirms 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.

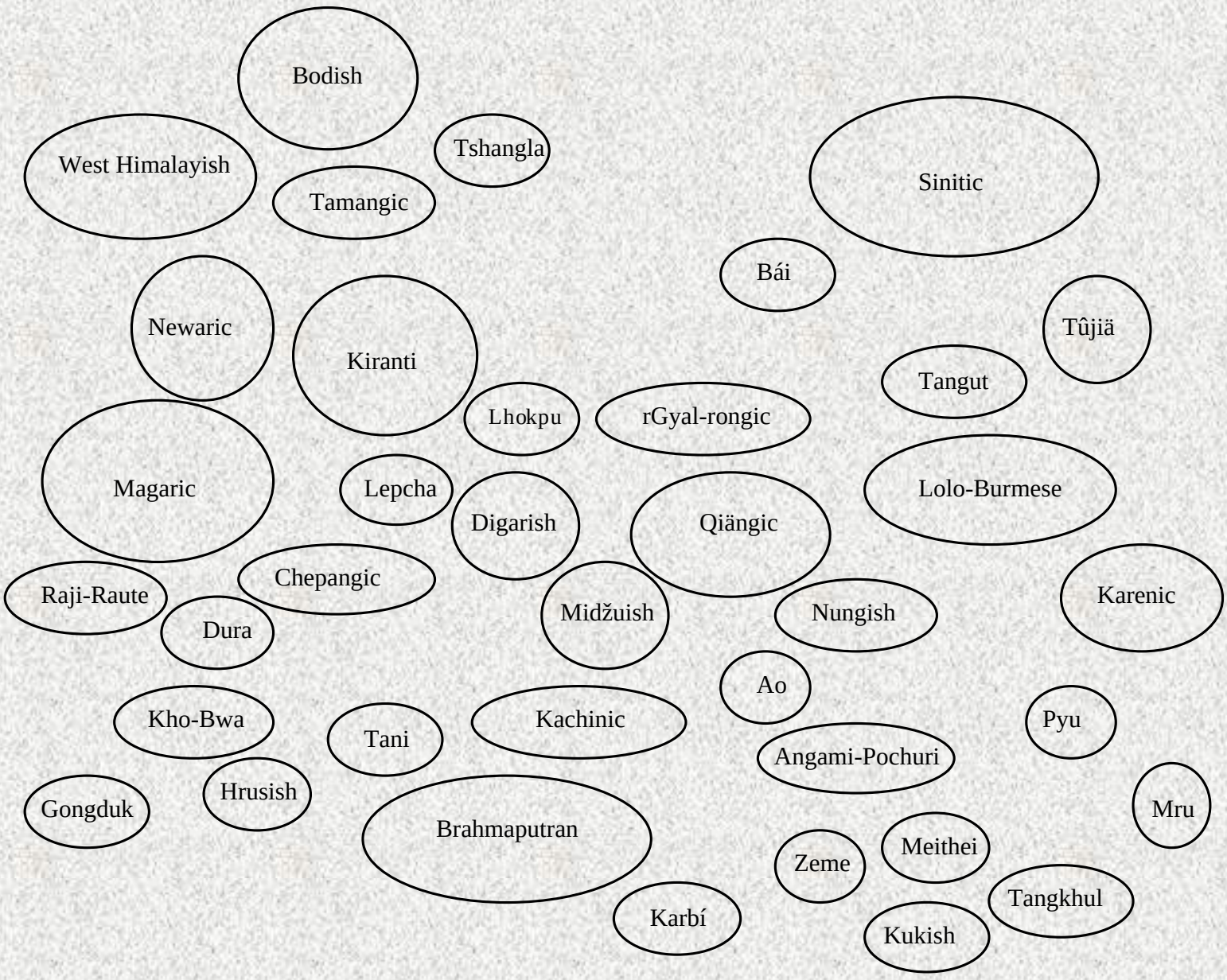
Vlq r OW le hwd q

Sino-Tibetan languages, from Van Driem (2001)

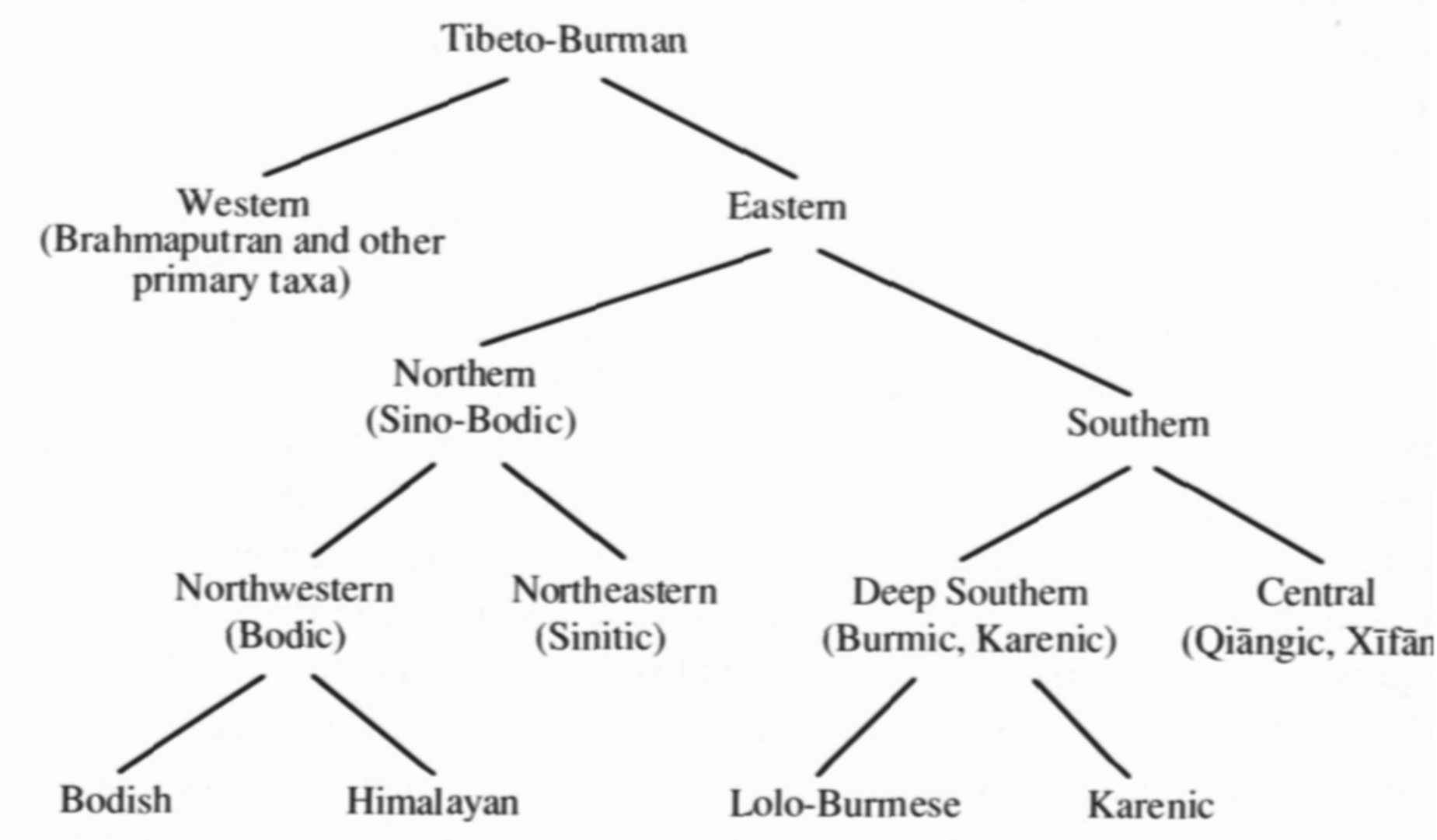
Sino-Tibetan according to Matisoff (2000)



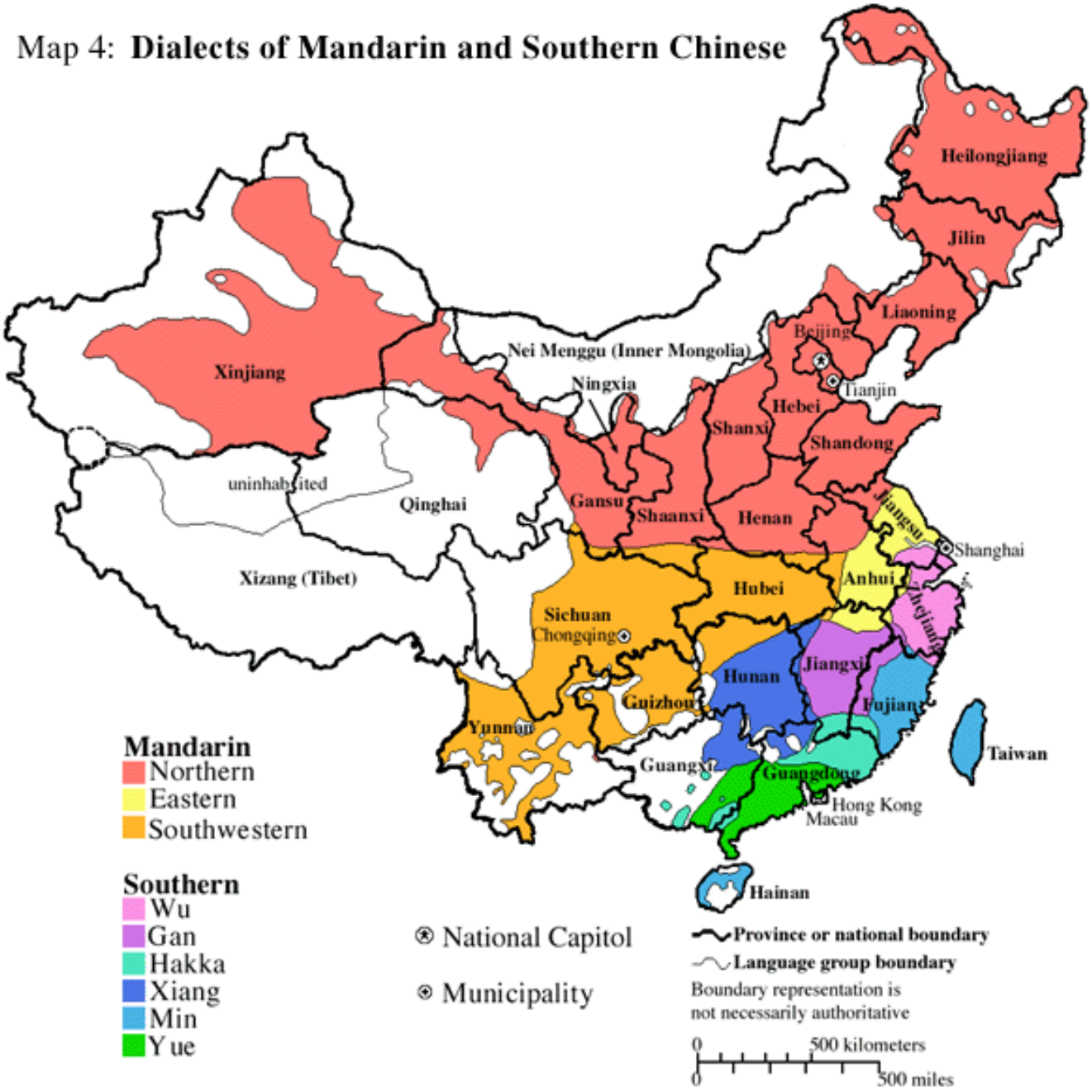
‘Fallen leaves’ model of Sino-Tibetan according to Van Driem (many places)



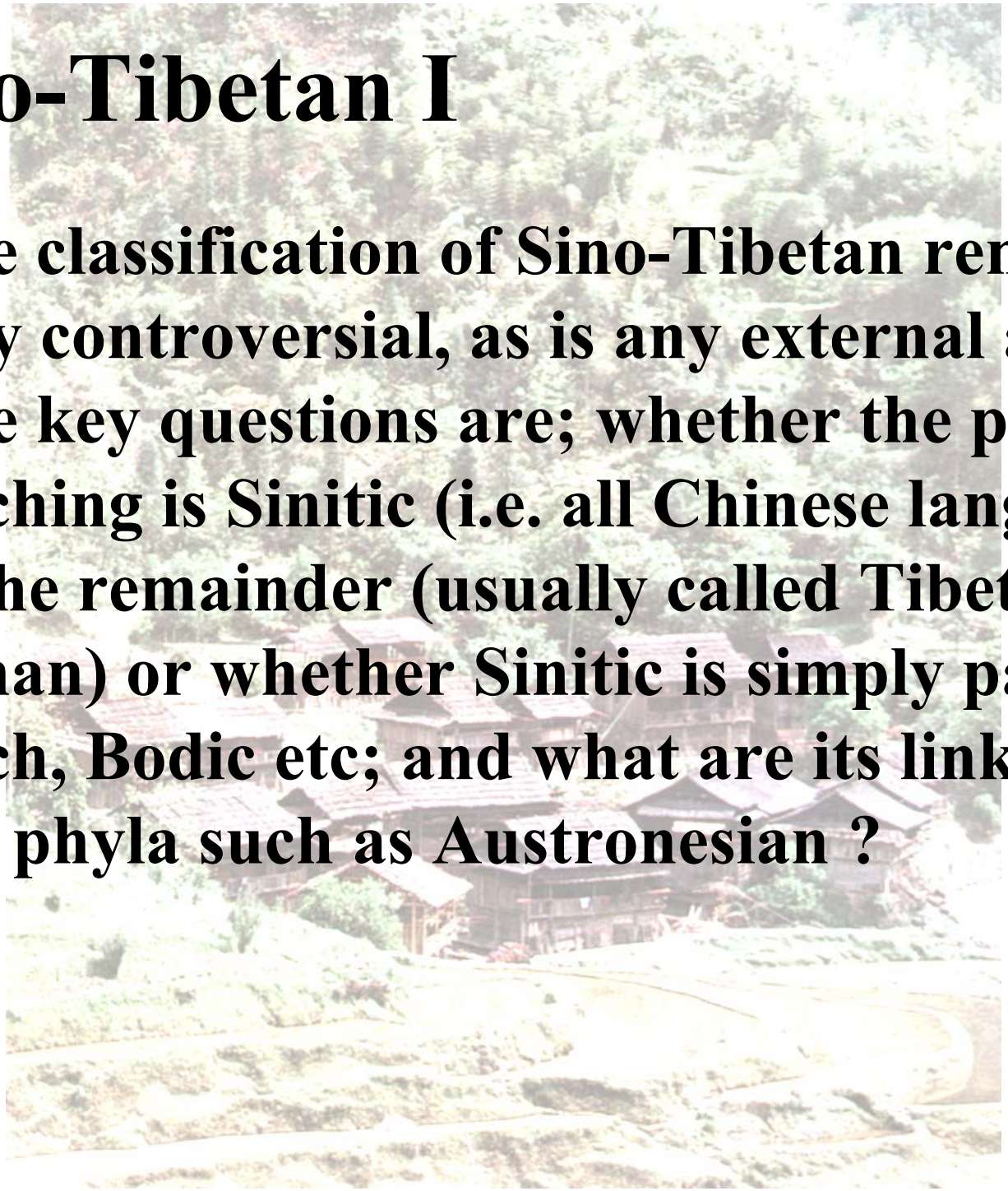
Tibeto-Burman [=Sino-Tibetan] according to Van Driem (2001)



Map 4: **Dialects of Mandarin and Southern Chinese**



Sino-Tibetan I



- The classification of Sino-Tibetan remains highly controversial, as is any external affiliation.
- The key questions 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, Bodic etc; and what are its links with other phyla such as Austronesian ?

Sino-Tibetan II

- Important distinction between Sinitic and Chinese. Sinitic is the language we arrive at by the reconstruction of modern living languages. Chinese is the totally attested in written texts.
- Chinese goes back to 1300 BC, possibly earlier. Sinitic only goes back to ca. 0 AD. There was therefore a ‘bottleneck’ or levelling at this time which eliminated a more diverse Sinitic. So actually dating the origin of Chinese is extremely difficult.

Sino-Tibetan III

- **Sinitic is not very diverse when compared with the rest of Sino-Tibetan. Even given the caveats just expressed it must be relatively recent. Presumably one among many competing ethnolinguistic groups that gained the advantage.**
- **Wherever it 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**

Sino-Tibetan IV

- China 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
- But it is difficult not to reach the conclusion that a major element in the *in situ* population was pre-Miao-Yao

Sino-Tibetan V

- **Sinitic is therefore *unlikely* to be identified with the earliest Neolithic communities in North China such as the Péilígăng or Císhān (6500 onwards). Perhaps these were Altaic speaking?**
- **The other side of the coin are the problematic 'remnant' languages of the Himalaya, Gongduk, Magaric etc. Either these are early branchings from the Sino-Tibetan tree or they are Kusundic, remnants of earlier language phyla that have been Sino-Tibetanised.**

So…

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 among them 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 peninsular, which is clearly associated with an alternation between land-mammal hunting and exploitation of marine resources. This might mean that the search for a ‘homeland’ is a chimaera.

The STEDT Gospel according to St. James

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 B.C.

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.

Locations of early archaeological cultures in China

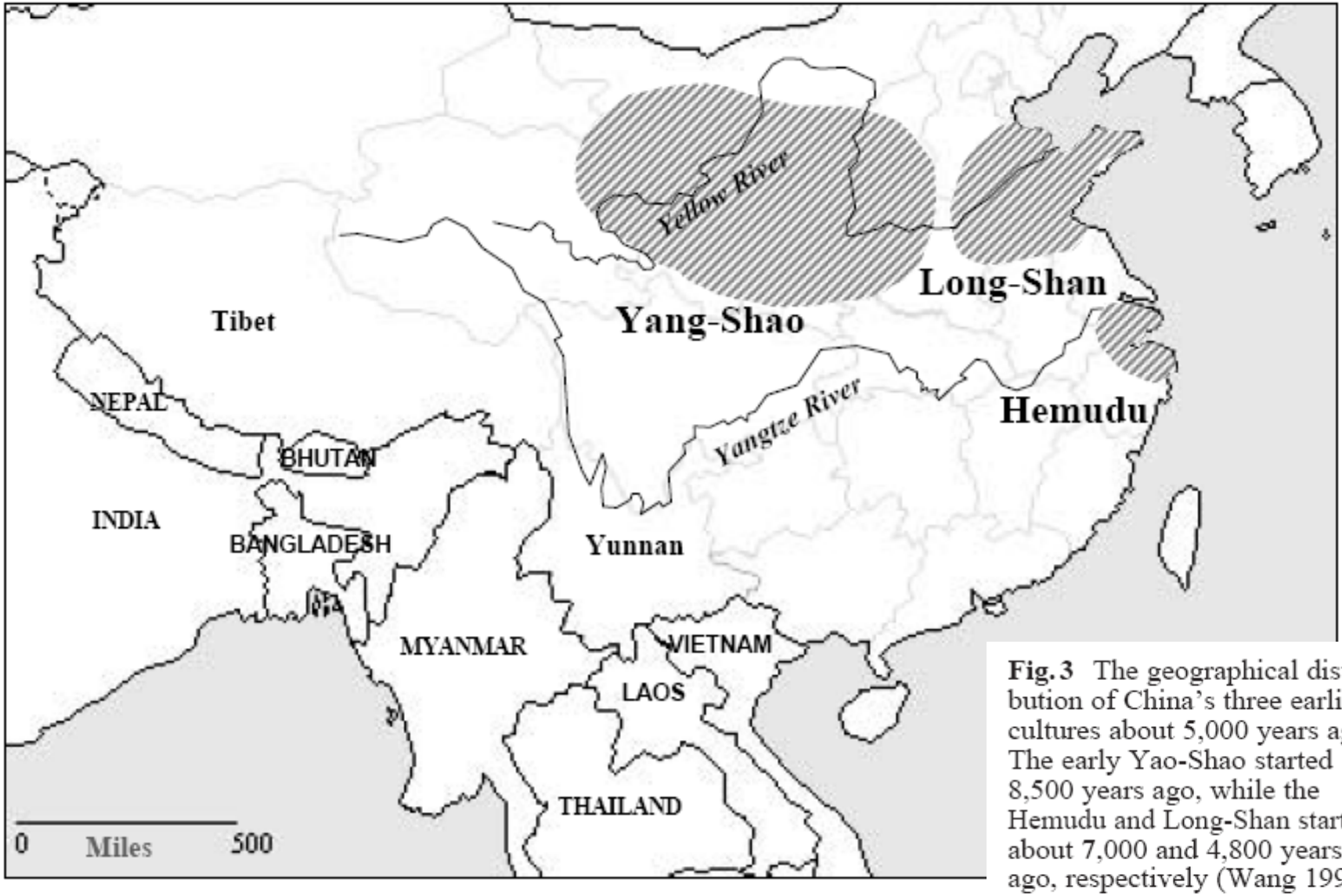
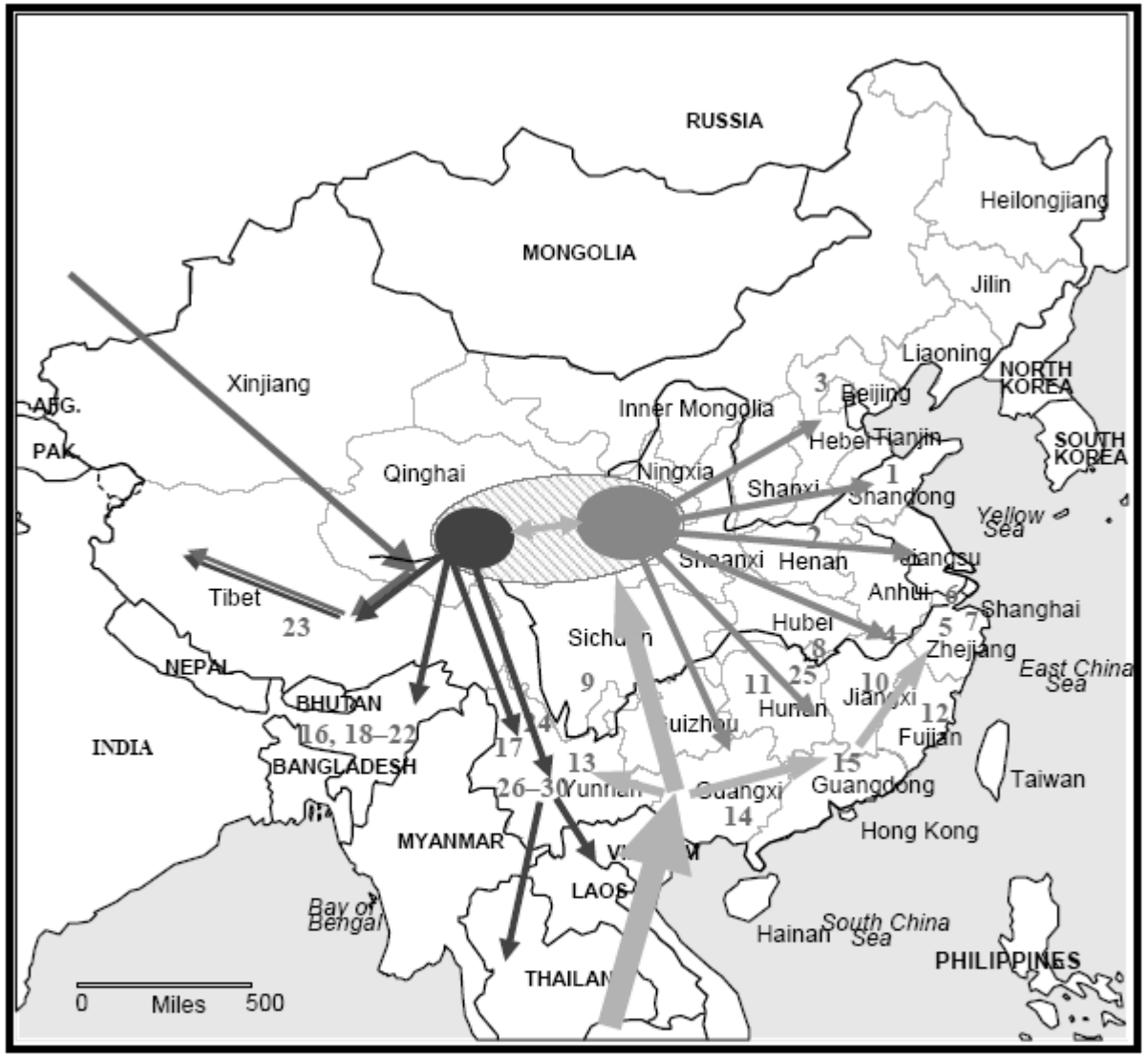


Fig. 3 The geographical distribution of China's three earliest cultures about 5,000 years ago. The early Yao-Shao started 8,500 years ago, while the Hemudu and Long-Shan started about 7,000 and 4,800 years ago, respectively (Wang 1994; Cavalli-Sforza and Piazza 1994)

Dispersal of Sino-Tibetan according to Su et al. (2000)

Fig. 4 The map of putative migration routes of the Sino-Tibetan populations. The *numbers* indicate the geographic locations of the Sino-Tibetan populations corresponding to the population numbers in Table 1



Peopling of China according to Chu et al. (1998)

P ILD R 0\ D R

Miao-Yao I

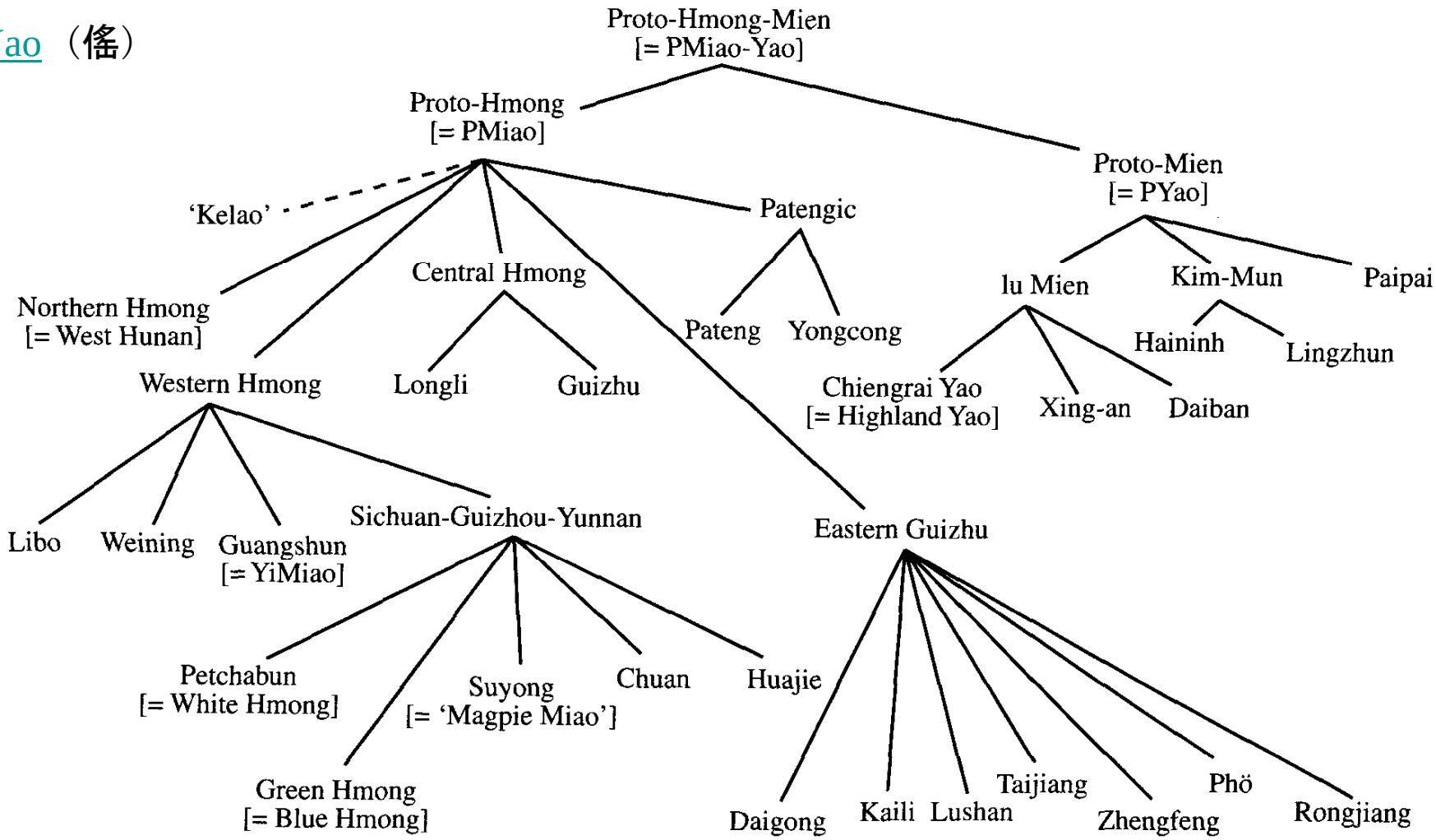
- The Miao-Yao [=Hmong-Mien] languages are spoken mostly in China with some groups also in Laos, Vietnam and Thailand. Their centre of gravity is between the Yangzi and the Mekong rivers. Miao-Yao languages are quite close to one another, and although the Ethnologue lists some 32 languages, many of these are mutually intelligible lects.

Miao-Yao [=Hmong-Mien] according to Matisoff (2000)

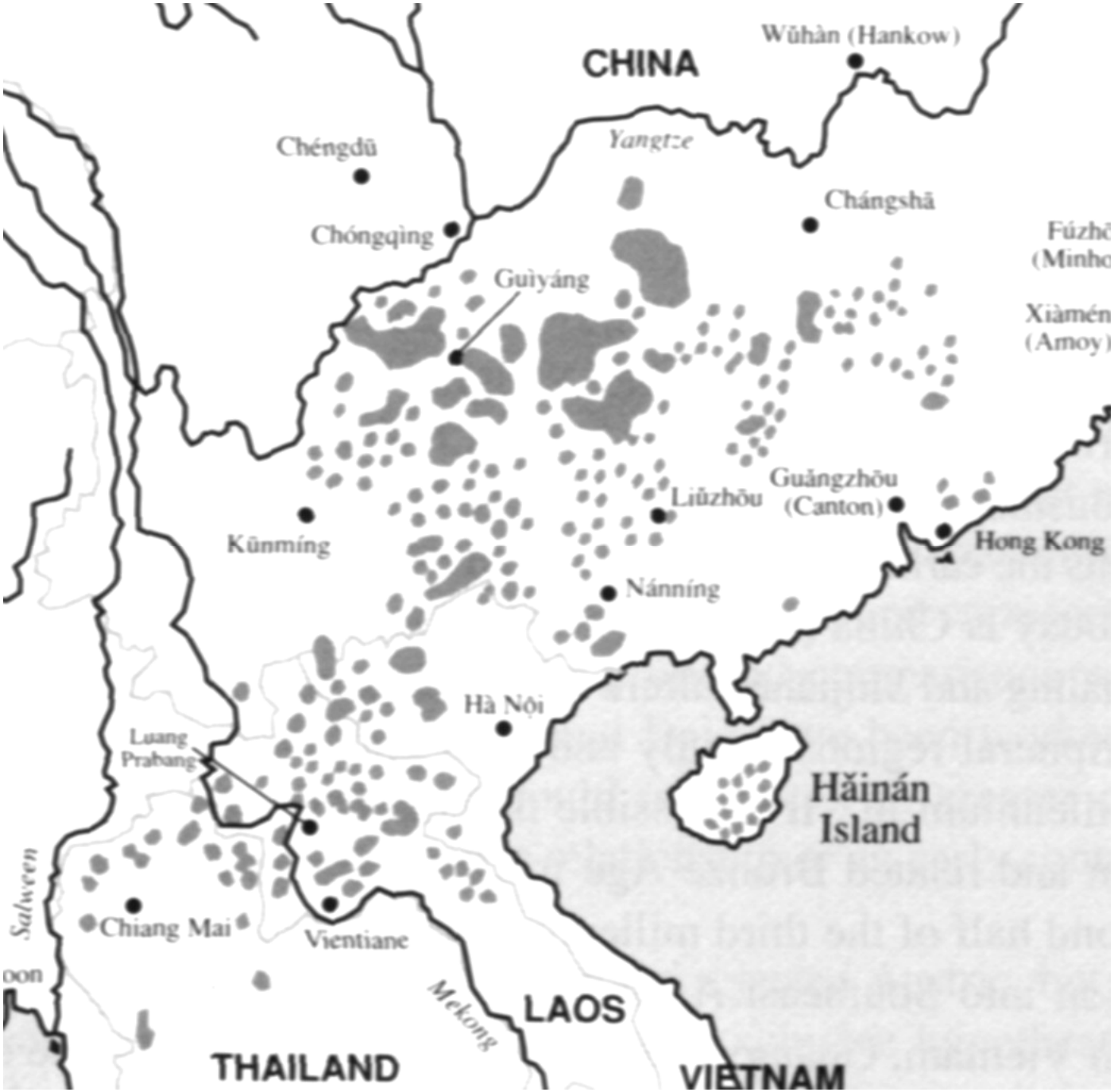
苗族 The *Miao* (**苗**)

+ She [=Ho Nte]

Yao (**傜**)



Miao-Yao languages, from Van Driem (2001)



Miao-Yao II

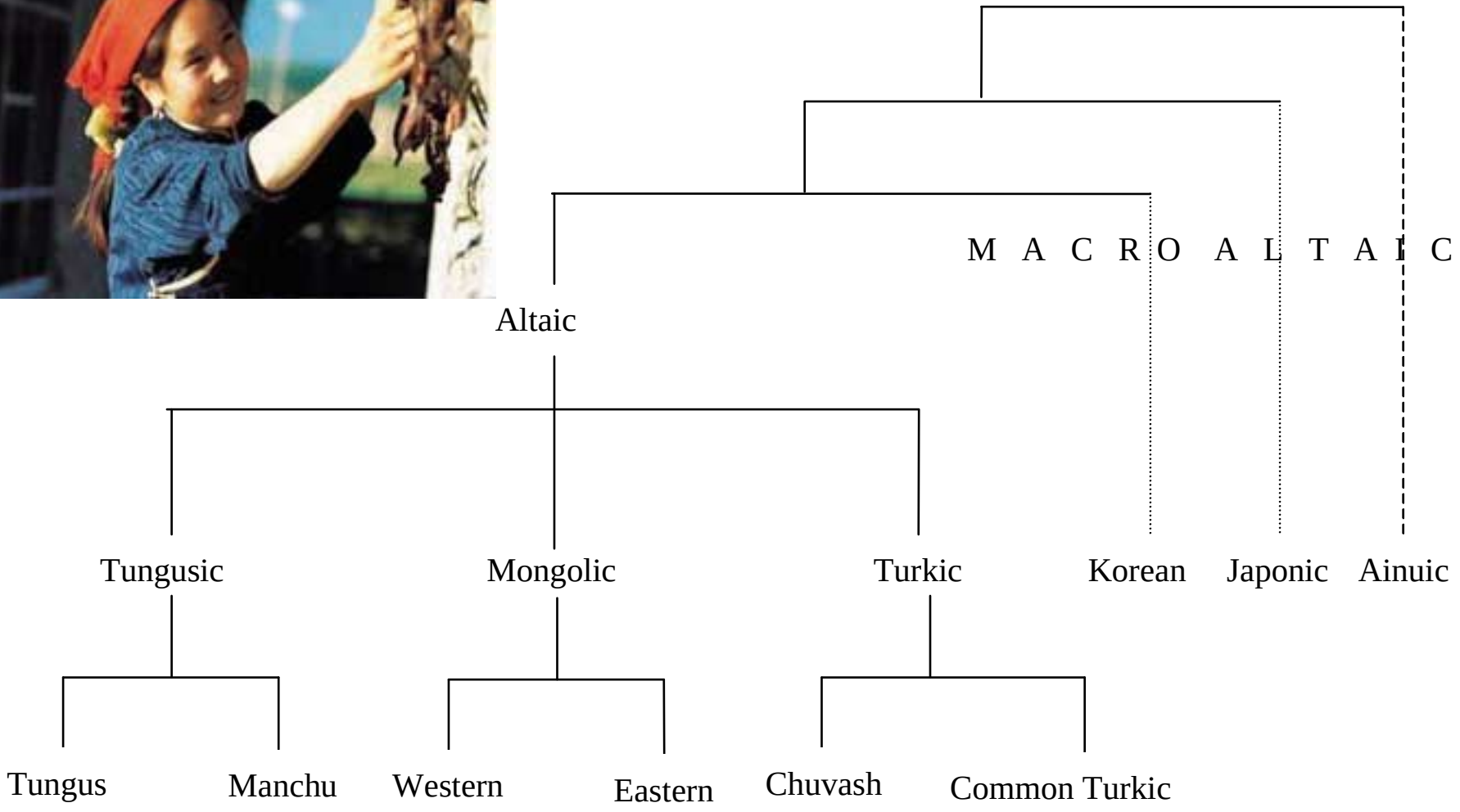
- Miao-Yao languages are also surprisingly undiverse. Hard to imagine the present languages have been diversifying for many millennia
- The linguistic geography of Miao-Yao suggests very strongly that these people were scattered by the incoming Han and probably forced southwards into Modern Laos and Thailand, probably in the last 3-4000 years

Miao-Yao II

- There has been a long-running argument about whether the rice-terminology of Miao-Yao was borrowed by the Sinitic speakers as they moved south.
- Japanese studies of the early cities, and palynological work is little-known but quite convincing. Liquidambar using Miao-Yao speakers were driven out of the river valleys by migrant Sinitic-speakers

D OWD LF

Altaic and Macro-Altaic



Altaic I

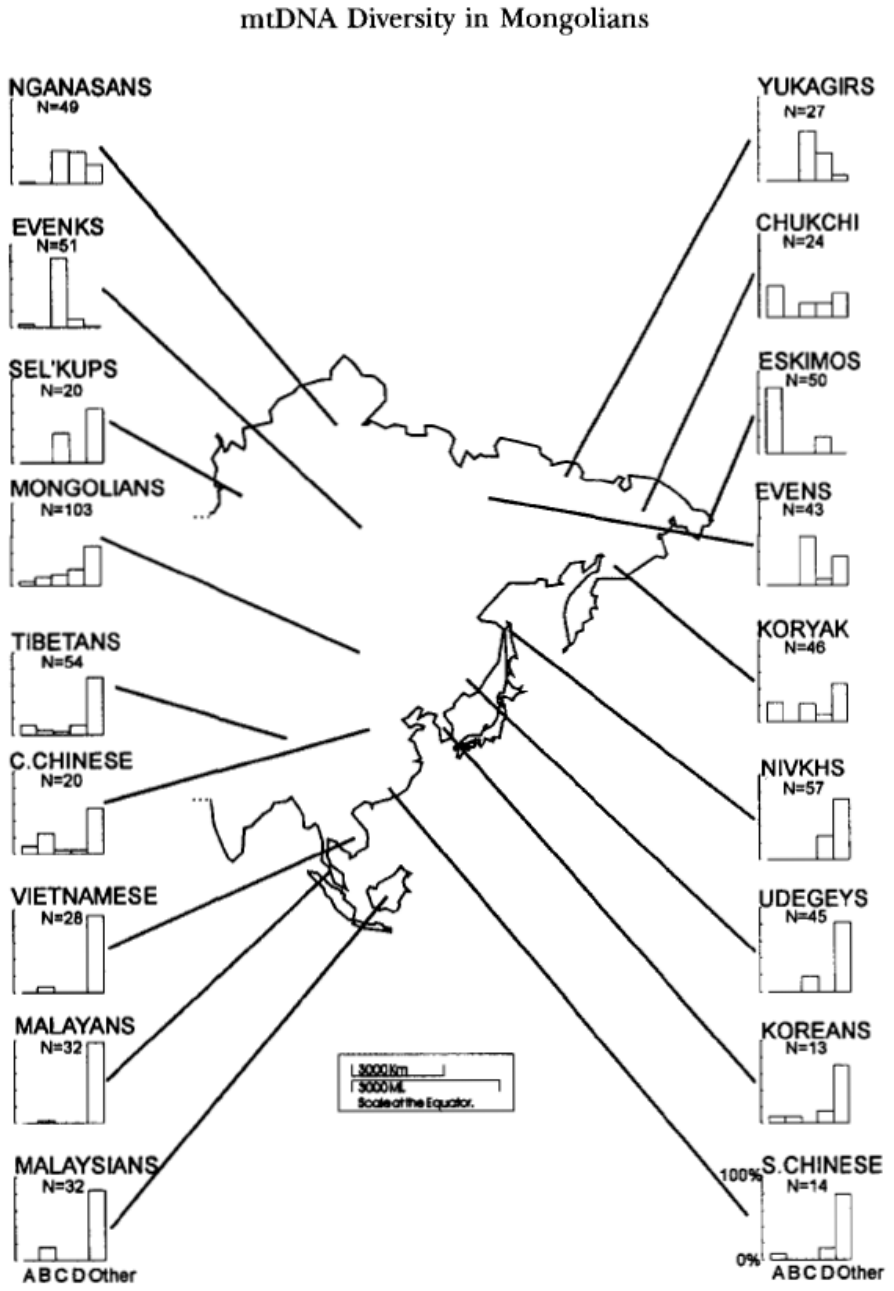
- Altaic is highly internally divided, so much so that some scholars claim it is not a phylum but a bundling of languages that have interacted.
- Macro-Altaic remains still more controversial although most scholars accept the membership of Korean, fewer Japanese
- Surprisingly, the individual members of Altaic, Turkic etc., are very undiverse and the dispersal of Turkic has largely taken place in historical time.

Altaic II

- Except for Manchu, Tungusic is now situated on the northern borders of China. But there is every reason to think that Tungusic speakers (who are quite likely the descendants of the LSA hunter-gatherers displaced by the rise of agriculture in North China) were once significantly further south. It is considered possible that Tungusic speakers were responsible for the introduction of 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).

Kolman et al. (1996)

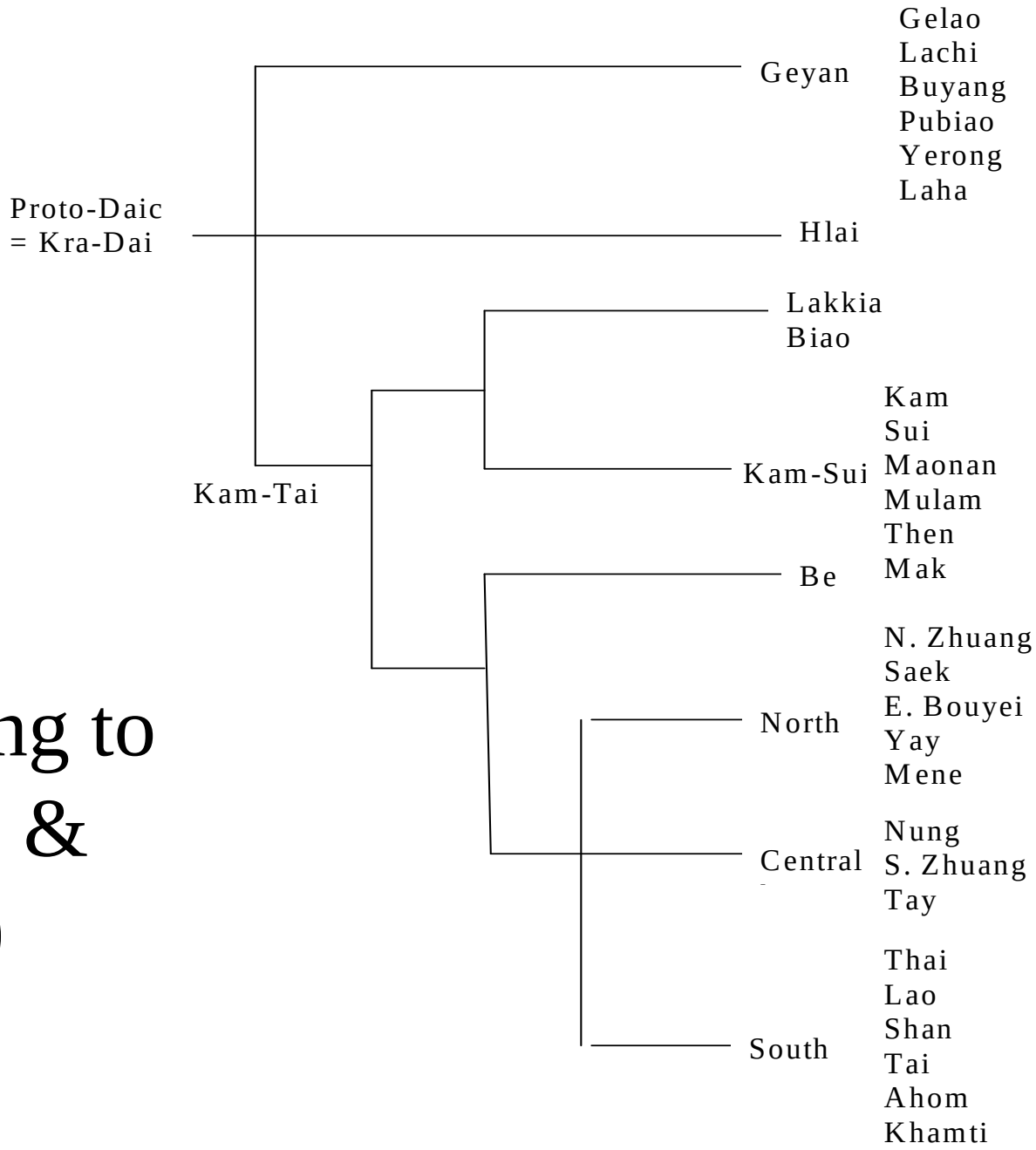
mtDNA diversity in Mongolians and other ‘northern’ peoples



G D LF

Major divisions of the Tai languages and related languages.

Daic according to Edmondson & Solnit (1997)



The drum tower of Kam Nationality, Guangxi Sanjiang Dong Autonomous County (1997)

G d lf#

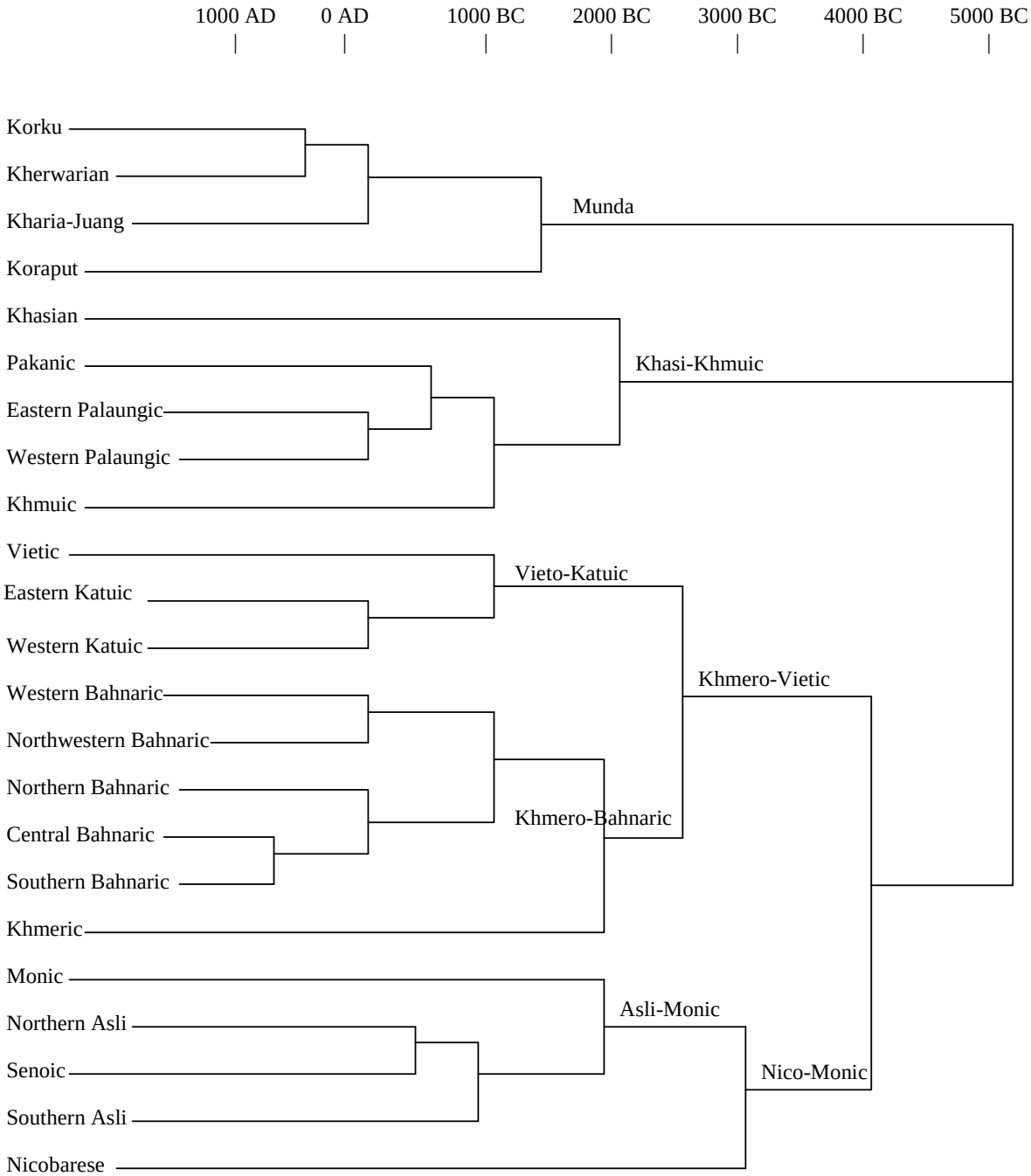
- All the diversity of Daic languages is in China: despite the southward extension of Thai today the likely origin of Daic is in Kweichow
- Daic languages are not all that diverse and almost certainly a candidate for a major agricultural expansion. Despite this, there is no obvious archaeological candidate to correlate with this
- The external affiliations of Daic have remained highly controversial, sharing as it does many features with surrounding language phyla. The recent proposal by Ostapirat on the relations between Daic with Austronesian will no doubt be controversial. Sagart has argued , and I accept, that this simply means the Daic speakers ‘are’ Austronesians, a branch of PMP that came back to the mainland.

G d lf#I

- Ostapirat assumes a simply model of split with Daic being the Austronesians who stayed at home.
- But... this seems unlikely, because in many ways 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 their language becoming radically restructured under the influence of Miao-Yao.

D X VWU R D VLD W LF

Distribution of the Austroasiatic languages.



Austroasiatic with a tentative calibration of time-depths for the various branches of the language family according to Diffloth

Austroasiatic I

- Austroasiatic languages are the most poorly researched of all those under discussion. Many are not documented at all and some recently discovered in China effectively not classified. Genetics of Austroasiatic speakers almost unresearched.
- Although there have been many promises, there are *no* justified proto-Austroasiatic reconstructions. Impossible to see whether faunal or crop names are really supported by a reconstructed proto-language
- But, Diffloth's claim that Austroasiatic speakers typically spread along river valleys seems to be justified, although they obviously became seagoing at some point

Austroasiatic II

- 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’

Austroasiatic III

- Diffloth has claimed that faunal reconstructions support a tropical origin; but no evidence for this is available and without a date, it is difficult to relate this to a dated palaeoenvironment
- It is worth noting, however, that 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?

Austroasiatic IV: genetics

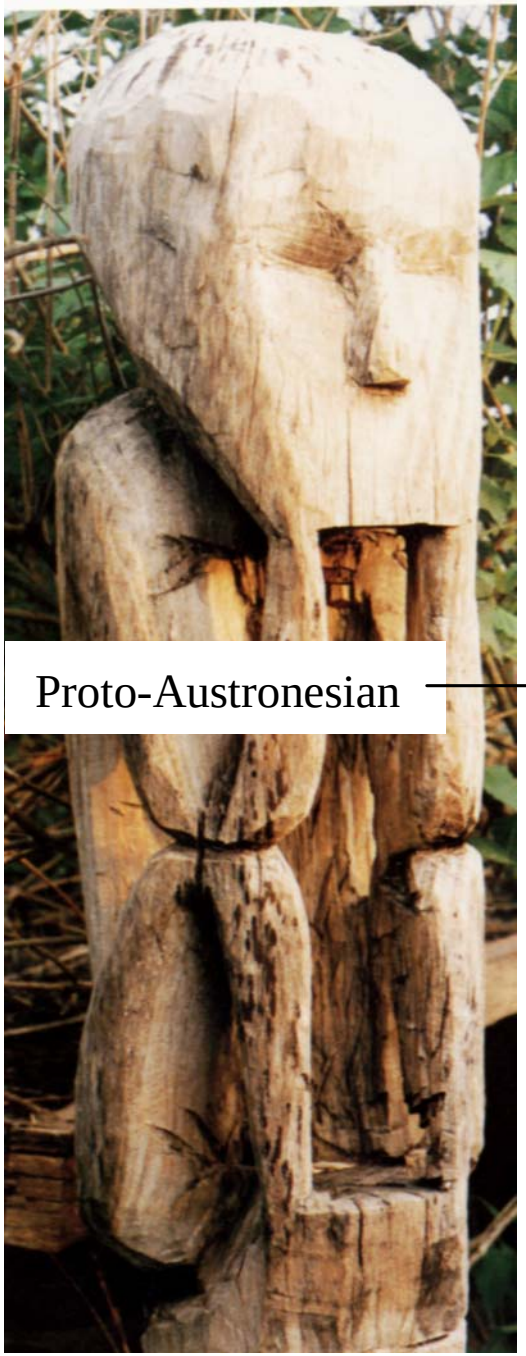
Our data indicate Austro-Asiatic speakers underwent population expansion about 17,000 years prior to the Elamo-Dravidian speakers and about 5,000 years prior to the Tibeto-Burman speakers. The confidence intervals of the expansion times of Austro-Asiatic and Tibeto-Burman speakers are non-overlapping 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

AUSTRONESIAN

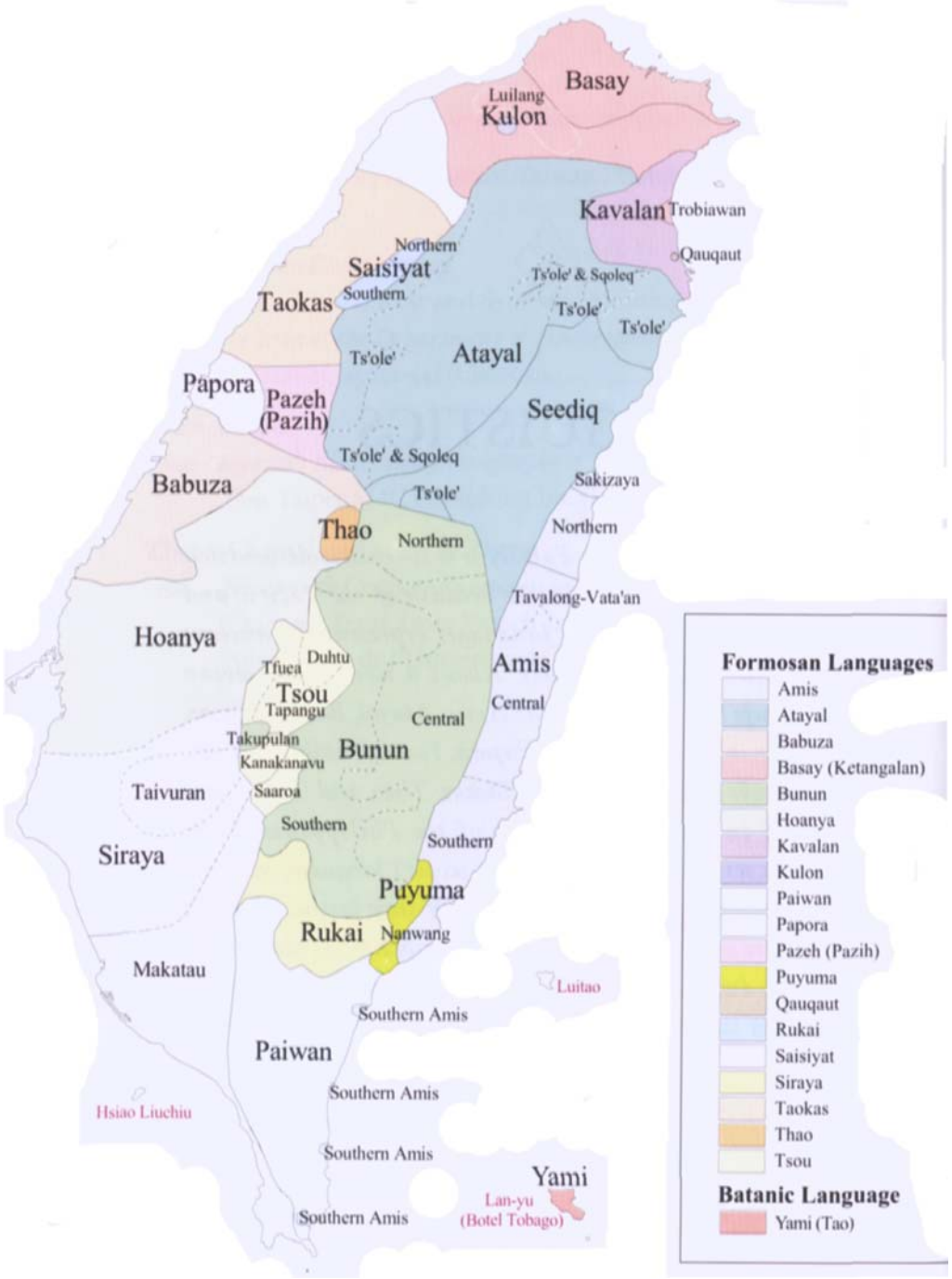

TAIWAN
Lan-yu
Hainan
VIETNAM
WMP
Mariana Islands
PHILIPPINES
MICRONESIA
Oc
MARSHALL ISLANDS
Hawaiian Islands
PALAU
Caroline Islands
MALAYSIA
Halmahera
SHWNG
Borneo
Sulawesi
PAPUA NEW GUINEA
MELANESIA
KIRIBATI
Nauru
POLYNESIA
INDONESIA
Seram
CMP
Timor
Solomon Is
TUVALU
Marquesas Islands
WMP
SAMOA
Tuamotu Archipelago
VANUATU
FIJI
Tahiti
TONGA
Cook Islands
MADAGASCAR
AUSTRALIA
New Caledonia
Easter Island
Oc
NEW ZEALAND

Austronesian according to Blust



Proto-Austronesian

- Atayalic
  - Atayal
  - Sediq
- East Formosan
  - Favorlang
  - Saisiat, etc.
- Puyuma
- Paiwan
- Rukai
- Tsouic
  - Tsou
  - Saaroa
  - Kanakanabu
- Bunun
- Western Plains
- Northwest Formosan
- Malayo-Polynesian


Basay
Luilang
Kulon
Kavalan
Trobiawan
Qauqaut
Northern
Saisiyat
Southern
Taokas
Ts'ole' & Sqoleq
Ts'ole'
Ts'ole'
Ts'ole'
Atayal
Papora
Pazeh
(Pazih)
Seediq
Ts'ole' & Sqoleq
Babuza
Ts'ole'
Sakizaya
Thao
Northern
Northern
Tavalong-Vata'an
Hoanya
Duhtu
Tfuea
Amis
Tsou
Central
Tapangu
Central
Takupulan
Bunun
Kanakanavu
Taivuran
Saaroa
Southern
Southern
Siraya
Puyuma
Rukai
Nanwang
Makatau
Luitao
Southern Amis
Paiwan
Southern Amis
Hsiao Liuchiu
Southern Amis
Yami
Lan-yu
(Botel Tobago)
Southern Amis
Formosan Languages
Amis
Atayal
Babuza
Basay (Ketangalan)
Bunun
Hoanya
Kavalan
Kulon
Paiwan
Papora
Pazeh (Pazih)
Puyuma
Qauqaut
Rukai
Saisiyat
Siraya
Taokas
Thao
Tsou
Batanic Language
Yami (Tao)

Austronesian I

- The 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. 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. The Hemudu site, northwest of Taiwan is usually identified as a possible source area.
- The Ta Peng Keng culture which links SE China and Taiwan represents the displacement of Pleistocene hunter-gatherers by incoming rice/marine people

Indo-European

Indo-European 1

- ‘Tajik’ in China consists of two languages, Sarikoli and Wakhi, both Indo-Iranian, languages of the Pamir branch.
- These are historically recent incursions, unconnected with the more puzzling issues of Tocharian and the Tarim Basin mummies. Tocharian is a language only known from written texts but which is Indo-European and surprisingly, probably related to Celtic and Italic, though this is disputed. Clearly the Tocharians interacted with a wide variety of languages of different phyla in their journey across Asia.

Indo-European 2

- 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?
- Actually there are loads of questions and many loony answers but I'll refrain just now
- One reason for thinking this are the 'Caucasian' features of the mummies and the striking features of dress including tartans and 'Welsh' conical hats. Assuming we are not dealing with stray Celtic supporters, it is reasonable to assume that these are Indo-European speakers and that they were hunter-gatherers who somehow wandered this long distance in pursuit of animals.

Indo-European 3

- 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 would be the simplest solution

Oota et al. 2002. 'Extreme mtDNA homogeneity in continental Asian populations.' **Am J Phys Anthropol 118:146-153.**

"Mitochondrial DNA (mtDNA) variation in continental Asia has not been well-studied. Here, we report 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."

Wang et al. (2000) Genetic Structure of a 2,500-Year-Old Human Population in China and Its Spatiotemporal Changes

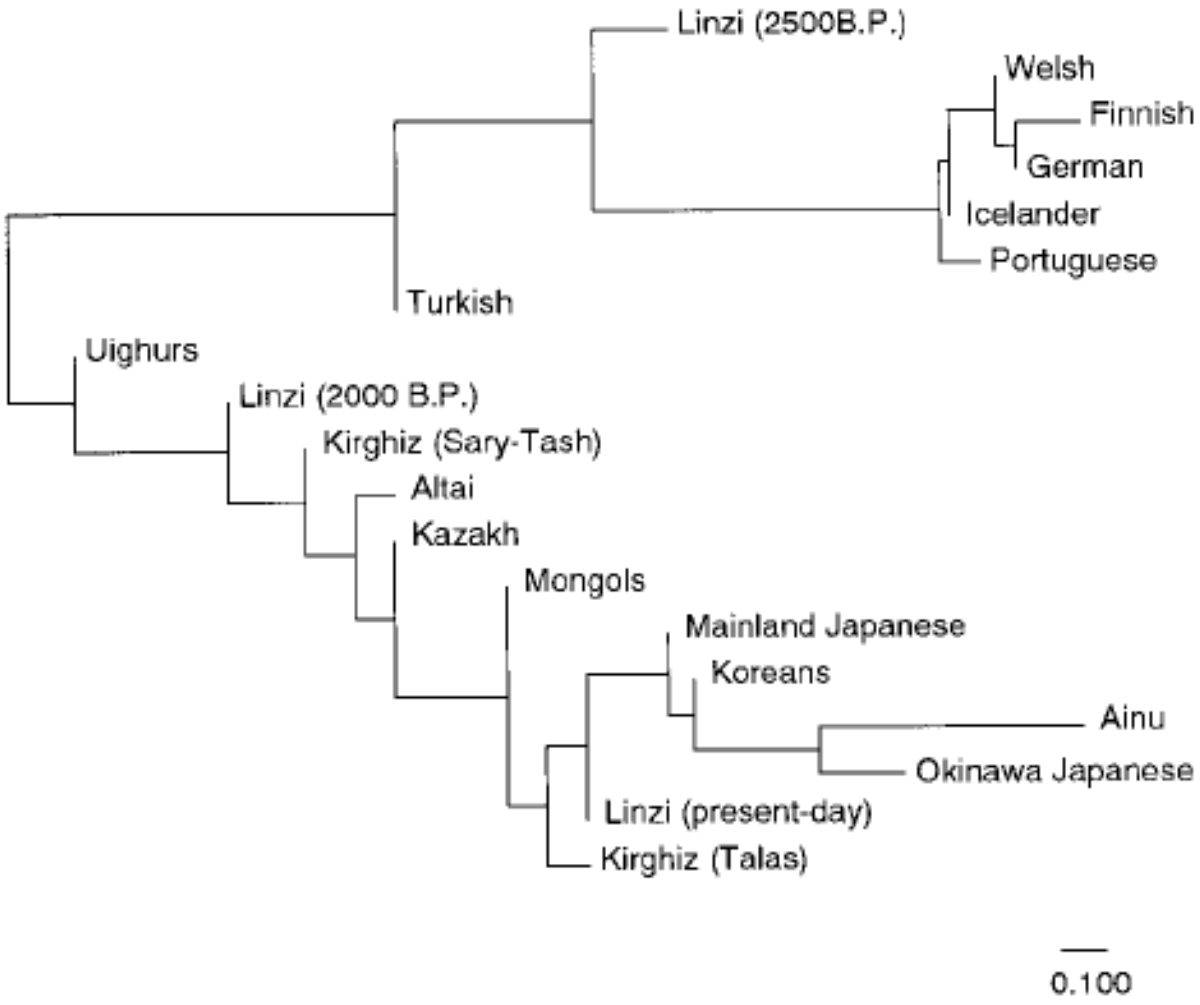
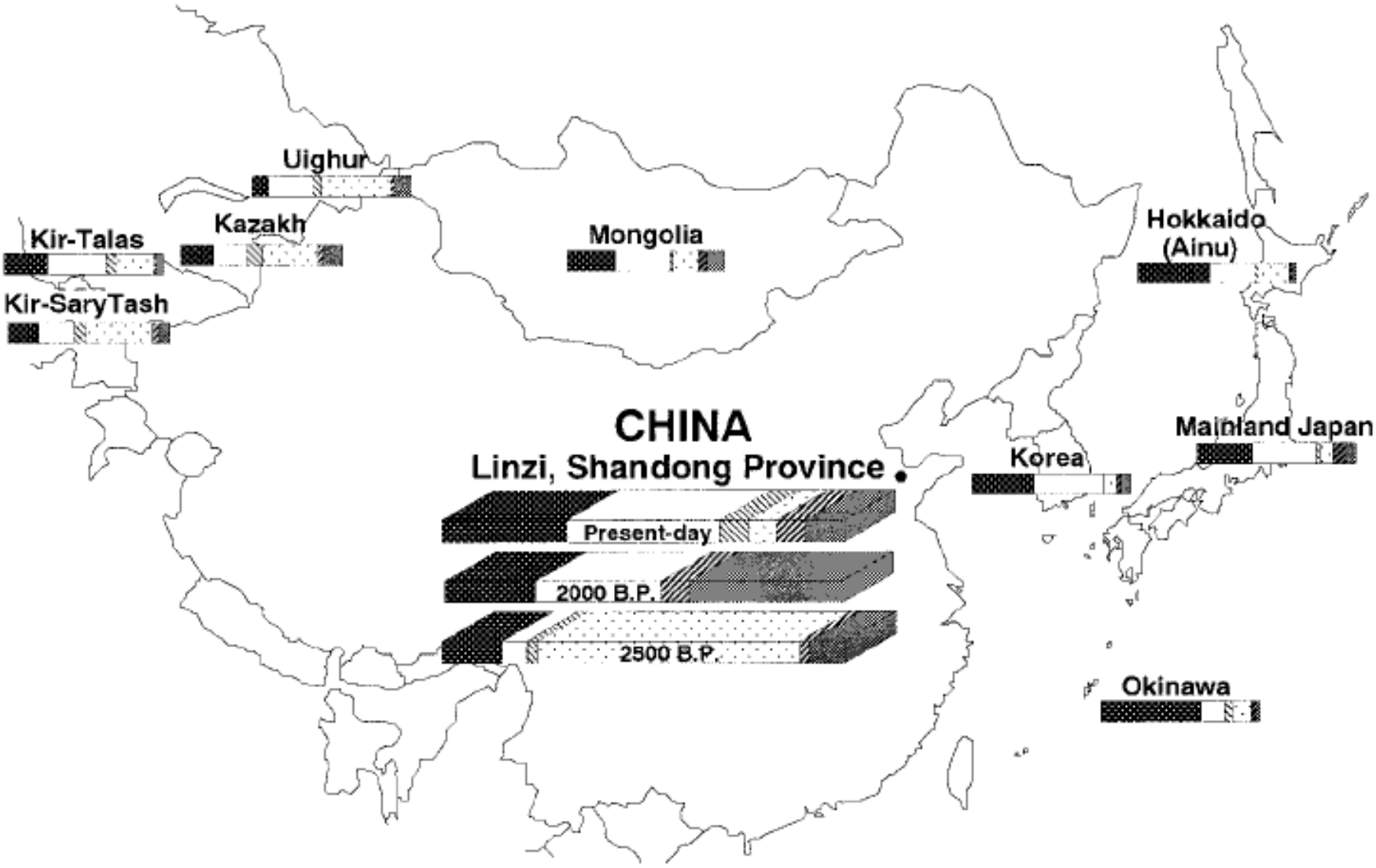


FIG. 3.—A neighbor-joining tree for 19 human populations. Branch lengths are proportional to genetic distances.

Wang et al. (2000) Genetic Structure of a 2,500-Year-Old Human Population in China and Its Spatiotemporal Changes

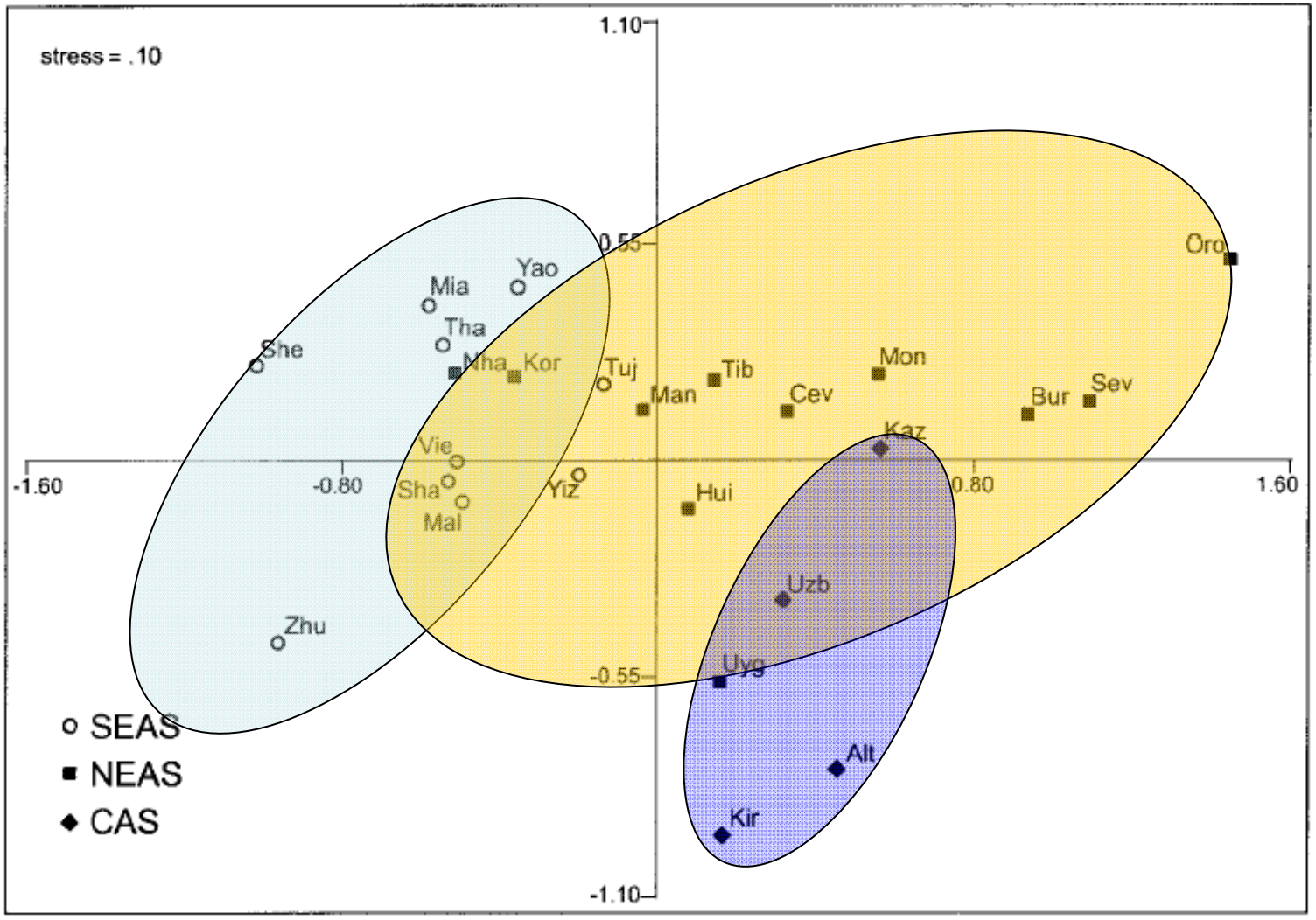


But maybe not….

In short, both the 2,500-year-old and the 2,000-year old Linzi populations had features in common with the modern populations from south China rather than any specific affinity to the European mtDNA pool.

Yao et al. 2003

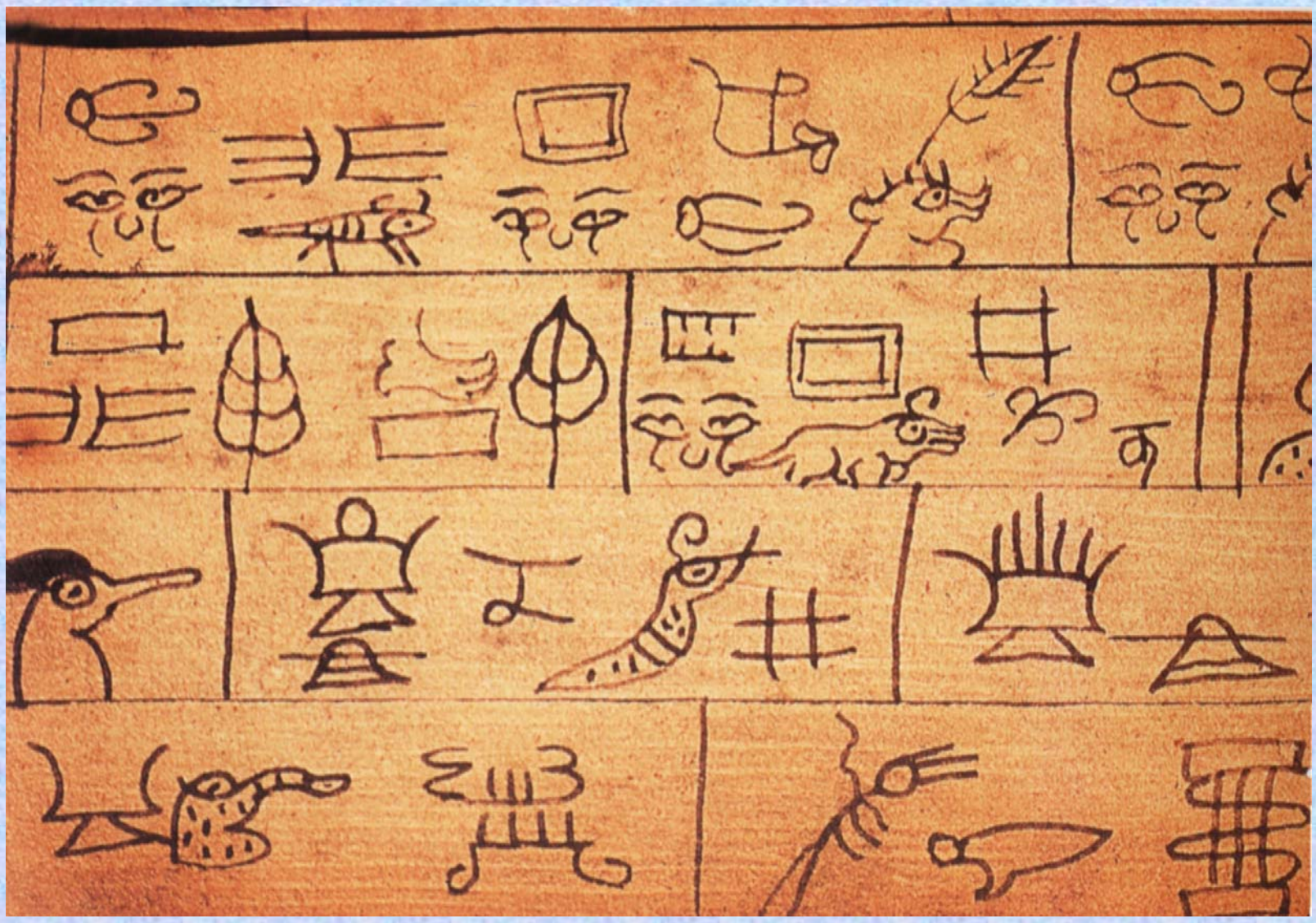
From: Karafet et al. (2001) Paternal Population History of East Asia: Sources, Patterns, and Microevolutionary Processes



MDS plot of 25 Asian populations, based on Φ_{ST} genetic distances. For three-letter population codes, see table 1.

From: Karafet et al. (2001) Paternal Population History of East Asia: Sources, Patterns, and Microevolutionary Processes

Conclusions



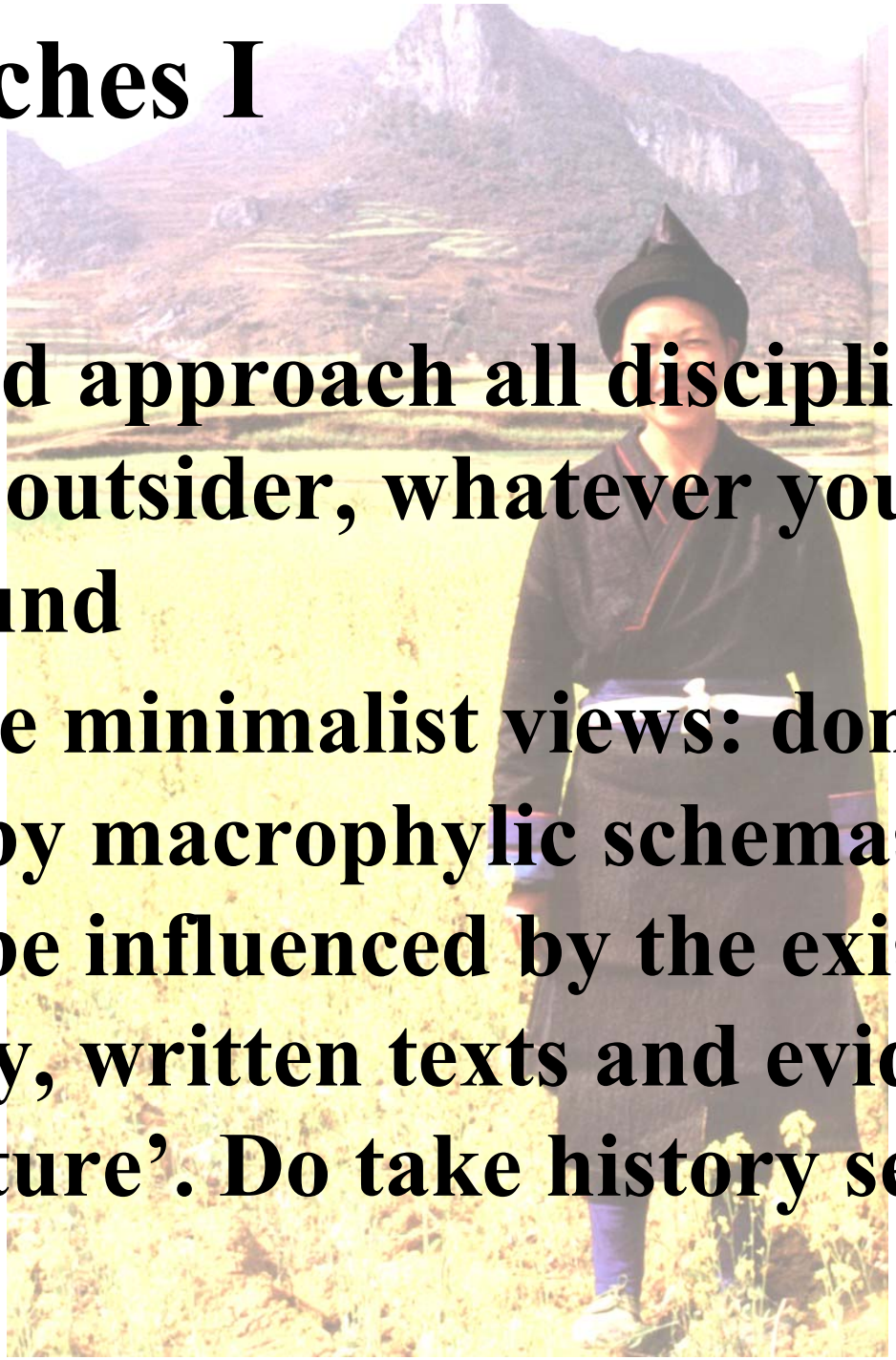
CONCLUSIONS I

- There is so far very little linkage between archaeological cultures and the patterns of different ethnolinguistic groupings.
- The antiquity of these groupings is highly controversial
- The 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

CONCLUSIONS II

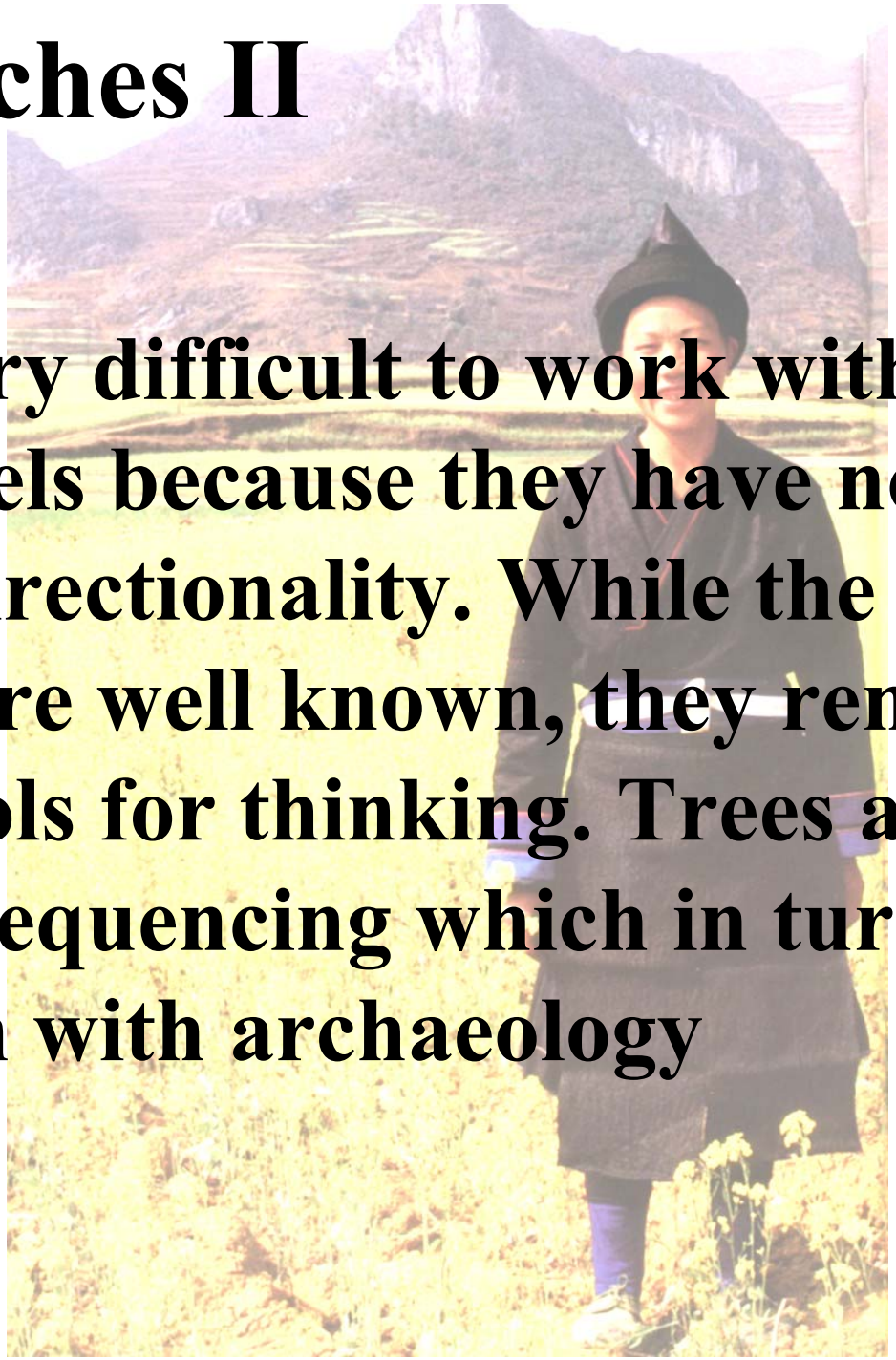
- Historical linguistics has a very long way to go especially in reconstructing lexical items that could be linked to archaeology. Some phyla remain very poorly served
- Archaeology remains very patchy with some areas well-known, others not.
- Genetics seems to be solving some large-scale problems about human settlement of the region. But it seems hard to know whether it can contribute to the problems of the interface of linguistics and archaeology
- But … collaboration such as in Austronesian studies is highly desirable

Approaches I



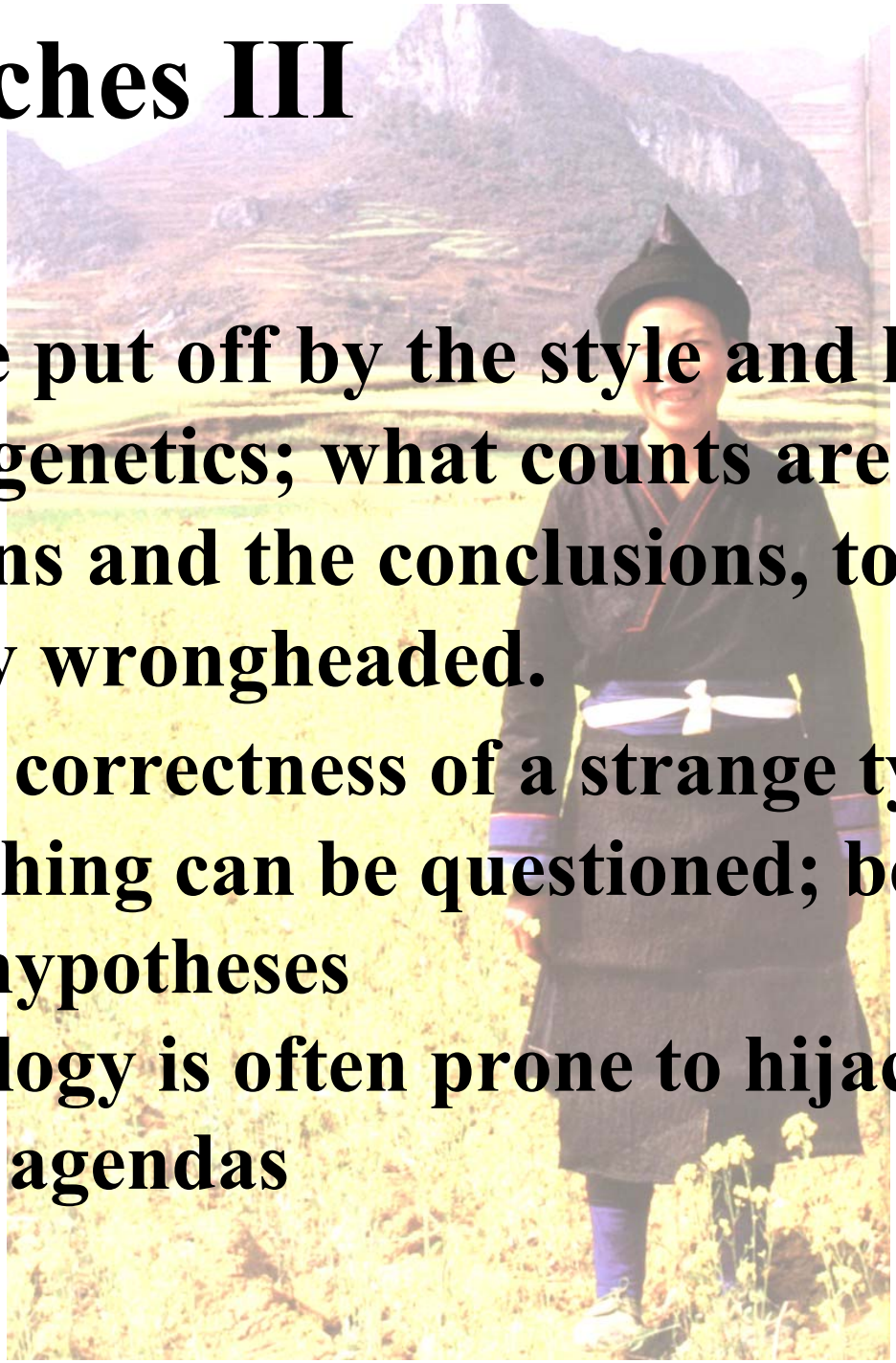
- Try and approach all disciplines as a sceptical outsider, whatever your background
- Assume minimalist views: don't be seduced by macrophylic schemas
- Don't be influenced by the existence of epigraphy, written texts and evidence of 'high culture'. Do take history seriously.

Approaches II



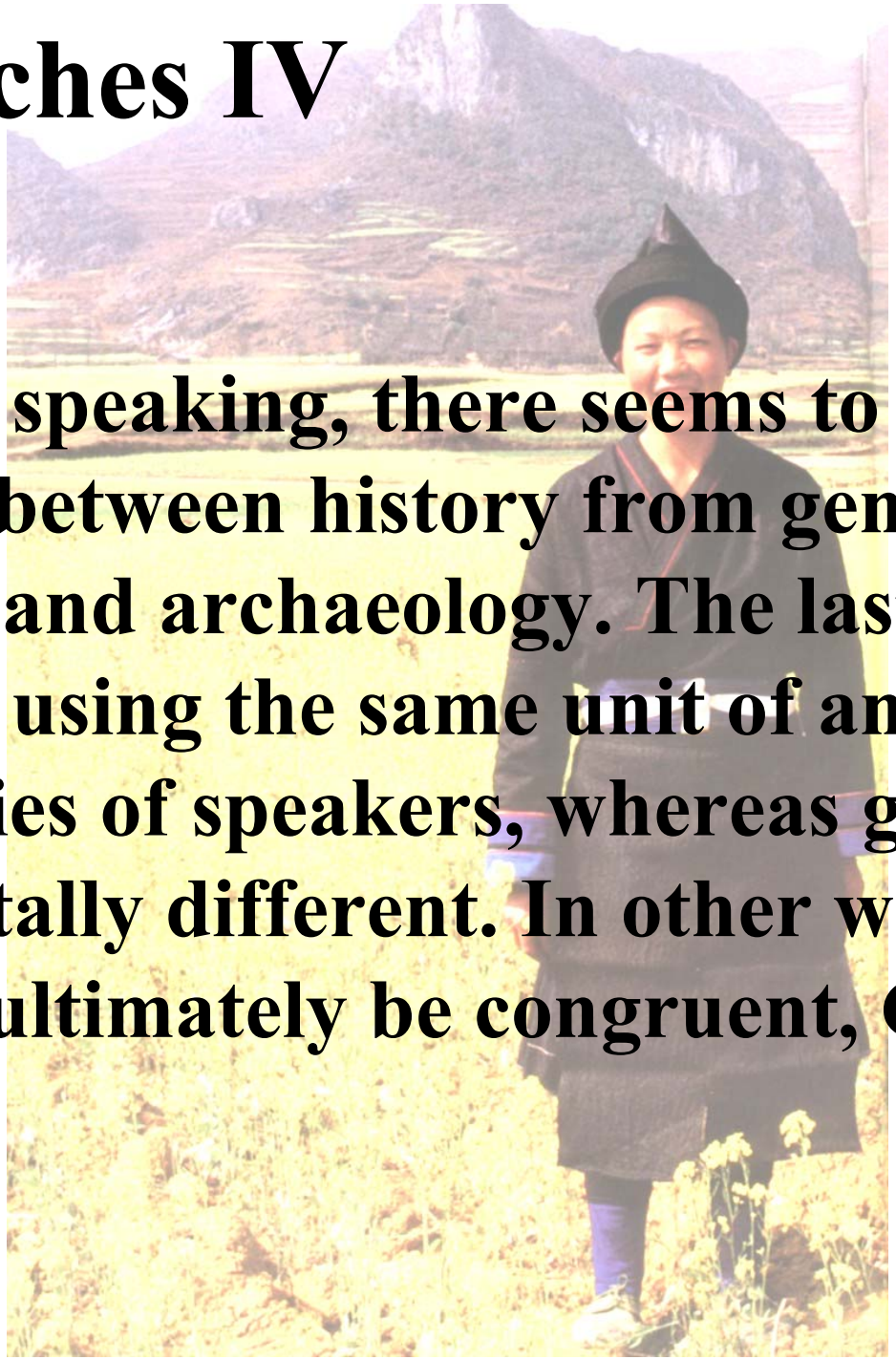
- It is very difficult to work with non-tree models because they have no dates and no directionality. While the problems of trees are well known, they remain useful tools for thinking. Trees also provide sequencing which in turn helps link them with archaeology

Approaches III



- Don’t be put off by the style and hard science of genetics; what counts are the assumptions and the conclusions, too often worryingly wrongheaded.
- Political correctness of a strange type means that everything can be questioned; better go with bold hypotheses
- Archaeology is often prone to hijacking by nationalist agendas

Approaches IV



- Broadly speaking, there seems to be a mismatch between history from genetics and linguistics and archaeology. The last two are essentially using the same unit of analysis, communities of speakers, whereas genetics is fundamentally different. In other words, L & A, should ultimately be congruent, G might never be.