

# 11 *Language history and culture groups among Austroasiatic-speaking foragers of the Malay Peninsula*

---

NICLAS BURENHULT, NICOLE KRUSPE AND MICHAEL DUNN

## 1 Introduction

The Malay Peninsula is a crossroads for people, languages and cultural influences, apparent in today's vibrant mix of Malay, Chinese, Indian, Thai and European. Yet this modern state of affairs all but conceals signals of much older situations of diversity. Thus, some 140,000 people grouped together under the label *Orang Asli* (Malay for 'aboriginal people') represent a range of cultural and biological adaptations and linguistic diversifications with roots far back in prehistory. These 20-plus ethnolinguistic groups represent a unique and vanishing window on the history of human diversity in the region, and they offer intriguing examples relevant to more general issues of the dynamics of human societies.

By synthesising the current ethnographic, linguistic and genetic body of knowledge about these groups with our own quantitative analyses of new lexical data from 27 language varieties, we explore the local historical relationships and interaction between languages and cultures. Specifically, we look at the relationship between a particular subsistence mode, namely nomadic foraging, and the Aslian branch of the Austroasiatic language stock. While foraging has been considered in many previous accounts to have a historically close connection to one particular sub branch of Aslian (Northern Aslian), we highlight several mismatches in this correlation and take a step toward disentangling a complex picture of linguistic history and contact.

## 2 The forager problem

The Northern Aslian (Aslian, Austroasiatic) languages of the Malay Peninsula have long been considered to be closely associated with a particular societal and economic tradition of nomadic foraging, upheld by the so-called Semang (Benjamin 1985; Rambo 1988; Fix 1995, 2002; Bulbeck 2004). The Semang comprise ten or so ethnolinguistic groups scattered over parts of the Malay Peninsula (Peninsular Malaysia and Isthmian Thailand), and together they number about 3000–4000 individuals. Along with proposed similar correlations of language and societal-economic system in other Aslian-speaking settings in the peninsula (Central Aslian with Senoi swidden cultures and Southern Aslian

with Aboriginal Malay collector-traders), the notion of a forager-Northern Aslian connection has shaped much of the discussion of peninsular prehistory.

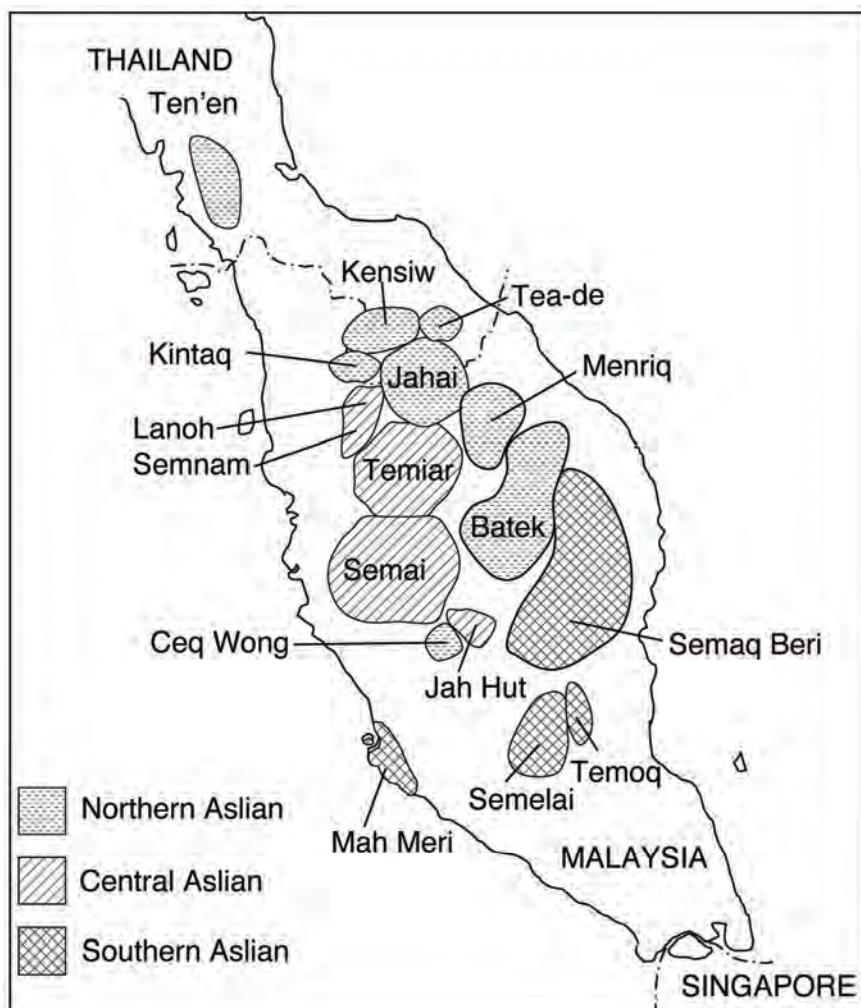
But the forager-Northern Aslian connection is only a near-match. A small group of foragers (c. 300) speak dialects of Lanoh, which belong to the Central Aslian branch of Aslian. Conversely, Ceq Wong, a geographical outlier of Northern Aslian, is spoken by a group of about 300 individuals who are not classified as Semang and whose subsistence is not focused on foraging. Furthermore, Semaq Beri is a Southern Aslian language spoken by a group of c. 2300 people who are not Semang but whose economy contains a significant component of nomadic foraging. Some earlier models tend to treat these ‘mismatches’ as exceptions, mixed societies which are presumably the results of recent departures from the main societal-linguistic categories.

The forager-Northern Aslian near-match is overlapped by a biologically defined category: for a long time, observers have maintained that there is a close connection between the Semang societal sphere and the physical features of its bearers. The reportedly short stature, dark skin and curly hair of the Semang led early anthropologists to classify them physically as ‘Negritos’ (see for example Schebesta 1952), a term which is still used to some extent. They were considered to represent the oldest human stratum in the peninsula. Superficial physical characteristics have been played down or questioned in subsequent accounts, and it has been shown that Semang history goes no further back than that of other indigenous groups in the peninsula (Bulbeck 2003). Still, recent genetic studies confirm that there is a close association between the Semang and an ancient local genetic lineage which goes far back into the Pleistocene and is represented only partially in other peninsular populations (Hill et al. 2006; Oppenheimer, this volume). Burenhult (forthcoming) proposes that continuous mobile foraging has been the best subsistence niche for preserving this ancient genetic lineage, and that today’s forager groups represent the current cultural exploiters of that niche. Yet the Semaq Beri—speakers of a Southern Aslian language who are not classified as Semang and whose relationship with the ancient genetic lineage has not been examined—also lead a predominantly mobile, foraging way of life.

Considering these categorical discrepancies, which genealogical histories and interactional dynamics can account for the relationship between language, culture and genes currently observed in this setting? With a focus on linguistic issues, the following sections revisit Aslian history and research (§3), introduce the ethnolinguistic categories concerned (§4), explore the nature and degree of linguistic contact between them (§5), and discuss the findings in light of current genetic and ethnographic knowledge (§6).

### 3 Aslian history and research

Aslian is a genealogically and geographically well-defined branch of the Austroasiatic language stock (see Figure 1). It is made up of some 20 languages spoken by minority groups in the Malay Peninsula, mainly in the rainforested areas of the interior. Lexicostatistics (Benjamin 1976) and comparative historical phonology (Diffloth 1975) have produced similar family trees for Aslian, with three main subgroups: Northern, Central and Southern Aslian. The three are considered to have branched off early from each other, soon after their common ancestral language arrived in the peninsula an estimated 4000–5000 years ago.



**Figure 1:** Map of the Malay Peninsula showing the approximate distribution of Aslian languages and subgroups (adapted from Benjamin 1976 and Burenhult, forthcoming).

The three Aslian clades coincide broadly with three ethnographically defined subgroupings of indigenous societal and economic features (Benjamin 1980, 1985). Thus, according to this widely accepted classification, Northern Aslian is by and large associated with the nomadic foragers known ethnographically as the Semang; Central Aslian is associated with semi-sedentary swidden horticulturalists referred to as the Senoi; Southern Aslian is linked to groups of collector-traders called Aboriginal Malay (and, in Benjamin's work, to a larger Malayic societal pattern). Benjamin's tripartite societal division is based primarily on distinct categories of specific institutionalised kinship regimes, namely marital patterns and kin-avoidance rules. Other societal and economic features map more or less well onto these categories, forming the more general and less robust Semang-Senoi-Malay patterns (Benjamin 1985, forthcoming). Importantly, some Aslian ethnolinguistic groups are difficult to classify according to this more general ethnographic division, either because they display a mix of societal-economic features, or because their linguistic identity does not match the expected societal-economic one (see further below).

Despite its complexities, the tripartite model of peninsular ethnography has provided the analytical categories and the basis for sampling in a number of subsequent studies of

Malayan indigenous cultures and history (see especially Rambo 1988; Fix 1995, 2002, this volume; Bulbeck 2004, this volume; Hill et al. 2006; Oppenheimer, this volume). Benjamin has always cautiously pointed to the complexities of his model. However, some authors' subsequent analytical employment of it has frequently focused on the main patterns and has taken the broad association between societal type and language group for granted. As a result, those ethnolinguistic groups which do not conform to the main patterns of the model have typically attracted less attention.<sup>1</sup>

Accordingly, most interpretations of the prehistory of the Malay Peninsula rely on Benjamin's 1985 classification and the associated Aslian family tree. Models by Rambo (1988) and Fix (1995) imply that Benjamin's categories can be projected into the past and suggest a common genetic, linguistic and cultural origin of all of the peninsula's indigenous groups (cf. Bulbeck 2004). This 'indigenist' perspective has largely developed in response to 'migrationist' models which emphasise demic diffusion and linguistic colonisation from the Southeast Asian mainland to the north during the Neolithic (see especially Bellwood 1985, 1993).<sup>2</sup>

Recent genetic studies pose a problem for the indigenist paradigm, essentially because they show that genetic lineages of the indigenous groups display various geographical origins and varying antiquity within the peninsula (Hill et al. 2006; Oppenheimer, this volume). Locally ancient (Pleistocene) haplotypes are present in all groups but show a particular association with the Semang. Furthermore, there is evidence of a considerable influx of lineages from Mainland Southeast Asia around the time of the arrival of agriculture, c. 5000 years ago. These lineages are particularly apparent in the Senoi group.

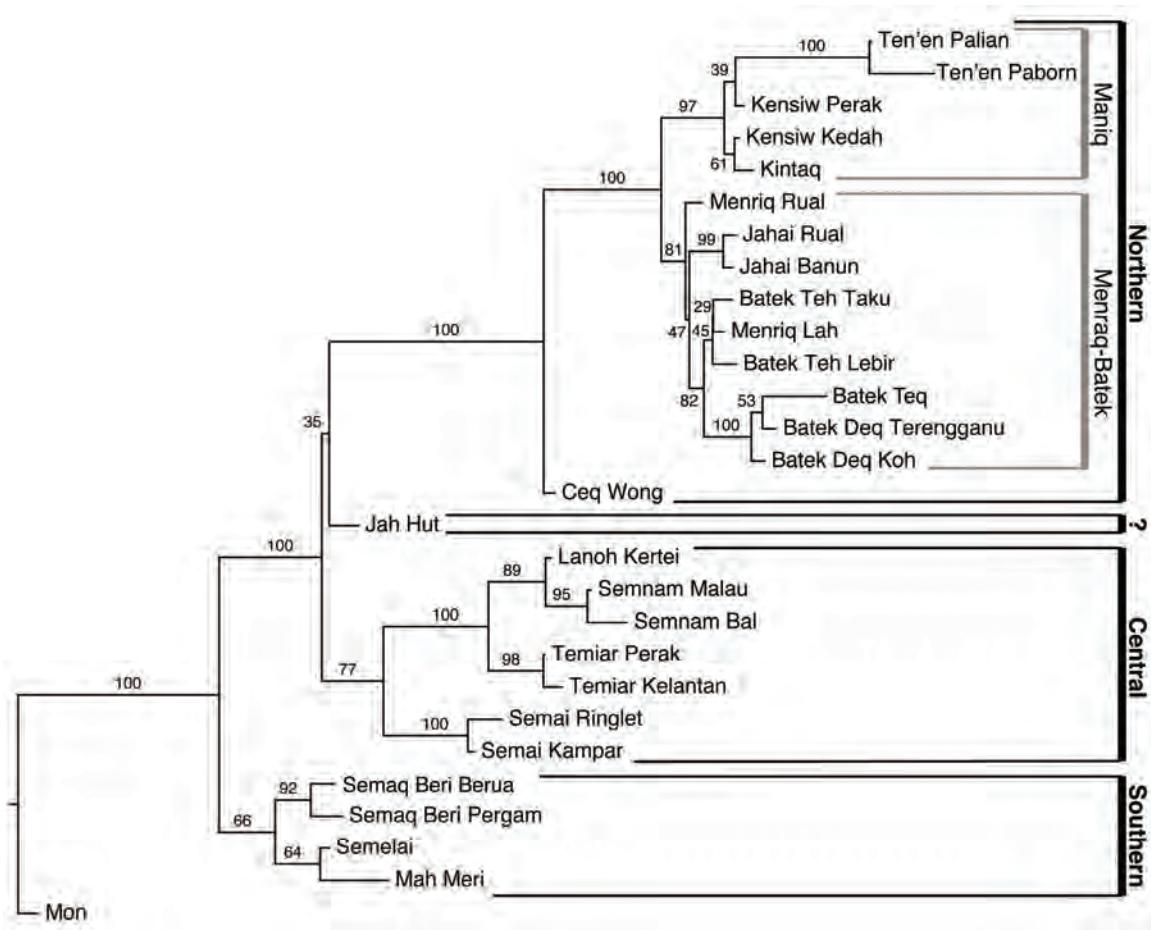
These advances have prompted significant reinterpretations of genetic and linguistic prehistory in the peninsula. For example, Fix's modelling (this volume) demonstrates how limited genetic influx during the Neolithic has affected the subsequent biological history of the indigenous groups. Also, Burenhult (forthcoming) elaborates a scenario combining elements of local genetic continuity with demic diffusion and language shift to account for the current distribution of Aslian languages, as well as genetic diversity of their speakers. According to this hypothesis, when the ancestor of the Aslian languages entered the peninsula in connection with the introduction of agriculture, some local foragers adopted the new economy and Aslian language, and they intermixed to a greater or lesser degree with the Neolithic immigrants (cf. Bellwood 1985). The hypothesis further proposes that it was in this diverse setting of intermixing that the Aslian sub branches split from the introduced proto-language, the clades coming into being in distinct situations of cultural and linguistic contact. Some foragers retained their economy and nomadic lifestyle (the predecessors of today's Semang), but at some point the pre-existing ties between them and the more settled, intermixed groups led to eventual language shift such that also the still foraging people spoke Northern Aslian languages.

Burenhult argues that the initial splitting up of three Aslian sub branches should not be automatically connected to the formation of the three societal-economic subgroupings and he places emphasis on the ethnolinguistic groups which do not match the general societal-

1 Bulbeck (2004) is a notable exception, where the significance of 'mismatches' like the Ceq Wong and Jah Hut is given prominence in interpreting peninsular prehistory. Benjamin himself has explicitly stated the need for ethnographic and linguistic research on some of these groups (Benjamin 1976, 1989).

2 Again, Benjamin has remained cautious about projecting the tripartite model back in time and has also called for a clear analytical distinction between linguistic, cultural and biological history (Benjamin 1989), a suggestion reiterated by Burenhult (forthcoming).

linguistic correlations. Thus, the idea that Northern Aslian crystallised in a non-forager setting is supported by the existence of one distinct and conservative Northern Aslian language—Ceq Wong—spoken by a semi-sedentary group with a mixed economy which does not belong to the Semang forager sphere. Another mismatch is a small group of Central Aslian languages—Semnam and Lanoh—spoken by foragers with both Semang and Senoi-like cultural features: most likely the result of a later language shift from Northern Aslian. The rest of the Central Aslian languages, as well as most Southern Aslian languages, are spoken by non-Semang. However, one Southern Aslian language (Semaq Beri, not discussed in Burenhult, forthcoming) is spoken by people with a clear focus on foraging, although they are not included in the Semang forager grouping (see §4.3).



**Figure 2:** Aslian family tree, rooted on Mon (from Dunn et al.). This is a Maximum Clade Consistency tree, summarising the 750 post-burn in trees of the Bayesian phylogenetic tree sample with branch length equal to the median length of all congruent branches found in the sample. Numbers on the branches indicate percentage of the tree sample supporting each bifurcation (for details, see Dunn et al.).

In a recent paper, Dunn et al. revisit Aslian history by analysing newly collected lexical data from 27 Aslian varieties with quantitative methods to produce measures of linguistic divergence as well as phylogenetic hypotheses.<sup>3</sup> While broadly reproducing the earlier

3 The same dataset is used in the present chapter to analyse post-split contacts, see §5.2.

proposed clades of Aslian genealogy, the phylogenetic aspect of this study also reveals that the three major clades show very unequal rates of lexical divergence: Southern Aslian is the most conservative branch; Central Aslian shows a bit more divergence; and most of Northern Aslian is contained within a clade which is highly divergent externally, but which has low internal diversity (suggesting a recent diversification; see Figure 2). It is the Northern Aslian languages spoken by Semang foragers which show great external divergence whereas the geographic and cultural outlier Ceq Wong comes out as an early and conservative split. So while the degree of lexical divergence coincides with societal distinctions, neither degree of divergence nor societal features coincide with the Northern Aslian clade. Specifically, the conservative nature of Ceq Wong suggests that the initial branching of Northern Aslian did not occur in a Semang-type cultural setting, nor did it coincide with its development. Instead, the Semang-Northern Aslian link and its accelerated lexical divergence is likely to be a later phenomenon. This lends support to the idea that the prehistoric environment in which Aslian spread and branching occurred was culturally, linguistically and biologically diverse.

## 4 The ethnolinguistic categories

In this chapter we examine the relationship between language and culture with a particular focus on the role of linguistic contact in analysing the history and development of peninsular societal-economic patterns. Acknowledging that today's broad Northern Aslian connection with foraging does not necessarily have deep historical pertinence, we sidestep analytically the traditional categories 'Northern-Central-Southern Aslian' and 'Semang-Senoi-Aboriginal Malay'. Instead, we examine a broader range of ethnolinguistic groups which cross-cuts these categories. Our dataset represents the Aslian branch as a whole, but our main area of analysis and discussion is groups whose subsistence mode has been described ethnographically or historically as displaying some degree of mobile foraging (see below). These include ethnolinguistic groups speaking languages of the Northern, Central and Southern Aslian branches. They also cross-cut the three conventional societal-economic groupings: Semang, Senoi and Aboriginal Malay. One advantage of this ethnolinguistically based sample is that groups which are difficult to classify according to the conventional tripartite paradigm (because they are in some sense 'mixed') are not anomalies which are best avoided, but can be treated on a par with those whose societal-economic characteristics are more faithful to the overall classification.

The following sections briefly outline the main linguistic and ethnographic features of four ethnolinguistically defined groups which are particularly pertinent to the ensuing analysis: Maniq/Menraq-Batek (§4.1), the Lanoh complex (§4.2), Semaq Beri (§4.3), and Ceq Wong (§4.4). The authors have conducted first-hand linguistic and ethnographic field work among these groups, of which most have not previously received linguistic attention. The descriptions are based on a variety of sources, including the authors' recent and unpublished findings.

### 4.1 Maniq/Menraq-Batek (MMB)

On linguistic grounds, this group ('Northern Aslian-speaking Semang') can be divided into two subgroups (Benjamin 1976; Dunn et al.). One subgroup comprises languages and dialects in Isthmian Thailand, such as Ten'en, Kensiw and Tea-de, as well as Kensiw and Kentaq in northernmost Peninsular Malaysia. Following Burenhult (forthcoming), these

varieties will here be referred to generically as Maniq (the preferred endonym for many of these groups, from their word for human being, *mani?* or *mni?*). The other subgroup is found predominantly in Malaysia (Perak, Kelantan and Pahang states) and comprises three languages: Jahai, Menriq and Batek. While Jahai and Menriq share a term for human being (*mnra?*), the Batek term is the same as the ethnonym (*batek*). This group is referred to here as Menraq-Batek. The Maniq/Menraq-Batek (MMB) are the only groups which combine the Semang societal tradition with Northern Aslian language, and they have a particularly close connection with locally ancient genetic lineages (as shown by samples analysed by Fucharoen et al. 2001 and Hill et al. 2006). The MMB languages show the greatest lexical divergence of all Aslian languages (Dunn et al., see Figure 2).

Traditionally, the MMB live in groups of about 15–50 people. Temporary camps of lean-tos or huts are inhabited for a few days to several weeks or months (sometimes even years), depending on the sustenance circumstances. The economy is based on foraging in a broad sense. Hunting, fishing and gathering form the backbone of their subsistence. But the MMB also make occasional swiddens (especially the Menraq-Batek groups), collect rainforest products for trade with outsiders, and seize any opportunity to engage in wage-labour, if such activities are considered to be economically advantageous at the time. Nowadays many MMB are permanently settled in resettlement villages established by the Malaysian and Thai governments, but some groups in both countries still pursue a mobile existence.

Semang society promotes this mobile lifestyle, its social structures encouraging dispersal and flux in space, time and human relations (Benjamin 1985). The conjugal family is the only persistent social unit, and bands and camps consist of several such families which co-exist on a voluntary basis. Strict cross-sex in-law avoidance rules apply, the filiative bias is patrifocal and residence is virilocal. Marriage is strictly exogamous. Society is egalitarian, and there is a moral obligation to share food with other members of a camp (van der Sluys 2000). For detailed accounts, see for example Schebesta (1952), Endicott (1979) and Lye (1997). Benjamin's 1985 classification suggests the Batek tend toward a Malayic societal pattern, have no dominant subsistence mode, and are indeed 'mixed' rather than Semang. Depending on which features are considered critical to classification, it could also be argued that other groups represent a mixed pattern at the present time, for example the Jahai and the Menriq.

#### 4.2 The Lanoh complex

Lanoh is a generic label for a cluster of dialects spoken historically along portions of the middle and upper Perak river, Peninsular Malaysia. Most of these dialects are extinct (for example Sabüm) or moribund (for example Yir); Semnam, spoken in one village, still has some 300 speakers (Burenhult & Wegener 2009). They belong to the Central Aslian sub branch of Aslian and are closely related to Temiar, one of the main languages of the Senoi societal sphere. Until recently the subsistence system of Lanoh speakers was based on nomadic foraging very similar to that pursued by the MMB (Dallos 2011; Burenhult, field notes 2006). In some early ethnographic accounts they are described as a particularly reclusive and mobile Semang group (see for example Schebesta 1928). Like the MMB, the Lanoh have a band-based society and the conjugal family is the basic social unit of

production. Their cross-sex in-law avoidance rules are similar to those of the MMB.<sup>4</sup> However, their filiative bias, which is cognatic rather than patrifocal, is shared with the Senoi rather than with the Semang. Benjamin therefore classifies them as belonging to the Senoi tradition instead of the Semang (Benjamin 1985:251, forthcoming).

While no genetic analyses have been carried out on Lanoh speakers, unsystematic phenotypic observations in the literature suggest a ‘Negritoid’ appearance similar to MMB speakers and thus possibly a close connection to the locally ancient genetic lineages. Their territory abuts on that of some MMB groups, and historically they were in close contact with the Kensiw, Kentaq and Jahai. They are currently undergoing assimilation by the Temiar. The fact that Lanoh belongs to the Central Aslian sub branch can possibly be attributed to a language shift among some previously Northern Aslian-speaking Semang (Burenhult, forthcoming).

#### 4.3 Semaq Beri

The Semaq Beri speak a language belonging to the Southern Aslian sub branch and traditionally inhabited an area north of the Pahang River in the upper reaches of the Kuantan, Kemaman, Dungun, Tembeling, Terengganu and Lebir Rivers. They have been characterised as ‘mixed’, an ‘ill-defined and heterogeneous group’ (Benjamin 1985). Southern Semaq Beri reportedly display the attributes of Benjamin’s Malayic grouping (see §3), while northern Semaq Beri are nomadic foragers whose significantly variable phenotypic features have led observers to suggest a mixed genetic lineage (Endicott 1975:4–5). The Semaq Beri-speaking foragers are traditionally not included in the Semang ethnographic grouping. However, it is clear that the northern Semaq Beri live like their MMB-speaking neighbours (the Batek) in small camps of lean-tos combining hunting, fishing and foraging with the occasional collection of forest produce for trade, or waged labour (Kuchikura 1987; Morris 1996). Societal features also have much in common with those of the Semang. Thus, the Semaq Beri society is egalitarian, band-based, and there is a strong moral obligation to share food with other members of one’s group. The conjugal family is the primary social unit. There is no fixed pattern of post-marriage residence, although there is some preference for virilocal residence in established marriages (Kuchikura 1987:23). Strict cross-sex in-law avoidance is observed. In addition, there is strict cross-sex avoidance between parents and children and between cross-sex siblings after the onset of puberty (Kruspe, field notes 2009).

---

4 In Benjamin’s typology of cross-sex relations (1985:252) the Lanoh pattern with Senoi groups, ‘sister-in-law’ relations being characterised by restraint with one’s spouse’s older siblings and by joking with one’s spouse’s younger siblings. However, recent work shows that the Semnam subgroup of Lanoh has ‘sister-in-law’ relations which are identical to those given by Benjamin for the MMB groups Kensiw and Kentaq. Here, avoidance is observed in relations with one’s spouse’s older siblings while relations with one’s spouse’s younger siblings are neutral (Burenhult, field notes 2008). Benjamin’s typology also characterises Lanoh brother/sister relations as neutral, in line with a common Senoi pattern and in contrast with a unified Semang pattern of avoidance. This is the case also in Burenhult’s recent Semnam data. However, with regard to brother/sister relations among the Semang Burenhult’s data diverge from Benjamin’s: the Jahai observe restraint rather than avoidance, which suggests that brother/sister avoidance is a less common Semang pattern than Benjamin’s classification indicates, at least at present.

Exogamous marriage forms an integral component of the foraging mode, and first cousin marriage is typically avoided.<sup>5</sup> All the Semaq Beri maintain close social relations with other bands across their extensive range, as well as close relations with their immediate Orang Asli neighbours. Presently, the northern Semaq Beri have close relations with the Batek Deq (Kuchikura 1987:9, 16–17; Kruspe, in prep), resulting in significant gene flow and cultural and linguistic influences, for example, inter-Aslian bilingualism which is absent in the southern group. The northern variety of Semaq Beri exhibits lexical and grammatical influence from Northern Aslian, which is less prevalent in southern varieties of Semaq Beri (§5 below, Kruspe, in prep). In the south the Semaq Beri have established social relations with their Aboriginal Malay neighbours the Jakun (Ramle 1993:43), giving rise to different genetic inflow, and cultural-linguistic influences. There are no genetic studies of the Semaq Beri.

There has been speculation as to the origins of foraging by the northern Semaq Beri. Benjamin (1976, 1985) treats their nomadism as a deviation from the defined cultural-linguistic division (collector-trader Southern Aslian) proposing that the Semaq Beri are ‘secondarily nomadic’ having shifted to foraging after splitting off from Semelai (another Southern Aslian language; Kruspe 2004). Endicott (1975) similarly provides a recent-convert scenario. However, ethnohistorical accounts from southern Semaq Beri confirm long-term engagement with nomadic foraging activities, only becoming sedentary in response to external pressure (Kruspe, in prep.; Evans 1915), suggesting that nomadic foraging may have been the dominant mode for all Semaq Beri in the past.

#### 4.4 Ceq Wong

The Ceq Wong are unique within Northern Aslian, geographically isolated from the remainder of the group and not part of the Semang societal sphere. They may represent a relic Northern Aslian population: phylogenetic analysis of the Ceq Wong lexicon clearly identifies the language as a conservative/relic variety (Dunn et al.; Burenhult, forthcoming; see also §5.2. below). All the other attested Northern Aslian languages are considerably divergent from it. The Ceq Wong practise a ‘mixed’ economic adaptation combining both foraging and swiddening, however their residence pattern is semi-sedentary like that of swidden horticulturalists. Furthermore, unlike the MMB, some of whom also engage in occasional swiddening but whose main activity is foraging, the Ceq Wong place emphasis on both activities. Despite this, they perceive themselves as subsistence foragers or ‘digging people’ (Howell 1989:13), in reference to their dependence on wild yams and ability to survive in the forest, unlike their Central Aslian neighbours the Jah Hut. Their simple swiddens in which manioc is grown appear to provide a supplement to foraging, and are often abandoned in favour of other economic activities. Ethnohistorical accounts from the Ceq Wong suggest that until recently they were more mobile (Howell 1989:21–22).

Ceq Wong society is egalitarian and there is a strong moral obligation to share food with other members of the group. The primary social unit is the conjugal family, which functions as an autonomous entity. Residence was traditionally in a camp or swidden in the forest with one or more families. These days some people choose to live in a government

---

<sup>5</sup> Kruspe’s ethnographic data differ from those of Benjamin (1985), according to whom cousin marriage is permitted among the Semaq Beri and cross-sex relations are characterised by restraint rather than avoidance.

settlement, but residence remains fluid. There are frequent movements between settlements, and between settlements and the forest, which may last from a few nights to several months or longer. Temporary lean-tos provide shelter in the forest, while more permanent houses are constructed in the swiddens.

In practice, most marriages are endogamous and first cousin marriage is currently the most common union, but curiously the Ceq Wong maintain the cultural ideal that marriage should be exogamous (with someone unrelated) otherwise it is incestuous (Howell 1989:28; Kruspe, field notes 2002). Both polygyny and polyandry are practised. There is uxorilocal residence immediately following marriage, after which residence becomes bilocal. The Ceq Wong observe restraint in cross-sex parent-in-law relations.

There are no genetic studies of the Ceq Wong. Unsystematic observation has seen the Ceq Wong classified phenotypically as Senoi, like the Central Aslian Jah Hut. There is evidence of intrusive genetic flow from sporadic intermarriage with the Jah Hut. A colonial report notes some Ceq Wong showing ‘much Negrito blood’ which was attributed to intermarriage with Batek Nong, a Semang group to the north (Ogilvie 1948:15, 29). The present day Ceq Wong are not aware of the existence of the Batek Nong.

## 5 Contact: lexical and other evidence

### 5.1 Sociolinguistics

Evidence suggests that some foraging societies of the Malay Peninsula share linguistic characteristics which cross-cut Aslian genealogical boundaries. For example, Benjamin (1976:74–76, 1980:4, 1985:234–235, 2001:111) has long argued that the Semang display distinct sociolinguistic features (see also Endicott 1997). The mobile lifestyle of the Semang, manifested in their system of intermarriage between individuals of widely dispersed bands, as well as in their pattern of group disintegration and regrouping into new constellations in response to changing subsistence conditions, is linked to particular patterns of individual language use. A speaker may move through several linguistic environments throughout his or her lifetime, which leads to a high rate of idiolectal change. At the same time, the diverse linguistic origins of members of a band also lead to marked variation in the language use of different speakers. Benjamin (2001:111) discusses this in terms of a mesh-like relation between language varieties which is idiolectal as much as dialectal. All of the MMB languages form a continuum of such linguistic interaction and, in the sociolinguistic sense, represent a unitary linguistic constellation. The participation of Lanoh, Semaq Beri and Ceq Wong in this constellation has so far been less clear.

### 5.2 Lexicon

In a previous study of linguistic divergence in Aslian, Neighbor-Net clustering of the lexicostatistical distance data, which has the advantage of being able to show ‘conflicting signal’, revealed a notable split between MMB-Lanoh on the one hand, and the rest of Aslian on the other (Dunn et al.).<sup>6</sup> This split partially cross-cuts genealogical boundaries,

6 The lexical data used in Dunn et al., also used in the new analysis of post-split contacts presented below, comprise lists of basic vocabulary from twenty-seven Aslian varieties. The bulk of these lists were collected recently (1990–2008) in the field by Burenhult and Kruspe, with additional wordlists provided by Neele Becker and Sylvia Tufvesson. For details, see Dunn et al.

showing evidence clustering the Lanoh complex with MMB (that is, Northern Aslian languages apart from Ceq Wong). This, it seems, is a lexically traceable variant of Benjamin's sociolinguistic category, and one that includes Lanoh. Although displaying some obvious MMB loans, Southern Aslian Semaq Beri does not participate in this clustering (see Figure 3).

The Neighbor-Net network is being used to represent the lexicostatistical distance, that is, proportion of shared and unshared cognates between each pair of languages in the list (where a distance of 0.0 indicates identity and 1.0 indicates that the lists are completely different; note that in lexicostatistics it is common to report proximity, the inverse of distance). In this chapter we take the analysis a step further by comparing this measure to a more realistic measure<sup>7</sup> of evolutionary distance, calculated from a phylogenetic tree. The Aslian family tree proposed in Dunn et al. is our best estimate of the phylogenetic relationships within the family given the data we have available. This language sample is large—containing data on more Aslian languages than have ever been presented together before<sup>8</sup>—but varies considerably in the amount of data which is feasibly obtained for each language. The common baseline is a form of the 200-word Swadesh list (used previously by Benjamin 1976), which was collected for all languages of the sample, as well as some outgroup languages (other Austroasiatic languages, not part of the Aslian family). These lists were then coded for probable cognacy on the basis of explicit criteria for identifying similarities in form-meaning mapping.<sup>9</sup>

The phylogenetic relations between the Aslian languages were estimated using computational methods from a family of techniques known as Bayesian phylogenetic inference. These methods allow rich inferences from lexical cognate data, by modelling the evolution of a language family as the gain and loss of reflexes of cognate sets, and are applicable without requiring for example a detailed understanding of regular sound change within the family. The result of this phylogenetic inference includes not only a tree topology (as would be produced by the linguistic comparative method, whenever an exhaustive reconstruction of Aslian should become available), but also includes a measure of statistical confidence for each branch in the tree, and (crucially for the argument below) a measure of the amount of evolutionary change on each branch. The details of our proposed family tree for the Aslian languages are presented in Dunn et al., and a summary tree of our best estimate of the phylogeny (rooted on Mon, a member of a sister clade to the Aslian family) is presented in Figure 2. From the tree sample underlying Figure 2, we calculated the median evolutionary distance between each pair of languages by summing the branch lengths between them in each tree in the sample.

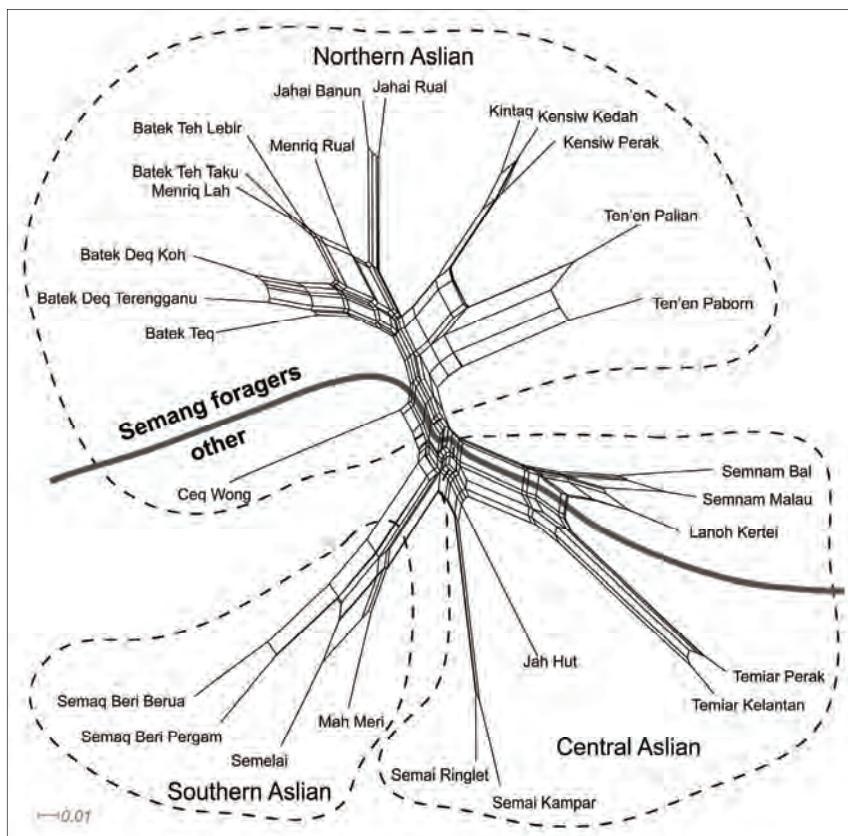
We are contrasting the model of evolution used in this attempt at phylogenetic inference—the Bayesian approach using explicit evolutionary models—with the lexicostatistical model. The Bayesian tree is a better model of the history of the languages under consideration. It considers the evolution of each individual cognate set, allows for different rates of changes, and implicitly can even handle low levels of family-internal borrowing (through allowing low rates of ‘spontaneous recreation’ of reflexes of cognate

7 It is a character-based method, which models the history of the individual reflexes of cognate sets, rather than a distance measure, which just models aggregate (dis)similarity.

8 The pioneering lexicostatistical study of Aslian by Benjamin (1976) involved a similar sample. Both samples are broadly representative of Aslian but differ in detail. For a full account of the differences, see Dunn et al.

9 We refer to Dunn et al. for the criteria used in identifying probable cognates.

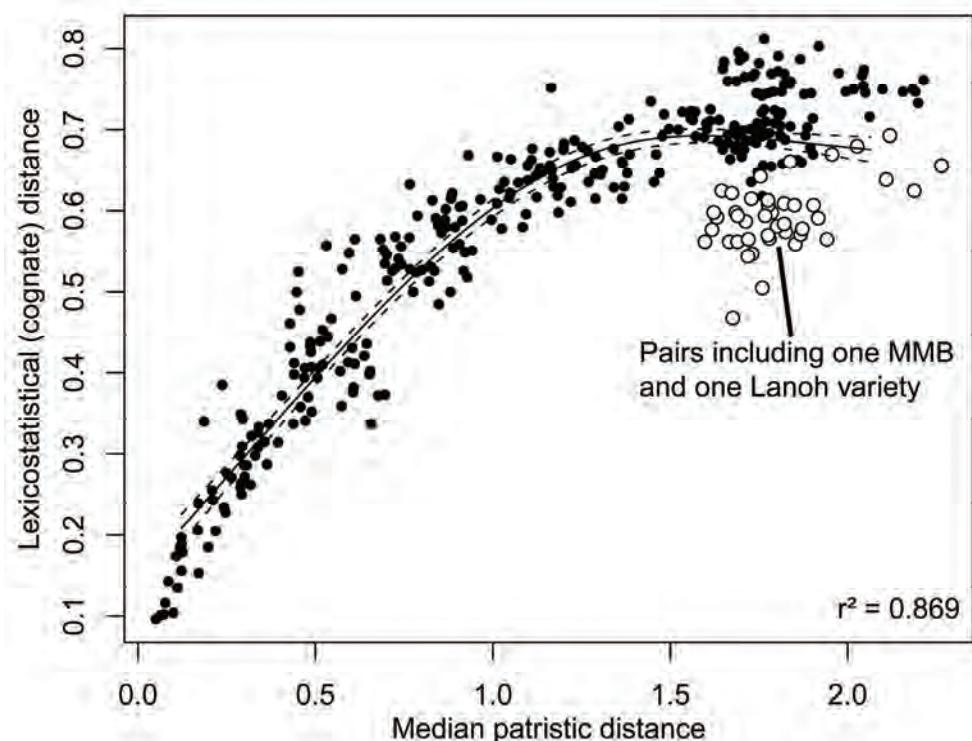
sets in branches where the set is extinct). Lexicostatistics measures pairwise similarity of languages in terms of proportion of shared cognates (or possible cognates) in a word list. As a model of historical relatedness, it assumes that similarity between these word lists is directly proportional to how closely the two languages are related. This is a very simple model of language change, and one that we know to be misleading in many circumstances. Lexicostatistical distances are a poor model of the historical processes involved in language change, and a set of lexicostatistical distances are only historically meaningful if one is willing to allow that lexical similarity has decayed at a constant rate, such that the proportion of non-cognate (or not recognisably cognate) words in a pair of languages should be proportional to the amount of time those languages have been historically distinct. The decay in lexical similarity can be due to lexical replacement, and (at deeper levels) also to processes of regular sound change, where the sum of sound changes has obscured the cognacy relationships. In particular, there is no space for family-internal borrowing in this model: borrowings of cognate terms cannot be distinguished from true, inherited cognate vocabulary.



**Figure 3:** Neighbor-Net clustering of lexicostatistical distances calculated from Aslian vocabulary lists (from Dunn et al.). The Neighbor-Net graph represents a matrix of distances without forcing the resolution of the major conflicts in the data.

In Figure 4 we compare lexicostatistical distance measures between each pair of languages to the sum of the branch lengths between these languages in a phylogenetic tree. If the lexicostatistical model of constantly decreasing similarity between languages accurately reflected the history of these languages, we would expect lexicostatistical and evolutionary patristic distance to be closely correlated.

Unsurprisingly, the correlation is strong ( $r = 0.87$ , meaning that 87% of the variance in one distance measure can be predicted from knowing the value of the other), since these are two different models of the same data. Thus, in Figure 4, the data points (representing language pairs) mostly fall along a smooth curve representing the phylogenetically expected similarities. But there are also clusters of points which do not fit the curve very well. The white dots mark pairs of languages which include one MMB language and one language from the Lanoh complex. These languages tend to have more terms identified as cognates than their phylogenetic relationships would predict, a clear signal of intra-Aslian borrowing and a pattern restricted to a MMB-Lanoh sphere. Importantly, these pairwise similarities are consistent throughout MMB and the Lanoh complex, and are not restricted to, say, members of each group which are currently contiguous or cohabitant (although such pairs indeed represent the most extreme examples of phylogenetically unexpected similarity). This suggests either (1) a distinct pattern of long-term lexical diffusion/exchange within MMB-Lanoh or (2) an ancient contact situation between proto-varieties of Lanoh and MMB leading to lexical congruence subsequently inherited throughout both groups (including a scenario of language shift from MMB to Proto Lanoh). The former explanation is probably the more likely one, since it largely links up to Benjamin's characterisation of sociolinguistic interaction. Also, the ancient contact scenario faces a chronological obstacle, since Proto-MMB probably is considerably older than Proto Lanoh. Any traces of contact-induced transfer of lexicon in Lanoh would therefore presumably not be shared with MMB as a whole, but with some subsection of it.

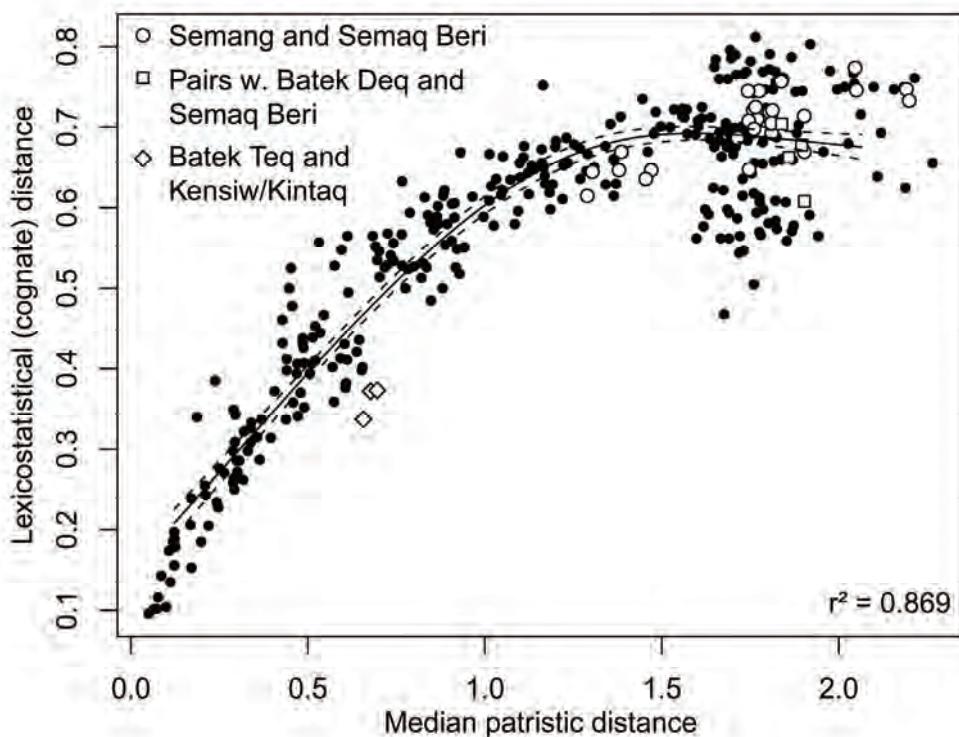


**Figure 4:** Phenetic x patristic distance (with pairs including one MMB variety and one Lanoh variety highlighted).

In Figure 5 we show the same data, but have marked pairs of languages including one MMB or Lanoh variety and one Semaq Beri (forager Southern Aslian) variety. For the

most part, these pairs occur along the expected curve, showing no indication of borrowing. However, a pair including Semaq Beri Berua and Batek Deq Terengganu shows a high degree of contact, whereas the pair including Semaq Beri Berua and another variety of Batek Deq does not. Pairs of one Batek Deq variety and one Semaq Beri variety are shown with squares instead of circles. This suggests very shallow contact between MMB and Semaq Beri, traceable only in one pair of varieties which are currently spoken in the same village.

Figure 5 also highlights a cluster of points which represents another set of pairs whose members are more similar than expected phylogenetically. These all include Batek Teq and some variety of Kensiw/Kintaq. This suggests the existence of a presumably old contact situation between these currently widely separated varieties, subsequently broken up geographically by the Menraq branch of MMB. Benjamin (1976:77) notes a similar geographic intrusion, but in the form of Jahai breaking up ancient contact between Kensiw and Menriq.



**Figure 5:** Phenetic x patristic distance (with pairs including one Semang variety and one Semaq Beri variety, as well as pairs with one Kensiw/Kentaq variety and Batek Teq, highlighted).

Another thing to note is that Ceq Wong shows no major identifiable pattern of post-split contact with any other language, instead behaving as expected from the phylogenetic analysis. This lends further support to the idea that Ceq Wong represents a conservative Northern Aslian relic with few traces of secondary contact. There is no evidence, for example, that the language has made an exit from the Semang sphere and subsequently undergone distinct patterns of lexical change due to a new contact situation. Claims of borrowing between Ceq Wong and Kensiw (Benjamin 1976:78) and between Ceq Wong

and Semaq Beri (Endicott 1975:7, citing Diffloth, p.c.) are not supported by the current dataset.

### 5.3 Grammar

There is less evidence for contact in grammar, perhaps stemming partly from the fact that detailed grammatical data are still lacking for many of the languages. However, at least one morphemic category present in both MMB and Lanoh is yet to be discovered elsewhere in Aslian. This is a causative infix surfacing in some MMB languages as *<ri>* and in the Semnam variety of Lanoh as *<yi>* (the *y* in the latter is an expected reflex, given the lack of a phoneme /r/ in Semnam). Another candidate is plural inflection in human nouns by means of an infix *<ra>* in some MMB languages and a corresponding *<ya>* in Semnam. A similar infix is present in Semaq Beri, which may have borrowed it from a MMB language, presumably Batek (Kruspe, in prep). It is also present in some Austronesian languages, though not currently productively in Malay. There is emerging evidence that some cognates of forms which belong to the class of expressives in other Aslian languages behave formally as stative verbs in MMB and Lanoh. However, this is a feature which may be shared by Ceq Wong (Kruspe, in progress). In fact, preliminary comparison based on Jahai and Ceq Wong suggests that the genealogical subgrouping Northern Aslian is a better predictor of grammatical similarities than sociolinguistic and societal-economic categories (Burenhult, forthcoming).

## 6 Conclusions

The indigenous communities of the Malay Peninsula represent a microcosm of human dynamics and complexity, and they provide a fascinating analytical setting for disentangling historical relationships between language, culture and genes. This paper has been especially concerned with the historical relationship between linguistic phenomena and a particular type of subsistence mode, namely nomadic foraging. The following conclusions can be drawn:

Firstly, there is evidence against a correlation between linguistic phylogeny (Northern Aslian) and societal type (Semang), so the current dominant societal mode of Northern Aslian speakers should not be taken as a reliable indicator of what the Proto Northern Aslian-speaking society was like. Our analysis of post-split contacts confirms that the Northern Aslian, non-Semang outlier Ceq Wong is a conservative relic which has not experienced significant secondary lexical exchange, and there is no indication that Ceq Wong made an exodus from the Semang cultural sphere. This suggests that the Northern Aslian clade crystallised in a non-Semang and possibly Ceq Wong-like setting, only secondarily spreading to the ancestral Semang (in connection with the branching off of MMB; see Burenhult, forthcoming). Yet, the forager-like aspects of Ceq Wong society point to complexities in the cultural settings in which Aslian was established and spread, and they evince the need to view some of the ‘mixed’ cultures as perhaps more archetypal and conservative than has typically been the case (cf. Bulbeck 2004).

Secondly, the lexical patterns revealed by the present study highlight a category which has typically gone unrecognised and unlabeled in recent literature. The MMB languages and those of the Lanoh complex show evidence of considerable secondary exchange of vocabulary, more so than any other Aslian setting. This pattern of linguistic exchange cross-cuts genealogical boundaries and Benjamin’s Semang-Senoi societal distinction, but

it is consistent with some other characterisations of the Lanoh as Semang-like. Incidentally, it also coincides with observed phenotypical characteristics. In all likelihood, the pattern reflects a distinct configuration of interaction, in line with Benjamin's notion of a Semang sociolinguistic entity but with the notable addition of Lanoh. Along with our new ethnographic data on kinship, these results call for a re-evaluation of the position of the Lanoh in relation to their neighbours.

Southern Aslian Semaq Beri, although partly spoken by foragers with very Semang-like societal features and currently co-existing with MMB neighbours, shows no deep traces of similar lexical contact. This, incidentally, suggests that Semaq Beri contact with the MMB is comparatively recent and that their foraging mode of subsistence possibly developed in response to this contact, potentially providing support to Endicott's (1975) and Benjamin's (1985) scenarios of secondary adoption of nomadic foraging. However, the presence of the foraging subsistence mode in all three branches of Aslian again brings into question the customary alignment of language and subsistence type. Semaq Beri foraging may be more conservative than previously assumed and recent linguistic contact with MMB may not necessarily be congruent with a switch in subsistence mode. The Semaq Beri may have traditionally foraged further to the south, possibly in the area where some southern Semaq Beri still reside. Recall the southern Semaq Beri ethnohistorical account of forager subsistence discussed in §4.3. The suggested recent Semaq Beri and Batek contact is possibly the result of the displacement of indigenous peoples in the historical period precipitated by the influx of Malay immigrants up the Pahang River and its tributaries and along other east coast rivers. Note again in this context the difficulty in defining the Lanoh in relation to the Semang and Senoi categories in purely societal terms—the Lanoh share societal features with both.

In addition to the lexically motivated category proposed here, genetics may offer similarly clear reflections of such a category, as indicated by Hill et al. (2006). However, in the absence of genetic data from the Ceq Wong, Lanoh and Semaq Beri, this cannot be substantiated. Genetic studies in these settings will be crucial to furthering our understanding of Aslian prehistory.

Thirdly, our analysis has revealed traces of ancient contact between geographically distant languages. The more-than-expected lexical similarity between Kensiw/Kentaq and Batek Teq suggests ancient geographical contiguity between Maniq and Batek varieties, possibly pointing to a past distribution of Aslian languages in the northeast of the peninsula (north Kelantan, southern Thailand) which was much wider than can be seen in the historical record. This contact seems to have come to an end with the expansion of Menraq languages northwards, and of Malay southward along the Kelantan river.

Finally, we hope to have shown that our conceptual approach—which sets out from the actual well-defined ethnolinguistic groups rather than generalised categories—provides a robust framework for analysis, and does justice to groups which have traditionally played a minor role in the exploration of Malayan indigenous history.

## Acknowledgements

We are grateful to David Bulbeck, Nick Enfield, Stephen Oppenheimer, Joyce White, and two anonymous reviewers for their helpful comments on earlier versions of the paper, as well as participants in the 19<sup>th</sup> International Conference on Historical Linguistics (Nijmegen, 2009) and the Dynamics of Human Diversity in Mainland Southeast Asia workshop (Siem Reap, 2009) for valuable comments on part of this work. The long-term

field research carried out by Burenhult was funded by the Max Planck Society, a European Community Marie Curie Individual Fellowship, and the Volkswagen Foundation's DOBES programme; and that of Kruspe by the Max Planck Society, the Research Centre for Linguistic Typology, the Hans Rausing Endangered Languages Programme, and the Volkswagen Foundation's DOBES programme. Dunn's research was sponsored by the Max Planck Society. We are also grateful to the Economic Planning Unit, the Jabatan Hal Ehwal Orang Asli, and Universiti Kebangsaan Malaysia for their helpful assistance, and we are especially indebted to the numerous Aslian-speaking communities who have generously shared their languages with us. Thanks also to Stephen Levinson, director of the Language and Cognition group at the Max Planck Institute for Psycholinguistics, Nijmegen, for providing the stimulating interdisciplinary research environment that enabled the authors to carry out this collaboration.

## References

Bellwood, Peter. 1985. *Prehistory of the Indo-Malaysian archipelago*. Sydney: Academic Press.

—. 1993. Cultural and biological differentiation in Peninsular Malaysia: the last 10,000 years. *Asian Perspectives* 32. 37–60.

Benjamin, Geoffrey. 1976. Austroasiatic subgroupings and prehistory in the Malay Peninsula. In Philip N. Jenner, Laurence C. Thompson & Stanley Starosta (eds) *Austroasiatic studies*, Part I, 37–128. Honolulu: The University Press of Hawaii.

—. 1980. *Semang, Senoi, Malay: Culture-history, kinship, and consciousness in the Malay Peninsula*. Unpublished manuscript.

—. 1985. In the long term: three themes in Malayan cultural ecology. In Karl L. Hutterer, Terry Rambo & George Lovelace (eds) *Cultural values and human ecology in South East Asia*, 219–278. University of Michigan: Center for South and Southeast Asian Studies.

—. 1989. Achievements and gaps in Orang Asli research. *Akademika* 35. 7–46.

—. 2001. Orang Asli languages: From heritage to death? In Razha Rashid & Wazir Jahan Karim (eds) *Minority cultures of peninsular Malaysia: Survivals of indigenous heritage*, 101–122. Penang: Academy of Social Sciences.

—. forthcoming. The current situation of the Aslian languages. In Hein Steinhauer & James T. Collins (eds) *Endangered languages of Southeast Asia*. London: School of Oriental and African Studies.

Bulbeck, David F. 2003. Hunter-gatherer occupation of the Malay Peninsula from the Ice Age to the Iron Age. In Julio Mercader (ed.) *The archaeology of tropical rain forests*, 119–160. New Brunswick: Rutgers University Press.

—. 2004. An integrated perspective on Orang Asli ethnogenesis. In Victor Paz (ed.) *Southeast Asian archaeology: Wilhelm G. Solheim II Festschrift*, 366–399. Quezon City: The University of the Philippines Press.

—. this volume, Biological and cultural evolution in the population and culture history of *Homo sapiens* in Malaya.

Burenhult, Niclas. (forthcoming). Foraging and the history of languages in the Malay Peninsula. In Tom Güldemann, Patrick McConville & Richard Rhodes (eds) *Historical linguistics and hunter-gatherer populations in global perspective*.

Burenhult, Niclas & Claudia Wegener. 2009. Preliminary notes on the phonology, orthography and vocabulary of Semnam (Austroasiatic, Malay Peninsula). *Journal of the Southeast Asian Linguistics Society* 1. 283–312.

Dallos, Csilla. 2011. *From equality to inequality: social change among newly sedentary Lanoh hunter-gatherer traders of Peninsular Malaysia*. Toronto: University of Toronto Press.

Diffloth, Gérard. 1975. Les langues mon-khmer de Malaisie: classification historique et innovations. *Asie du Sud-Est et Monde Insulindé* 6(4). 1–19.

Dunn, Michael, Niclas Burenhult, Nicole Kruspe, Neele Becker & Sylvia Tufvesson. Aslian linguistic prehistory: a case study in computational phylogenetics. *Diachronica*.

Endicott, Kirk M. 1975. A brief report on the Semaq Beri of Pahang. *Federated Museums Journal (new series)* 20. 1–23.

—. 1979. *Batek Negrito religion: the world-view and rituals of a hunting and gathering people of Peninsular Malaysia*. Oxford: Clarendon.

—. 1997. Batek history, inter-ethnic relations and subgroup dynamics. In R. Winzeler (ed.) *Indigenous peoples and the state: politics, land, and ethnicity in the Malayan Peninsula and Borneo*, 46. 30–50. New Haven: Yale Southeast Asia Studies Monograph.

Evans, Ivor H. N. 1915. Notes on Some Aboriginal Tribes of Pahang. *Journal of the Federated Malay States Museum* 5(4). 192–219.

Fix, Alan G. 1995. Malayan paleosociology: implications for patterns of genetic variation among the Orang Asli. *American Anthropologist* 97(2). 313–323.

—. 2002. Foragers, farmers, and traders in the Malayan Peninsula: origins of cultural and biological diversity. In Kathleen D. Morrison & Laura L. Junker (eds) *Forager-traders in South and Southeast Asia*, 185–202. Cambridge: Cambridge University Press.

—. this volume, Origin of genetic diversity among Malaysian Orang Asli: An alternative to the demic diffusion model.

Fucharoen, Goonnapa, Supan Fucharoen & Satoshi Horai. 2001. Mitochondrial DNA polymorphisms in Thailand. *Journal of Human Genetics* 46(3). 115–125.

Hill, Catherine, Pedro Soares, Maru Mormina, Vincent Macaulay, William Meehan, James Blackburn, Douglas Clarke, Joseph Maripa Raja, Patimah Ismail, David Bulbeck, Stephen Oppenheimer & Martin Richards. 2006. Phylogeography and ethnogenesis of Aboriginal Southeast Asians. *Molecular Biological Evolution* 23(12). 2480–2491.

Howell, Signe. 1989. *Society and Cosmos: Chewong of Peninsular Malaysia*. Chicago and London: University of Chicago Press. Original publ. 1984, Oxford: Oxford University Press.

Kruspe, Nicole. 2004. *A grammar of Semelai*. Cambridge: Cambridge University Press

—. (in prep.). *A preliminary report on Semaq Beri*.

—. (in progress). *A grammar of Ceq Wong*.

Kuchikura, Yukio. 1987. Subsistence Ecology among Semaq Beri Hunter-gatherers of Peninsula Malaysia. Hokkaido. *Behavioral Science Report Series E*, No. 1. Sapporo: Hokkaido University.

Lye, Tuck-Po. 1997. *Knowledge, forest, and hunter-gatherer movement: the Batek of Pahang, Malaysia*. Unpublished doctoral dissertation. University of Hawai'i: Department of Anthropology.

Morris, Kimball. 1996. Forest Products for Market: Rattan and the Semaq Beri of Ulu Tembeling. *Akademika* 48. 93-112.

Ogilvie, Charles S. 1948. More of the Che Wong. *Malayan Nature Journal* 3(1). 15-29.

Oppenheimer, Stephen, this volume, MtDNA Variation and Southward Holocene Human Dispersals within Mainland Southeast Asia.

Rambo, A. Terry. 1988. Why are the Semang? Ecology and ethnogenesis of aboriginal groups in Peninsular Malaysia. In A. Terry Rambo, Kathleen Gillogly & Karl L. Hutterer (eds) *Ethnic diversity and the control of natural resources in Southeast Asia*, 19-35. Ann Arbor: University of Michigan Press.

Ramle, Abdullah. 1993. *Semaq Beri, Komuniti Orang Asli di Terengganu*. Kuala Terengganu: Hakcipta Kolej Agama Sultan Zainal Abidin.

Schebesta, Paul. 1928. *Among the forest dwarfs of Malaya*. Kuala Lumpur: Oxford University Press.

—. 1952. *Die Negrito Asiens: Geschichte, Geographie, Umwelt, Demographie und Antropologie der Negrito*. Vienna: St. Gabriel Verlag.

Sluys, Cornelia M.I. van der. 2000. Gifts from the immortal ancestors: cosmology and ideology of Jahai sharing. In Peter P. Schweitzer, Megan Bieseck & Robert K. Hitchcock (eds) *Hunters and gatherers in the modern world: conflict, resistance and self-determination*, 427-454. New York: Berghahn.