# *Mitochondrial DNA Diversity in Southeast Asian Populations*

THEODORE G. SCHURR[1] AND DOUGLAS C. WALLACE[2]

*Abstract* In a previous study of Southeast Asian genetic variation, we characterized mitochondrial DNAs (mtDNAs) from six populations through high-resolution restriction fragment length polymorphism (RFLP) analysis. Our analysis revealed that these Southeast Asian populations were genetically similar to each other, suggesting they had a common origin. However, other patterns of population associations also emerged. Haplotypes from a major founding haplogroup in Papua New Guinea were present in Malaysia; the Vietnamese and Malaysian aborigines (Orang Asli) had high frequencies of haplogroup *F*, which was also seen in most other Southeast Asian populations; and haplogroup *B*, defined by the Region V 9-base-pair deletion, was present throughout the region. In addition, the Malaysian and Sabah (Borneo) aborigine populations exhibited a number of unique mtDNA clusters that were not observed in other populations. Unfortunately, it has been difficult to compare these patterns of genetic diversity with those shown in subsequent studies of mtDNA variation in Southeast Asian populations because the latter have typically sequenced the first hypervariable segment (HVS-I) of the control region (CR) sequencing rather than used RFLP haplotyping to characterize the mtDNAs present in them. For this reason, we sequenced the HVS-I of Southeast Asian mtDNAs that had previously been subjected to RFLP analysis, and compared the resulting data with published information from other Southeast Asian and Oceanic groups. Our findings reveal broad patterns of mtDNA haplogroup distribution in Southeast Asia that may reflect different population expansion events in this region over the past 50,000–5000 years.

Over the past 15 years, both restriction fragment length polymorphism (RFLP) analysis and direct DNA sequencing have been used to study the patterns of mitochondrial DNA (mtDNA) variation in Asian populations. RFLP analysis has been the primary method of detecting sequence variation in the coding regions of the mtDNA genome (Horai et al. 1984; Horai and Matsunaga 1986; Cann et al. 1987; Harihara et al. 1988; Stoneking et al. 1990; Ballinger et al. 1992; Torroni et al. 1993a, 1994a; Passarino et al. 1993; Kivisild et al. 1999; Schurr et al. 1999),

[1]Department of Anthropology, University of Pennsylvania, Philadelphia, PA 19104-6398.
[2]Center for Molecular Medicine, Emory University School of Medicine, Atlanta, GA 30322.

*Human Biology,* June 2002, v. 74, no. 3, pp. 431–452.


KEY WORDS: HAPLOTYPE, HAPLOGROUP, RFLP, CONTROL REGION, SOUTHEAST ASIA, ORANG ASLI, AUSTRONESIAN, SINO-TIBETAN, PAPUA NEW GUINEA

whereas sequencing studies have primarily targeted the noncoding control region (CR) to detect nucleotide diversity among individuals (Lum et al. 1994; Sykes et al. 1995; Horai et al. 1996; Kolman et al. 1996; Lum and Cann 1998; Richards et al. 1998; Redd and Stoneking 1999). Collectively, these studies have clearly shown that a number of distinctive mtDNA lineages, or haplogroups, are present in these populations, with some of them having undergone considerable differentiation. Furthermore, many sequence polymorphisms are known to have certain ethnic groups or geographic regions (e.g., Papua New Guinea [PNG] populations; Stoneking et al. 1990; Ballinger et al. 1992; Redd and Stoneking 1999), thereby permitting the reconstruction of the genetic history of these mtDNAs, and potentially that of the populations in which they occur.

Other types of sequence polymorphisms that have been used to reveal population affinities in Asia are small insertion-deletion mutations occurring in the noncoding regions of the mtDNA genome, where small repeat sequences are located (designated Regions I through VII; Cann and Wilson 1983). One of these length polymorphisms (LPs) has been of particular importance for tracing population movements through Asia. This 9-base-pair (bp) deletion occurs in the intergenic region between the COII (*MTCOX*2*) and $tRNA^{Lys}$ (*MTTK*) genes, also known as Region V (Cann and Wilson 1983). The Region V deletion was first detected in populations of Asian descent by Cann and Wilson (1983) and Cann et al. (1984), and since this time has been observed in mtDNAs from populations inhabiting East Asia (Horai and Matsunaga 1986; Harihara et al. 1992; Horai et al. 1996; Yao et al. 2000), South Asia (Passarino et al. 1993; Torroni et al. 1994a; Watkins et al. 1999; Schurr et al. 2000), Southeast Asia (Ballinger et al. 1992; Melton et al. 1995, 1998, Redd et al. 1995), Australia and PNG (Wrischnik et al. 1987; Stoneking et al. 1990; Betty et al. 1996), and Polynesia and Melanesia (Hertzberg et al. 1989; Stoneking et al. 1990; Lum et al. 1994; Sykes et al. 1995; Lum and Cann 1998; Richards et al. 1998; Merriwether et al. 1999). This broad distribution of the Region V marker indicates that recent population expansions from Southeast Asia into Polynesia probably have included many deletion mtDNAs from the same extended lineage (Ballinger et al. 1992).

The Region V deletion has also been detected in non-Asian populations (Cann et al. 1987; Chen et al. 1995; Torroni et al. 1995; Soodyall et al. 1996), raising the possibility that all deletion mtDNAs are part of the same mtDNA lineage. However, these non-Asian deletion mtDNAs have been shown to have different haplotypic (mutational) backgrounds than those occurring in Asian populations, suggesting that the Region V deletion has occurred independently in them. Similarly, Ballinger et al. (1992) detected at least two independent occurrences of the Region V deletion in haplotypes that did not belong to haplogroup *B*, a result suggesting that not all deletion mtDNAs present in Asian populations were closely genealogically related. This same pattern has since been seen in aboriginal Australians (Betty et al. 1996), South Asians (Watkins et al. 1999), and Native Americans (Torroni et al. 1993b), implying that the Region V 9-bp deletion occurs relatively frequently in human mtDNAs and that it involves different repeat

units (Lum and Cann 1998; Handoko et al. 2001). However, when contextualized with other genetic information, this LP remains a useful marker with which to trace Asian population affinities.

Unfortunately, it has been difficult to integrate studies of mtDNA variation in Asian populations because the type and level of resolution of the molecular genetic methods used to characterize their haplotypic diversity has varied considerably. As a result, the data from papers using RFLP haplotyping methods have often not been directly comparable to those employing CR sequencing, or to some employing limited RFLP analysis with CR sequencing. This, in turn, has meant that the data from these studies have remained largely uncollated in any significant way, with a few exceptions (e.g., Kivisild et al. 1999; Forster et al. 2001; Huoponen et al. 2001). This study represents an attempt to bridge the gap between RFLP and CR sequencing studies of Asian mtDNA variation, and assess the larger scale patterns of genetic diversity in Southeast Asia that emerge from this approach. As such, it is largely a summary of existing data sets, although it also adds new data from ongoing work to help illuminate these trends.

This paper also focuses on several different mtDNA lineages that are present in Southeast Asia, namely, haplogroups *B, F,* and *M*. These haplogroups are present in Vietnamese and aboriginal groups from Malaysia (Orang Asli) and Borneo to varying degrees, as well as in other Southeast Asian groups. Their distribution in this geographic region allows us to make certain inferences about the genetic relationships of the populations that possess them. Furthermore, these mtDNA lineages appear to demarcate major expansions of Asian peoples, whether the initial colonizers of this region or the most recent expansion of Austronesian speakers.

## Samples and Methods

The data discussed in this paper derive mostly from published sources. The RFLP haplotype data come from Horai et al. (1984), Horai and Matsunaga (1986), Ballinger et al. (1992), Passarino et al. (1993), Torroni et al. (1993a, 1994a), Kolman et al. (1996), and Kivisild et al. (1999). The HVS-I sequence data come mostly from Torroni et al. (1993a), Lum et al. (1994), Melton et al. (1995, 1998), Redd et al. (1995), Horai et al. (1996), Kolman et al. (1996), Lum and Cann (1998), Redd and Stoneking (1999), and Yao et al. (2000). The unpublished HVS-I sequence data for Nepalese groups were first reported in Schurr et al. (2000), while the preliminary HVS-I sequence data for the Bornean, Malaysian, and Vietnamese samples were first reported in Wallace and Schurr (2001). All of these HVS-I sequences were determined by standard methods (Schurr et al. 1999).

Nearly all of these samples share one thing in common: they have been screened for the Region V 9-bp deletion that characterizes haplogroup *B* (Ballinger et al. 1992; Torroni et al. 1992). Among the papers from which deletion frequencies were obtained are Horai and Matsunaga (1986), Cann et al.

(1987), Harihara et al. (1988), Hertzberg et al. (1989), Stoneking et al. (1990), Ballinger et al. (1992), Passarino et al. (1993), Torroni et al. (1993a, 1994a), Lum et al. (1994), Melton et al. (1995, 1998), Redd et al. (1995), Betty et al. (1996), Kolman et al. (1996), and Lum and Cann (1998).

## Results and Discussion

**Haplogroup Nomenclature.** Due to the historical sequence in which major mtDNA clusters were named as haplogroups, there has been confusion regarding the exact phylogenetic relationships of the RFLP clusters named in different papers. In some cases this confusion has arisen because of the different laboratory methods used to characterize human mtDNAs, while in others it has arisen because researchers attempting to follow an established precedent used the same derived nomenclature to classify different clusters. Furthermore, it is now apparent that the variety of already identified haplogroups represent different hierarchical levels of the human mtDNA phylogeny (e.g., Macaulay et al. 1999). Consequently, a clarification of the current system of nomenclature being used to classify mtDNA clusters in world populations will help us understand the nature of haplotypic diversity in Southeast Asian populations.

In Ballinger et al. (1992), a number of clusters were observed in Southeast Asian populations (Table 1, Panel A). At this time, haplogroup *M* was defined as having only the +*Dde*I 10394 site, whereas both haplogroups *E* and *F* had the +*Dde*I 10394 and +*Alu*I 10397 sites (+*Dde*I/+*Alu*I sites), with these latter two differing by the presence or absence of the +*Hae*III 16517 site. The remaining Southeast Asian mtDNAs were also placed in clusters that were given lettered names (*A, C, D, H, J, N-Q, S*, and *T*), including haplogroup *A*, which is now called haplogroup *F* (Torroni et al. 1994a).

In the ensuing revision of this haplogroup nomenclature, a number of these clusters were given new letter designations. For example, haplogroups *K, L,* and *R* in Ballinger et al. (1992) were renamed *G, D,* and *C* in Torroni et al. (1992, 1993a, 1994a) (Table 1, Panel A). In addition, the mtDNAs possessing the +*Dde*I/+*Alu*I sites were placed in haplogroup *M* (Chen et al. 1995). Most of the remaining letters used to classify Southeast Asian mtDNAs by Ballinger et al. (1992) were appropriated and used to define other haplogroups in European, Asian, and African populations (Torroni et al. 1992, 1993a, 1994a, 1994b, 1996; Chen et al. 1996) (Table 1, Panel B). As a result, a number of discrete clusters in Southeast Asian populations were never given any new designation. However, some of these older designations (e.g., *M*) may not have defined a specific haplogroup per se, but instead a general subset of mtDNAs having a particular mutation (+*Dde*I 10394). Since this time, the classification of mtDNA haplogroups in other world populations has more or less followed this revised nomenclature (Table 1, Panel B).

One other point needs to be made here. As shown in several analyses of mtDNA haplotypes (Kivisild et al. 1999; Macaulay et al. 1999; Quintana-Murci

**Table 1.** Mitochondrial DNA Haplogroup Designations

Panel A: Ballinger et al. (1992) Nomenclature

| *Old Haplogroup* | *Haplogroup-identifying RFLPs* | *New Haplogroup* | *References* |
|---|---|---|---|
| *A#* | –*Hinc*II 12406, –*Hpa*I 12406 | *F* | 1 |
| *B#* | –*Hinc*II 7853, +*Dde*I 10394, +*Alu*I 10397 | *TBD* | – |
| *C#* | –*Dde*I 3534, –*Alu*I 3537, +*Dde*I 10394, –*Hin*fI 15234, +*Mbo*I 15235 | *B* | 2 |
| *D#* | +*Hae*III 16517 | *B* | – |
| *E* | +*Dde*I 10394, +*Alu*I 10397, +*Hae*III 16517 | *M* | 1, 3 |
| *F#* | +*Dde*I 10394, +*Alu*I 10397 | *M* | 3 |
| *G* | –*Hin*fI 7598, +*Hin*fI 16389, –*Ava*II 16390, +*Dde*I 10394, +*Alu*I 10397 | *E* | 1 |
| *H* | +*Hae*III 663 (+*Hae*III 16517) | *A* | 2 |
| *I* | –*Mbo*I 951, +*Hae*II 1063, +*Hae*II 9326, +*Hha*I 9329, +*Alu*I 10143, +*Dde*I 10394, +*Alu*I 10397 | *TBD* | – |
| *J* | –*Dde*I 1715, –MspI 4711, +*Dde*I 10394, –*Hin*fI 11403, +*Mbo*I 13180 | *TBD* | – |
| *K* | +*Hae*II 4830, +*Hha*I 4831, +*Dde*I 10394, +*Alu*I 10397 | *G* | 1 |
| *L* | –*Alu*I 5176, +*Dde*I 10394, +*Alu*I 10397 | *D* | 2 |
| *M* | +*Dde*I 10394 | *TBD* | – |
| *N* | –*Hae*III 16517 | *TBD* | – |
| *O* | –*Ava*II 13366, +*Mbo*I 13367, +*Bam*HI 13368 | *TBD* | – |
| *P* | +*Hinc*II 12026, +*Hpa*I 12026 | *TBD* | – |
| *Q* | –*Hinc*II 1004 | *TBD* | – |
| *R* | –*Hinc*II 13259, +*Alu*I 13261, +*Dde*I 10394, +*Alu*I 10397 | *C* | 2 |
| *S* | +*Hinc*II 206, +*Hpa*I 206, +*Alu*I 15606 | *P* | 4 |
| *T* | +*Hin*fI 16389, –*Ava*II 16390, +*Hae*III 16517 | *TBD* | – |

*Note*: # = presence of Region V deletion; TBD = to be determined; because these letters were subsequently used to specify new mtDNA clusters, the remaining groupings of Southeast Asian mtDNAs have not been renamed. References: 1 = Torroni et al. 1994a; 2 = Torroni et al. 1992; 3 = Chen et al. 1995; 4 = Forster et al. 2001.

Panel B: Current mtDNA Haplogroup Nomenclature

| *Haplogroup Designations* | *Region(s) in which Haplogroup Primarily Occurs* | *Reference* |
|---|---|---|
| *A* | Asia, Americas | Torroni et al. 1992 |
| *B* | Asia, Americas | Torroni et al. 1992 |
| *C* | Asia, Americas | Torroni et al. 1992 |
| *D* | Asia, Americas | Torroni et al. 1992 |
| *E* | Asia | Torroni et al. 1994a |
| *F* | Asia | Torroni et al. 1994a |
| *G* | Asia | Torroni et al. 1994a |
| *H* | Europe, Middle East, Asia | Torroni et al. 1994d |
| *I* | Europe | Torroni et al. 1994d |
| *J* | Europe | Torroni et al. 1994d |

**Table 1.** Continued

| *Haplogroup Designations* | *Region(s) in which Haplogroup Primarily Occurs* | *Reference* |
|---|---|---|
| *K* | Europe | Torroni et al. 1994a |
| *L* | Africa | Torroni et al. 1992 |
| *M* | Africa, Asia | Chen et al. 1995 |
| *N* | Eurasia | Forster et al. 2001 |
| *O* | ? | – |
| *P* | PNG | Forster et al. 2001 |
| *Q* | PNG | Forster et al. 2001 |
| *R* | Eurasia | Macaulay et al .1999 |
| *S* | ? | – |
| *T* | Europe, Middle East | Torroni et al. 1996 |
| *U* | Europe, Middle East, Asia, Africa | Torroni et al. 1996 |
| *V* | Europe, Middle East | Torroni et al. 1996 |
| *W* | Europe | Torroni et al. 1996 |
| *X* | Europe, Middle East, Asia | Torroni et al. 1996 |
| *Y* | Asia | Schurr et al. 1999 |
| *Z* | Asia | Schurr et al. 1999 |

et al. 1999), haplogroup *M* is actually a macrohaplogroup, meaning that *M* represents the founding or stem haplogroup from which all subsequent haplogroups bearing the +*Dde*I/+*Alu*I sites evolved. In the same way, haplogroup *L* encompasses all *L1*and *L2* mtDNAs found in African populations (Chen et al. 1995, 2000; Watson et al. 1997). Therefore, any mtDNA with the +*Dde*I/+*Alu*I sites can be said to belong to this macrohaplogroup. Thus, haplogroups *C, D, E, G,* and *Z* (Torroni et al. 1992, 1993a, 1994a; Schurr et al. 1999) can be considered smaller branches of haplogroup *M*. Most of these smaller haplogroups have retained their independent status as mtDNA lineages because the designations for haplogroups *A–L* preceded the naming of macrohaplogroup *M*. However, some researchers have been renaming these smaller haplogroups as variants of *M* itself (*M1–M7*) (e.g., Quintana-Murci et al. 1999; Kivisild et al. 1999; Forster et al. 2001), a trend that may ultimately supplant the other designations.

**Haplogroup *M*.** With this clarification of haplogroup nomenclature, we can now examine the frequency distribution of haplogroup *M* in Southeast Asian populations. Based on recent work, it now appears that this ancient lineage had its origins in east Africa (Passarino et al. 1998; Quintana-Murci et al. 1999; Donham et al. 2000). From there, it was dispersed into East Asia by way of the Indian subcontinent, with a diverse array of haplotypes evolving in South Asia since this time (Passarino et al. 1996; Kivisild et al. 1999). As a consequence of this dispersal pattern, haplogroup *M* has been part of the initial expansions of modern human groups into Southeast Asia, with most of the mtDNAs present in extant Asian populations being divided into two general clusters based on the presence

or absence of the +*DdeI*/+*AluI* sites (Ballinger et al. 1992) and the 16223T mutation in the HVS-I (Macaulay et al. 1999).

Evidence for the antiquity of haplogroup *M* in this region comes from recent estimates of its age in Asia. Whether based on RFLP haplotype or HVS-I sequence data, these estimates indicate that haplogroup *M* arrived in East Asia some 65,000–45,000 years ago (Table 2). Such early dates for the entry of haplogroup *M* into Southeast Asia are not surprising in light of the considerable diversity seen in the mtDNA haplotypes belonging to this large cluster (Schurr et al. 1990; Stoneking et al. 1990; Ballinger et al. 1992; Torroni et al. 1992, 1993a, 1994a; Kolman et al. 1996; Passarino et al. 1996; Kivisild et al. 1999; Forster et al. 2001). In fact, haplogroup *M* occurs in all Southeast Asian populations at varying frequencies (25%–45%), with the highest frequencies occurring in the Malays and Sabah Aborigines (~60%) (Figure 1).

Additional evidence for the antiquity of haplogroup *M* is the number of unnamed and diverse clusters present in Southeast Asian populations (Table 3). These include clusters present in the Orang Asli of Malaysia, the Kadazan of Borneo, and highland PNG populations (Stoneking et al. 1990; Ballinger et al. 1992). As yet, these groupings, as well as a number of other clusters in Asian populations (Torroni et al. 1993a, 1994a; Kivisild et al. 1999), have not been classified using the revised haplogroup nomenclature. Undoubtedly more of them will be detected once regional population samples (e.g., Indonesian and Melanesian) are analyzed for RFLP haplotype variation in addition to HVS-I sequence diversity.

Furthermore, haplogroup *M* haplotypes tend to have a specific distribution in Southeast Asian populations. As noted above, many unique clusters of *M* haplotypes appear in different aboriginal populations from Malaysia (including Borneo) (Table 3), with these populations generally inhabiting the interior areas of the peninsulas and islands of the region. In addition, highland populations of Melanesia (PNG and Solomon Islanders: Stoneking et al. 1990; Ballinger et al. 1992; Friedlaender et al. 1995; Merriwether et al. 1999) show predominantly *M* mtDNAs, with the coastal populations of these areas having mostly non-*M* haplotypes. This distribution suggests that the earliest populations who brought haplogroup *M* mtDNAs to various parts of Southeast Asia and Melanesia initially settled the interior portion of these areas, while the later-arriving Austronesian populations, who possessed largely non-*M* haplotypes, mostly settled the coastal areas and/or pushed existing populations further into the interior.

**Table 2.** Age of Haplogroup *M* in Africa and Asia

| *Population* | *HVS-I Data* | *RFLP Data* | *Reference* |
|---|---|---|---|
| East Africans | 36,000 ± 11,000 | 48,000 ± 15,000 | Quintana-Murci et al. 1999 |
| Indians | 53,000 ± 7000 | 56,000 ± 7000 | Quintana-Murci et al. 1999 |
| Mongolians | 67,000 ± 5000 | – | Kolman et al. 1996 |
| SE Asians + Tibetans | – | 77,182 – 55,517* | Chen et al. 1995 |

*Note:* * = based on a 2.2%–2.9% per myr mtDNA evolutionary rate (Torroni et al. 1994).

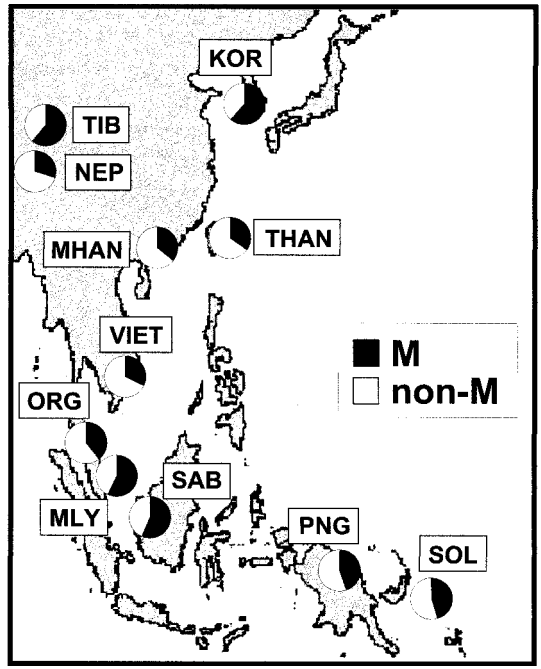


**Figure 1.** Frequencies of haplogroup *M* haplotypes in East and Southeast Asian populations. Data are taken from Stoneking et al. (1990), Ballinger et al. (1992), Torroni et al. (1994a), Merriwether et al. (1999), and Schurr et al. (2000).

**Table 3.** Unique mtDNA Clusters in Southeast Asian Populations

| *Population* | *Haplogroup* | *Polymorphic Restriction Sites* |
|---|---|---|
| Semai (Malaysia) | *M** | –*Mbo*I 951, +*Hae*III 1063, +*Hae*II 9326, +*Hha*I 9327, +*Alu*I 10143, +*Dde*I 10394, +*Alu*I 10397 |
| Kadazan (Borneo) | *M** | –*Hae*III 4563, +*Hin*fI 10054, +*Dde*I 10394, +*Alu*I 10397 |
| Kadazan (Borneo) | ? | –*Mbo*I 951, –*Mbo*I 7859, +*Alu*I 8484, +*Hae*II 11001, +*Hha*I 11002 |
| Jeni (Borneo) | ? | –*Dde*I 1715, –*Hpa*II 4711, +*Dde*I 10394, –*Hin*fI 11403, +*Mbo*I 12528, +*Mbo*I 13180 |
| Highland PNG | *P* | +*Hpa*I 207, +*Hinc*II 207, +*Alu*I 15606 |
| Highland PNG | *Q* | +*Dde*I 10394, +*Alu*I 10397, +*Taq*I 16178 |

*Note:* The polymorphic restriction sites defining these clusters are reckoned as site gains (+) or losses (–) relative to the Cambridge Reference Sequence (CRS) (Anderson et al. 1981), as corrected by sequencing studies (Andrews et al. 1999). The asterisk (*) means that these clusters belong to the haplogroup *M* but have not been given further taxonomic identifiers.

**Haplogroup *F*.** Another mtDNA lineage that is fairly widespread throughout Southeast Asia is haplogroup *F*. Haplogroup *F* mtDNAs are defined by the associated –*Hin*cII 12406 and –*Hpa*I 12406 polymorphisms, with many of these also having the associated –*Hae*II 9052 and –*Hha*I 9053 mutations (Ballinger et al. 1992; Torroni et al. 1993a, 1994a). Judging from initial HVS-I sequence analysis of Southeast Asian RFLP haplotypes (AS33, AS39, AS71, AS91, AS98, and AS99 from Ballinger et al. 1992), haplogroup *F* mtDNAs lack the 16223 C→T mutation seen in haplogroup *M* mtDNAs and possess the 16304 T→C transition. Based on this information, it is possible to examine the polymorphic nucleotide composition of published HVS-I sequences from Asian populations that have not been subjected to high-resolution RFLP analysis and infer the frequencies of this mtDNA lineage in those populations. Accordingly, the haplogroup frequencies shown in Figure 2 are taken from RFLP haplotype data (Horai et al. 1984; Ballinger et al. 1992; Harihara et al. 1992; Torroni et al. 1993a, 1994a; Ivanova et al. 1999; Schurr et al. 2000) or inferred haplogroup status based on HVS-I sequences (Horai et al. 1996; Kolman et al. 1996). In this regard, it should be pointed out that the haplogroup *F* denoted in Kolman et al. (1996) for Mongolians is not analogous with that defined in Tibetans by Torroni et al. (1994a), since the Mongolian samples were not screened for the –*Hin*cII 12406 and –*Hpa*I 12406 polymorphisms.

Two other caveats should also be made here. First, it is known that the –*Hin*cII 12406 and –*Hpa*I 12406 polymorphisms occasionally occur through mutational events that are independent of the one that gave rise to haplogroup *F* (Ballinger et al. 1992; Torroni et al. 1993b, 1994b, 1996; Chen et al. 1995). However, because of their relative rarity, and because of the clear genealogical relatedness of the mutations appearing in Southeast Asian mtDNAs (Ballinger et al. 1992), these occurrences should not cause an overestimation of the frequency of haplogroup *F* in Asia. The second caveat is that the 16304C mutation sometimes occurs in mtDNAs having different haplotypic backgrounds (e.g., Watson et al. 1997; Chen et al. 2000). However, the 16304 site is not known to be hypervariable (e.g., Gurven 2000; Stoneking 2000), meaning the problem of recurring mutations can essentially be ignored when making estimates of haplogroup *F* frequencies.

Having said this, let us return our attention to Figure 2. As seen in this figure, haplogroup *F* appears in a number of Asian populations, including Filipinos (Cann et al. 1987) and Aboriginal Taiwanese (Melton et al. 1995, 1998), and occurs at its highest frequencies in Southeast Asia, specifically in the Vietnamese (Ballinger et al. 1992). Relatively high frequencies of this haplogroup also occur in the Orang Asli of Malaysia, to whom the Vietnamese show linguistic ties (Austro-Asiatic family; Bellwood 1979). The frequency distribution of this mtDNA lineage shows that it decreases in both northerly and southeasterly directions from Southeast Asia, but is widespread in East Asia itself, being seen as far north as central Siberia (Evenks) (Torroni et al. 1993a) and as far south as Borneo (Kadazan). Because many island Southeast Asian and Melanesian populations

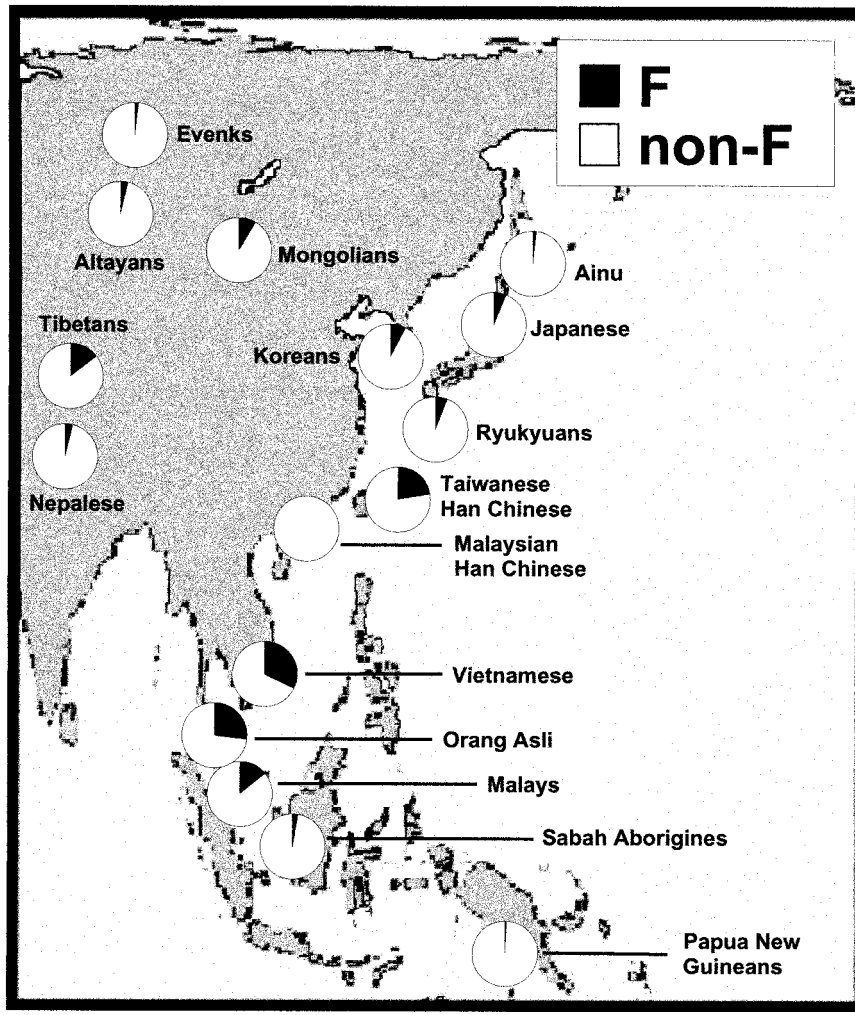


**Figure 2.** Frequencies of haplogroup *F* in East Asia. Data are taken from Horai et al. (1984), Ballinger et al. (1992), Harihara et al. (1992), Torroni et al. (1993a, 1994a), Horai et al. (1996), Kolman et al. (1996), Ivanova et al. (1999), and Schurr et al. (2000).

have not been typed for RFLP markers, it is unclear whether this haplogroup is present in these populations. However, its detection in coastal PNG groups (Stoneking et al. 1990) suggests that this mtDNA lineage could be present in other Melanesian populations.

Based on its distribution, haplogroup *F* may have been disseminated throughout East Asia through some kind of a population expansion. It does not appear to be as diverse a lineage as haplogroup *M*, based on the RFLP composi-

tion of its haplotypes (Horai et al. 1984; Harihara et al. 1988; Ballinger et al. 1992; Torroni et al. 1993a, 1994a), and, in fact, is found within the geographic range of haplogroup *M*. In addition, while the distribution of haplogroup *F* overlaps that of haplogroup *B*, its greatest haplotypic diversity occurs in Vietnam, not in Taiwan, the Philippines, and Indonesia, as is the case for haplogroup *B* (Melton et al. 1995, 1998; Redd et al. 1995; Richards et al. 1998). Thus, the spread of haplogroup *F* in East Asia does not appear to be directly related to the expansion of Austronesian speakers into Indonesia, Melanesia, and Polynesia, although some of its haplotypes may have been dispersed through this event. An alternative possibility is that haplogroup *F* was spread throughout East Asia during the expansion of Sino-Tibetan languages some 8000–6000 YBP, most of which are now found in Southeast Asia and China. However, this hypothesis will have to be tested with additional data from various populations speaking Sino-Tibetan languages.

**Haplogroup *B*.** As shown by various studies, deletion mtDNAs from haplogroup *B* are broadly distributed in Asian populations. Recent population expansions associated with the spread of Austronesian languages appear to have brought these haplotypes from Southeast Asia into Polynesia around 5000–1000 YBP. Many Polynesian haplogroup *B* mtDNAs also possess a set of polymorphisms that distinguishes them from similar types in other Asian populations (Hagelberg and Clegg 1993; Hagelberg et al. 1994; Lum et al. 1994; Melton et al. 1995, 1998; Redd et al. 1995; Sykes et al. 1995; Lum and Cann 1998; Richards et al. 1998). This set of mutations, which includes the 16217 T→C, 16247 A→G, and 16261 C→T transitions (or "CGT"), has been called the "Polynesian Motif" because of its uniqueness to Polynesian and related populations (Hagelberg and Clegg 1993; Hagelberg et al. 1994; Lum et al. 1994). The Polynesian Motif evolved from mtDNAs bearing the ancestral 16189C and 16217C sequence motif through a series of mutational steps (CAC→CAT→CGT) that probably began in Taiwan and continued as populations spread south into the Philippines, Indonesia, and Melanesia (Melton et al. 1995, 1998; Redd et al. 1995), although some favor an Indonesian source for this lineage (Richards et al. 1998).

Haplogroup *B* mtDNAs are also found in Vietnamese, Malaysian, and Bornean populations (Ballinger et al. 1992). However, none of the HVS-I sequences identified in these deletion haplotypes show the Polynesian Motif (CGT), and only one (Sabah Aborigine, SA26; Ballinger et al. 1992) has the hypothesized intermediate state, CAT. These findings are generally concordant with those of Melton et al. (1995, 1998), who observed the Polynesian Motif at the highest frequency in Polynesians and coastal Papua New Guineans, and at modest frequencies in East Indonesians and Malays, with the intermediate form (CAT) occurring at the highest frequencies in Taiwanese aborigines and moderate frequencies in Filipinos and east Indonesians.

Ballinger et al. (1992) also suggested that two deletion haplogroups were present in Asian populations, with these being called *C** and *D**. Haplogroup *D**

has since been renamed haplogroup *B* (Torroni et al. 1992), but *C** has received little attention since its first description. So as not to confuse it with haplogroup *C*, which is not closely related to either deletion lineage, haplogroup *C** will henceforth be called haplogroup *B**, since this letter designation will associate it with the other major deletion haplogroup. With respect to their mutational composition, haplogroup *B* is characterized by the Region V 9-bp deletion and the +*Hae*III 16517 mutation, while, in addition to these polymorphisms, haplogroup *B** also has the –*Dde*I 3534, –*Alu*I 3537, +*Dde*I 10394, –*Hin*fI 15234; and +*Mbo*I 15235 mutations (Ballinger et al. 1992; Passarino et al. 1993) (Table 4).

When subjected to phylogenetic analysis, the haplotypes from both of these haplogroups clustered together but formed separate branches (Figure 2 in Ballinger et al. 1992). However, this association occurred in part because of the inclusion of the 9-bp deletion as an RFLP character in the haplotype data matrix. If this marker is eliminated from the data set, then the two branches are positioned in separate portions of the Asian mtDNA tree because of the presence of the +*Dde*I 10394 site in *B** mtDNAs (data not shown). Regardless of their positions, however, haplogroup *B** had long branches connecting mutationally diverse haplotypes, whereas haplogroup *B* had shorter and shallower branches of haplotypes. These data suggested that *B** could be an older deletion lineage, or else a highly divergent subbranch of haplogroup *B*. By contrast, the shallower branches of shorter lengths for haplogroup *B* suggested that it could be a younger deletion lineage that had rapidly diversified and been spread in East Asia in relatively recent prehistory.

However, both of these putative deletion lineages are approximately the same age. Based on the maximum likelihood estimates of haplogroup diversity presented in Ballinger et al. (1992), haplogroup *B* arose between 33,500–16,750 YBP, while haplogroup *B** emerged between 44,500–22,250 YBP (Table 5). These age estimates are supported by a recent median network analysis of RFLP and HVS-I sequence data from haplogroup *B* in which its expansion in Asia was calculated to be 29,100 ± 7100 YBP (Forster et al. 2001). Such findings suggested that haplogroup *B* could possibly have evolved from haplogroup *B**, or, alternatively, that both haplogroups represented independent occurrences of the Region V 9-bp deletion in different mtDNA lineages that were subsequently

**Table 4.** Deletion Haplotypes in Southeast Asia

| *Haplogroup* | *RFLP Haplotype* | *HVS-I Sequence* |
|---|---|---|
| *B* | +*Hae*III 16517, 9-bp deletion | 16189C, 16217T, 16519C |
| *B** | –*Dde*I 3534, –*Alu*I 3537, +*Dde*I 10394, –*Hin*fI 15234, +*Mbo*I 15235, +*Hae*III 16517, 9-bp deletion | 16140C, 16189C, 16519C |

*Note:* Haplogroup *B** haplotypes have been detected in Nepalese (Passarino et al. 1993; Schurr et al. 2000), Taiwanese Han, Koreans, Kadazan [Borneo], and Vietnamese (Ballinger et al. 1992) populations through RFLP analyses.

**Table 5.** Age of mtDNA Deletion Lineages in Asia

| *Haplogroup* | *Age or Divergence Time (YBP)* | *References* |
|---|---|---|
| *B* | 33,500–16,750# | Ballinger et al. 1992 |
| | 40,455–30,690^ | Ballinger et al. 1992 |
| *B** | 44,500–22,500# | Ballinger et al. 1992 |
| | 30,455–23,105^ | Ballinger et al. 1992 |
| *B* | 30,000 | Lum et al. 1994 |
| *B* | 55,000 | Melton et al. 1998 |
| *B* | 58,000 | Redd et al. (1995) |
| *B* | 72,000 | Soodyall et al. (1996) |

*Note:* Ballinger et al. (1992) estimated the maximum likelihood (ML) divergences of haplogroups *B* (0.067%) and *B** (0.089%) in Southeast Asia, but not their ages. However, they did estimate the age of certain populations using intrapopulational ML diversity values and an mtDNA mutation rate of 2.0%–4.0% per myr (#). The ages of *B* and *B** shown here are thus estimated with this mutation rate. When a 2.2%–2.9% per myr mtDNA evolution rate (Torroni et al. 1994) is used, as indicated by the caret (^), the time depths of these mtDNA clusters increase.

distributed in Asia at similar times. To discriminate between these possibilities, it will be necessary to determine whether the +*Dde*I 10394 site present in *B** haplotypes is the same phylogenetically ancient site that is present in African mtDNAs, or, instead, a secondary occurrence of this polymorphism in mtDNAs that originally lacked the +*Dde*I/+*Alu*I sites.

Surprisingly, these RFLP-based age estimates for haplogroups *B* and *B** are largely inconsistent with comparable estimates based on HVS-I sequence data. The only HVS-I estimate that is similar to those obtained from RFLP data (30,000 YBP) comes from the Polynesian data of Lum et al. (1994) (Table 5). Otherwise, the other HVS-I estimates suggest that haplogroup *B* arose much earlier than indicated by RFLP data (~62,000 YBP) (Redd et al. 1995; Soodyall et al. 1996; Melton et al. 1998). While these early dates may simply be measuring the coalescence date of all Asian haplotypes bearing the Region V deletion, they may also be overestimates of the lineal age of haplogroup *B*. This possibility is suggested by the fact that Redd et al. (1995) estimated the expansion of Austronesian speakers from insular Southeast Asia into the Pacific at around ~12,000 YBP, a date that is twice as old as any archeological dates for this expansion.

To glean additional insights into the origins of haplogroup *B**, we sequenced the HVS-I region of some of the Southeast Asian mtDNAs (AS55, AS59, and AS60 from Ballinger et al. 1992) and two Nepalese mtDNAs (Schurr et al. 2000) that could be placed in this haplogroup on the basis of their RFLP composition, and compared the resulting data with HVS-I sequences from haplogroup *B* mtDNAs. We found that all of these *B** mtDNAs possessed the 16140 T→C mutation in addition to other polymorphisms typically found in haplogroup *B* mtDNAs. Using this information, we scanned the published HVS-I sequences from Asian populations to find other deletion haplotypes that had the 16140C mu-

tation. Those that possessed this mutation were classified as haplogroup *B** mtDNAs, and those that did not were placed in haplogroup *B*. This partitioning of deletion mtDNAs into two haplogroups then allowed us to approximate their frequency in other Asian populations.

Interestingly, we found that quite a few deletion mtDNAs from various Southeast and East Asian populations had the 16140C mutation. Haplogroup *B** mtDNAs occurred primarily in East and Southeast Asians, whereas haplogroup *B* mtDNAs had a much wider distribution in Asia and the Pacific (Figure 3). In addition, both *B* and *B** have been observed in Aboriginal Taiwanese and Taiwanese Han populations (Ballinger et al. 1992; Melton et al. 1995, 1998). Based on both RFLP and HVS-I sequence data (Passarino et al. 1993; Torroni et al. 1994a; Kolman et al. 1996; Schurr et al. 2000), we have detected haplogroup *B** mtDNAs in Nepalese and Mongolian populations, but, surprisingly, not in Tibetans. In the case of Mongolians, two of their deletion haplotypes (#8.12 and #17.25) were included in *B** because of having the 16140C mutation and the +*Dde*I 10394 site, although they were not screened for the other RFLPs that define this putative haplogroup (Kolman et al. 1996).

Both types of deletion mtDNAs also appeared in Chinese populations (Yao et al. 2000) (Figure 4). Interestingly, the majority of the haplogroup *B** mtDNAs were found in Chinese minority populations, such as the Tu and Miao, but were also present in several ethnic Han Chinese populations. In fact, the HVS-I sequences having the 16140C mutation formed a large cluster that was distinctive from the one containing typical haplogroup *B* sequences (Yao et al. 2000) (Figure 5). This distribution suggested that, whether a separate lineage or a subbranch of haplogroup *B*, haplogroup *B** mtDNAs were dispersed in a somewhat different manner than haplogroup *B* itself. Further examination of deletion mtDNAs throughout East Asia should help to clarify both the phylogenetic relationships between *B* and *B** mtDNAs, and their dispersal patterns in this region.

**Mitochondrial DNA Diversity in Southeast Asia.** This general overview of the pattern of mtDNA diversity in Southeast Asia allows us to make several inferences about its population history. First, several major mtDNAs are shared across large parts of this region. In total, haplogroups *B, F,* and *M* constitute the majority of all mtDNAs in Southeast Asian groups (Table 6). Their distribution in Southeast Asia, and the pattern of haplotype sharing by populations that inhabit this region, probably reflects the process by which it was settled. One of the first haplogroups to enter Southeast Asia was clearly (macro-) haplogroup *M*. Its antiquity in this region is evident by its estimated age (Table 2) and the many unique clusters of derived haplotypes that are present in Southeast Asian, PNG, and Melanesian groups (Tables 3 and 7), particularly in ethnic groups representing some of the first populations to enter this region (e.g., PNG highland populations). Haplogroups *F* and *B* are also present throughout Southeast Asia, but are not as genetically diverse as haplogroup *M*; thus, they must have arisen after the latter mtDNA lineage was brought into this region and began diversifying. Both of the former haplogroups appear to be associated with major population expan-

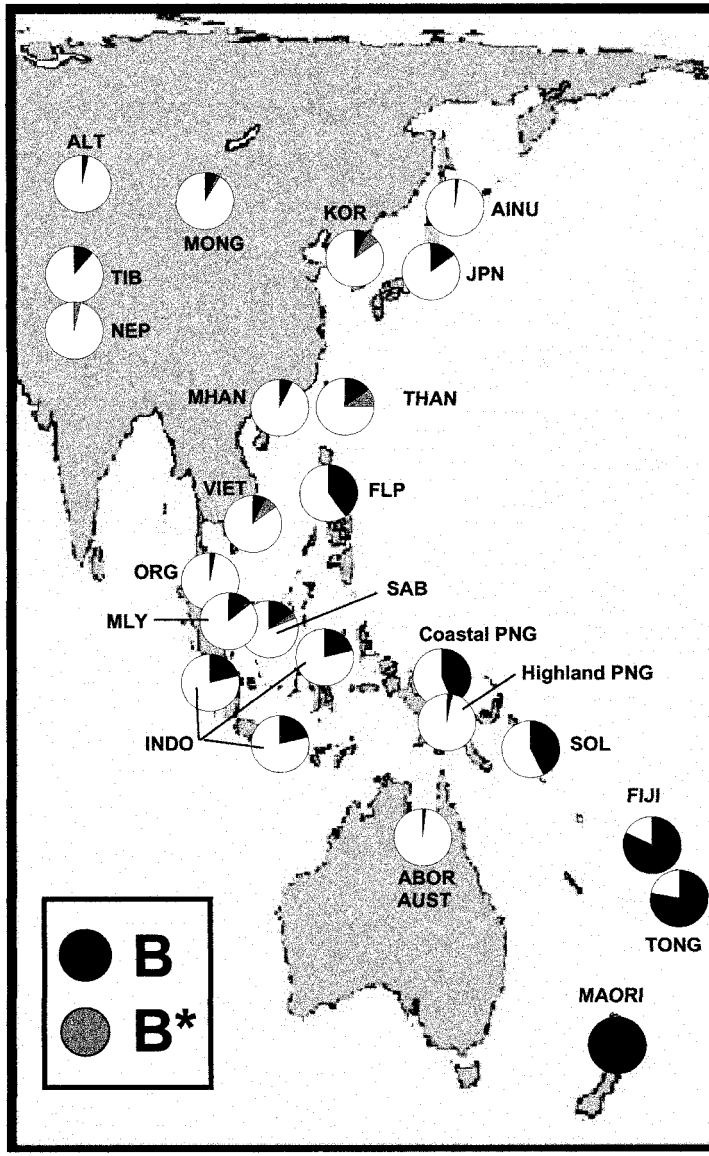


**Figure 3.** Frequencies of haplogroups *B/B** in East Asia. Data are taken from Hertzberg et al. (1989), Stoneking et al. (1990), Ballinger et al. (1992), Harihara et al. (1992), Torroni et al. (1994a), Melton et al. (1995), Betty et al. (1996), Horai and Matsunaga (1996), Kolman et al. (1996), Sukernik et al. (1996) Merriwether et al. (1999), and Schurr et al. (2000). Population abbreviations are as follows: ALT = Altayans; MONG = Mongolians; TIB = Tibetans; NEP = Nepalese; KOR = Koreans; JPN = Japanese; THAN = Taiwanese Han; MHAN = Malaysian Han; VIET = Vietnamese; FLP = Filipinos; ORG = Orang Asli; MLY = Malays; SAB = Sabah Aborigines; INDO = Indonesians; PNG = Papua New Guinea; SOL = Solomon Islanders; ABOR AUST = Aboriginal Australians; FIJI = Fijians; and TONG = Tongans.

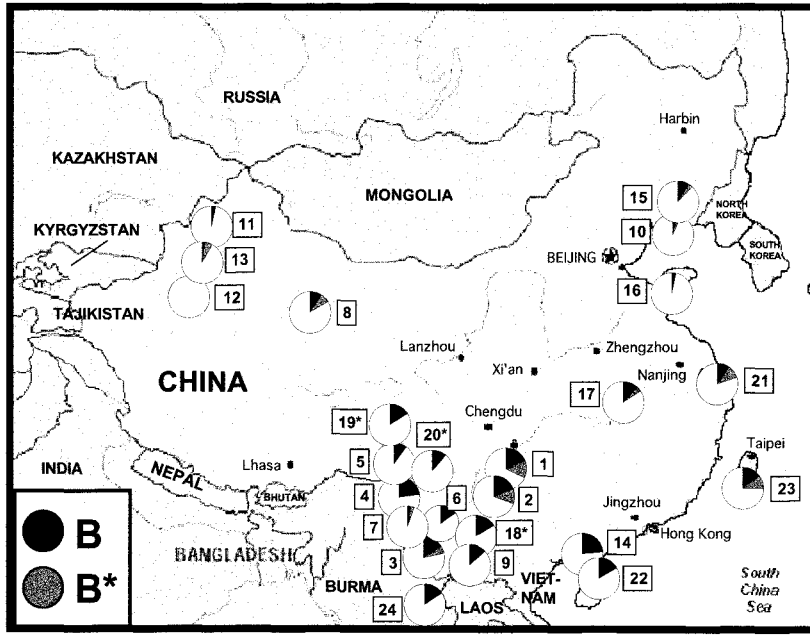


**Figure 4.** Distribution of haplogroups *B/B** mtDNAs in China. The HVS-I sequence data are taken from Yao et al. (2000). The population numbers are identical with those shown in Table 1 of Yao et al. (2000): 1 = Miao; 2 = Buyi; 3 = Dai; 4 = Naxi; 5 = Nu; 6 = Sali; 7 = Wa; 8 = Tu; 9 = Zhuang; 10 = Man; 11 = Uighur (Wei); 12 = Kazakh; 13 = Xinjiang Han; 14 = Guangdong Han; 15 = Liaoning Han; 16 = Qingdao Han; 17 = Wuhan Han; 18 = Yunnan Han; 19 = Sichuan Han; 20 = Guizhou Han; 21 = Shanghai Han; 22 = Cantonese; 23 = Taiwanese Han; and 24 = Thai. The asterisk (*) for populations 18-20 indicates that too few mtDNAs were sequenced to be able to ascertain the frequencies of *B* and *B** mtDNAs in them. As a result, only total deletion haplotype frequencies are presented for these groups.

sions in Southeast Asia, although probably from different source areas and at different times.

From a population standpoint, we can see these larger trends as well as more recent contacts between populations living in this region. Peninsular Malaysian populations show genetic affinities with populations from Borneo, while the Vietnamese also show affinities with populations from Malaysia (Table 7) (Ballinger et al. 1992). In addition, the Orang Asli, Malays, and Bornean Aborigines exhibited some similarities to highland PNG populations, with insular Southeast Asians appearing genetically closer to the coastal PNG populations (Ballinger et al. 1992). These data, along with the distribution of haplogroup *B* in the region, suggest that the expansion of Austronesian speakers into the Pacific Basin genetically influenced populations from Malaysia and Indonesia, as well as Melanesia, an interpretation also supported by HVS-I sequence data (Stoneking et al. 1990; Ballinger et al. 1992; Lum et al. 1994, 1996; Sykes et al. 1995; Redd et al. 1995; Melton et al. 1995, 1998; Richards et al. 1998; Merriwether et al. 1999).

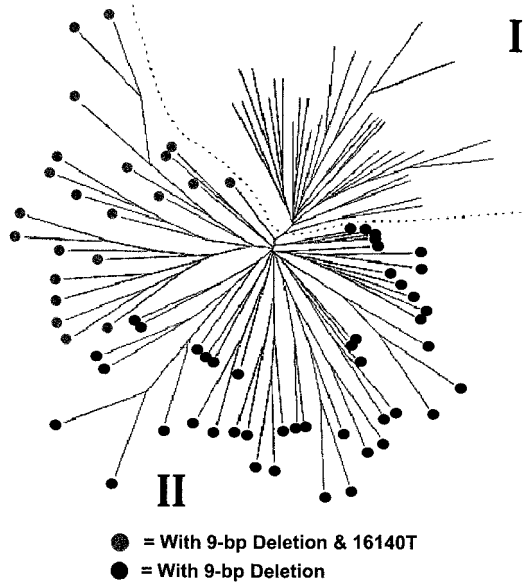


**Figure 5.** Phylogeny of HVS-I sequences from Chinese deletion mtDNAs (modified after Figure 3 from Yao et al. 2000). Haplogroups B and B* are indicated in the tree by the black and gray circles, respectively.

**Table 6.** Major Haplogroup Frequencies in Southeast Asian Populations

| *Population* | *n* | *B/B** | *F* | *M* | *Other* |
|---|---|---|---|---|---|
| THAN | 20 | 7.1 | 22.2 | 35.0 | 35.7 |
| MHAN | 14 | 32.0 | – | 35.7 | 32.3 |
| VIET | 28 | 17.9 | 32.1 | 32.1 | 17.9 |
| ORG | 31 | 3.1 | 27.3 | 39.4 | 30.2 |
| MLY | 14 | 14.3 | 14.3 | 57.1 | 14.3 |
| SAB | 32 | 18.6 | 2.9 | 56.3 | 22.2 |
| PNG | 119 | 41.8 | 1.0 | 44.5 | 12.7 |
| KOR | 13 | 15.4 | 7.7 | 61.5 | 15.4 |
| TIB | 54 | 11.1 | 14.8 | 61.1 | 13.0 |
| NEP | 50 | 4.0 | 4.0 | 30.0 | 62.0 |

*Note:* "Other" encompasses all mtDNAs that do not fall into haplogroups *B, F,* and *M*. Population abbreviations are as follows: THAN = Taiwanese Han; MHAN = Malaysian Han; VIET = Vietnamese; ORG = Orang Asli; MLY = Malays; SAB = Sabah Aborigines; PNG = Papua New Guinea; KOR = Koreans; TIB = Tibetans; and NEP = Nepalese.

To supplement this qualitative appraisal of genetic diversity in Southeast Asia, we re-examined the genetic distances estimated from high-resolution RFLP haplotypes for Southeast Asian populations (Ballinger et al. 1992). These distance measures were used to create a population tree of Southeast Asian populations to determine their relative similarities to each other (Figure 6). The resulting

**Table 7.** Genetic Links Between Southeast Asian Populations

| *Haplogroup* | *Polymorphisms* | *Populations* |
|---|---|---|
| *B* | Region V 9-bp deletion | Vietnamese, Orang Asli, Malays, Bornean Aborigines, Coastal PNG, Indonesians |
| *F* | –*Hpa*I 12104, –*Hin*cII 12104 | Vietnamese, Orang Asli, Malays, Bornean Aborigines, Coastal PNG |
| *P*^ | +*Hpa*I 207, +*Hin*cII 207, +*Alu*I 15606 | Malays, Coastal and Highland PNG |
| *E** | –*Hha*I 7598, –*Hin*fI 10830, +*Hin*fI 16389/–*Ava*II 16390 | Malays, Bornean Aborigines, Coastal PNG |
| *M** | –*Hin*cII 7853, +*Hha*I 5351, +*Hin*fI 9820 | Vietnamese, Bornean Aborigines |

^Placed in haplogroup *S* by Ballinger et al. (1992), but now called *P* by Forster et al. (2001).
*All *M* haplotypes have the +*Dde*I 10394 and +*Alu*I 10397 sites.

tree reveals that Malays and Sabah (Borneo) Aborigines are closer to each other than to Malaysian Chinese and Vietnamese, with the Orang Asli being somewhat intermediate between the Malaysian and Vietnamese populations. These associations are consistent with the pattern of haplotype sharing amongst these populations (Ballinger et al. 1992) and with other linguistic and cultural data (e.g., Iskandar 1976; Bellwood 1979). Given these associations, one might predict that Indonesian populations will cluster more closely with peninsular Malaysian and Bornean populations, and that Indonesian groups will also show some affinities with PNG populations.

Recent data have also illuminated the relationships between PNG and East Asian populations. Most PNG mtDNAs belong to one of three major clusters. The first cluster, identified by Stoneking et al. (1990), which was called haplogroup *S* by Ballinger et al. (1992) and has recently been renamed haplogroup *P* by Forster et al. (2001) (Table 7), represents at least a third of all highland PNG haplotypes, but also occurs in both coastal PNG and Malay populations. We now know that haplogroup *P* is synonymous with the HVS-I sequence cluster PNG1 of Redd and Stoneking (1999), because the Malay mtDNA belonging to this haplogroup (AS90; Ballinger et al. 1992) shares the 16357C mutation with PNG1 mtDNAs.

The second cluster is defined by the +*Taq*I 16178 and +*Dde*I/+*Alu*I sites, and is now called haplogroup *Q* (Forster et al. 2001) (Table 2). It constitutes the largest proportion of highland PNG mtDNAs, while also appearing at low frequencies in coastal PNG populations. This haplogroup is identical with HVS-I cluster PNG2 of Redd et al. (1999), and possesses a unique combination of HVS-I mutations (16129A-16144C-16148T-16223T-16241G-16265C-16311C-16343G) that is not seen in other Asian haplogroups.

The third RFLP cluster consists of deletion mtDNAs from haplogroup *B*, which are seen nearly exclusively in coastal PNG populations (Stoneking et al. 1990; Redd and Stoneking 1999). In addition, as mentioned previously, haplogroup *F* is observed in PNG populations at very low frequencies, while hap-

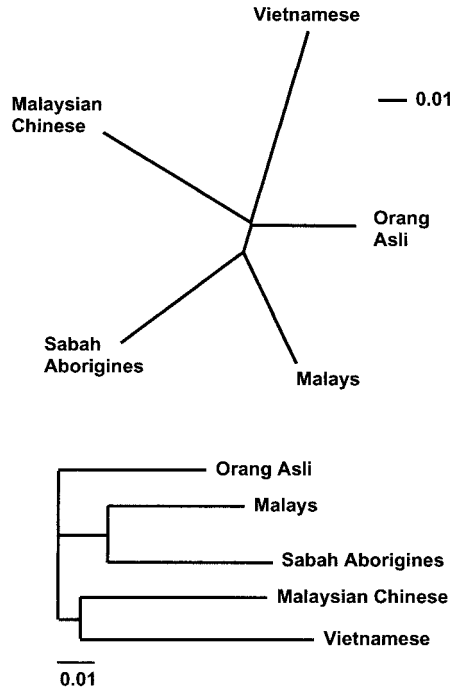


**Figure 6.** A Southeast Asian population tree. The genetic distances between populations are based on the maximum likelihood estimates reported in Table 3 of Ballinger et al. (1992). The Neighbor program in PHYLIP 3.573c (Felsenstein 1998) was used to produce the neighbor-joining tree (Saitou and Nei 1987), which was drawn in both unrooted and cladogram format by TreeView (Page 1996).

logroup *E* mtDNAs defined by the +*Hin*fI 7598 site (Torroni et al. 1994a) also seem to be present (Table 7). Thus, PNG populations are characterized by both a set of largely unique mtDNA haplogroups and others originating in Southeast Asia.

These data, as well as the comparisons of PNG and aboriginal Australian mtDNA haplotypes (Cann et al. 1987; Stoneking et al. 1990; Ballinger et al. 1992; van Holst Pellekaan et al. 1998; Redd and Stoneking 1999; Huoponen et al. 2001), have demonstrated that PNG and Australian populations are genetically distinct from each other. Consequently, the original colonists of these two geographic areas appear to have arisen from different ancient migrations into the Sahul region.

*Acknowledgments* This work was supported by The National Science Foundation through grant NSF 9414900 to T.G.S., and by the National Institutes of Health through grants NS21328, HL45572, and AG13154 to D.C.W.

*Received 19 October 2001; revision received 1 February 2002.*

## Literature Cited

Anderson, S., A.T. Bankier, B.G. Barrell et al. 1981. Sequence and organization of the human mitochondrial genome. *Nature* 290:457–465.

Andrews, R.M., I. Kubacka, P.F. Chinnery et al. 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.* 23:147.

Ballinger, S.W., T.G. Schurr, A. Torroni et al. 1992. Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient Mongoloid migrations. *Genetics* 130:139–152.

Bellwood, P. 1979. *Man's Conquest of the Pacific: The Prehistory of Southeast Asia and Oceania.* Oxford, UK: Oxford University Press.

Betty, D.J., A.N. Chin-Atkins, L. Croft et al. 1996. Multiple independent origins of the COII/tRNA$^{Lys}$ intergenic 9-bp deletion in aboriginal Australians. *Am. J. Hum. Genet.* 58:428–433.

Cann R.L., W.M. Brown, and A.C. Wilson. 1984. Polymorphic sites and the mechanism of evolution in human mitochondrial DNA. *Genetics* 106:479–499.

Cann, R.L., M. Stoneking, and A.C. Wilson. 1987. Mitochondrial DNA and human evolution. *Nature* 325:31–36.

Cann, R.L., and A.C. Wilson. 1983. Length mutations in human mitochondrial DNA. *Genetics* 104:699–711.

Chen, Y.-S., A. Olckers, T.G. Schurr et al. 2000. Mitochondrial DNA variation in the South African Kung and Khwe and their genetic relationships to other African populations. *Am. J. Hum. Genet.* 66:1362–1383.

Chen, Y.-S., A. Torroni, L. Excoffier et al. 1995. Analysis of mtDNA variation in African populations identifies the most ancient of all human continent-specific haplogroups. *Am. J. Hum. Genet.* 57:133–149.

Donham, B.P., T.G. Schurr, D.L. Donham et al. 2000. Mitochondrial DNA variation in Ethiopian populations: Evidence of regional diversification of East Africa (abstract). *Am. J. Phys. Anthropol. Suppl.* 30:141–142.

Felsenstein, J. 1998. *PHYLIP (Phylogeny Inference Package), version 3.573c.* Distributed by author, University of Washington, Seattle, WA.

Forster, P., A. Torroni, C. Renfrew et al. 2001. Phylogenetic star contraction applied to Asian and Papuan mtDNA evolution. *Mol. Biol. Evol.* 18:1864–1881.

Friedlaender, J., D.A. Merriwether, and M.I. Kamboh. 1995. New indications of Bougainville genetic distinctiveness (abstract). *Am. J. Phys. Anthropol. Suppl.* 20:93.

Gurven, M. 2000. How can we distinguish between mutational "hot spots" and "old sites" in human mtDNA samples? *Hum. Biol.* 72(3):455–471.

Hagelberg, E., and J.B. Clegg. 1993. Genetic polymorphisms in prehistoric Pacific islanders determined by analysis of ancient bone DNA. *Proc. R. Soc. Lond. B.* 252:163–170.

Hagelberg, E., S. Quevedo, E. Turbon et al. 1994. DNA from ancient Easter Islanders. *Nature* 369:25–26.

Handoko, H.Y., J.K. Lum, Gustiani et al. 2001. Length variations in the COII-tRNA(Lys) intergenic region of mitochondrial DNA in Indonesian populations. *Hum. Biol.* 73:205–223.

Harihara, S., M. Hirai, Y. Suutou et al. 1992. Frequency of a 9-bp deletion in the mitochondrial DNA among Asian populations. *Hum. Biol.* 64:161–166.

Harihara, S., N. Saitou, M. Hirai et al. 1988. Mitochondrial DNA polymorphism among five Asian populations. *Am. J. Hum. Genet.* 43:134–143.

Hertzberg, M., K.N.P. Mickleson, S. Serjeantson et al. 1989. An Asian-specific 9-bp deletion of mitochondrial in frequently found in Polynesians. *Am. J. Hum. Genet.* 44:504–510.

Horai, S., T. Gojobori, and E. Matsunaga. 1984. Mitochondrial DNA polymorphism in Japanese. I. Analysis with restriction enzymes of six base pair recognition *Hum. Genet.* 68:324–332.

Horai, S., and E. Matsunaga. 1986. Mitochondrial DNA polymorphism in Japanese. II. Analysis with restriction enzymes of four or five base pair recognition. *Hum. Genet.* 72:105–117.

Horai, S., K. Murayama, K. Hayasaka et al. 1996. mtDNA polymorphism in East Asian populations, with special reference to the peopling of Japan. *Am. J. Hum. Genet.* 59:579–590.

Huoponen, K., T.G. Schurr, Y.-S. Chen et al. 2001. Mitochondrial DNA variation in an Australian Aboriginal tribe: Evidence for genetic isolation and regional differentiation. *Hum. Immunol.* 62:954–969.

Iskandar, C. 1976. *Orang Asli: The Aboriginal Tribes of Peninsular Malaysia*. Oxford, UK: Oxford University Press.

Ivanova, R., A. Astrinidis, V. Lepage et al. 1999. Mitochondrial DNA polymorphism in the Vietnamese population. *Eur. J. Immunogenet.* 26(6):417–422.

Kivisild, T., M.J. Bamshad, K. Kaldma et al. 1999. Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineage. *Curr. Biol.* 9:1331–1334.

Kolman, C.J., N. Sambuughin, and E. Bermingham. 1996. Mitochondrial DNA analysis of Mongolian populations and implications for the origins of New World founders. *Genetics* 142:1321–1334.

Lum, J.K., and R.L. Cann. 1998. mtDNA and language support a common origin of Micronesians and Polynesians in Island Southeast Asia. *Am. J. Phys. Anthropol.* 105:109–119.

Lum, J.K., O. Rickards, C. Ching et al. 1994. Polynesian mitochondrial DNAs reveal three deep maternal lineage clusters. *Hum. Biol.* 66:567–590.

Macaulay, V., M. Richards, E. Hickey et al. 1999. The emerging tree of West Eurasian mtDNAs: A synthesis of control-region sequences and RFLPs. *Am. J. Hum. Genet.* 64:232–249.

Melton, T., S. Clifford, J. Martinson et al. 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.

Melton, T., R. Peterson, A.J. Redd et al. 1995. Polynesian genetic affinities with Southeast Asian populations as identified by mtDNA analysis. *Am. J. Hum. Genet.* 57:403–414.

Merriwether, D.A., J.S. Friedlaender, J. Mediavilla et al. 1999. Mitochondrial DNA variation is an indicator of Austronesian influence on island Melanesia. *Am. J. Phys. Anthropol.* 110:243–270.

Page, R.D.M. 1996. TREEVIEW: An application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences* 12:357–358.

Passarino, G., O. Semino, G. Modiano et al. 1993. COII/tRNALys intergenic 9-bp deletion and other mtDNA markers clearly reveal that the Tharus (Southern Nepal) have Oriental affinities. *Am. J. Hum. Genet.* 53:609–618.

Passarino, G., O. Semino, G. Modiano et al. 1996. mtDNA provides the first known marker distinguishing proto-Indians from the other Caucasoids: It probably predates the diversification between Indians and Orientals. *Ann. Hum. Biol.* 23:121–126.

Passarino, G., O. Semino, L. Quintana-Murci et al. 1998. Different genetic components in the Ethiopian population, identified by mtDNA and Y-chromosome polymorphisms. *Am. J. Hum. Genet.* 62:420–434.

Quintana-Murci, L., O. Semino, H.J. Bandelt et al. 1999. Genetic evidence of an early exit of *Homo sapiens sapiens* from Africa through eastern Africa. *Nat. Genet.* 23:437–441.

Redd, A.J., and M. Stoneking. 1999. Peopling of Sahul: mtDNA variation in aboriginal Australian and Papua New Guinean populations. *Am. J. Hum. Genet.* 65(3):808–828.

Redd, A.J., N. Takezaki, S.T. Sherry et al. 1995. Evolutionary history of the COII/tRNALys intergenic 9 base pair deletion in human mitochondrial DNAs from the Pacific. *Mol. Biol. Evol.* 12:604–615.

Richards, M., S. Oppenheimer, and B. Sykes. 1998. mtDNA suggests Polynesian origins in eastern Indonesia. *Am. J. Phys. Anthropol.* 63:1234–1236.

Saitou, N., and M. Nei 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4:406–425.

Schurr, T.G., C. Panter-Brick, and D.C. Wallace. 2000. Mitochondrial DNA variation in Nepal: Ge-

netic differences among Himalayan ethno-linguistic groups (abstract). *Am. J. Hum. Biol.* 12:276.

Schurr, T.G., R.I. Sukernik, Y.B. Starikovskaya et al. 1999. Mitochondrial DNA diversity in Koryaks: Evidence for ancient and recent population expansions in the Okhotsk Sea - Bering Sea region. *Am. J. Phys. Anthropol.* 108:1–40.

Soodyall, H., L. Vigilant, A.V. Hill et al. 1996. mtDNA control-region sequence variation suggests multiple independent origins of an "Asian-specific" 9-bp deletion in sub-Saharan Africans. *Am. J. Hum. Genet.* 58(3):595–608.

Stoneking, M. 2000. Hypervariable sites in the mtDNA control region are mutational hotspots. *Am. J. Hum. Genet.* 67(4):1029–1032.

Stoneking, M., L.B. Jorde, K. Bhatia et al. 1990. Geographic variation in human mitochondrial DNA from Papua New Guinea. *Genetics* 124:717–733.

Sukernik, R.I., T.G. Schurr, Y.B. Starikovskaya et al. 1996. Mitochondrial DNA variation in native Siberians, with special reference to the evolutionary history of American Indians. Studies on restriction endonuclease polymorphism. *Genetika* 32:432–439.

Sykes, B., A. Leiboff, J. Low-Beer et al. 1995. The origins of the Polynesians: An interpretation from mitochondrial lineage analysis. *Am. J. Hum. Genet.* 57:1463–1475.

Torroni, A., K. Huoponen, P. Francalacci, et al. 1996. Classification of European mtDNAs from an analysis of three European populations. *Genetics* 144:1835–1850.

Torroni, A., J.A. Miller, L.G. Moore et al. 1994a. 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., M.T. Lott, M.F. Cabell et al. 1994b. Mitochondrial DNA and the origin of Caucasians: Identification of ancient Caucasian-specific haplogroups, one of which is prone to a recurrent somatic duplication in the D-loop region. *Am. J. Hum. Genet.* 55:760–776.

Torroni, A., M. Petrozzi, P. Santolamazza et al. 1995. About the "Asian"-specific 9-bp deletion of mtDNA. *Am. J. Hum. Genet.* 57:507–508.

Torroni, A., T.G. Schurr, C.-C. Yang et al. 1992. Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations. *Genetics* 130:153–162.

Torroni, A., R.I. Sukernik, T.G. Schurr et al. 1993a. mtDNA variation of aboriginal Siberians reveals distinct genetic affinities with Native Americans. *Am. J. Hum. Genet.* 53:591–608.

Torroni, A., T.G. Schurr, M.F. Cabell et al. 1993b. Asian affinities and continental radiation of the four founding Native American mtDNAs. *Am. J. Hum. Genet.* 53:563–590.

van Holst Pellekaan, S.M., M. Frommer, J.A. Sved et al. 1998. Mitochondrial control-region sequence variation in Aboriginal Australians. *Am. J. Hum. Genet.* 62:435–449.

Wallace, D.C., and T.G. Schurr. 2001. mtDNA diversity in Vietnamese, Malaysian and Bornean populations (abstract). *Am. J. Phys. Anthropol.* 32:159.

Watkins, W.S., M. Bamshad, M.E. Dixon et al. 1999. Multiple origins of the mtDNA 9-bp deletion in populations of south India. *Am. J. Phys. Anthropol.* 109:147–158.

Watson, E., P. Forster, M. Richards et al. 1997. Mitochondrial footprints of human expansions in Africa. *Am. J. Hum. Genet.* 61:691–704.

Wrischnik L.A., R.G. Higuchi, M. Stoneking, et al. 1987. Length mutations in human mitochondrial DNA: Direct sequencing of enzymatically amplified DNA. *Nucleic Acids Res.* 15:529–542.

Yao, Y.G., W.S. Watkins, and Y.P. Zhang. 2000. Evolutionary history of the mtDNA 9-bp deletion in Chinese populations and its relevance to the peopling of East and Southeast Asia. *Hum. Genet.* 107:504–512.