

Submitted as a Research Article to *Molecular Biology and Evolution*

Phylogeography and Ethnogenesis of Aboriginal Southeast Asians

Catherine Hill^{1,2}, Pedro Soares¹, Maru Mormina¹, Vincent Macaulay³, William Meehan², James Blackburn², Douglas Clarke², Joseph Maripa Raja⁴, Patimah Ismail⁴, David Bulbeck⁵, Stephen Oppenheimer⁶, Martin Richards¹

¹ Institute of Integrative and Comparative Biology, Faculty of Biological Sciences, University of Leeds, Leeds, UK; ² Department of Chemical and Biological Sciences, University of Huddersfield, Huddersfield, UK; ³ Department of Statistics, University of Glasgow, Glasgow, UK; ⁴ Department of Biomedical Science, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, Malaysia; ⁵ School of Archaeology and Anthropology, Australian National University, Canberra, ACT 0200, Australia; ⁶ Department of Anthropology, University of Oxford, Oxford, UK.

Running head: Phylogeography and ethnogenesis in Southeast Asia

Keywords: Mitochondrial DNA, Phylogeography, Southeast Asia, Orang Asli, Paleolithic, Neolithic

Corresponding author: Martin Richards

Institute of Integrative and Comparative Biology

Faculty of Biological Sciences

L.C. Miall Building

University of Leeds

Leeds, LS2 9JT, UK

Phone: +44 0113-343 2984

Fax: +44 0113-343 2835

E-mail: m.b.richards@leeds.ac.uk

Abstract

Studying the genetic history of the Orang Asli of Peninsular Malaysia can provide crucial clues to the peopling of Southeast Asia as a whole. We have analyzed mitochondrial DNA control-region and coding-region markers in 447 mitochondrial DNAs (mtDNAs) from the region, including 260 Orang Asli, representative of each of the traditional groupings, the Semang, the Senoi and the Aboriginal Malays, allowing us to test hypotheses about their origins. All of the Orang Asli groups have undergone high levels of genetic drift, but phylogeographic traces nevertheless remain of the ancestry of their maternal lineages. The Semang have a deep ancestry within the Malay Peninsula, dating to the initial settlement from Africa >50,000 years ago. The Senoi appear to be a composite group, with approximately half of the maternal lineages tracing back to the ancestors of the Semang, and about half to Indochina. This is in agreement with the suggestion that they represent the descendants of early Austroasiatic speaking agriculturalists, who brought both their language and their technology to the southern part of the peninsula ~4000 years ago, and coalesced with the indigenous population. The Aboriginal Malays are more diverse, and although they show some connections with island Southeast Asia, as expected, they also harbor haplogroups that are either novel or rare elsewhere. Contrary to expectations, complete mtDNA genome sequences from one of these, R9b, suggest an ancestry in Indochina around the time of the Last Glacial Maximum, followed by an early-Holocene dispersal through the Malay Peninsula into island Southeast Asia.

Introduction

It has long been recognized that the population history of the indigenous people of Peninsular Malaysia should provide crucial insights into the prehistory of Southeast Asia as a whole. The Orang Asli (literally “original people”) encompass an astounding range of phenotypic diversity, even though they make up only 0.5% of the local population (Rashid 1995). Theories of their origins developed during Malaysia’s colonial period portrayed them as relics of a sequence of colonization events, prior to the establishment of the mainstream (*Melayu*) Malays as the dominant ethnic group. Since Malaysia’s independence, scholars have however incorporated notions of *in situ* divergence (Bulbeck and Lauer 2006).

Benjamin (1985; 2002a; 2002b) has outlined a framework of three intergrading Orang Asli traditions (Figure 1). Two language groups are involved: Aslian, a well-defined branch of the Austroasiatic family, which includes most of the Orang Asli languages; and dialects of Malay, which is an Austronesian language. The Semang tradition is associated with Northern Aslian languages, rainforest foraging in small bands, egalitarianism, patrilineal descent, and people of “Negrito” appearance. The Senoi tradition, as best represented by the Semai and Temiar, is associated with Central Aslian languages, slash-and-burn farming at higher altitudes, descent groups residing in long houses, egalitarianism, cognatic descent, and a variably “Negrito” to “Mongoloid” appearance. The Aboriginal Malay tradition involves Malay dialects (apart from southern Aslian amongst the Semelai), social ranking, expertise in collecting and trading rainforest produce, a stubborn resistance against Islam and other markers of a *Melayu* identity, and an association with a “Mongoloid” appearance.

The traditional “layer-cake” theory assumed successive waves of Semang, Senoi and Aboriginal Malays (Cole 1945; Carey 1976; see also Birdsell 1993). On the basis of superficial anatomy and their foraging lifestyle, the Semang were grouped with the Philippine Aeta and Andaman Islanders, as well as Melanesians, Tasmanians and certain tropical Australian rainforest foragers into a “Negrito race”. These were thought to have originated in Africa and spread through Southeast Asia before colonizing the southwest Pacific. The second wave of immigrants, the Senoi, was believed to have its origins in South Asia, amongst the Veddas and other small-bodied, forest foragers in South Asia, along with the Toaleans of South Sulawesi, and most mainland Australian Aborigines. The arrival of the Aboriginal Malays supposedly represented the first influx of “Mongoloids” into Peninsular Malaysia, as part of the colonization of the Indo-Malaysian Archipelago by the light-skinned, straight-haired “Proto-Malays”. The subsequent evolution and expansion of the “Deutero-Malays” was registered in Peninsular Malaysia with its colonization by the *Melayu* Malays (Harrower 1933; Carey 1976).

Bellwood (1993; 1997) simplified the number of migrations to two, whilst suggesting an explanation for how such migrations may have occurred and their relation to language distribution. He stressed the advantages of an agricultural economy in supporting large populations which could absorb forager groups. Southeast Asia’s “Negritos”, including the Semang, would represent the relict descendants of Southeast Asia’s original “Australo-Melanesian” foragers. Both Austroasiatic and Austronesian languages had their origins in South China, and were introduced to Southeast Asia during the middle Holocene with the Neolithic expansion of farmers of “Mongoloid” physical type. Austroasiatic took a mainland route southwards, while Austronesian expanded along the island arc from Taiwan to

the Philippines, and then Indonesia and Malaysia. Interaction between immigrant farmers and resident foragers resulted in the mixed phenotype of certain groups, notably the Senoi, as well as language shift by the Semang to Aslian.

An alternative group of models attempts to explain the differences between the Orang Asli groups as a product of local differentiation. Rambo (1988) suggested that the Semang and Senoi developed from the same ancestral population but differentiated through adaptation to the distinct ecological niches they came to occupy. Fix (1995; 2002), citing his research into hemoglobin E and ovalocytosis markers, proposed that the three traditions described by Benjamin represent the crystallization of divergent yet complementary lifestyles, following the initial establishment of sedentary populations within Malaya at around 5000 years ago.

By Indo-Malaysian standards, the archaeology and human skeletal record of the Malay Peninsula are very well documented, but without a clear resolution to the debates described above. Malaya and northern Sumatra are the southern outposts of the pebble-based Hoabinhian assemblages, considered to be associated with a terminal-Pleistocene to mid-Holocene forager economy, which also occur throughout Indochina to South China in the north (Van Heekeren 1972; Bellwood 1997; Forestier et al. 2005). The Malayan Neolithic is marked by the appearance, ~4,000 years ago, of pottery and modest amounts of polished stone artifacts, with similarities to their counterparts in sites in central Thailand such as Ban Kao. It is believed to be associated with the introduction to Malaya of agriculture, which then spread during the Early Metal Phase (Bellwood 1993; Bellwood 1997; Bulbeck 2003). However, the Neolithic burials provide minimal evidence for any population incursion (Bulbeck 2000; Bulbeck 2005; Bulbeck et al. 2005), and even the Early Metal Phase burials are ambiguous in terms of registering a “Mongoloid” presence (Matsumura 2005;

Bulbeck and Lauer 2006). Further, the osteological evidence is equivocal as to whether the Semang or the Senoi would be the Orang Asli most closely related to the Peninsula's Hoabinhian inhabitants (Bulbeck et al. 2005; Bulbeck and Lauer 2006).

No detailed genetic investigations using high-resolution non-recombining marker systems have been carried out to date. We have, however, recently investigated the position in the global mtDNA phylogeny of complete genome sequences of eight haplogroups found primarily in the Malay Peninsula, showing that most of them branch directly from the Eurasian mtDNA ancestor lineages ~60,000 years ago and are indigenous and unique to the Peninsula (Macaulay et al. 2005). We here extend this analysis with a phylogeographic study of mtDNA variation in the Peninsula, in order to test the population history hypotheses outlined above.

Materials and Methods

Subjects

We collected buccal cells from 260 maternally unrelated Orang Asli after obtaining informed consent. The study was passed by ethical panels in both the UK and Malaysia, and formally approved by the relevant administrative bodies at both local and national level. DNA was extracted using the InstaGene matrix (BioRad). The samples encompassed all three Orang Asli groups (locations shown in Figure 1): 112 Semang (29 Batek, 51 Jahai and 32 Mendriq), 52 Senoi (51 Temiar and 1 Semai) and 96 Aboriginal Malay (61 Semelai, 33 Temuan and 2 Jakun). For comparison 180 Sumatrans (42 from Medan, 52 from Pekanbaru, 34 from Bangka, 24 from Padang and 28 from Palembang) and seven new *Melayu* Malays (additional to a published set)

were also included in the study. DNAs from the Sumatran samples were from the MRC Molecular Haematology Unit in Oxford, while the *Melayu* Malays were collected from members of the sampling team and from Orang Asli with maternal *Melayu* ancestry. To resolve the R9b tree as far as possible, we further analyzed the complete mtDNA genome of a further Aboriginal Malay sample, four Indonesian, three Vietnamese and one Thai mtDNAs identified as putative R9 or R9b on the basis of HVS-I motifs. We also sequenced two pre-R9b lineages from Vietnam that lack the HVS-I motif but share the coding-region transition at np 1541 with R9b.

Comparative mtDNA data from Thailand, peninsular Malaysia, Taiwan (Han and aboriginals), the Philippines, Sabah, East Indonesia, Papua New Guinea, Pacific islands, the Nicobars, the Andamans, Hong Kong, mainland China, Japan, Mongolia, Korea and Central Asia were taken from the literature (Hertzberg et al. 1989; Horai and Hayasaka 1990; Redd et al. 1995; Sykes et al. 1995; Betty et al. 1996; Horai et al. 1996; Kolman, Sambuughin, and Bermingham 1996; Lee et al. 1997; Comas et al. 1998; Lum et al. 1998; Melton et al. 1998; Pfeiffer et al. 1998; Seo et al. 1998; Nishimaki et al. 1999; Fucharoen, Fucharoen, and Horai 2001; Oota et al. 2001; Prasad et al. 2001; Qian et al. 2001; Kivisild et al. 2002; Yao et al. 2002a; Yao et al. 2002b; Yao and Zhang 2002; Endicott et al. 2003; Zainuddin and Goodwin 2003; Thangaraj et al. 2005) and from the authors' unpublished data (Indonesia, Indochina and Singapore). Most comprise only HVS-I sequences, often supplemented only with the status of the 9-bp deletion in the COII/tRNA^{Lys} region. East Eurasian mtDNAs often cannot be assigned unambiguously to haplogroups when only the HVS-I sequence is available, but in some cases sufficient motif information is present to include them in phylogenetic analyses of particular haplogroups.

Sequencing and RFLP typing

We amplified and sequenced HVS-I using the primers conH1 (5'-CCTGAAGTAGGAACCAGATG-3') and conL1 (5'-TCAAAGCTTACACCAGTCTTGTAACC-3')—minimum length sequenced 16019-16401; maximum length sequenced 16004-16497; average length sequenced 16012-16497. The HVS-II of selected samples was amplified and sequenced using primers conL4 (5'-GGTCTATCACCTATTAACCAC-3') and conH4 (5'-CTGTAAAAGTGCATACCGCCA-3') minimum length sequenced 060–420; maximum length sequenced 039–426. PCR products were purified using QIAquick PCR purification columns (Qiagen) and sequenced using an ABI 3700 capillary sequencer (Dundee University sequencing service) or a Beckman-Coulter CEQ8000 sequencer. Sequence traces were unambiguous and checked by two individuals, and were in addition checked phylogenetically using the recommendations of Bandelt et al. (2002). Haplogroup status was clarified by screening RFLPs diagnostic of particular haplogroups, as follows: haplogroup M (+10397 *AluI*, +10394 *DdeI*), N (–10397 *AluI*, –10394 *DdeI*), M7 (+9824 *HinfI*), D (–5176 *AluI*), E (–7598 *HhaI*), G (+4831 *HhaI*), U (+12308 *HinfI*, in the presence of a mismatch primer, as described in Torroni et al. 1996), I (+10032 *AluI*). In addition, the haplogroup B affiliation was checked by screening for the 9-bp deletion in the COII/tRNA^{Lys} region (Hertzberg et al. 1989), and haplogroup F affiliation by sequencing position 10310 within the fragment 10270–10991.

Complete sequencing of haplogroup R9 mtDNA genomes was performed using the protocol of Torroni et al. (2001). In an attempt to obtain a more resolved tree the coding region variants 4017 and 7849 were respectively RFLP screened with *MnII* and *PsyI* in fourteen R9b North Thai samples with the 16288 variant, and one Indonesian sample with the 16288 and 16192 variants.

Phylogenetic and population analyses

We constructed reduced median networks (Bandelt et al. 1995) of HVS-I sequences within each haplogroup using the program Network 4.1. The time to the most recent common ancestor (TMRCA) of a haplogroup was estimated using the statistic ρ combined with an estimated mutation rate of 1 transition every 20,180 years in HVS-I between nucleotide positions (nps) 16090 and 16365 (Forster et al. 1996) and 1 substitution every 5,140 years in the coding region between nucleotide positions (nps) 577 and 16023 (Mishmar et al. 2003), with a heuristic estimate of SE (σ) following Saillard et al. (2000). Haplotype diversity was calculated as $1 - \sum_i x_i^2$, where x_i is the relative frequency of the i th haplotype in the sample (Torroni et al. 2001).

Results

HVS-I analysis and RFLP testing indicated that the majority of mtDNA lineages in the Orang Asli, whilst falling into the three major non-African haplogroups M, N and R, did not closely resemble any known mtDNA lineages. We resolved the relationships between the major clades and their place in the global mtDNA phylogeny by means of complete sequence analysis (Macaulay et al. 2005). A

skeleton of the resulting phylogeny that incorporates previously known East Eurasian variation is shown in Figure 2. It is immediately clear that there are a number of indigenous lineages in the Orang Asli that branch directly from the Eurasian founder haplotypes, at the roots of haplogroups M, N and R. It is also clear that there are striking differences between the distribution of mtDNAs in the different Orang Asli groups (Table 1 and Supplementary Material). The predominant Orang Asli clades are M21a in the Mendriq and Batek Semang, R21 in the Jahai Semang and Temiar Senoi, F1a in the Temiar Senoi, N21 in the Semelai, M22 in the Temuan and R9b in both groups of Aboriginal Malays. Of these, only F1a is commonly found outside the Malay Peninsula.

Haplotype diversity and PCA

The limited number of sequence types and high levels of haplotype sharing suggest that all Orang Asli groups have lost diversity through drift, with the Aboriginal Malays more diverse than the Semang or Senoi. Haplotype diversity values confirm this impression (Table 2). The least diverse group is the Mendriq Semang, who today number only a few hundred individuals, in which >84% of their sequences belong to haplogroup M21a. The most diverse are the Temuan Aboriginal Malays, and the Temiar Senoi fall between the two extremes. This difference is reflected in the values for the three Orang Asli groups as a whole: the Semang are the least diverse and Aboriginal Malays the most, with the Senoi in between. All are substantially less diverse than the five Sumatran groups studied.

A principal components analysis of haplogroup frequencies (Figure 3) supports the traditional clustering of Batek and Mendriq Semang and Semelai and

Temuan Aboriginal Malays, but clusters the Jahai Semang towards the Temiar Senoi, as a result of their sharing of the indigenous haplogroup R21 (see below). The Aboriginal Malays were closest to the Sumatrans in the first component (24.3%), with the Batek and Mendriq at the opposite pole, and the Jahai and Temiar between the two extremes.

Orang Asli mtDNAs from the north

The two familiar and widespread Southeast Asian mtDNA haplogroups, are haplogroup B and R9, the latter encompassing haplogroup F (Torroni et al. 1994; Kivisild et al. 2002; Yao et al. 2002a; Yao and Zhang 2002; Kong et al. 2003). Although previously identified at high frequency in the Semai Senoi (Melton et al. 1995), haplogroup B is only present at low frequencies amongst the Orang Asli groups that we sampled, except for a single B5b type elevated to high frequency, presumably by drift, in the Batek. This type seems most likely to have been introduced fairly recently from island Southeast Asia, since it is a derived type only present in one ethnic group, and the ancestral sequence is found in both Sumatra and eastern Indonesia, and not in Indochina (authors' unpublished data).

The two main branches of haplogroup R9 are R9b and F (Kong et al. 2003), which diverged from R9 ~53,000 years ago (Macaulay et al. 2005). Several clades within these are represented in the Orang Asli, each with a different distribution. The distribution of R9b is especially interesting, because it is much less widely distributed in Southeast Asia than haplogroup F, thus potentially opening a window onto the time of early settlement. Within the Orang Asli, R9b is found only in the Aboriginal Malays (both Semelai and Temuan), and is largely represented by just one frequent

HVS-I type, present in both groups. R9b is rare elsewhere, but is found at low frequencies in Vietnam, Thailand and Indonesia (authors' unpublished data) and in the Yunnan and Guangxi provinces of South China (Yao and Zhang 2002).

In order to clarify the phylogeographic pattern, which is ambiguous in the HVS-I sequences, we sequenced the complete mtDNA genome of two Aboriginal Malays harboring distinct R9b sequences, as well as four from Indochina and four from Indonesia. Furthermore, we sequenced two pre-R9b lineages from Vietnam that lack the HVS-I motif but share the coding-region transition at np 1541 with R9b. The early divergence of Vietnamese lineages in the reconstructed phylogeny (figure 4) suggests an ancient divergence of pre-R9b $\sim 29,000$ ($\pm 6,600$) years ago in Indochina, and divergence of R9b $\sim 19,000$ ($\pm 5,400$) years ago in Vietnam/South China (published Chinese HVS-I R9b lineages also cluster at this point). There is then a derived subclade from which the Thai, Aboriginal Malay and Indonesian R9b lineages all emerge that dates to $\sim 9,000$ ($\pm 2,700$) years ago. Many of the Indonesians fall into a further derived subclade and there are no nesting relationships between the Aboriginal Malays and Indonesians other than common ancestry. This overall pattern suggests that R9b diversified in Indochina and then spread southwards into the Malay Peninsula at least 9,000 years ago, with some lineages subsequently dispersing throughout island Southeast Asia.

F1a, a common and widespread Southeast Asian clade, is found largely in the Senoi, where only the derived subclade F1a1a is present (Figure 5). It is found in both groups of Senoi sampled: almost half of our Temiar sample, and also in the single Semai individual; yet it is entirely absent from all three Semang groups and also from the Temuan, although it is present at low frequencies in the Semelai. The root type of F1a1a is shared with subjects from Indonesia, Taiwan and China. However, it is

especially frequent in both published data from Thailand (~10%) (Fucharoen, Fucharoen, and Horai 2001), and in our unpublished data from north-west Thailand (~21%) and Vietnam (~20%). Derived types are found in five Aboriginal Malays (all Semelai) and, on a separate branch, seven Senoi. An interesting link emerges with some Nicobarese, who also possess F1a1a at high frequency. However, given the high frequency of the root type in Indochina, this probably reflects a shared common ancestry of some Senoi and Nicobarese lineages in Indochina, rather than any specific links between the two. We estimated the MRCA of F1a1a to be ~10,700 (\pm 4,500) years old from complete sequences (Macaulay et al. 2005), whereas from control-region diversity we estimate an age of ~7700 (\pm 3000) years in Indochina. This suggests an arrival of new people in the Malay Peninsula from a northern source (most likely in Indochina), and intermarriage with the ancestors of the Semang, within that time.

The widespread, if uncommon, mainland East Asian haplogroup N9a is found in the form of the derived subclade N9a6a in all three main Orang Asli groups at similar, low frequencies, although again it is distributed very unevenly and is diverse only in the Aboriginal Malays. N9a6a has an age estimated at ~5500 (\pm 2600) years and is shared with *Melayu* Malays and Indonesians. It nests within N9a6, which is otherwise found largely in South China, Indochina and Sumatra. Its distribution suggests a history similar to that of R9b, with a deep ancestry in mainland Southeast Asia and a more recent expansion through Malaysia into island Southeast Asia.

Orang Asli mtDNAs from off-shore

The Asian macro-haplogroup M is present in all of the Orang Asli groups, but is also very unevenly distributed. Of particular interest, 9% Aboriginal Malays (all Semelai) belong to the root type of the starlike subclade M7c1c, which is also found in *Melayu* Malays and throughout Austronesian-speaking populations in Taiwan, island Southeast Asia and as far east as Micronesia. The ancestral M7c1* is most common and diverse in China, consistent with a dispersal from South China to island Southeast Asia, and thence (more recently) into the Malay Peninsula. The age of M7c1c, which should post-date the dispersal into island Southeast Asia, is estimated at ~8300 (\pm 2400). This signal is therefore consistent with an expansion of Austronesian speakers, mariners-cum-agriculturalists, or both in the mid-Holocene as proposed by Bellwood (2004), and may indicate a subsequent small-scale dispersal into the Malay Peninsula from Indonesia.

Haplogroup N21 is characterized by the ancestral (L3) state at the haplogroup N motif position 8701, along with a HVS-I transition at np 16193 (figure 6). This is most likely due to a reversion within haplogroup N of the diagnostic position 8701, since such a reversion occurs independently in the data set of Fuku et al. (2002). This would suggest an age of up to ~63,000 years, the age of the MRCA of haplogroup N (Macaulay et al. 2005). N21 is found only in Aboriginal Malays (both Semelai and Temuan), several *Melayu* Malays, and in Indonesia. Although the latter are rare, they are much more diverse than the lineages in the Aboriginal Malays, which are also highly derived within the N21 tree, suggesting an origin in island Southeast Asia and a recent dispersal into the Malay Peninsula. Four Temuan also belong to another novel clade, N22, which again is rare but more diverse in Indonesia.

Indigenous Orang Asli mtDNAs

Within haplogroup M, there is an ancient and yet highly localized clade, M21, with three derived sister subclades (figure 7). M21 is ~57,000 years old (Macaulay et al. 2005). M21a is most common in the Semang (reaching its highest levels at ~84% in the Mendriq) and is also present in the “Maniq” Semang of southern Thailand (Fucharoen, Fucharoen, and Horai 2001), suggesting that it arose in ancestors of the Semang. The Thai Semang samples belong to a single derived sequence type, also found in a minority of the Semang in our sample (Batek and Jahai). The most common type is present in 38 Semang, four *Melayu* Malays, one Senoi and one Aboriginal Malay, and derivatives are present in both Semang and Senoi. Curiously, the root type of M21a is seen only in two Aboriginal Malays, in one *Melayu* Malay, and in a single individual from a sample of 89 individuals from southern Borneo. The most likely explanation for this pattern is gene flow from Semang or Senoi into both the Aboriginal Malays and into Borneo (cf. Adelaar 1995).

The much rarer M21b, which shares a common ancestor with M21a (labeled M21a'b) ~44,000 years ago, may also be an indigenous Malay Peninsula haplogroup. It is present in both Semang and Senoi, with a very derived subclade shared between six Aboriginal Malays and several individuals from Island Southeast Asia. M21c, a sister clade to M21a'b, is even rarer than M21b, having been sampled in only two Semelai. However, it is an intriguing indicator of possible long-standing (perhaps even pre-glacial) relationships between the apparently distinct aboriginal groups.

M22, which diverged directly from the MRCA of haplogroup M ~63,000 years ago, is found in 16% of Temuan Aboriginal Malays and two Thais. The very

few remaining unclassified haplogroup M samples have been grouped together for convenience as M* (except for one belonging to the East Asian haplogroup M9; see Figure 2), and their phylogeographic distributions are as yet undetermined for lack of similar sequence types in the HVS-I database.

A possible sister haplogroup of haplogroup R9 is the novel clade R21, present only in the Semang (mainly Jahai) and Temiar Senoi, the majority of whom share a single HVS-I sequence type. Using coding-region information only, R21 diverged from the common haplogroup R ancestor ~60,000 years ago (Macaulay et al. 2005), although the putative control-region link with haplogroup R9 (at np 16304) may imply a slightly younger common ancestor. The only plausible neighboring sequence types are in a sample from Singapore and possibly two north-east Chinese, which share the reversion at 10398 and one of which also shares the 16304 variant. R21 may therefore, like M21, be largely indigenous to the Semang/Senoi, and may represent another component of deep Upper Pleistocene ancestry within the Malay Peninsula that has not succeeded in dispersing more widely.

Discussion

The mtDNA variation shows strong evidence for indigenous origins of the Orang Asli within the Malay Peninsula, dating back to ~60,000 years ago—probably within only a few thousand years of the dispersal from East Africa (Macaulay et al. 2005). This is suggested strongly by haplogroups M21 and R21, which predominate in the Semang and Senoi; whereas N21 and N22, which appear to be largely restricted to the Aboriginal Malays, may represent gene flow from Island Southeast Asia. Gene flow from the outside world has been by no means negligible: all three groups have seen

the Holocene arrival of N9a lineages; the Senoi have a substantial Holocene component from Indochina in F1a1a; the Batek Semang have B5b, probably from island Southeast Asia; and in addition to N21 and N22 the Aboriginal Malays have lineages, such as M7c1c, indicating recent arrivals from off-shore, perhaps associated with the arrival of Austronesian-speaking people.

The Semang appear to be the most direct descendants of the original inhabitants of the Peninsula, and to have experienced only minor subsequent gene flow from outside, probably in the recent past. However, the three Semang groups are somewhat different from each other in their haplogroup distributions, and the Jahai in particular resemble the Temiar rather more than they do the other Semang groups. Further, the mitochondrial relationships of the Semang seem not to correlate with linguistic classifications, in which Jahai and Mendriq are sister languages which then relate to Batek (Benjamin 2002a). Most importantly, none of the Semang resemble the Andamanese who have their own indigenous haplogroup M mtDNAs (Endicott et al. 2003; Thangaraj et al. 2003). Based on these considerations, and classical marker data on the Philippine Aeta (Omoto 1995), the genetic evidence refutes the notion of a specific shared ancestry between the “Negrito” groups of the Andaman Islands, Malay Peninsula, and Philippines.

A different demographic signal appears to be indicated by the distribution of haplogroup R9b, found at high frequency only in the Aboriginal Malays, and perhaps also N9a. Our complete sequences suggest a Pleistocene origin to the north in Indochina, with an early-Holocene dispersal southwards through the Malay Peninsula and into island Southeast Asia. This runs counter to the prevailing view that regards the Aboriginal Malays as having arrived in the Peninsula from island Southeast Asia only in the mid-Holocene, as a result of the putative expansion of Austronesian

speakers in the archipelago (e.g. Bellwood 1997). The only echo in the archaeological literature, of which we are aware, is the suggestion of Van Heekeren (1972) that the Hoabinhian had originated in South China before spreading south to Malaya and north Sumatra (at around the Pleistocene/Holocene boundary). On the other hand, haplogroups N21, N22 and M7c1c suggest an equally large off-shore component, dating to the mid/late Holocene, in the ancestry of the Aboriginal Malays.

Perhaps the most striking signal is the presence of F1a1a, which aside from the apparently indigenous R21 is the most common haplogroup in the Senoi, carried by almost half of the individuals sampled. This haplogroup, which is of early to mid-Holocene age, has been observed elsewhere at high frequencies only in Indochina and probably dispersed there from South China (where it is less frequent but more diverse, and where its one-step ancestor is found) during the Holocene. This suggests that almost half of the maternal lineages of the Senoi may trace back to an origin in Indochina at some point within the last 7000 years or so. This is consistent with the view of Bellwood (1993), that the Neolithic was brought into Peninsular Malaysia by groups from central Thailand (associated with the Ban Kao Neolithic culture), which intermarried with indigenous groups to create the ancestors of modern Senoi. These people may also, as Bellwood (1993) suggests, have brought the Austroasiatic languages to the Malay Peninsula.

It should be remembered that all three groups have been subject to considerable genetic drift, as indicated by both the mtDNA diversity patterns and osteological data (Bulbeck and Lauer 2006). This drift places limits on the robustness of any phylogeographic analysis. The survival of small, semi-isolated Orang Asli populations in recent times is also suggested by the ethnographic data (e.g. Benjamin 2002b). The Semang as a whole numbered ~3200 individuals in the year 2000, with

the Senoi and Aboriginal Malays at approximately 49,000 and 40,000, respectively (Benjamin 2002a, p.22)—substantially more than only a few decades ago (Carey 1976). They exhibit less extreme patterns of drift than the Semang, though the Senoi, now the largest group in terms of census size, appear to have undergone more drift than the Aboriginal Malays. This may have been due to the initial processes of ethnogenesis, or subsequent founder effects, such as the proposed expansion of the Temiar eastwards in recent times (Benjamin 2002a).

Phylogeographic analysis suggests at least four detectable colonization events that affected the Orang Asli, respectively dated to over 50,000 years ago, ~10,000 years ago, the middle Holocene, and the late Holocene. Although this brings to mind the traditional “layer cake” theory, the latter’s assumption of unchanged relicts of earlier population waves is completely unfounded. All three Orang Asli groups have local roots that reach back to ~50,000 years ago, and all have been affected to a greater or lesser extent by subsequent migrations to the Peninsula. Nonetheless, the differences between the Orang Asli groups do reflect a distinct ancestry to a greater degree than Rambo’s (1988) model of local ethnogenesis would imply. Bellwood’s model of a melting pot combining elements from distinct forager and agriculturalist occupations is also too simple; there appears to have been a detectable immigration from the north, perhaps associated with the appearance of the Hoabinhian, thousands of years before Bellwood’s proposed Neolithic immigration to the Peninsula (which our evidence confirms). It is important to acknowledge the role of local evolution for all three groups, from at least the early Holocene onwards, whilst allowing for some immigration—perhaps several waves from the north, affecting the gene pools of both the Senoi and Aboriginal Malays, and from island Southeast Asia, primarily affecting the Aboriginal Malays (cf. Rayner and Bulbeck 2001; Fix 2002).

It appears, then, that the Orang Asli may indeed represent in microcosm demographic processes that are likely to be seen more widely in Southeast Asia: some maternal lineages that trace back to the first settlement, more than 50,000 years ago; some representing late-glacial and early Holocene dispersals; and some pointing to Neolithic or post-Neolithic shifts of population—perhaps also, as has been widely suggested, involving the spread of languages. In particular, our evidence of a ~10,000 year old migration into the Peninsula may open new insights into the interpretation of the Hoabinhian, possible dispersals of Southeast Asian foragers adapted to different vegetation regimes prior to the Holocene sea-level rises (Bird, Taylor, and Hunt 2005), and the osteological variability shown by Southeast Asia's late Pleistocene and early Holocene human remains (Oppenheimer 2003, pp. 266–268).

Electronic Database Information

The sequence data for this paper appear in GenBank under accession nos. AY966035–AY966044, AY966046–AY966051, AY966053–AY966077, AY966080–AY966086, AY966088–AY966101, AY966103–AY966108, AY966110–AY966400, DQ973531–DQ973623, DQ979791–DQ979795, and DQ981465–DQ981475.

Acknowledgments

We thank Granada Television/Discovery Channel, the University of Huddersfield, the British Academy and the Bradshaw Foundation for financial support; Adi Taha and the Muzium Negara Malaysia and the Universiti Putra Malaysia for support and sponsorship before and during the sampling expeditions; Norazila Kassim Shaari for help with sample collection; John Clegg and A.S.M. Sofro for Indonesian samples;

Andy Cassidy and Graeme Scott for HVS-I sequencing; and the people of the Malay Peninsula for generously participating by providing cheek-swab samples for this study.

Literature Cited

- Adelaar KA. 1995. Borneo as a cross-roads for comparative Austronesian linguistics. In: Fox JJ, Tryon D, and Bellwood P, Eds. *The Austronesians: Comparative and Historical Perspectives*. The Australian National University, Canberra. p. 75-95
- Bandelt H-J, Forster P, Sykes BC, and Richards MB. 1995. Mitochondrial portraits of human populations using median networks. *Genetics* 141:743-753.
- Bandelt H-J, Quintana-Murci L, Salas A, and Macaulay V. 2002. The fingerprint of phantom mutations in mitochondrial DNA data. *Am. J. Hum. Genet.* 71:1150-1160.
- Bellwood P. 1993. Cultural and biological differentiation in Peninsular Malaysia: the last 10,000 years. *Asian Perspectives* 32:37-60.
- Bellwood P. 1997. *Prehistory of the Indo-Malaysian archipelago*. University of Hawaii Press, Hawaii.
- Bellwood P. 2004. The origins and dispersal of agricultural communities in Southeast Asia. In: Glover I, and Bellwood P, Eds. *Southeast Asia from Prehistory to History*. Routledge Curzon, London. p. 21-40
- Benjamin G. 1985. In the long term: three themes in Malaysian cultural ecology. In: Hutterer KL, Rambo AT, and Lovelace G, Eds. *Cultural values and human ecology in Southeast Asia*. Michigan University Press, Ann Arbor, Michigan. p. 219-278

- Benjamin G. 2002a. On being tribal in the Malay world. In: Benjamin G, and Chou C, Eds. Tribal communities in the Malay world: historical, cultural and social perspectives. Institute for Southeast Asian Studies, Singapore. p. 7-76
- Benjamin G. 2002b. Orang Asli languages: from heritage to death? In: Rashid R, and Wazir JK, Eds. Minority cultures of Peninsular Malaysia: survivals of indigenous heritage. AKASS (Malaysian Academy of Social Science), Penang. p. 99-120
- Betty DJ, Chin-Atkins AN, Croft L, Strahl M, and Esteban S. 1996. Multiple independent origins of the COII/tRNA^{Lys} intergenic 9-bp mtDNA deletion in aboriginal Australians. *Am. J. Hum. Genet.* 58:428-433.
- Bidang Integrasi Pengolahan dan Diseminasi Statistik (ed). 2001. Sumatera Selatan dalam angka Sumatera Selatan in figures 2001. Palembang: Badan Pusat Statistik Propinsi Sumatera Selatan. p. 189
- Bidang Integrasi Pengolahan dan Diseminasi Statistik (ed). 2002. Riau dalam angka in figures 2002. Pekanbaru: Badan Pusat Statistik Propinsi Riau. p. 147.
- Bidang Integrasi Pengolahan dan Diseminasi Statistik (ed). 2002. Sumatera Utara dalam angka Sumatera Utara in figures 2002. Medan: Badan Pusat Statistik Provinsi Sumatera Utara. p. 34, 44.
- Bird MI, Taylor D, and Hunt C. 2005. Palaeoenvironments of insular Southeast Asia during the Last Glacial Period: a savanna corridor in Sundaland? *Quaternary Science Reviews* 24:2228-2242.
- Birdsell JB. 1993. Microevolutionary patterns in Aboriginal Australia. A gradient analysis of clines. Oxford University Press, Oxford.
- BPS Propinsi Kepulauan Bangka Belitung (ed). 2001. Kepulauan Bangka Belitung dalam angka Kepulauan Bangka Belitung in figures 2001, p.109. Badan Pusat

Statistik Propinsi Kepulauan Bangka Belitung/Badan Perencanaan Daerah
 Propinsi Kepulauan Bangka Belitung, Tanjung Pandan.

BPS Propinsi Sumatera Barat (ed). 2002. Sumatera Barat dalam angka West Sumatera
 in figures 2002, p.41. Badan Pusat Statistik Propinsi Sumatera Barat, Padang.

Bulbeck D. 2000. Dental morphology at Gua Cha, West Malaysia, and the
 implications for "Sundadonty". *Bulletin of the Indo-Pacific Prehistory
 Association* 19:17-41.

Bulbeck D. 2003. Hunter-gatherer occupation of the Malay Peninsula from the Ice
 Age to the Iron Age. In: Mercader J, Ed. *Under the Canopy: The Archaeology
 of Tropical Rain Forests*. Rutgers University Press, New Brunswick. p. 119-
 160

Bulbeck D. 2005. The Gua Cha burials. In: Zuraina M, Ed. *The Perak Man and other
 prehistoric skeletons of Malaysia*. Penerbit Universiti Sains Malaysia, Penang.
 p. 253-309

Bulbeck D, Rahimah AK, Lauer A, Radzi Z, and Rayner D. 2005. Tooth sizes in the
 Malay Peninsula past and present: insights into the time depth of the
 indigenous inhabitants' adaptations. *International Journal of Indigenous
 Research* 1:41-50.

Bulbeck D, and Lauer A. 2006. Human variation and evolution in Holocene
 Peninsular Malaysia. In: Oxenham M, and Tayles N, Eds. *Bioarchaeology of
 Southeast Asia*. Cambridge University Press, Cambridge. p. 131-171

Carey I. 1976. *Orang Asli: the aboriginal tribes of peninsular Malaysia*. Oxford
 University Press, Kuala Lumpur.

Cole F-C. 1945. *The Peoples of Malaysia*. D. van Nostrand Co., Princeton.

- Comas D, Calafell F, Mateu E et al. 1998. Trading genes along the Silk Road: mtDNA sequences and the origin of central Asian populations. *Am. J. Hum. Genet.* 63:1824-1838.
- Endicott P, Gilbert MT, Stringer C, Lalueza-Fox C, Willerslev E, Hansen AJ, and Cooper A. 2003. The genetic origins of the Andaman Islanders. *Am. J. Hum. Genet.* 72:178-184.
- Fix A. 2002. Foragers, farmers and traders in the Malayan peninsula: origins of cultural and biological diversity. In: Morrison KD, and Junker LL, Eds. *Forager-traders in South and Southeast Asia*. Cambridge University Press, Cambridge. p. 185-202
- Fix AG. 1995. Malayan paleosociology: implications for patterns of genetic variation amongst the Orang Asli. *American Anthropology* 97:313-323.
- Forestier H, Simanjuntak T, Guillaud D, Driwantoro D, Wiradnyana K, Siregar D, Awe RD, and Budiman. 2005. Le site de Togi Ndrawa, ile de Nias, Sumatra Nord: les premieres traces d'une occupation hoabinhienne en grotte en Indonesie. *C.R. Palevol* 4:727-733.
- Forster P, Harding R, Torroni A, and Bandelt H-J. 1996. Origin and evolution of Native American mtDNA variation: a reappraisal. *Am. J. Hum. Genet.* 59:935-945.
- Fucharoen G, Fucharoen S, and Horai S. 2001. Mitochondrial DNA polymorphisms in Thailand. *J. Hum. Genet.* 46:115-125.
- Fuku N, Oshida Y, Takeyasu T, Guo LJ, Kurata M, Yamada Y, Sato Y, and Tanaka M. 2002. Mitochondrial ATPase subunit 6 and cytochrome b gene polymorphisms in young obese adults. *Bioch. Biophys. Res. Comm.* 290:1199-1205.

- Harrower G. 1933. Skeletal remains from the Kuala Selinsing excavations, Perak, Malay Peninsula. *Journal of the Malay Branch Royal Asiatic Society* 11:190-210.
- Hertzberg M, Mickleson KNP, Serjeantson S, Prior JF, and Trent RJ, . 1989. An Asian-specific 9-bp deletion of mitochondrial DNA is frequently found in Polynesians. *Am. J. Hum. Genet.* 44:510-540.
- Horai S, and Hayasaka K. 1990. Intraspecific nucleotide sequence differences in the major noncoding region of human mitochondrial DNA. *Am. J. Hum. Genet.* 46:828-842.
- Horai S, Murayama K, Hayasaka K et al. 1996. mtDNA polymorphism in East Asian populations, with special reference to the peopling of Japan. *Am. J. Hum. Genet.* 59:579-590.
- Kivisild T, Tolk H-V, Parik J, Wang Y, Papiha SS, Bandelt H-J, and Villems R. 2002. The emerging limbs and twigs of the east Asian mtDNA tree. *Mol. Biol. Evol.* 19:1737-1751 (erratum 1720:1162).
- Kolman C, Sambuughin N, and Bermingham E. 1996. Mitochondrial DNA analysis of Mongolian populations and implications for the origin of New World founders. *Genetics* 142:1321-1334.
- Kong Q-P, Yao Y-G, Sun C, Bandelt H-J, Zhu C-L, and Zhang Y-P. 2003. Phylogeny of East Asian mitochondrial DNA lineages inferred from complete sequences. *Am. J. Hum. Genet.* 73:671-676 (erratum 675:157).
- Lee SD, Shin CH, Kim KB, Lee YS, and Lee JB. 1997. Sequence variation of mitochondrial DNA control region in Koreans. *Foren. Sci. Int.* 87:99-116.

- Lum JK, Cann RL, Martinson JJ, and Jorde LB. 1998. Mitochondrial and nuclear genetic relationships among Pacific island and Asian populations. *Am. J. Hum. Genet.* 63:613-624.
- Macaulay V, Hill C, Achilli A et al. 2005. Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science* 308:1034-1036.
- Matsumura H. 2005. Bioanthropological significance of prehistoric human skeletal remains in Malaysia. In: Zuraina M, Ed. *The Perak Man and other prehistoric skeletons of Malaysia*. Penerbit Universiti Sains Malaysia, Penang. p. 425-446
- Melton T, Peterson R, Redd AJ, Saha N, Sofro ASM, Martinson J, and Stoneking M. 1995. Polynesian genetic affinities with Southeast Asian populations as identified by mtDNA analysis. *Am. J. Hum. Genet.* 57:403-414.
- Melton T, Clifford S, Martinson J, Batzer M, and Stoneking M. 1998. Genetic evidence for the proto-Austronesian homeland in Asia: mtDNA and nuclear DNA variation in Taiwanese aboriginal tribes. *Am. J. Hum. Genet.* 63:1807-1823.
- Mishmar D, Ruiz-Pesini E, Golik P, Macaulay V, Clark AG, Hosseini S, Brandon M, Easley K, Chen E, Brown MD, Sukernik RI, Olckers A, and Wallace DC. 2003. Natural selection shaped regional mtDNA variation in humans. *Proc. Natl. Acad. Sci. USA* 100:171-176.
- Nishimaki Y, Sato K, Fang L, Ma M, Hasekura H, and Boettcher B. 1999. Sequence polymorphism in the HV1 region in Japanese and Chinese. *Legal Medicine* 1:238-249.

- Omoto K. 1995. Genetic diversity and the origin of the "Mongoloids". In: Brenner S, and Hanihara K, Eds. Origins and past of modern humans as viewed from DNA. World Scientific, Singapore. p. 92-109.
- Oota H, Kurosaki K, Pookajorn S, Ishida T, and Ueda S. 2001. Genetic study of the palaeolithic and neolithic Southeast Asians. *Hum. Biol.* 73:225-231.
- Oppenheimer S. 1998. *Eden in the east*. Weidenfeld and Nicholson, London.
- Oppenheimer S. 2003. *Out of Eden*. Constable and Robinson Ltd, London.
- Pfeiffer H, Steighner R, Fisher R, Mörnstad H, Yoon CL, and Holland MM. 1998. Mitochondrial DNA extraction and typing from isolated dentin - experimental evaluation in a Korean population. *Int. J. Legal Med.* 111:309-313.
- Prasad BVR, Ricker CE, Watkins WS, Dixon ME, Rao BB, Naidu JM, Jorde LB, and Bamshad M. 2001. Mitochondrial DNA variation in Nicobarese islanders. *Hum. Biol.* 73:715-725.
- Qian YP, Chu Z-T, Dai Q, Wei C-D, Chu YJ, Tajima A, and Horai S. 2001. Mitochondrial DNA polymorphism in Yunnan nationalities in China. *J. Hum. Genet.* 46:211-220.
- Rambo AT. 1988. Why are the Semang? In: Rambo AT, Gillogly K, and Hutterer KL, Eds. *Ethnic diversity and the control of natural resources in Southeast Asia*. Michigan papers on South and Southeast Asia, Ann Arbor, Michigan. p. 19-35
- Rashid R. 1995. Introduction. In: Rashid R, Ed. *Indigenous Minorities of Peninsular Malaysia: Selected Issues and Ethnographies*. Intersocietal and Scientific Sdn. Bhd., Kuala Lumpur. p. 1-17
- Rayner D, and Bulbeck D. 2001. Dental morphology of the "Orang Asli" aborigines of the Malay Peninsula. In: Henneberg M, Ed. *Causes and effects of human*

- variation. Australasian Society for Human Biology, University of Adelaide, Adelaide.
- Redd AJ, Takezaki N, Sherry ST, McGarvey ST, Sofro ASM, and Stoneking M. 1995. Evolutionary history of the COII/tRNA(Lys) intergenic 9-base-pair deletion in human mitochondrial DNAs from the Pacific. *Mol. Biol. Evol.* 12:604-615.
- Saillard J, Forster P, Lynnerup N, Bandelt H-J, Nørby SS. 2000. mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. *Am J Hum Genet* 67:718-726.
- Seo Y, Stradmann-Bellinghausen B, Rittner C, Takahama K, and Schneider PM. 1998. Sequence polymorphism of mitochondrial DNA control region in Japanese. *Foren. Sci. Int.* 97:155-164.
- Sykes B, Leiboff A, Low-Beer J, Tetzner S, and Richards M. 1995. The origins of the Polynesians—an interpretation from mitochondrial lineage analysis. *Am. J. Hum. Genet.* 57:1463-1475.
- Thangaraj K, Singh L, Reddy AG, Rao VR, Sehgal SC, Underhill PA, Pierson M, Frame IG, and Hagelberg E. 2003. Genetic affinities of the Andaman Islanders, a vanishing human population. *Curr. Biol.* 13:86-93.
- Thangaraj K, Chaubey G, Kivisild T, Reddy AG, Singh VK, Rasalkar AA, and Singh L. 2005. Reconstructing the origin of Andaman Islanders. *Science* 308:906.
- Torrioni A, Miller JA, Moore LG, Zamudio S, Zhuang JG, Droma T, and Wallace DC. 1994. Mitochondrial DNA analysis in Tibet: implications for the origin of the Tibetan population and its adaptation to high altitude. *Am. J. Phys. Anthropol.* 93:189-199.

- Torroni A, Huoponen K, Francalacci P, Petrozzi M, Morelli L, Scozzari R, Obinu D, Savontaus M-L, and Wallace DC. 1996. Classification of European mtDNAs from an analysis of three European populations. *Genetics* 144:1835-1850.
- Torroni A, Rengo C, Guida V, Cruciani F, Sellitto D, Coppa A, Calderon F L, Simionati B, Valle G, Richards M, Macaulay V, Scozzari R. 2001. Do the four clades of the mtDNA haplogroup L2 evolve at different rates? *Am J Hum Genet* 69:1348-1356
- Van Heekeren HR. 1972. *The Stone Age of Indonesia*, 2nd edition. Martinus Nijhoff, The Hague.
- Yao Y-G, Kong Q-P, Bandelt H-J, Kivisild T, and Zhang Y-P. 2002a. Phylogeographic differentiation of mitochondrial DNA in Han Chinese. *Am. J. Hum. Genet.* 70:635-651.
- Yao Y-G, Nie L, Harpending H, Fu Y-X, Yuan Z-G, and Zhang Y-P. 2002b. Genetic relationship of Chinese ethnic populations revealed by mtDNA sequence diversity. *Am. J. Phys. Anthropol.* 118:63-76.
- Yao Y-G, and Zhang Y-P. 2002. Phylogeographic analysis of mtDNA variation in four ethnic populations from Yunnan Province: new data and a reappraisal. *J. Hum. Genet.* 47:311-318.
- Zainuddin Z, and Goodwin W. 2003. Mitochondrial DNA profiling of modern Malay and Orang Asli populations in peninsular Malaysia. In: Doutremepuich C, and Morling N, Eds. *Progress in Forensic Genetics*. Elsevier Science, Amsterdam. p. 428-430

Table 1. Haplogroup Frequencies in Orang Asli, Malay and Sumatran Populations

	Semang			Senoi	Aboriginal Malay			Melayu ^a	Sumatra				
	Batek	Jahai	Mendriq	Temiar	Semelai	Temuan	Jakun		Medan	Pekanbaru	Bangka	Padang	Palembang
A	0.02
B*	0.09	.	.	.	0.04	0.03	0.17	.
B4a	0.03	.	.	0.01	.	0.15	.	0.06	0.04
B4b	0.01	.	0.02	.	.	.
B4c	0.06	0.02	0.12	0.06	0.04	0.07
B5a	0.03	.	.	0.02	.	.	.	0.09	0.05	0.04	.	.	0.11
B5b	0.45	.	0.06	0.01	0.10
C	0.01
D	0.01
E	0.06	0.07	0.02	0.15	0.08	0.04
F1a*	0.06	0.02	0.10	0.03	0.25	0.07
F1a1*	0.04
F1a1a	.	.	.	0.43	0.07	.	0.50	0.08	0.02	0.04	0.03	.	0.18
F3	0.01
F4	0.05	0.02	.	.	.
H	0.01
I	0.02	.	.	.
M*	.	.	.	0.04	0.02	0.03	0.5	0.19	0.17	0.17	0.38	0.13	0.21
M7*	0.07	0.02	.	0.04	.
M7b/M7b1	0.04	0.10
M7c1a	0.02
M7c1c	0.13	.	.	0.06	0.10	0.10	0.03	.	0.21
M9	0.01
M12	0.03
M21a	0.48	0.16	0.84	0.06	0.03	0.03	.	0.05
M21b	.	0.04	0.03	0.02	0.07	0.06	.	.	0.02
M21c	0.03
M22	0.18
N*	0.01	.	.	0.12	0.04	.
N21	0.31	0.15	.	0.02	0.04

	Semang			Senoi	Aboriginal Malay				Sumatra				
	Batek	Jahai	Mendriq	Temiar	Semelai	Temuan	Jakun	Melayu ^a	Medan	Pekanbaru	Bangka	Padang	Palembang
N22	0.12
N9a6	.	0.18	.	0.06	0.02	0.12	.	0.03	.	.	0.09	0.13	.
P	0.01
Q	0.02
R*	0.03
R9*	0.01	0.02	0.04	0.03	.	.
R9b	0.28	0.21	.	0.01	.	0.04	.	0.04	.
R21	0.03	0.63	0.06	0.37	.	.	.	0.02
R22	0.02
U7	0.02	.	.	.
Y2	0.02	0.17	0.06	0.06	.	.
Z	0.04

^a Data includes that of Zainuddin and Goodwin (2003)

Table 2. Haplogroup diversity and population sizes in the sampled populations

POPULATION GROUP ^a	POPULATION	HAPLOTYPE DIVERSITY	POPULATION (2000) ^b
Semang	Batek	0.675	960
	Jahai	0.554	1049
	Mendriq	0.535	145
	<i>All</i>	0.760	2154
Senoi	Temiar	0.784	15,122
	<i>All</i>	0.780	15,122
Aboriginal	Semelai	0.841	4103
Malay	Temuan	0.889	16,020
	<i>All</i>	0.889	20,123
Sumatrans	Medan	0.959	1.25 million
	Pekanbaru	0.952	0.53 million
	Bangka	0.951	0.45 million
	Palembang	0.940	1.10 million
	Padang	0.939	0.71 million
	<i>All</i>	0.982	~40 million (all Sumatra)

^a See figure 1.

^b Population size estimates: Orang Asli for the year 2000 from (Benjamin 2002a); Sumatran Muslims for the years 2001/2002 from Bidang Integrasi Pengolahan dan Diseminasi Statistik (2001; 2002); BPS Propinsi Kepulauan Bangka Belitung (2001); BPS Propinsi Sumatera Barat (2002).

Figure Legends

Figure 1. Map of peninsular Malaysia showing the locations of Orang Asli groups sampled. Based on Oppenheimer (1998), p.206.

Figure 2. Schematic tree of major Southeast Asian mtDNA haplogroups, contextualizing those found in the Malay Peninsula. Diagnostic HVS-I and coding-region markers tested are indicated; additional coding-region motif positions are shown in parentheses. Underlined mutations occur more than once in the tree. Shaded haplogroups are those found in the Orang Asli.

Figure 3. Plot of first two principal components of Malaysian and Sumatran mtDNA haplogroup frequencies. Malay data includes that of Zainuddin and Goodwin (2003). PC1 = 24.3%; PC2 = 18.2%.

Figure 4. Reduced-median network of haplogroup R9b, based on complete mtDNA sequences. Branches are labelled with the nucleotide position (np) of mutations. Letters following positions indicate transversions; others are transitions. Mutations that happened more than once in the tree are underlined. The polarity of evolution at np 16192 cannot be determined so that these branches are collapsed for the purposes of phylogeographic interpretation. Published HVS-I data indicate that further Chinese lineages are found at the node that is derived at np 183 and ancestral at 143, 4017, 7849 and 16288; our typing of 14 further Thai samples and one Indonesian sample derived at np 16288 indicates that they emerge from the node that is also derived at nps 4017 and 7849.

Figure 5. Reduced-median network of haplogroup F1a1a, based on HVS-I sequences in the region 16090–16365. Labeled as above.

Figure 6. Reduced median network of haplogroup N21, based on HVS-I sequences in the region 16090–16365. Labeled as above.

Figure 7. Reduced median network of haplogroup M21, based on HVS-I sequences in the region 16090–16365. Labeled as above.

Figure 1

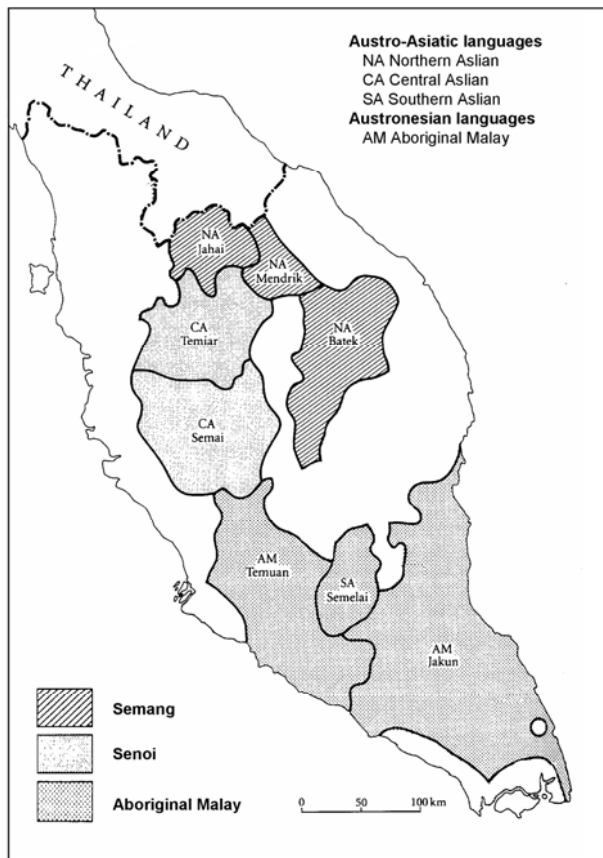


Figure 2

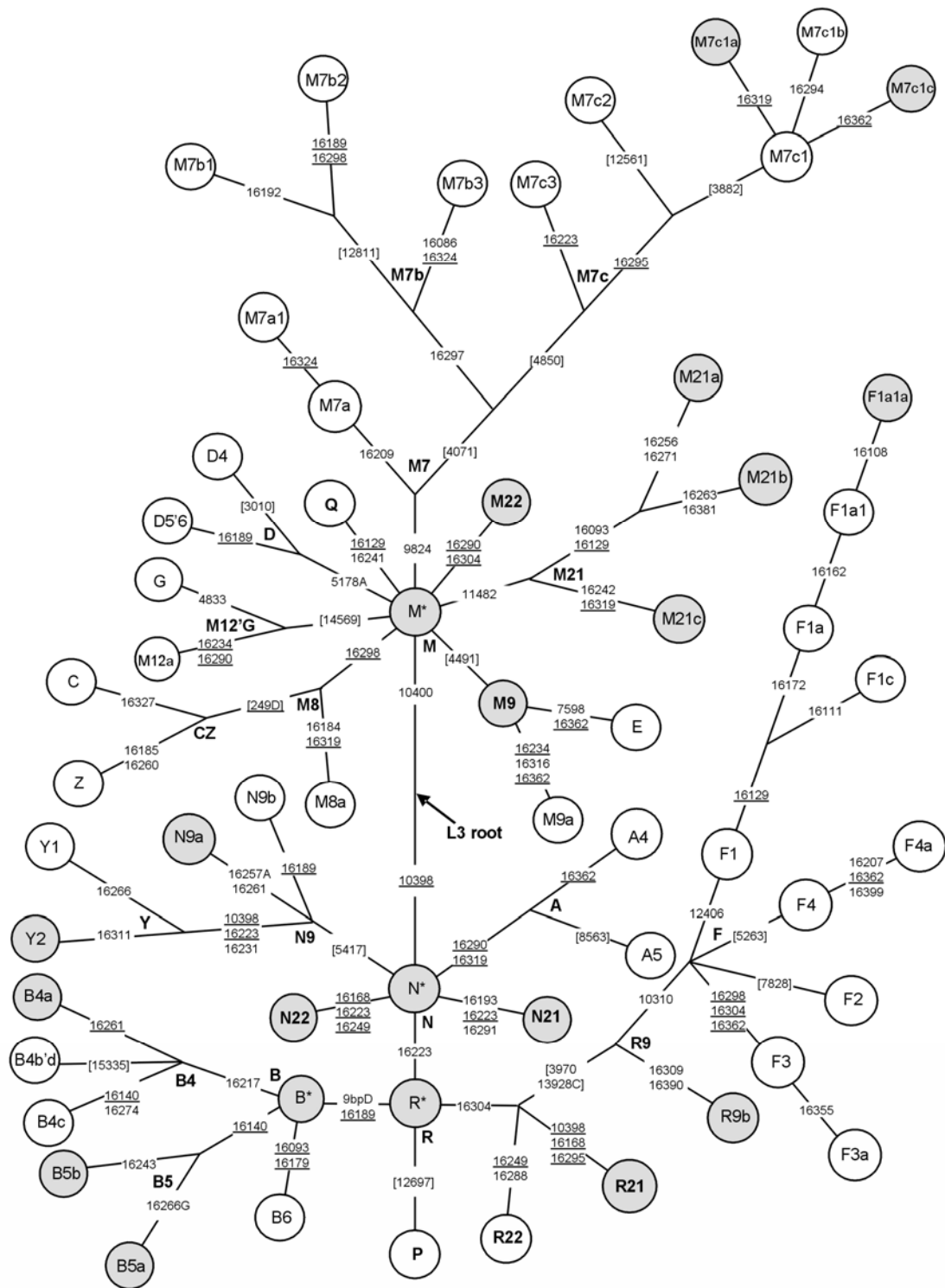


Figure 3

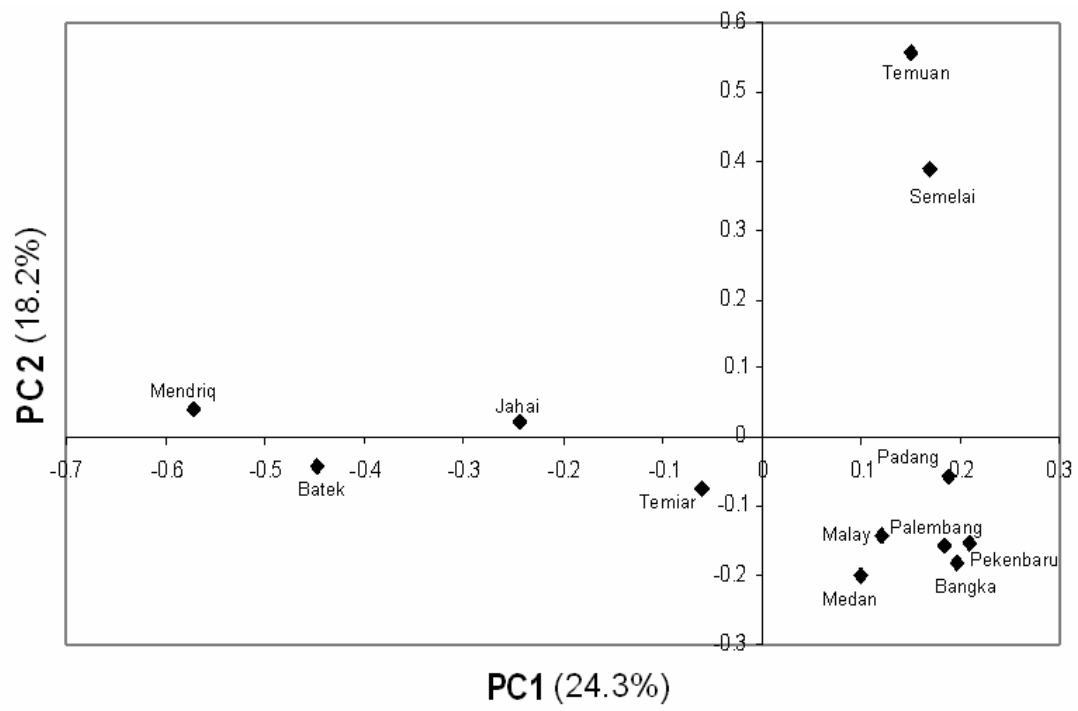


Figure 4

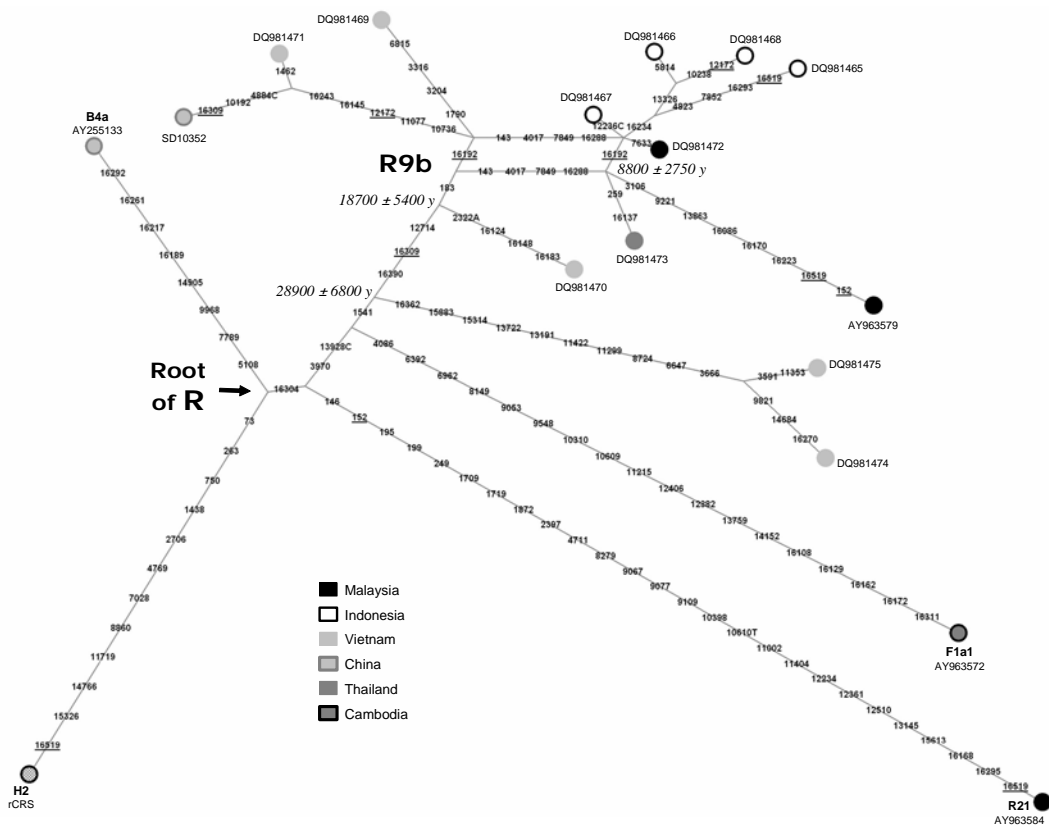


Figure 5

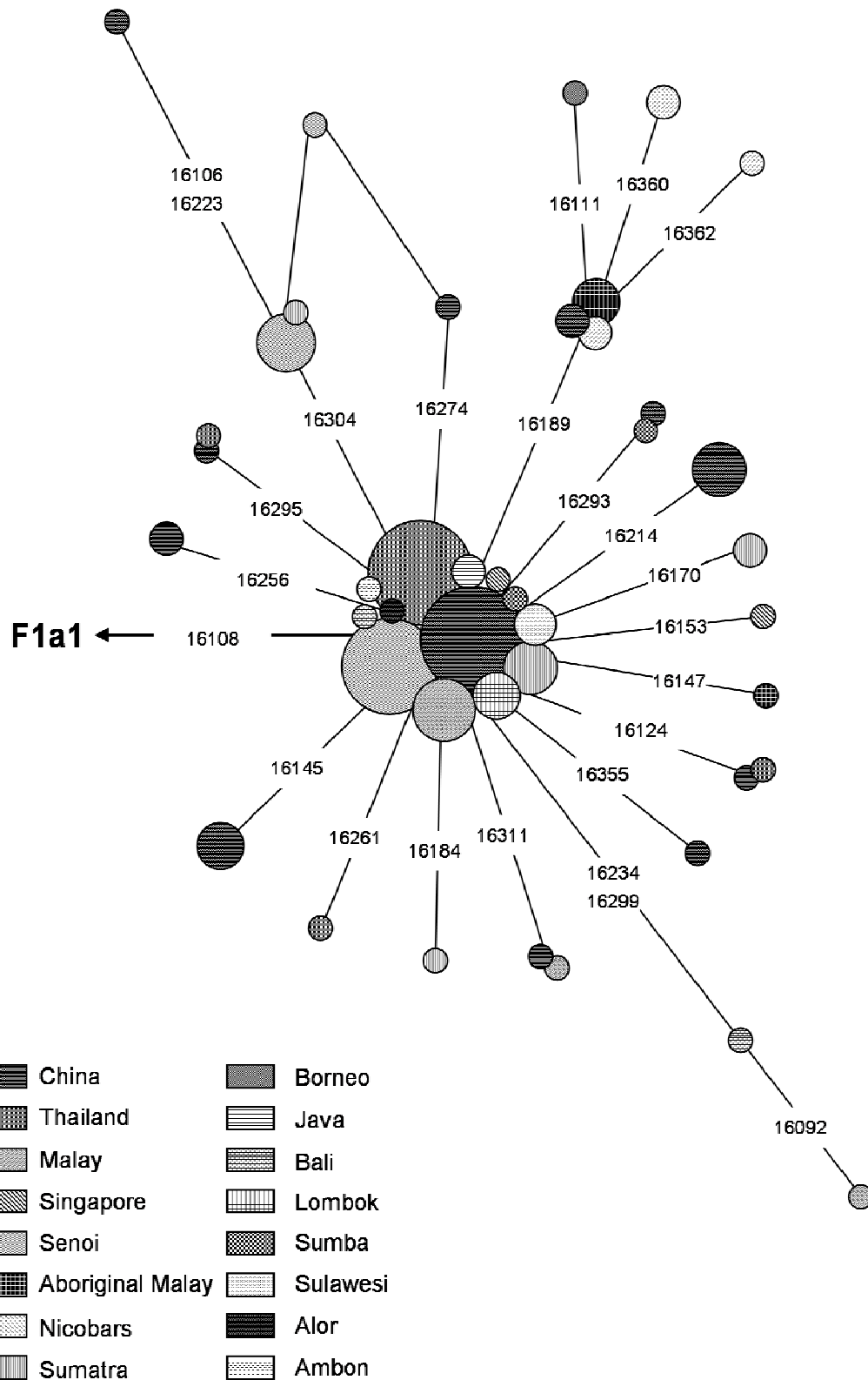


Figure 6

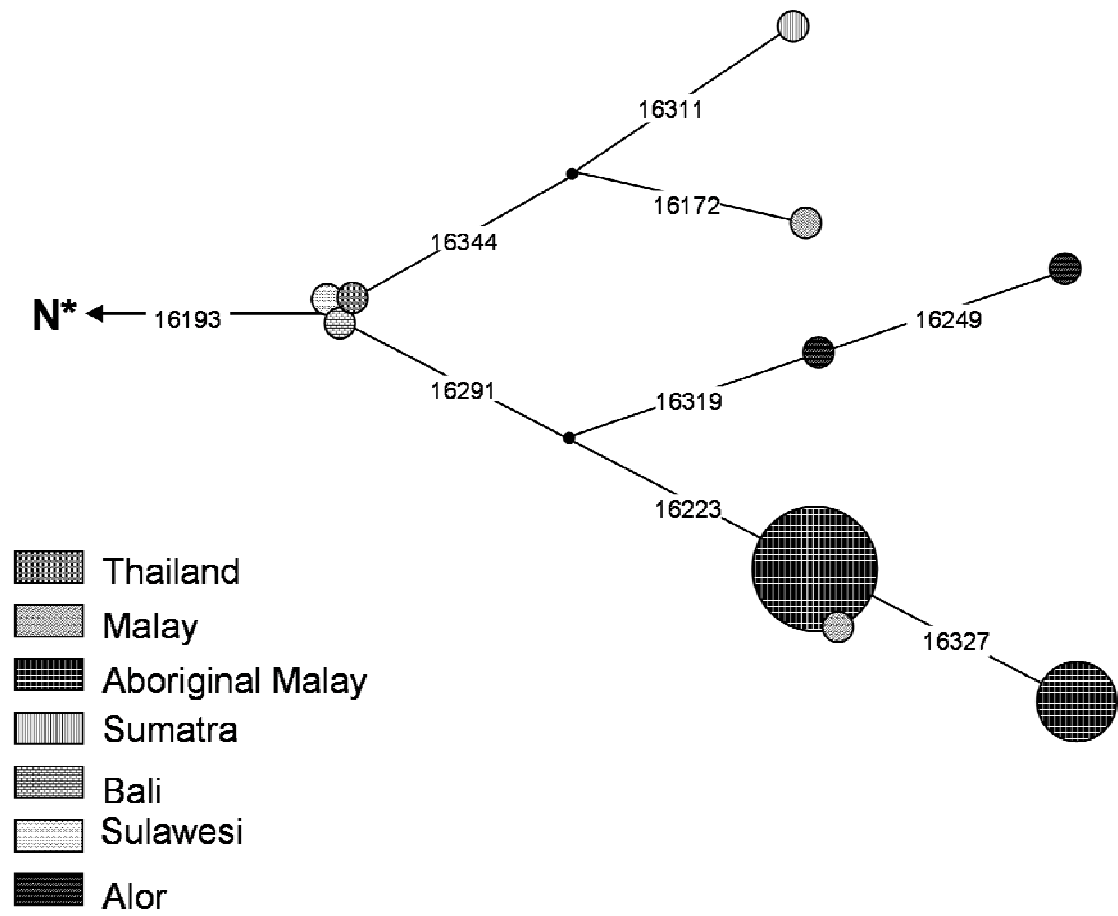


Figure 7

