

Part IV

*Origins and diversification:
the case of Austroasiatic groups*

14 *The Austroasiatic Urheimat: the Southeastern Riverine Hypothesis*

PAUL SIDWELL & ROGER BLENCH

1 Introduction¹

The Austroasiatic language phylum is situated in the heartland of MSEA and yet today is remarkably fragmented, its individual branches scattered from Northeast India to the Malay Peninsula. Interwoven territorially with much more geographically coherent phyla such as Daic and Hmong-Mien, the narrative of its dispersal is central to our general understanding of the ethno-cultural history of Southeast Asia. Although comparative Austroasiatic linguistics is now more than a century old, limited progress has made towards a consensus on the homeland or *Urheimat* of Austroasiatic languages. The Austroasiatic phylum is generally considered to be the oldest identifiable language grouping of that region (excluding perhaps Andamanese). A model for its origins and migration paths that could account for the present distribution of the languages is crucial for the linguistic history and ethnography of Southeast Asia. However, the linguistic literature relating to this too often presents confident claims that invoke unpublished materials, paying little heed to evaluating alternative hypotheses. Published studies are not always transparent, especially problematic when they lack adequate data that readers might assess and analyse for themselves.

In this chapter we focus on linguistic arguments for a likely Austroasiatic homeland, and possible correlations with the—still emerging—archaeological record. The orientation of the chapter is linguistic; we assume no necessary equation between linguistic entities and archaeological assemblages or genetic profiles. It is evident that the ethno-history of Southeast Asia has often involved multilingualism and various radical language shifts among communities large and small, and this must also have occurred among prehistoric communities, especially in the context of the early expansion of agriculture. Consequently, when we talk about a linguistic homeland or *Urheimat*, we do not wish to imply that the cultural complex which radiated from that centre necessarily largely *originated* in that location, only that it began to diversify and spread from there. In other words, it is the last location in which the speaker community presented a linguistic unity.

Among the various suggestions for the Austroasiatic centre of dispersal offered over the years, there are three broad trends:

¹ Acronyms and conventions used in this chapter: C (any consonant), V (any vowel), # (quasi-reconstruction, i.e. form based on rapid inspection of cognates), MSEA (Mainland Southeast Asia)

N. J. Enfield, editor.

Dynamics of Human Diversity, 317-345. Pacific Linguistics, 2011

© This edition vested with Pacific Linguistics.

1. A western origin, in northeastern India or in the vicinity of the Bay of Bengal (Van Driem 2001)
2. A northern origin, in central or southern China (Norman & Mei 1976)
3. A central origin, within Southeast Asia (Von Heine-Geldern 1923)

None of these proposals has been supported by sufficient evidence, and there has been a broad failure to correlate the antiquity and location of the homeland of Proto-Austroasiatic with known archaeological data. It is taken as a given that the reconstructibility of lexical items in Austroasiatic related to subsistence must be congruent with current understanding of the archaeology, ecology and palaeoclimatology of the region. This chapter will argue that:

- a) None of the internal sub-families proposed for Austroasiatic have been demonstrated unambiguously and that the provisional model for its internal structure must be a largely flat array.
- b) The widely-accepted division between Muṅḍā and the remainder of Mon-Khmer is spurious. Although Muṅḍā shows significant typological differences from other Austroasiatic languages, this cannot be taken as evidence for the antiquity or primacy of the split.
- c) A flat array in turn points to the diversification of a dialect chain and therefore implies a relatively younger age (ca. 4000 BP) for Austroasiatic than is usually advanced.
- d) The proto-Austroasiatic lexicon has terms for crops requiring a humid climate (taro and rice) as well as other items (boats, river fauna) which suggest an aquatic/riverine environment.
- e) Congruence between these elements can be achieved with the assumption that early Austroasiatic initially diverged somewhere in the Middle Mekong, and its initial dispersal was along river valleys, exploiting both aquatic resources and humid soils.

Ultimately we suggest that various facts concerning Austroasiatic languages can well be explained by convergence and contact that continued after a relatively late break-up into distinct branches. We propose a model to correlate the linguistic results with evidence from archaeology and anthropology, which we call the Southeastern Riverine Hypothesis. This model has precursors, first advanced nearly a century ago. Von Heine-Geldern (1923), anthropologist and student of Wilhelm Schmidt (founder of comparative Austroasiatic studies) advanced a *Kulturkreise* theory in the 1920s which modelled the dispersal of Austroasiatic out of Southeast Asia. In the 1970s, on the basis of lexicostatistical studies (for example Thomas 1973, Huffman 1978), it was suggested that the phylum dispersed from Indo-China/Northeast Thailand as recently perhaps as 3800 BP.

2 General considerations

The three basic proposals for the Austroasiatic homeland have been advanced on the bases of the following types of argument:

- a) lexical isoglosses and/or typological affinities interpreted as indicating contact, and thus proximity, of otherwise unrelated language families

- b) correlation of lexical reconstructions with archaeological facts/hypotheses or features of the natural world ('paleo-linguistics')
- c) 'centre-of-gravity' arguments that identify supposed zones of higher linguistic diversity with greater time depth

Lexical and/or structural similarities between Austroasiatic and other languages of insular and mainland Asia are readily found (Enfield, this volume). This has been the source of a series of genetic hypotheses relating it to other phyla. For example, Shorto (1976a) argued that Austroasiatic-Austronesian comparisons demonstrate the validity of Austric, a hypothesis which goes back at least to Schmidt (1905). Blench (this volume) discusses the history of various macrophyla hypotheses in more detail, but we believe that none of them has been reliably demonstrated and they will not be further considered here.

Another source of quasi-genealogical hypotheses has been extensive lexical borrowing between phyla. It has long been argued that Austroasiatic was once much more widespread in China and was driven south by the expansion of the Han (Norman & Mei 1976). It is claimed that some names of zodiacal animals, and the Old Chinese words for 'river' and 'tiger' are borrowings (Norman 1988:18). Schuessler (2007) expanded this in his *Etymological dictionary of Old Chinese* which asserts that Austroasiatic forms underlie many Sinitic etyma. The subtext is a general identification of northern regions as the homeland of Austroasiatic, but these authors fail to note that many roots are typically widespread in the region occurring in multiple phyla. For example, tigers were historically common across the region, although today they are confined to a few small reserve areas. Tigers have a crucial role in spiritual beliefs of many peoples, which may account for the distribution across language phyla of a key lexeme, #kVla. Table 1 shows the regional reflexes of #kVla.

It is clear that the name for 'tiger' has been freely borrowed between phyla and is apparently ancient in Sino-Tibetan, Daic and Austroasiatic. Such words cannot be used in genealogical classifications and certainly not in arguments about the location of homelands.

Along similar lines, it has been widely claimed that, especially since Norman & Mei (1976), the name of the Yangtze itself is of Austroasiatic origin. This is based on the casual resemblance between Old Chinese *k^hroŋ² and Austroasiatic forms such as Old Mon *kruñ* /kruŋ/ 'river'. While suggestive, no compelling reasons have been put forward to show that this is anything other than a lexical coincidence, such as one may find comparing any pair or group of languages. In a similar vein Van Driem (2001:290) reminds us that 'Toponyms and especially river names [...] have suggested to researchers such as Hermann Berger and Manfred Mayrhofer that Austroasiatic is an old ethnic substrate in the north of the Indian subcontinent.' Despite these claims, no specific comparisons have yielded a decisive body of isoglosses. In the Indian context, it is often no more than the claim that various Indic words appear to show prefixes, which is also an Austroasiatic characteristic, although without even suggesting that the forms have specific Austroasiatic cognates (see for example Witzel 1999).

2 As presently reconstructed by Baxter and Sagart, see:

http://sitemaker.umich.edu/wbaxter/old_chinese_reconstructions&record

Table 1: The #kVla root for ‘tiger’ in Southeast Asian languages

Phylum	Branch	Language	Attestation
Sino-Tibetan	Sinitic	OCM	*hlâ?
Sino-Tibetan	Burmic	Old Burmese	klya
Sino-Tibetan	Loloish	Mantsi Meo Vac	qua ⁵³
Sino-Tibetan	Luish	Cak	ka θa
Sino-Tibetan	Luish	Lushai	sa kei
Sino-Tibetan	Bodish	Monpa	khai-la
Sino-Tibetan	Naga	Mao Naga	okhe
Austroasiatic	Palaungic	Shinman	ka? ⁴ vai ³
Austroasiatic	Khmeric	Angkorian Khmer	khlaa
Austroasiatic	Bahnaric	Sedang	klá
Austroasiatic	Katuic	Pacoh	kulaa
Austroasiatic	Katuic	Ir	kala?
Austroasiatic	Katuic	So	kula
Austroasiatic	Monic	Proto-Monic	*klaa?
Austroasiatic	Aslian	Sakai	kla
Austroasiatic	Khasian	War Jaintia	k ^h la
Austroasiatic	Munḍā	Munḍā	kula
Daic	Tai	Thai	-k ^h aan (Year of ~)

Scholars have also cited characteristics of the palaeo-environment when seeking the likely homeland. Diffloth (2005) observed that since names for various tropical species can be reconstructed for the proto-language, a high humidity ecology is indicated. Hence the suggestion of the shores of the Bay of Bengal urged by Van Driem (2001:290). Yet Sagart (2008, citing Chang Kwang-chih 1986) notes that the mid-Holocene climate of central China was 2° to 5° warmer than today. Consequently, the potential zone of tropical flora and fauna likely encompassed any and all proposed Austroasiatic homelands. By contrast, Peiros (1998) and Peiros & Shnirelman (1998) assert that the Austroasiatic lexicon indicates a non-tropical, non-coastal location, but present no evidence to support such a claim.

There is also a related tendency to assert that Austroasiatic is of great antiquity. Peiros (2004) offers a glottochronological calculation of 8300 BP for the initial branching. Diffloth (2005) proposes approximately 7000 BP, and likewise Blust (1996:132) estimates that ‘By 7000–7500 BP, PAA had separated into western (Munḍā) and eastern (Mon-Khmer) dialect areas’. Now it seems the antiquity of Austroasiatic falls into the category of facts so well known it can be repeated without qualification or explanation (for example Van Driem 2007:10). Since for each of these authors the homeland was located somewhere significantly removed from where the majority of Austroasiatic languages are spoken today, by necessity a long period of time must have elapsed for the languages to disperse over a wide area, and for others (Sinitic, Indo-Aryan, etc.) to occupy the original site.

A common feature of the diverse claims for, variously, China or South Asian origins, is that they are vague; failing to make predictions that may be readily tested. Similar types of evidence, such as resemblances in place names, are invoked to support mutually incompatible models. In response we claim that it is better to begin with reviewing what we know about the Austroasiatic languages, and seek the simplest explanation (the ‘fewest moves’) which may account for those facts. Such a model should be the starting point for discussion of Austroasiatic linguistic origins.

3 The Received Classification of Austroasiatic Languages

Current thinking about the classification of the branches of Austroasiatic can be traced to the comparative and typological studies of Pinnow (1959, 1960, 1963). While earlier studies (for example Grierson 1919; Przyluski 1924) treated Muṅḍā ('Kolarian') as a distinct division, scholars in the latter part of the 20th century relied directly upon Pinnow for making explicit the notion of a binary split between Muṅḍā and Mon-Khmer. Pinnow's (1963:278) scheme of a Western group (Nahali-Muṅḍā) and Eastern Group (Khmer-Nicobar) is reproduced in Figure 1.

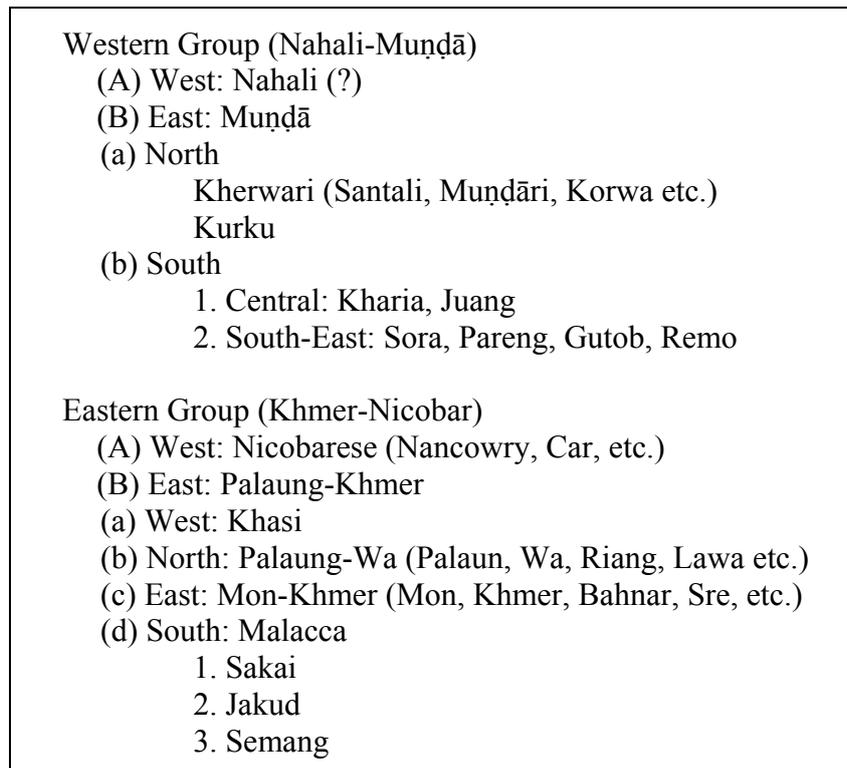


Figure 1: Pinnow's Austroasiatic classification.

Nahali is now recognised by almost all scholars to be a language isolate, albeit one that has come under Muṅḍā (Korku) influence (Blench 2008).

Pinnow (1963:150) recognised the limitations of the typological approach as a means of discerning historical relations. He was not especially confident that the 'Khmer-Nicobar' languages formed a unity in the same way as Muṅḍā, and stated that they may even be historically 'independent of one another and traceable solely to Proto-Austroasiatic'. In this case Khmer-Nicobar 'would have only structural and geographical justification'. But this caveat was almost universally ignored in favour of a genealogical reading. Before Pinnow, there had been no clear guide to the likely structure of the family; suddenly we could speak with confidence and authority about Austroasiatic, and a new orthodoxy was adopted. Since then a primary Muṅḍā Mon-Khmer split has largely been assumed. For example, recently we find: 'The primary split in the family is between the Muṅḍā languages in central and eastern India and the rest of the family.' (Anderson 2008:598) And with a twist: 'The Austroasiatic language family is conventionally divided into three branches or sub-families, viz. the Muṅḍā, the Nicobarese and the Mon-Khmer languages.' (Van Driem 2001:262).

Programmatically the desirable approach would have been to build on Pinnow (1959) and other comparative studies (Schmidt 1904, 1905, Skeat & Blagden 1906, Schafer 1952, 1965, Haudricourt 1965 etc.) to produce a working model of proto-Austroasiatic in the 1960s–70s, so that a classification based upon shared innovations could have been discussed. Shorto attempted this, but only published a fragment at the time (Shorto 1976b).³ Diffloth began the task but has not released his results. Consequently the field turned to other strategies, specifically lexicostatistics. Lexicostatistics is a widely used heuristic for revealing likely language relationships, and is especially favoured when one is more or less restricted to using lexical data (cf. Burenhult, Dunn and Kruspe this volume). The method counts shared forms on a limited basic wordlist to produce a crude index of similarity. The method has been strongly criticised, especially in respect of studies that have placed a high value on lexicostatistical results for creating family trees, and the present writers share those concerns.

Notwithstanding the poor image and well understood limitations for the method, it was lexicostatistical studies that distinguished the dozen or so Austroasiatic branches recognised today, and confirmed by subsequent comparative analyses. The most important early lexicostatistical study was by Thomas & Headley (1970). They concluded (p. 405): ‘The Austroasiatic phylum would appear to be composed of at least four families: Muṅḍā, Mon-Khmer, Malacca, Nicobarese.’

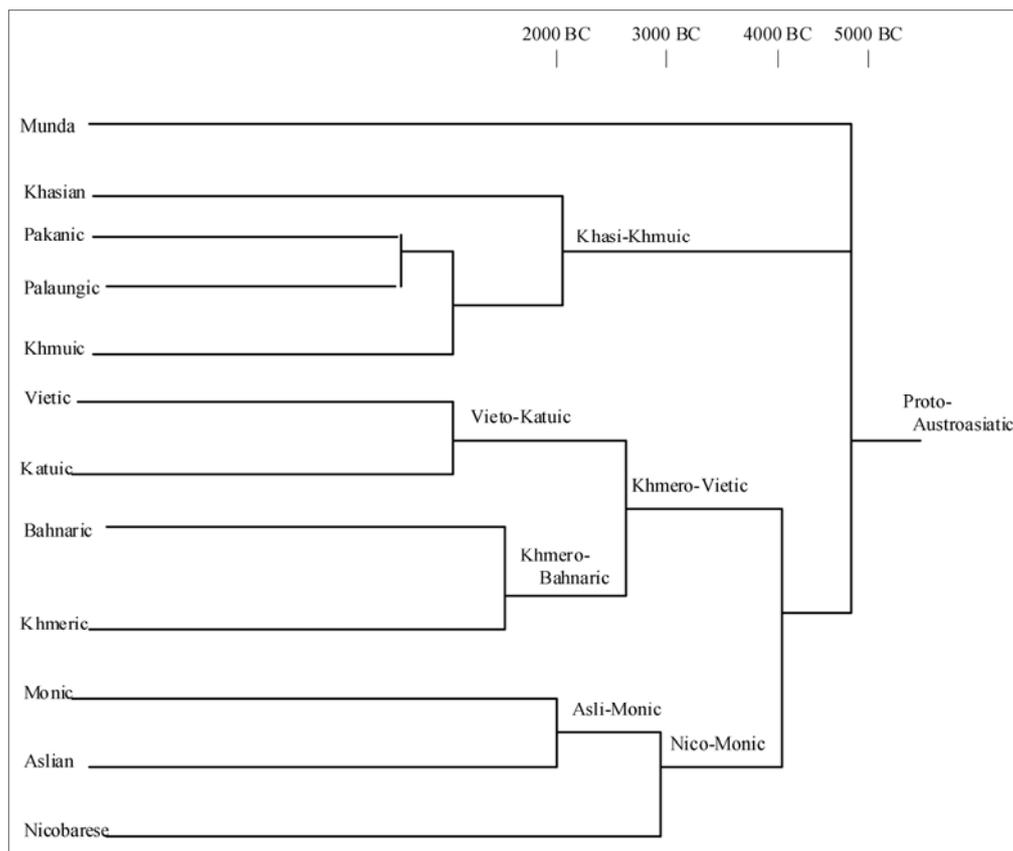


Figure 2: Diffloth's model of Austroasiatic (adapted from Diffloth 2005).

3 Shorto pursued a comprehensive comparative reconstruction, which was published posthumously in 2006. In the course of that effort he was unable to justify any nested sub-grouping of AA branches by historical phonology, and instead conducted several lexicostatistical investigations (at least four attempts) which are discussed in Sidwell (2009).

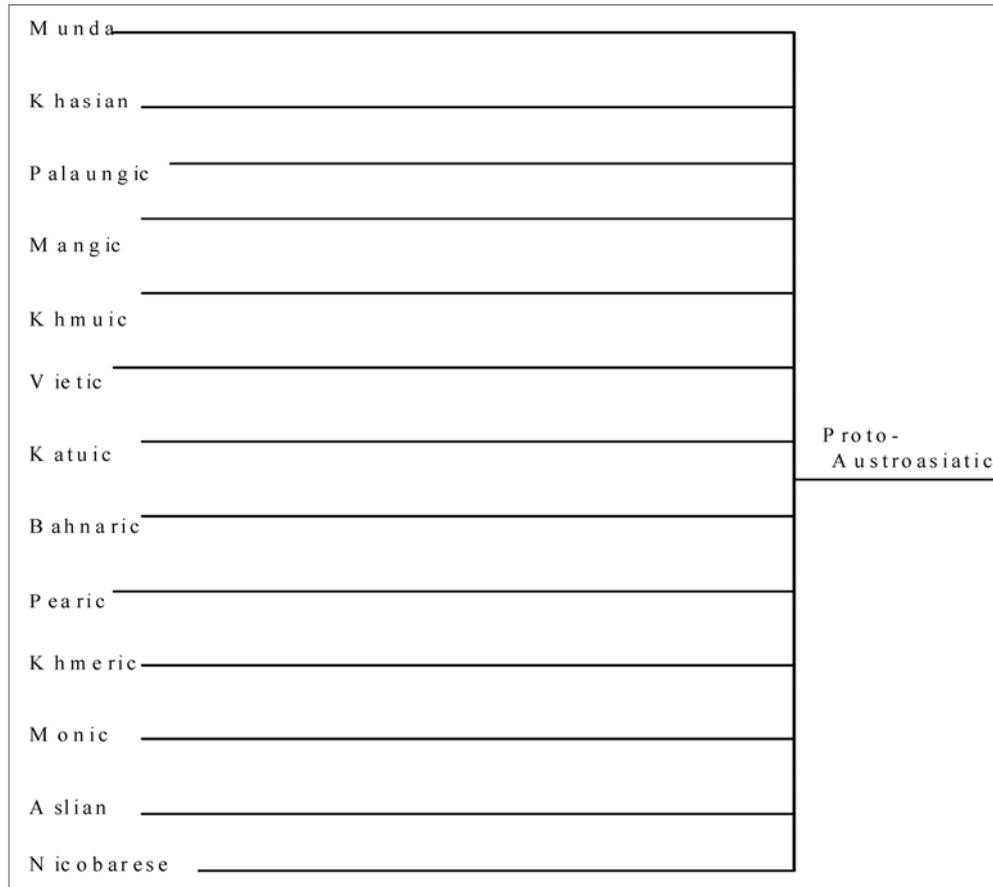


Figure 3: Flat-array structure for Austroasiatic (Sidwell 2008).

Concretely, the separation of Malacca (Aslian) and Nicobarese from Mon-Khmer was shown by ~~a~~-cognate scores of 11~16% between Temiar and Mon-Khmer, and 6~12% between Nicobarese and Mon-Khmer. Above these scores varying between 18% and 35% were taken as indicating nine distinct Mon-Khmer branches: Pearic, Khmer, Bahnaric, Katuic, Khmuic, Monic, Palaungic, Khasi, and Viet-Muong. Another widely cited source (Parkin 1991:6) presents Thomas & Headley’s scheme, quoting their lexicostatistical figures, but credits it to Diffloth, saying, ‘This breakdown is based mainly on Diffloth, though he included Aslian in with Mon-Khmer.’ However, a general view later emerged that the very low percentages found for Aslian and Nicobarese are caused by special factors, and that these are Mon-Khmer groups isolated from the rest of the family. The notion that a similar explanation might also account for Muṅḍā has also surfaced recently.

The present situation is that there are two main competing classifications of Austroasiatic, that given by Diffloth (2005) which can be contrasted with Sidwell (2008) shown in Figure 3. Figure 2 shows a version of Diffloth (2005) with sub-branches omitted so that it can be directly compared with Figure 3. Diffloth proposed dates attached to individual nodes, but these are only applicable to the earlier period once individual sub-branches are merged.

	Mundā		Khasian		Palaungic		Khmuic		Vietic		Katuic		Bahnaric		Monic		Pearic		Khmeric		Aslian		Nicobarese				
	Mun	Sor	Khs	War	Dea	Wa	Kmu	Mal	Muo	Ruc	Bru	Nge	Kat	Bah	Jeh	Jru	Mon	Nya	Cho	Pea	Sur	Cam	Sem	Sml	Jah	Car	Nan
Mundāri	37		18	13	15	17	18	22	21	23	25	24	22	17	20	23	24	23	18	20	16	16	16	16	11	10	15
Sora		37	14	10	13	14	20	20	21	24	23	19	19	17	18	18	21	20	11	15	12	13	17	13	12	8	12
Khasi	18	14		60	23	20	22	26	22	21	24	24	23	21	23	22	21	20	19	20	20	21	20	15	16	13	16
War-J	13	10	60		17	16	16	24	14	14	19	21	17	13	16	16	15	16	16	18	17	17	17	14	10	8	10
De'ang	15	13	23	17		53	26	20	15	18	19	22	22	19	24	17	23	24	16	19	17	18	20	14	14	12	14
Wa	17	14	20	16	53		26	22	16	19	26	26	26	20	26	21	23	26	22	27	17	18	21	18	16	13	18
Khmu'	18	20	22	16	26	26		35	17	22	27	26	24	20	21	21	21	23	21	25	18	19	21	16	15	12	14
Mal	22	20	26	24	20	22	35		18	23	31	31	28	22	27	22	21	24	19	25	21	21	22	20	17	12	13
Muong	21	21	22	14	15	16	17	18		50	27	23	25	24	22	25	22	22	15	18	14	14	20	16	14	11	14
Ruc	23	24	21	14	18	19	22	23	50		30	29	27	26	27	30	23	23	17	20	19	20	20	18	14	14	17
Bru	25	23	24	19	19	19	26	27	31	30		57	57	41	40	43	28	31	30	34	23	24	29	28	20	17	21
Nge'	24	19	24	21	22	26	26	31	23	29	57		61	38	36	38	30	33	30	34	27	26	25	31	23	18	20
Katu	22	19	23	17	22	26	24	28	25	27	57	61		43	39	40	29	30	30	34	26	26	28	28	20	16	19
Bahnar	17	17	21	13	19	20	20	22	24	26	41	38	43		52	51	23	26	30	34	24	22	29	30	18	13	19
Jeh	20	18	23	16	24	26	21	27	22	27	40	36	39	52		55	23	24	27	27	22	23	31	31	23	14	18
Jru'	23	18	22	16	17	21	21	22	25	30	43	38	40	51	55		27	29	29	33	27	26	27	30	21	15	19
Mon	24	21	21	16	23	23	21	21	22	23	28	30	29	23	23	27	84	23	28	24	23	24	27	23	18	15	18
Nyahkur	23	20	20	15	24	26	23	24	22	23	31	33	30	26	24	29	84	23	28	24	23	27	27	24	19	15	17
Chong-W	18	11	19	16	16	22	21	19	15	17	30	30	30	30	27	29	23	23	61	22	20	22	22	22	11	14	18
Pear-KPT	20	15	20	18	19	27	25	25	18	20	34	34	34	34	27	33	28	28	61	30	28	27	29	18	16	22	
Surin	16	12	20	17	17	17	18	21	14	19	23	27	26	24	22	27	24	24	22	30	89	22	21	15	12	14	
Cambodian	16	13	21	17	18	18	19	21	12	20	24	26	26	22	23	26	23	23	20	28	89	20	19	17	13	15	
Semai	16	17	20	17	20	21	21	22	20	20	29	25	28	29	31	27	27	27	22	27	22	20	38	36	14	19	
Semelai	16	13	15	14	14	18	16	20	16	18	28	31	28	30	31	30	23	24	22	29	21	19	38	28	11	16	
Jahai	11	12	16	10	14	16	15	17	14	14	20	23	20	18	23	21	18	19	11	18	15	17	36	28	12	14	
Car	10	8	13	8	12	13	12	12	11	14	17	18	16	13	14	15	15	15	14	16	12	13	14	11	12	49	
Nancowry	16	12	16	10	14	18	14	13	14	17	21	20	19	19	18	19	18	17	18	22	14	15	19	16	14	49	

Figure 4: Austroasiatic lexicostatistical matrix (Sidwell 2009)

Diffloth has not offered a general rationale for his scheme, only fragments. For example, he (personal communication⁴) has suggested that the metathesis of proto-Austroasiatic **pti(i)s* ‘mushroom’ to Khmer *psət*, Bahnaric, for example Sre *bsit*, and Pearic, for example Chong *psiit*, must be a unique historical accident indicative of common ancestry. But subgrouping arguments that hinge on a single form rather than a collection of common but unrelated events must admit various possible explanations. After all, if Khmer, Bahnaric and Pearic did share a unique period of common development, we would reasonably expect to find indications in the core vocabulary, much as can be found, for example, in respect of Khasi-Palaungic (discussed below). But one or even several isoglosses found amongst thousands of comparanda cannot be a convincing sub-grouping argument.

A challenge to Diffloth’s model has been put forward by Sidwell (2008), who argues instead for a flat array, rejecting not only the Muṅḍā/Mon-Khmer split, but other proposed internal nodes. This is based on a review of the phonological and lexical correspondences at branch level, which failed to find innovations that would justify the kind of nested sub-branching shown in the model in Figure 2. Sidwell’s revised model is shown in Figure 3 with the addition of Mangic (Pakanic in Diffloth’s terminology), now also considered a distinct branch.

Blench (in press a) argues that the language of the Shom Pen, foragers in the Nicobar Islands, may also constitute a separate branch of Austroasiatic, but this has yet to be assessed by the scholarly community and will not be included in this analysis.

The flat array suggests that Austroasiatic first fanned out into a dialect chain in which neighbours would have remained intercomprehensible while cumulative differences would have been greatest at the geographical extremes. Such a model has implications for both the dating of Austroasiatic and its likely directions of spread.

4 Sidwell’s Lexicostatistics

Having failed to identify phonological innovations that would support Diffloth’s nested branching model, Sidwell revisited the lexicostatistical analysis of Austroasiatic. There were 28 languages compared using the standard Swadesh 100 list and a matrix (Figure 4) generated using the programme *glottpc.exe*⁵. The results show remarkable similarities to those of Huffman (1978) in terms of the relative distributions of percentages. The biggest difference is in the direct Katuic-Bahnaric comparison, with an average cognacy of only 40%, rather than the 47% found by Huffman. This is still high, and all other things being equal, it would be consistent with a Katuic-Bahnaric sub-grouping. Yet comparative reconstruction makes it clear that Katuic and Bahnaric do not sub-group (see discussion in section 8, below).

Both Sidwell’s and Huffman’s figures show a remarkable pattern, highlighted in the shaded parts of Figure 4. Rather than inter-branch percentages indicating neat patterns of nested sub-branching, distinct branches show elevated scores with Katuic-Bahnaric, but a much flatter spread of scores if Katuic-Bahnaric is removed from consideration. The common factor is geography, with scores declining as one gets further afield from the middle Mekong. There is also a weaker but discernibly similar effect between Mon and Nyah Kur and the rest of Austroasiatic. Among the Aslian languages it is the southern sub-

4 Diffloth also briefly discussed this at the 2008 SEALS meeting in Malaysia.

5 The data, cognate assignments, and computational analyses of this and other lexicostatistical studies by Sidwell can be retrieved online at: <http://people.anu.edu.au/~u9907217/lexico>.

branch that show more cognates with Katuic-Bahnaric (and to the lesser extent Monic). For example, Semai and Semelai show 31% agreement with Jeh (Bahnaric), yet within Aslian a lower score (28%) is counted between Jahai (North Aslian) and Semelai (South Aslian).

How are these figures to be interpreted? Since lexicostatistical methods were first pioneered by Swadesh in the early 1950s (for example 1950, 1952 and *passim*:-) they have been consistently savaged by critics who have focussed on unreliability in sub-grouping and dating results (for example Hoijer 1956, Bergsland & Vogt 1962, Holm 2003 and others). Those criticisms focus mainly on the fallacy of a constant rate of change, and the problem of distinguishing inherited from borrowed vocabulary, which we acknowledge here. We offer these lexicostatistical results specifically as a heuristic for diagnosing and investigating borrowing between languages, an otherwise crucial but under-investigated aspect of the Austroasiatic homeland question. We do not attempt to construct a traditional lexicostatistical tree, but instead compare the differences in cognate percentages with the indications of historical phonology. Our underlying assumption is that the latter is necessarily the gold standard for determining branching relations, especially in the present case where it is not possible to compare, for example, morphological paradigms of the complexity found in, say, Indo-European. It is clear that our present results, similar to Huffman's results of a generation before, show a clear geographical correlation: cognate percentages are generally higher when compared to languages located in the middle Mekong zone, and higher the closer one gets to that zone. Our interpretation is that the figures are so contaminated by borrowings it is unsafe to posit anything other than roughly equidistant branches on lexical grounds.

At Sidwell's request the lexicostatistical matrix for 28 AA languages was subjected to a Bayesian phylogenetic analysis by Russell Gray and Simon Greenhill at the University of Auckland. Subsequently they calculated the Neighbor-Net tree reproduced here in Figure 5. The method is different in kind to a traditional lexicostatistical analysis, and is widely used today in biology and genetics in particular, and increasingly in linguistics (cf. Burenhult, Dunn & Kruspe this volume).

The Bayesian analysis is strongly consistent with our hypothesis of roughly equidistant branching. The Neighbor-Net shows a strongly tree-like signal with 12 branches unambiguously distinguished. The lexical proximity of Katuic-Bahnaric is found, although one may note the cross-linking lines in the net that indicate signal interference (probably borrowing) between Aslian, Katuic, Bahnaric and Pearic. There is a weak indication of sub-branching between Palaungic, Khmuic and Khasi, and also between Vietic and Muṅḍā. However, these are merely 'best-fits' and have extremely low statistical weight. To put it another way, while a Northern sub-group of Palaungic, Khmuic and Khasi would be an unremarkable—even pleasing—result, it is on the basis of this data no more likely than a Muṅḍā-Vietic sub-branch, a geographically absurd prospect.

Both our interpretations of the matrix (Figure 4) and the Bayesian analysis would be unremarkable if the languages formed a geographic contiguity, such that borrowings could readily propagate through the speaker community. Yet Austroasiatic branches are today distributed in discontinuous pockets, sometimes hundreds of kilometres apart. The especially high agreement between Katuic and Bahnaric can be explained by ongoing borrowing, since they were never really separated, but the same is not evident for the elevated cognate scores they share with Monic, Pearic, Palaungic, Khmuic, Khmer, and even South and Central Aslian. We suggest that in the initial stages of its dispersal, Austroasiatic had the character of a contiguous dialect chain, along which borrowing could readily spread. As various branches became more isolated, the mechanism of lexical

convergence ceased to function, and lexical change would then have been driven by internal and novel external factors. The centre of that chain was located on the middle Mekong, with the most northerly and southerly extremities ultimately becoming the Muṅḍā and Nicobaric branches respectively (see Figure 7 for a broad representation). On balance, the above hypothetical scenario readily explains these lexical data.

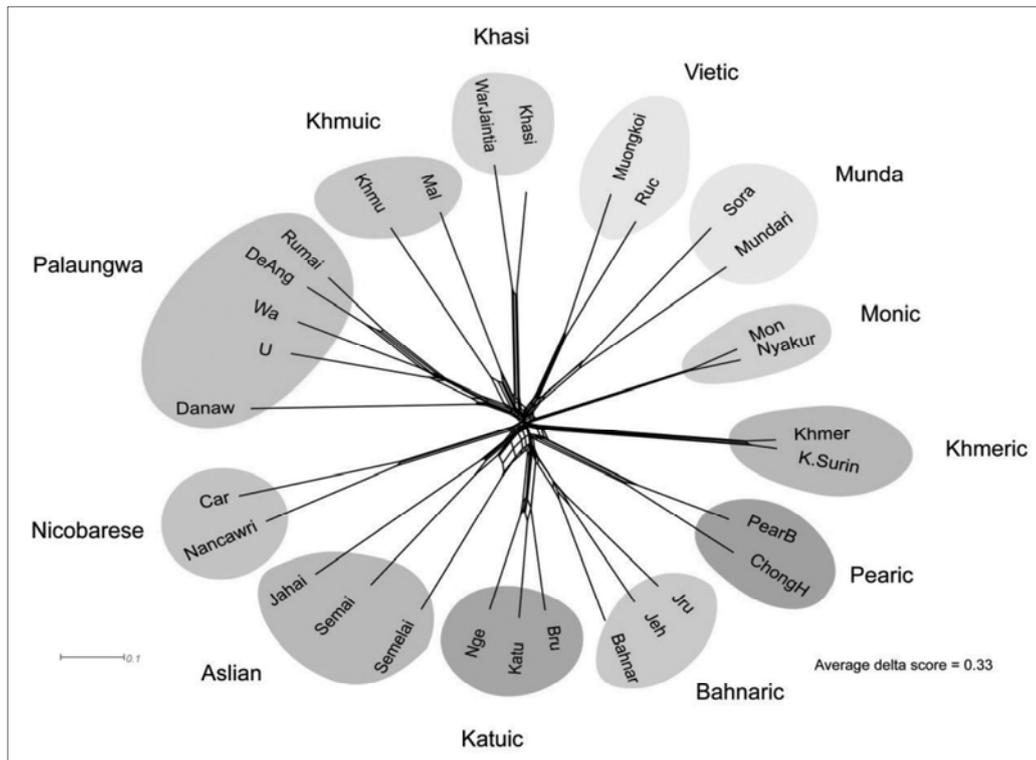


Figure 5: Neighbor-Net based on data in Figure 4 (by Russell Gray & Simon Greenhill).

5 Is Muṅḍā a primary division?

In relation to Muṅḍā, the most important question must be whether its typological character (suffixing, highly synthetic) is innovative or conservative. If the latter, the Mon-Khmer languages (non-suffixing, analytical) must constitute the innovative group, since they are unlikely to have all independently undergone the same typological restructuring. Pinnow wrote:

... the Muṅḍā languages are undoubtedly ~~are~~ more similar to Proto-Austroasiatic than the other members of the family. From a morphological viewpoint they are far more conservative than Nicobarese and Khasi, and from the standpoint of vocabulary they surpass the Mon-Khmer languages in their preservation of ancient word stems and word forms. (Pinnow 1963:150)

Subsequent writers have appropriated these views, for example Van Driem (2001:299). But it is much more likely that Muṅḍā is the innovator, and that the other languages retain, more or less unchanged, the typological character of Austroasiatic. The most persistent advocates of this latter view are Donegan & Stampe (for example 1983, 2004; Donegan 1993). They argue that the characteristics that make Muṅḍā distinctive are innovative, and that the restructuring was from isolating to synthetic typology, a reversal of Pinnow's formulation. Donegan & Stampe posit a shift from rising to falling accent in pre-Muṅḍā,

which would explain the restructuring of Austroasiatic sesquisyllables into disyllabic roots and the rise of suffixation in Muṅḍā languages. In the course of restructuring, new vowels were inserted to break up initial clusters, forming new initial syllables in Muṅḍā. The original contrast of long versus short vowel was lost as long vowels were effectively split in two to create the new syllables. Where the proto-vowel was short it did not change, and the quality of the new vowel was determined by other factors. Consequently Muṅḍā initial syllable vowels are predictable when the proto-main vowel was long; this is much easier to explain if the MK pattern is original. The alternative hypothesis, to derive MK word structure from Muṅḍā-type disyllabic roots, defies the usual rules of sound change. Examples showing this root structure correspondence are compiled from Shorto (2006) and presented in Table 2.

Table 2: Root structure correspondence: Mon-Khmer and Muṅḍā

PMK	Mon-Khmer	Muṅḍā
*jliiŋ ‘long’	Old Mon: <i>jliñ</i>	Muṅḍāri: jiliŋ
*kluu? ‘tortoise’	Mon: <i>klao</i>	Kharia: <i>kulu</i>
*brii? ‘forest’	Bahnar: <i>bri:</i>	Muṅḍāri: <i>bir</i>
*kjaal ‘air, wind’	Old Mon: <i>kyāl</i>	Muṅḍāri: <i>hɔjɔ</i>
*rk[aw]? ‘husked rice’	Khmer: <i>ʔəŋkɔ:</i>	Sora: ‘ <i>ruŋku:</i>
*kmuu? ‘dirty’	Khmer: <i>khmau</i>	Muṅḍāri: <i>humu</i>
*smuul ‘shadow, soul’	Khmer: <i>sɾəmaol</i>	Muṅḍāri: <i>umbul</i>
*kra? ‘road, way’	Praok: <i>kra</i>	Muṅḍāri: <i>hora</i>
*kla? ‘tiger’	Old Mon: <i>kla</i> 	Muṅḍāri: <i>kul ~ kula</i>
*[hj]mu? ‘name’	Old Mon: <i>jamo</i> ’, <i>himo</i> ’	Kurku: <i>jumu ~ jimu</i>
*jə[ə]ŋ ‘foot/leg’	Old Mon: <i>juñ</i>	Muṅḍāri: <i>jaŋga</i>
*ba? ‘paddy’	Bahnar: <i>ba:</i>	Muṅḍāri: <i>baba</i>

Interestingly, Donegan & Stampe (2004) favour a South Asian origin for Austroasiatic. Their supposition is that such profound structural change within Muṅḍā must have taken a long time, perhaps even more time than Austroasiatic languages appear to have been in Southeast Asia. We speculate that a rapid restructuring could well have occurred if Muṅḍā had gone through a bottle-neck event, perhaps as a small population of emigrants arriving in South Asia.

Both Sagart (this volume) and Sidwell (2009) have suggested that the Donegan-Stampe model, characterising Muṅḍā as a restructured Mon-Khmer type language, bears precisely upon the sub-classification of Austroasiatic, since it removes the rationale for Pinnow’s West-East division. It would thus appear there is no strong basis to the widely received notion of Muṅḍā versus Mon-Khmer coordinate branches. Put more strongly, it appears to represent a fundamental methodological error to assume that a branch with the most complex typology must somehow be its most ancient representative. This points strongly to a cultural classification based on fragmentary early documentation being erected into a genealogical theory without any strong evidential base. The parallel with the spurious division between Sinitic and Tibeto-Burman that is supposed to characterise Sino-Tibetan comes to mind (see, for example, discussion in Van Driem 2001:316).

6 Proposals for Northern Mon-Khmer

Beyond the position of Muṅḍā, the most important characteristic of the various published classifications has been the identification of a Northern Mon-Khmer division, consisting principally of Palaungic, Khasi and Khmuic. The idea was hinted at by Thomas & Headley (1970:404) who found that a case ‘might possibly be able to be made for a northern vs. southern grouping on the basis of the Khmuic figures, but this would be hard to sustain in the face of the rest of the figures.’ And later lexicostatistical studies, especially Headley (1976) and Peiros (1998, 2004:23), have weakly suggested a grouping of Palaungic and Khmuic counting some 26% cognates, while the same studies show lower percentages with Khasi (21~23%). However, the most important evidence for a Northern clade appears to be phonological, with two specific sound changes discussed.

Diffloth (1977) argued for a loss of medial *-h- in Palaungic, Khmuic and Khasi, and this appears to have informed his initial formulation of Northern Mon-Khmer. That sound change is now known to be restricted specifically to Khmuic (for example Khmu *maam* ‘blood’ cf. Semelai *maham* ‘id.’, Khmu *bi?* ‘sated’ cf. Danaw *əi'ʰ*, Palaung *hu?*, Semai *bahe* ‘id.’) and has been abandoned by Diffloth. Then, for a seminar delivered in Moscow in 1989⁶; Diffloth discussed a correspondence of preconsonantal *s- in those languages to a *t- in the rest of Mon-Khmer. Shorto had also discussed the same correspondence in a note drafted in the 1970s, quoted in the introduction to his posthumous 2006 handbook:

The whole of this group is characterized by a shift of *t in initial position in structures *CCVC (in some cases) to a sibilant, *prima facie* via an affricate stage. This minor shift is interesting because it apparently extends to Muṅḍā. Its incidence may be conditioned by the lost (in Mon-Khmer) V1 of Proto-Austroasiatic *CVCVC, or it may entail reconstructing an additional proto-phoneme (*t₁), *t₂. Thus we find ‘taro’, Khm. tra:v, Ste traw; RL ʔsəroʔ, Khs. shriew, Sora ‘saro:-gai-ən, Muṅḍāri sāru, Santali saru; ‘sun, day’, Old Mon tney, Khm. thṅay; KY səŋiʔ, RL ʔsəŋiʔ, Khs. sngi, Muṅḍāri siŋi. Contrast (a bad example since it uses infix forms, but with a Muṅḍā cognate) ‘new’, Khm. thmxy ~ Middle Mon t/a/mi, RL ʔtʰən/meʔ, Khs. th/ym/mai, Kharia ʔtʰən/mɛ. (Shorto 2006:x-xi)

This appears to be a real correspondence, although of the 14 examples in Shorto (2006) perhaps only five are viable. According to Sidwell’s provisional analysis, Shorto’s *t₁-/*t₂- is unnecessary, and the correspondence in question is the regular outcome of *t- before a non-labial continuant. The suggestion by Shorto that vowel assimilation may be involved runs directly counter to the Donegan-Stampe model of word-structure, and would also contradict the Pinnow inspired treatment of Muṅḍā as a separate division. Presently, it would seem to be a case of the lenition of a stop in a particularly weak position, which may or may not have occurred independently. As it is, the number of tokens is so small, it is difficult to assess its significance.

The general issue of a Northern or Khasi-Palaungic sub-family, and whether it includes Khmuic, and/or the Mangic/Pakanic languages (Mang of Vietnam, Bolyu and Bagan of China) is crucial to the question of Austroasiatic diversity. Returning to our **principle view** that real sub-grouping ought to be evident in the basic vocabulary, one can readily offer significant observations. Of most immediate importance, it is apparent that between Khasi and Palaungic there are some eight isoglosses on the Swadesh 100 list that can be treated

6 ‘Sub- and supra-classification of Mon-Khmer’ at the Institute of Far-Eastern Studies in Moscow. A copy of the handout is kept in the Cornell Library manuscript collection.

as innovations (lexical replacements, semantic shifts, loans). This is approximately 1/3 of the basic vocabulary they have in common, which we take as strongly indicative of sub-grouping, especially given the great geographical isolation between these two groups. These isoglosses are given in Table 3. Also shown are Mangic and Khmuic data, demonstrating that they do not share these innovations. Unfortunately our lexical sources for Mangic (Ferlus ms.), are not as extensive as we would like, yet even with the sparse data available a clear pattern of independent lexical evolution is evident.

Table 3: Innovations suggesting Khasi-Palaungic

Gloss	Khasi	Palaungic [#]	Mang	Bolyu	Bugan	Khmu Chuang	PMK (Shorto)
blood	snam	*snaam	ham	sa:m	kaŋ ³¹	ma:m	-
							*jhaam
claw/nail	tirsim	*rnsiim	dɔj	ma:i ¹³ ti ⁵⁵	biaŋ ³⁵	tm ^h mɔ:ŋ	-
							*t ₁ m[uə]ŋ?
hair	ʃno:ʔ	nok (Danaw)	hók	suk ⁵³	sak ⁵⁵	gləʔ	*snuuk
		*suk					*suk
man/male	trme (Amwi)	*-meʔ (Riang k ^ə rmeʔ ²)	cuj	qɔ ³¹ pɔ ³³	piau ³⁵	cmbroʔ/ gleʔ	-
							-
rain	slɛ (Amwi)	*sələʔ	maʔ	qɔ ⁵⁵	ʔa ³³	kmaʔ	*gmaʔ
swim	dʒŋi:	*ŋɔj				kljɔ:ŋ	*[l]ŋuj
two	ʔa:r	*lʔaar	zuəi	mbi ⁵⁵	bi ³¹	ba:r	*ʔaar
							(?)
water	ʔum	*ʔoom	zum	nde ⁵³	da ³⁵		*ʔ[o]m
							(?)
							*d ^ə ak

[#] Proto-Palaungic reconstruction by Sidwell, published online at sealang.net/monkhmer.

The Khasi-Palaungic innovations are neither shared with Khmuic nor Mangic. Furthermore, where Mangic and Khmuic agree, it is in retention of AA vocabulary. Two etyma stand out as requiring special comment: ‘blood’ and ‘water’.

- Ferlus (2009) reconstructs a proto-Vietic root **saam* ‘to bleed’ (for example Viet. *tuom* ‘ooze, exude’) which has a direct cognate in Mangic, and is uniquely infixated in Khasi and Palaungic. It is not clear how this connects to other Austroasiatic forms which indicate a prevocalic /h/ (note regular loss of /h/ in Khmuic).
- the root **ʔ[ɔ]m* ‘water’ replaced proto-Austroasiatic **daak*, perhaps by development from a root meaning ‘to bathe’ (cf. proto-South Bahnaric **ʔum* ‘bathe’). Yet reflexes of **ʔoom* are not general in Khmuic, but restricted to Khmu’, Khang, and Bit (Bit may be Khmuic or Palaungic, sources conflict). Other Khmuic languages have diverse forms for ‘water’, for example Iduh *paj*, Ksingmul *hɔ:t*, Mlabri *wək*, Pray *ʔɔ:k*.

There are no unambiguous indications in the basic lexicon that might link either Mangic or Khmuic to Khasi-Palaungic. Comparative phonology suggests a sound change in which proto-Austroasiatic **tC-* shifted to **sC-* in Muṅḍā, Khasi, Palaungic and Khmuic (Mangic languages regularly lose this segment so we cannot take them into account here), but the implications of treating it as a single change are too great; it would reduce Muṅḍā to a branch of one of several Mon-Khmer sub-families, and challenge too many other facts. On balance it appears that we are obliged to abandon the notion of a Northern-Mon-Khmer clade beyond a likely Khasi-Palaungic sub-grouping. And it seems appropriate at this stage to treat Mangic as an independent branch.

7 Proposals for Nuclear Mon-Khmer

The third coordinate division of Diffloth’s (2005) *stammbaum* includes those groups previously characterised as Southern and Eastern Mon-Khmer or Nuclear Mon-Khmer, Khmero-Vietic and Nico-Monic on Figure 3. This resembles the ‘Central Branch’ of Peiros (2004) although he also includes Nicobarese. Again there is little in print to support such a grouping, although there has been some discussion of the putative lower level groups. One of these is Vieto-Katuic, first proposed in Diffloth (1991). The evidence is supported by an **-h-* / **-s-* correspondence based on six apparent cases. Alves (2005) compiled some 40 Vieto-Katuic isoglosses, some of which may be loans, or retentions from a higher node of Austroasiatic. Nonetheless, it is clear that Vieto-Katuic presents a promising line of inquiry.

In side remarks concerning Nicobarese and Aslian in a paper on Palaungic vowels, Diffloth (1991:14) explicitly mentions that: ‘...Nancowry Nicobar (Radakrishnan 1981:25) is described even today as also having five diphthongs: /iá/, /ía/, uá/, /úa/ and /úua/, which seem to correspond with what we can reconstruct for Proto-Aslian.’

Diffloth (personal communication 2008) gave several etyma that appear to show these correspondences. Unfortunately it is not possible to assess the significance of this correspondence in isolation from a complete reconstruction of the respective vocalic systems. Further to this, Diffloth (personal communication 2009) indicated an isogloss for ‘wife’ showing a supposed lexical innovation uniting Monic and Aslian, for example: Old Mon *kəndɔ:r*, Semelai *kərdɔ:r*. So far as we can tell, there is only one isogloss in the basic lexicon that unites all three branches of the supposed southern clade: **btəm* ‘night’ (Shorto 2006), for example Semelai *pətəm*, Written Mon *btəm*, Car Nicobar *hatəm*. Such isoglosses might equally well be explained by contact or exclusive retention, keeping in mind that the lexicostatistics indicates that we should find some examples of inter-branch borrowing. We would expect that if a southern clade is real, it should have a strength of lexical support similar to what we find for Khasi-Palaungic.

The Bahnaric and Katuic branches, located more or less on the Khorat plateau region, have linguistic histories that are now relatively well researched, with numerous comparative studies since the 1960s (for example Blood 1966; Diffloth 1982; Efimov 1990; Efimov 1983; Gainey 1985; Peiros 1996; Shorto 2006; Sidwell & Jacq 2003; Sidwell 1998, 2000, 2005b; Smith 1972; Theraphan L-Thongkum 2001; Thomas & Smith 1967; Thomas 1967). So far comparative research reveals only one apparent lexical innovation linking these two branches: the etymon for which Shorto (2006) reconstructs PMK **knʔiəs* ‘nail, claw’, and suggests that it may be an infixated form of **kiəs* ‘to scrape, scratch’. Sidwell (2005b) reconstructs proto-Katuic **krnias*, and Theraphan (2001) reconstructs proto-Bahnaric **k(ə)rnias*.

There are many examples of transparent loans back and forth between the two groups, indicated by the asymmetries in their geographical distributions. For example; Bahnaric **liam* ‘good’ and **ʔηkee* ‘horn’ are restricted in Katuic to Katu *liem* and *tage*, while Katuic **səəŋ* ‘five’ and **kmɔɔ* ‘year’ have replaced proto-Bahnaric **pdam* and **cnam* in West Bahnaric. Such examples are numerous, the real problem being to identify early-stage loans. Various prominent etyma illustrate the dramatically different lexical, phonological and morphological histories of Bahnaric and Katuic. For example, Katuic **ηhaaŋ* ‘bone’ and Bahnaric **ktsiŋ* ‘bone’ (for example. Bahnar *ktiŋ*, Laven *ktiəŋ*, Jeh *ksiaŋ*) each show idiosyncratic developments from proto-Austroasiatic **ʔaaŋ*; neither Katuic **ktiak* ‘earth’ nor Bahnaric **teh/tneh* ‘earth’ can be reconciled with proto-Austroasiatic **tiʔ* by regular phonological correspondences; Katuic innovates a medial /s/ in **ksaj* ‘moon’ without obvious motivation (cf. Bahnaric **khaj*, Nancowry Nicobar *kahé* etc.); and Bahnaric **ʔuŋ* ‘fire’ has a short vowel and a final nasal while Katuic and the rest of Austroasiatic indicate a long vowel and a final /s/ (for example proto-Katuic **ʔuus*, Semai *ʔɔs*, Car Nicobar *ʔh* etc.).

In addition to the above, there is admittedly speculative evidence for additional branches of Austroasiatic which no longer exist. Their existence can be inferred indirectly from vocabulary in modern languages which seems to be etymologically distinct from the historical lexicon of those languages as currently classified. Such vocabulary would be borrowings or assimilations from languages no longer spoken. Blench (2009) has presented the idea that there were once three or perhaps four such subgroups. These are:

- a) The language of the Shompen on the Nicobar islands. Some documentation for this language has only recently become available, and a preliminary publication argues that it may be a language isolate (Blench 2008). However, additional evidence (Diffloth personal communication) points to some cognates with mainland Austroasiatic not shared with other Nicobarese languages (Blench in press a). Shompen might therefore represent an earlier and distinct migration to the islands.
- b) Acehnese. This language is usually classified as Chamic (Thurgood 1999; Sidwell 2005a). However, it does have a great deal of ‘residual’ vocabulary whose origins are unclear. This might either result from substrate languages of unknown affiliation or possibly from the Chamicisation of a prior Austroasiatic language (Diffloth personal communication).
- c) Pre-Chamic. In the same way, Chamic itself has distinctive unetymologised vocabulary, which could result from the absorption of resident populations on the Vietnamese mainland.
- d) Borneo Austroasiatic. There is considerable archaeological and cultural evidence for intensive maritime contact between western Borneo and the SE Asian mainland.

Adelaar (1995) says ‘The Land Dayak languages have a few striking lexical and phonological similarities in common with Aslian languages. This suggests that Land Dayak originated as the result of a language shift from Aslian to Austronesian, or that both Land Dayak and Aslian have in common a substratum from an unknown third language’. These connections were observed as early as Skeat & Blagden (1906) and the question remains unresolved. Blench (in press c) has reviewed the lexical evidence for this hypothesis and finds some striking similarities between Borneo languages and mainland Austroasiatic.

It should be underlined that these hypotheses have yet to be fully substantiated, and it may well be that, for example, not all these substrates are Austroasiatic, or that c) and d) turn out to be the same ‘lost’ branch. However, in the broader picture, they all point in the same direction, to linguistic diversity in the Southeast Asian heartland, which is the underlying thrust of the current argument.

8 Centre of diversity

There is no evidence of shared innovations that would justify grouping Mon-Khmer branches into one or two families co-ordinate with Muṅḍā. The most parsimonious explanation for the typological divergence of Muṅḍā is that it was restructured from isolating to synthetic type due to South Asian contact influences. This conclusion alone would place the centre of Austroasiatic diversity outside South Asia.

This chapter has argued that the default classification of Austroasiatic is an essentially flat array. With no basis for treating the non-Muṅḍā Austroasiatic languages as a single clade, we must abandon the idea of two main branches with a notional geographic centre towards South Asia. Even if we accept some or all the specific sub-grouping proposals outlined above, the centre of diversity would still be radially aligned within Indo-China and the lower Mekong. Groups with few or no indications of such borrowings moved away first, and came under other unrelated influences. Those in an intermediate position, such as Monic or Khmuic, either left later, or migrated early but remained in intermittent contact with the heartland. Perhaps Huffman (1978) was on the right track when he suggested this as the homeland location. Bahnaric and Katuic, resident in that region for the longest period, show the highest level of mutual borrowings. In this context it strikes us as relevant that the languages of this region exhibit strong marks of contact-driven lexical convergence.

9 The lexicon of subsistence

Diffloth (2005) has claimed that reconstructions of words for fauna support a tropical western origin for Austroasiatic. Notwithstanding the difficulty in assessing those reconstructions, such an argument must include considerations of the range and distribution of such environments at the time depths being considered for proto-Austroasiatic (that is, 4–8000 BP). Crucial also to the argument concerning the origin and dispersal of Austroasiatic is the reconstruction of the subsistence lexicon. The other uniting feature is the reckoning of the role of rice in proto-Austroasiatic society. ~~Sagart (this volume) asserts the common position when he writes that, ‘The Austroasiatic homeland question is intimately linked with the domestication of rice.’~~ Zide & Zide (1976) pointed out that Muṅḍā rice vocabulary shows cognates with other branches of Austroasiatic, and Diffloth (2005:78) has added additional lexical items to show rice cultivation was present

in the earliest period (see also Blench 2005). Below in Table 4 we list Diffloth's reconstructions, plus the equivalents from Shorto (2006).

Table 4: Proposed PMK reconstructions for rice terminology

Diffloth (2005)	Shorto (2006)
#(kə)ḃa:ʔ 'rice plant'	*ḃaʔ 'paddy'
#rəŋko:ʔ 'rice grain'	*rk[aw]ʔ (& *rkaawʔ?) 'husked rice'
#cəŋka:m 'rice outer husk'	*skaamʔ 'chaff, husks of paddy'
#kəndək 'rice inner husk'	*lʔək 'rice-bran'
#pʰe:ʔ 'rice bran'	*[p]heʔ 'husked rice'

Diffloth (2005:80) asserts that the reconstruction of such a lexicon indicates 'that PAA is a very old language whose speakers were well acquainted with rice cultivation (at least dry hillside cultivation)'. But this does not necessarily follow: we would expect that much or all of this terminology was simply transferred, by semantic extension, from previous usage that need not have involved rice previously. For example, Ferlus (1996) discusses at length the adaptation of the Austroasiatic word for 'taro' (reconstructed *t₂rawʔ/*t₂raaw[] by Shorto) in Mon, Khmer and Katuic, to designate the rice associated with intensive irrigated paddy, as it became increasingly important in Northeast Thailand from around 2000 BP. It is also surely significant that, despite numerous studies (Zide & Zide 1976; Revel 1988; Ferlus 1996; Bradley 1997; Peiros & Shnirelman 1998; Vovin 1998; Sagart 2003; Blench 2005), no convincing correspondence have been found between rice-related vocabulary in Austroasiatic and the terminologies of rice in other language phyla.

Moreover, a major controversy exists over the antiquity of rice cultivation in East Asia. Normile (1997) reports rice remains before 11,500 BP in Hunan and Hubei in central China and dates of similar antiquity regularly occur in the literature. Jiang & Liu (2006) review a series of dates for this region between 8000 and 13,500 BP. However, Fuller et al. (2008) argue that many of these finds are either poorly dated or refer to wild rice and arise from a misunderstanding of the phenotypic characters of the grain and that true domestic rice only occurs from about 6000 BP onwards. However, as Zong et al. (2007) point out, by 7700 BP there is good evidence for landscape management, through fire and flood control, consistent with paddy cultivation. Whatever the situation in China, it is important to note that no remains of unambiguously domesticated rice have been found in MSEA; these have only been dated to around 4000 BP.

One claim made by Diffloth (2005) appears to us to be uncontroversial; that Austroasiatic speakers typically spread along river valleys, seeking swampy ground to cultivate taro (although they obviously also became seagoing at least once, viz. settlement on the Nicobars). This is consistent with the suggestion made in 1943 by Haudricourt & Hédin, when they proposed that rice in Southeast Asia began as a weed in those boggy taro fields. A close association between taro and rice persists among Austroasiatic farmers, some of whom still ritually plant taro in their rice fields even though they now depend on the rice crop (such as Condominas (1957) described for the Mnong Gar). Generally the indications are strong that taro was the original crop and that rice was superimposed upon it. The extension of rice agriculture into new niches over time, such as the steep hillsides, would have greatly extended the potential range of those early communities.

Another point arises from consideration of Diffloth (2005:79). Discussing faunal terms, he remarks: 'These words are morphologically opaque, suggesting long-term familiarity with the items in question.' The same logic must necessarily apply generally, and

significantly we find that terms associated with the farming and processing of rice for consumption are not morphologically opaque. The following analyses can be offered in respect of the following quasi-reconstructions:

#təmpal ‘mortar’

This word is clearly an instrumental nominalisation with the *-p- infix. The root is *tal with a meaning associated with pounding/striking, reflected in modern Khmer /dal/ ‘to punch, pound’.

#jənreʔ ‘pestle’

Ferlus (2008) reconstructs this word as an infixed derivative of proto-Vietic *tʃeʔ ‘to dig, excavate’ (via *tʃ-r-eʔ), which diffused through Austroasiatic in association with the pestle itself.

#jəmpiər ‘winnowing tray’

The medial -mp- suggests a nominalisation of a hypothetical *jiər; possible reflexes include Khasi *jiar* ‘to drain off, to filter’ and Prao *ci:a* ‘to fall, shed’. On the other hand, Shorto reconstructs the primary meaning as verbal, on the basis of forms such as Central Nicobarese *ifuə* ‘to blow’.

#jərmuəl ‘dibbling-stick’

Diffloth and Shorto disagree on whether the primary meaning is nominal or verbal. The medials -rm- strongly suggest a nominalisation of a hypothetical *juul, very likely reflected in proto-Pearic *co:l ‘to plant’ (Headley 1985).

The above are consistent with the idea that methods of farming and preparing harvested rice for consumption were relatively new to proto-Austroasiatic speakers. These words could even have been coined, and diffused through the speaker community, after the linguistic break-up had begun, but while speakers were still in contact (the dialect chain stage).

10 Aquatic subsistence

If early Austroasiatic speakers were spreading up river valleys, then their way of life must have been strongly associated with boats and the exploitation of aquatic resources. Potential reconstructions suggesting a riverine environment for Austroasiatic are numerous. Table 5 compiles PMK reconstructions proposed by Shorto, consistent with an original aquatic subsistence strategy.

Table 5: PMK reconstructions from Shorto suggesting an aquatic subsistence strategy

Gloss	PMK	Selected Reflexes
boat/canoe	*duuk	Khmer <i>tù:k</i> , Bru <i>tùəʔ</i> , Viet. <i>nóc</i>
boat	*d ₂ luŋ	Old Mon <i>dluñ</i> , Lawa <i>ʔloŋ</i> , Bahnar <i>pluŋ</i> , Khmu <i>cəlɔːŋ</i>
crab	*kt ₁ aam	Khmer <i>kda:m</i> , Khasi <i>tham</i> , Bahnar <i>kətaam</i> , Khmu <i>ktɑːm</i>
prawn	*[k]nt ₁ a[i]s	Khmu <i>cntah</i> , Riang-Lang <i>kəntas</i>
prawn, shrimp	*suum	Katu <i>suam</i> , Viet. <i>tôm</i>
shrimp	*knbis	Khmu <i>kəmpùh</i> , Chrau <i>kəmvih</i>
catfish	*[t]kɔʔ	Mon <i>həkɔʔ</i> , Chrau <i>kɔː</i>

Gloss	PMK	Selected Reflexes
eel	*nduŋ	Khmu <i>ʔəntùəŋ</i> , Chrau <i>nduŋ</i> , Mon <i>dalun</i> , Chong <i>kəmlɔːŋ</i>
serpent headed fish [#]	*knl(ua)n	Bru <i>kluàn</i> , Bahnar <i>rəlɔːn</i> , Mon <i>kanan</i> , Nancowry <i>lúan</i>
otter	*bheʔ	Khmer <i>phèː</i> , Semelai <i>bəheʔ</i> , Khasi <i>kəsiʔ</i>

[#]From Diffloth (1979).

One term for boat is attested in both Austroasiatic and Austronesian, reflexes are laid out in Table 6.

Table 6: A SEA regional term for ‘boat’

Phylum	Subgroup	Language	Attestation	Gloss
Austronesian	PAN		* <i>qabaŋ</i>	boat, canoe
	Taiwan	Siraya	<i>avaŋ</i>	canoe
	Taiwan	Favorlang	<i>abaŋu</i>	boat
	Philippines	Magindanao	<i>kaban</i>	boat
	Philippines	Sulu	<i>guban</i>	boat
	Malayic	Malay	<i>kəbang</i>	vessel
	Barrier	Nias	<i>owo</i>	boat
	Barrier	Sichule	<i>ofo</i>	boat
Austroasiatic	Bima-Sumba	Sawu	<i>kowa</i>	boat
	PMK		* <i>kbaŋ</i>	ship
	Monic	Old Mon	<i>kbañ</i>	ship
	Bahnaric	Biat	<i>baŋ</i>	coffin
	Aslian	Jehai	<i>kupon</i>	boat
	Nicobarese		<i>kopòk</i>	boat

The lack of Muṅḍā and Khasi cognates makes it difficult to assign this term to proto-Austroasiatic; nonetheless the Nicobarese and Aslian forms are clearly not just Malay borrowings, and it must be assigned to an early period in Austroasiatic expansion. Mahdi (1999) has identified the links, both cultural and lexical, between coffins and boats, such as is attested in Bahnaric. The cognacy with Austronesian is puzzling but must be evidence for significant early contact between these two phyla (Blench in press c). An aspect of a fisher-forager strategy is that boat-using populations can move far and fast. This is accepted for Austronesian, where the rapid diversification of proto-Malayo-Polynesian is strongly associated with a major upstep in sailing technology, something reflected in the terminology for sails and other parts of the boat (Pawley & Pawley 1994). A similar innovation in boat-forms on the Mekong would account both for the rapid dispersal of Austroasiatic and the absence of nested subgroups.

These examples from the lexical data should be regarded as provisional. Shorto’s evidence is often rather more scattered than a proposed reconstruction to a proto-language might warrant. Nonetheless, it points strongly to the importance of aquatic subsistence in conjunction with wet-zone agriculture. As a consequence, speakers must have been agriculturalists and thus any date or place proposed for the expansion of Austroasiatic must be congruent with archaeological evidence for the MSEA Neolithic. The extent to which this was demic diffusion as opposed to language expansion is largely irrelevant; whether small numbers of speakers spread and influenced in situ foragers to adopt their subsistence strategies or whether all Austroasiatic groups are the result of relatively large population movements is yet to be determined.

The archaeological evidence points to a rapid expansion of the Neolithic in the Yunnan/Northern Vietnam borderland, some 4000 years ago (Higham 2002:85 ff.). Higham (2004:47) notes: ‘The pattern of intrusive agriculturalists settling inland valleys in southern China, while the coast continued to be occupied by affluent foraging groups, is repeated in the Red River area and the contiguous coast of Vietnam.’

The most well-known site of this type is Phung Nguyen, about 200 km inland from Halong Bay. Dates remain problematic, but the adjacent site of Co Loa has been dated to 2000 BC (Lai Van Toi 1999). In summarising the situation, Higham says:

We find agricultural settlements being founded in the lower Red River valley, along the course of the Mekong and its tributaries, and in the Chrao Phraya valley...The dates for initial settlement, as far as they are known, are approximately the same with none earlier than about 2300 BC. Most intriguingly, the pottery vessels in many of the sites over a broad area have a similar mode of decoration. The sites reveal extended inhumation graves and an economy incorporating rice cultivation and the raising of domestic stock. (Higham 2002:352)

And Bellwood (2005:132) remarks on the wide distribution of ‘incised and zone-impressed’ pottery ‘across parts of far southern China, northern Vietnam and Thailand after about 2500 BC’. In relation to the spread of this tradition: ‘Peninsular Neolithic pottery has cord-marked decoration with rare incision and red-slipping, often with tripod feet or pedestals...Gua Cha in Kelantan also has fine incised pottery with zoned punctuation dating to about 1000 BCE.’

Finally, Rispoli in the most recent, wide-ranging review of ‘incised and impressed’ pottery says:

The main peculiarity of the incised & impressed pottery style is its sudden appearance around the second half of the 3rd millennium BCE in Neolithic sites distributed in the major river plains of mainland Southeast Asia Incised & impressed pottery style, moreover, does not appear in isolation, but it is associated recurrently with: small polished stone tools; stone or shell bracelets and necklace beads. (Rispoli 2008:238)

We suggest that the sudden expansion of this distinctive pottery style and associated toolkit and decorative elements is a marker of the Austroasiatic expansion.

A new chronology is now being developed for the beginning of the Neolithic in MSEA, based on the new C14 chronology of Ban Non Wat (Higham & Higham 2009). This proposes that the older radiocarbon dates need to be revised and the period 1800/1700 to 1100 BC is more credible. Rispoli (personal communication) observes that the recent C14 dated excavations at An Son and Da Kai in South Việt Nam are fully consonant with this. On the basis of comparisons between our sites in Central Thailand and most of the other Neolithic sites in Thailand, Vietnam and Yunnan the new dates ‘put all the tiles in the right place’, linking MSEA with Yunnan as well as Guangxi/Guangdong.

By what mechanisms could Austroasiatic have spread so far and so fast that all the apparent branches were effectively dispersed and in place by 2–3000 years ago? Probably the answer lies in the kind of transition the Austroasiatic speakers went through as they expanded fisher/forager/vegeculturalist practices to incorporate irrigated and dry paddy. But they are not congruent with a date of 7000 BP. If rice agriculture in Indo-China is ca. 4200 BP, the initial dispersal of proto-Austroasiatic should not be earlier than this. If this is the case, then the phylum is unlikely to have an intricate nested structure, because this

would not allow sufficient time for such a structure to develop. The ‘flat array’ model of Austroasiatic is thus more plausible and congruent with the lexical data.

The Southeastern Riverine Hypothesis propounded here attributes the original homeland of Austroasiatic, or at least a principal phase of its expansion, to a region along the Mekong River. This is partly a ‘centre of gravity’ view, permitting the different groups of Austroasiatic to have become established in their various secondary homelands by a series of least radical moves, that is moves which are characterised by short distances and plausible directions. But it also suggests an important transformation in our views of the subsistence strategies of early speakers of Austroasiatic languages. It seems reasonable to hypothesise that they were based along the river and were primarily fisher-forager populations⁷.

The cultural innovation of adopting rice into their repertoire, including the facility to farm dry rice in areas upland from main waterways, could well have facilitated the outward East-West spread overland, as opposed to the mainly North-South orientation of the Mekong (and Chao Phraya, Irrawaddy etcetera) of peoples who had previously established themselves along the riverine environment as fisher-forager-tuberculturalists. The stimulus for this innovation could have been contact with another population practising agriculture, and the obvious candidate is the Daic speakers, who would have been moving into the relevant area, from the Pearl River Delta to the Red River, Northern Laos and the Upper and Mid-Mekong at around four thousand years ago (Blench in press b). This is not to say that Austroasiatic speakers adopted an agricultural package wholesale, but rather that by stimulus diffusion, the concept of a more intensive use of riverbank and immediate upland environments would have been transmitted.

11 A model for the early history of Austroasiatic

Taken together, these elements suggest that we can reconstruct the early history of Austroasiatic as follows;

- a) ca. 3800 years ago, a new pottery style begins to spread rapidly throughout the region. This is associated with beginnings of the Neolithic in the region.
- b) early Austroasiatic speakers, already practising taro cultivation, are situated on the middle Mekong and adopt rice and also get access to improved types of boat.
- c) this subsistence revolution stimulates them to move both up and down the Mekong but also to spread westward to parallel river systems, seek new areas for their taro fields.
- d) a significant movement westward (perhaps to the Tonle Sap system and/or Chao Phraya Basin) allows the development of a south-western nucleus, the origin of Monic, Nicobarese and Aslian.
- e) the rapidity of this movement accounts for the difficulty in finding well-supported nested structures in the phylogenetic tree.
- f) subsequent expansions, particularly of the Daic, Sino-Tibetan and Austronesian language phyla fragmented the chain of Austroasiatic languages leading to their comparative geographic isolation in many outlying areas.

7 It is intriguing that this is also an evolving characterisation of the Austronesians (Bulbeck 2008) in the light of the conspicuous absence of archaeological evidence for the ‘Neolithic’ agricultural package previously held to drive their expansion.

- g) Muṅḍā languages underwent a typological shift in contact with South Asian languages, but this was limited to a single branch rather than indicative of an early two-way division in the phylum.

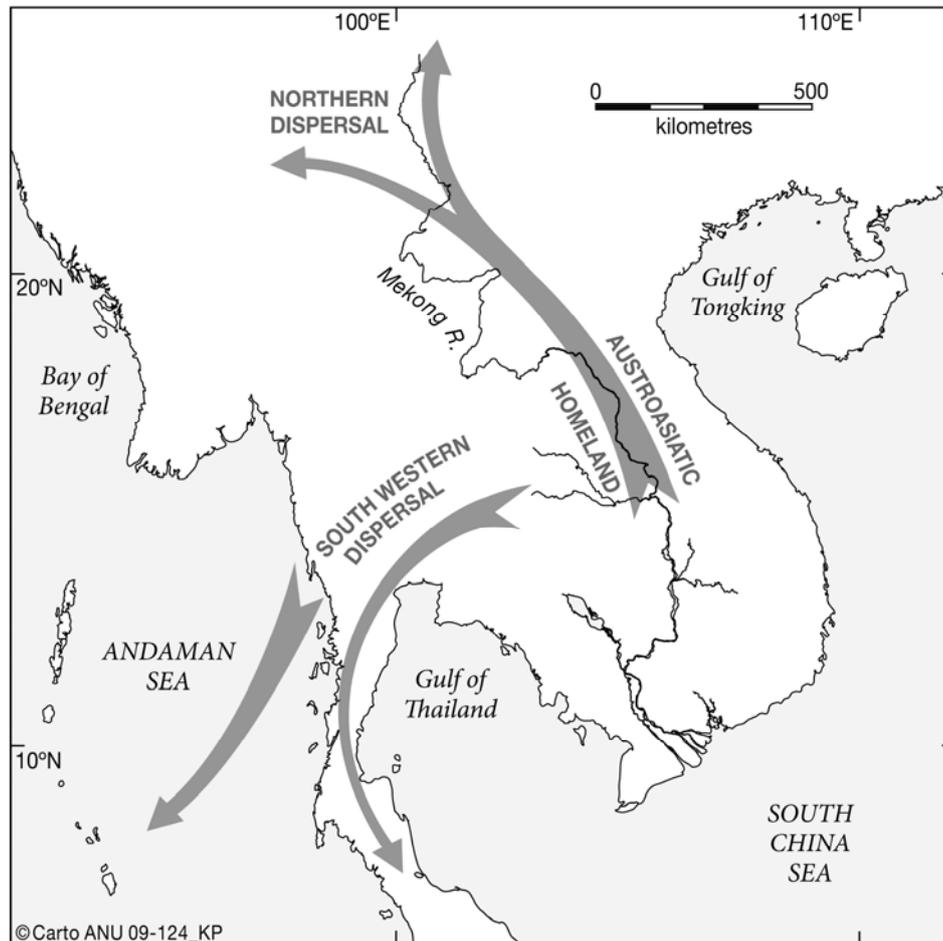


Figure 6: The Southeastern Riverine hypothesis for the Austroasiatic dispersal.

Figure 6 presents a map of the possible pattern of the dispersal of AAS according to the Southeastern Riverine Hypothesis. The model posits an early broad North-South differentiation along the Mekong valley, followed by a pattern of outward migrations, represented here by broad-brush arrows.

Figure 7 presents the possible subsequent movements of other language phyla which would have acted to fragment Austroasiatic branches, according to this scenario.

See also the paper by Blench (this volume) for the broader picture of language phylum interaction in MSEA.

It is important to underline the provisional nature of this hypothesis. The direct archaeological evidence for agriculture in mainland Southeast Asia is sparse and almost entirely confined to rice, the domestic status of which is the subject of debate. Linguistic pointers to other cultigens and broader indications of subsistence remain without archaeobotanical correlates.

The other aspect of this hypothesis, which cannot be emphasised too strongly, is to develop strong arguments either for or against the unification of individual branches of Austroasiatic. This chapter has reviewed a number of proposals and found them wanting.

Nonetheless, this does not exclude the possibility that Monic-Nicobarese-Aslian, Vietic-Katuic or Palaungic-Khasian subgroups will eventually be demonstrated. But both lexicostatistics and the absence of shared phonological innovations suggest that Austroasiatic will never prove to have the complex nested structures of previous proposals, nor will it have the antiquity these are taken to imply.

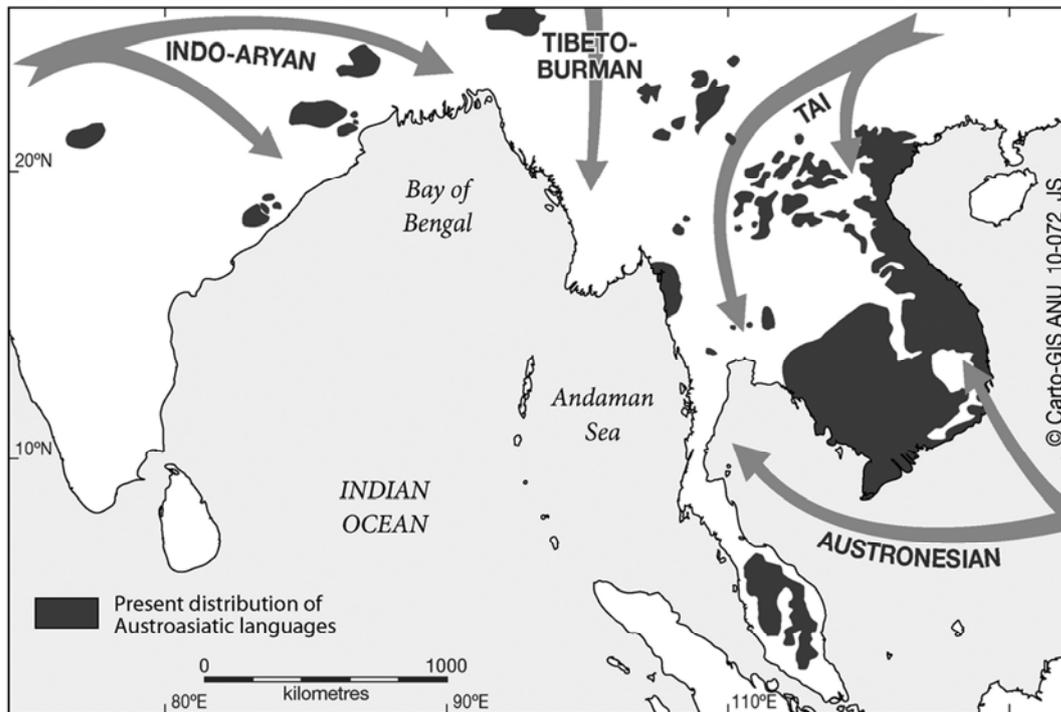


Figure 7: The intrusion of outside language groups leading to isolation of Austroasiatic communities.

Acknowledgements

Sidwell's contribution to this chapter was facilitated by assistance from National Endowment for the Humanities (Washington). Any views, findings, conclusions or recommendations expressed in this publication do not necessarily represent those of the National Endowment for the Humanities.

References

- Adelaar, Alexander K. 1995. Borneo as a Cross-Roads for Comparative Austronesian Linguistics. In Peter Bellwood, James J. Fox & Darell Tryon (eds) *The Austronesians: Historical and Comparative Perspectives*, 81–102. Canberra: ANU.
- Anderson, Gregory (ed.). 2008. *The Munda Languages*. London/New York: Routledge.
- Alves, Mark. 2005. The Vieto-Katuic Hypothesis: Lexical Evidence. In Paul Sidwell (ed.) *SEALS XV Papers from the 15th Annual Meeting of the Southeast Asian Linguistics Society 2003*, 169–176. Pacific Linguistics, Research School of Pacific and Asian Studies, The Australian National University.
- Bellwood, Peter. 2005. *First Farmers: The Origins of Agricultural Societies*. Oxford: Blackwell.

- Bergsland, Knut & Hans Vogt. 1962. On the validity of glottochronology. *Current Anthropology* 3. 111–153.
- Blench, Roger M. 2005. From the mountains to the valleys: understanding ethnolinguistic geography in SE Asia. In Laurent Sagart, Roger M. Blench & Alia Sanchez-Mazas (eds) *The peopling of East Asia*, 31–50. London: Routledge.
- 2008. Re-evaluating the linguistic prehistory of South Asia. In Toshiki Osada & Akinori Uesugi (eds) *Occasional Paper 3: Linguistics, Archaeology and the Human Past*, 159–178. Kyoto: Indus Project, Research Institute for Humanity and Nature.
- 2009. *Are there four additional unrecognised branches of Austroasiatic?* Presentation at ICAAL-4. Bangkok, October 29–30, 2009.
- in press a. *Is Shom Pen a distinct branch of Austroasiatic?* Paper given at the IVth ICAAL, Bangkok, November 2009 and in press for a special edition of Mon-Khmer Studies.
- in press b. *The expansion of the Daic peoples: cultural and iconographic correlations.* Presentation at the XIVth EURASEAA meeting. Leiden, September 2008.
- in press c. *Was there an Austroasiatic presence in island SE Asia prior to the Austronesian expansion?* Paper given at the 19th IPPA, Hanoi December 2009, and subsequently submitted for publication.
- in press d. *Vernacular names for taro in the Indo-Pacific region and their possible implications for centres of diversification.* Paper submitted for a special IPPA volume on wet taro systems. David Addison & Matthew Spriggs (eds).
- this volume, The role of agriculture in the evolution of mainland Southeast Asian language [phya](#).
- Blood, Henry F. 1966. *A Reconstruction of Proto-Muong (Including Tentative Reconstruction of Proto-South-Bahnaric)*. M.A. Thesis, Department of Linguistics Indiana University. Published by SIL in 1976.
- Blust, Robert. 1996. Beyond the Austronesian homeland: The Austric hypothesis and its implications for archaeology. In Ward H. Goodenough (ed.) *Prehistoric Settlement of the Pacific*, 117–137. Transactions of the American Philosophical Society 86.5.
- Bradley, David. 1997. What did they eat? Grain crops of the Burmic peoples. *Mon-Khmer Studies* 27. 161–170.
- Bulbeck, David. 2008. An integrated perspective on the Austronesian diaspora: the switch from cereal agriculture to maritime foraging in the colonisation of island SE Asia. *Australian Archaeology* 67. 31–52.
- Burenhult, Niclas, Michael Dunn & Nicole Kruspe, this volume, Language history and culture groups among Austroasiatic-speaking foragers of the Malay Peninsula.
- Condominas, Georges. 1957. *Nous avons mangé la forêt [...]*. Paris: Mercure de France.
- Diffloth, Gérard. 1977. Mon-Khmer Initial Palatals and “substratumized” Austro-Thai. *The Mon-Khmer Studies Journal* 6. 39–57.
- 1979. Aslian languages and Southeast Asian prehistory. *Federation Museums Journal* 24. 3–16.
- 1982. Registres, dévoisement, timbres vocaliques: leur histoire en Katouïque. *Mon-Khmer Studies* 11. 47–82.
- 1991. Palaungic Vowels in Mon-Khmer Perspective. In Jeremy H.C.S. Davidson (ed.) *Austroasiatic Languages, Essays in honour of H. L. Shorto*, 13–28. School of Oriental and African Studies, University of London.

- . 2005. The contribution of linguistic palaeontology to the homeland of Austro-asiatic. In Laurent Sagart, Roger Blench & Alicia Sanchez-Mazas (eds) *The Peopling of East Asia: Putting Together Archaeology, Linguistics and Genetics*, 79–82. Routledge/Curzon.
- Donegan, Patricia J. 1993. Rhythm and Vocalic Drift in Muṅḍā and Mon-Khmer. *Linguistics of the Tibeto-Burman Area* 16/1.
- Donegan, Patricia J. & David Stampe. 1983. Rhythm and holistic organization of language structure. In John F. Richardson, Mitchell Marks & Amy Chukerman (eds), 337–353.
- . 2004. Rhythm and the synthetic drift of Muṅḍā, in Rajendra Singh (ed.) *The Yearbook of South Asian Languages and Linguistics 2004*, 3–36.
- Driem, George van. 2001. *Languages of the Himalayas: An Ethnolinguistic Handbook of the Greater Himalayan Region: Containing an Introduction to the Symbiotic Theory of Language*. Brill.
- . 2007. Austroasiatic phylogeny and the Austroasiatic homeland in light of recent population genetic studies. *Mon-Khmer studies: a journal of Southeast Asian languages and cultures* 37. 1–14.
- Efimov, Aleksandr. 1983. *Problemy fonologicheskoy rekonstrukcii proto-katuicheskogo jazyka*. Kandidat Dissertation, Institute of Far Eastern Studies Moscow.
- . 1990. *Istoricheskaja Fonologija Juzhnobaxnaricheskix Jazykov*. Moskva: Nauka.
- Ferlus, Michel. 1996. Du taro au riz en Asie du Sud-Est, petite histoire d'un glissement sémantique. *Mon-Khmer Studies* 25. 39–49.
- . 2008. Proto-Vietic reconstruction available online at <http://sealang.net/monkhmer/>.
- . 2009. Le mot “sang” en austroasiaque. *Mon-Khmer Studies* 38. 25–39.
- Fuller, Dorian, Lin Qin & Emma Harvey. 2008. Evidence for the late onset of agriculture in the Lower Yangtze region and challenges for an archaeobotany of rice. In Alicia Sanchez-Mazas, Roger M. Blench, Malcom D. Ross, Ilia Peiros & Marie Lin (eds) *Human migrations in continental East Asia and Taiwan. Matching archaeology, linguistics and genetics*, 40–83. London: Routledge.
- Gainey, Jerry. 1985. *A comparative study of Kui, Bruu and So phonology from a genetic point of view*. Bangkok, Thailand: Chulalongkorn University thesis.
- Grierson, George A. 1919. *Linguistic survey of India*. vol. VIII, Part 1. Calcutta.
- Haudricourt, André-Georges. 1965. Mutation consonantique en Mon-Khmer. *Bulletin de la Société Linguistique de Paris* 60. 160–72.
- Haudricourt, André-Georges & Louis Hédin. 1943. L'homme et les plantes cultivés. In *Geographie humaine* 19. Paris: Gallimard.
- Headley, Robert K. 1976. Some Considerations on the Classification of Khmer. In Philip N. Jenner, Laurence C. Thompson & Stanley Starosta (eds) *Austroasiatic Studies*, 431–452. Honolulu: The University Press of Hawaii.
- . 1985. Proto-Pearic and the Classification of Pearic. In Thomas Ratanakul & Premisrat (eds) *Southeast Asian Linguistic Studies presented to André-G Haudricourt*, 428–478. Bangkok: Mahidol University.
- Heine-Geldern, Robert von. 1923. Südostasien. In Georg Buschan (ed.) *Illustrierte Völkerkunde*. Stuttgart: Strecker und Schröder.
- Higham, Charles. 2002. *Early cultures of mainland Southeast Asia*. Bangkok: River Books.

- . 2004. Mainland Southeast Asia from the Neolithic to the Iron age. In Ian Glover & Peter Bellwood (eds) *Southeast Asia: from prehistory to history*, 41–67. Abingdon: RoutledgeCurzon.
- Higham, Charles & T. Higham. 2009. Dating Southeast Asian Prehistory: New Radiometric Evidence from Ban Non Wat. *Antiquity* 83. 125–144.
- Hoijer, Harry. 1956. Lexicostatistics: a critique. *Language* 32. 49–60.
- Holm, Hans. 2003. The proportionality trap or what is wrong with lexicostatistical subgrouping. *Indogermanische Forschungen* 108. 38–46.
- Huffman, Franklin E. 1978. *On the centrality of Katic-Bahnaric to Austroasiatic*. Communication à la 2ème Conférence Internationale sur les langues Austroasiatiques, Mysore (Inde) In *Proceedings of the International Congress of Anthropological and Ethnological Sciences*, 10(2). 78–79.
- Jiang, Leping & Li Liu. 2006. New evidence for origins of sedentism and rice domestication in the Lower Yangzi River. China. *Antiquity* 80. 355–361.
- Lai Van Toi. 1999. The initial inhabitants of Co Loa [in Vietnamese]. *Khao Co Hoc* 3. 39–54.
- Mahdi, Waruno. 1999. The dispersal of Austronesian boat forms in the Indian Ocean. In Roger Blench & Matthew Spriggs (eds) *Archaeology and Language III*. 144–79. London: Routledge.
- Norman, Jerry. 1988. *Chinese*. Cambridge: Cambridge University Press.
- Norman, Jerry & Mei Tsu-lin. 1976. The Austroasiatics in Ancient South China: some lexical evidence. *Monumenta Serica* 22. 274–301.
- Normile, Dennis. 1997. Archaeology: Yangtze seen as earliest rice site. *Science* 275. 309–310.
- Pawley, Andrew & Medina Pawley. 1994. Early Austronesian terms for canoe parts and seafaring. In Andre K. Pawley & Malcolm D. Ross (eds) *Austronesian terminologies: continuity and change*, 329–61. Canberra: PL Series C-127.
- Peiros, Ilia. 1996. *Katic comparative dictionary*. Pacific Linguistics C-132.
- . 1998. *Comparative Linguistics in Southeast Asia*. Pacific Linguistics Series C-142. Canberra: Australian National University.
- . 2004. *Genetičeskaja klassifikacija avstroaziatskix jazykov*. Moskva: Rossijskij gosudarstvennyj gumanitarnyj universitet (doktorskaja dissertacija).
- Peiros, Ilia & Shnirelman, Victor. 1998. Rice in Southeast Asia: a regional interdisciplinary approach. In Roger Blench & Matthew Spriggs (eds) *Archaeology and Language, archaeological data and Linguistic hypotheses* Vol. II. 379–389.
- Pinnow, Heinz-Jürgen. 1959. *Versuch Einer Historischen Lautlehre Der Kharia-Sprache*. Wiesbaden: Otto Harrassowitz.
- . 1960. Über den Ursprung der voneinander abweichenden Strukturen der Muṅḍā- und Khmer-Nicobar Sprachen. *Indo-Iranian Journal* 4(1). 81–103.
- . 1963. The position of the Muṅḍā languages within the Austroasiatic language family. In Harry L. Shorto (ed.) *Linguistic Comparison in Southeast Asia and the Pacific*, 140–152. London: SOAS.
- Przyluski, Jean. 1924. Les langues austroasiatiques. In Meillet, Antoine & Marcel Cohen (eds) *Les Langues du Monde (Collection linguistique publiée par la société de linguistique de Paris)*, 16. 335–403. Paris: Librairie Ancienne Edouard Champion.

- Radhakrishnan, R. 1981. *The Nancowry word: phonology, affixal morphology and roots of a Nicobarese language*. Current inquiry into language and linguistics 37, Edmonton, Alberta, Linguistic Research Inc.
- Revel, Nicole (ed.). 1988. *Le riz en Asie du sud-est*. [3 vols.] Paris: EHESS.
- Rispoli, Fiorella. 2008. The incised and impressed pottery of Mainland Southeast Asia: following the paths of Neolithization. *East and West* 57. 235–304.
- Sagart, Laurent. 2003. The vocabulary of cereal cultivation and the phylogeny of East Asian languages. *Bulletin of the Indo-Pacific Prehistory Association* 23 (Taipei Papers, Volume 1). 127–136.
- 2008. The expansion of *Setaria* farmers in East Asia: a linguistic and archaeological model. In Alicia Sanchez-Mazas, Roger M. Blench, Malcolm D. Ross, Ilia Peiros & Marie Lin (eds) *Human migrations in continental East Asia and Taiwan. Matching archaeology, linguistics and genetics*. 133–158. London: Routledge.
- this volume, *The Austroasiatics: East to West or West to East?*
- Schmidt, Wilhelm. 1904. Grundzüge einer Lautlehre der Khasi-Sprache in ihren Beziehungen zu derjenigen der Mon-Khmer-Sprachen. Mit einem Anhang: die Palaung-Wa-, und Riang-Sprachen des mittleren Salwin. *Abh. Bayrischen Akademie der Wissenschaft* 1.22.3. 677–810.
- 1905. Grundzüge einer Lautlehre der Mon-Khmer-Sprachen. *Denkschrift der Akademie der Wissenschaften, Wien, Philologisch-Historische Klasse* 51. 1–233.
- Schuessler, Axel. 2007. *ABC Etymological Dictionary of Old Chinese*. University of Hawaii Press.
- Shorto, Harry L. 1976a. In Defense of Austric. *Computational analyses of Asian and African languages* 6. 95–104.
- 1976b. The Vocalism of Proto-Mon-Khmer. In Philip N. Jenner, Laurence C. Thompson & Stanley Starosta (eds) *Austroasiatic Studies*. Honolulu: University of Hawaii (Oceanic Linguistics, Special Publication, No. 13). Part II. 1041–1067.
- 2006. *A Mon-Khmer Comparative Dictionary*. Canberra, Pacific Linguistics 579.
- Sidwell, Paul. 1998. *A Reconstruction of Proto-Bahnaric*. PhD Thesis, University of Melbourne.
- 2000. *Proto South Bahnaric: a reconstruction of a Mon-Khmer language of Indo-China*. Canberra: Pacific Linguistics.
- 2005a. Acehnese and the Aceh-Chamic Language Family. In Anthony Grant & Paul Sidwell (eds) *Chamic and beyond: Studies in mainland Austronesian languages*, 211–246. Canberra: Pacific Linguistics PL 569.
- 2005b. *The Katuic Languages: classification, reconstruction and comparative lexicon*. Munich: Lincom Europa.
- 2008. Is Mon-Khmer dead? Long live Austroasiatic! Paper read to *18th Annual Meeting of the Southeast Asian Linguistics Society*, Bangi: Universiti Kebangsaan Malaysia.
- 2009. *The Austroasiatic Central Riverine Hypothesis*. Keynote address read to the *19th Annual Meeting of the Southeast Asian Linguistics Society*, University of Social Sciences and Humanities, Vietnam National University-Ho Chi Minh City, 28 May 2009.
- Sidwell, Paul & Pascale Jacq. 2003. *A Handbook of Comparative Bahnaric: Volume 1, West Bahnaric*. Canberra: Pacific Linguistics 551.
- Skeat, Walter William & Charles Otto Blagden. 1906. *Pagan Races of the Malay Peninsula*. 2 Vols. London: Macmillan.

- Smith, Kenneth, D. 1972. *A phonological reconstruction of Proto-North-Bahnaric*. Dallas: Language Data Series, Summer Institute of Linguistics.
- Swadesh, Morris. 1950. Salish internal relationships. *International Journal of American Linguistics* 16. 157–167.
- 1952. Lexico-statistical dating of prehistoric ethnic contacts: With special reference to North American Indians and Eskimos. *Proceedings of the American Philosophical Society* 96. 452–463.
- Theraphan, Luang-Thongkum. 2001. *Languages of the Tribes in Xekong Province Southern Laos*. The Thailand Research Fund.
- Thomas, David. 1973. A Note on the Branches of Mon-Khmer. *Mon-Khmer Studies* 4. 139–40.
- Thomas, David & Marilyn Smith. 1967. Proto-Jeh-Halang. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung*, 20. 157–175.
- Thomas, David & Robert Headley. 1970. More on Mon-Khmer subgroupings. *Lingua* 25. 398–418.
- Thomas, Dorothy M. 1967. *A Phonological Reconstruction of Proto-East-Katuic*. MA thesis, University of North Dakota.
- Thurgood, Graham. 1999. *From ancient Chamic to modern dialects: Two thousand years of language contact and change*. Oceanic Linguistics Special Publications No. 28. Honolulu: University of Hawai'i Press.
- Vovin, Alexander. 1998. Japanese rice agriculture terminology and linguistic affiliation of Yayoi culture. In Roger M. Blench & Matthew Spriggs (eds) *Archaeology and Language II*, 366–378. London: Routledge.
- Witzel, Michael. 1999. Early sources for South Asian substrate languages. *Mother Tongue Special Issue*, 1–76.
- Zide, Arlene R.K. & Norman H. Zide. 1976. Proto-Muṅḍā cultural vocabulary: evidence for early agriculture. In Philip N. Jenner, Laurence C. Thompson & Stanley Starosta (eds) *Austro-Asiatic Studies, Part II*, 1295–1334. Honolulu: University of Hawai'i.
- Zong, Y, Z Chen, JB. Innes, C Chen, Z Wang & H Wang. 2007. Fire and flood management of coastal swamp enabled first rice paddy cultivation in east China. *Nature* 449. 459–462.