How can we use the 'Words and things' approach in understanding the prehistory of Sino-Tibetan languages?

La Trobe University,
Melbourne, 2nd September, 2015

Roger Blench
Visiting Fellow, Museu Emilio Goeldi, Belem, Brazil
Chief Research Officer, Kay Williamson Educational Foundation
McDonald Institute for Archaeological Research, Cambridge
Visiting Professor, University of New England, Armidale
Academic Visitor, Rajiv Gandhi University, Arunachal Pradesh

Sino-Tibetan

- Despite being the language phylum with the most speakers (currently 1.5 billion) the internal classification of Sino-Tibetan is poorly understood and marked by a conspicuous lack of consensus.
- Underlying this may be a characteristic focus of the ‘major’ languages, such as Sinitic, Tibetic and Burmese, all of which, while historically important, probably represent low-level branches.
- The diversity of Sino-Tibetan is concentrated between Arunachal Pradesh and Nepal and there seems little doubt that an approach to reconstruction which took these languages as primary would propose a markedly different range of proto-forms.
- One indicator of this is that reconstructions in Tibeto-Burman include words for ‘iron’, ‘trousers’ and other words indicative of relatively late urban culture.

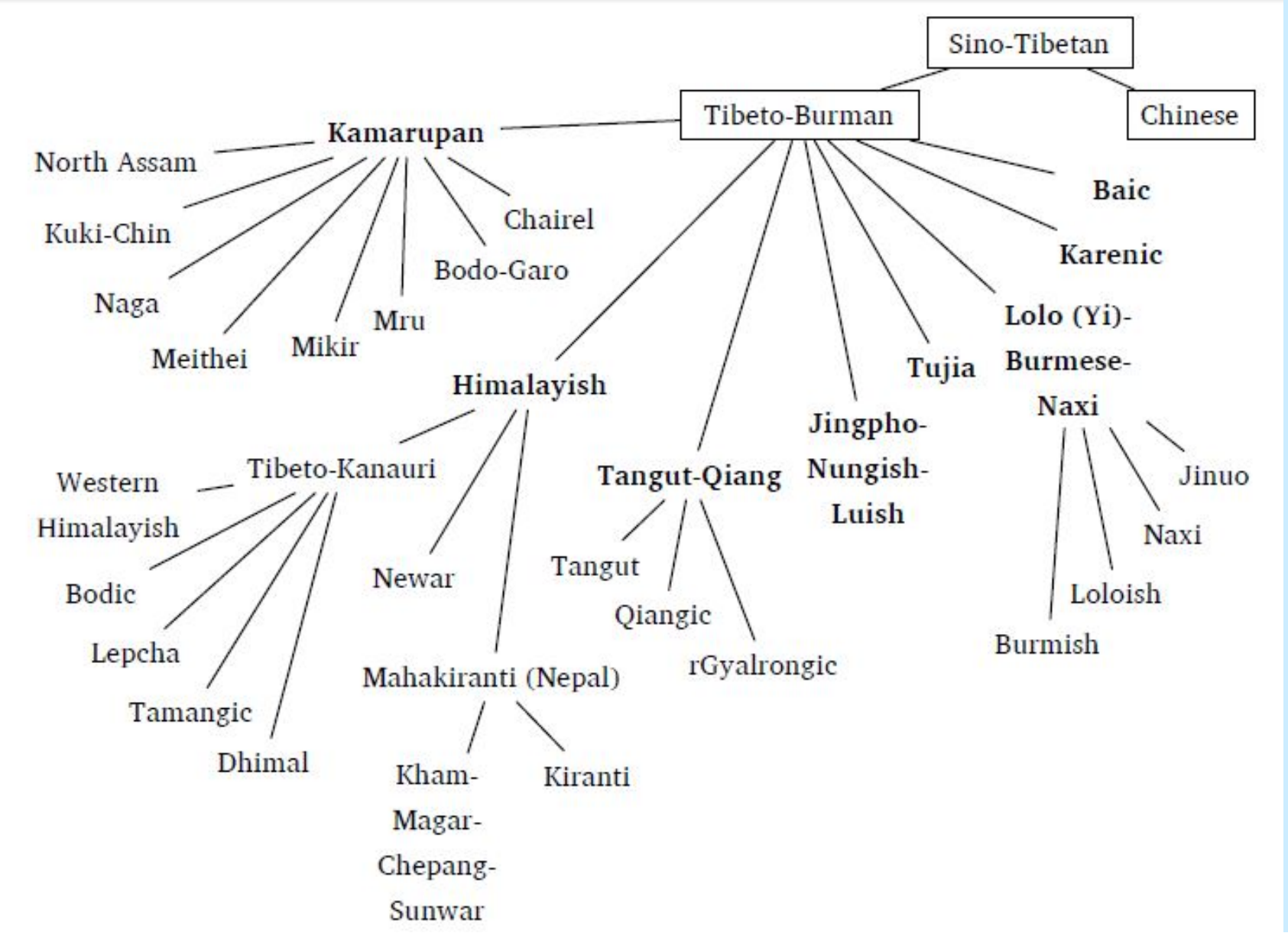
Sino-Tibetan II

- As most people will know there has been a lengthy and unresolved debate over the name of the phylum
- 'Tibeto-Burman' and 'Trans-Himalayan' have been proposed
- The presentation uses Sino-Tibetan, but with no presupposition of any particular model of the language family
- There is, however, increasing consensus that Sinitic lies *within* Sino-Tibetan and does *not* constitute a primary branch
- This was rather a cultural classification like Semitic within Afroasiatic since no evidence has ever been presented to support it.
- More crucial, however, is the sort of model we present on internal structure, since this affects out reconstructions and thus hypotheses about the genesis and diversification of the phylum
- It is also worth noting that no evidence has been presented for the unity of the phylum as a whole. Linguistics by assertion?

Sino-Tibetan III

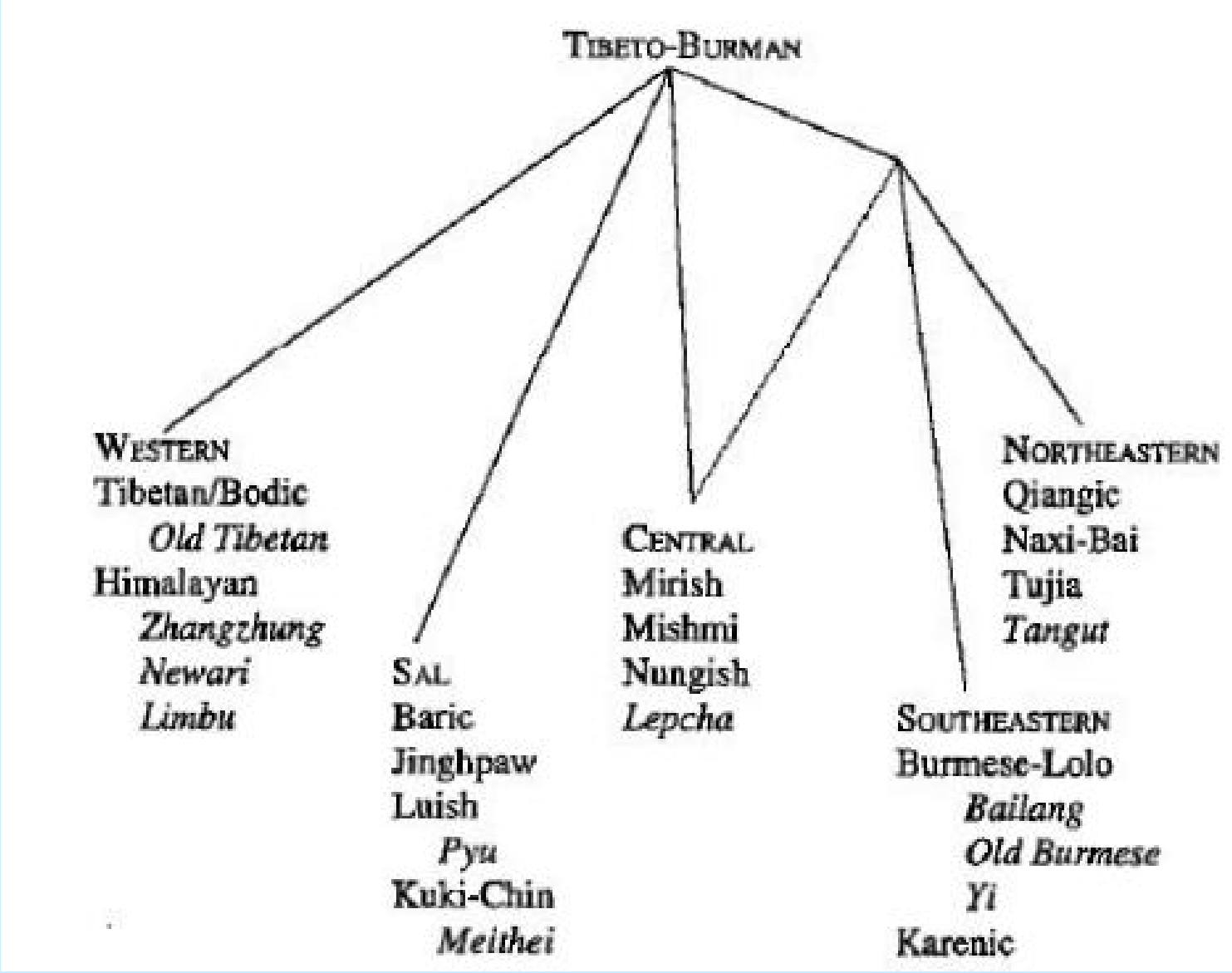
- Unfortunately, the STEDT school makes no attempt to provide an internal classification, retaining the primary Sinitic split and a vague 'Kamarupan' admitted not be genetic
- The alternative, the 'fallen leaves' of George Van Driem, takes no view about genetic affiliation and has as many as forty independent groups with no argument as to the relations between them
- And both of these have nothing to say about the 'North Assam' languages, the highly diverse languages of Arunachal Pradesh, whose Sino-Tibetan affiliation is very uncertain and not demonstrated anywhere in the literature.
- Proof that Gongduk and Lhokpu are Sino-Tibetan also seems to be sketchy in the extreme. We need to be able to consider this possibility that these languages are not Sino-Tibetan at all, but isolates.
- Some examples of proposed classifications

Matisoff (2008)



GvD's 'fallen leaves'

Bradley (2002)



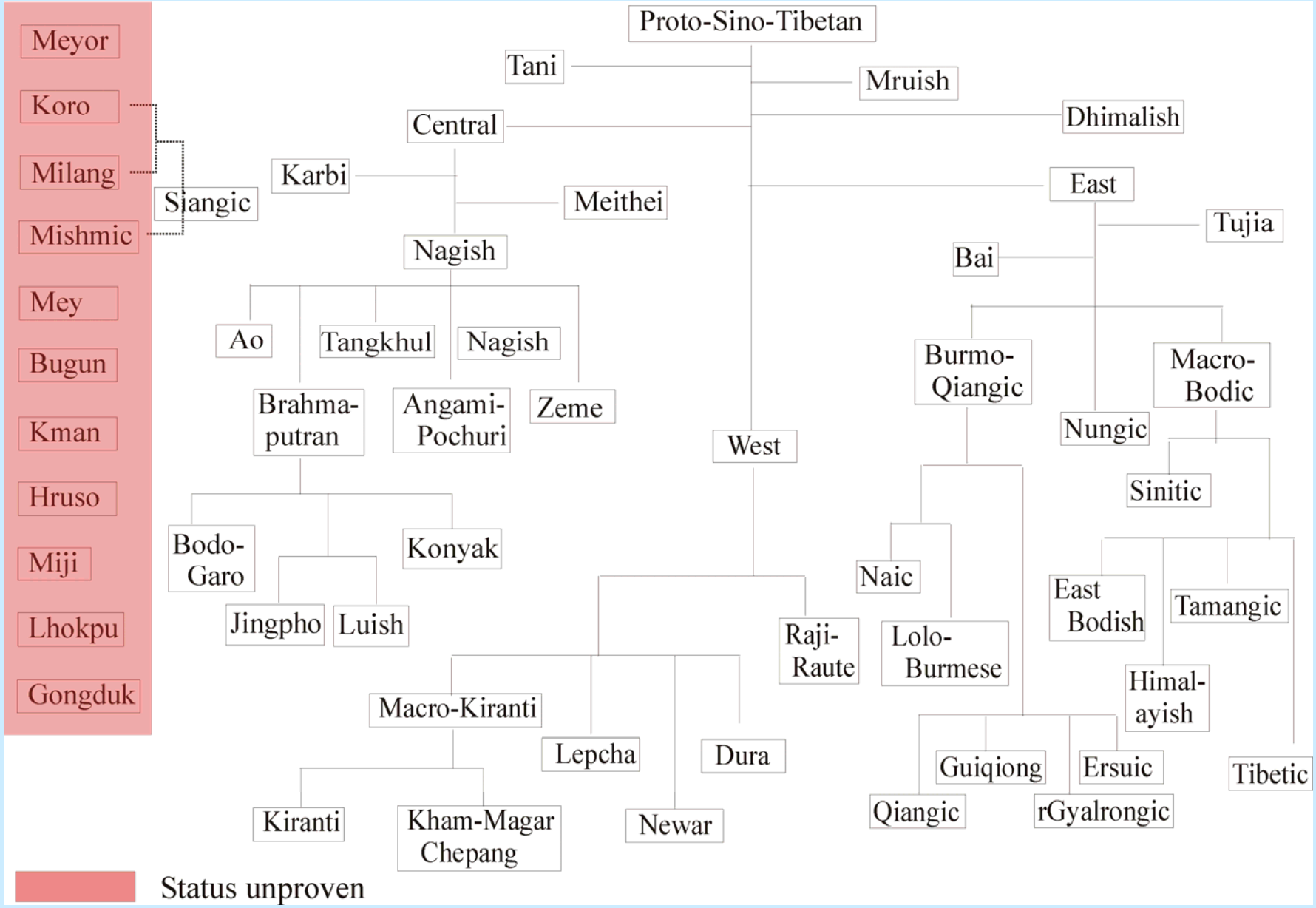
Sino-Tibetan IV

- The STEDT school retains the primary Sinitic split and a vague Kamarupan' admitted not be genetic
- The alternative, the 'fallen leaves' of George Van Driem, takes no view about genetic affiliation and has as many as forty independent groups with no argument as to the relations between them
- The 'greatest diversity' principle would certainly place the origin of Sino-Tibetan in NE India, unless…

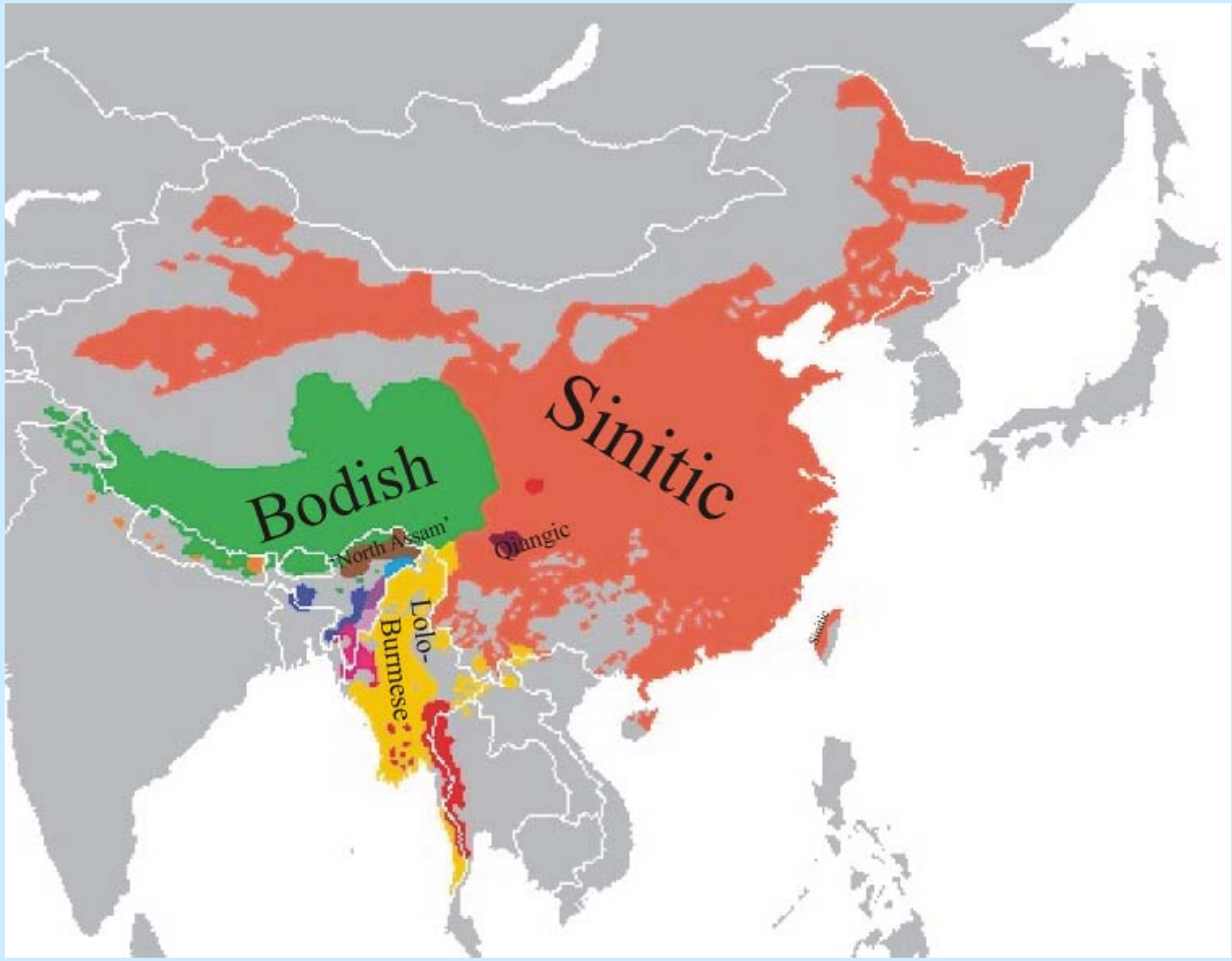
What sort of tree for Sino-Tibetan?

- I have tried to put together a more useable tree, excluding individual languages where the evidence for Sino-Tibetan affiliation is weak or 'unproven'
- These are marked separately awaiting further argument
- But probably these issues cannot be resolved until be decide what counts as 'core' Sino-Tibetan and also what percentage of core lexemes must a language possess to acquire membership
- Which alos requires us to esstablish
- And suggesting that all the 'core' languages are at the base of the tree and highly diverse ones as primary branches
- This provides a *tool for thinking* about the issue of regional lexemes

Sino-Tibetan configuration



Sino-Tibetan map



How old is Sino-Tibetan?

- Apart from the internal structure, it would be useful to know how old Sino-Tibetan is. For example, in Matisoff (2008: xxxvi) advances a date of 6000 BP, although on what basis is unclear.
- This would be well before metals, and also before agriculture in some parts of the Sino-Tibetan area
- So clearly such a date has to correspond to both to the proposed reconstructions and reflect a hypothesis about internal structure
- On the basis that agriculture cannot be reconstructed for Sino-Tibetan, nor metal use, one might want to put the earliest dates even older.
- Whatever the case, only quality reconstructions supported with data tables and excluding loanwords will help us assign a date

Sino-Tibetan V

- But really we need branch by branch reconstruction and then a gradual construction of mesolects. Some of this exists with Tangkhulic, Chin, Ersuic, Yi etc.
- However, in relation to big historical questions, such as how old is Sino-Tibetan, where did it originate and were its first speakers farmers, foragers or vegeculturalists, we are yet to make much progress
- And of course big linguistic questions such as was it originally tonal, did it have long complex verb forms as in Kiranti and Qiangic or are these later developments?
- Despite the current enthusiasm for Bayesian phyologenies, I believe the 'words and things' (*Worter und Sachen*) or 'linguistic palaeontology' still has mileage
- But we face two problems in this area;
- The first is the highly patchy archaeology. China is well-covered, Thailand and Việt Nam fairly well and Nepal, NE India, barely at all

Archaeology and Sino-Tibetan

- Even where we have archaeology, it is highly oriented towards tombs, cemeteries, and archaeobotany focuses on cereals, typically rice.
- Suppose early speakers were vegeculturalists, as ethnography suggests, then a lack of starch and phytolith work means that we wouldn't have evidence for this.
- The second problem is reconstruction procedure. It seems the occurrence of an apparently cognate root in literally a few languages qualifies a word to be considered PTB.
- A couple of examples, 'iron' and 'crossbow' illustrate the sort of problems we can encounter

Sino-Tibetan reconstruction issues: iron

- Iron is given as PTB **syam*. We know that iron-smelting appears in China 500 BC, and smelting only spreads to small-scale societies further south after 0 AD.
- We know that the Chinese trade knives, axes and tools before iron production spreads
- This completely *disqualifies* iron from being PTB and the forms must be a chain of loanwords.
- And indeed, consultation of STEDT shows a huge diversity of forms

Sino-Tibetan reconstruction issues: crossbows

- ‘Crossbow’ is another example, which is reconstructed as **s.na* in PTB. When we look up the actual evidence it turns out the only evidence is OC *nǔ*.
- Baxter and Sagart reconstruct OC *C.nˤa ʔ*, which looks very complicated and entirely improbable given that it is a borrowing from Austroasiatic **sna*, which is attested in almost all branches except Munda.
- The PTB reconstruction has been ‘fixed’ of course since it is not a true reconstruction based on evidence but a construct from prior knowledge.
- We actually know when crossbows appear in China, at around 200 BC, since we find the firing mechanisms in bronze, or even whole crossbows, in tombs

Sino-Tibetan reconstruction issues: crossbows

- Early Qin crossbow, ca. 200 BC

Sino-Tibetan environmental reconstruction

- So the message is, if we want to use a ‘words and things’ approach we have to be serious about evidence and loanwords, informed archaeologically and looking outside the phylum
- A classic feature of this type of approach is environmental reconstruction. Do reconstructible forms for plants, animals, climatic features indicate the possible homeland of the phylum?
- For example, if the homeland lay in a high montane region, we would expect a could reconstruction for ‘snow, ice’. If in a flat plain, perhaps no words for ‘mountain, valley’ (at the proto-language, only in mesolects)
- Similarly with animals; reconstructions compared with biogeography should provide clues to the location of a homeland.
- All these are established procedures with Indo-European; and indeed this is the point of conflict with the ‘Bayesian’ model which favours Anatolia as opposed to the Kurgan model

Sino-Tibetan ‘ice, snow’

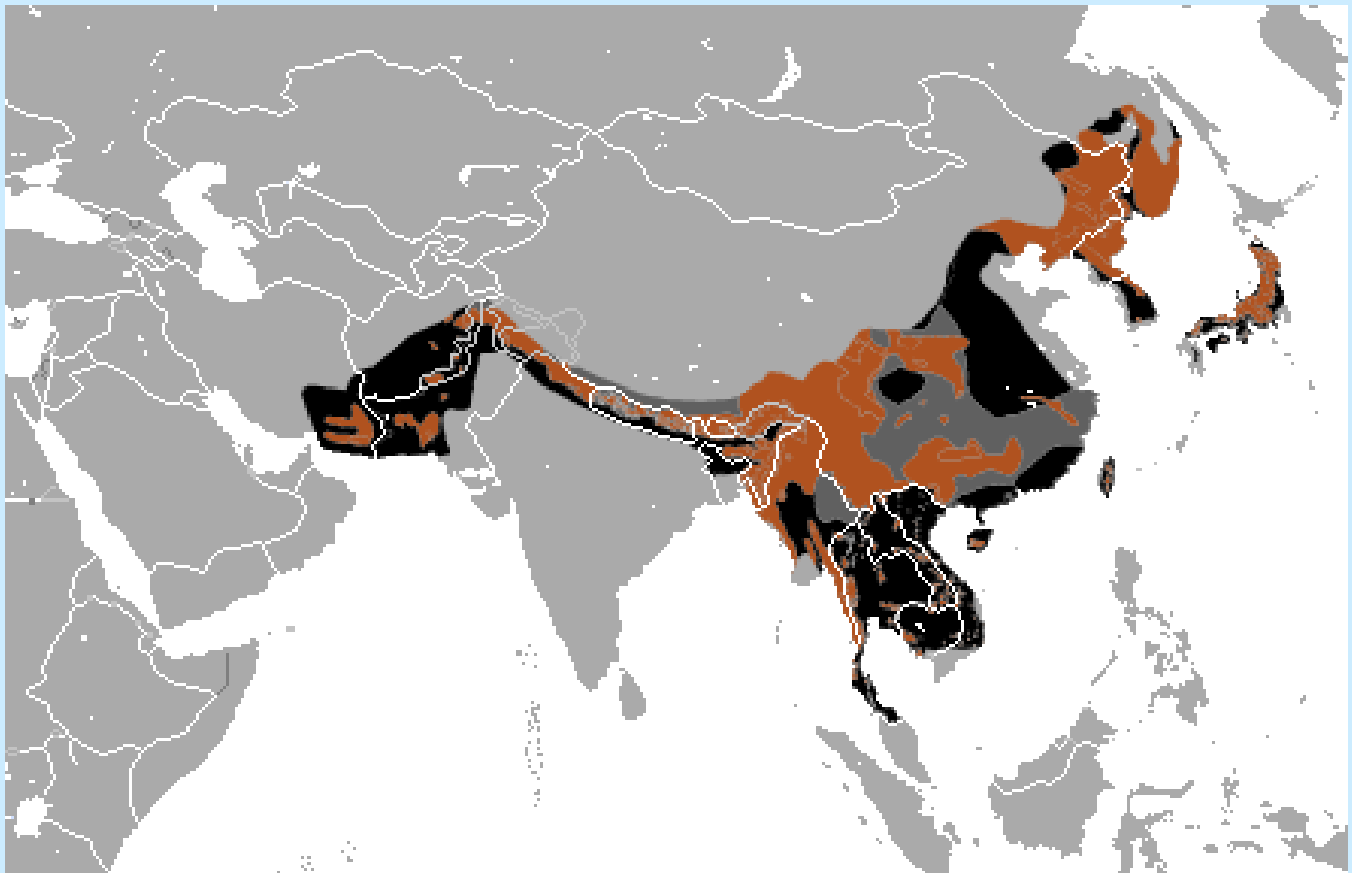
- If we tabulate words for ‘ice, snow’ we find about five or six distinct roots, scattered across the region, including *#kliN, #shü, #kyam, #pham* (see Appendix to Blench & Post 2013)
- Even within a single branch several different roots may occur
- This suggests that snow and ice did not feature as an essential part of the climate in the original zone of Sino-Tibetan languages, but that new words were adopted, invented and borrowed as languages moved around

Sino-Tibetan subsistence

- It seems that fish were a major item in the diet of early Sino-Tibetan speakers.
- Almost every branch has a clear reflex of **ŋ̊a/ ŋa** suggesting that fish were salient in the lexicon
- Hunting was certainly important, since reflexes of N-la(k)/B-la(k) 'arrow' occur in every branch of Sino-Tibetan except Baiic, assuming OC *lək* 'shoot arrow' is cognate. The root for 'bow' is similarly widespread, #-*li(k)*, or else the same root as 'arrow'
- By contrast the word for 'spear' seems not to reconstruct at all. The PTB form given in Matisoff (2003) **m.dung* is *not* attested in most branches and there is no other widespread root

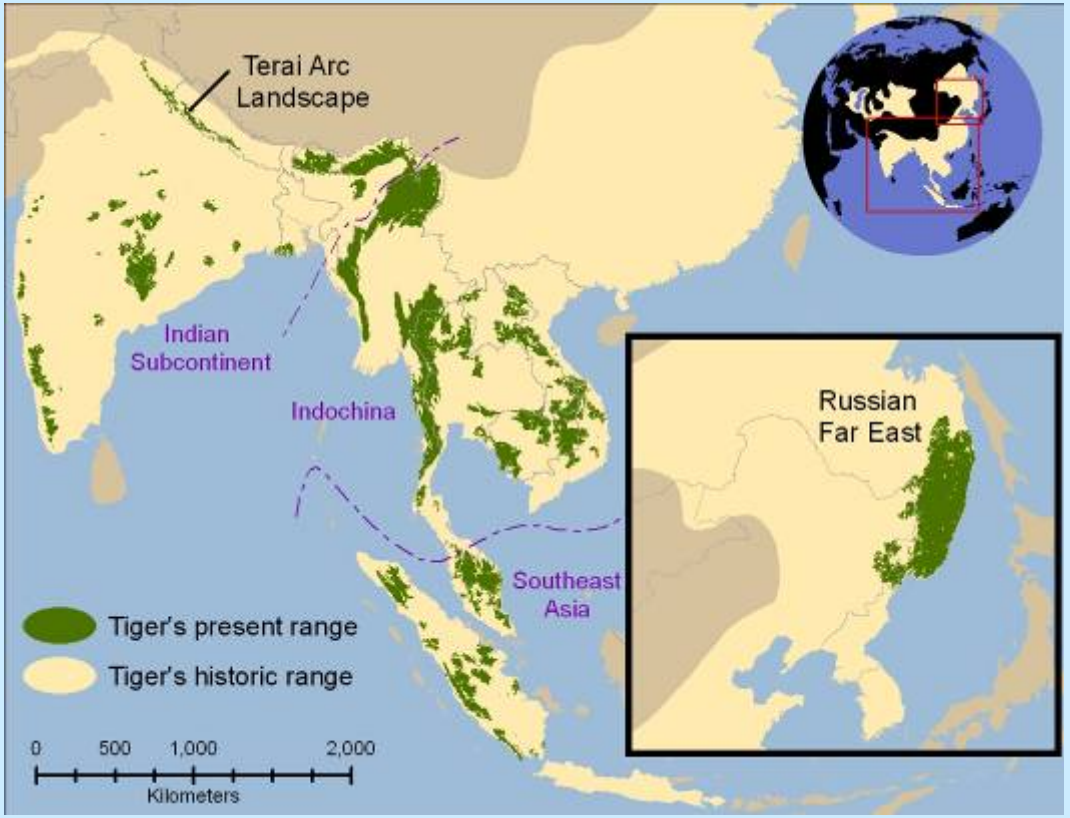
Sino-Tibetan fauna

- There are several species of bear in the range of Sino-Tibetan but only one which covers the entire area, the Asian black bear, *Ursus thibetanus*
- This is probably the base referent of the root *#tom*, which is attested in every Sino-Tibetan branch (I think)



Sino-Tibetan fauna: the tiger

- The tiger is also spread across the range of Sino-Tibetan languages and beyond into Siberia and ISEA
- Sino-Tibetan has two major roots for 'tiger', *d.key* and *#kV.la*, the latter widespread in all four regional language phyla
- This probably points to the ritual importance of the tiger in SE Asian culture



Sino-Tibetan agriculture I?

- One of the problems with assuming that early Sino-Tibetan speakers were a Neolithic society is a complete inability to reconstruct settled agriculture linguistically
- The current consensus is that millets were domesticated in North China around 9000 BP but that rice is managed in the Yangtse Valley until 6500 BP and only then properly domesticated
- And that whoever these people were, they were *not* Sino-Tibetan speakers
- ‘Grain-growing agriculture was the basis of early Sinitic and every large early TB society; however, the actual words for these grain crops do not show clear and regular cognates across the full range of ST and TB languages’. (Bradely 2011)
- .

Sino-Tibetan agriculture II?

- Ethnographic evidence suggests that in some areas the transition to farming might be quite late in some regions, with hunting and gathering remaining an important element in subsistence until recently
- In particular the ‘small’ branches of Sino-Tibetan show no cognacy at all.
- Suggesting that the cereal-growing societies are a significantly later development
- Archaeobotany has recently indicated that the millets (foxtail and broomcorn) spread to the foothills of the Tibetan Plateau around 5500 BP, i.e. a great deal later than first domesticated in North China (ca. 8-10000 BP)
- So, millet is probably a chain of borrowing, spreading south and west and not to be reconstructed.

Thinking about NE India

- From the Palaeolithic onwards the region must have been inhabited by highly diverse hunter-gatherers. These would undoubtedly have spoken comparably diverse languages, which have largely disappeared today, although evidence for them may survive as substrates in existing languages.
- Only in Arunachal Pradesh, where many languages are difficult to classify, such as Puroik, Mey, Bugun, Koro, Hruso and Miji, are there probable survivals from this period.
- Elsewhere, such as in the Khasi Hills and the Assam plains, the subsequent expansion of incoming populations has eliminated the traces of the languages of foragers.
- In addition, in Arunachal Pradesh we find evidence that even populations who farm today, such as the Puroik and Milang, remained partly dependent on semi-wild plants, such as the sago palm and the tree-fern, until recent times.

Ethnolinguistic map of NE India

Sago processors

- Ethnographic accounts of populations such as the Puroik (Sulung) suggest that they *are* still largely hunters and sago-exploiters and the Milang were until 1 or 2 generations ago .

Milang washing sago log

Mithuns

- The mithun (*Bos frontalis*) is the characteristic bovid in NE India, the most prized cultural species. Linguistically it is the only species which is embedded, with the same lexeme across many languages. Words for all other livestock species are derived from it.

Vegetative crops, plantains and taro

- Plantains are both widely grown and semi-wild types exploited throughout the region, along with yams and taro

Sino-Tibetan expansion: a new model I

- The earliest speakers of Sino-Tibetan were highly diverse foragers living in an arc between the eastern slopes of the Himalayas and regional lowland jungles up to 9,000 years ago and practising arboriculture (sago)
- Some spoke early Sino-Tibetan languages, others unknown languages now present only as substrates and perhaps surviving as Kusunda

- Seasonal foragers exploit the high Tibetan Plateau from 7500 BP
- Perhaps 6-5000 BP 'livestock revolution' takes place in the mid-level Himalayas. Yak herders move up and settle the Tibetan Plateau permanently.

Sino-Tibetan expansion: a new model II

- Gathering of wild cereals (buckwheat etc.) and tubers (high-altitude taro) leads to proto-agriculture in the mid-level Himalayas
- Foragers who will become the Naga complex began to practise vegeculture (taro, plantains) (NE India) and animal management (mithun) by 6000 BP possibly, through contact with Austroasiatic speakers
- By 5000 BP diverse early Sino-Tibetan groups in the Himalayas begin spreading eastwards to China. Sinitic is *not* a primary branch, but simply the language of one of many migratory groups
- Proto-Tujia, proto-Bai and probably others meet unknown populations (Hmong-Mienic? Austronesians?) with domestic pigs, millet, while also cultivating and beginning to domesticate rice

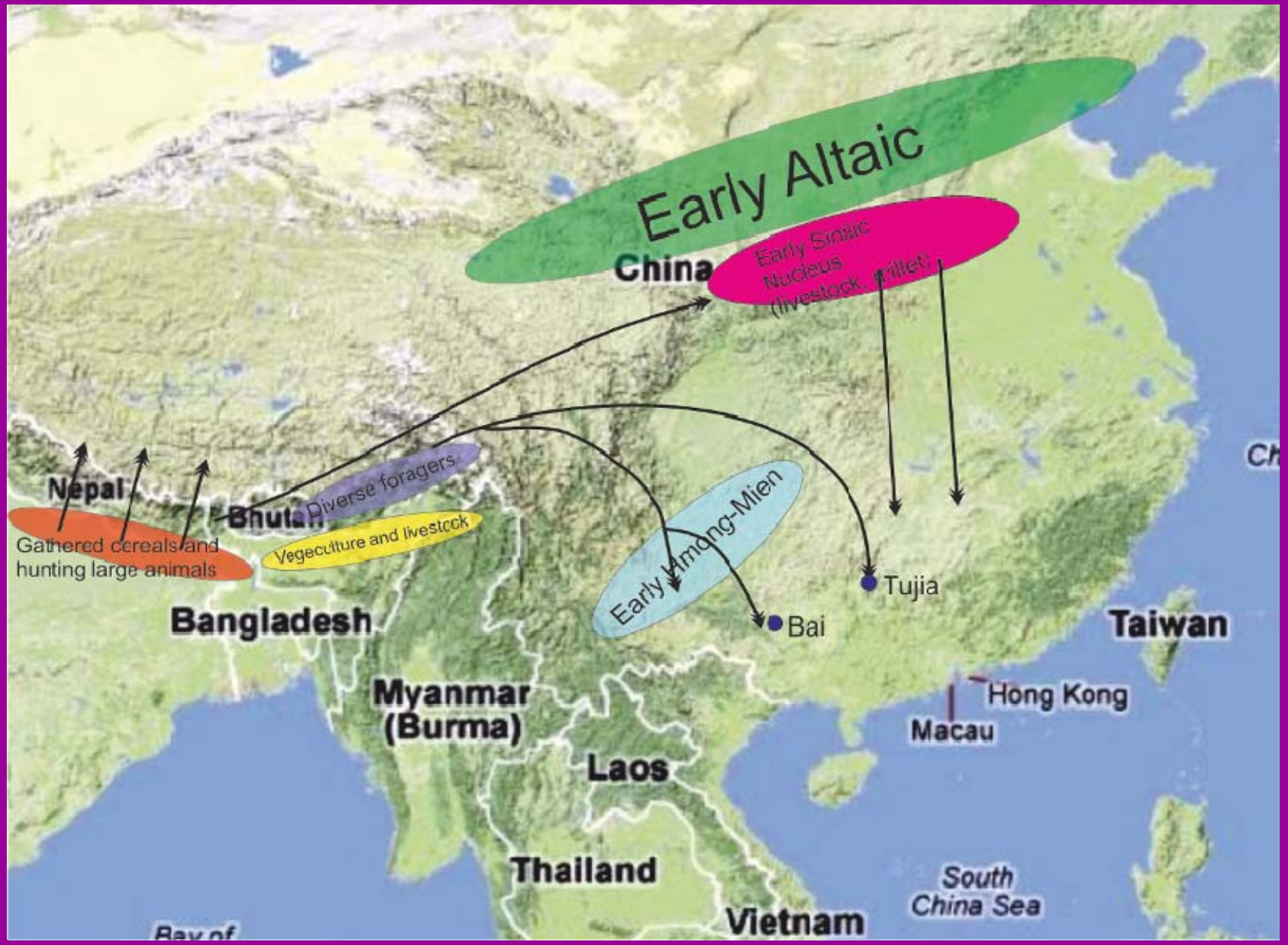
Sino-Tibetan expansion: a new model III

- The Sinitic languages expand southwards, assimilating or encapsulating many small groups. They encounter Hmong-Mien speakers with rice and switch millet terminology to rice
- Rice moves up from India but also westwards from China (hence hybridised types) and overlays older cereals where ecologically possible
- Ruminants (cows, sheep, goats) spread downwards into China from Central Asia 4400 BP (? Altaic for small ruminants but not cattle)

Sino-Tibetan expansion: a new model IV

- Tibetic speakers undergo a major expansion (when? maybe very late ca. 800 AD) assimilating linguistic diversity on the Plateau
- Rice invades the lowland vegecultural zones rather later, pushing taro into residual systems
- Groups such as early Burmic spread southwards, fragmenting Austroasiatic-speaking peoples

Mapping the Sino-Tibetan Expansion



Some conclusions I

- Sino-Tibetan languages are *relatively* well-documented, lexically at least and a good representation of the data is available on the internet
- Grammatical systems far less so, but progress is being made
- Nonetheless, arguments about the coherence of the phylum and its internal classification are in chaos
- A large number of completely untrustworthy 'reconstructions' are in the literature
- We can't make any sense of this unless we can develop a more sophisticated view of borrowing and cognacy
- And be prepared to listen to the findings from other disciplines, especially archaeology
- It also suggests that Sino-Tibetan is highly unsuitable for the mathematical phylogenies currently being promoted

Some conclusions II

- We also need to develop a ‘gold standard’ for membership of Sino-Tibetan and include or exclude language based on evidence not geography
- Given all this, if we can clean the data of a great accumulated weight of preconceptions, the inappropriate focus on major languages, archaeologically improbable assertions..
- Then we will be able to see the phylum more clearly and begin the process anew
- And if you take a dynamic view of grammar, this type of history also allows you to reconstruct contact scenarios and explain better the facts of the present

What does this mean for other phyla?

- The linguistic palaeontology approach began with Indo-European and is still dominant, despite encroachments from other methods
- Indo-European suffers from some of the same problems as Sino-Tibetan, inappropriate focus on written texts etc.
- Austronesian is in many ways the best type example of this approach where high-quality linguistics is combined with trustworthy archaeology
- Afroasiatic ('Hamito-Semitic') is probably the worst case, where an obsession with 'high' languages (scriptural), a confusion with racial issues and so on has led to bad linguistics and worse correlations with archaeology

Palaeosociolinguistics I

- There is plenty of evidence for the interface between social and political structures and language change in the present
- There is no reason to think anything was different in the past; the same processes apply
- The difference is that in the past we can't necessarily see processes as they occur, so we have to deduce them retrospectively from the state of the language at present
- Palaeosociolinguistics is the hypothetical modelling of language change as affected by social process.
- For example, many language phyla have undergone major bottlenecks or levelling episodes.
- For example, Sinitic, Hmong-Mien, Berber, Malagasy
- Hence, paradoxically, Old Chinese is not a representation of proto-Sinitic, which undergoes a bottleneck at the end of the Warring States period

Palaeosociolinguistics II

- Only the careful analysis of individual language relations can help unravel these processes
- Similarly, we can assume that technical innovations play an essential role in the expansion of language phyla.
- For example, the introduction of the Asian composite bow across the Bering Straits some 2000 years ago seems to have powered the Athabaskan expansion which revolutionised the linguistic geography of North America.
- And innovative watercraft the remarkable expansion of the Austronesian-speaking peoples across the Pacific
- Exploring these topics provides insights into key topics such as langugage diversification and change
- Not to be encapsulated by comparing lists of 200 words of basic vocabulary

Is it relevant for other areas of linguistics?

- Obviously this approach is relevant for historical linguistics. I would argue, however, that it is also relevant for the analysis of synchronic grammar.
- One of the discoveries of contemporary linguistics is the lack of consonance between different subsystems in the classification of languages. Lexicon, phonology and morphosyntax do not necessarily tell the same story.
- The reason is almost certainly complex patterns of contact and interaction, as well as certain types of social structure
- French linguists used to call this 'ethnolinguistics' before the syntax steamroller came and flattened out the tales we tell about language
- Culture and language are inextricably intertwined…..

THANKS

- Thanks to Stephen Morey for the invitation to Melbourne
- And to my many informants in Northeast India