

THE AUSTRONESIAN SETTLEMENT OF MAINLAND SOUTHEAST ASIA

Robert Blust

1. The linguistic macro-history of mainland Southeast Asia. Five distinct language families are historically attested in mainland Southeast Asia: 1. Austroasiatic (AA), 2. Sino-Tibetan (ST), 3. Tai-Kadai (TK), 4. Hmong-Mien (HM; in the older literature: Miao-Yao), and 5. Austronesian (AN). In this section I will briefly summarize the subgrouping and geographical distribution of languages belonging to each of these families, and from this information draw a set of inferences about the relative antiquity and primary center of dispersal of each genetic group within the region.¹ These inferences follow from a fundamental principle: areas of greater diversity are more likely to have been settled longer than areas of lesser diversity. This principle has a variety of applications. It is perhaps best known to linguists in its application to the determination of primary centers of dispersal, or homelands, as first formulated by Sapir (1916 [rptd: 1968]: esp. 452ff), and later formalized by Dyen (1956).² In this context it applies to collections of languages which have a single common ancestor, that is, to members of the same language family.

In considering the distribution of languages belonging to *different* language families, the diversity principle permits inferences about the most likely centers of dispersal of each family, but cannot tell us the order of priority by which these families arrived in their attested locations. To justify inferences about the relative antiquity of different language families in a given geographical region we need information not only about the internal subgrouping of each language family (relative diversity), but also about the separation times of the primary branches in each family (absolute diversity).

The latter consideration applies, for example, in situations where a continuous distribution of related languages, A, intersects an interrupted distribution of related languages, B. There are two subtypes of such a distribution: (1) A has greater absolute diversity than B, (2) B has greater absolute diversity than A. Type

(1) is most simply explained by a hypothesis that A was in place prior to B, and that the latter distribution is a product of migration around (or through) A without dividing the territory of the latter. On the other hand, Type (2) implies that B was in place prior to A, and that the latter split the territory of B through a more recent migration. For convenience we can call the type of migration inferred from distribution type (1) an indirect, or circuitous migration, and that inferred from distribution type (2) a direct, or splitting migration. To judge from the small number of reasonably clear-cut cases in the published literature, splitting migrations are the more common type. Such a situation appears to be found, for example, in aboriginal California, where the scattered pockets of distantly related Hokan languages contrast markedly with: 1. a solid block of more closely related Penutian languages which separates many of the Hokan languages north of San Francisco Bay from each other, and from the Esselen, Salinan, and Chumashan branches of Hokan further to the south, and 2. a solid block of even more closely related Uto-Aztecan languages (all belonging to the Takic branch of that family) which separates Esselen, Salinan, and Chumashan from the distantly related Yuman branch still further to the south (Shipley 1978).

1.1. Austroasiatic. According to Ruhlen (1987) there are some 155 AA languages distributed from northeastern India in the west to Vietnam in the east, and from Yunnan in the north to the Malay Peninsula and the Nicobar Islands in the south. By universal agreement the primary split within AA separates the Munda languages of northeastern India from the typologically dissimilar and distantly related Mon-Khmer languages of mainland Southeast Asia. Ruhlen (1987:148) describes Munda and Mon-Khmer as 'subfamilies' of Austroasiatic. Within Mon-Khmer he recognizes three primary branches: North (including Khasi, Palaungic-Khmuic, and Viet-Muong), East (including Katuic, Bahnaric, Khmer, and Pearic), and South (including Monic, Aslian, and the languages of the Nicobar Islands). As the useful map accompanying Lebar, Hickey and Musgrave (1964) shows clearly, the distribution of AA languages is very fragmentary; only in Vietnam and the interior of the Malay Peninsula are there relatively unbroken blocks of AA speakers. Given this

subgrouping picture, the most economical hypothesis would be that Proto-Austroasiatic was spoken in the region of present-day northern Burma or Assam.

1.2. Sino-Tibetan. Ruhlen (1987) lists 258 ST languages, which fall into two primary divisions: Sinitic, with 12 languages, and Tibeto-Karen, with 246 languages. The latter division in turn contains two primary branches: Karen, with 14 languages, and Tibeto-Burman, with 232 languages. Tibeto-Burman further divides into 1. Tibetic (75 languages in the Himalayas), 2. Baric (16 languages in the Himalayan foothills), and 3. Burmic (141 languages mostly in the hill country of the Assam-Burma borderland). In striking contrast with AA, the Tibeto-Karen languages extend in a relatively narrow, but solid band from north of the 25th parallel in the Naga Hills of Assam, to south of the 11th parallel in peninsular Burma and Thailand. Unfortunately, information on absolute separation times for AA and Tibeto-Karen languages is lacking. However, impressionistically the separation of Munda from Mon-Khmer appears to be greater than that of Karen from Tibeto-Burman, and far greater than the separation of the Burmic languages from their immediate common ancestor.³

Given the above observations it appears simplest to assume that the AA languages already occupied the Assam-Burma borderland (and adjacent regions to the east and west) at the time that Tibeto-Karen languages began to expand southward from the eastern Himalayas. As a result of the latter splitting migration the Munda languages and Khasi in the west became separated from the remaining members of the AA family further to the east and south.

1.3. Tai-Kadai. According to Ruhlen (1987) there are some 57 TK (called 'Daic') languages, divided into two primary subgroups: 1. Lati-Gelao, and 2. Li-Kam-Tai. Almost all of the most divergent members of this group are located in southern China (Guizhou, southwest Hunan, Guangxi, Guangdong, border region of Yunnan and Vietnam). The closely related Central and Southwestern Tai languages occupy an almost continuous block of lowland ricelands from southern China to the northern half of the Malay Peninsula, with an extension into the Shan states of

Burma. This distribution strongly suggests a relatively recent expansion from southern China, and the inferential evidence from language is closely supported by the evidence of documentary history (cf. Hall 1955, Chapter 7, or Wyatt 1984: Chs. 1, 2, who sketch a picture of gradual Tai infiltration of the Salween, Menam, and Mekong basins over the past millenium).

1.4. Hmong-Mien. Ruhlen (1987) lists four HM (called 'Miao-Yao') languages, of which three are in the Yao group. These are scattered in small pockets over much of southern China (Hunan, Guangxi, Yunnan) and northern mainland Southeast Asia. They are closely related, and appear to have moved southward from southern China over the past several centuries.

1.5. Austronesian. The AN languages of mainland Southeast Asia fall into three discrete groups: 1. Chamic, 2. Malay, and 3. Moken-Moklen.

There are ten reported Chamic languages, as follows: 1. Utset/Utsat (also called 'Huihui' = 'Moslem' in the Chinese sources), on Hainan Island, China, 2. Jarai, on the Darlac Plateau of Vietnam, with extensions into eastern Kampuchea, 3. Rhade, on the Darlac Plateau south of the Jarai, also with extensions into eastern Kampuchea, 4. Haroi, east of Cheo Reo in Vietnam (hence between the Jarai and the coast), 5. Eastern Cham, mostly along the south-central coastal plain of Vietnam from about Phan Rang to Phan Thiet, 6. Western Cham, mostly around the Tonle Sap in central Kampuchea, 7. Chru, near the valley of Dran, between Dalat and Phan Rang, Vietnam, 8. Southern Raglai, inland from Phan Rang, Vietnam, 9. Northern Raglai, in the mountains inland from Nhatrang, Vietnam, and 10. Caccia Raglai, near the Pacific Ocean, in Ninh Thuan Province, Vietnam. Lebar, Hickey and Musgrave (1964) also mention Bih, Krung, Noang, and Rai, which are not included by Ruhlen, and Lee (1966) mentions Bih and Rai, but not Krung or Noang. The Chamic languages are closely interrelated, and have undergone far-reaching typological adaptations to their neighbors. Both Utset (Benedict 1984, Thurgood and Maddieson 1992), and Eastern Cham (Edmondson and Gregerson to appear) have developed contrastive contour tones, the former probably through contact with one or both of the indigenous TK languages of

Hainan (Be and Hlai), and the latter through contact with Vietnamese. All other Chamic languages have undergone such extensive typological convergence with their Mon-Khmer neighbors that they were mistakenly misclassified by Schmidt (1926:140) as Austroasiatic 'mixed' languages.⁴

Cham was the language of the Indianized state of Champa, first reported in Chinese historical sources under the name 'Lin-yi' in 192 A.D. (Maspéro 1928, Hall 1955:26). At that time it was located just south of Hué in what is now Quang-nam province, Vietnam. Following the destruction of Champa by the Vietnamese in 1471 a portion of the population ancestral to the modern Western Cham fled to the region of the Tonle Sap in Kampuchea; another part of the Cham-speaking population ancestral to the modern Eastern Cham remained in what is today central Vietnam. Although TK languages have been in greater Southeast Asia far longer than AN languages, then, AN languages appear to have a longer history in mainland Southeast Asia south of China.

Malay is spoken in a number of dialects on the Malay Peninsula, in Sumatra, Borneo, and in various widely scattered locations elsewhere in both island and mainland Southeast Asia. Beginning at least as early as the 7th century with the rise of the Indianized state of Srivijaya in southern Sumatra, Malay became an important lingua franca in the major trade centers of island Southeast Asia. For over a millenium it has been the major source of borrowing throughout this region, contributing substantial numbers of loanwords (Malay, and in some cases, Sanskrit and Arabic) to languages as widely separated as Malagasy (Adelaar 1989), Javanese (Nothofer 1975, Blust 1981a), and Tagalog (Wolff 1976), among many others. For some centuries the political fortunes of Malay-speaking and Javanese-speaking centers of influence alternated in western Indonesia, with consequent changes in the pattern of borrowing, although Malay lexical influence appears to have been both more long-lasting and much more widespread.

Moken is spoken on the islands of the Mergui Archipelago in peninsular Burma and Thailand, and is reportedly non-tonal. The closely related Moklen is spoken on portions of the adjacent

mainland, and apparently has contrastive contour tones (Michael D. Larish, p.c.).

1.6. Summary. Subgrouping and distributional evidence suggests the following order of historical priority for the five attested language families of mainland Southeast Asia: 1. Austroasiatic has the longest history in the region, with a probable homeland in the Assam-Burma border region by perhaps 6,000 B.P. or earlier, 2. by perhaps 4,000 B.P. Tibeto-Karen expanded southward from the eastern Himalayas, separating the Munda languages and Khasi on the west from the rest of Austroasiatic on the east, 3. at a much later period, probably during the last two or three centuries before the Christian era, the Chamic languages and Malay became established in mainland Southeast Asia along the coasts of the South China Sea; at an undetermined period probably postdating the movement of ancestral Chamic and Malayic speakers to mainland Southeast Asia speakers of Moken-Moklen settled the coastal islands of mainland Southeast Asia facing the Andaman Sea, 4. beginning perhaps a millenium ago the Central and Southwestern Tai languages began to expand out of southern China into the major river basins of mainland Southeast Asia, 5. perhaps as recently as the past three or four centuries speakers of Hmong-Mien languages began to move southward out of China into the hill country of mainland Southeast Asia.⁵

2. The Austronesian homeland. In 1889 the eminent Dutch Indologist and Austronesianist Hendrik Kern proposed that the homeland of the widely dispersed Austronesian-speaking peoples 'probably was situated in Champa, Cochin-China, Cambodia and the neighboring regions along the sea' (Kern 1917:120). Kern's conclusion was based on several considerations, among them: 1. the evidence of cognate sets relating to flora and fauna, which indicated a center of origin within the tropics and near the sea, 2. additional cognate sets relating to agriculture, domesticated animals, and metallurgy, which favored a homeland in Indonesia or 'on the east coast of Further India' (= mainland Southeast Asia) rather than in Melanesia, Micronesia, or Polynesia, 3. the use of the term *selatan*

in Malay and Acehnese to indicate 'south', which Kern analyzed etymologically as *selat* 'straits' plus a locative suffix, hence 'toward the straits (of Malacca)', pointing to a place of origin north of the historically known Malay- and Acehnese-speaking areas, 4. the almost universal Austronesian use of a directional parameter contrasting 'toward the interior/ toward the sea', which according to Kern is more likely to have arisen in a physical environment characterized by a coastline on one side and the sea on the other than in an insular environment where the coastline is opposed to the sea on all sides, 5. the sparse native population of the island of Borneo, which would seem to rule it out as a center of origin, and 6. the existence of numerous Austronesian loanwords in such unrelated mainland languages as Cambodian, Vietnamese, and Thai, which indicate a longstanding Austronesian presence on the Southeast Asian mainland.

From a modern perspective perhaps the most remarkable feature of Kern's argument is its total silence with regard to subgrouping. Had Kern taken subgrouping into account he could not have justified the conclusion he reached, since there is relatively little genetic diversity among the Austronesian languages of mainland Southeast Asia. Indeed, the higher-level subgrouping of Austronesian, in close agreement with the ever-accumulating archaeological record, indicates an Austronesian homeland far to the north of that proposed by Kern, on or near the island of Taiwan (Haudricourt 1965, Dahl 1976, Blust 1977, 1984/85, Bellwood 1985). With mainland Southeast Asia no longer either linguistically or archaeologically a plausible candidate for the Austronesian homeland the question naturally arises 'when, how, and from where did the Austronesian languages of mainland Southeast Asia arrive in their historical locations?'.

3. Malayo-Chamic. In the remainder of this paper I will attempt to answer the question just raised. It will be seen that the data of comparative linguistics can be used to justify only a part of the answer that I propose. In the interest of stimulating further discussion and debate I elaborate on the core linguistic argument through extrapolations from documentary history and appeal to

the archaeological record. The result of this integrative approach is a set of four theses: 1. The closest linguistic relatives of Chamic are Acehnese and the Malayic languages, a subgroup that I call 'Malayo-Chamic' (MC), 2. Moken-Moklen does not subgroup closely with the MC group, 3. the MC homeland was in southwest Borneo, 4. mainland southeast Asia was settled twice by speakers of Austronesian languages: first by speakers of MC expanding out of southwest Borneo in a single (though undoubtedly generations-long) migration into much of eastern Sumatra, the Malay Peninsula, the Gulf of Siam, and the coast of Vietnam north to the 16th or 17th parallel, and second by the ancestral Moken-Moklen, probably from northern Sumatra.

Most of the elements of the present hypothesis have been advocated individually by other writers. The possible subgrouping relationship between Chamic and Acehnese (spoken on the northernmost tip of Sumatra) was proposed in 1891 by Niemann, who suggested that the Acehnese are descendants of immigrants from Champa. Blagden (1929) found additional support for this thesis, and proposed further that Austronesian-speakers had once lined the entire southern and western coasts of Indo-China, in contact with Mon-Khmer populations further inland (see Dyen 1971, for a convenient summary). More recently the thesis of an Acehnese-Chamic subgroup has been defended repeatedly by Cowan (1948, 1974, 1991), and his general views regarding the close relationship of Acehnese and Chamic are shared by a number of other writers, including I. V. Collins (1975), Shorto (1975) and Durie (1985:3ff). Similarly, the close relationship of Chamic with Malay has been defended by Marrison (1975), Blust (1981a), and James T. Collins (1992).⁶

In the interest of brevity I will not attempt to add to the evidence for an Acehnese-Chamic subgroup, but will accept the published arguments as sufficiently persuasive, and instead present evidence that what Shorto (1975) has called 'Achino-Cham' subgroups immediately with what Adelaar (1992) has called 'Malayic'. The subgroup that I defend thus has the following membership and structure:⁷

I. Malayo-Chamic

1. Achino-Cham

a. Acehnese

b. Chamic

2. Malayic

A position very close to the one defended here is expressed by Oey (n.d.) who, however, presents no lexical evidence, and whose phonological evidence is less focused on truly distinctive innovations.

Despite the agreements noted above there are conflicting views regarding the position of the Chamic languages. Although Dyen (1965) included Chru (written 'Cru'), together with Malay, Madurese, and Acehnese, as part of his lexicostatistically-defined 'Sundic Hesion', he did not regard Chru as any more closely related to Malay than are the Batak languages of northern Sumatra, and he regarded the implied relationship between Chamic as a whole and the 'Malayan Subfamily' (Malay, Minangkabau, Kerinci) as more distant than that between the Malayan Subfamily and the Barito languages of southeast Borneo, Gayō, Sundanese, Lampung, Javanese, Balinese, and Sasak. Dyen's treatment of Chamic as more distantly related to Malay than are many of the languages of western Indonesia is further reflected in Nothofer (1975). I believe that this position is fundamentally in error, since 1. MC languages share a number of innovations exclusively of other AN languages which are not likely to be borrowed, and 2. although some innovations are shared exclusively by Malayic, Madurese, Sundanese, and Javanese, many of these are best regarded as loanwords, given the known history of extensive borrowing from Malay in western Indonesia.

Most recently Ruhlen (1987) has attempted a synthesis of the sometimes conflicting views of various specialists. He recognizes a 'Malayic' group with four coordinate branches: 1. Madurese, 2. Malayic Dayak (Iban, Malayic Dayak), 3. Malay-Moklen (Moken-Moklen, and a 'Malayan' group which includes Malay, Minangkabau, Urak Lawoi', and Rejang), and 4. Chamic (divided into Utset = Huihui vs. the rest). At least two

features of this classification seem quite clearly to be in error: 1. it is evident from inspection that Iban is more closely related to Malay than are either Rejang or Moken-Moklen, and 2. Chamic subgroups with Malay and its closest congeners apart from Moken and Moklen. The evidence for these claims will be presented in three subsections: 1. phonological innovations, 2. morphological innovations, and 3. lexical innovations.

3.1. Malayo-Chamic phonological innovations. Three general phonological innovations are shared by Malay, Acehnese and the Chamic languages: (1) PMP *q shifted to PMC *h, (2) PMP *R merged with PMP *r as PMC *r, but remained distinct from all other proto-phonemes, and (3) PMP *w disappeared in initial position. As a consequence of the loss of initial *w, a fourth change occurred which was lexically specific: (4) historically secondary /h/ was introduced in the reflex of PMP *waRi 'day', initially in the construction *mata hari 'sun' (= 'eye of the day'), where it served to prevent contraction of the sequence of like vowels which arose following the change of PMP *maCa ni waRi to pre-PMC *mata ari. Table 1 schematizes the presence (+) or absence (-) of each of these innovations (numbered 1-4) in the Chamic languages, Malay and several other languages which share some of the same changes, or which have been associated with Malay, with Chamic, or with both in past classifications.

TABLE 1

Malayo-Chamic phonological innovations and their distribution in other languages of western Indonesia (PC = Proto-Chamic)⁸

LANGUAGE	(1)	(2)	(3)	(4)
1. PC	+	+	+	+
2. Aceh	+	+	+	?
3. Malay	+	+	+	+
4. Maloh	-	+	-	-
5. Moken	-	-	-	-
6. Gayō	+	-	-	?
7. Karo Batak	+	+	-	-

8. Nias	+	-	-	?
9. Rejang	-	-	-	-
10. Lampung	+	-	-	?
11. Sundanese	+	-	-	?
12. Javanese	+	-	-	-
13. Madurese	?	+	?	?
14. Balinese	+	-	-	-
15. Sasak	-	+	+	?
16. Ngaju Dayak	?	-	?	?
17. Malagasy	?	-	-	?

Table 1 is to be read as follows. In column (1) a '+' indicates that the language in question reflects *q as /h/ in at least some environments. A '?' indicates a zero reflex which is ambiguous, since it could result either from the changes *q > *h > zero, or the direct loss of *q. A '-' indicates that the language in question reflects *q as a glottal stop, and so probably has never passed through a *q > *h stage. In column (2) a '+' indicates that the language in question reflects *R as /r/ unmerged with *l, *y, zero, etc. A '-' indicates that although *R and *r may have merged, the distinction between them and other PMP phonemes has not been retained. The subgrouping value of the *R/r merger is weak, since it is widespread, and its distinctiveness in the present context is based on common retention rather than shared innovation. In column (3) a '+' indicates that the language in question retains a non-zero reflex of initial *w. The evidence from Madurese is contradictory, since *w- sometimes disappears (*wada > *adha?* 'not have, all gone', *waRi > *are* 'day'), and sometimes appears as /b/ (*wada > *badha* 'be, have', *walu > *ballu?* 'eight', while no clear non-borrowed reflexes of initial *w are known in Ngaju Dayak. A '-' indicates that initial *w has a non-zero reflex, normally /w/. In column (4) a '+' indicates that PMP *waRi is reflected idiosyncratically with initial /h/. A '?' indicates either that both *q- and *w- have disappeared and a reflex of *waRi is available, or that a reflex of *waRi is not available.

(1) Shift of PMP *q to PMC *h. It is clear that the Chamic languages, Acehnese and Malay all reflect PMP *q as /h/.

Through a later change Acehnese lost initial /h/ in native words (but reacquired it in some Malay loans), and /h/ sometimes disappeared, or optionally disappeared in certain environments in Malay (Dyen 1953). The question whether PMP *q and *h merged as PMC *h is more difficult. Although initial *q and *h evidently merged unconditionally in Chamic (cf. PMP *huab > Rhade *həap* 'yawn') and Malay, this was not the case in final position, and does not appear to have been the case intervocalically either in Chamic (Blood 1962, Thomas 1963) or in Malay (Dyen 1953). Table 2 cites reflexes of (1) PMP *qabu 'ash', (2) *taqun 'year' and (3) *(um)-utaq 'vomit' as supporting evidence for the distribution reported in column (1) of Table 1.⁹

TABLE 2

Reflexes of PMP *q in Malayo-Chamic and other languages of western Indonesia

LANGUAGE	(1)	(2)	(3)
1. PC	*həbəw	*thun	*pətah
2. Aceh	abèè	tahon	meuntah
3. Malay	(h)abu	tahun	muntah
4. Maloh	—	taun	—
5. Moken	kaboy	takon	motak
6. Gayō	—	tun	—
7. Karo Batak	abu	tahun	utah
8. Nias	(h)awu	—	muta
9. Rejang	abew	tawen	utea?
10. Lampung	—	tahun	utah
11. Sundanese	hawu	ta(h)un	utah
12. Javanese	awu	tahun	mutah
13. Madurese	abu	taon	ota
14. Balinese	abu	tahun	utah
15. Sasak	awu-awu	taun	uta?
16. Ngaju Dayak	(kawo)	—	uta
17. Malagasy	(mavo)	—	—

This innovation acquires significance from its rarity. There are more than 900 Austronesian languages, and outside of Chamic the change $*q > /h/$ is attested in only two areas: 1. western Indonesia, where it is found in about a dozen languages or close-knit subgroups and 2. in Lakalai of New Britain. Since all of these languages except Lakalai occupy a nearly contiguous area, it is possible that innovation (1) defines a larger subgroup in western Indonesia of which Malayo-Chamic is a single branch. Such a group, if accepted, could include Malayo-Chamic, Gayō, the Batak languages, Nias, Lampung, Sundanese, Javanese, and Balinese. However, it could not include Maloh of southwest Borneo, Rejang of south-central Sumatra, or Sasak of Lombok, where $*q > /h/$ in at least some environments. Since Sasak appears to be closely related to Javanese and Balinese, this discrepancy implies at least one historically independent change $*q > /h/$ in the latter two languages. Given this conclusion we might hypothesize a parallel change in several other languages of the region. However, a hypothesis of parallel evolution fails to explain why a change which hardly occurred elsewhere tended to occur repeatedly in a geographically contiguous area of western Indonesia.

On balance it appears that contact may well have played a significant part in the spread of this distinctive innovation across subgroup boundaries in Malaya-Sumatra-Java-Bali. Contact provides a far less satisfactory explanation for the shared change in all attested Chamic languages, including the oldest inscriptional documents for Cham (Marrison 1975), and Utset of Hainan island, where earlier final $/h/$ conditioned the tone which Thurgood and Maddieson (1992) write '55'. Jarai, Rhade, Chru and Roglai are spoken in the highlands of Vietnam, separated by over 400 miles from the nearest Malay-speaking areas, and although Malay contact with Champa is historically attested, all indications are that it involved occasional commercial or political association between equals, and hence was not likely to lead to massive lexical borrowing which could have triggered an independent change $*q > /h/$.¹⁰

Perhaps the greatest value of innovation (1) is that it enables us to exclude Moken-Moklen from the Malayo-Chamic

group, since in this set of languages *q and *k merged as /k/: *qenay > *kenay* 'sand', *kami > *kamoi* 'we (excl.)', *taqun > *takon* 'year', *luka > *loka* 'a cut, wound', *tanaq > *tanak* 'earth', *manuk > *manok* 'fowl'.

(2) Merger of PMP *r and *R as PMC *r. This merger sharply distinguishes the MC languages from Moken-Moklen, in which *r/R merged with zero or *l in initial position, with *l intervocalically, and with *n word-finally (*r/R > *l > -/n/, perhaps under Thai influence); from Rejang, in which *R and apparently *r merged with *l, /ɾ/ or zero intervocalically, but otherwise disappeared; from Nias, in which *R and apparently *r disappeared unconditionally; from Gayō, and Sundanese, in which *R generally merged with *y as /y/; from Lampung, in which *R merged with *d in initial position, with *d or *y intervocalically, and with *y in final position; from Javanese, in which *r appears as /r/, but *R as zero; from Balinese, in which *R and apparently *r became /h/; from Ngaju Dayak, in which *r became /r/, but *R became /h/ in the native speech stratum; and from Malagasy, in which *r became /r/, but *R became zero, /z/, or occasionally other segments. However, the same merger has occurred in a number of languages outside western Indonesia. Documentation of most of these reflexes is available in Dempwolff (1934-38), Blood (1962), Thomas (1963), Nothofer (1975) or Blust (1984).¹¹

(3) Merger of *w with zero in initial position. This change, which is highly distinctive, may also have involved the loss of *w in certain intervocalic environments, although these will not be discussed here. Table 3 cites reflexes of (1) PMP *wahiR 'fresh water; stream', (2) *wakaR 'root', and (3) *waRi 'day' as supporting evidence for the distribution reported in column (3) of Table 1.

TABLE 3

Reflexes of PMP initial *w in Malayo-Chamic and other languages of western Indonesia¹²

LANGUAGE	(1)	(2)	(3)
1. PC	—	—	*hrəy

2. Aceh	—	—	uròë
3. Malay	air	akar	hari
4. Moken	win	y-akan	waley
5. Gayō	weih	—	—
6. Karo Batak	—	—	waluh
7. Nias	—	waʔa	waʔu
8. Rejang	bioa	—	biley
9. Lampung	way	—	—
10. Sundanese	cai	akar	—
11. Javanese	—	—	—
12. Madurese	—	—	are
13. Balinese	wé	akah	wahi
14. Sasak	aiʔ	akah	—
15. Ngaju Dayak	—	—	—
16. Malagasy	—	—	—

The evidence that PMP *w- was lost in Chamic is indirect, and is, moreover, critically dependent on a single comparison. Two other comparisons which point to the same conclusion, are problematic: *wakaR > JAR *akha* (with unexplained -h-) 'root', and *wahiR > *wayar > PC *ʔya (with unexplained *ʔ) 'water'. Oey (n.d.) claims that PMP initial *w (written *u) became Proto-Chamic *h in all forms, but of the four etymologies which he cites in support of this claim only one (*waRi) can be justified. It is noteworthy that initial *w was not lost in Moken-Moklen or in any of the other languages of Table 1 except Sasak. Despite its high lexicostatistical percentage with Malay (Dyen 1965, Nothofer 1975), Madurese shows contradictory reflexes which suggest that it did not lose initial *w. MAD *are* 'day' may, then, be a loanword from Malay.

Perhaps the key piece of subgrouping evidence which emerges from Table 3 is the irregular initial consonant in the Proto-Chamic and Malay reflexes of *waRi 'day'. Because a change *w > /h/ is phonetically unlikely, whereas the loss of *w is attested in other languages, there are grounds for believing that /h/- has been acquired secondarily in PC *hrəy. Dempwolff (1937:70d) described the *h-* of MAL *hari* as an 'analogical' development in the sequence *mata ari 'sun' ('eye of the day'),

but stated no proportional basis for an analogy. Although Wilkinson (1959) lists (*h*)*ari* 'day' for standard Malay without reference to morphological environment, in Sarawak Malay there are two allomorphs: 1. *ari*, occurring in isolation, and 2. *hari*, occurring in the collocation *mata hari* 'sun'. Since recurrent sound changes would lead us to expect MAL ***mata ari*, it appears that /h/ was inserted in this environment to prevent coalescence of the sequence of like vowels which resulted from loss of *w, and that the allomorph /hari/ was then generalized to all positions. The probability is small that Malayic and Chamic would independently undergo this same sequence of lexically idiosyncratic aspirate insertion and generalization. Moreover, since the PMP collocation was **mata ni waRi* 'sun', it follows that the genitive marker **ni* had already disappeared in Proto-Malayo-Chamic prior to the development of an epenthetic aspirate.

3.2. *Malayo-Chamic morphological innovations.* The disappearance of the PMP genitive marker **ni* in the Malayo-Chamic languages is a morphological innovation which distinguishes this group from, e.g., the Batak languages. Other morphological innovations which appear to be distinctive (not widely shared) are loss of the morpheme boundary in the following forms: 1. PMP **um-inum* > PMC **minum* 'drink': CHAM, JAR *meñūm* (with secondary palatalization prior to the centralization of the penultimate vowel), ACH *minòm*, MAL, Kenindjal *minum*, MIN *minun*;¹³ 2. PMP **ma-qalesem* > PMC **masem* 'sour': CHAM *məthəm* 'vinegar', JAR *masam* 'acid, sour', RHA *msam* 'sour', Roglai *masap* 'vinegar', BJR, IBAN *masam* 'acid, sour, sharp', MAL *masam* 'acid; sour', MIN *masam* 'sour'; and 3. PMP **ma-qasin* > PMC **masin*: JAR *məsin*, RHA *msin*, Roglai *masit* 'salty', IBAN, BJR *masin* 'salty, salted', MAL *m-asin* 'brackish; salt; briny', MIN *masin* 'salty, brackish'. The first of these changes is particularly distinctive, as it did not occur in, e.g., Madurese (*enom* 'drink'), although it does appear in Rejang *méném* 'drink', where I take it to be an independent development.¹⁴

3.3. *Malayo-Chamic lexical innovations.* A small number of lexical innovations shared by the Chamic and Malayic languages

with a restricted number of other languages in western Indonesia was listed in Blust (1981a). These include, most notably, innovations for the PAN numerals *pitu 'seven' (replaced by *tuzuq), *walu 'eight' (replaced by *dua lapan), and *siwa 'nine' (replaced by *salapan/sambilan). The first of these innovations is widely distributed in central, western and northern Borneo, and also appears in Moken-Moklen and Sundanese. The innovations for 'eight' and 'nine' have a much more limited distribution, and consequently appear to be more precise indicators of subgroup membership.

Shorto (1975) cites a number of lexical innovations shared by Acehnese with what he calls 'mainland Austronesian', and many of these have cognates in the Malayic languages. He does not, however, make a serious attempt to establish that the innovations in question are restricted to the languages he compares, since this would require searching the available dictionaries of many other languages. The difficulty of finding cognate sets which are confined exclusively to the MC languages becomes apparent once such a search is undertaken. Clearly, if MC is a valid subgroup it would be reasonable to expect some set of lexical innovations to be shared by the MC languages and no others. What complicates this theoretical expectation almost beyond resolution is the overwhelming influence of Malay throughout the region for at least the past millenium. Lexical borrowing from Malay varying from light to very heavy is a widespread feature of the languages of western Indonesia. Since many loans have been nativized, and some are items of basic vocabulary, lexical isoglosses which reflect subgrouping relationships are sometimes very difficult to separate from those which reflect centuries-long borrowing from Malay. A few examples should suffice to illustrate.

The comparison CHAM *pathəy/pithəy*, JAR *pasəy*, RHA *msey*, ACH: *beusòë*, IBAN, MAL *besi* 'iron (metal)' points to *besi, a form which (contra Dempwolff) cannot be assigned to Proto-Austronesian, but rather appears to be a relatively late innovation in western Indonesia. The question naturally arises, 'how late an innovation?'. If MC is a valid subgroup and *besi was innovated in its immediate common ancestor we would expect reflexes of this etymon to be confined to the MC

languages. If MC is a valid branch within a larger subgroup defined by reflexes of *tuzuq 'seven' (see above) and *besi was innovated in the common ancestor of this larger group, we would expect reflexes of *besi to be confined to many of the languages of northern, central and western Borneo, Moken-Moklen, Sundanese and MC, but this is not the case. In addition to the MC languages, reflexes of *besi appear among others in Moken-Moklen, the Batak languages and Rejang of Sumatra, Sundanese, Old Javanese, modern Javanese, Madurese, Balinese, Sasak, some of the Land Dayak languages of southwest Borneo, and apparently in Malagasy (where it is irregular). Since all of these languages except Moken-Moklen, Rejang and Land Dayak retain a reflex of *pitu 'seven', we are forced to conclude that either some reflexes of *tuzuq 'seven', or some reflexes of *besi 'iron' have spread by borrowing. A case such as this is relatively easy to decide: iron is a useful item of material culture, and one which is likely to be traded by those who have it to those who do not. Such forms as Moken *bechoi*, Karo Batak *besi*, Old and modern Javanese *wesi*, Balinese, Sasak *besi*, Malagasy *vi* 'iron' are therefore all likely Malay loanwords. The clear implication of such an interpretation, however, is that Malay loanwords were already present in Old Javanese (the language of east Javanese texts from roughly the 9th to 15th century), and had entered Malagasy before the migration out of Southeast Asia (cf. Adelaar 1989 for many more examples of the latter). What complicates the interpretation of such lexical distributions is this: if reflexes of *besi are interpreted as loans in non-MC languages their presence in Acehnese or Chamic could be attributed to borrowing as well.

To circumvent the foregoing problem we might choose a cognate distribution for which borrowing provides a less likely explanation. A good example is the word for 'tooth'. In PMP the meaning 'tooth' was represented by several phonologically similar forms, of which the most important were *ipen and *nipen. In Proto-Malayo-Chamic these forms were replaced by a lexical innovation *gigi, of unknown provenance: JAR *tə-gəy* (with body-part prefix seen also in *susu > *tə-səw* 'breast', and *kakay > *tə-kay* 'foot/leg'), RHA *êgey*, ACH *gigòè*, IBAN, MAL *gigi* 'tooth, tooth-like projection'. A possible external cognate is KB

gigi 'point of a plough, coulter', but even if this etymology is accepted we could maintain that **gigi* was a MC innovation in the meaning 'tooth'. The problem with this conclusion is that a reflex of **gigi* meaning 'tooth' also shows up in three languages outside the MC group: MAD *gigi* 'tooth', BAL *gigi* 'tooth, tusk'; *gigi kilap* 'thunderbolt', SAS *gigi* 'tooth', *gigi-n gergaji* 'teeth of a saw', *peŋ-gigi* 'point of a plough'. While Madurese may ultimately prove to subgroup immediately with Malayo-Chamic, this does not appear likely for Balinese or Sasak (cf. Table 1). Unusual as it may be for body-part terms to be borrowed, then, borrowing appears to offer the most likely explanation for the distribution of reflexes of **gigi* 'tooth' outside the MC group. Although this explanation inevitably raises the question whether the Chamic terms themselves are loans, borrowing into Chamic seems unlikely for at least two reasons: 1. the use of the Jarai word with the an innovative prefix /te/-, and 2. the relative distance and remoteness of Jarai or Rhade from centers of Malay influence in comparison to the Batak languages, Balinese or Sasak. Keeping these provisos in mind, Appendix 1 offers a set of likely MC lexical innovations, some of which also appear in other languages of western Indonesia, presumably as a result of borrowing from Malay.

The MC hypothesis is greatly strengthened by epigraphical evidence. As Marrison (1975) has noted, although most inscriptions associated with the kingdom of Champa are in Sanskrit, a few include native material, and these constitute the oldest written documents in any Austronesian language. According to Marrison the earliest Cham inscription, though undated, can be assigned to the late 4th century A.D. This short text, associated with a well at Dong-yen-chau near the Cham capital of Indrapura, 'reveals a language close to Malay in grammar and vocabulary'. Among distinctive lexical items cited are *punya* 'property', *nari* 'from', and *dengan* 'with' (cp. modern Malay *punya* 'ownership, possession', *dari* 'from', *dengan* 'with'). More striking still is an inscription on the first stele of Dong Duong (875 A.D.) which records in Sanskrit the founding of Indrapura by Indravarman II, and contains an appended list of field names in Cham. Among the latter Marrison cites *jraung*

apaung salavang, and *damanuv vlur candang*, which he associates with modern Malay *jerong apong selempang* 'shark stranded sideways', and *danau balur chandang* 'field by the lake of the aggressive *tenggiri* fish'. These agreements between a form of Cham spoken in the late 9th century, and modern Malay are striking and, given their quasi-idiomatic character, not likely to have been borrowed.

The wider relationships of the MC languages remain debatable. Blust (1981a: Table 1) proposed that Malayic, Sundanese, Maloh and Rejang form a subgroup apart from Achino-Chamic, but in view of the evidence considered here the innovations in the numerals which were used to include Rejang and Maloh in this group are perhaps best attributed to borrowing. It still appears likely to me that Sundanese may be coordinate with MC in a larger group which includes no other languages.¹⁵

4. The Malayo-Chamic settlement of mainland Southeast Asia. The existence of a Malayo-Chamic subgroup, with member languages distributed from Hainan island to southwest Borneo raises a number of culture-historical questions. Among the more fundamental of these are: 1. how did the MC languages acquire their present distribution?, and 2. what is the time-frame relevant to the MC dispersal? Unfortunately, the linguistic evidence provides only partial and approximate answers to these questions. In the hope of filling in some important gaps I will attempt to supplement the linguistic record with the admittedly fragmentary and sometimes ambiguous evidence of documentary history, and the independent evidence of the archaeological record. My conclusions in this and the following section must inevitably remain conjectural, but in explicitly stating a position and what I regard as supporting evidence for it they may nonetheless benefit future inquiry into the subject.

4.1. The Malayo-Chamic homeland. As noted by Adelaar (1992:207) the Proto-Malayic homeland remains in dispute, but the single most promising location is southwest Borneo, where a number of divergent Malay-like languages still are spoken by indigenous longhouse-dwelling animists who show little evidence of having reached their present locations from the Malay

Peninsula or Malay-speaking regions of eastern Sumatra (Hudson 1970). From Southwest Borneo Malayic speakers could easily have reached their historically attested locations on Sumatra (Minangkabau, primarily on the west coast of central Sumatra; 'Middle-Malay', along the east coast of central and southern Sumatra; various other forms of Malay in scattered locations around the island), the Malay Peninsula, the islands of Bangka and Belitung, the Riau Archipelago, Tioman island, the Anambas islands and Natuna through a major population movement probably extending over several generations. We know that they were a seafaring people, and that such minor water barriers would not have prevented their free movement in large numbers.

4.2. *Documentary history of Cham and Malay.* As already noted, Cham was the language of the Indianized state of Champa, first reported in Chinese historical sources under the name Lin-yi, in 192 A.D. (Maspéro 1928, Hall 1955:26). At that time it was located just south of Hué, in what is now Quang-nam province, Vietnam, and for centuries it occupied roughly the same area despite pressure due to the southward expansion of the Vietnamese. A language generally identified as 'Old Malay' appears in three inscriptions associated with the Indianized state of Sriwijaya in south Sumatra, two from the region of Palembang, and one from the neighboring island of Bangka, dated at 683, 684 and 686 A.D. (Teeuw 1961:9ff). We can thus say with some confidence that Cham was already spoken in the coastal regions of south-central Vietnam by the end of the 2nd century A.D., and that within 500 years of this date Malay was spoken in southeastern Sumatra. In both cases the attested language was that of an Indianized state, and almost certainly was in the region for some time prior to Indianization. If the Proto-Malayic homeland was in southwest Borneo the same probably was true of Proto-Malayo-Chamic, since there is no basis in the wider linguistic relationships for deriving PMC from Vietnam. The question then arises whether the migration of the Chamic-speaking peoples and the Malayic-speaking peoples from southwest Borneo were separate historical events, or part of a single demographic expansion.¹⁶

4.3. *A conjectural history of the Malayo-Chamic migrations.*

The simplest and internally most consistent interpretation of the attested language distributions and proposed subgrouping relationships considered above favors the following conjectures: (1) the Malayic and Chamic languages arrived on the Asian mainland through a population movement which involved speakers of one (dialectally complex) language which originated in southwest Borneo, probably in or near the basin of the Kapuas River (2) a MC dialect chain probably developed along the 2,000 miles of coastline reaching from the east coast of the Malay Peninsula to the region of Hue in Vietnam; (3) the northern portion of this dialect chain would have been in contact with speakers of AA languages, but the southern portion initially would have had few contacts with speakers of other languages, (4) the MC dialect chain was first broken by the southward expansion of the Khmers, who overthrew the Indianized state of Funan in the middle of the 6th century A.D., and presumably led to the extinction of its language, (5) beginning probably in the 13th century the physical and linguistic separation of Chamic and Malayic was increased further by the steady southward expansion of the Thai, at the expense of whatever AN-speaking populations had previously inhabited the Gulf of Siam and the eastern coast of the northern Malay Peninsula, (6) certainly prior to 1500, and probably prior to 1200 one group of Chamic speakers migrated to northern Sumatra, giving rise to the Acehnese. Each of these conjectures will be considered in turn.

Conjecture (1) is motivated by parsimony. The presence of Malayic and Chamic languages on the Asian mainland can be accounted for by separate migrations, or by a single migration. If MC is a valid linguistic subgroup it is simplest to posit a single migration to explain the attested language distribution. This position is strengthened by the (indirect) documentary evidence that Cham was already spoken in Vietnam by 192 A.D. Given the relative closeness of their relationship today, it is improbable that Chamic and Malayic could have been more than dialectally distinct 1,800 years ago. Rather than posit separate movements of what must have been essentially the same people departing from the same center of dispersal, I assume a single demographic expansion

that was possibly driven by innovations in metallurgy which took place in southwest Borneo during the last few centuries before the Christian era (Bellwood 1985:289ff). In all likelihood the initial migration of MC speakers would not have preceded the earliest records of Champa by more than four or five centuries, thus placing the expansion out of southwest Borneo in the period 200-300 B.C.

Conjecture (2) is motivated by a consideration of the settlement pattern of AN-speaking peoples generally. Like other early AN groups, speakers of PMC were a maritime people who favored coastal environments. In spreading out from southwest Borneo they would almost certainly have explored uncontested coastal environments to the fullest extent possible. Initially this would have included the east coast of Sumatra, and the eastern coast of the Malay Peninsula north along the Gulf of Siam and the Mekong delta to the historically attested region of Champa. Quaritch Wales (1976:10), citing a complaint of the Chinese Buddhist pilgrim Fa-hien, reports that piracy had already become a threat to navigation in the seas surrounding the Malay Peninsula by the 5th century A.D. This observation is important for two reasons. First, piracy can exist only where there is a thriving maritime commercial network, and such writers as Wolters (1967), and Kenneth Hall (1985) have left no doubt that this was the case in western Indonesia during the early Srivijaya period. Second, pirates in the region of the Malay Peninsula in the 5th century A.D. could hardly have been other than Austronesian-speaking peoples. Although the rise of a thriving maritime commercial network undoubtedly owed its impetus to the beginnings of the India-China trade, there is every reason to believe that historically attested commercial empires such as Srivijaya were an elaboration of networks of maritime traffic and localized trade which had existed in the pre-Indian period.

There are some scattered indications that the west and east coasts of the Malay Peninsula were not equally attractive to seafarers. Quaritch Wales (1976:8) points out that the west coast presents greater hazards to navigation, and is more overgrown with 'forbidding mangrove swamps', whereas the east coast generally presents 'long white beaches bordered by thin belts of

feathery casuarinas.' This subjective impression is borne out by the historical pattern of entrepôts which grew up along the more northerly parts of the Malay Peninsula (Langkasuka, Tāmbraṅga, C'aiya), since the descriptions of the Chinese traveller Chao Ju-kua in 1225 A.D., appear to place all of the important commercial centers of the Malay Peninsula on the east coast bordering the Gulf of Siam (Wheatley 1961:64).

Today Malay speakers in the Malay Peninsula extend north only to the region of Songkhla (roughly 7 degrees north latitude) on the east coast, and the region of Phuket (roughly 8 degrees north latitude) on the west coast. The language boundary between Thai and Malay is thus located anywhere from 100 to 150 miles north of the political boundary between Thailand and Malaysia. North of approximately 7 degrees north latitude the east coast of the Malay Peninsula is now exclusively Thai-speaking. Since we know from documentary history that this is a relatively recent development certainly not preceding the 13th century, the people of Langkasuka, Tāmbraṅga and C'aiya were not Thais.¹⁷ This leaves essentially two possibilities: 1. they were AA speakers, or 2. they were AN speakers.

Wheatley (1961:Chs. 6,7), drawing mainly on the early records of Chinese Buddhist pilgrims, locates the states of Lang-ya-ssu-chia, Tan-ma-ling, Chia-lo-hsi and Teng-liu-mei on the east coast of the Malay Peninsula in 1225 A.D. The first of these is identified with Langkasuka 'a name which runs thread-like through Malayan history from the second to the early sixteenth century'. Langkasuka was located near modern Songkhla, at about 7 degrees north latitude, and judging from Malay traditional history, was Malay-speaking. Wheatley (map, p. 64) places Tan-ma-ling (Tāmbraṅga) to the north of Langkasuka at about 8-9 degrees, Chia-lo-hsi (C'aiya) just south of the Isthmus of Kra in the Takuapa-Bandon Bay region at between 9 and 10 degrees, and Teng-liu-mei far to the north at about 13 degrees north latitude. Teng-liu-mei is described as a dependency of the Khmer-speaking state of Chen-la, and the remaining states of the Malay Peninsula as dependencies of Srivijaya. But this was not always the case. Quaritch Wales (1976:10) notes that from the 3rd

to the 6th century Langkasuka, Tāmbraḷiṅga and C'aiya were dependencies of the powerful state of Funan in the Mekong delta. Following the decline of Funan in the late 6th century these peninsular states were independent briefly before becoming tributary to Srivijaya early in the 8th century.

The mere fact that certain of the 13th century east peninsular states mentioned in the Chinese sources were tributary to Srivijaya does not necessarily imply that they were Austronesian-speaking. However, two considerations favor this interpretation. First, with few exceptions the recorded history of warfare in mainland Southeast Asia has pitted distinct ethnolinguistic groups against one another (Burmese against Mon and Thai, Thai against Mon and Khmer, Khmer and Cham against Vietnamese, etc.). Although various of these groups were incorporated with others at different times, it was invariably through force, and predictably resulted in later rebellion. By contrast, Funan, of unknown linguistic affiliation, was commonly allied with Champa, apparently on friendly terms. Speculative as it must remain, this does suggest that the people of Funan and Champa shared a common cultural and linguistic heritage. If so, the seemingly peaceful rule first of Funan and later of Srivijaya over the small states of the Malay Peninsula would not be anomalous, since political control in this case as well would have benefitted from a shared cultural and linguistic heritage. Second, Malay traditional history suggests that Langkasuka was Malay-speaking. Given the apparently close political and trade connections which tied it to Tāmbraḷiṅga and C'aiya, it appears likely that the east coast of the Malay Peninsula early in the 13th century was Malay-speaking at least as far north as the Isthmus of Kra. This information takes us neither as far north nor as far back in time as we would like, but it does lend credence to the view that Austronesian-speaking peoples occupied much of the coastline of the Gulf of Siam prior to the expansion of the Thais into the Malay Peninsula.

These fragmentary historical indications of an earlier Austronesian presence around the Gulf of Siam are further supported by the archaeological record. Bellwood (1985:292) maintains that the culture complex associated with the slab graves and iron industry of Malaya 'provide the oldest archaeological

evidence which might tentatively be associated with the Austronesian coastal settlement of Malaya.' The sites to which he refers are located on both sides of the Malay Peninsula from Selangor in the south to Ongbah Cave at the head of the Gulf of Siam, where a similar iron industry dates to somewhat over 2,000 years B.P.

Conjecture (3) is based on inferences reached from the language distributions discussed in section 1. Two thousand years ago there were no Thai speakers on the Gulf of Siam or the northern neck of the Malay Peninsula, and apart from recent Austronesian-speakers who had arrived by sea, the eastern coast of the northern Malay Peninsula probably supported a sparse population of hunter-gatherers. However, north of the Mekong basin a substantial population of Austroasiatic-speaking peoples had been established for millenia. As Marrison (1975) notes, even the earliest documentary records for Cham show evidence of typological convergence toward a Mon-Khmer model.

Conjecture (4) is based on the sketchy early history of the Indianized state of Funan. According to Hall (1955:85) 'Funan proper stretched over southern Cambodia and Cochin China of modern times.' Its port, Oc Eo, probably was founded by the first century A.D. in the Mekong delta, a country 'intersected with innumerable channels which made it possible for Chinese travellers to 'sail across Funan' on their way to the Malay Peninsula.' The Chinese dynastic histories relating to Southeast Asia attest to the vigor and importance of this kingdom. There is no unambiguous record of its language, although Hall speculates (1955:23) that the people of Funan 'were Indonesians who were in a tribal state at the dawn of its history.' By the middle of the 6th century Funan had fallen to the Khmers.¹⁸

The fall of Funan probably has a special significance for understanding the history of AN languages on the Southeast Asian mainland. If Funan was AN-speaking, in the early centuries of the Christian era a single dialect chain would have extended almost unbroken from the southern tip of the Malay Peninsula to Champa. The expansion of the Khmers into the region of the Mekong delta would then have divided an earlier language continuum into two separate and smaller dialect chains, the

northern portion evolving into Chamic and the southern portion into at least a large part of Malayic.

Conjecture (5) is based on the historical records for the relatively late arrival of the Thais in the Malay Peninsula. According to Hall (1955:103) by the first half of the 12th century the Thais 'had begun to infiltrate into the Menam valley and had settled in the state of Lavo (Lopburi)', but apparently had not yet settled areas further to the south. Following the decline of Srivijayan hegemony over the peninsular states in the 13th century, the establishment of the Thai capital at Ayut'ia in 1350, and the sack of Angkor in 1431, there were no further obstacles to Thai expansion southward into the Malay Peninsula.

Conjecture (6) is based on the often noted evidence of Mon-Khmer typological and lexical influence on Acehnese. Although it is possible that this influence was acquired from a MK substratum on Sumatra itself, there is no direct evidence that MK speakers were ever established on the Sumatran mainland (as opposed to the Nicobar Islands to the north). Given the likelihood of an Achino-Chamic subgroup, the simplest hypothesis would appear to be that the Acehnese are a splinter group which migrated back from the Chamic-speaking region on the mainland to island Southeast Asia. It probably is not without significance that similar migrations of Cham speakers are attested within the historical period (Marrison 1951).

The hypothesis that Acehnese represents a back-migration from the Chamic-speaking region of mainland Southeast Asia raises a number of unanswered questions: 1. when did the migration occur?, 2. does Acehnese subgroup with Cham proper, or did it become separated at a time before the other Chamic languages had become clearly distinct from their immediate common ancestor?, 3. were there continuing contacts between Acehnese and Chamic after the migration to Sumatra? With regard to the first question, Durie (1985:2ff) notes that Marco Polo visited north Sumatra in 1292 and named six kingdoms, each described as having its own language. Five of these were on the Acehnese coast. Although Durie does not speculate as to whether any of the north Sumatran kingdoms described by Marco Polo were Acehnese-speaking, it would be surprising if this were not

the case, since the port of Ferlec (Perlak) was a commercial center which had already embraced Islam, and the name 'Aceh', 'which the Acehnese give to their language and culture' appears not long after 1500 with reference to the port kingdom at Banda Aceh. It seems likely, then, that Acehnese and Chamic have been separated for at least seven centuries.

With regard to the second question we have little information upon which to base an assessment of the internal divergence times of the Chamic languages themselves. Lee (1966) proposes a division into a northern branch represented by Jarai and Rhade, and a southern branch represented by the other languages. However, he does not consider Utset. Without presenting the supporting evidence Lee cites an averaged lexicostatistical percentage of 68 percent connecting Rhade with Cham, various forms of Roglai, Rai and Chru. This suggests a separation time of no more than a millenium, yet the highland Chamic peoples are animists, and show no indication of having ever embraced a state religion. Since Champa evidently was an Indianized state in the late 2nd century A.D. the lexicostatistical evidence clashes with that from other sources. To date no one who has taken note of the special similarity between Acehnese and Chamic has singled out a particular Chamic language for comparison. It thus appears that the separation occurred at a time when Chamic still formed a relatively undifferentiated dialect network. Little can be said with the evidence now available concerning question 3.

5. The Moken-Moklen settlement of mainland Southeast Asia. Since Moken-Moklen evidently does not subgroup with Malayo-Chamic and is not among its closest external relatives, the question naturally arises as to what the wider linguistic connections of this language might be.

The Moken are 'sea gypsies' who live in the island world of the Mergui Archipelago. As described by White (1922) they appear to be a fugitive population, avoiding contact with outsiders whenever possible, and often becoming victims of economic exploitation when it becomes impossible to maintain their isolation. The Moklen are former sea gypsies who have settled on the mainland of peninsular Thailand, and are rapidly being

assimilated culturally and linguistically to the dominant Thai (Larish p.c.). Both groups are neighbors of the Urak Lawoi', boat-dwelling speakers of a Malay dialect who inhabit the island chain to their south (Hogan 1988), and they have long been in contact with peninsular Malays as well (White 1922). Given these geographical relationships it is not surprising that Moken-Moklen contains a number of Malay loanwords. This observation is important as a clue to the more recent history of the Moken-Moklen, but it tells us nothing about their primary derivation.

A cursory inspection of the basic vocabulary of Moken, Acehnese, Gayō, Toba Batak, Rejang and Lampung shows little promise of producing a body of exclusively shared innovations linking Moken to any of the other languages of Sumatra. Nonetheless there are a few tantalizing indications of past contact with Acehnese, as with MOK *cicum* 'bird', ultimately of AA origin, but showing the diagnostic disyllabism of ACH *cicém*, and MOK *gelaŋ*, ACH *glaŋ* 'earthworm' (also MAL *gelaŋ-gelaŋ*, REJ *gloŋ* 'tapeworm', LPG *geloŋ* 'worm').

Perhaps the most salient character of the Moken as described by White (1922) is their elusiveness, seemingly the result of a deliberate avoidance of contact with outsiders. This trait marks them as fundamentally different from such sea nomads as the Sama Bajaw of the Sulu Archipelago, and apparently the Urak Lawoi' in the islands immediately to their south. Indeed, the Sama Bajaw have spread to a number of widely scattered locations in Indonesia, and have historically been actively engaged in trade with neighboring land peoples (Pallesen 1985). One could easily pass over this difference as simply another example of cultural variation, but we might also consider the possibility that the difference has a historical cause. The Acehnese now occupy the extreme northern tip of Sumatra, and apparently have done so at least since the introduction of Islam in the 13th century. As noted already, it appears that they represent a back-migration from the Chamic-speaking region of mainland Southeast Asia. The present linguistic diversity on the island of Sumatra and the Barrier Islands to its west cannot easily be explained unless we assume that Austronesian speakers settled the area at least a millenium before

the MC expansion. Since large parts of Sumatra are now occupied by speakers of MC languages it is worth asking whether the MC settlement of Sumatra could have displaced any of the earlier populations. Were the ancestral Moken-Moklen driven from the northern tip of Sumatra by the Acehnese? The Moken certainly are prime candidates for a refugee group, although it would perhaps be surprising that an event which could only have happened centuries ago would be in a sense institutionalized in the culturally-transmitted behavior of the people.

6. Epilogue: The longhouse and outrigger as ethnohistorical markers. The historical explanation offered here for the attested distribution of Malayo-Chamic peoples has made use of information from several disciplines, including comparative linguistics, archaeology, and documentary history. It may, therefore, not be out of place to conclude by considering the culture historical implications of two important trait distributions from the domain of material culture. In so doing we will look again, and in somewhat greater detail, at some of the major events in the prehistory of mainland Southeast Asia which preceded the advent of Austronesian speakers in the region.

More than 75 years ago Sapir (1968) showed convincingly that the distribution of culture traits can be resolved into a limited number of patterns, or types, each of which is optimally associated with a particular type of historical explanation. This method was developed in Blust (1981b), and applied predictively in Blust (1991) with regard to the distribution of traits involving beliefs about the spirit world. It can be applied with equal validity to the distribution of material culture traits.

House types in the AN world are quite varied. One of the most distinctive is the longhouse, a communal structure raised on pilings, often several hundred feet long, and divided along its length into a public gallery on the one side and a succession of individual family compartments on the other. Among Austronesian speakers, apart from the Gayō and Kerinci of Sumatra, the longhouse is attested exclusively from the island of Borneo, but it is not universal on that island. Lebar (1972) provides summary data on house types for the whole of Indonesia which shows an

interesting pattern: the longhouse is almost continuously distributed in southwest Kalimantan, Sarawak and Sabah, but is absent in the eastern (Tidong) and southeastern portions of the island (Ngaju Dayak, Ma'anyan). As noted in Blust (1987) there is neither linguistic nor distributional evidence that the longhouse is ancient in the Austronesian world.

Elsewhere Lebar, Hickey and Musgrave (1964) document the longhouse in three discontinuous parts of mainland Southeast Asia: 1) in a more-or-less continuous band running through southern Laos and neighboring regions of Vietnam and eastern Kampuchea, where it is used by such Mon-Khmer-speaking peoples as the Alak, Cao, Cheng, Halang, Halang Doan, Jeh, Kasseng, Katang, Ngeh, Ma, Sedang, So, Tau-oi and Mnong, and by such Austronesian-speaking peoples as the Jarai and Rhade, 2) among the hill Karens of Burma, and 3) among the Senoic peoples of the Malay Peninsula. What is most striking about this distribution is the concentration of the great majority of attested cases in an area in which Mon-Khmer and Chamic-speaking peoples have been in contact for perhaps two millenia. Such a distribution type is most simply explained through diffusion (Blust 1981b). The incorporation of Karen and Senoic changes the distribution type, but does not favor independent invention as an explanation, since the longhouse is ethnographically unattested outside a fairly compact region in Southeast Asia. If the distribution of the longhouse in the Laos-Vietnam-Kampuchea borderland is a product of diffusion, was it a trait that was introduced by AN-speakers to their MK neighbors, or vice versa? To answer this question we must consider the wider distributional picture. Since Senoic speakers are geographically isolated from other longhouse-using groups, it is difficult to explain the presence of the longhouse among them as a product of diffusion. Rather, the attested distribution among AA speakers suggests that the longhouse was in use among the common ancestors of at least the East and South Mon-Khmer-speaking peoples. Following the break-up of Proto-Mon-Khmer the longhouse was introduced into the Malay Peninsula by the Senoic peoples, who continued to construct it as they settled new territories in their southward migrations.

Given this interpretation we have answered our question regarding the direction of borrowing between Mon-Khmer and Chamic peoples: the latter acquired the longhouse from their Mon-Khmer neighbors. Although this analysis of the distributional facts is motivated by general principles of interpretation (Blust 1981b:286), it leaves the presence of the longhouse in Borneo unexplained. Independent invention is a possibility, but as noted above if the longhouse had been independently invented in Borneo we would expect a similar invention in other parts of the world.¹⁹ Since the ethnographic record shows no such distribution we are forced to accept diffusion as the best explanation for the presence of the longhouse in Borneo. But diffusion from where? As noted above, the attested distribution of the longhouse in Borneo includes only the northern and central portions of the island from about the Kapuas basin to Sabah --- in other words, that part which faces the South China Sea and, of course, some 600 miles across it the coast of southern Vietnam.

Human migrations seldom involve the removal of an entire population at one time, or the termination of contact between those who have left and those who have remained behind. The expansion of the Malayo-Chamic peoples out of southwest Borneo was a movement of maritime people. If the sea could be crossed in one direction it clearly could be crossed in the other direction as well, and it is very unlikely that there was no continuing contact between the Austronesian settlers of mainland Southeast Asia and those who remained behind in Borneo. Today the longhouse is found among many Bornean peoples who do not speak MC languages. But if a newly-acquired preference for living in longhouses had been brought back to Borneo by MC speakers who had acquired it from their MK neighbors the preference could have spread as easily in Borneo as it evidently did on the mainland.²⁰

Although it is not central to our discussion, we have omitted one detail from the above scenario which may shed light on the wider ethnohistory of mainland Southeast Asia, and which will lead us from a consideration of the longhouse to a consideration of the outrigger canoe. If the longhouse was an ancient Mon-Khmer house type which was introduced into the Malay

Peninsula in prehistoric times by the Senoic peoples, its presence among the Tibeto-Burman speaking Karen implies a period of significant contact between ancestral South Mon-Khmer and Karenic peoples. A similar inference might be drawn from the fact that the order of major sentence constituents in Karen, like that in neighboring Mon and Thai, is SVO, whereas in all other Tibeto-Burman languages except the heavily Sinicized Bai, the order is SOV (Matisoff 1992:481ff).²¹

Bellwood (1985:258-70) conveniently summarizes the archaeological literature on 'the Ban Kao Neolithic culture of southern Thailand and Malaya.' The type site of Ban Kao is a burial ground with associated habitations located in present-day Thailand near the head of the Malay Peninsula. Although the burials themselves have not been directly dated 'habitation layers in the site have a convincing spread of radiocarbon dates between the late third and mid-second millennia BC.' It has been suggested that the economy of the Ban Kao culture was based on rice agriculture, the domestication of pigs, probably fowls, and less certainly, cattle. Among articles of material culture are 1. a distinctive pottery made on a slow wheel, with many early forms having tripod bases, ring feet or high pedestals, 2. shouldered adzes, 3. stone (including nephrite) bracelets, 4. bone fish-hooks and combs, 5. clay bark cloth beaters and spindle whorls, and some possible semi-lunar stone reaping knives.

In the Malay Peninsula no open burial grounds of the Ban Kao type have been found, but a number of cave sites have been excavated with burials in association with a material culture that is clearly derived from the Ban Kao tradition to the north. The discussion in Bellwood (1985:258-70) leaves little doubt that prior to about 2,000 B.C. the Malay Peninsula was inhabited almost exclusively by hunter-gatherers of Hoabinhian culture, and probable Negrito physical type. Then, quite abruptly bearers of a Neolithic culture which included 1. grain agriculture, 2. pottery made on a slow wheel, 3. bark cloth and perhaps also woven cloth, and 4. probably some domesticated animals other than the dog, entered the region from the north. Craniometric and other physical measurements show that the bearers of the Ban Kao culture in Malaya were phenotypically different from those at Ban

Kao in Thailand, suggesting early population mixture between the incoming Neolithic migrants and the indigenous hunter-gathers. Who were the bearers of the Ban Kao culture? It is not difficult to eliminate 1. the indigenous Negrito population of Southeast Asia, 2. Tai-Kadai speakers (who were still far to the north), and 3. Austronesian speakers (who had a distinctively different material culture, and reached the Malay Peninsula by sea).

Diffloth (cited in Ruhlen 1987:156) recognizes a South Mon-Khmer subgroup with three primary branches: 1. Monic, 2. Aslian, 3. Nicobar Islands. Given these divisions, and the distribution of other Mon-Khmer languages further to the north, we would have to place the homeland of Proto-South Mon-Khmer around the head of the Malay Peninsula --- that is, in the matrix region of the Ban Kao archaeological culture. Since there is no other candidate for a modern continuation of the Ban Kao population in Malaya, we have no choice but to associate the Senoic peoples with the earliest Neolithic manifestations in Malaya.²²

The foregoing conclusions are based on 1. the archaeological record, and 2. the culture-historical implications of linguistic subgrouping. But why would an agricultural population have migrated southward from a relatively less forested region into the Malay Peninsula, for the most part a region of dense jungle? Distributional evidence suggests that the southward movement of Tibeto-Burman speakers out of the eastern Himalayas was spearheaded by the Karenic peoples. I have already suggested that this migration was well underway by perhaps 4,000 years ago, and Bellwood (1985:258-70) reports that the southward movement of the Ban Kao culture into the Malay peninsula (at least as far as northern Malaya) probably began around 2,000 B.C. There is thus reason to believe that two major north-to-south migrations were in progress at about the same time, and that one (Karenic) ended roughly where the other (Senoic) began.

The southward expansion of Tibeto-Burman speakers into mainland Southeast Asia was earlier described as a 'splitting' migration which separated the Munda languages and Khasi on the west from the rest of Austroasiatic on the east. However, from the standpoint of those who were in the direct path of the expansion,

it could equally well be described as a 'pushing' migration. In this larger context the movement of the Senoic peoples into the Malay Peninsula thus acquires a cause: circa 4,000 B.P. the region around the head of the Malay Peninsula was occupied by ancestral South Mon-Khmer speakers, known archaeologically as the Ban Kao culture. Southward-moving Tibeto-Burman speakers, spearheaded by the Karen, came into contact with this population and displaced a major part of it into the Malay Peninsula, after which it had no possible migration direction (at least by land) except southward. Another part remained in the general region of the South Mon-Khmer homeland, or was displaced westward, giving rise to the modern Mon, whose traditional territory included the deltas of the Irrawaddy, Sittang, and Salween Rivers, around the Gulf of Martaban (Lebar, Hickey and Musgrave 1964:95).

Unlike the later arrival of Malayic peoples by sea, the arrival of MK speakers in the Malay Peninsula was from the beginning an invasion of the interior rainforest (Bellwood 1985:160, map). North of the Kra ecotone is a zone of seasonal forest which yields to non-seasonal rainforest as one moves into the southern half of the peninsula. Southward-moving agriculturalists thus would have been exposed to some transitional environments before encountering the permanent non-seasonal rainforest of northern Malaya. While Malayic speakers settled the coasts throughout the lower half of the peninsula, maintained a littoral orientation, and only gradually built their hamlets further upriver, the Senoic settlers evidently advanced by paddling and poling canoes along the interior rivers where, by early in the second millenium B.C. they had introduced grain agriculture, riverbank longhouse settlements, pottery, and probably several domesticated animals including the dog, pig, and chicken. Here they came into significant and early contact with the indigenous Negrito hunter-gatherer population, which in time probably entered into a symbiotic relationship with them based on the exchange of jungle produce for manufactured goods. Over a period of several millenia there was cultural assimilation in both directions, and all of the Negrito groups of the peninsula adopted Mon-Khmer

languages, much as the Negrito hunter-gatherers of the Philippines adopted Austronesian languages (Reid 1987).

Despite the archaeological and distributional evidence for early Senoic settlement of the interior rainforest of Malaya, it would be surprising if no South Mon-Khmer-speaking groups which moved into the Malay Peninsula followed the coast. We would expect that any coastal migration would have followed the west coast rather than the east coast, since 1. historically the Mon were concentrated around the Gulf of Martaban, some 200 miles northwest of the northern extremity of the Malay Peninsula, and 2. the third coordinate branch of South Mon-Khmer comprises the languages of the Nicobar Islands, in the Bay of Bengal, some 320 miles west of the central Malay Peninsula. The separate geographical position of the Nicobarese naturally raises questions about 1. how long they have been in their present position, and 2. how they got there.

Archaeologically both the Nicobar Islands and the Andamans, some 100 miles to the north of them, are virtually a blank. However, if the Nicobarese languages are correctly classified in the South Mon-Khmer group, it is very unlikely that the Nicobars were settled by MK speakers earlier than perhaps 3,500 B.P. At the same time, if Nicobarese does form a *primary* branch of South Mon-Khmer it is unlikely that it has been separated from other South Mon-Khmer languages for less than 3,500 years. The advent of archaeological research in the area will eventually reveal whether this time frame is at all accurate. For now we can perhaps make more progress in considering how the Nicobarese reached their islands.

So far as can be determined the outrigger canoe was invented once in human history, by speakers of Austronesian languages either in Taiwan, or in the northern Philippines. It clearly was part of the cultural repertoire of speakers of Proto-Malayo-Polynesian by about 5,000 B.P. Since the outrigger canoe was the principal physical means which permitted the wide geographical expansion of the Austronesian-speaking peoples, its presence serves as a useful trace of prehistoric Austronesian contact among non-Austronesian-speaking groups. Doran (1981:78-79) conveniently maps the distribution of traditional single- and

double-outrigger canoes worldwide. The outrigger is completely absent from the Americas, Europe, East Asia, the Saudi Arabian Peninsula, and almost all of Africa. Elsewhere it is associated exclusively with Austronesian-speaking peoples, except in 1. coastal New Guinea, 2. eastern portions of the Cape York Peninsula of Australia, 3. the Nicobar Islands, 4. the Andaman Islands, 5. Sri Langka and the Malabar coast of India, and 6. parts of east Africa adjacent to Madagascar, in particular the Bajun and Comoro Islands, Zanzibar, and portions of the Mozambique coast (Murdock 1959:209). Although Doran includes the Nicobars and Andamans in his map of the distribution of the single-outrigger canoe, he does not discuss these areas. In each of the other areas in which the outrigger canoe is found in association with peoples who do not speak Austronesian languages the presence of the *outrigger is attributed to contact. There thus seems little reason to doubt that the use of the outrigger in the Nicobar and Andaman Islands is also a product of contact with Austronesian-speaking peoples. The crucial questions are: with which Austronesian-speaking peoples, and when?*

The distribution of the outrigger in Sri Langka, the Malabar coast of India, and the Mozambique coast and adjacent insular areas of east Africa is most plausibly seen as a cultural trail marking the route of the Malagasy migration from Southeast Borneo. As Adelaar (1989) has argued forcibly, there is reason to believe that after their departure from Southeast Borneo the ancestral Malagasy stopped first in south Sumatra, or the adjacent islands of Bangka and Billiton, where they were exposed to the Indianized culture of the maritime trading state of Srivijaya. Since the closest linguistic relatives of the Malagasy today are interior riverine peoples without ocean-going canoes, it is possible that the Malagasy acquired large ocean-going boats built in Srivijaya before departing from Southeast Asia. From there they evidently passed through the Strait of Malacca, following the west coast of the Malay Peninsula northward to the region of the Gulf of Martaban and the Irrawaddy delta, and thence followed the coast of India, the Saudi Arabian Peninsula, and east Africa to Madagascar. In areas that already had a well-developed tradition of overseas voyaging (as the Arabian Peninsula) the outrigger was

not borrowed; in areas that had no such tradition the outrigger was borrowed, provided that the contact situation was sufficiently long and meaningful for such cultural exchange to take place. The fact that Doran (1981) reports no instances of the outrigger on the Coromandel coast of India suggests that the Malagasy passed through this area relatively quickly. They then appear to have stayed long enough for significant borrowing to take place in Sri Lanka, and left an outrigger legacy along much of the Malabar coast from the southernmost tip of India to west of the Indus River in modern Pakistan.

Since this trail of material culture encircles the Bay of Bengal, one would naturally be inclined to look to the Malagasy as the source of the outrigger canoes of the Nicobar and Andaman Islands. However, I believe that such an interpretation would be misguided. First, there is every indication that the Malagasy followed the coast on their 8,000 mile voyage from Southeast Borneo to Madagascar. Such a sailing tradition is reflected linguistically in several of the languages of western Indonesia, including Malagasy (Malay *rantau* 'coast along a bay; coastal zone as opposed to the land', *me-rantau* 'to travel; to range; to emigrate in search of a living', Minangkabau *rantau* 'coastal zone; foreign country', *ma-rantau* 'travel, go to a foreign country', Malagasy *ranto* 'trade, commerce, traffic. Used of a journey to a distant part for commercial business', *mi-rànto* 'make an excursion to seek one's fortune'). If a coastal route had been followed, there is no reason why the Nicobars or Andamans would have been contacted during a migration from southeastern Borneo (or southern Sumatra) to Madagascar. Second, even if the Malagasy had passed through the Nicobars and Andamans, it is unlikely that the outrigger would have been borrowed without a significant period of contact. The outrigger canoe is a relatively complex piece of pre-industrial technology (Horridge 1987:ix compares the order of mechanical complexity to that of windmills or farm wagons), and probably could not be successfully reproduced through casual contact. It is difficult to escape the impression that the direction of the Malagasy migration was determined by prior knowledge of India, and it is equally difficult to imagine what could have motivated the Malagasy to remain in

either the Nicobar or the Andaman Islands long enough for a significant transfer of technology to take place.

A second possible source for the outrigger in the Nicobars and Andmans is contact with Malays. Since at least the nineteenth century, and surely for some time before it Malay traders have been in contact with the Nicobar Islands (Kloss 1903). However, there is no indication that the Nicobarese outrigger canoe is a recent acquisition. Writers as disparate as Kloss (1903:53ff) and Justin (1990:34ff) treat the outrigger as no less a part of traditional Nicobarese culture than pottery or horticulture. If the Nicobars were settled by a part of the South Mon-Khmer migration which followed the west coast of the Malay Peninsula, we have every reason to believe that these settlers arrived in the Nicobars with some form of agriculture, pottery, the use of large permanent houses built on stilts, and some domesticated animals including, but not limited to the dog.²³ Although Bellwood (1985:261) reports the discovery of 'two wooden oars carbon dated to within the first millenium B.C.' at the Dengkil site in Selangor, all that we can safely infer about the boating skills of early Senoic peoples in the Malay Peninsula is that they had canoes which could be paddled or poled along the rivers. When and where was the outrigger acquired by the ancestral Nicobarese? Neither linguistic evidence nor the evidence of traditional history suggests a recent introduction, and it appears at least a plausible hypothesis that the outrigger canoe was the means by which the Nicobar Islands were settled. This hypothesis is strengthened by the probability that the outrigger canoe of the Andaman Islands was acquired prehistorically from the Nicobars.²⁴

As seen already, Malayic speakers probably did not reach the Malay Peninsula until the 2nd or 3rd century B.C. If the Nicobarese represent a part of the South Mon-Khmer population of the Malay Peninsula which did not migrate to the Nicobar Islands until the arrival of Malayic speakers some two millenia ago, there should be clear linguistic evidence of an Aslian-Nicobarese subgroup. Since this does not appear to be the case we have little choice but to assume that the Nicobarese separated from Aslian speakers at time when proto-South

Mon-Khmer was still a single dialectally-complex language. If they did this as a result of acquiring the use of the sea-going outrigger canoe it must have been from Austronesian speakers who reached the Malay Peninsula and northern Sumatra well before the Malayo-Chamic migrations.

The non-Malayic languages of Sumatra include: 1. Gayō, 2. the Batak languages, 3. Simeulue, 4. Nias-Sichule, 5. Mentawai, 6. Enggano, 7. Rejang, and 8. Lampung. Although the exact classification of these languages remains unsettled, it is clear that few if any of them subgroup closely with Malayo-Chamic. Impressionistically it appears unlikely that the Batak languages, Simeulue, Nias, or Mentawai share a common ancestor with Malay which existed as a unity more recently than about 2,000 B.C., and the archaeological dating of Neolithic farming activities in Sumatra suggests a similar time frame for Austronesian settlement of the area (Bellwood 1985:231).

Although the only Austronesian languages recorded in the Malay Peninsula within the historical period are Malay dialects, Skeat and Blagden (1906:2:435ff) reported what they called a 'Generically Malayan element' in various of the Mon-Khmer languages, viz. a number of lexical items which appear to have cognates in Austronesian languages other than Malay. Given the probable antiquity of Austronesian settlement on Sumatra, it would be surprising if the Malay Peninsula did not also have an Austronesian-speaking population prior to the MC migrations. Since the Strait of Malacca separating Sumatra from the Malay Peninsula is no more than 30 miles wide at its narrowest point, it is also possible that some South Mon-Khmer speakers settled northern Sumatra. The ancestral Nicobarese, then, could well have acquired the outrigger canoe through contact with the earliest Austronesian-speaking arrivals on the west coast of the Malay Peninsula, or on the north coast of Sumatra. In either case the settlement of the Nicobars would have been an easy matter, involving a sailing distance of about 320 miles in the former case, or barely 100 miles in the latter.

Finally, to return to the Austronesian focus which is the central concern of this paper, it may not be amiss to briefly consider the MC migration in relation to the Malagasy migration

which took place some centuries later from southeast Borneo. Whereas a plausible case can be made that the latter was triggered by Indian contact, whether directly or mediated through the Indianized Malay-speaking maritime-trading states of South Sumatra (Adelaar 1989, Dahl 1991) this apparently was not true of the earlier migration. If the movement of MC peoples out of Borneo began in the 2nd or 3rd century B.C. as I have suggested, this would be too early to appeal to Indian contact as a significant factor. Moreover, it is difficult to escape the impression that the Malagasy were motivated by a desire to reach some specific destination (probably in India, although for reasons that may never be known they continued beyond). By contrast the MC migration appears to have been a product of rapid population expansion and the rise or intensification of inter-island trade connections. Its orientation was consistently toward the shorelines facing southwest Borneo, most notably Sumatra and the east coast of the Malay Peninsula. Metaphorically it could be said that the Malagasy migration described a long distance linear trajectory toward a specific target (which was not Madagascar), while the expansion of the MC peoples out of Borneo was a scattering migration concerned less with destination than with departure.

Appendix 1

Proposed Malayo-Chamic Lexical Innovations

The following set of reconstructions is offered as a sampling of likely MC lexical innovations. In nearly every case a cognate can be found in at least one other language of western Indonesia, and these are appended in notes to the primary comparisons. However, all of the languages in which such cognates occur are known to have borrowed extensively from Malay. I make the general covering assumption that in each of these cases the forms in MC are native, while those in non-MC languages with the possible exception of Sundanese and perhaps Madurese are Malay loans.

1. *baraŋ-baraŋ ‘things, belongings’

Chamic: CHAM *pìrāŋ-pìrāŋ* ‘everything’

ACH: *baraŋ* ‘goods, wares’

Malayic: IBAN *baraŋ* ‘things, goods, belongings, luggage’, MAL *baraŋ* ‘thing; stuff; wares; goods’, *baraŋ-baraŋ* ‘things of all sorts’

NOTE: Also KB *baraŋ*, *baraŋ-baraŋ* ‘goods, belongings, possessions’ (cited as a Malay loan), LPG *baxaŋ*, SND *baraŋ* ‘goods’, BAL *baraŋ* ‘thing, property, matter’, SAS *baraŋ* ‘thing, goods (cited as a Malay loan)’.

2. *binantu/*minantu ‘child-in-law’

Chamic: CHAM *mətəw* ‘son-in-law’, JAR *pətəw*, RHA *mtəw* ‘daughter-in-law, son-in-law’

Malayic: IBAN *menantu* ‘child-in-law; parent-in-law’, MAL *benantu/menantu*, MIN *binantu/minantu* ‘son- or daughter-in-law’

NOTE: Also SND *minantu*, NgD *manantu*, MLG *vinánto* ‘son- or daughter-in-law’.

3. *cium ‘kiss’ (PMP *hajek)

Chamic: JAR, RHA *cūm* ‘to kiss’

Malayic: BJR *cium* ‘kiss’, MAL *cium* ‘inhaling fragrance; kissing (in the Malay way)’, MIN *cium* ‘kiss’

NOTE: Also SND *cium*, MAD *ceom*, NgD *sium* ‘nose kiss’.

4. *cucu ‘grandchild’ (PMP *makempu)

Chamic: CHAM, RHA *co*, JAR *te-co* 'grandchild'

ACH: *cucò* 'grandchild'

Malayic: MAL *cucu* 'grandchild (of either sex)'

NOTE: Also REJ *cucu-ŋ* 'grandchild' (said to be a modern word), BAL, *Ba'amang cucu* 'grandchild'.

5. **diri*? 'body; self'

Chamic: CHAM *trəy*, JAR *drəy* 'body'

Malayic: Iban *diri?*, MAL *diri* 'self, oneself', MIN *diri* 'person, individual'

NOTE: Also KB *diri* '(rare) = *iba* 'self', SND *diri* 'self', JAV *diri* 'oneself, one's identity', MAD *diri?* 'self, oneself', BAL *dih* 'self, person' (said to be 'the true Balinese form of the commoner *diri*', the latter borrowed from Javanese), SAS *diri?* 'self'.

6. **empus* 'to blow' (PMP **hiup*)

Chamic: RHA *puh* 'to blow (of wind)'

Malayic: MAL *empus* 'blowing hard; panting and puffing; making a current of air'

NOTE: PMP **hiup* survived in Chamic: JAR *ayup* 'to blow'.

7. **hian* 'deity' (PMP **qatuan*)

Chamic: JAR, RHA *yāŋ* 'deity, god'

Malayic: IBAN *yaŋ* 'tutelary spirit appearing in dream or vision to bard and shaman', MAL *hian* 'divinity; Godhead'

NOTE: Also SND *hian* 'disappear'; SND *yan* 'originally animistic (Hindu) divinity'.

8. *kakay 'foot, leg' (PMP *qaqay)

Chamic: JAR *tə-kay* 'foot'

ACH: (*gaki*) 'foot'

Malayic: IBAN, MAL *kaki* 'leg, foot'

NOTE: Also SND, NgD *kaki* 'foot, leg'. Hardeland (1859) gives the Ngaju Dayak word as a Malay loan, and the same appears to be true of ACH *gaki*.

9. *kenal 'recognize a person' (PMP *kilala)

Chamic: RHA *knāl* 'recognize'

Malayic: IBAN *kenal* 'know, recognize, be acquainted with'; MAL *kenal* 'knowing by the use of the eyes; slight acquaintance'

NOTE: Also LPG *kenal* 'know someone', JAV (but not OJ) *kenal* 'know, be acquainted with', MAD *kennal* 'acquainted with'.

10. *kesuh 'breathe heavily, sigh'

Chamic: RHA *ksuh* 'breathe out, exhale; blow into (a straw, musical instrument)'

Malayic: IBAN *kesoh* 'breath, puff of wind; puff, blow, sigh', MAL *kesoh-kesoh* 'heavy sibilant breathing'

11. *kujur 'fish spear'

Chamic: RHA *kju* 'spear'

Malayic: MAL *kujur* 'fish-spear'

NOTE: Also DPB *kujur*, TB *hujur* 'lance'.

12. *males 'lazy'

Chamic: JAR, RHA *alah* 'lazy'

ACH: *malaiḥ* 'lazy, sluggish, slow'

Malayic: IBAN *malas* 'unwilling; lazy', MAL *malas*, MAL (Jakarta) *males* 'lazy'

NOTE: Also REJ *malas* 'lazy, lackadaisical', SAS *males* 'lazy' (said to be a Malay loan, but if so the last-syllable vowel is unexplained). Possibly **alas*, with a fossilized stative/attributive prefix **ma-* outside Chamic.

13. *malu 'shy, ashamed' (PMP **ma-hiaq*)

Chamic: CHAM, JAR *mələw* 'shame, ashamed', RHA *mləw* 'ashamed, embarrassed'

ACH: *malèë* 'ashamed, bashful, timid, shy'

Malayic: IBAN *malu* 'shy, modest; shame, ashamed; hurt someone's feelings', MAL *malu* 'feelings of shame, modesty or shyness (both in a good and a bad sense)'

NOTE: Also KB *ke-malū-n* 'shame, put to shame; the genitals', REJ *malew*, MAD *malo* 'ashamed', MLG *málo* 'bashfulness, shamefacedness'. SAS *malu* 'shy', *ke-malu-an* 'shyness; genitals' is explicitly marked as a Malay loan (Goris 1938:185).

14. *pulaw 'island' (PMP **nusa*)

Chamic: CHAM *palaw* 'island', RHA *plaw* 'large island'

ACH: *pulò* 'island'

Malayic: IBAN *pulau* 'island; copse of trees; isolate, separate, surround', MAL *pulau* 'isolated patch; island'

NOTE: Also MOK *polaw* 'island', KB *pulo* 'island; woods in which the village lies hidden', REJ *pulew*, SND, OJ, JAV, BAL *pulo*, MAD *polo* 'island', NgD *pulao* 'island; thicket, copse'. Whatever its antiquity *pulaw probably meant 'thicket, copse'. I assume that the meaning 'island', which has either replaced PMP *nusa or given rise to competing forms meaning 'island' in a number of languages, was a PMC innovation which later spread to other languages of western Indonesia through borrowing from Malay.

15. *sayap 'wing' (PMP *panij)

Chamic: JAR, RHA *siap* 'wing'

ACH: *sayeuëb* 'wing'

Malayic: IBAN, MAL *sayap* 'wing'

NOTE: Also REJ *sayap*, BAL (High) *sayap* 'wing' (said to be a Malay loan; cp. Low Balinese *kampid*). Li (1966) proposes PC *chiāp 'wing'.

16. *senan̄ 'satisfied, happy, contented'

Chamic: RHA *ênāŋ* 'happy, not angry, kind'

ACH: *seunan̄* 'satisfied, contented'

Malayic: IBAN *senan̄* 'comfortable, at ease, well-to-do', MAL *senan̄* 'ease, restfulness'

NOTE: Also KB *senan̄* 'find something good; enjoy, relax', REJ *senan̄* 'pleased, happy, glad', LPG *senan̄* 'happy', SND *senan̄* 'quiet, peace, calm', MAD *senen̄* 'ease, restfulness', NgD *sanan̄* 'quiet, calm'.

17. *tepuŋ ‘flour’

Chamic: JAR *təpuŋ*, RHA *kpuŋ*, ROGLAI *tupuk* ‘flour, meal’

ACH: *teupōŋ* ‘meal, rice flour’

Malayic: IBAN *tepong* ‘flour, esp. of rice, meal, paste’, MAL *tepoŋ* ‘flour; meal’

NOTE: Also KB, SAS *tepuŋ*, MAD *teppoŋ* ‘flour’, BAL *tepuŋ* ‘flour, meal’, NgD *tepoŋ* ‘bread, cake’.

18. *teraŋ ‘clear, distinct’

Chamic: RHA *klāŋ* ‘clear, easy to understand (of speech, sound, etc.)’

ACH: *traŋ* ‘clear, distinct’

Malayic: IBAN *teraŋ* ‘clear, bright; evident’, MAL *teraŋ* ‘clear; bright; obvious’

NOTE: Also KB *teraŋ* ‘light, day’, SND *teraŋ* ‘light, bright, clear’, JAV *teraŋ* ‘clear, cleared up’, MAD *terraŋ* ‘bright, clear’, BAL (High) *teraŋ* ‘clear, bright, light, visible’ (cp. Low Balinese *cedaŋ* ‘light’), SAS *teraŋ* ‘illuminated, bright’, NgD *tarāŋ* ‘bright, clear’.

19. *teruŋ ‘eggplant: *Solanum* spp.’

Chamic: JAR, RHA *troŋ* ‘eggplant’

Malayic: MAL *terōŋ* ‘eggplant; brinjal; aubergine’

NOTE: Also KB *teruŋ* ‘name of many edible and inedible plants of the cucumber and tomato groups’, LPG *tiuŋ* ‘eggplant’ (possibly a native word), SND *téroŋ* ‘name of a plant, the

nightshade; the fruits are eaten raw and cooked', MAD *terroŋ*, TAG *talóŋ* 'eggplant'.

20. *tikus 'mouse, rat' (PMP *labaw)

Chamic: CHAM *təkuh*, JAR *təkuih*, RHA *k'kuih* 'mouse, rat'

ACH: *tikōʔh* 'mouse'

Malayic: MAL *tikus* 'rat, mouse', MIN *tikuhi* 'mouse, rat'

NOTE: Also MOK *tiku:*, TB *tihus*, MAD *tekos*, GAYŌ, REJ, OJ, JAV *tikus*, BAL (High) *tikus* (cp. Low Balinese *bikul*) SAS *tikus* 'rat, mouse'.

21. *uraŋ 'person, human being' (PMP *tau)

Chamic: CHAM *uraŋ* 'person', JAR *araŋ araŋ* 'person (indefinite)', RHA *araŋ* 'one, someone'

Malayic: MAL *oraŋ* 'person, human being'

NOTE: Also KB *uraŋ* 'person, in the sense of belonging to an area, country, ethnic group or class (e.g. *uraŋ Aceh* 'an Acehnese')', OJ, JAV *woŋ* 'person, human being', MAD *(u)reŋ* 'person', BAL *woŋ* 'person belonging to a group (village, tribe, caste, country, etc.)'.

ACKNOWLEDGEMENT

I am indebted to Leonard Andaya, Geoffrey Benjamin, James T. Collins, and Gerard Diffloth for valuable criticism of an earlier version of this paper. The responsibility for any remaining errors of fact or interpretation is mine alone.

NOTES

1. Subgroupings are based on Ruhlen (1987), supplemented or qualified in some cases by additional information, as noted.
2. Sapir is careful to distinguish between *number of languages* and *number of primary branches*. It is only the latter which count toward determining the linguistic diversity of an area, and hence toward inferring what Sapir called 'the linguistic center of gravity'. Dyen (1956) formalized the notion under the name 'principle of least moves'.
3. In both cases typological differences (Munda vs. Mon-Khmer, Karen vs. Tibeto-Burman) may have given rise to exaggerated claims of the degree of actual divergence. Diffloth (p.c.), for example, has noted that the Munda languages share several phonological innovations uniquely with the North branch of Mon-Khmer, a fact which may in time lead to revision of the subgrouping picture for Austroasiatic. Based on lexicostatistical counts, Benjamin (1976:83) suggested that the Aslian group of South Mon-Khmer began to separate from other Austroasiatic languages by 5970 BP. In view of the relative closeness of the Aslian languages to one another, and the far greater divergence time that it implies for deeper splits within the AA group this date is surprisingly early, and in need of reassessment. The same date is reasserted in Benjamin (1986:4), although more recently Benjamin (p.c., dated April 20, 1993) has stated 'I no longer go along with the very great time-depths for Aslian that I published in 1976. Now, I would link the spread of Aslian to the Ban Kao/'Tripod' culture in the west of the peninsula around 3,000 B.P.'.
4. As a result of a recently completed linguistic reconnaissance of Mon-Khmer languages Diffloth (p.c.) now reports another Austronesian language, Chvea, spoken near Sisophon in Battambang province, Kampuchea. According to his information source this language, of which nothing further is known, is 'not Cham'.
5. In addition to these five historically attested language families it is likely that there was at least one other language family in mainland Southeast Asia which coexisted with Austroasiatic in the prehistoric period. The traditionally nomadic populations of the Malay Peninsula today speak AA languages, with heavy and apparently increasing borrowing from Malay (Benjamin 1976). Yet both their hunting-gathering economy and the (mixed) Negrito physical type of some groups suggest that their linguistic affiliation is a result of a change of language allegiance at some time in the prehistoric past. This view is broadly consistent with the occurrence of Negrito hunter-gatherers in the Philippines who now

speak AN languages, but apparently do so as a result of a similar change in language allegiance during the prehistoric period (Reid 1987), and with the occurrence of a physically more distinctive Negrito population in the Andaman Islands which has no close linguistic relatives anywhere. For further suggestions of a Negrito linguistic substratum in island Southeast Asia cf. Blust (1981b), and Reid (to appear).

6. James T. Collins (p.c.) disputes the immediate subgrouping connection of Acehnese and Chamic, maintaining that Acehnese probably subgroup with Malay, and the Malayo-Aceh group with Chamic. This point is, however, of little consequence in terms of the culture-historical schema I propose. Thomas and Healey (1962) combined the Chamic languages and Malay in their 'Malay Stock', a proposed genetic unit co-ordinate with: (1) their 'Philippine Superstock' and (2) their 'Southern Mindanao Family'. However, since the Chamic languages and Malay were the only non-Philippine languages in their study, their conclusion expresses nothing stronger than the claim that the Chamic languages are more closely related to Malay than to any language in (1) or (2).

7. Shorto (1975) places the weight of his claim for an 'Achino-Chamic' subgroup on three types of evidence: 1. the systematic correspondence of ACH /a/ to Cham /a/ and of ACH *ə*/ to Cham /ā/, 2. the occasional appearance of an unexpected intervocalic glottal stop in cognate forms in Acehnese and Chamic, and 3. the distinctive reduction of CVCVCV reduplications to CVC monosyllables both in Acehnese and in Chamic. Collins (1992) questions Shorto's subgrouping claim, maintaining that vowel lengthening in Chamic and breaking in Acehnese are historically unrelated, and hence not evidence of an exclusively shared phonological innovation. There are problems with Shorto's analysis which space does not permit us to address here, but this is no less true of the alternative which Collins offers, since it proposes rules of historical development with a number of unexplained exceptions. All in all the argument that Acehnese and Chamic have undergone a period of common development apart from all other AN languages seems to be well supported by: 1) the evidence of exclusively shared vocabulary, including the appearance of common Mon-Khmer loanwords (ACH *cicēm*, CHAM *cīm*, JAR, RHA *cim* 'bird'), 2) the reflexes of *R (/r/ in non-final position, zero in final position), 3) the distinctive treatment of CVCVCV reduplications, and 4) the loss of unstressed penultimate vowels in certain phonological environments, e.g., **puļuq* > ACH *plōh*, CHAM, JAR, RHA *pluh* 'ten'; **paqit* > ACH *phét*, CHAM, JAR *phi?*, RHA *phi?* 'bitter'; **telu* > ACH *lhèè*, CHAM, JAR *kləw*, RHA *tləw* 'three' (with different resolutions of the difficult initial cluster in Acehnese, Cham and Jarai).

8. Language abbreviations and principal sources of material are: PC : Proto-Chamic (Lee 1966, with modifications and additions), Cham (Moussay 1971), JAR : Jarai (Lafont 1968), RHA : Rhade (Tharp and Buon-ya 1980), Utset (Benedict 1984, Thurgood and Maddieson 1992), MAL : Malay (Wilkinson 1959), Kenindjal (Hudson 1970); ACH : Acehnese (Kreemer 1931), MOK : Moken-Moklen (Lewis 1960, Larish n.d., with diacritics omitted from the latter), KB : Karo Batak (Neumann 1951), Nias (Sundermann 1905), REJ :

Rejang (Jaspan 1984, Blust 1984), LPG : Lampung (Walker 1976), SND : Sundanese (Coolsma 1930), JAV : Javanese (Horne 1974), OJ : Old Javanese (Zoetmulder 1982), MAD : Madurese (Safioedin 1977), BAL : Balinese (Barber 1979), SAS : Sasak (Goris 1938), NgD : Ngaju Dayak (Hardeland 1859), MLG : Malagasy (Richardson 1885). In citing forms from the Chamic languages I write shwa as /ə/, and depart from the common practice of marking shortness with a superposed breve in favor of marking length with a superposed macron. For most other languages I cite forms in the orthography of the sources.

9. For non-medial *q in Maloh cf. *qatay > *até* 'liver', *taneq > *tanaʔ* 'earth'. For non-medial *q in Gayō cf. *qitem > *item* 'black', *putiq > *putih* 'white'. Nias has lost all original final consonants. Other examples of *q > /h/ in non-final position are *qumaŋ > *huma* 'hermit crab', and *paqa > *faha* 'thigh'. For initial reflexes of *q in Lampung cf. *qañud > *hañuʔ* 'swept away (by current of river)', and *qudip > *huxiʔ* 'live'.

10. James T. Collins (p.c.) has pointed out to me that 'There is very strong evidence that Malay was a very influential language in late 16th and 17th century Champa. Po Darma of the French research institute (EFEO) has shown me five separate Cham-Malay lexicons in manuscripts that date from that period! Apparently as A. Reid suggests, Malay was widely used in the ports of SEA; Po Darma adds that Malay was viewed as an essential language in magic and court rituals (hence the lexicons written in Chamic script). The fact that the Chams were well on the way to conversion to Islam in that period (the Bani sect in C. Vietnam) also suggests the role of Malay in that area.' These observations are important, and well-taken, but they have little bearing on the central issue of whether there is convincing evidence for a linguistic subgroup that includes Malayic, Chamic, and Acehnese. The shared innovations which are taken as evidence for such a subgroup are found throughout Chamic, and include many items of basic vocabulary. Malay may well have been an important contact language among the Chams (and even other Chamic-speaking groups) after the fall of Champa, but the fundamental linguistic affiliation of Chamic has not been obscured by borrowing from Malay any more than the fundamental linguistic affiliation of English was obscured by wholesale borrowing from French during the centuries of the Norman Conquest.

11. In both Chamic and Acehnese PMC *r disappeared in final position (*ikur > CHAM *iku*, JAR, RHA *aku*, ACH *iku* 'tail', *ulaR > CHAM *ula*, RHA, JAR *ala*, ACH *ula* 'snake'). For Moklen cf. *Ratus > *lato* 'hundred', *laŋit > *la: gét* 'sky' (also *ribut > Moken *lébot* 'storm'), *Rumaq > *omak* 'house', *uRat > *olat* 'tendon', *ulaR > *olan* 'snake', and *ikuR > *ikun* 'tail'; for Gayō cf. *daReq > *rayoh* 'blood', *baqeRu > *ayu* 'new', *kahiW > *kayu > *kayu* 'wood'; for Nias cf. *Ratus > *otu* 'hundred', *daReq > *do* 'blood', *baqeRu > *bohau* 'new'; for Lampung cf. *ribu > *xibu* 'thousand', *Ratus > *xatus* 'hundred', *duha > *xua* 'two', *baqeRu > *baxu* 'new', *baReq > *bayoh* 'swollen', *ulaR > *ulay* 'snake', *qatelur > *tahluy* (> Met.) 'egg'; for Balinese cf. *Ratus > *atus* 'hundred', *uRat > *uat* 'sinew, muscle, vein', *tabaR > *tabah* 'insipid, tasteless'.

12. The following material is intended to supplement that in Table 3: PMP *wani > KB *wani* 'honeybee'; *walu > LPG *walu* 'eight'; *wada > JAV *ora* 'not, no', *walu > JAV *wolu* (OJ *wwalu*) 'eight'; *wada > MAD *adha?* 'not have, be gone', *badha* 'have, exist', *walu > *ballu?* 'eight'. JAR *akha* 'root' appears to be cognate with ACH *ukheuē* 'root', and hence evidently does not reflect *wakaR. SND *akar* almost certainly is a loan from Malay, and the same may be true of BAL *akah*.

13. Adelaar (1992:194) notes that six of the seven Malayic isolects which he compared have a fossilized initial nasal in the reflex of *um-inum. The seventh, Banjar Hulu, has *kinum* 'drink', thus suggesting that the affix in PMP *um-inum was still separable in the Proto-Malayic reflex. In view of the overwhelming agreement of the other reflexes I assume that BH *kinum* is a product of borrowing or reshaping.

14. Evidently reflecting an affixed form of the doublet *inem (Blust 1983/84, no. 109). Jaspán (1984) also gives *mbu?biowa* 'drink' (lit. 'eat water').

15. Madurese may also subgroup closely with Malayo-Chamic, although it is difficult to distinguish exclusively shared lexical innovations from what may have amounted to massive relexification through centuries of contact with Malay.

16. According to Bellwood (1985:292) 'Linguistically...an origin for the Malays, and also for the other smaller *orang Melayu asli* groups such as the Temuan and Jakun...may perhaps be located in Sumatra...or western Borneo during the first millenium B.C.'. Of these alternatives the linguistic evidence clearly favors Borneo, and more particularly the region of the Kapuas River basin in the southwestern quarter of the island as the primary center of dispersal for the Malayic (and hence Malayo-Chamic) peoples. Given its proximity to southwestern Borneo, southeastern Sumatra must have been one of the areas settled earliest from the Bornean homeland.

17. Wyatt (1984:50ff) suggests that the Thai expansion into the Malay Peninsula began sometime after A.D. 1200 out of a complex of petty chiefdoms and principalities 'centered on the old towns of Suphanburi and Phetburi' (around the head of the Gulf of Thailand). The town of Nakhon Si Thammarat, mid-way down the east coast of the peninsula, apparently came under Thai control in the thirteenth century, although ethnically it almost certainly contained many non-Thai elements.

18. Leonard Andaya (p.c.) informs me that my depiction 'of the southward expansion of the Khmers overthrowing the Indianized state of Funan is contrary to historians' reconstruction of the period.' In particular, he maintains that some Southeast Asian historians believe there was a historical continuity from Funan through Chen-la to Angkor (hence forcing us to conclude that Funan was Khmer-speaking). Neither of the sources to which he refers me could be obtained, but recent writers such as Kenneth Hall (1985:76ff) suggest that Funan was ultimately dismembered by the Chams coming in from the east, and by 'the peoples of the Khmer domain known to the Chinese as Chen-la', who moved against Funan 'from modern Thailand and Cambodia to the north.' In short, the linguistic affiliation of Funan continues to be a major puzzle in the ethnolinguistic

history of mainland Southeast Asia, and nothing that is definitely known prevents us from assuming that this first Southeast Asian state was Austronesian-speaking.

19. Communal houses were, of course, traditionally used in some other parts of the world, as among the members of the League of the Iroquois in upstate New York, the Puebloan peoples of the American Southwest, or such South American tropical forest tribes as the Nambicuara (Bennett 1963). The culture-historical value of the Southeast Asian longhouse derives from its specific and distinctive construction, which includes: 1. a long rectangular form, 2. the use of pilings on which the floor rests, usually at least eight feet above the ground, and 3. the common (although not universal) longitudinal division into a public gallery and private family compartments.

20. The sole exception to this general pattern of distribution which Lebar (1972) lists is the Ot Danum of the Barito River basin in southeast Borneo who, unlike their closest linguistic relatives the Ngaju Dayak and Ma'anyan, live in longhouses. However the Ot Danum are the most interior of all Barito Family peoples (Hudson 1967), and hence were those Barito speakers most likely to have come in contact with the Malayic-speaking peoples of the Kapuas drainage, or with other longhouse-dwelling peoples from the northern portions of the island. In addition Loeb (1972:248) reports that the *umah*, or dwelling house of the Gayō of northern Sumatra, is a communal structure inhabited by a number of families, and Lebar (1972:29) states that the *umah* among the Kerinci of central Sumatra 'may be classed in the genre of Southeast Asian longhouses'. I assume that the longhouse in these two Sumatran societies was acquired in prehistoric times through contact, ultimately (although not necessarily directly) with speakers of Mon-Khmer languages.

21. Matisoff implies that the divergent word-order of Karen is a result of contact within the relatively recent past. I would like to leave open the possibility that an SVO order in Karen is much older, dating from the initial period of contact with a still largely undifferentiated Proto-South Mon-Khmer linguistic community located near the head of the Malay Peninsula.

22. Benjamin (1986:20ff), claiming support from several genetic studies, believes that the Malayan aborigines are a single polymorphic physical type, descended from a similar ancestral population. This view fails to explain 1. why a 'Negrito' physical type is also found in the Philippine, and Andaman Islands, 2. why this physical type shows intergradations with neighboring non-Negrito peoples both in the Philippines and the Malay Peninsula, where contact with non-Negrito peoples has been in progress for centuries, but not in the Andamans, which have remained relatively isolated, 3. why the Negrito physical type is associated with a traditional hunter-gatherer economy throughout Southeast Asia, whereas adjacent non-Negrito peoples are associated with agriculture and permanent settlements, or 4. why physically Negrito peoples speak Austronesian languages in the Philippines and Austroasiatic languages in Malaya, while the languages of the Andamans are internally highly divergent, and have no well-established relatives anywhere else. The most plausible account of these and other observations is that the modern Negritos of Southeast Asia continue an

- indigenous population of hunter-gatherers which was widespread throughout the region prior to the introduction of the Neolithic by what Bellwood (1985) calls 'Southern Mongoloid' peoples. Where there was contact between the two populations (as in the Philippines and Malaya), occasional intermarriage led to some intergradation of physical types; where contact was minimal (as in the Andamans) the Negrito physical type was not affected by Mongoloid admixture.
23. The materially more backward, semi-nomadic, forest-dwelling Shompen of Great Nicobar Island almost certainly represent a historically secondary hunter-gatherer adaptation due to environmental pressures operating on a small, originally agricultural population, much like the Tasaday of the Philippines, the Penan of Borneo, or the Kubu of Sumatra.
24. Radcliffe-Brown (1964:474ff, 492ff) speculated that the Andamanese possessed both pottery and the single outrigger canoe from a very early time, predating the isolation of Little Andaman from Great Andaman, if not actually predating the settlement of the Andamans from the Asian mainland. Although the relevant archaeology remains to be done, linguistic indications suggest that the Andamans have been settled for upwards of 10,000 years, and it appears most likely to me that both pottery and the outrigger canoe reached the Andamans from the Nicobars at some time after the settlement of the latter circa 3,500 B.P. How the Andamanese reached their islands in the first place remains a mystery (much like that surrounding the initial settlement of Australia), but the use of bamboo sailing rafts is a distinct possibility.

REFERENCES

- Adelaar, K.A. 1989. Malay influence on Malagasy: Linguistic and culture-historical implications. *Oceanic Linguistics* 28:1-46.
- . 1992. Proto-Malayic: The reconstruction of its phonology and parts of its lexicon and morphology. *Pacific Linguistics* C-119. Canberra: Department of Linguistics, Research School of Pacific Studies, The Australian National University.
- Aymonier, Étienne, and Antoine Cabaton. 1906. *Dictionnaire Cham-Français*. Paris: Imprimerie Nationale.
- Barber, C.C. 1979. *Dictionary of Balinese-English*. 2 vols. University of Aberdeen.
- Bellwood, Peter. 1985. *Prehistory of the Indo-Malaysian Archipelago*. Sydney: Academic Press.
- Benedict, Paul K. 1984. Austro-Tai parallel: A tonal Chamic language on Hainan. *Computational Analyses of Asian & African Languages*, 22:83-86. Tokyo: Project on Lexicological Analysis, National Inter-University Research Institute of Asian & African Languages & Cultures.
- Bennett, Wendell C. 1963 [1947]. *Habitations*. In Julian H. Steward, ed., *Handbook of South American Indians*, Vol. 5, *The comparative ethnology of South American Indians*:1-51. Smithsonian Institution, Bureau of

- American Ethnology, Bulletin 143. New York: Cooper Square Publishers, Inc.
- Benjamin, Geoffrey. 1976. Austroasiatic subgroupings and prehistory in the Malay Peninsula. In Philip N. Jenner, Laurence C. Thompson, and Stanley Starosta, eds., *Austroasiatic Studies, Part I. Oceanic Linguistics Special Publication No. 13*: 37–128. Honolulu: The University Press of Hawaii
- . 1986. *Between isthmus and islands: Reflections on Malayan palaeo-sociology*. Singapore: National University of Singapore, Department of Sociology Working Papers, No. 71.
- Blagden, C.O. 1929. *Achinese and Mon-Khmer. Feestbundel uitgegeven door het Koninklijk Bataviaasch Genootschap van Kunsten en Wetenschappen bij gelegenheid van zijn 150 jarig bestaan 1778–1928*. Vol. 1:35–38. Weltevreden: G. Kolff.
- Blood, Doris Walker. 1962. Reflexes of Proto-Malayo-Polynesian in Cham. *Anthropological Linguistics* 4.9:11–20.
- Blust, Robert. 1977. The Proto-Austronesian pronouns and Austronesian subgrouping: A preliminary report. *Working Papers in Linguistics* 9.2:1–15. Honolulu: Department of Linguistics, University of Hawaii.
- . 1981a. The reconstruction of Proto-Malayo-Javanic: An appreciation. *Bijdragen tot de Taal-, Land- en Volkenkunde* 137:456–69.
- . 1981b. Linguistic evidence for some early Austronesian taboos. *American Anthropologist* 83.2:285–319.
- . 1983/84. Austronesian etymologies - II. *Oceanic Linguistics* 22/23:29–149.
- . 1984. On the history of the Rejang vowels and diphthongs. *Bijdragen tot de Taal-, Land- en Volkenkunde* 140:422–51.
- . 1984/85. The Austronesian homeland: A linguistic perspective. *Asian Perspectives* 26.1:45–67.
- . 1987. Lexical reconstruction and semantic reconstruction: The case of Austronesian “house” words. *Diachronica* IV.1-2:79–106.
- . 1991. On the limits of the ‘thunder complex’ in Australasia. *Anthropos* 86.4/6:517–28.
- Collins, I.V. 1975. *The Austro-Asiatic substratum in Acehnese*. Unpublished Ph.D. dissertation. Berkeley: Department of Linguistics, University of California.
- Collins, James T. 1992. Chamic, Malay, and Acehnese: the Malay world and the Malayic languages. In *Le Campā et le monde Malais: Actes de la conférence internationale sur le Campā et le monde Malais*: 109–121. Paris: Publications du Centre d’histoire et Civilisations de la Péninsule Indochinoise.
- Coolsma, S. 1930 [1884]. *Soendaneesch-Hollandsch woordenboek*. 2nd ed. Leiden.
- Cowan, H.K.J. 1948. Aanteekeningen betreffende de verhouding van het Atjéhsch tot de Mon-Khmer talen. *Bijdragen tot de Taal-, Land- en Volkenkunde* 104:429–514.

- . 1974. Evidence of long vowels in early Acehnese. *Oceanic Linguistics* 13:187–212.
- . 1991. Acehnese dialects in connection with Chamic migrations. In Ray Harlow, ed., *VICAL 2: Western Austronesian languages and contact languages. Papers from the Fifth International Conference on Austronesian Linguistics, Parts 1 & 2*: 53–83. Auckland: Linguistic Society of New Zealand.
- Dahl, Otto Chr. 1976 [1973]. *Proto-Austronesian*. Scandinavian Institute of Asian Studies Monograph Series, No. 15. London: Curzon Press.
- . 1991. *Migration from Kalimantan to Madagascar*. Oslo: The Institute for Comparative Research in Human Culture.
- Dempwolff, Otto. 1934–38. *Vergleichende Lautlehre des austronesischen Wortschatzes*, I. Band: Induktiver Aufbau einer indonesischen Ursprache, II. Band: Deduktive Anwendung des Urindonesischen auf austronesische Einzelsprachen, III. Band: Austronesisches Wörterverzeichnis. *Zeitschrift für Eingeborenen-Sprachen*, Beihefte 15, 17, 19. Berlin: Reimer.
- Doran, Edwin, Jr. 1981. *Wangka: Austronesian canoe origins*. College Station: Texas A&M University Press.
- Durie, Mark. 1985. *A grammar of Acehnese on the basis of a dialect of North Aceh*. *Verhandelingen van het Koninklijk Instituut voor Taal-, Land- en Volkenkunde* 112. Dordrecht: Foris Publications.
- Dyen, Isidore. 1953. *The Proto-Malayo-Polynesian laryngeals*. Baltimore: Linguistic Society of America.
- . 1956. Language distribution and migration theory. *Lg.* 32:611–26.
- . 1965. A lexicostatistical classification of the Austronesian languages. *IJAL Memoir* 19. Baltimore: Waverly Press.
- . 1971. The Chamic languages. In Thomas A. Sebeok, ed., *Current Trends in Linguistics*, vol. 8: *Linguistics in Oceania*: 200–210. The Hague: Mouton.
- Edmondson, Jerold A. and Kenneth J. Gregerson. to appear. *Western Cham as a register language*. To appear in an *Oceanic Linguistics Special Publication on tonality*.
- Goris, R. 1938. *Beknopt Sasaksch-Nederlandsch woordenboek*. Singaradja, Bali.
- Hall, D.G.E. 1955. *A history of South-East Asia*. London: Macmillan & Co.
- Hall, Kenneth R. 1985. *Maritime trade and state development in early Southeast Asia*. Honolulu: University of Hawaii Press.
- Hardeland, August. 1859. *Dajacksch-Deutsches Wörterbuch*. Amsterdam: Frederik Muller.
- Haudricourt, André G. 1965. Problems of Austronesian comparative philology. *Lingua* 14:315–29.
- Hogan, David W. 1988. *Urak Lawoi': Basic structures and a dictionary*. Pacific Linguistics C-109. Canberra: Department of Linguistics, Research School of Pacific Studies, Australian National University.
- Horne, Elinor Clarke. 1974. *Javanese-English dictionary*. New Haven: Yale University Press.

- Horridge, Adrian. 1987. Outrigger canoes of Bali and Madura, Indonesia. Bishop Museum Special Publication 77. Honolulu: Bernice P. Bishop Museum.
- Hudson, A.B. 1967. The Barito isolects of Borneo: A classification based on comparative reconstruction and lexicostatistics. Southeast Asia Program Data Paper No. 68. Ithaca: Cornell University.
- . 1970. A note on Selako: Malayic Dayak and Land Dayak languages in western Borneo. *The Sarawak Museum Journal* 18:301–18.
- Jaspan, M.A. 1984. Materials for a Rejang-Indonesian-English dictionary. Pacific Linguistics D-58. Canberra: Department of Linguistics, Research School of Pacific Studies, Australian National University.
- Justin, Anstice. 1990. The Nicobarese. Calcutta: Seagull Books (on behalf of the Anthropological Survey of India).
- Kern, Hendrik. 1917 [1889]. Taalkundige gegevens ter bepaling van het stamland der Maleisch-Polynesische volken. In *Verspreide Geschriften* 6:105–20. The Hague.
- Kloss, C. Boden. 1903. In the Andamans and Nicobars: The narrative of a cruise in the schooner "Terrapin," with notices of the islands, their fauna, ethnology, etc. London: John Murray.
- Kremer, J. 1931. *Atjéhsch handwoordenboek (Atjéhsch-Nederlandsch)*. Leiden: Brill.
- Lafont, Pierre-Bernard. 1968. *Lexique Français-Jarai-Viêtnamien (parler de la province de Plei Ku)*. Publications de l'école Française d'Extrême-Orient. Paris.
- Larish, Michael D. n.d. *Moklen-English vocabulary*. Ms., 19pp.
- Lebar, Frank M. 1972. Ethnic groups of insular Southeast Asia, vol. 1: Indonesia, Andaman Islands, and Madagascar. New Haven: Human Relations Area Files Press.
- , Gerald C. Hickey and John K. Musgrave. 1964. Ethnic groups of mainland Southeast Asia. New Haven: Human Relations Area Files Press.
- Lee, Ernest Wilson. 1966. Proto-Chamic phonologic word and vocabulary. Unpublished Ph.D. dissertation, Indiana University. Ann Arbor: University Microfilms International.
- Lewis, M. Blanche. 1960. Moken texts and word-list: A provisional interpretation. *Federation Museums Journal* (n.s.), vol. 4. Kuala Lumpur.
- Loeb, Edwin M. 1972 [1935]. *Sumatra: its history and people*. Oxford in Asia paperbacks. Kuala Lumpur: Oxford University Press.
- Marrison, G.E. 1951. The Chams of Malacca. *JMBRAS* 24.1:90–98.
- . 1975. The early Cham language and its relationship to Malay. *JMBRAS* 48.2:52–60.
- Maspéro, Georges. 1928. *Le royaume de Champa*. Paris & Bruxelles: G. van Oest.
- Matisoff, James A. 1992. Sino-Tibetan linguistics: Present state and future prospects. *Annual Review of Anthropology* 20:469–504.
- Moussay, Gérard. 1971. *Dictionnaire Cam-Viêtnamien-Français*. Phan Rang: Centre Culturel Cam.

- Murdock, George Peter. 1959. *Africa: Its peoples and their culture history*. New York: McGraw-Hill.
- Neumann, J.H. 1951. *Karo Bataks-Nederlands woordenboek*. Medan, Sumatra.
- Niemann, G.K. 1891. Bijdrage tot de kennis der verhouding van het Tjam tot de talen van Indonesië. *Bijdragen tot de Taal-, Land- en Volkenkunde* 40:27-44.
- Nothofer, Bernd. 1975. The reconstruction of Proto-Malayo-Javanic. *Verhandelingen van het Koninklijk Instituut voor Taal-, Land- en Volkenkunde*, 73. The Hague: Nijhoff.
- Oey, Eric M. n.d. The linguistic position of the Chamic languages. Ms., 44pp.
- Pallesen, A. Kemp. 1985. Culture contact and language convergence. Manila: Linguistic Society of the Philippines.
- Quaritch Wales, H.G. 1976. *The Malay Peninsula in Hindu times*. London: Bernard Quaritch.
- Radcliffe-Brown, A.R. 1964 [1922]. *The Andaman Islanders*. The Free Press of Glencoe.
- Reid, Lawrence A. 1987. The early switch hypothesis: Linguistic evidence for contact between Negritos and Austronesians. *Man and Culture in Oceania* 3 (Special Issue):41-59.
- . to appear. Possible non-Austronesian lexical elements in Philippine Negrito languages. *Oceanic Linguistics*.
- Richardson, J. 1885. *A new Malagasy-English dictionary*. Antananarivo: London Missionary Society.
- Ruhlen, Merritt. 1987. *A guide to the world's languages*, vol. 1: classification. Stanford University Press.
- Safioedin, Asis. 1977. *Kamus Bahasa Madura-Indonesia*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa
- Sapir, Edward. 1968 [1916]. Time perspective in aboriginal American culture: A study in method. In David G. Mandelbaum, ed., *Selected Writings of Edward Sapir in language, culture and personality*: 389-467. Berkeley: University of California Press.
- Schmidt, Wilhelm. 1926. *Sprachfamilien und Sprachenzirkel der Erde*. Heidelberg.
- Shipley, William F. 1978. Native languages of California. In William C. Sturtevant, ed., *Handbook of North American Indians*, vol. 8: California (Robert F. Heizer, volume editor): 80-90. Washington: Smithsonian Institution.
- Shorto, H.L. 1975. Achinese and mainland Austronesian. *BSOAS* 38:81-102.
- Skeat, Walter William and Charles Otto Blagden. 1906. 2 vols. *Pagan races of the Malay Peninsula*. London: Macmillan and Co. Ltd.
- Sundermann, H. 1905. *Niassisch-Deutsches Wörterbuch*. Moers.
- Teeuw, A. 1961. A critical survey of studies on Malay and Bahasa Indonesia. Koninklijk Instituut voor taal-, land- en volkenkunde Bibliographical Series 5. The Hague: Nijhoff.

- Tharp, J.A. and Y-Bham Buon-ya. 1980. A Rhade-English dictionary with English-Rhade finderlist. *Pacific Linguistics* C-58. Canberra: Department of Linguistics, Research School of Pacific Studies, The Australian National University.
- Thomas, David, and Alan Healey. 1962. Some Philippine language subgroupings: A lexicostatistical study. *Anthropological Linguistics* 4.9:21-33
- Thomas, Dorothy. 1963. Proto-Malayo-Polynesian reflexes in Rade, Jarai, and Chru. *Studies in Linguistics* 17:59-75.
- Thurgood, Graham and Ian Maddieson. 1992. A radical typological shift: From disyllabic and atonal to monosyllabic and tonal. Paper presented at SEALS II, Arizona State University, 1992.
- Walker, Dale F. 1976. A grammar of the Lampung language: The Pesisir dialect of Way Lima. NUSA. Jakarta.
- Wheatley, Paul. 1961. The golden Khersonese: Studies in the historical geography of the Malay Peninsula before A.D. 1500. Kuala Lumpur: University of Malaya Press.
- White, Walter G. 1922. The sea gypsies of Malaya. London: Seeley, Service.
- Wilkinson, R.J. 1959. A Malay-English dictionary (romanised). 2 parts. London: Macmillan & Co.
- Wolff, John U. 1976. Malay borrowings in Tagalog. In C.D. Cowan and O.W. Wolters, eds., *Southeast Asian history and historiography: Essays presented to D.G.E. Hall*: 345-67. Ithaca: Cornell University Press.
- Wolters, O.W. 1967. *Early Indonesian commerce: A study of the origins of Srivijaya*. Ithaca: Cornell University Press.
- Wyatt, David K. 1984. *Thailand: A short history*. New Haven: Yale University Press.
- Zoetmulder, P.J. 1982. *Old Javanese-English dictionary*. 2 vols. The Hague: Nijhoff.