Mitochondrial DNA Variation in Island Southeast Asia

Catherine E. Hill B.Sc. (Hons), M.Sc.

University of
HUDDERSFIELD

Thesis submitted in partial fulfilment of the requirements for the Degree of Doctor of Philosophy

Department of Chemical and Biological Sciences,
University of Huddersfield

June 2005

Contents

Appendices

List of Tables and Figures

Tables

Figures

List of Abbreviations

AMOVA	Analysis of molecular variance
ATP	Adenosine triphosphate
B.C.	Before Christ
BP	Before present
bp	Base pairs
CI	Confidence interval
COII	Cytochrome c oxidase II
CRS	Cambridge reference sequence
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
EDTA	Ethylenediaminetetraacetic acid
Hb-E	Haemoglobin E
HVS	Hypervariable segment
G6PD	glucose-6-phosphate dehydrogenase
kb	Kilobase
MRCA	Most recent common ancestor
mtDNA	Mitochondrial DNA
NaOH	Sodium hydroxide
np	Nucleotide position
PC	Principal component
PCR	Polymerase chain reaction
RFLP	Restriction fragment length polymorphism
RNA	Ribonucleic acid
rRNA	Ribosomal RNA
SE	Standard error
SNP	Single nucleotide polymorphism
STR	Short tandem repeat
Taq	*Thermus aquaticus*
TMRCA	Time to the most recent common ancestor
TPK	Ta-p'en-k'eng
tRNA	Transfer RNA

UV	Ultra-violet
v/v	volume/volume
w/v	weight/volume

Acknowledgements

I would like to thank Martin Richards and Dougie Clarke for the supervision and help they have provided throughout this research. I would also like to thank Vincent Macaulay and Antonio Torroni for the assistance they have provided with different aspects of the project.

I would also like to say thank you to my family, especially my parents and Jane, and all my friends, especially Esther, Emily, Anna, Lou, Natalie and Ruth for all their help and support and for putting up with me during the more stressful times!

THANK YOU

Abstract

It is known that Island Southeast Asia was colonised relatively early in the history of modern humans; however, it is still a matter of some debate as to whether the modern inhabitants of Island Southeast Asia are descended from these original inhabitants or are the result of some later migration. Currently, the prevailing theory in both archaeology and linguistics is that the modern inhabitants of Island Southeast Asia are largely descended from an agricultural people who originated in China and Taiwan around 6,000 years ago. From there they are thought to have migrated through the Philippines and into Eastern Island Southeast Asia around 2,500-1,500 B.C. assimilating or replacing the indigenous peoples. However, other researchers have suggested that a model of regional continuity is more suitable for Island Southeast Asia and that the modern inhabitants are the direct descendents of the original Pleistocene inhabitants. Still others have suggested that intermediate models would be more appropriate.

This study aimed to use mitochondrial DNA to test the validity of these models. A secondary aim was to look at the mitochondrial DNA of the indigenous Orang Asli groups of the Malay Peninsula in an attempt to reconstruct a picture of the early Pleistocene variation of Southeast Asia. To this end, mitochondrial DNA was obtained and sequenced from 885 individuals from various locations in Island Southeast Asia and also 259 Orang Asli individuals.

This study has demonstrated that the populations of Island Southeast Asia contain a high level of genetic diversity, including a number of novel haplogroups. Significant differences have also been found between Eastern and Western populations suggesting that they have been established long enough to become regionally specific. Most Island Southeast Asian haplogroups date to the Pleistocene or early Holocene which suggests that they are mostly indigenous to the area. Those which could have a connection to Taiwan seem too old to have been part of an 'out of Taiwan' event as it has been traditionally visualised. Only ~13% of mtDNA types (belonging to haplogroups M7c1c, D5 and Y2) could be linked to such an event suggesting that if a migration did occur it was demographically minor.

A number of novel haplogroups were also found in the Orang Asli which form strong support for the theory that that at least the Semang, if not all Orang Asli groups in part, are descended from the original Pleistocene inhabitants of the Malay Peninsula. These novel haplogroups diverge from the same set of founder types as the haplogroups found across the rest of Eurasia; that they diverge from close to the roots of these founder types suggests they are of considerable antiquity. This, along with expansion dates of ~60,000 obtained in this study, suggests that only a single, early 'out of Africa' event took place which led to the peopling of the rest of the world by modern humans.

1. Introduction

1. Introduction

The main aim of this study is to remedy the relative paucity of mitochondrial DNA (mtDNA) data on the indigenous peoples of Island Southeast Asia and the Malay Peninsula and to use this data to construct a genealogical tree for the maternal side of Southeast Asian ancestry. It should then be possible to estimate the number and timings of different colonisation events using the geographic origin of the samples and the time depth of lineages on the genealogical tree.

1.1 Mitochondrial DNA

Mitochondria are cytoplasmic organelles which are the source of ATP (and therefore energy) production in eukaryotic cells. They have their own genetic system which consists of a 16,569 base pair circular genome in humans (Anderson *et al.* 1981; Andrews *et al.* 1999) and are thought to have originated as endosymbiotic bacteria which were taken up into eukaryotic cells around 1.5 billion years ago (Margulis 1981).

The mitochondrial genome codes for a number of genes including: two rRNAs, 22 tRNAs, cytochrome c oxidase subunits I, II and III, ATPase subunit 6 and cytochrome b (Anderson *et al.* 1981). mtDNA is unusual in that there are either no or very few non-coding bases between the genes; however, the genome does include a section known as the control region (or D-loop) which is non-coding and is therefore the least conserved section of the genome (Anderson *et al.* 1981). mtDNA is also unusual in that its genetic code differs the universal vertebrate nuclear genome code: UGA codes for tryptophan instead of being a 'stop' codon, AUA codes for methionine not isoleucine, AGA and AGG are 'stop' codons instead of coding for arginine, and AUA and AUU are used as initiation codons (Anderson *et al.* 1981).

mtDNA is particularly useful for studying human origins as not only is it present in multiple copies in each cell (Clayton 1982) (making it relatively easy to extract large amounts of DNA), but it is effectively maternally inherited and therefore does not recombine, thus creating a clear maternal genealogy. Spermatozoa do contain ~50-75 mitochondria within their midpiece which enters the oocyte at fertilisation (Ankel-

Simmons and Cummins 1996). However, there appears to be a stringent process in mammalian fertilised eggs which excludes paternal mtDNAs and ensures absolute maternal inheritance (Shitara *et al.* 1998). Paternal inheritance has been detected in mice (Gyllensten *et al.* 1991); however, further study has shown that when this does occur, not enough mtDNA is leaked for it to be distributed to all tissues and it is not transmitted through the germline to the next generation (Shitara *et al.* 1998). Mitochondrial DNA also mutates at a faster rate than nuclear DNA (Brown *et al.* 1979) enabling relatively recent prehistoric events to be studied. Furthermore, these mutations accumulate at a relatively constant rate allowing demographic events to be dated (Ayala 1995).

1.2 Mitochondrial DNA and the Evolution of Modern Humans

mtDNA has now been used for over 20 years in an attempt to answer questions of human origins, migrations and settlement patterns. One of the first studies in the field was that of Brown (1980) who used 18 restriction enzymes to study the mtDNA of 21 humans of different geographic and ethnic backgrounds. The results of this work suggested that the small amount of sequence heterogeneity found could have accumulated in the last 180,000-360,000 years and that modern-day humans could have evolved from a small population which existed at that time. This study also raised the possibility of population-specific restriction fragment length polymorphisms (RFLPs).

This work was elaborated on by Cann *et al.* (1987) and Vigilant *et al.* (1991) amongst others. They extended the number and geographic range of the samples studied and developed the methodology used in studying them to include sequencing of the two hypervariable segments of the control region (Vigilant *et al.* 1991). Both studies supported an age of around 200,000 years for the most recent common ancestor (MRCA) of modern-day human populations and suggested that, furthermore, this ancestor lived in Africa. The African origin was supported by the fact that, in both studies, the genealogical tree of mtDNA types consisted of one branch leading exclusively to African mtDNA types and a second leading to both African and non-African types. This led to mtDNA studies joining the controversy concerning modern human origins

and added further support to the 'out of Africa' position as opposed to that of 'multi-regionalism'.

Exponents of these models of modern human evolution have been debating their merits for the past few decades. According to followers of multiregionalism, *Homo erectus* (which ranged from Africa to Asia) evolved independently in each of the locations it had spread to, with gene flow occurring between the different populations leading to the gradual and parallel development from *Homo erectus* to *Homo sapiens* (Wolpoff and Thorne 1991). In contrast, the 'out of Africa' hypothesis states that modern *Homo sapiens* arose in Africa around 200,000 years ago before spreading around the rest of the world replacing the more archaic forms of man (Stringer and Andrews 1988). The proponents of both theories claim that the fossil record supports them.

Multiregionalists such as Milford Wolpoff and Alan Thorne believe that regional features can be followed through the fossil record to modern humans with no need for an influx of Africans. For example, they feel that the same features can be traced from the *Homo erectus* specimens found in Asia to the earliest known Australians of ~60,000 years ago and on to modern Australians. They also apply this theory to Europe. Over half of Neanderthals had the opening to their mandibular nerve canal covered by a broad, bony bridge, a trait which is also found in modern humans, in particular Europeans. This is seen by Wolpoff and Thorne (1991) as direct evidence that Neanderthals contributed to the modern human gene pool as this appears more likely than the trait evolving twice.

However, this hypothesis depends on the occurrence of numerous migrations and interbreeding between populations from different continents. There are no obvious parallels to this and it is hard to equate with the current existence of different species or subspecies of other animals in different regions (Ayala 1995). The alternative is the 'out of Africa' hypothesis which is now by far the dominant view and which does not require the concept of parallel evolution. This is based on the evidence that, according to some palaeontologists, late Pleistocene fossils from China resemble European and African middle Pleistocene hominids more than their "supposed local ancestors" (Stringer and Andrews 1988). Also the earliest *Homo sapiens* fossils are found in

Africa and the Levant, no Neanderthal/*Homo sapiens* transitional fossils have been found in Europe despite the excellent fossil record, and modern humans seem to have been present in the Levant before Neanderthals (Stringer and Andrews 1988). Work on cranial morphology has also supported the view of Neanderthals and modern humans as separate species (Ponce de Léon and Zollikofer 2001). The recent discovery of a possible new hominin species in Flores, Indonesia, is further evidence against the multiregional hypothesis. If confirmed as a new species of *Homo*, this hominin (known as *Homo floresiensis*), which lived until as recently as 18,000 years ago, must have evolved (probably from *Homo erectus*) without any gene exchange with other hominins and also made no contribution to the gene pool of modern *Homo sapiens* (Brown *et al.* 2004; Morwood *et al.* 2004).

As stated above, early mtDNA work appeared to support the 'out of Africa' model. However, these early studies suffered problems due to their lack of resolution. For example, a reanalysis of the Cann *et al.* (1987) data showed that many phylogenetic trees could be constructed which were more parsimonious than the original (Templeton 1993). Some of the more parsimonious trees contained a purely African primary branch as did the original, while others had a mixed African-Asian primary branch. The same problems were found with the Vigilant *et al.* (1991) data. This did not prove that the MRCA was non-African, or even that the new trees were the most parsimonious. It only showed that more parsimonious trees existed, and that a more complete analysis was needed. However, these problems have since been mostly overcome by the use of more extensive sample sets, combined control region sequencing with coding region RFLP typing (e.g. Torroni *et al.* 1996; Macaulay *et al.* 1999), complete mtDNA sequencing (e.g. Ingman *et al.* 2000; Herrnstadt *et al.* 2002) and the use of better phylogenetic analysis (e.g. Bandelt *et al.* 1995; Penny *et al.* 1995; Watson *et al.* 1997).

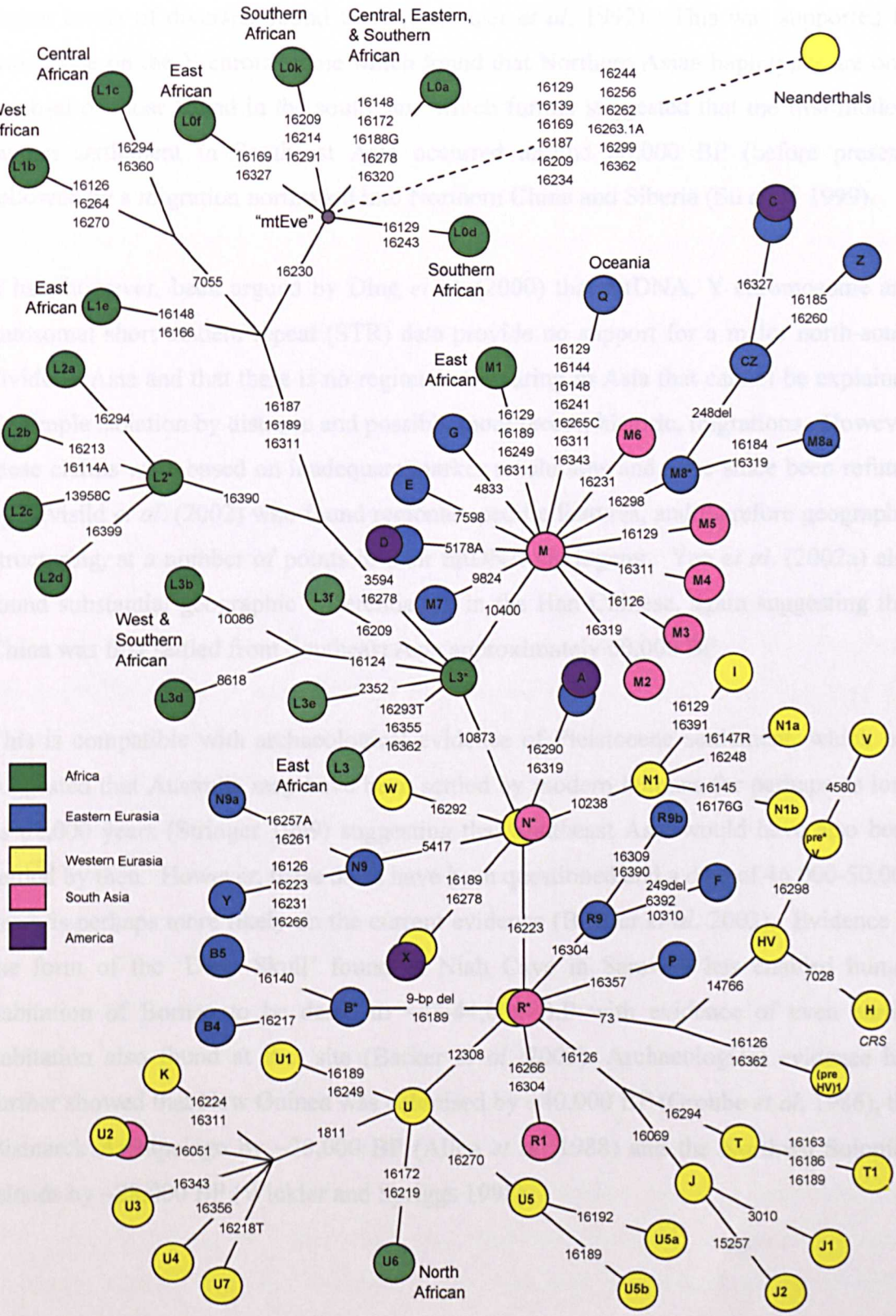
More detailed studies have confirmed the African origin of modern humans and have shown that all non-African mtDNAs form subclusters of an African clade termed L3 (Watson *et al.* 1997) which is defined by the lack of a *Hpa*I site at nucleotide position (np) 3592 relative to the Cambridge Reference Sequence (or CRS [Anderson *et al*, 1981]) and which expanded from East Africa approximately 60,000 years ago (Mountain *et al.* 1995; Watson *et al.* 1997). These non-African subclusters of L3

themselves form 'macrohaplogroups' M (defined by a transition at np 10400 and the resultant gain of an *Alu*I site at np 10397) and N (defined by a transition at np 10873 and the subsequent loss of a *Mnl*I site at np 10871) (Torroni *et al.* 1994; Quintana-Murci *et al.* 1999), haplogroups from both of which are found across Asia. The general structure of the world-wide tree has been confirmed by the use of complete mtDNA sequences, see figure 1. Ingman *et al.* (2000) used 53 complete mtDNA sequences to create a tree which is almost bifurcating and in which the three oldest clades are entirely African.

Ancient mtDNA has also been used to support the out-of-Africa hypothesis. Neanderthal mtDNAs have been shown to form a separate clade from all modern human mtDNAs. Comparison between Neanderthal and modern human mtDNA has suggested that the Neanderthals diverged from anatomically modern humans ~500,000 years ago, further undermining the multiregional hypothesis (Krings *et al.* 1997; Krings *et al.* 1999a; Ovchinnikov *et al.* 2000).

mtDNA has also been used more specifically to study the settlement processes in different continents and countries, particularly Europe and the Americas. The use of high-resolution restriction analysis showed that most mtDNA types could be classified into different haplogroups (e.g. Schurr *et al.* 1990; Ballinger *et al.* 1992; Chen *et al.* 1995; Torroni *et al.* 1996). Most of these haplogroups were, as predicted by Brown (1980), restricted to certain geographic areas. These haplogroups have since been refined by the addition of control region information (e.g. Macaulay *et al.* 1999) and, more recently, by the use of conformation-sensitive gel electrophoresis. Finnilä *et al.* (2001) used the latter technique to produce a network of European mtDNAs which, although providing many new markers for identifying haplogroups, still agreed predominantly with the earlier classifications found by restriction analysis.

Figure 1 – Skeleton tree of mtDNA haplogroups, showing their geographic distribution (after M. Richards)



Most of the work done on Asia has so far concentrated on migration routes into China and East Asia, and also out into Polynesia. Early work on mtDNA in Asia suggested that South China was the centre of modern human expansion in Asia based on the higher levels of diversity found there (Ballinger *et al.* 1992). This was supported by work done on the Y chromosome which found that Northern Asian haplotypes are only a subset of those found in the south, and which further suggested that the first modern human settlement in Southeast Asia occurred around 60,000 BP (before present) followed by a migration northward into Northern China and Siberia (Su *et al.* 1999).

It has, however, been argued by Ding *et al.* (2000) that mtDNA, Y chromosome and autosomal short tandem repeat (STR) data provide no support for a major north-south divide in Asia and that there is no regional structuring in Asia that cannot be explained by simple isolation by distance and possibly more recent, historic, migrations. However, these claims were based on inadequate marker resolutions and have since been refuted by Kivisild *et al.* (2002) who found regional-specific features, and therefore geographic structuring, at a number of points in their mtDNA phylogeny. Yao *et al.* (2002a) also found substantial geographic differentiation in the Han Chinese, again suggesting that China was first settled from Southeast Asia approximately 60,000 BP

This is compatible with archaeological evidence of Pleistocene settlement, which has suggested that Australia may have been settled by modern humans for perhaps as long as 62,000 years (Stringer 1999) suggesting that Southeast Asia would have also been settled by then. However, these dates have been questioned and a date of 46,000-50,000 years is perhaps more likely on the current evidence (Bowler *et al.* 2003). Evidence in the form of the 'Deep Skull' found in Niah Cave in Sarawak has enabled human habitation of Borneo to be dated to ~43-44,000 B.P with evidence of even earlier habitation also found at that site (Barker *et al.* 2002). Archaeological evidence has further showed that New Guinea was colonised by ~40,000 BP (Groube *et al.* 1986), the Bismarck Archipelago by ~33,000 BP (Allen *et al.* 1988) and the Northern Solomon Islands by ~28,000 BP (Wickler and Spriggs 1998).

1.3 The Origins of Modern Southeast Asians – Archaeology and Linguistics

As stated above, archaeological evidence has shown that Island Southeast Asia was colonised relatively early in the history of modern humans. However, controversy remains about whether the modern inhabitants of Island Southeast Asia are descended from these original inhabitants or are the result of some later migration. This controversy is linked to the colonisation of Polynesia as, despite this early colonisation of Southeast Asia and Melanesia, Polynesia was the last main area of the world to be colonised by humans with Easter Island being colonised by 500 A.D., Hawaii by 600 A.D. and New Zealand not until 1100 A.D. It has been suggested in the past that the Polynesians may have originated in South America (Heyerdahl 1950); however, as will be discussed later in this section, this theory is not supported by any of the archaeological, linguistic or genetic evidence. It would therefore seem obvious that these colonists would at least have had to pass through Island Southeast Asia (see figure 2 for a map of the area under discussion) if they did not originate there.

Theories concerning the history of Island Southeast Asia and the settlement of Polynesia are generally linked to the origin of the Austronesian language family. There are around 1200 Austronesian languages (www.ethnologue.com) which are spoken throughout Island Southeast Asia and Polynesia and also in coastal Melanesia and Madagascar. However, they are not spoken in highland New Guinea (where non-Austronesian, Papuan languages are spoken) or in mainland Southeast Asia (apart from in a small enclave in interior Vietnam) (www.ethnologue.com).

Austronesian languages are generally thought to be derived from languages spoken in mainland Asia. It has been suggested that they form part of an "Austro-Tai" superfamily along with the Kadai languages (spoken in Hainan and Southern China), the Hmong-Mien or Miao-Yao languages (also spoken in Southern China), and Thai (Benedict 1975). A partial 'Proto-Austro-Tai' vocabulary has also been attempted which includes words for field, plough, rice, sugarcane, cattle and canoe. If this theory is correct then it links the ancestry of the Austronesian-speakers to the Neolithic communities of Southern China. However, an alternative hypothesis links Austronesian

and Austroasiatic languages (Blust 1996b); the latter comprise the Munda languages of Eastern India and the Mon-Khmer languages of mainland Southeast Asia. Blust (1996b) suggests that the homeland of this 'Austric' superfamily was located in the North-western area of what is now the Yunnan province of China, where the Salween, Mekong and Yangzi rivers run parallel to each other. Blust (1995) also includes Tai-Kadai within an expanded Austric superfamily.

Figure 2 – Map of Southeast Asia

Map courtesy of the CIA World Factbook

One of the main theories of Island Southeast Asian and Polynesian origins is the 'out of Taiwan' model championed by the archaeologist Peter Bellwood (Bellwood 1997) and the linguist Robert Blust (1996a) This model, also known as the 'express train to Polynesia' (Diamond 1988), argues that the ancestors of the modern Austronesian speakers (including, of course, the Polynesians) were a rice-growing, agricultural people who arose on the Southern Chinese mainland and Taiwan approximately 6,000 years ago. From their original homeland, these people migrated through the Philippines and into Eastern Island Southeast Asia around 2,500-1,500 B.C. from where they moved into both Western Island Southeast Asia and Melanesia around 1,500-500 B.C. and eventually reached Polynesia starting in around 300 A.D. A further aspect of this model is that, due to their technological 'superiority', these Neolithic immigrants would have largely replaced the indigenous hunter-gatherer populations of Island Southeast Asia apart from in highland New Guinea where they have survived as the modern Papuan-speakers.

This model is derived partly from the fact that at least 9 primary branches of Austronesian are spoken only by Taiwanese aborigines (Blust 1999). The other branch is argued to be the source of all the Malayo-Polynesian branches of the Austronesian family which are spoken across Island Southeast Asia and Polynesia but are not spoken in Taiwan. Malayo-Polynesian languages are, in turn, classified into the following subgroups: Western Malayo-Polynesian (spoken in the Philippines, Vietnam, Madagascar, Peninsular Malaysia, Sumatra, Java, Borneo, Sulawesi, Bali, Lombok and Western Sumbawa), Central Malayo-Polynesian (spoken in the Lesser Sundas eastward from Eastern Sumbawa and in the Moluccas except Halmahera) and Eastern Malayo-Polynesian which is itself made up of two subgroups. The first of these is spoken in South Halmahera and West New Guinea and the second is spoken across the rest of Melanesia, Micronesia and Polynesia. This is interpreted as meaning that the Austronesian languages must have developed in Taiwan and spread from there across their current distribution with the various branches of Malayo-Polynesian separating along the voyage (Blust 1996a). It should be noted that Western Malayo-Polynesian is not characterised by any phonemic innovations; it is rather comprised of all the Malayo-Polynesian languages which do not have the diagnostic features of Central or Eastern Malayo-Polynesian (Blust 1999).

This was supported by Gray and Jordan (2000) who carried out a parsimony analysis on a linguistic dataset of Austronesian languages. To do this they converted cognate sets from 77 Austronesian languages into a binary matrix. The most parsimonious tree found in this analysis was a close fit to that expected with the 'express train' model; however many of the branches were not well supported. Also Gray and Jordan (2000) assume a Taiwanese root for their data, a Filipino root seems equally plausible. In a more recent work, Greenhill and Gray (in press) created a maximum likelihood tree with the same data which had better resolution and was better supported. This tree also appeared to support the 'express train' model, although some aspects of the branching order were not as expected. For example, the languages included from North Borneo and Brunei occurred basally next to the Taiwanese languages.

It has also been noted by Bellwood (2003) that comparisons with historical data suggest that large-scale language spread does not occur without substantial levels of colonisation. Furthermore, it has been estimated that the maximum time that shared cognates will remain in any language family is ~7,000-10,000 years (Nichols 1998); this in turn suggests an origin for the Austronesian language family sometime in the Holocene and therefore around the time of the beginnings of agriculture (Bellwood 2005). However, this date rests on a branch of linguistics known as glottochronology which assumes that languages change at a constant rate over time. This theory has been criticised on a number of points and is rejected by a number of authors (e.g. Renfrew 1987).

Bellwood (1997) also uses archaeological evidence to trace connections between the Chinese mainland, Taiwan, the Philippines and Island Southeast Asia. For example, the Ta-p'en-k'eng (TPK) culture, which began in Taiwan around 4,300 B.C., has supposed antecedents in Guangdong and Fujian on the Chinese mainland which date to 5,200-4,200 B.C. It is possible that these cultures can in turn be traced back to the Dawenkou tradition of Shandong (which dates from ~6,000 B.C. to ~2,500 B.C.) which has been found to include similar burial practices to those found in Fujian and Taiwan, and which also featured red-slipped pottery (Bellwood 2005). The TPK culture possessed

rice and foxtail millet production and also had reaping knives, spindle whorls and barkcloth beaters (Bellwood 2005).

The TPK culture was followed in Northern Taiwan by the Yüan-shan culture (2,500 B.C. until the first millennium B.C.) which has parallels with some of the early Neolithic sites found in Luzon in the Northern Philippines. For example, artefacts such as pottery earrings and slate projectile points are found in both Taiwan and the Philippines (Bellwood 1997). There is also evidence of a Neolithic trade in jade from Fengtian in Eastern Taiwan to the Philippines (Bellwood 2005). The presence of red-slipped pottery and shell artefacts in the Philippines and across Island Southeast Asia has been linked by Bellwood (1997) to the prehistoric Lapita culture of Melanesia and Polynesia and is therefore seen as a further marker of Southeast Asian influence. However, Szabó and O'Connor (2004) have questioned these links, noting that shell ornaments were extremely diverse, both in morphology and mode of manufacture, across Island Southeast Asia and that many of these techniques were different from those used in the Lapita culture. Szabó and O'Connor (2004) also observe that a supposed similarity does not necessarily indicate a "familial" relationship. It could instead be due to a variety of cultural interactions such as trade and exchange, diffusion of an idea or a technology, or the persistence of an already existing culture.

It has also been suggested by Bellwood (1997) and Blust (1996a) that by reconstructing Proto-Austronesian words it is possible to demonstrate the subsistence basis of these expansions from Taiwan. To illustrate this they cite such examples as the words for rice, sugarcane, dog, pig, fowl, sky and head louse which are found in similar forms from Taiwan to Easter Island. Words such as thresh, granary, pestle, mortar and harvest which are obviously related to cultivation of grain crops can also be traced back to proto-Austronesian (Pawley 2003). However, the words for various boat-building activities and also the words for most Pacific Austronesian foods (such as breadfruit, yams, bananas, sago, betel-nuts and chicken) can only be reconstructed to Proto-Malayo-Polynesian and can therefore only be traced back as far as the Philippines at the most (Bellwood 1997). It has also been emphasised by Szabó and O'Connor (2004) that the presence of a term in a reconstructed proto-language does not indicate anything

about the relative importance of the item/practice or its exclusive association with the speakers of that proto-language.

The main alternatives to the 'out-of Taiwan' model suggest that the Austronesian languages, and therefore the ancestral Polynesians, evolved somewhere in Island Southeast Asia or Melanesia. If this is the case then the modern inhabitants of Island Southeast Asia are the direct descendents of the original Pleistocene inhabitants. As stated above, Austronesian languages are now spoken throughout Island Southeast Asia and Melanesia except highland New Guinea, some parts of New Britain and New Ireland, and the New Georgia archipelago where non-Austronesian, Papuan languages are spoken (Foley 1986). These languages are also spoken, to a lesser extent, in central and eastern Timor, Alor, Pantar, Morotai and northern Halmahera (Foley 1986); these may be due to migrations within the last ~4,000 years (Pawley 2003). These Papuan languages are extremely diverse with 750 of them being present on New Guinea alone. This is mainly due to the combination of the length of human habitation on the island (~40,000 years according to the evidence of Groube *et al.* [1986]) and the physical barriers between different groups created by the difficult terrain of New Guinea (Foley 1986).

The relationship between these Papuan languages is unclear. Foley (1986) has argued that they are fragmented into around sixty separate families and that they are more appropriately termed 'non-Austronesian' languages due to their negative categorization. However, Wurm (1983) has postulated the existence of a major Trans-New Guinea phylum which comprises over 500 languages and which stretches across more than four-fifths of New Guinea and extends to Timor. He further suggests that the other Papuan languages can be classified into a further four major and six minor groups with only just over half a dozen remaining as unrelated isolates (Wurm 1983).

Dyen (1962) used lexicostatistics to argue that the highest levels of diversity within the Austronesian language family were found in the New Britain area of Melanesia and on the east coast of New Guinea. To him this suggested that the language family must have arisen there. However, this method of analysis has since been criticised as the

Austronesian languages of Melanesia and New Guinea may have become more varied due to their proximity to the Papuan languages (Blust 1995; Bellwood 2005).

It has been argued by Terrell (1981) that the geographic distribution of Austronesian and non-Austronesian speakers on New Guinea can be explained by the divergence of one source population due to "competition and displacement" rather than the arrival of a second, more-'advanced' migrating group replacing some of the indigenous inhabitants. He has also emphasised that Austronesian and Papuan speakers cannot be distinguished unambiguously on cultural grounds and that the linguistic evidence is inconclusive. The Austronesian languages of Papua New Guinea and its adjacent Melanesian islands show great lexical diversity and also have many similarities with non-Austronesian languages, these are usually attributed to borrowing but could reflect a common origin. If this is the case then it seems most likely that the diversity within both Austronesian and non-Austronesian languages has arisen within their present distribution (Terrell 1981).

More recently, Terrell (2004) has argued that the apparent dichotomy between Austronesians and Papuans can be explained by changing environmental factors. In the Late Pleistocene, most of coastal Papua New Guinea would have been uninhabitable and so once humans arrived there they would have had to move inland and would have found it difficult to keep in contact with the areas from which they had originated. At this time, due to lower sea levels, what is now New Guinea was attached by a land bridge to the north of what is now Australia. This would have created a barrier between the island groups of Wallacea in the west and the Bismarks and the Solomons in the east, thus enabling the divergence of culture, speech and physical appearance between the peoples of Southeast Asia and Wallacea and those of New Guinea and the islands to the east. However, by ~6,000 BP, sea levels had risen and created new floodplains and lagoons which would have been much better suited to human habitation. Therefore, by the mid-Holocene, New Guinea would no longer have been isolated and trade could have begun across the whole of Island Southeast Asia.

Further to this, it has also been proposed that, due to the relatively small distances between its islands, much of Melanesia would have been part of an ancient 'voyaging corridor' which stretched from mainland Southeast Asia to the end of the Solomon

Islands (Irwin 1992). This would have made it possible for the indigenous peoples to learn the navigational and survival skills necessary to colonise the more distant Pacific. This 'voyaging corridor' may also have enabled trade and interaction between groups over long distances.

Terrell and Welsch (1997) have also used archaeological evidence to criticise the prevailing 'out of Taiwan' model. As stated earlier, the Lapita culture is often cited as proof of a trail of pottery styles and archaeological evidence leading back to the Philippines and Taiwan. However, Lapita pottery has only ever been found in Melanesia and Polynesia (Tonga and Samoa) and not in mainland Asia, Taiwan, the Philippines or Indonesia. In fact it is now known that the Lapita style was developed in the Bismarck Archipelago, therefore any links between Lapita assemblages in Melanesia or Polynesia can only have "generic, not specific" links to early ceramic assemblages from Indonesia (Terrell and Welsch 1997). Terrell and Welsch (1997) also argue that there is no evidence showing Lapita as a widespread and unified culture. The modern-day communities on the 700 km stretch of New Guinean coast studied by Terrell and Welsch (1997) speak 60 different languages which belong to 24 language families; despite this, their material cultures are remarkably similar. This may also have been the case in prehistory, meaning that the presence of related artefacts at Lapita sites does not automatically indicate that the people living at those sites belonged to the same "linguistic group, ethnic group or 'people' " (Terrell and Welsch 1997).

Furthermore, there appears to be no evidence that social and economic contacts between Lapita communities were any different to those which had been developing in the 'voyaging corridor' prior to this time as would perhaps be expected if a new group of people had suddenly arrived in the vicinity. It has often been claimed that the domestication of plants and animals enabled the Lapita expansions into Polynesia (Bellwood 1997). However, Terrell and Welsch (1997) suggest that Lapita subsistence was more likely based on both wild and managed food resources and that there is also no way of showing that the food-management processes seen at some Lapita sites were specific to them or if they were also found at 'non-Lapita' sites. Terrell and Welsch (1997) conclude that the Lapita culture was a product of indigenous development rather

than an introduction by a foreign people and that there was possibly a cultural or social barrier which prevented it being transmitted to mainland New Guinea.

Moreover, there is evidence that the people of Melanesia had already developed ground stone, shell tool and ceramic technologies, horticulture and efficient sailing techniques before the beginnings of the Lapita culture (Allen 1996). There is also evidence that the peoples of the 'voyaging corridor' and highland New Guinea had begun to develop agriculture by 9,000 BP Furthermore there is little evidence of rice agriculture in Island Southeast Asia during the period of the putative Austronesian expansion. In two of the three sites where rice and pottery are found together during this period (Gua Sireh in Western Sarawak and Ulu Leang in Southern Sulawesi), the pottery is not of red-slipped, and therefore potentially Lapita related, design (Paz 2003). Taking this into account, and as little Oceanic agriculture has an unambiguous Asian origin, it seems unlikely that the introduction of agriculture from a extraneous source is needed to explain the expansion of the Lapita 'culture'.

Meacham (1984-1985) has also argued against an Austronesian homeland in South China. He argues that while it may be true that the Taiwanese branch of the Austronesian language family may appear to diverge first, this does not necessarily mean that the language family arose in Taiwan. Instead, he contends that the Austronesian languages may have originally had a much greater diversity, much like that seen in the Papuan languages of New Guinea today. This diversity would only have been lost as mobility and communications between groups increased. The Taiwanese languages may have maintained this early diversity if they were not part of this trend of increased mobility. Meacham (1984-1985) also emphasises the fact that there is no evidence that Austronesian languages were ever spoken on the Chinese mainland. Blust (1995), however, suggests that this may not be an issue of concern as the expansion of Han Chinese over the past ~2,000 years may have caused many languages to disappear.

Meacham (1984-1985) further asserts that there is no archaeological evidence for any population pressure in the Neolithic of South China which is generally supposed to be the catalyst for any Austronesian expansion (Bellwood 1997). This should be apparent

in a high density of archaeological sites in the relevant locales; however, the density of sites is always relatively low and even in the Bronze Age "great tracts of lowland appear to have been unoccupied" (Meacham 1984-1985). The same general cultural sequence also seems to be found along much of the coast, suggesting no significant population movements.

The archaeological sequence in Taiwan is seen by Meacham (1984-1985) as being very different from that on the mainland and to support a long period of isolation with only minor contacts. While a number of artefacts are shared between Taiwan and the mainland, they appear in extremely localized contexts in Taiwan and hundreds or thousands of years after they appear on the mainland. Unlike Bellwood (1997), Meacham (1984-1985) suggests that the TPK culture of Taiwan did not originate in China but rather came from the tropics by 5,000-4,000 B.C. or had its origins in the Pleistocene peoples of Taiwan. According to Meacham (1984-1985), the prevailing picture of Taiwanese prehistory is one of isolation.

His theory of Taiwanese prehistory can be summarised as follows: boat development would have been earlier in Greater Sundaland as it broke up into different islands, then as Sundaland broke up Taiwan would have been entered by boat from the south. Several factors suggest that Taiwan was part of 'Austronesia': lack of an Early Neolithic, the continuance of the Palaeolithic, the absence of painted pottery, the presence of horticulture but no rice cultivation, and the presence of Austronesian languages. To Meacham (1984-1985), this suggests that there was one initial migration to Taiwan from the south at some point between 10,000 B.C. and 5,000 B.C with little contact afterwards.

Solheim *et al.* (in press; personal communication) have also argued for a non-Chinese origin for Austronesian languages. They have proposed the term 'Nusantao' for the natives of Southeast Asia (including South China) who were part of a maritime trading and communications network which originated around the South Philippines and East Indonesia. From there, the Nusantao spread north through the Philippines to Southern Taiwan and South-East China at some point before 5,000 B.C. These people would have aided the development of Austronesian languages from proto-Austronesian and

possibly started a river trade in South China from where they introduced the TPK culture back into Taiwan. Like Meacham (1984-1985), Solheim *et al.* (in press; personal communication) suggest that at some point Taiwan became isolated, enabling a number of separate Austronesian languages to develop there.

Solheim *et al.* (in press; personal communication) propose that the Nusantao are represented in the archaeological record by a pottery culture which has been termed 'Sa Huynh-Kalanay'. This has been found in the Kalanay cave site in Masbate island in the Philippines, Sa Huynh in Vietnam (the oldest of the sites), the east coast of Thailand, Malaysia, Sarawak and Indonesia. Solheim *et al.* (in press; personal communication) also suggest that the Nusantao network spread from South-East China, through Indonesia and out to the Bismarck Archipelago; there the Sa Huynh-Kalanay pottery was incorporated into the existing Melanesian culture to create the Lapita culture.

A number of researchers have proposed an intermediate history which allows for a Neolithic migration from Taiwan but which includes more interaction with indigenous groups than is proposed by Bellwood (1997). For example, in a review of the current dates of the Neolithic transition, Spriggs (2003) has suggested that, while the spread of the Neolithic from mainland China to Taiwan and on through the Philippines to Sulawesi was relatively straightforward, the picture is not as clear from then on. It is in that area that a number of cultural changes seem to have occurred. For example, rice was the staple crop of the Taiwanese Neolithic but was replaced by root crops outside the Philippines and possibly Borneo. Therefore, Spriggs (2003) suggests that the beginnings of the Neolithic in Southern Wallacea were mainly due to the integration of incoming Neolithic groups with the indigenous pre-Neolithic populations which may have been "pre-adapted" to Neolithisation by the adoption of subsistence practices from New Guinea.

It has also been suggested that the domestic pig species which was introduced to the area east of Sulawesi was a hybrid between the common Eurasian species (*Sus scrofa*) and the indigenous Sulawesi species (*Sus celebensis*) (Groves 1981). However, this has not been supported by recent mtDNA evidence (Larson *et al.* 2005).

This idea of integration has been formalised by Green (1991) as part of his 'intrusion/ innovation/integration' or 'Triple-I' hypothesis. He notes that many portable, non-ceramic aspects of the Lapita assemblage have no precedent in the Bismarcks, thus suggesting some level of 'intrusion' by an incoming population. However, other aspects of Lapita culture appear to have been 'integrated' into it from the indigenous cultures of the region. These include certain domesticated plants such as taro, bananas, breadfruit, sugarcane and coconuts; the obsidian exchange system also seems to have been carried over from the pre-Lapita inhabitants. Advances in sailing technology show that 'innovation' occurred once the incoming and indigenous cultures became in contact with one another.

Bulbeck *et al.* (2001) also emphasised continuation between the 'Mesolithic' populations of Sulawesi and Austronesian incomers from Taiwan. One of the main pre-Austronesian technologies in South Sulawesi is termed the 'Toalean' and is applied to microlithic assemblages which date from ~8,000 BP to ~1,500 BP and which include a range of stone points and blades as well as bone points and shell objects. The Toalean people appear to have had a hunting culture which included the use of spears and bows and arrows. The area over which the Toalean sites are found is roughly the same as the area in which the Austronesian, Makassar languages are spoken today. Bulbeck *et al.* (2001) provide a number of pieces of evidence which suggest that the Toalean hunter-gatherers interacted directly with the Austronesian farmers. For example, earthenware potsherds are found in late Toalean sites which suggests the use of ceramics had been adopted from the newly-arrived population. Rice and bovid bones have also been found associated with typical Toalean assemblages, suggesting they too were adopted by the indigenous peoples. In contrast, no evidence of farming sites has been found in South Sulawesi prior to ~2,000 BP which implies that (contrary to Bellwood 1997) the possession of a farming culture did not enable any Neolithic immigrants to simply supplant the indigenous hunter-gatherers (Bulbeck *et al.* 2001).

1.4 The Origins of Modern Southeast Asians – Cranial Morphology

The analysis of cranial morphology generally involves studying a number of standard cranial measurements from a number of populations and applying multivariate statistical

analysis to ascertain the relationships between the various populations. Work on cranial morphology in Asia has suggested that populations from Island Southeast Asia cluster with those from mainland Southeast Asia and that Polynesian groups form a separate cluster between Southeast Asia and Melanesia. They do not appear to be connected with the Taiwanese and Chinese groups studied (Pietrusewsky 1997).

1.5 The Origins of Modern Southeast Asians – Genetic Evidence

In recent years this debate has been added to by an increasing amount of genetic data. Most of this work has concentrated on the Polynesian peoples, little work has so far been done on the potential source populations. Work on globin genes used deletions in α-globin genes, which lead to forms of α-thalassemia, as potential migration markers (Hill *et al.* 1985). The most common deletion in Vanuatu and Polynesia was found to be $-\alpha^{3.7}$III (a deletion of 3.7 kb between two normal α-globin genes) which is also found in coastal Papua New Guinea but not, so far, in Southeast Asia, which suggests that it arose in Melanesia. It seems likely that this event happened a substantial amount of time ago as some Melanesian individuals possess a further mutation on top of the $-\alpha^{3.7}$III deletion, leading to Hb J Tongariki (Old *et al.* 1978). This evidence combines to suggest that, even if the ancestral Polynesians did not originate in Melanesia, they must have passed through it slowly enough to allow time for them to pick up the $-\alpha^{3.7}$III deletion. However, the most common deletion in coastal Papua New Guinea, which leads to a 4.2 kb deletion, is not found in Polynesia suggesting that the ancestral Polynesians may have assimilated or evolved the $-\alpha^{3.7}$III deletion somewhere in island Melanesia.

Early work on mtDNA in Polynesia found an extremely high level (93%) of a 9 base pair deletion in the COII/$tRNA^{Lys}$ intergenic region (Hertzberg *et al.* 1989) which had previously been suggested as an East Asian marker (Wrischnik *et al.* 1987) and which is now known to characterise haplogroup B (Torroni *et al.* 1992). The deletion was also found in Japan, Korea, the Philippines (Harihara *et al.* 1992), Fiji, in the Tolais of New Britain and in coastal Papua New Guinea but was absent in highland New Guinea (Hertzberg *et al.* 1989). This roughly mirrors the modern distribution of Austronesian languages and was used to suggest that the deletion could be used as a marker for the

Austronesian expansions. The populations of highland Papua New Guinea would therefore be representative of the indigenous pre-Austronesian groups, an inference supported by Stoneking *et al.* (1990) who found that populations from highland Papua New Guinea had higher levels of mtDNA diversity than coastal populations and were therefore likely to be older.

Partial sequencing of the first hypervariable segment (HVS-I) of the mtDNA control region enabled Hagelberg and Clegg (1993) to identify transitions at nucleotide positions 16217, 16247 and 16261 in four modern Tahitians and also in ancient DNA from the remains of two prehistoric Polynesians (one of whom dated to 400 BP and another which was undated but pre-European) all of whom also exhibited the 9 base pair deletion. However, the 9 base pair deletion and its related substitutions were not found in earlier remains associated with the Lapita culture which is traditionally linked to the Austronesian expansions. Hagelberg and Clegg (1993) used these findings to suggest that the first inhabitants of the central Pacific originated in Melanesia and that the Lapita culture should be connected to indigenous Melanesians and not Austronesian-speaking migrants thus supporting Terrell (1981). However, these findings were hampered by a small sample size and the fact that the use of ancient DNA is fraught with difficulty and potential contamination issues (see e.g. Kolman and Tuross 2000).

The three transitions discussed above, along with a fourth transition at nucleotide position (np) 16189, are now popularly known as the 'Polynesian motif' (Redd *et al.* 1995) and are characteristic of haplogroup B4a1. They were also found in Hawaii, Samoa, Tonga and Micronesia by Lum *et al.* (1994), the ancestral sequences consisting of either transitions at np 16189, np 16217 and np 16261, or just at np 16189 and np 16217 were also found in Polynesia, Malaysia, Indonesia, China and Japan. Lum *et al.* (1994) also found a group of Polynesian and Papua New Guinean mtDNAs with transitions at np 16223, np 16241 and np 16311 and also an A to C transversion at np 16265 (now known to be characteristic of haplogroup Q), the distribution of which also supports a genetic link between Melanesia and Polynesia.

Redd *et al.* (1995) found haplogroup B at a frequency of 21% (22% of which had the Polynesian motif) in their samples from the Moloccas (Hiri and Ternate) and Nusa

Tenggara (Alor, Flores, Roti and Timor). They also found it at a frequency of 0% in highland Papua New Guinea, 42% in coastal Papua New Guinea (74% of which had the motif) and 100% in Samoa (79% with the motif). This data (along with an average substitution rate for the control region of $1.142 \pm 0.333 \times 10^{-7}$ substitutions per site per year per lineage) was used to obtain an age of 58,000 years (95% confidence interval 12,000-104,000 years) for haplogroup B and one of 12,000 years (CI 900-23,000 years) for the Polynesian motif itself. Also found was a general decrease in sequence diversity from Indonesia through coastal Papua New Guinea to Samoa suggesting a sequence of successive founder events. The extreme lack of diversity in Samoa also reflects the recent nature of Polynesian colonisation. Redd *et al.* (1995) concluded that an initial population expansion carrying the deletion reached as far as Indonesia and that the final mutation of the motif (at np 16247) arose in Indonesia between 900 and 23,000 BP subsequent to which there was a final expansion out across Polynesia beginning in around 5,500 BP

Since then the Polynesian motif has been found in Vanuatu, the Cook Islands, the Marquesas and New Zealand (Sykes *et al.* 1995), Madagascar (Soodyall *et al.* 1995), the Trobriand islands near the New Guinea coast (Hagelberg *et al.* 1999), Kiribati, Nauru, Palau, the Marshall Islands and several more locations in Micronesia (Lum and Cann 2000). The motif was also found in ancient DNA extracted from the remains of 12 Easter Islanders by Hagelberg *et al.* (1994). It has, however, never been found in highland Papua New Guinea, West New Guinea (Irian Jaya) (Tommaseo-Ponzetta *et al.* 2002), Taiwan or the Philippines although the ancestral states are very common in the latter two.

Melton *et al.* (1995) again found haplogroup B at near fixation in Polynesia with the motif at 79.2%. The immediate precursor of the motif was found at its highest frequency and diversity in three Aboriginal groups of Taiwan, suggesting that this intermediate motif evolved there. Melton *et al.* (1995) suggested that their data supported the 'out of Taiwan' theory of Bellwood (1997) with the transition at np 16261 occurring in Taiwan and being carried out through the Philippines and into Island Southeast Asia around 6,000 BP The final transition at np 16247 would then have occurred somewhere in East Indonesia before the final migration eastward into

Polynesia. In this case, most modern Island Southeast Asians would be descended from these Taiwanese immigrants.

The 'out of Taiwan' theory also seemed to be supported by the data of Sykes *et al.* (1995) who again found haplogroup B at extremely high levels and extremely low diversity in Polynesia with the highest diversity in Taiwan where it had a minimum divergence estimate of 31,000 years thus suggesting an origin in Southeast Asia. However, Sykes *et al.* (1995) did draw attention to their caveat that the most common Polynesian haplotype is not found in Taiwan or the Philippines. Some Melanesian lineages were also found in Polynesia in the form of haplogroup Q, also found by Lum *et al.* (1994). This was found in all locations surveyed in Polynesia, except the Marquesas, at an overall level of 3.8% and was also found in Papua New Guinea and Vanuatu where it displays considerably higher levels of diversity and where it has a minimum divergence time of 22,000 BP suggesting it may have been introduced to Melanesia during the Pleistocene colonisation of the area.

Richards *et al.* (1998) added to the debate by performing a re-analysis of the Redd *et al.* (1995) data. Using the statistic ρ (Forster *et al.* 1996), the haplotype ancestral to the Polynesian motif was dated to approximately 30,000 years in Taiwan with the motif itself being dated to ~17,000 years in Eastern Indonesia, ~5,000 years in coastal Papua New Guinea, ~3,000 years in Samoa and ~1,000 years in the Cook Islands. All these estimates, however, did have wide 95% confidence limits and the number of samples available to date the motif in Eastern Indonesia was extremely limited. Nevertheless the results did not provide much support for the 'out of Taiwan' model. Instead they suggested that, as argued by Solheim *et al.* (in press; personal communication), the Polynesian expansion originated somewhere in Island Southeast Asia, probably somewhere between Southeastern Borneo and the Moloccas; modern Island Southeast Asians would, therefore, be descended from the original Pleistocene inhabitants.

Lum and Cann (2000) also argued in favour of Polynesian origins in Island Southeast Asia on the basis of lineage sharing – the majority of lineages they found in Remote Oceania were shared with populations from the Philippines or Indonesia but not mainland Asia.

Lum *et al.* (1998) used both mtDNA and nuclear genome short tandem repeat (STR) data in an attempt to broaden the picture of Polynesian ancestry. Using principal component (PC) analysis, they were able to show differences in the clustering patterns when using mtDNA and STRs. When looking at mtDNA, they found that all their Polynesian populations clustered together, with their closest affinities being to Island and mainland Southeast Asia. However, in the STR PC map, Polynesian populations were found to be intermediate between the Asian populations and those from Papua New Guinea. Lum *et al.* (1998) contend that this supports the 'out of Taiwan' model but that the STR data shows extensive male-biased, post-colonisation gene flow. This latter is explained by the hypothesis that most post-colonisation contact (and gene flow) would have been biased in favour of the male portion of the societies who would have been in control of trade.

Y chromosome studies have also cast doubt on the 'out of Taiwan' theory. Su *et al.* (2000) found only one haplogroup (their H6 – defined by the single nucleotide polymorphism [SNP] known as M122; now referred to as haplogroup O3) shared between Taiwan and Polynesia. They also found no evidence of a Melanesian contribution to the Polynesian gene pool, and suggested an origin for the Polynesians somewhere in Island Southeast Asia. Again this implied that modern Island Southeast Asians would be descended from the original Pleistocene inhabitants; however, this study looked at a limited number of SNPs and therefore may suffer from problems of ascertainment.

Kayser *et al.* (2000) found only three Y chromosome haplogroups in the Cook Islands. One of these was defined by a mutation at position 711 of the RPS4Y gene (now known as haplogroup C), which was found at a frequency of 82% in the Cook islands, 26% in coastal Papua New Guinea, 10-15% in Eastern Indonesia and 9-12% in island Papua New Guinea. In the Cook Islands and Papua New Guinea the mutation was consistently associated with the microsatellite deletion DYS390.3; only in Indonesia was the mutation found without the deletion. The age of the deletion was estimated to be 11,500 years (albeit with large confidence intervals and using a very fast mutation rate).

Kayser *et al.* (2000) used this to infer that this major Polynesian haplogroup originated in Melanesia, although an Eastern Indonesian origin seems equally likely.

The second major haplogroup found by Kayser *et al.* (2000) in the Cook Islands was haplogroup O3, defined by the SNP M122, which made up 7.1% of their sample. This haplogroup was also found at high frequencies in East and Southeast Asia and at lower levels in Melanesia but was absent in highland Papua New Guinea, therefore mirroring the distribution of the mtDNA 'Polynesian motif'. The age of the M122 mutation was estimated to be 11,100 years with an expansion signal at 6,000 years (again dated using a very fast mutation rate). It was therefore suggested that this second haplogroup represents an expansion from Southeast Asia, possibly with the Austronesian language family, in which Austronesian speakers moved slowly across Melanesia, mixing sufficiently with the indigenous inhabitants to pick up many 'Melanesian' genes.

Capelli *et al.* (2001) also found haplogroup C at high frequencies in Polynesia, as well as in Melanesia and Southeast Asia. It was also found in samples from mainland Asia, but the diversity in those samples was far lower, suggesting an origin in the south. This haplogroup was dated to more than 12,000 years indicating again that it was probably carried by the indigenous peoples of the region.

Haplogroup O3 (termed haplogroup L in their work) was also found at high frequencies in Polynesia, Island Southeast Asia and Taiwan by Capelli and colleagues; again, most of its diversity was found to be in the north. Capelli *et al.* (2001) used a neighbour-joining tree to suggest an origin in the north for haplogroup O3. This was based on the fact that chromosome types from the Ami, a Taiwanese aboriginal group, were distributed through all major groups in the tree. However, this type of analysis may not always be accurate as trees such as this do not allow ambiguities in the dataset to be visualised as would be the case with a network approach. Under the geographic origins postulated by Capelli *et al.* (2001) 93% of their Melanesian samples (which include both Papuan and Austronesian speakers) are assigned as having an indigenous, Pleistocene ancestry with an input of only 3.6% from the north. In Island Southeast Asia the contributions are 32% indigenous and 19.5% northern and in Polynesia they are 64% southern and 32% northern. Therefore over 60% of the Y chromosomes in the

region as a whole have an indigenous, pre-Neolithic origin and, in Melanesia at least, the acquisition of Austronesian languages seemed to have been mainly a cultural process, at least on the paternal line of descent.

Kayser *et al.* (2003) looked at Y-chromosome diversity in West New Guinea (Irian Jaya), which was found to be much lower than that of other populations, including Papua New Guinea. The only populations with a comparably low level were aboriginal Taiwanese and a group from the Cook Islands. Approximately 70% of West New Guinea males were found to belong to haplogroup M*, the main Melanesian Y-chromosome type, which is also found in Eastern Indonesia. Kayser *et al.* (2003) dated this haplogroup to ~8,200 years with evidence of a population expansion ~4,400 years ago. Two of the groups from the central/western highlands (the Dani and the Lani) were found to be almost fixed for haplogroup C2(M208) which is also found in Papua New Guinea, the Trobriand Islands and the Cook Islands. The Melanesian haplotypes from this haplogroup were found to date to ~4,800 years with an expansion at ~1,500 years. Haplogroup C2 was found in 9% of coastal/lowland West New Guinea samples but was not found in the highlands; this haplogroup has also been found in Papua New Guinea, the Moluccas and the Nusa Tenggaras. Haplogroup K(M230) was found in two samples from West New Guinea; this haplogroup is the major haplogroup of highland Papua New Guinea and is also found in coastal Papua New Guinea, New Britain, the Moluccas and the Nusa Tenggaras.

Kayser *et al.* (2003) also found that the Y-chromosome diversity in West New Guinea was much lower than the mtDNA diversity. They suggested that this could be due to extreme patrilocality and/or polygyny, both of which are known to occur in New Guinea society. Another possible explanation is recurrent warfare which would involve the deaths of many more male than female members of a group. Kayser *et al.* (2003) proposed that the Y-chromosome haplogroups M*, K(M230) and C2 are indigenous to Melanesia, and that haplogroups O1 and O3 are 'Austronesian'. The latter are present in Papua New Guinea (where they are more common on the coast) but not West New Guinea, thus mirroring a putative Austronesian expansion.

To complement the above work on human genetics in Island Southeast Asia and the Pacific, work has also been done on the mtDNA of the Pacific rat (*Rattus exulans*) (Matisso-Smith and Robins 2004). *R. exulans* is thought to originate in Island or peninsular Southeast Asia, and first appears in Remote Oceania associated with Lapita settlements. It is a distinct species to those introduced later by Europeans and so does not interbreed with them. Matisoo-Smith and Robins (2004) found three major haplogroups in their sample; which included ancient and modern samples from across Island Southeast Asia, Melanesia and Polynesia. One of these haplogroups was found only in the Philippines, Borneo and Sulawesi with no indication that it had ever spread east of this. The second was found from the Philippines through Halmahera and Melanesia out to the Santa Cruz islands. Matisoo-Smith and Robins (2004) suggest this could be connected with the prehistoric obsidian trade which is known to have occurred across this area. Finally, the third haplogroup was found across Polynesia and was only found elsewhere in Halmahera. Matisoo-Smith and Robins (2004) propose that this is consistent with the area of Wallacea around Halmahera being this origin of the Lapita culture, and therefore human expansions into the Pacific.

Work on the mtDNA of wild and domestic pigs has also suggested that the area around Halmahera may have been important for the dispersal of pigs into Near Oceania (Larson *et al.* 2005). The mtDNA of pigs from Halmahera, New Guinea, Vanuatu and Hawaii have been shown to form a monophyletic clade which is well separated from any other wild or domestic clusters. No connections have been found between this group and the Taiwanese wild boar, therefore providing no support for the 'out of Taiwan' hypothesis.

1.6 Malaysia and the Orang Asli

How is it possible for us to know what the early Pleistocene variation of Southeast Asia would have been like? This seems to be an important question to answer in light of the above arguments. The indigenous Semang groups of the Malay Peninsula are often characterised as the first inhabitants of the region and so could hold the key to answering this question.

The Malay Peninsula is home to an astonishing range of human biological diversity. The indigenous Orang Asli tribes, who make up 0.5% of the population, have traditionally been classified on the basis of language, culture, geographic location and anatomical traits (particularly hair type and skin colour) into three groups: Semang, Senoi and Aboriginal Malay (Carey 1976). Most Orang Asli speak Aslian languages which form part of the Southeast Asian branch of the Austroasiatic family. This family also includes languages spoken in North-Eastern India and Burma, Thailand, Indochina and the Nicobar Islands immediately north of Sumatra (Ruhlen 1987). The Semang speak Northern Aslian languages and live, or did until recently, in small, nomadic hunter-gatherer groups in the lowland rainforests. They resemble the Andamanese and Filipino Aeta in that they are short in stature with dark skin and woolly hair and are often classed with them as 'Negritos' (Bellwood 1997). The Senoi are traditionally swidden (slash and burn) farmers who live at higher altitudes than the Semang, speak Central Aslian languages, and are described as being taller in stature than the Semang, with lighter skin and wavy hair. The Aboriginal Malays are horticulturalists and fishers who resemble physically the *Melayu* Malays - the most numerous group in Peninsular Malaysia. Some Aboriginal Malay groups, such as the Semelai, speak Southern Aslian languages, while the majority, including the Temuan, speak Austronesian languages, as do the *Melayu* Malays.

There are a number of distinct models for the prehistory of the Malay Peninsula (Rayner and Bulbeck 2001), which differ in particular on the origins of the Senoi. The traditional 'layer-cake' view (summarised in Carey 1976) sees the three groups as being the products of three separate migrations into the peninsula. The first immigrants were thought to be the ancestors of the Semang who were perhaps also related to the colonisers of New Guinea and the Andaman Islands. The second wave brought the population ancestral to the Senoi who were argued to be related to either the Veddas of Southeast Asia or the Australian Aborigines. The Aboriginal Malays, who resemble other modern Southeast Asians, arrived in the final wave from Island Southeast Asia (Carey 1976; Fix 1995; Rayner and Bulbeck 2001).

Bellwood (1993, 1997) has used archaeological and linguistic evidence to propose that the Semang are descended from the pre-Neolithic Hoabinhian peoples who occupied the

interior rainforests of the peninsula from at least 10,000 BP These would have been part of a much broader 'Australo-Melanesian' or 'Old Melanesian' substratum throughout much of Southeast Asia who were later largely displaced by Neolithic newcomers from South China/Taiwan. Archaeological evidence shows that the Hoabinhians also occupied a number of sites on the western coast but the absence of any marine items at inland sites such as Gua Cha suggests that small groups of Hoabinhians lived inland all year round. Bellwood (1993) has argued that the beginning of the Neolithic at around 1,200–2,000 B.C coincided with a dramatic cultural change in methods of burial and also introduced a wide range of artefacts which have no precedent in the Hoabinhian culture. This led him to suggest that the Senoi are descended from Hoabinhian tribes who interbred with incoming Neolithic farmers who probably originated in the Ban Kao culture of Southern and Central Thailand and also brought with them the Aslian languages currently spoken by most Orang Asli groups.

The Austronesian languages spoken by some Aboriginal Malays are closely related to the modern Malay language and to other Malay–Chamic languages of West Borneo, Sumatra and coastal Vietnam, suggesting that they represent a separate migration from Island Southeast Asia (Bellwood 1993). There is some evidence for forest clearances in Sumatra from around 2,000 B.C. but little evidence of a cultural change in the peninsula until the arrival of bronze and iron metal-working and new artifactual styles after 500 B.C. It is therefore suggested by Bellwood (1993) that the Aboriginal Malays represent the descendents of a migration from Sumatra at some point after 2,000 BP

By contrast, several models argue the case for a largely indigenous origin for the Senoi. Solheim (1980) argued that the Semang are descended from Hoabinhians who lived on the coast, and that the Senoi are descended from those who lived inland, with some subsequent admixture from the incoming Aboriginal Malays who passed to them their Neolithic culture. Rambo (1988), on the other hand, believes that the ancestors of both the Semang and Senoi lived on the coast. The first groups to inhabit the interior mountains would have been subject to new environmental and lifestyle pressures as they adapted to swidden farming, thus causing them to diverge into a distinctive group, the Senoi. The Semang evolved from the populations that remained in the lowland forests and traded forest products for tools and food. The Aboriginal Malays represent a

comparatively recent migration from Indonesia who have undergone some gene flow with the Semang and Senoi. Benjamin (1985, 1996) has similarly argued the case for local continuity.

Dental, skeletal and cranial data have been used to investigate the biological affinities of the Orang Asli tribes (Bulbeck 1996, 2000; Rayner and Bulbeck 2001). Bulbeck (1996) found that both Hoabinhian and Ban Kao individuals were substantially taller than hinterland Neolithic and modern Orang Asli populations. He also used cranial evidence to suggest that the Semang and Senoi had a common origin and began diverging in the early Holocene as a result of differing selection pressures, following the adoption of agriculture and also gene flow between the Senoi and the expanding Malays. Later 'Mongoloid' genetic contributions to the Senoi led to their characteristic cranio-facial traits. Dental evidence also showed similarities between the Hoabinhians and Neolithic populations, suggesting that the latter represented an expansion of a local population. Again, some 'Mongoloid' inheritance was postulated for the Senoi (Bulbeck 2000).

More recently, Rayner and Bulbeck (2001) proposed that the Semang and New Guinea populations have retained an ancient dental morphology similar to that of Europeans and North Africans. As other Southeast Asian and Pacific populations seem to have diverged from this pattern, this supports the idea that the Semang represent the first settlers of the Malaysian Peninsula. The Aboriginal Malays closely resembled other Island Southeast Asian, Polynesian and Micronesian populations, indicating that they represent a later migration into the Peninsula, while the Senoi were intermediate between the two groups, again suggesting that they are the product of Southeast Asian influence on proto-Semang populations.

Bulbeck (2004) has recently proposed another model for the evolution of the Orang Asli. He suggests that the Hoabinhian peoples foraged along well-defined jungle trails which were disrupted by later incoming populations. In his view, the Semang are descended from groups which adapted to these reduced trails and maintained a living in the jungle. The Senoi are descended from groups which established remote farming communities in the gaps created in the Hoabinhian trails. Finally the Aboriginal Malays are descended

from the groups which lived along the rivers and coasts and who were involved in the growing international trade routes.

Few studies have examined the mtDNA of Orang Asli populations. Ballinger *et al.* (1992) included samples from 32 Orang Asli (mostly Senoi) in their study. Five of their samples formed one of the few population-specific clusters in their phylogeny, a subgroup of haplogroup M defined by the gain of an *Alu*I site at np 10143. Other samples belonged to haplogroups M*, N*, B and F. Ballinger *et al.* (1992) suggest that their results show close similarities between the Orang Asli, the Austronesian-speaking Sabah aborigines of Northern Borneo and the inhabitants of coastal Papua New Guinea which implies that at least some of the Orang Asli could be the result of Austronesian migrations into the peninsula.

Melton *et al.* (1995) included 30 Senoi in their study. The 9 base pair deletion in the COII/tRNA^{Lys} intergenic region which is characteristic of haplogroup B was found in 36.7% of these samples, all of these belonged to the subgroup B4a which is further characterised by a transition at np 16217. In contrast only 3% of the Orang Asli samples studied by Ballinger *et al* (1992) belonged to haplogroup B. Melton *et al.* (1995) present a neighbour-joining tree based on their haplogroup B samples which indicates that the Orang Asli samples cluster most closely with the Filipino samples followed by the Malay and Taiwanese, again suggesting a possible Austronesian influence.

Su *et al.* (1999) showed that the most frequent Y-chromosome clade in a small group of unidentified Orang Asli was haplogroup O2a (nomenclature of the Y Chromosome Consortium [2001]) a group present throughout Southeast Asia. Saha *et al.* (1995) and Gajra *et al.* (1997) used classical markers to propose that the Semai Senoi had undergone a long period of isolation. Saha *et al.* (1995) looked at red blood cell enzymes and plasma protein polymorphisms in 349 Semai Senoi. They found possible private alleles of both red cell glucose-6-phosphate dehydrogenase (G6PD) and 6-phosphogluconate dehydrogenase in the Senoi which suggests that they have been somewhat isolated genetically and have a long population history. This was supported by Gajra *et al.* (1997) who studied polymorphic sites on the apolipoprotein B100 gene

in 195 Semai Senoi and found a high frequency of one of the ancestral haplotypes which they suggest shows a long population history for the Orang Asli. Saha *et al.* (1995) also used genetic distance and principal component analysis to show that the Senoi cluster with Khmer and Javanese populations and have no real link with the Vedda. This supports the linguistic evidence which indicates a close relationship between the Senoi and the Mon-Khmer. The Senoi do not appear to be particularly close to the nearby Malays, Chinese and Indians which again suggests some degree of genetic isolation.

Fix (1995) used haemoglobin E (Hb-E) and ovalocytosis, both of which are thought to protect against malaria, to study the population structure of the Orang Asli. Both conditions are found in all three groups of Orang Asli albeit at different frequencies. According to Fix (1995), Hb-E appears to be a Mon-Khmer gene which again reflects the linguistic connection between the two groups. Ovalocytosis, however, occurs at high frequencies in small areas from Sulawesi to Papua New Guinea and may have been introduced into the Aboriginal Malays (in whom it is most common) through intermarriage with traders of Island Southeast Asia. It may then have passed into the Senoi and Semang via further gene flow from the Aboriginal Malays (Fix 1995; Fix 2000).

1.7 Island Southeast Asian Groups Investigated in this Study

The group from **Padang** (a city in west central Sumatra) belong to the Minangkabau ethnic group. They practise Islam and speak Minangkabau which is part of the Para-Malay branch of Malayan languages and which is currently spoken by around 6.5 million people. The Malayan languages make up part of the Malayic subgroup of Sundic languages which are themselves a branch of Western Malayo-Polynesian (www.ethnologue.com). The Minangkabau economy is based upon wet-rice agriculture and dry-plot market gardening. It is possible that the Minangkabau were originally of the same ethnic group as the Malays of the Malay Peninsula and Eastern Sumatra but have become differentiated due to isolation (LeBar 1972).

The group from **Medan**, the capital of Northern Sumatra, belong to the Batak ethnic group. These people were relatively isolated until the 19th Century and even now maintain a more traditional lifestyle than many Indonesian groups (Lebar 1972). They speak one of a number of Batak languages; these languages as a whole are spoken by ~5.8 million people and belong to the Sumatran group of Sundic languages (www.ethnologue.com). Some Batak groups practise Christianity, while others practise Islam and still others practise indigenous religious beliefs. The Batak economy is predominantly based on rice agriculture (LeBar 1972).

All other Sumatran groups included in this study (from **Pekanbaru, Palembang and Bangka**), as well as the group from **Banjarmasin** in Southern Borneo, are ethnic Malays. It has been suggested, based on the high level of linguistic diversity found there, that Western Borneo is the original homeland of the Malay. From there it is thought that Malayic speakers must have spread to Southern Sumatra and eventually across to the Malay Peninsula (Adelaar 2004). Ethnic Malays are fairly widespread across Western Indonesia and tend to speak types of Local Malay which are part of the Malayan group of languages, these in turn form part of the Malayic group of the Sundic languages. Banjar Malay (the type of Local Malay spoken in Banjarmasin) is spoken by ~2.1 million people, while Palembang Malay is spoken by ~500,000 people. Bangka is a dialect of Malay (which in turn is a branch of Local Malay); Malay is spoken by ~10 million people in Indonesia as a whole and ~40,000 people in Bangka (www.ethnologue.com).

The **Tenggerese** group included in this study are one of the minority groups of Java and live on the volcanic slopes of the Tengger Mountains in East Java. Traditionally they are thought to have descended from refugees who fled the fall of the Madjapahit kingdom in the early 16th Century (LeBar 1972). Speakers of the Tenggerese language, which is part of the Javanese branch of Sundic languages, number ~500,000 (www.ethnologue.com). The Tenggerese economy is mainly based on maize agriculture and on the sale of vegetables and potatoes, they have maintained their indigenous religious beliefs and have not yet adopted Islam or Christianity (LeBar 1972).

The group from Denpasar (the capital of **Bali)** which was included in this study belong to the Balinese ethnic group. The Balinese language is part of the Bali-Sasak branch of Sundic languages and is spoken by ~3.8 million people (www.ethnologue.com). Unlike most of Indonesia, Bali is predominantly Hindu; the economy is based on rice cultivation (LeBar 1972).

The group from **Mataram** (the capital of Lombok) belonged to the Sasak ethnic group which is the major ethnic group on the island. The Sasak language is part of the Bali-Sasak branch of the Sundic languages and is spoken by ~2.1 million people (www.ethnologue.com). The Sasak are divided up into the Waktu Telu and the Waktu Lima. The former tend to live in the more remote central and mountain villages and have maintained more of their traditional belief system than the latter who are Islamic, more numerous and tend to live in larger settlements (LeBar 1972).

The group from **Manado** in North-East Sulawesi belong to the Minhasa ethnic group. The Minhasa language (also known as Tombulu) is spoken by ~60,000 people. It forms part of the Minhasan group of languages, which in turn are one of the Sangir-Minahasan branches of Sulawesi languages; the Sulawesi languages are themselves a branch of Western Malayo-Polynesian (www.ethnologue.com). The Minhasans are mostly Christian and have a mainly rice-based economy (LeBar 1972).

The **Toraja** are the indigenous inhabitants of mountainous central Sulawesi. They speak Toraja-Sa'dan languages which are part of the Northern group of South Sulawesi languages, these in turn form part of the Sulawesi branch of Western Malayo-Polynesian (www.ethnologue.com). The group from **Palu** (the capital of central Sulawesi) can be considered to be Western Toraja (LeBar 1972) and speak Kaili. This makes up part of the Kaili-Pamona group of the West Central branch of Central Sulawesi languages, these again form part of the Sulawesi branch of Western Malayo-Polynesian (www.ethnologue.com).

Ujung Padang is also known as Makassar and is one of the major cities of South Sulawesi. The Makarrese speak the language of the same name which is spoken by ~1.6 million people in total and which forms part of the Makassar branch of South

Sulawesi languages. As stated above, these form part of the Sulawesi branch of Western Malayo-Polynesian (www.ethnologue.com). The Makarrese are thought to be descended from Toraja groups who have been undergone admixture with later incomers, mainly from Malay and Javanese sources (LeBar 1972).

The group from **Waingapu** (the largest town in Sumba) belong to the Sumbanese ethnic group. They speak Kambera (or East Sumbanese) which has ~200,000 speakers in total. It belongs to the Bima-Sumba group of Central Malayo-Polynesian languages (www.ethnologue.com). The Sumbanese mainly practice small-scale farming and seem to have maintained their indigenous belief systems (LeBar 1972).

The group from **Alor** belong to the ethnic group of the same name. They speak Alorese which is spoken by ~25,000 people in total and which belongs to the Flores-Lembata group of languages. These in turn belong to the Timor branch of Central Malayo-Polynesian (www.ethnologue.com). Maize agriculture is the main subsistence basis, with pig hunting still occurring in the dry season. There is a complex trade and exchange pattern established amongst mountain villages. The Alorese have mainly maintained their indigenous belief systems (LeBar 1972).

Ambon is the capital of Maluku province (also known as the Moluccas or Spice Islands). Ambonese, which is the primary language, is spoken by ~200,000 people and is a Malay-based Creole (www.ethnologue.com).

No detailed information was available about the ethnicity of the individuals from **Kota Kinabalu** or the **Philippines** used in this study.

1.8 Aims of this Study

The main aim of this study is to use mtDNA variation to try to clarify some of the issues discussed in the above sections. As discussed in section 1.5, much work has been done on the mtDNA of Polynesians; however, there is currently not much data on the potential source populations in Island Southeast Asia. Mitochondrial DNA has therefore been obtained from a number of populations from across Indonesia which

should enable any potential indigenous and Taiwanese maternal markers to be identified. As discussed in section 1.6, it should also be possible to use mtDNA from the Orang Asli groups of peninsular Malaysia to study the late Pleistocene variation of Southeast Asia.

The main laboratory techniques which will be used in this project are: the polymerase chain reaction (PCR), DNA sequencing and restriction fragment length polymorphism (RFLP) analysis. More specifically, PCR will be used to amplify HVS-I in all samples which will then be sequenced and aligned to the CRS to identify any mutations. If necessary, RFLP analysis, coding region and HVS-II sequencing will then be used to clarify haplogroup status.

The resulting data will be analysed in a number of ways in order to elucidate the picture of Island Southeast Asian ancestry. Principal component analysis will be used to visualise the distance between each population. Analysis of molecular variance (AMOVA) will be used to look for significant differences between the different groups studied in both the Malay Peninsula and Island Southeast Asia. Tajima's *D* will be used to detect any potential population expansions. These will also be visualised using mismatch distributions. Finally, phylogenetic networks will be constructed for each haplogroup which will enable phylogeographic analysis to be carried out. This data will be used to infer the timing of dispersal and expansion events in Southeast Asia and will be compared to the models for Southeast Asian ancestry discussed in the previous sections.

2. Materials and Methods

2. Materials and Methods

All chemicals and reagents were obtained from Sigma Aldrich (Poole, Dorset) except where stated.

2.1 Subjects

Buccal cells were obtained from 259 individuals from the Malay Peninsula using cytological brushes (Flowgen, Nottingham) after obtaining informed consent. 885 anonymous unlinked DNA samples were also obtained by Professor ASM Sofro from across Island Southeast Asia. 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. The number of samples obtained from each locale/ethnic group is shown in tables 1 and 2. These locations can be seen on the map in figure 3.

Table 1 – Number of samples obtained from each ethnic group from the Malay Peninsula

Ethnic Group	Number of Samples
Semang	
• Jahai	50
• Mendriq	31
• Batek	29
Total	110
Senoi	
• Temiar	51
• Semai	1
Total	52
Aboriginal Malay	
• Semelai	60
• Temuan	32
• Jakun	2
Total	94
***Melayu* Malay**	3

Table 2 – Number of samples obtained from each location in Island Southeast Asia

Location	Number of Samples
Sumatra	
• Padang (sample code PAD)	24
• Medan (MED)	42
• Pekanbaru (PEK)	52
• Palembang (PLB)	28
• Bangka (BGK)	34
Total	180
Java	
• Tengger (TGR)	36
Borneo	
• Kota Kinabalu (KK)	68
• Banjarmasin (BAN)	89
Total	157
Bali	
• Denpasar (BAL)	65
Lombok	
• Mataram (MTR)	44
Sulawesi	
• Ujung Padang (UJP)	46
• Toraja (TOR)	64
• Palu (PAL)	38
• Manado (MND)	89
Total	237
Sumba	
• Waingapu (WAI)	50
Ambon (AMB)	43
Alor (ALO)	45
Taiwanese Aboriginals	
• Bunun (BUN)	8
• Paiwan (PAI)	1
Total	9
Philippines (FIL)	19

Figure 3 – Map showing the locations from which samples were obtained

DNA was extracted from the Island Southeast Asian samples at the MRC Molecular Haematology Unit in Oxford. All post-extraction work on these samples was performed by myself. 90% of the Orang Asli samples were extracted at the University of Huddersfield by Will Meehan and James Blackburn, around 10% of the RFLP tests were carried out by Mike Ward as part of a final year project. The remainder of the extractions and RFLP tests were performed by myself.

2.2 DNA Isolation

DNA was extracted from buccal cells using the InstaGene matrix (BioRad, Hemel Hempstead) and the following protocol: the cytology brush (Flowgen, Nottingham) containing the sample of buccal cells was placed in a microfuge tube and 500 µl 50 mM NaOH added to it. The sample was then vortexed for 60 seconds before being incubated at 95 °C for 10 minutes. 50 µl 1 M Tris pH 8.0 was then added and the sample vortexed for 30 seconds before the addition of 200 µl InstaGene matrix. The sample was then incubated at 56 °C for 30 minutes before being vortexed for 10 seconds and incubated at 100 °C for a further 8 minutes. The sample was then vortexed for an additional 10 seconds before the cytology brush was removed; the remaining sample was centrifuged at 12,000 rpm for 3 minutes. The supernatant containing the DNA was removed for use in PCR and subsequent analyses.

2.3 PCR Amplification

The HVS-I of all the samples was amplified using the primers conH1 (5′- CCTGAA GTAGGAACCAGATG-3′) and conL1 (5′-TCAAAGCTTACACCAGTCTTGTAA ACC-3′). HVS-II was also amplified in selected samples using the primers L4 (5′ - GGTCTATCACCCTATTAACCAC-3′) and H4 (5′-CTGTTAAAAGTGCATACC GCCA-3′).

Selected samples were also amplified from np 10270 to np 10579 to check the status of np 10310; this was done using the primers mitM-F (5′-TCCTTTTACCCCTACCAT GAG-3′) and mitM-R (5′-ATTATTCCTTCTAGGCATAGTAG-3′). Other samples were amplified from np 8196 to np 9163 to check the status of np 8701; this was done using the primers mitB-F (5′-ACAGTTTCATGCCCATCGTC-3′) and mitWK-R (5′-CCTAGCCATGGCCATCC-3′). The conditions shown in table 3 were used when amplifying all these sections, and when amplifying the other fragments described in sections 2.5 and 2.6.

Table 3 – Protocol for general PCR

Reagent	Final Concentration
DNA template (diluted 1/10)	1 µl
Each dNTP	125 µM
Each primer	0.2 µM
10 x PCR buffer (including 15 mM $MgCl_2$)	1 x
Taq DNA Polymerase	1 unit
Final Volume	25 µl

	Temperature	Time
Initial Denaturation	94 °C	4 minutes
Denaturation	94 °C	1 minute
Annealing	53 °C	1 minute
Extension	72 °C	1 minute
		35 cycles
Final Extension	72 °C	8 minutes

The PCR products were electrophoresed for 15 minutes at 100 V on a 1.5% w/v agarose gel in 1 x Tris borate EDTA buffer (89 mM Tris borate, pH approx. 8.3, containing 2 mM EDTA). The gel was stained in a 1µg/ml ethidium bromide solution and visualised on an ultra-violet (UV) transilluminator.

2.4 DNA Sequencing

PCR products were purified using QIAquick PCR purification columns (Qiagen, Crawley, West Sussex). The Orang Asli samples and those from Medan and Pekanbaru were sequenced by the University of Dundee sequencing service using an ABI 3700 capillary sequencer. These samples were aligned to the Cambridge Reference Sequence (Anderson *et al.* 1981) using the SeqEd (Accelrys Inc, Cambridge) and Chromas 1.62 (http://www.technelysium.com.au/chromas.html) programs. All other samples were sequenced by myself using a Beckman-Coulter CEQ8000 sequencer. Purified PCR products were reamplified using the relevant forward or reverse primer and the mastermix (known as 'Quickstart') provided by Beckman-Coulter which contains DNA

polymerase, pyrophosphatase, buffer, dNTPs and dye terminators. This was done under the conditions shown in table 4.

Table 4 – Protocol for sequencing PCR

Reagent	
Purified PCR product	~ 50 ng
Primer	3.2 pmol
Quickstart mastermix	4 µl
Final volume	20 µl

	Temperature	**Time**
Denaturation	96 °C	20 seconds
Annealing	50 °C	20 seconds
Extension	60 °C	4 minutes
		} 35 cycles

The PCR products were then treated in the following way to precipitate out the DNA: 2 µl 3 M sodium acetate, 2 µl 100 mM EDTA, 1 µl glycogen and 60 µl 95% v/v ethanol were added to each sample, samples were then centrifuged at 13,000 x g for 15 minutes at 4 °C. The supernatant was removed and 200 µl 70% v/v ethanol added before the samples were centrifuged at 13,000 x g for 2 minutes at 4 °C. Again, the supernatant was removed and a further 200 µl 70% v/v ethanol added before a second centrifugation at 13,000 x g for 2 minutes at 4 °C. The supernatant was removed and the samples dried until all ethanol was removed. 40 µl of the 'Sample Loading Solution' provided by Beckman-Coulter was then added before transferring the samples to a 96-well plate for sequencing. Samples were then sequenced using the Beckman-Coulter CEQ8000 capillary sequencer. Samples were aligned to the CRS (Anderson *et al.* 1981) using the software provided on the CEQ8000.

2.5 Testing for the 9 Base Pair Deletion within the COII/tRNA^{Lys} Intergenic Region

The 9bp deletion which is characteristic of haplogroup B was detected by PCR amplification with the primers mitB-F (5′-ACAGTTTCATGCCCATCGTC-3′) and mitB-R (5′-ATGCTAAGTTAGCTTTACAG-3′), these amplify the region from np 8196 to np 8316 (Wrischnik *et al.* 1987). PCR products were run on a 3% w/v agarose gel in 1 x Tris borate EDTA buffer (89 mM Tris borate, pH approx. 8.3, containing 2 mM EDTA) for at least 30 minutes at 80 V. The gel was stained in a 1μg/ml ethidium bromide solution and visualised on a UV transilluminator. The PCR products amplified from mtDNAs carrying the deletion form 112 bp fragments while the PCR products from mtDNAs without the deletion form 121 bp fragments.

2.6 Restriction Fragment Length Polymorphism Tests

RFLP screening was used to resolve haplogroup status in hierarchical fashion as follows: haplogroup M (+10397 *Alu*I, +10394 *Dde*I), N (–10397 *Alu*I, –10394 *Dde*I), M7 (+9824 *Hin*fI), D (-5176 *Alu*I), E (-7598 *Hha*I), G (+4830 *Hha*I), P (+15606 *Alu*I), M15 (-9052 *Hha*I), M16 (+5351 *Hha*I), M17 (+10054 *Hin*fI), M18 (+10143 *Alu*I), U (+12308 *Hin*fI), I (+10032 *Alu*I). M15, M16, M17 and M18 were identified from Ballinger *et al.* (1992). NB. '+' means the gain of a restriction site whereas '–' means the loss of a restriction site.

The samples were amplified using the relevant primers (shown in table 5) and the conditions shown in table 3. The PCR products were electrophoresed for 15 minutes at 100 V on a 1.5% w/v agarose gel in 1 x Tris borate EDTA buffer (89 mM Tris borate, pH approx. 8.3, containing 2 mM EDTA). The gel was stained in a 1μg/ml ethidium bromide solution and visualised on a UV transilluminator. Successful PCR products were then digested with the relevant restriction enzyme (New England Biolabs, Hitchin, Hertfordshire) (see table 5).

Table 5 – Details of RFLP tests

Digest	Primers	Region Amplified	Enzyme Used	Positive Result	Negative Result
M	mitM-F = 5`-TCCTTTTACCCC TACCATGAG –3` mitM-R = 5`-ATTATTCCTTCT AGGCATAGTAG-3`	np 10270 – np 10579	*Alu*I	Bands of 179 bp and 131 bp	One band of 310 bp
N	As M	As M	*Dde*I	Bands of 222 bp and 87 bp	Bands of 181 bp, 87 bp and 41 bp
M7	mitM7-F = 5`-CGCATCAGG AGTATCAATCACC-3` mitM7-R = 5`-TATTAGTTG GCGGATGAAGC-3`	np 9620 – np 9878	*Hin*fI	Bands of 132 bp, 72 bp and 54 bp	Bands of 132 bp and 126 bp
D	mitD-F = 5`-CTACTATCTCGC ACCTG-3` mitD-R = 3`-TAGGAGTAG CGTGGTAA-3`	np 5154 – np 5480	*Alu*I	One band of 326 bp	Bands of 304 bp and 22 bp
E	mitE-F = 5`-CTCCATAAACCT GGAGTG-3` mitE-R = 5`-GTAAAGGAT GCGTAGGGATG-3`	np 7367 – np 7840	*Hha*I	One band of 473 bp	Bands of 242 bp and 231 bp
G	mitGV-F = 5`-GGAGCTTAA ACCCCCTTA-3` mitGV-R = 5`-GGATAAGAT TGAGAGAGT G-3`	np 4326 – np 4934	*Hha*I	Bands of 505 bp and 103 bp	One band of 608 bp
P	mitP-F = 5`-CTTACTTCTCTT CCTTCTCTCC-3` mitP-R = 5`-TTAGAATGA GGAGGTCTGCGGC-3`	np 15439 – np 15752	*Alu*I	Bands of 167 bp and 146 bp	One band of 313 bp
M15	mitWK-F = 5`-CCTAGCCAT GGCCATCC-3` mitWK-R = 5`-GGCTTACTA GAAGTGTGAAAA C-3`	np 8846 – np 9163	*Hha*I	One band of 317 bp	Bands of 206 bp and 111 bp
M16	mitD-F = 5`-CTACTATCTCGC ACCTG-3` mitD-R = 3`-TAGGAGTACGT GGTAA-3`	np 5154 – np 5480	*Hha*I	Bands of 197 bp and 129 bp	One band of 326 bp
M17	mitSM-F = 5`-CTGTATGTC TCCATCTAT TG -3` miSM-R = 3`-TTAGTGGCA GGTTAGTTGTT -3`	np 9960 – np 10317	*Hin*fI	Bands of 263 bp and 94 bp	One band of 357 bp
M18	As M17	As M17	*Alu*I	Bands of 183 bp, 89 bp and 85 bp	Bands of 272 bp and 85 bp
U	mitU-F = 5`-CTCAACCCC GACATCATTACC-3` mitU-R = 5`-ATTACTTTTATT TGGAGTTGCACCAAGATT-3`	np 12104 – np 12338	*Hin*fI	Bands of 204 bp and 30 bp	One band of 234 bp
I	mitI-F = 5`-TTCGAAGCCGCC GCCTGATACTGG-3` mitI-R = 5`-GTAGTAAGGCTA GGAGGGTG-3`	np 9909 – np 10107	*Alu*I	Bands of 123 bp and 75 bp	One band of 198 bp

This was done by incubating the following at 37 °C for 2 hours:

• DNA sample	5 - 8.5 µl (depending on brightness of band)
• Restriction enzyme	0.5 µl
• Buffer	1 µl
Total	10 µl

Most digests were then run on 1.5% w/v agarose gels for at least 30 minutes before staining and visualising as before. The only exceptions were the tests for haplogroups M7, D and U which were run on 3% w/v agarose gels.

2.7 Verification of the Dataset

All sequence traces were read by two people to ensure accuracy of reading. The results were compared to the existing worldwide database (Martin Richards, personal communication) and, where possible, included in the existing scheme of worldwide haplogroups. Any unusual mutations, such as transversions or transitions at sites with a low relative mutation rate were rechecked. The data from each haplogroup was run through the programme 'netmat' (courtesy of Vincent Macaulay, University of Glasgow) and all sites between np 16051 and np 16365 which undergo transitions at least as fast as the average transitional rate were filtered out. A median network of the remaining sites was drawn using Network 4.1 (http://www.fluxus-engineering.com/sharenet.htm). Any mutations which led to reticulations in the networks were rechecked and the dataset amended if necessary.

2.8 Analysis

2.8.1 Comparative Data

Comparative data has been taken from the literature, and mostly comprises HVS-I sequence data, often with only the 9-bp deletion in the COII/tRNA^{Lys} region included (the data of Yao *et al.* 2002a, Yao *et al* 2002c and Kivisild *et al.* 2002 excluded). HVS-I data alone cannot always be resolved clearly into mtDNA haplogroups and was therefore not included in the PC analyses, although in many cases sufficient motif

information is present to include them in phylogenetic analyses of particular haplogroups or subclades. The data used included samples from: Thailand, Malaysia, Taiwanese aboriginals, the Philippines, Sabah, East Indonesia, Papua New Guinea, Pacific islanders, the Nicobars, Taiwanese Han, Hong Kong Han, China, Japan, Mongolia, Korea, Central Asia, and unpublished data from Singapore and Irian Jaya (Betty *et al.* 1996; Comas *et al.* 1998; Fucharoen *et al.* 2001; Hertzberg *et al.* 1989; Horai and Hayasaka 1990; Horai *et al.* 1996; Kivisild *et al.* 2002; Kolman *et al.* 1996; Lee *et al.* 1997; Lum *et al.* 1998; Melton *et al.* 1998; Nishimaki *et al.* 1999; Oota *et al.* 2001; Pfeiffer *et al.* 1998; Prasad *et al.* 2001; Qian *et al.* 2001; Redd *et al.* 1995; Seo *et al.* 1998; Sykes *et al.* 1995; Yao *et al* 2000; Yao *et al* 2002a; Yao *et al.* 2002b; Yao *et al* 2002c; Zainuddin and Goodwin 2004; Martin Richards, pers. comm.; Peter Forster, pers. comm.).

2.8.2 Measures of Diversity

Intragroup heterozygosity was calculated as $1 - \Sigma x_i$, where x_i is the relative frequency of the *i*th haplotype (Torroni *et al.* 2001). This was calculated for the region between np 16090 and np 16365.

Haplogroup diversity was calculated using ρ, which is defined as the average number of sites which differ between a set of sequences and their common ancestor (Forster *et al.* 1996; Saillard *et al.* 2000). This was calculated using a mutation rate of one transition every 20,180 years in the region from np 16090 to np 16365.

2.8.3 Principal Component Analysis

Principal component analysis was used to visualise the distance between each population. This was performed using the programme POPSTR (courtesy of Henry Harpending, University of Utah). The analysis was performed on the following combinations of data:

- All data from Island Southeast Asia found in this study along with the Chinese data of Kivisild *et al.* (2002), Yao *et al.* (2000), Yao *et al.* (2002a),

Yao *et al.* (2002b), Yao *et al.* (2002c), the Taiwanese data of Tajima *et al.* (2003) and Melton *et al.* (1998) which has also been elaborated on in this study, and unpublished data from the Philippines (Martin Richards, personal communication).

- All data from Island Southeast Asia, Taiwan and the Philippines.
- All Orang Asli and Sumatran data from this study along with unpublished data from Irian Jaya (Martin Richards, personal communication) and the Malay data of Zainuddin and Goodwin (2004).
- All Orang Asli and Sumatran data from this study along with the Malay data of Zainuddin and Goodwin (2004).

For each cycle of principal component analysis performed, a second analysis was carried out to plot the contribution of each haplogroup to each principal component.

2.8.4 Tajima's *D*

Tajima's *D* (Tajima 1989) is a test of selective neutrality which can also be used to detect population expansions. It compares the number of segregating sites (S) to π (nucleotide diversity). Under neutrality these should be equal and therefore *D* should equal zero. If a system departs from neutrality then *D* becomes either positive or negative; a significantly negative result can also indicate that a population expansion has occurred. Tajima's *D* was calculated for all populations using the programme DnaSP (http://www.ub.es/dnasp/).

2.8.5 Mismatch Distributions

Mismatch distributions also enable population expansions to be detected. They are constructed from a matrix of pairwise distances; the mismatch distribution is a histogram plotting the number of pairwise differences against their respective frequencies. The shape of the distribution gives insights into the population's history in that a smooth, bell-shaped distribution indicates a population expansion while a ragged, multimodal distribution indicates constant population size (Rogers and Harpending

1992). The raggedness statistic (r) is used to quantify the shape of the distribution – the lower r is, the smoother the distribution. r is defined as the sum of the squared difference between neighbouring peaks in the distribution. Mismatch distributions were constructed using DnaSP (http://www.ub.es/dnasp/) which was also used to calculate r.

Any expansions found in the mismatch distributions were dated using the statistic τ. $\tau = 2ut$, where t is time in generations, τ measures time in units of $1 / 2u$ generations, and u is the mutation rate of the total region of DNA being studied (Rogers and Harpending 1992).

2.8.6 Analysis of Molecular Variance

Analysis of molecular variance (or AMOVA) enables population diversity to be calculated for different hierarchic levels: among groups, among populations within groups, and within populations (Excoffier *et al.* 1992). AMOVA takes into account both haplotype frequencies and the genetic distance between haplotypes.

Successive cycles of AMOVA were performed on the following groups constructed from the dataset (significance was calculated using 1000 permutations):

- All Island Southeast Asian populations compared to each other.
- Eastern Island Southeast Asian populations (those from Mataram, Ujung Padang, Toraja, Palu, Manado, Waingapu, Ambon and Alor) compared to those from West of the Wallace line (Pekanbaru, Medan, Padang, Palembang, Bangka, Tengger, Banjarmasin, Kota Kinabalu and Bali).
- Eastern groups (Ujung Padang, Toraja, Palu, Manado, Waingapu, Ambon and Alor) compared to those from central (Tengger, Banjarmasin, Kota Kinabalu, Bali and Mataram) and Western (Pekanbaru, Medan, Padang, Palembang and Bangka) areas.
- Groups speaking Western Malayo-Polynesian languages compared to those speaking Central Malayo-Polynesian languages and the population from Ambon who speak a Malay-based Creole.
- A second language-based analysis with the Ambonese included in the Western Malayo-Polynesian group.

- All Orang Asli groups compared to each other.
- Orang Asli groups compared to those from Island Southeast Asia.

Matrices of pairwise F_{ST} values were also constructed to analyse the differences between individual Orang Asli and Island Southeast Asian populations.

All above analyses were performed using Arlequin (Schneider *et al.* 2000).

2.8.7 Phylogeographic Analysis

A phylogeographic approach was used to study the distribution of mitochondrial DNA lineages within haplogroups. This integrates the genetic data with its geographic distribution. Examining the geographic distribution of lineages in a network often enables conclusions to be made about prehistoric migrations and other demographic events (see e.g. Macaulay *et al.* 1999 and Torroni *et al.* 2001).

To do this each haplogroup was studied individually. For each haplogroup a reduced median network was constructed manually and verified using the program Network 4 .1 (http://www.fluxus-engineering.com/sharenet.htm). In each network, the size of a node is proportional to the number of samples it represents. If there is an incompatibility between two or more samples, this is shown as a reticulation in the network. These can range from 4-cycles which are used to represent incompatibilities between four sequence types to cubes which can show incompatibilities between up to eight sequence types. In a reticulation, the nucleotide pair at which a parallel mutation occurs is included on only one of the relevant branches.

Each haplogroup was dated as a whole using the statistic ρ, which was described above, again this was calculated using a mutation rate of one transition every 20,180 years in the region from np 16090 to np 16365. If the network was very reticulated then nucleotide positions 16093, 16129, 16189, 16223, 16311 and 16362 were downweighted by one-fifth to try to resolve the network. The geographic distribution of lineages over the network was examined and any subclades within its distribution which were specific to Island Southeast Asia were also dated. It is obviously of interest to

calculate the ages of certain haplogroups in different locations; therefore, where there was sufficient data, haplogroups were dated separately in South China, Taiwan, the Philippines and Island Southeast Asia (and also, in certain cases, in Thailand and Papua New Guinea).

3. Results – Laboratory Work

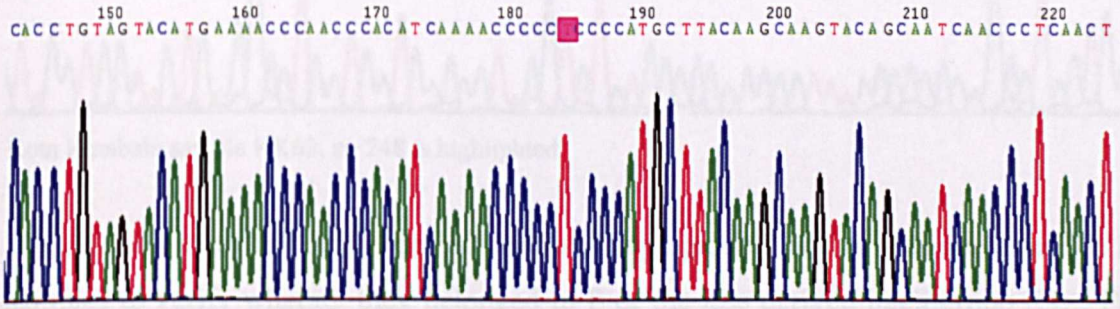
3. Results – Laboratory Work

3.1 DNA Sequencing

As stated in section 2.1, 1144 samples were sequenced and/or restriction tested during this study. The HVS-I of all samples was sequenced as described in section 2.3 and was aligned to the CRS (Anderson *et al.* 1981) using either the SeqEd and Chromas 1.62 programs (Orang Asli, Medan and Pekanbaru samples) or using the software provided on the Beckman Coulter CEQ8000 (all other samples).

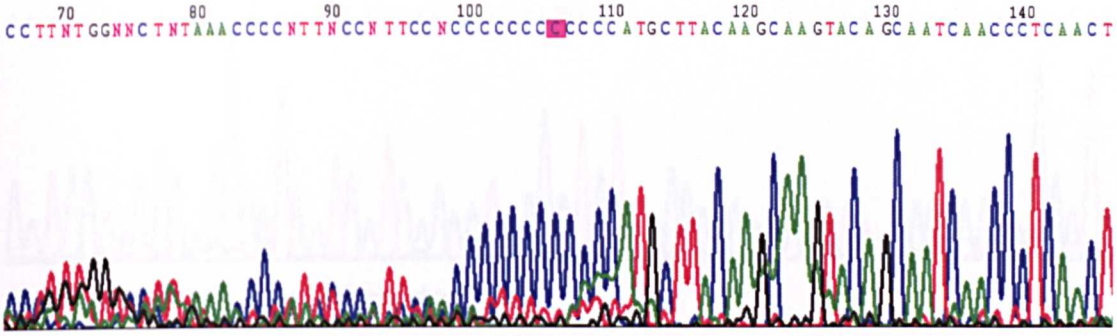
Roughly 10% of the samples were resequenced to act as checks/controls. For most of the samples sequenced (~70%) it was only necessary to sequence the forward strand of HVS-I. However, the remaining ~30% of the samples possessed a transition at np 16189 (the defining HVS-I mutation in haplogroup B but also found in other haplogroups) which causes the sequencing trace to be unreadable after this point (see figures 4 and 5), meaning that the reverse strand also needed to be sequenced.

Figure 4 – Chromatograph illustrating a sample without the np 16189 transition



Orang Asli sample 38A, np 16189 highlighted

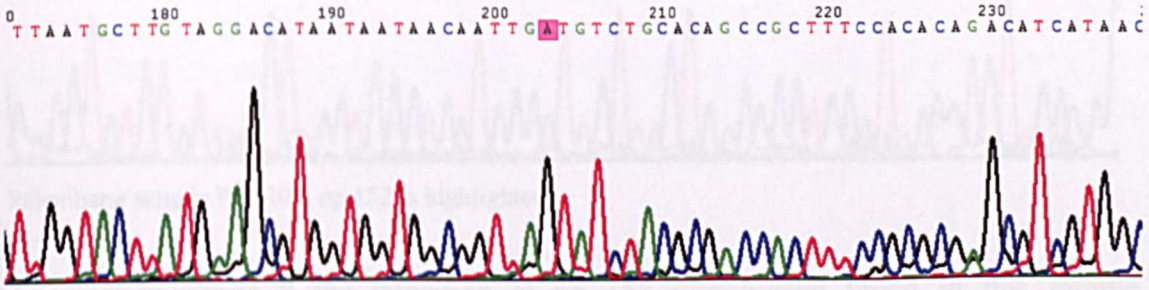
Figure 5 – Chromatograph illustrating the np 16189 transition



Pekanbaru sample PEK21, reverse strand, np 16189 highlighted

Other areas of the mitochondrial genome were also sequenced in certain samples. Figure 6 shows a section of the second hypervariable region (HVS-II) of sample KK63, sequenced to look for the deletion of an adenine at np 249. This deletion was indeed found and thus the sample was confirmed as belonging to haplogroup C.

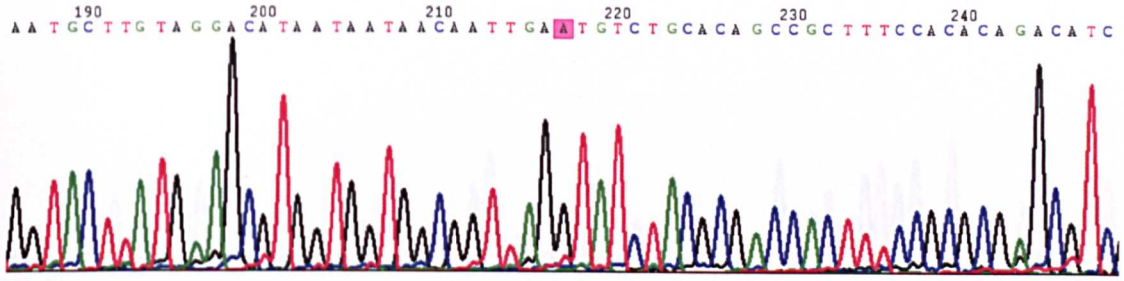
Figure 6 – Chromatograph illustrating the np 249 deletion



Kota Kinabalu sample KK63, np 248 is highlighted

The deletion at np 249 is also characteristic of haplogroup F and so was sequenced in 12 samples to verify whether they belonged to F or the less derived haplogroup R9. For example, sample ALO63 was found to lack the deletion at np 249 and so was assigned to haplogroup R9 (figure 7).

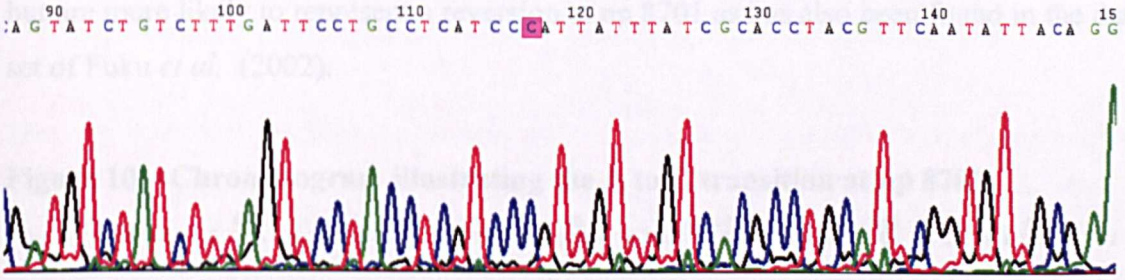
Figure 7 – Chromatogram illustrating the presence of np 249



Alor sample ALO63, np 249 is highlighted

The HVS-II of one sample (PLB108) which was thought to belong to haplogroup Z was also sequenced. Haplogroup Z shares the deletion at np 249 with haplogroup C and has a further transition at np 152 (figure 8).

Figure 8 – Chromatogram illustrating the T to C transition at np 152

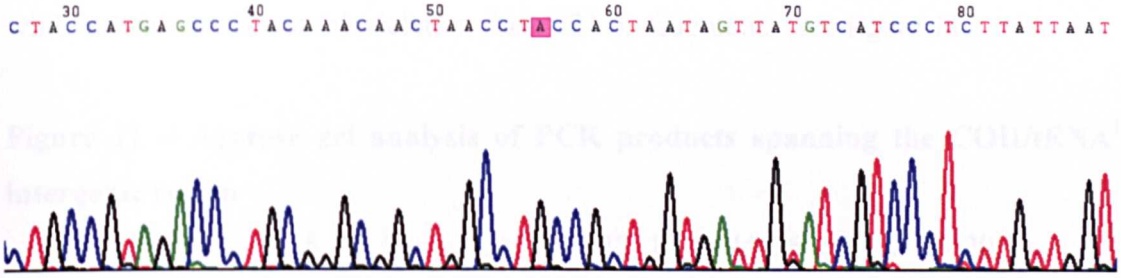


Palembang sample PLB108, np 152 is highlighted

As seen in figure 8 the transition at np 152 was indeed found in this sample, demonstrating that it does belong to haplogroup Z.

The segment from np 10270 to np 10579 was also sequenced in three samples to look for the presence of a transition at np 10310 which is also characteristic of haplogroup F (see figure 9). This section of chromatogram highlights the presence of the np 10310 transition in sample MND54 which therefore belongs to haplogroup F.

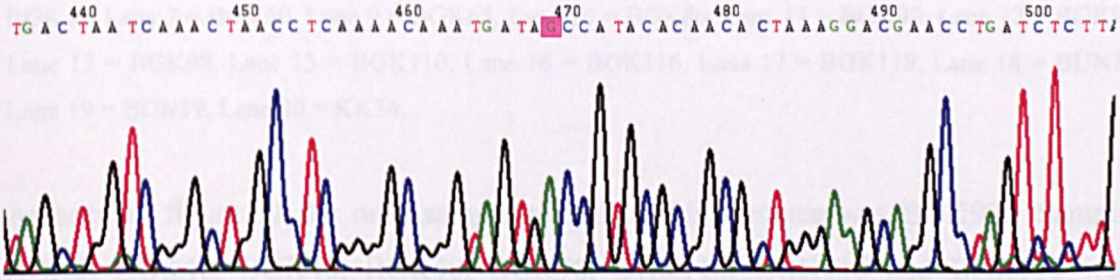
Figure 9 – Chromatogram illustrating the G to A transition at np 10310



Manado sample MND54, np 10310 highlighted

Finally, the stretch between np 8196 and np 9184 was sequenced in a further three samples (Orang Asli samples: 114A, 115A and 122B) which were thought to belong to a previously unknown branch of macrohaplogroup N. This showed that they all have a transition at np 8701 relative to the CRS, and therefore have the ancestral state at that position (figure 10). They could therefore be a one-step ancestor of macrohaplogroup N but are more likely to represent a reversion at np 8701 as has also been found in the data set of Fuku *et al.* (2002).

Figure 10 – Chromatogram illustrating the A to G transition at np 8701



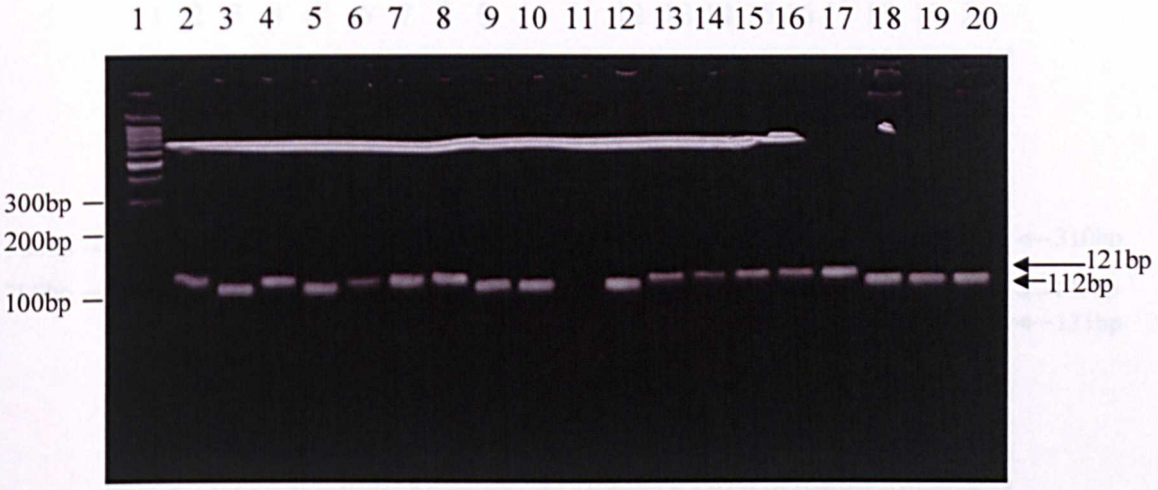
Orang Asli sample 114A, np 8701 highlighted

3.2 Detecting the 9 Base Pair Deletion in the COII/tRNA^{Lys} Intergenic Region

As discussed in section 3.1, haplogroup B is defined by a transition at np 16189. However, as np 16189 has a relatively high mutation rate, the same transition is also found in other haplogroups. Therefore, some potential haplogroup B samples were also tested for the 9 bp deletion in the COII/tRNA^{Lys} intergenic region which is also characteristic of haplogroup B (Hertzberg *et al.* 1989). This was detected in this study by PCR amplification with the primers mitB-F and mitB-R and electrophoresis of the products on a 3% w/v agarose gel (see section 2.5; Wrischnik *et al.* 1987 and Hertzberg

et al. 1989). In this process, the mtDNAs hosting the deletion form 112 bp fragments while mtDNAs without this deletion form 121 bp fragments (see figure 11).

Figure 11 – Agarose gel analysis of PCR products spanning the COII/tRNA^{Lys} intergenic region



Lane 1 contains a 100 bp DNA ladder, lanes 8 and 14 contain negative (non-B) DNA samples (from Czech individuals). Lane 2 = BGK5, Lane 3 = BGK30, Lane 4 = BGK40, Lane 5 = BGK42, Lane 6 = BGK46, Lane 7 = BGK50, Lane 9 = BGK65, Lane 10 = BGK86, Lane 11 = BGK92, Lane 12 = BGK93, Lane 13 = BGK98, Lane 15 = BGK110, Lane 16 = BGK116, Lane 17 = BGK118, Lane 18 = BUN18, Lane 19 = BUN19, Lane 20 = KK24.

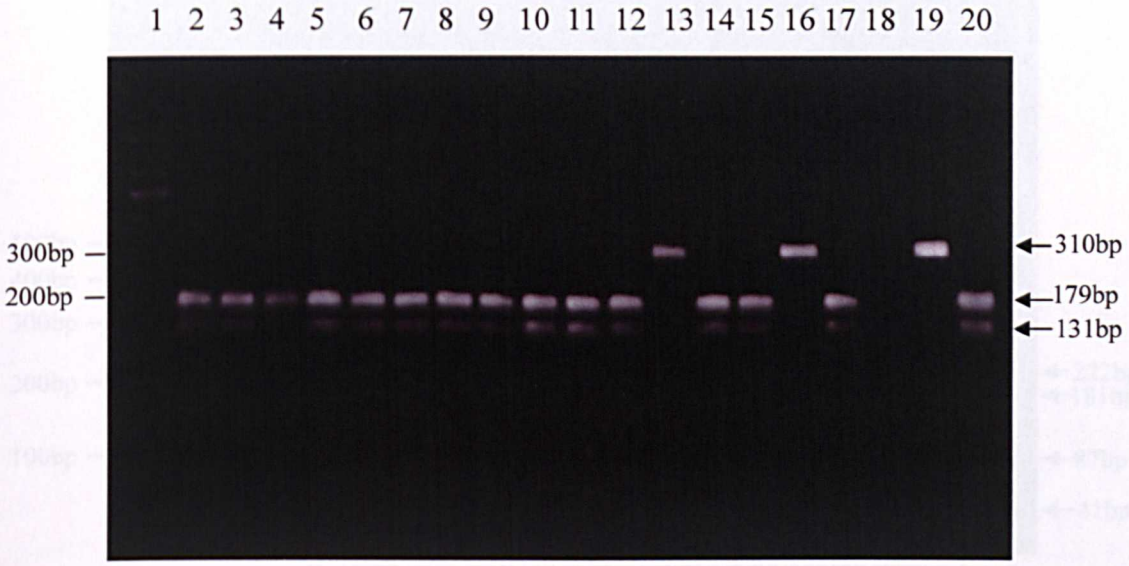
As seen in figure 11 the only sample to fail in this instance was BGK92. Samples BGK5, BGK40, BGK46, BGK50, BGK98, BGK110, BGK116 and BGK118 were found to have both copies of the 9 bp repeat and therefore were not haplogroup B. However, samples BGK30, BGK42, BGK65, BGK86, BGK93, BUN18, BUN19 and KK24 did have the 9 bp deletion and were therefore confirmed as belonging to haplogroup B.

3.3 Restriction Fragment Length Polymorphism Tests

RFLP screening was also used to resolve haplogroup status. As described section 2.6, this was done in a hierarchical fashion. The samples were first checked for their M and N status (M = +10397 *Alu*I +10394 *Dde*I, N = −10397 *Alu*I −10394 *Dde*I), and any further tests carried out depended on this result (NB. '+' means the gain of a restriction

site and '–' means the loss of a site). As seen in figure 12, digestion of the M fragment with *Alu*I gives one band of 310 bp if the sample is N and two fragments of 131 bp and 179 bp if it is M.

Figure 12 – *Alu*I RFLP analysis of PCR products spanning np 10270 – 10579

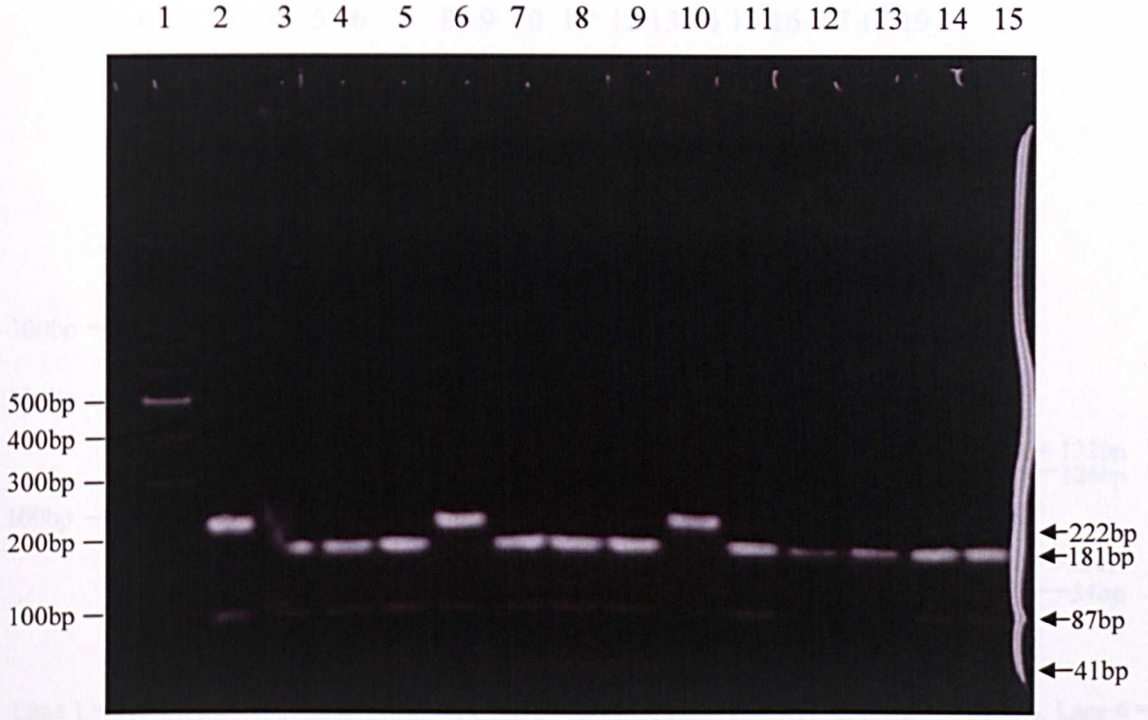


Lane 1 = 100 bp ladder, Lane 2 = AMB77, Lane 3 = AMB94, Lane 4 = BAL6, Lane 5 = BAL7, Lane 6 = BAL8, Lane 7 = BAL10, Lane 8 = BAL12, Lane 9 = BAL15, Lane 10 = BAL16, Lane 11 = BAL18, Lane 12 = BAL21, Lane 13 = BAL31, Lane 14 = BAL35, Lane 15 = BAL38, Lane 16 = BAL43, Lane 17 = BAL47, Lane 18 = BAL50, Lane 19 = BAL51, Lane 20 = BAL52.

As seen above, samples AMB77, AMB94, BAL6, BAL7, BAL8, BAL10, BAL12, BAL15, BAL16, BAL18, BAL21, BAL35, BAL38, BAL47 and BAL52 have bands at 131 bp and 179 bp and are therefore M. However, samples BAL31, BAL43, BAL50 and BAL51 have only one band at 310 bp and are therefore macrohaplogroup N.

The results of most *Alu*I digests were verified by digesting the same sample fragment with *Dde*I. As seen in figure 13, this digestion gives bands of 181 bp, 87 bp and 41 bp in M samples and bands of 222 bp and 87 bp in N samples.

Figure 13 – *Dde*I RFLP analysis of PCR products spanning np 10270 – 10579

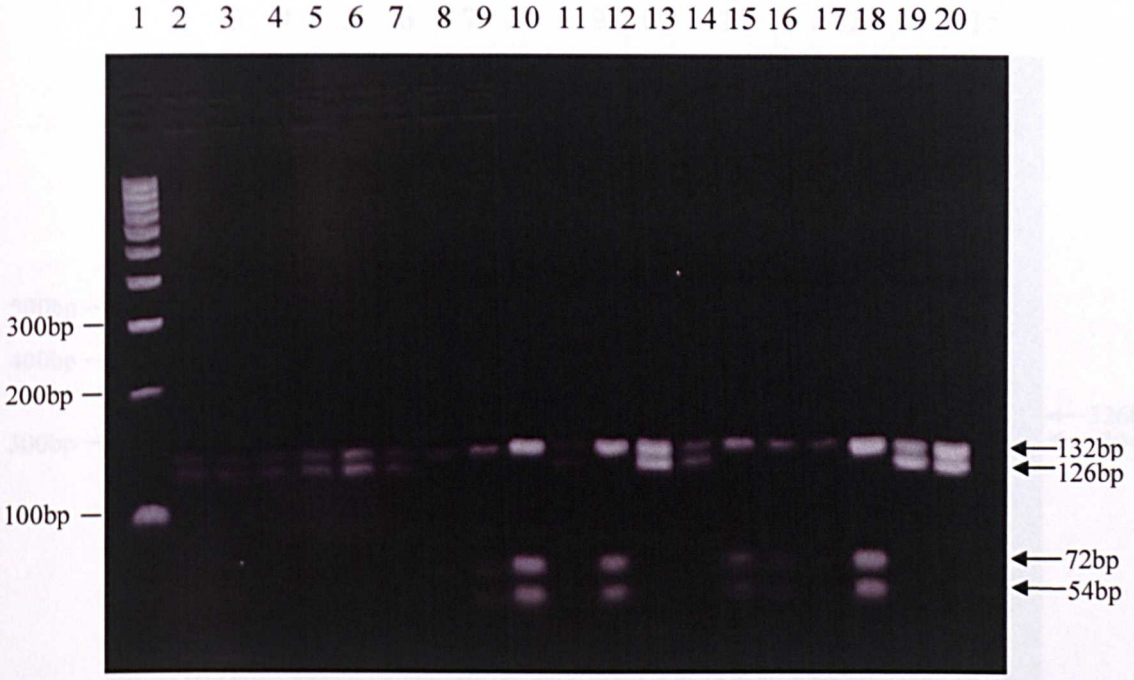


Lane 1 = 100 bp ladder, Lane 2 = BAL13, Lane 3 = BAL19, Lane 4 = BAL30, Lane 5 = BAL33, Lane 6 = BAL34, Lane 7 = BAL60, Lane 8 = BAL63, Lane 9 = BAL65, Lane 10 = BAL80, Lane 11 = BGK25, Lane 12 = BGK45, Lane 13 = BGK78, Lane 14 = BGK111, Lane 15 = PAD31.

As seen above, samples BAL19, BAL30, BAL33, BAL60, BAL63, BAL65, BGK25, BGK45, BGK78, BGK111 and PAD31 had a band at 181 bp and therefore belonged to macrohaplogroup M. On the other hand, samples BAL13, BAL34 and BAL80 had a band at 222 bp, these samples are therefore part of macrohaplogroup N.

Samples which were found to belong to macrohaplogroup M were further tested to assign them at the haplogroup level; some of these are illustrated below. Haplogroup M7 is defined by the gain of a *Hin*fI site at np 9824; relevant samples were amplified between np 9620 and np 9878 in order to study this site. As there is also a *Hin*fI site at np 9752, non-M7 samples have bands of 126 bp and 132 bp while M7 samples give bands at 54 bp, 72 bp and 132 bp (see figure 14).

Figure 14 – *Hin*fI RFLP analysis of PCR products spanning np 9620 - 9878

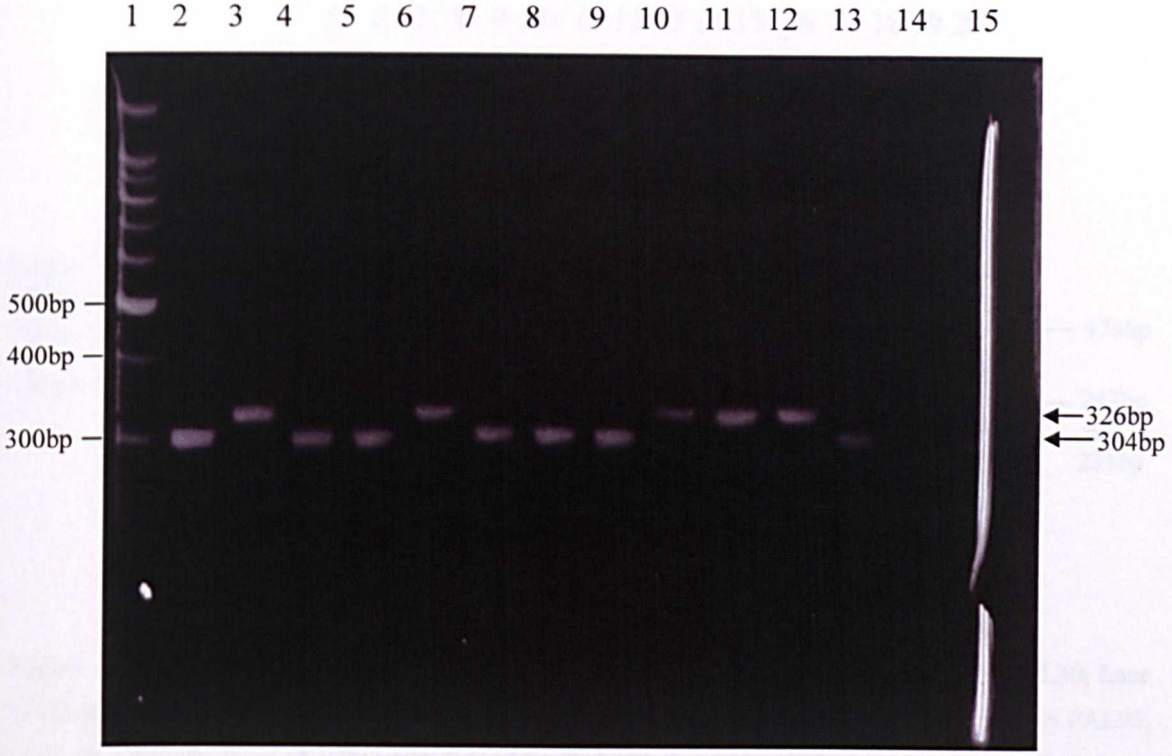


Lane 1 = 100 bp ladder, Lane 2 = KK148, Lane 3 = KK149, Lane 4 = KK151, Lane 5 = KK154, Lane 6 = KK157, Lane 7 = KK167, Lane 8 = KK169, Lane 9 = KK172, Lane 10 = KK174, Lane 11 = KK178, Lane 12 = MND4, Lane 13 = MND11, Lane 14 = MND12, Lane 15 = MND33, Lane 16 = MND38, Lane 17 = MND39, Lane 18 = MND43, Lane 19 = MND52, Lane 20 = MND55.

As seen here samples KK169, KK172, KK174, MND4, MND33, MND38, MND39 and MND43 do belong to haplogroup M7 while samples KK148, KK149, KK151, KK154, KK157, KK167, KK178, MND11, MND12, MND52 and MND55 do not.

Samples were also tested for their *Alu*I status at np 5176; loss of an *Alu*I site at this position is characteristic of haplogroup D. To evaluate this position, samples were amplified between np 5154 and np 5480. Samples belonging to haplogroup D have a single band of 326 bp whereas non-D samples have two bands of 22 bp and 304 bp (see figure 15).

Figure 15 – *Alu*I RFLP analysis of PCR products spanning np 5154 – 5480

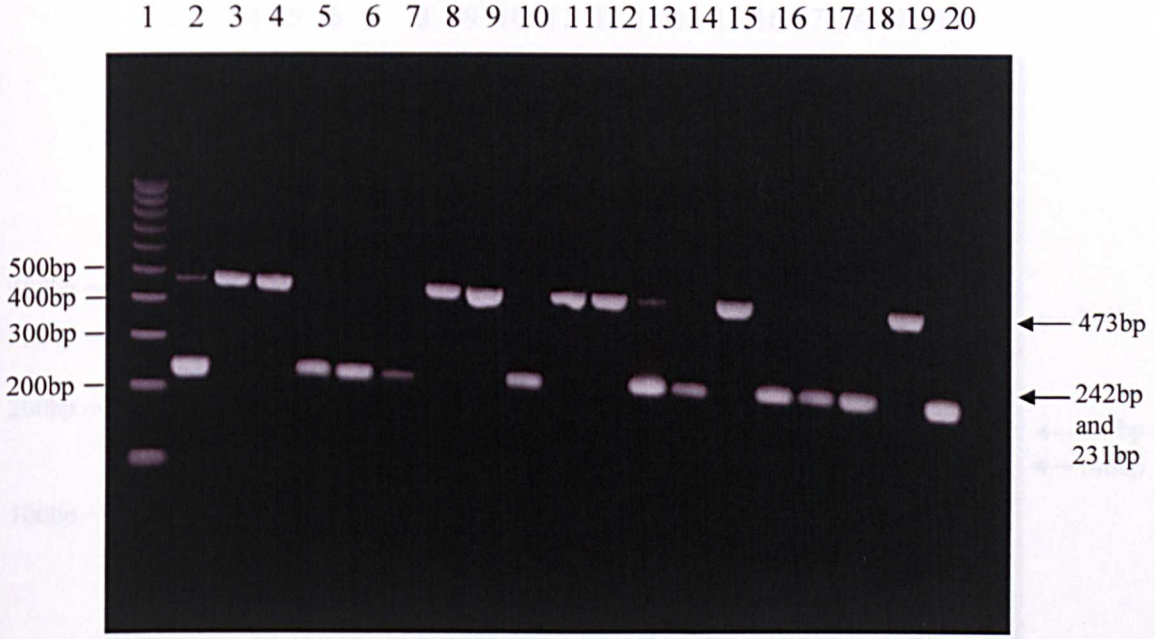


Lane 1 = 100 bp ladder, Lane 2 = KK157, Lane 3 = KK161, Lane 4 = KK167, Lane 5 = KK169, Lane 6 = positive control known to be haplogroup D, Lane 7 = KK171, Lane 8 = KK172, Lane9 = KK174, Lane 10 = MND1, Lane 11 = positive control known to be haplogroup D, Lane 12 = MND5, Lane 13 = MND11, Lane 15 = MND16.

As seen above, samples KK161, MND1 and MND5 have the band at 326 bp and are therefore haplogroup D. However, samples KK157, KK167, KK169, KK171, KK172, KK174, MND11 and MND16 have the band at 304 bp and therefore do not belong to haplogroup D. The band at 22bp is too small to be seen on the gel.

To test whether samples belonged to haplogroup E, they were amplified between np 7367 and np 7840 and digested with *Hha*I. Haplogroup E is defined by the loss of a *Hha*I site at np 7598 and so has a single band of 473 bp, non-E samples have two bands at 231 bp and 242 bp (see figure 16).

Figure 16 – *Hha*I RFLP analysis of PCR products spanning np 7367 – 7840

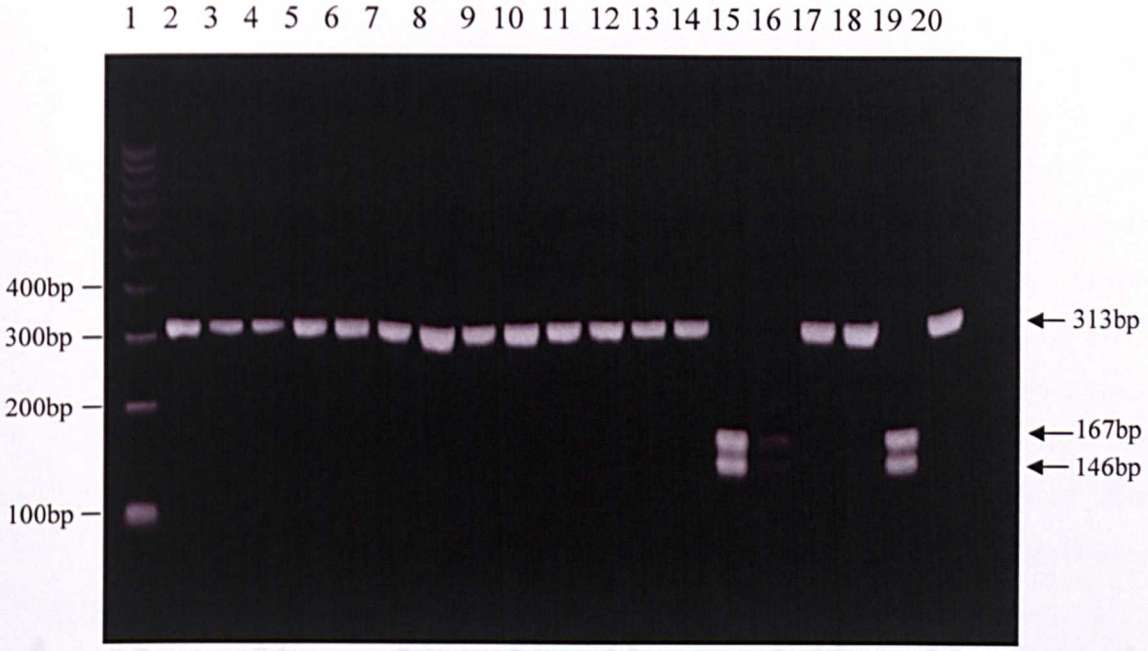


Lane 1 = 100 bp ladder, Lane 2 = PAD100, Lane 3 = PAD94, Lane 4 = PAD112, Lane 5 = PAL30, Lane 6 = PAL36, Lane 7 = PAL37, Lane 8 = PAL39, Lane 9 = PAL54, Lane 10 = PAL59, Lane 11 = PAL95, Lane 12 = PAL99, Lane 13 = PAL108, Lane 14 = PAL155, Lane 15 = PAL166, Lane 16 = PLB10, Lane 17 = PLB19, Lane 18 = PLB47, Lane 19 = PLB58, Lane 20 = PLB63.

As seen above, samples PAD94, PAD112, PAL39, PAL54, PAL95, PAL99, PAL166 and PLB58 have the band at 473 bp and are therefore E. However, samples PAD100, PAL30, PAL36, PAL37, PAL59, PAL108, PAL155, PLB10, PLB19, PLB47 and PLB63 do not belong to haplogroup E. The bands at 242 bp and 231 bp have not been fully separated due to the poor resolution of the gel.

Some samples from macrohaplogroup N were tested for their *Alu*I status at np 15606, the gain of an *Alu*I site at this position is diagnostic of haplogroup P (see figure 17). When digested, haplogroup P samples have bands of 146 bp and 167 bp whereas non-P samples have only a single band at 313 bp.

Figure 17 – *Alu*I RFLP analysis of PCR products spanning np 15439 – 15752



Lane 1 = 100 bp ladder, Lane 2 = ALO1, Lane 3 = ALO115, Lane 4 = AMB25, Lane 5 = AMB50, Lane 6 = BAL13, Lane 7 = BAL34, Lane 8 = BAL51, Lane 9 = BAL65, Lane 10 = BAL80, Lane 11 = BGK25, Lane 12 = BGK26, Lane 13 = BGK78, Lane 14 = BGK111, Lane 15 = MND9, Lane 16 = MND11, Lane 17 = MND79, Lane 18 = MND84, Lane 19 = MND97, Lane 20 = MTR122.

In the example above, samples MND9, MND11 and MND97 have the two smaller bands and are therefore haplogroup P. All the other samples (ALO1, ALO115, AMB25, AMB50, BAL13, BAL34, BAL51, BAL65, BAL80, BGK25, BGK26, BGK78, BGK111, MND79, MND84 AND MTR122) were not digested and therefore do not form part of haplogroup P.

4. Results – Evaluation of the Dataset

4. Results - Evaluation of the Dataset

As stated in section 2.1, 1144 samples were sequenced / RFLP typed as part of this study. Obviously such a substantial data set needs to be checked to verify the authenticity of the sequences. This is highlighted by the high level of mistakes which have made their way into the literature. These can usually be identified in an excess of transversions/indels or the sharing of unusual mutations between several haplogroups/lineages. For example, the data set of Lee *et al.* (1997) displays a great excess of transversions, most of which are unique in the literature. One of their samples in fact shows three such transversions. This suggests that most of these mutations are artefactual, possibly due to phantom mutations (e.g. systematic appearances of false mutations which may be due to biochemical problems) or base shifts due to mis-scoring (Bandelt *et al.* 2001).

Mistakes such as these have led to unnecessary confusion about the occurrence of recombination in mtDNA. For example, Hagelberg *et al.* (1999) claimed to have found evidence of mtDNA recombination in a population from the island of Nguna in Vanuatu. They appeared to have found a rare transition at np 16076 in five separate mtDNA lineages from the same island – including what are now defined as haplogroups B4a1, P and Q. As there is no evidence that np 16076 is hypervariable, they postulated that mtDNA recombination was the only explanation for their results and that it must happen relatively often to explain the frequency of this mutation in their results. However, these mysterious results were in fact a consequence of a ten nucleotide shift in the Nguna sequences relative to the reference sequence; the transitions at np 16076 were in fact transitions at the much more variable np 16086, meaning that the recombination conclusion was "no longer tenable for these data" (Hagelberg *et al.* 2000).

Therefore it is important to check a data set thoroughly before utilising it in any further analysis. The most obvious way to start is to ensure that the all sequence data are read by two people and then to attempt to fit the results into the scheme of existing haplogroups. In this study, this was done by comparing the new data with the existing worldwide database and identifying the occurrence of haplogroup-specific mutations.

Networks were drawn by hand and any unusual mutations, particularly at sites with a low relative mutation rate, were rechecked with the original sequence data. Any unusual transversions were also rechecked.

Another method used to check the validity of the data set is to filter out any frequent mutations and use a (quasi-)median network to visualise incompatibilities (Bandelt *et al.* 2002). In this study, each major haplogroup (plus macrohaplogroups M and N as a whole) was analysed individually, by using the methods of Bandelt *et al.* (2002), and the programme 'netmat' (courtesy of Vincent Macaulay, University of Glasgow). This process filters out the sites between np 16051 and np 16365 which undergo transitions at least as fast as the average transitional rate. A network of the remaining sites was drawn using Network 4.1 (Shareware Phylogenetic Software Website).

The following haplogroups gave perfect trees (i.e. no homoplasy in the filtered data): B4, D, E, G, M21, N9, N*, M*, P, Q and R9; some of these are shown in figures 18-24. This suggests that no major errors were made in collecting at least the results for these haplogroups.

Figure 18 – Network showing the filtered data from haplogroup D

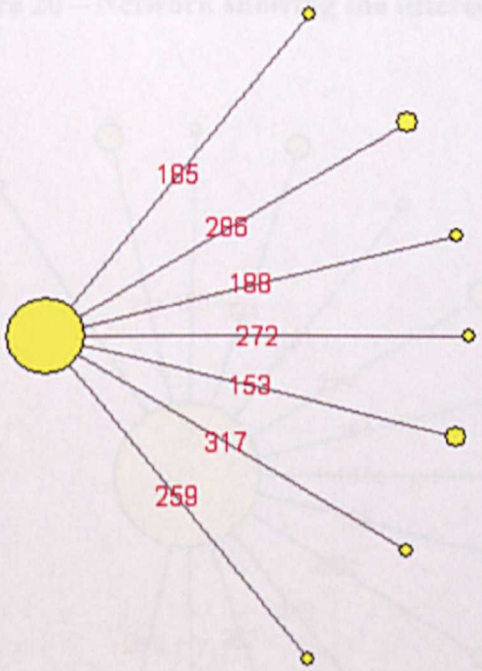


Figure 19 – Network showing the filtered data from haplogroup B4

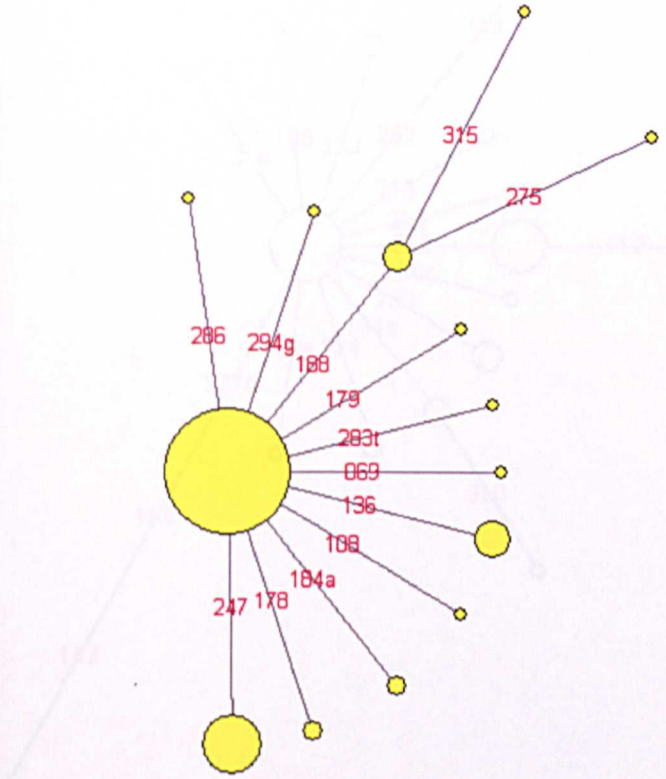


Figure 20 – Network showing the filtered data from haplogroup E

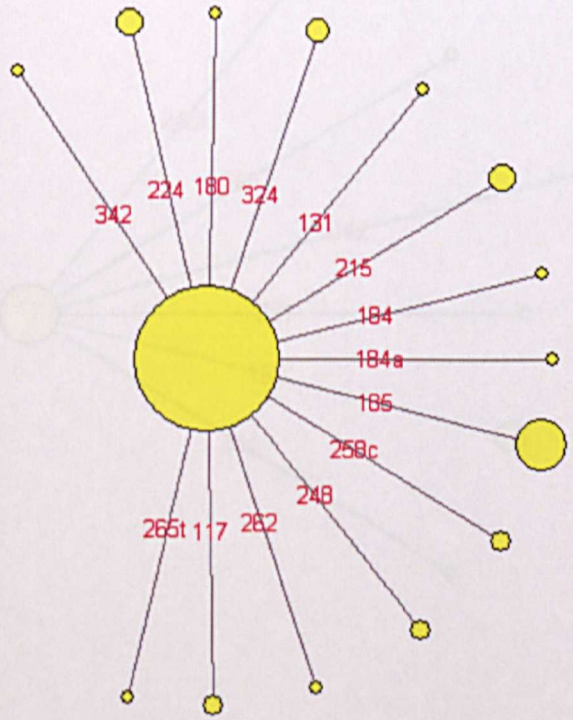


Figure 21 – Network showing the filtered data from haplogroup G

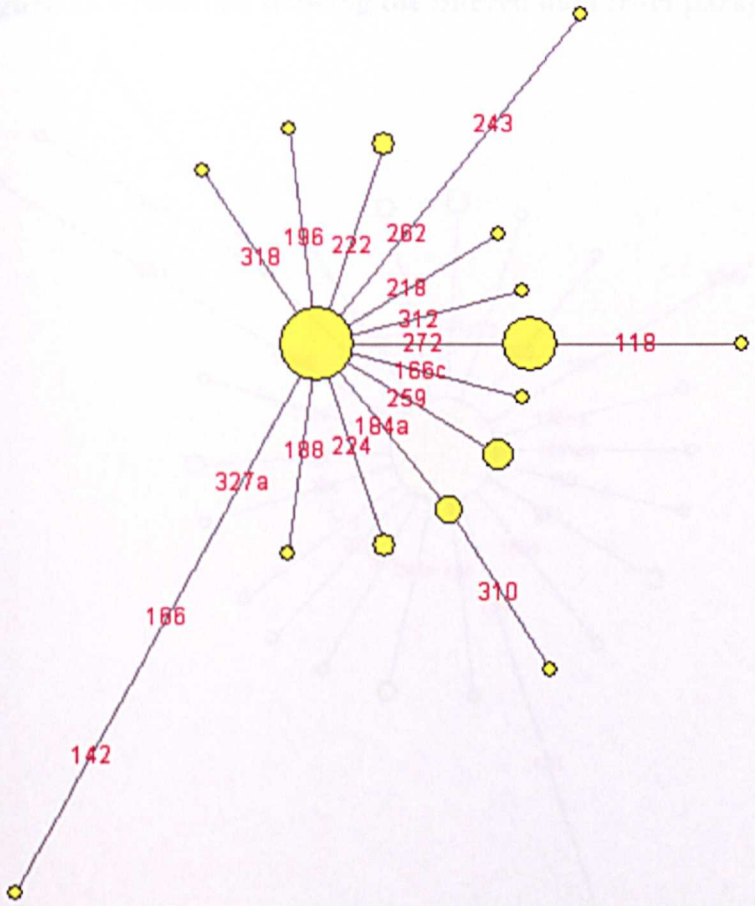


Figure 22 – Network showing the filtered data from paragroup N*

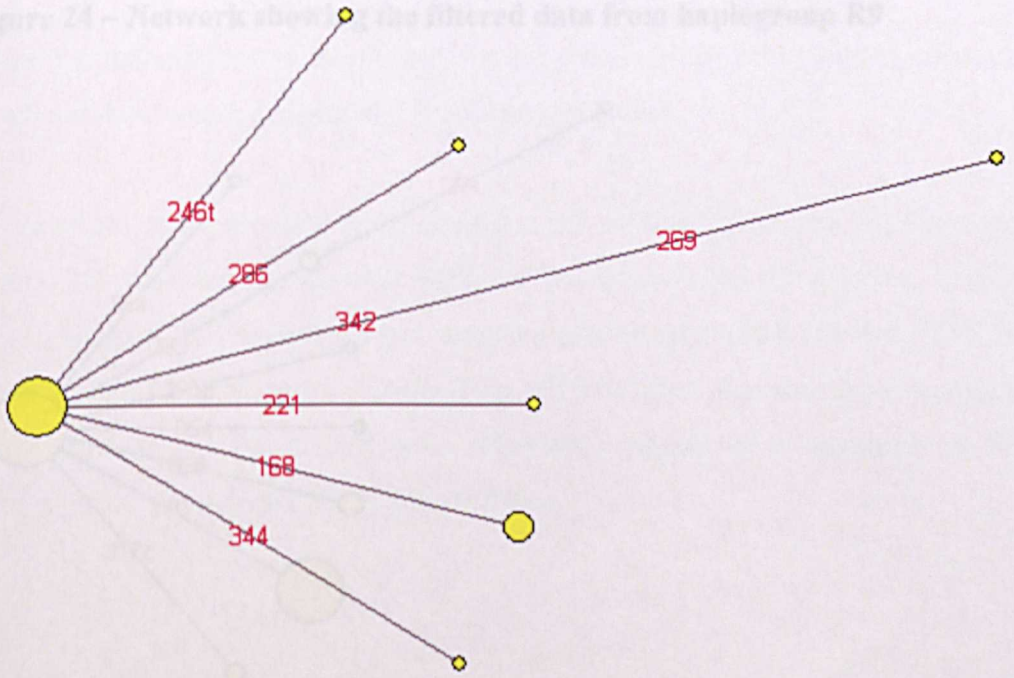


Figure 23 – Network showing the filtered data from paragroup M*

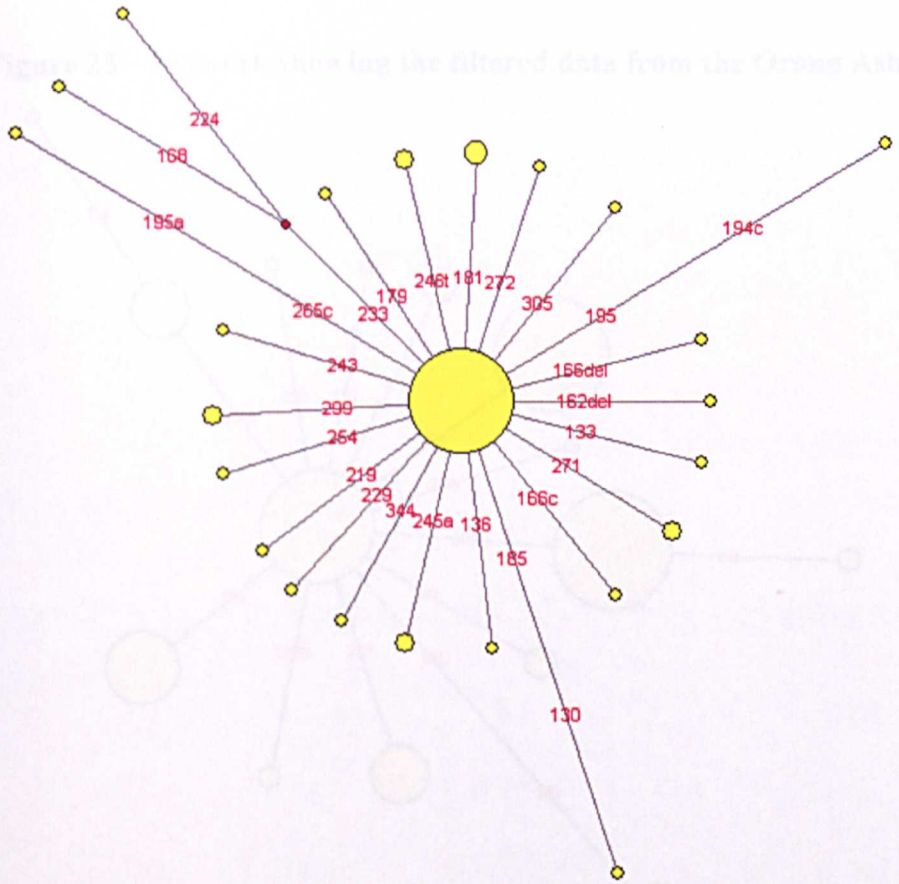
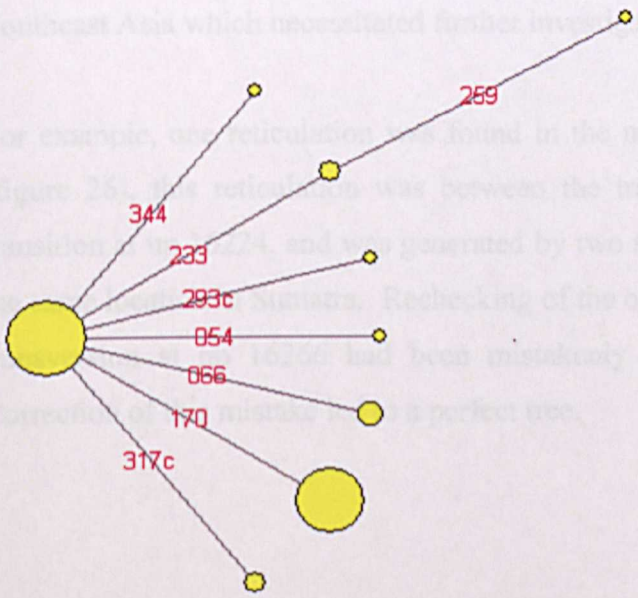
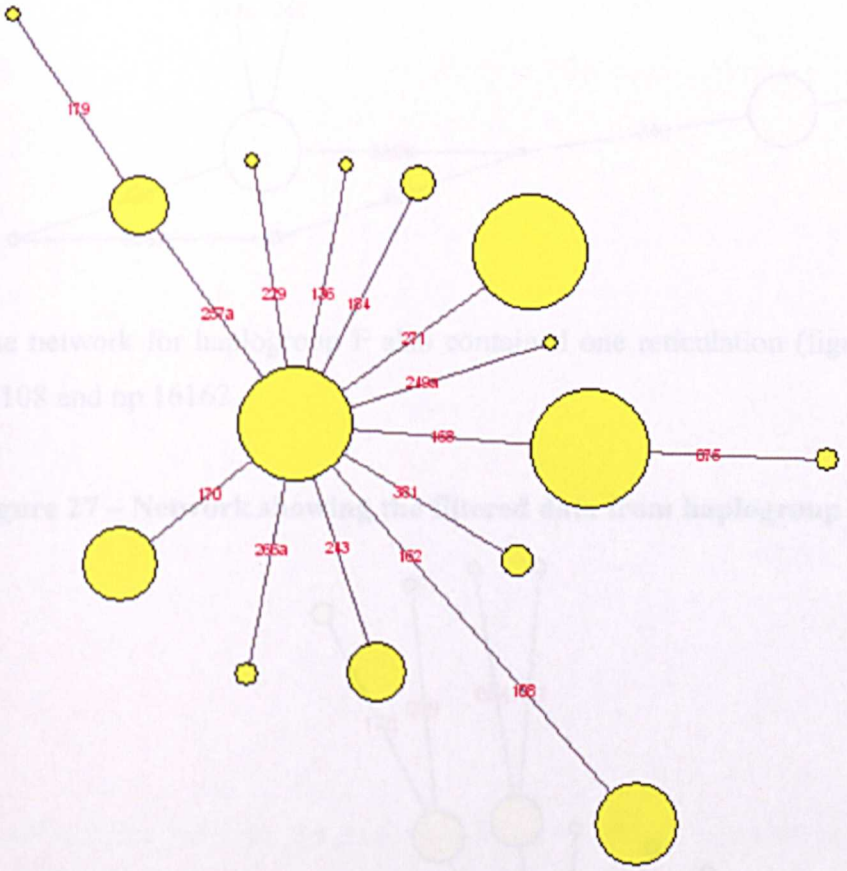


Figure 24 – Network showing the filtered data from haplogroup R9



A network was also constructed of the Orang Asli data which again gave a perfect tree (figure 25).

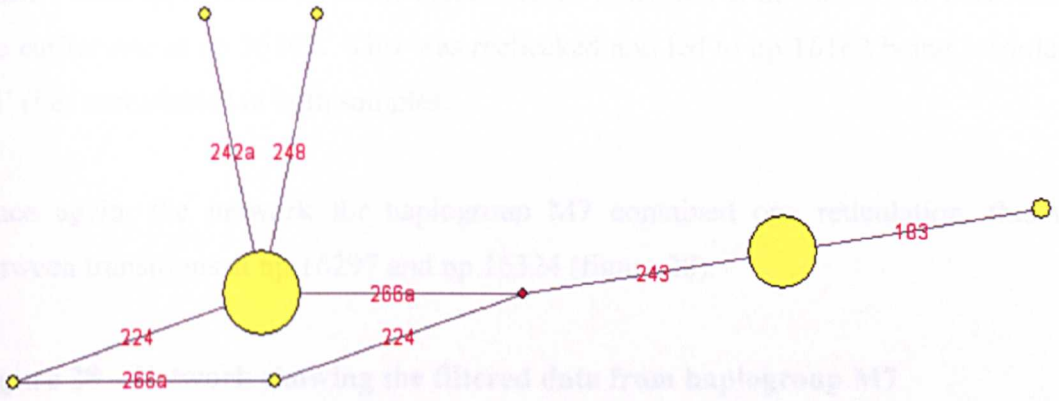
Figure 25 – Network showing the filtered data from the Orang Asli samples



However, reticulations were apparent in the networks for other haplogroups in Island Southeast Asia which necessitated further investigation.

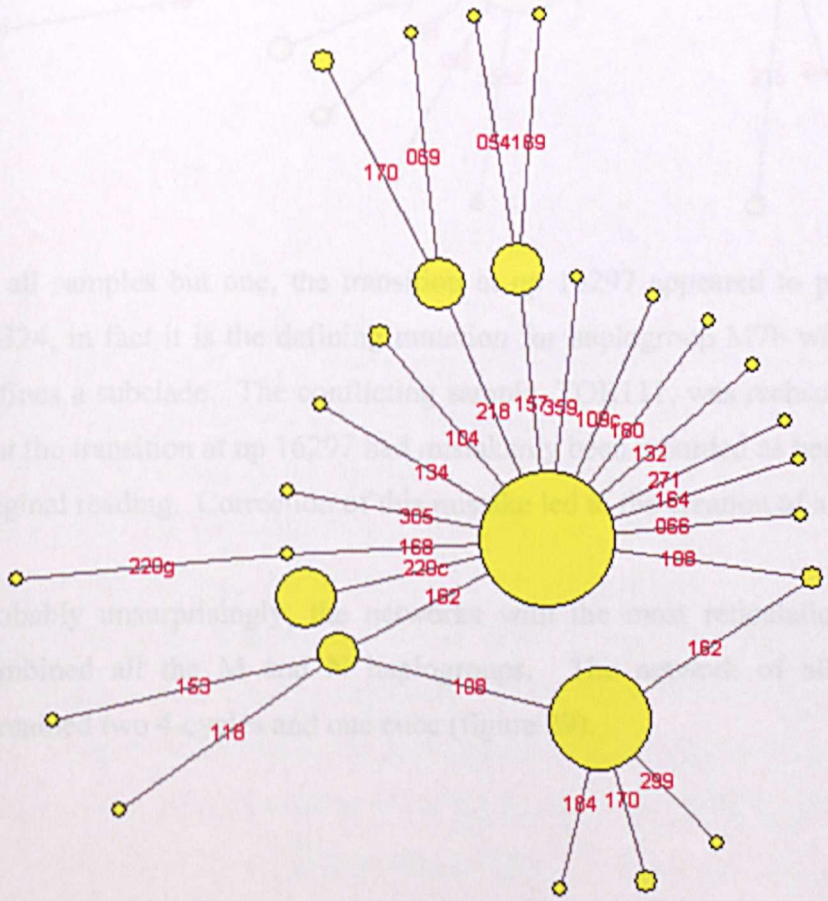
For example, one reticulation was found in the network generated for haplogroup B5 (figure 26), this reticulation was between the transversion to A at np 16266 and a transition at np 16224, and was generated by two samples (PEK116 and PEK119) from the same location in Sumatra. Rechecking of the original sequence data showed that the transversion at np 16266 had been mistakenly scored as a transition in PEK119. Correction of this mistake led to a perfect tree.

Figure 26 – Network showing the filtered data from haplogroup B5



The network for haplogroup F also contained one reticulation (figure 27), between np 16108 and np 16162.

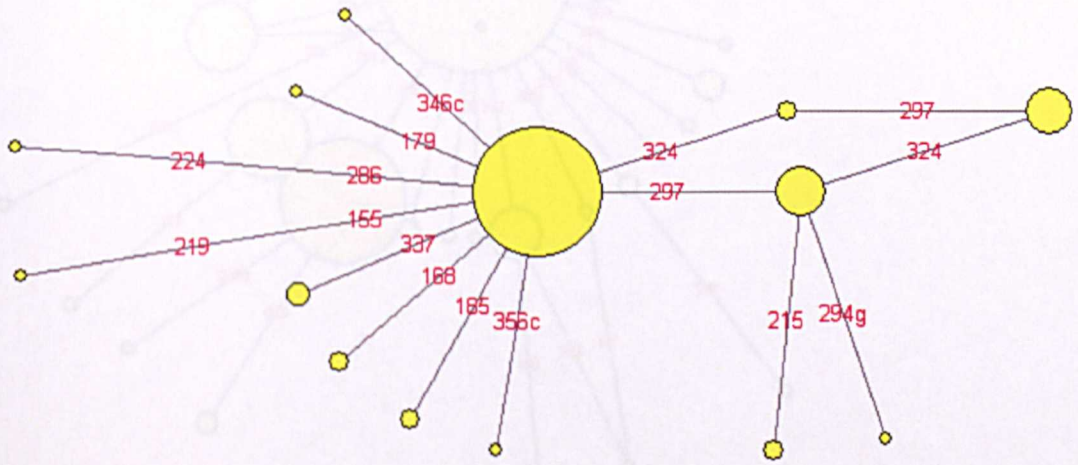
Figure 27 – Network showing the filtered data from haplogroup F



These transitions define haplogroups F1a1a and F1a1 respectively, samples PLB31 and PLB56 both appeared to have the more derived transition at np 16108 but were missing the earlier one at np 16162. This was rechecked and led to np 16162 being recorded as 'N' (i.e. unreadable) in both samples.

Once again, the network for haplogroup M7 contained one reticulation, this time between transitions at np 16297 and np 16324 (figure 28).

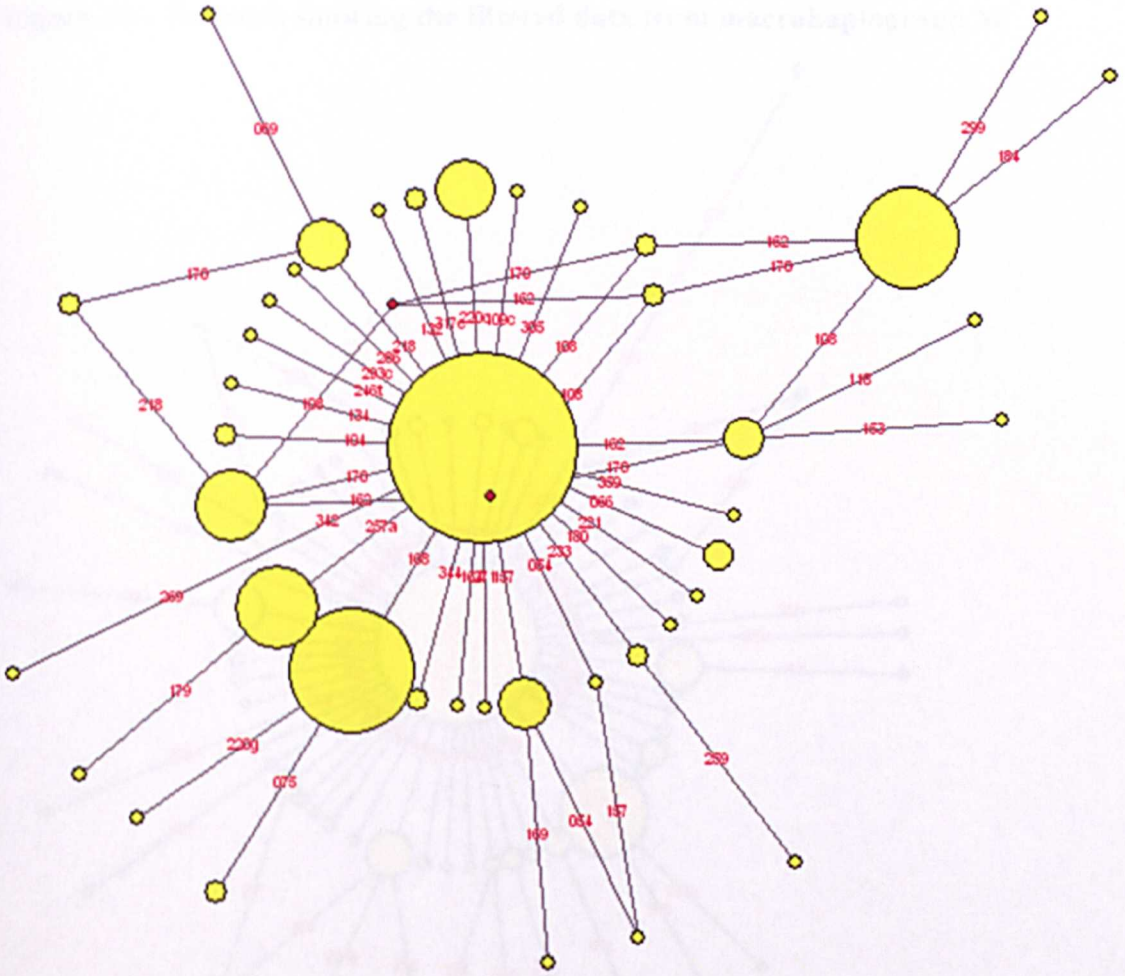
Figure 28 – Network showing the filtered data from haplogroup M7



In all samples but one, the transition at np 16297 appeared to precede the one at np 16324, in fact it is the defining mutation for haplogroup M7b whereas np 16324 only defines a subclade. The conflicting sample, TOR111, was rechecked and it was found that the transition at np 16297 had mistakenly been recorded as being at np 16294 in the original reading. Correction of this mistake led to the creation of a perfect tree.

Probably unsurprisingly, the networks with the most reticulation were those which combined all the M and N haplogroups. The network of all the N haplogroups contained two 4-cycles and one cube (figure 29).

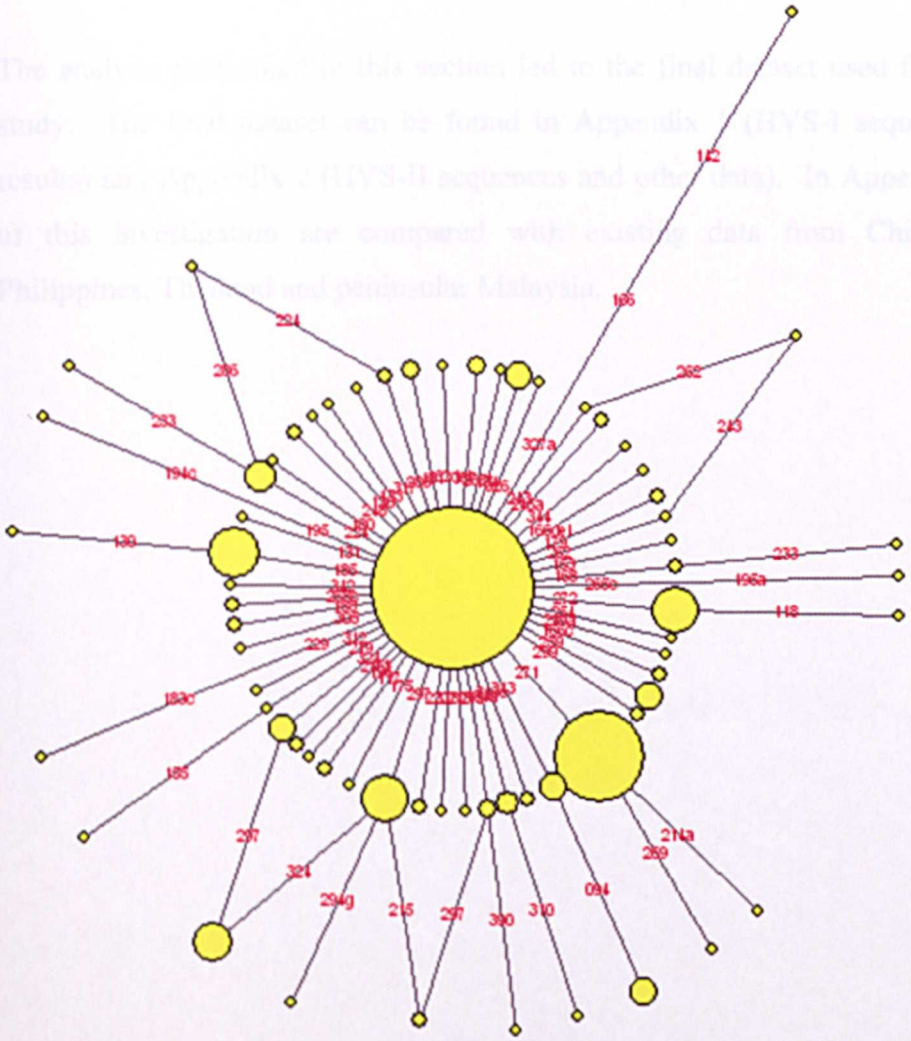
Figure 29 – Network showing the filtered data from macrohaplogroup N



One of the 4-cycles was between np 16054 and np 16157, rechecking this led to both ALO58 and ALO64 being re-classified as 'N' at 16054. The second 4-cycle and the cube were generated by the 16108-16162 reticulation which was dealt with in 'F' above, and also by the presence of a transition at np 16170 in both F* (PEK7) and F1a1a (PLB44 and PLB45). The existence of these transitions was verified by rechecking the original sequence data and it was also found in MED8 (a match for PEK7) which was missed when analysing the sequences the first time. Transitions at np 16170 have also been found by other studies in haplogroups H, K, L1, L2 and L3 (Vigilant *et al.* 1991; Watson *et al.* 1997; Parson *et al.* 1998; Krings *et al.* 1999b) providing further support for their occurrence on different backgrounds.

The network of all the M haplogroups contained four 4-cycles (figure 30).

Figure 30 – Network showing the filtered data from macrohaplogroup M



The first of these was between np 16243 and np 16262, investigation of which led to the correction of the sequence for MND30 from a transition at np 16262 to one at np 16261. The second 4-cycle was between np 16224 and np 16286. Examination of the sequences containing one or both of these mutations led to the discovery that a transition at np 16287 in MND84 had been mistakenly classified as being at np 16286. The final two 4-cycles were generated by transitions at np 16215, np 16297 and np 16324. In this case (apart from the mistake which led to the reticulation in the M7 network which was dealt with above) the occurrence of these transitions was verified by rechecking the original sequences. Therefore, in the Island Southeast Asian sample, both np 16215 and np 16324 are found in both haplogroup E1 and haplogroup M7b/M7b1. However, both mutations have also been found in other European, Asian

and African haplogroups (e.g. Richards *et al.* 1996; Watson *et al.* 1997; Yao *et al.* 2000) suggesting again that they are true, recurrent mutations.

The analysis performed in this section led to the final dataset used for the rest of the study. The final dataset can be found in Appendix 1 (HVS-I sequences and RFLP results) and Appendix 2 (HVS-II sequences and other data). In Appendix 3 the results of this investigation are compared with existing data from China, Taiwan, the Philippines, Thailand and peninsular Malaysia.

5. Results – Analysis

5. Results – Analysis

5.1 Heterozygosity

From even a cursory look at the sequence data found in this study (see Appendix I), it seems clear that the Island Southeast Asian populations studied have retained much more diversity than the Orang Asli groups. This was confirmed using a measure of heterozygosity (as detailed in section 2.8.2) to quantify the intragroup divergence (table 6).

As seen in table 6, the least diverse group are the Mendriq Semang (heterozygosity = 0.543) which is explained by the fact that over 84% of their sequences belong to haplogroup M21a, most of which are a single haplotype. By contrast the most diverse group of the Orang Asli are the Temuan (heterozygosity = 0.889). This difference is maintained in the divergence of the three Orang Asli groups as a whole: the Semang are the least diverse (heterozygosity = 0.768), the Aboriginal Malays the most (heterozygosity = 0.899), with the Senoi in between (heterozygosity = 0.803).

The least diverse of the Island Southeast Asian populations is the one from Tengger in Java (heterozygosity = 0.904). This is probably due to the fact that haplogroups F1a, M* and the putative M10 each make up approximately 20% of the population. However, even that Island Southeast Asian population is more variable than the most diverse Orang Asli group. This can be most likely explained by the fact that all the Island populations have been consistently larger over long periods of time and so will have undergone less drift than the Orang Asli groups.

Table 6 – Intragroup diversity in the populations under investigation

	Sample size (n)	Heterozygosity ($1 - \Sigma_i x_i^2$)
Batek	29	0.675
Jahai	50	0.561
Mendriq	31	0.543
Total Semang	**110**	**0.768**
Temiar	51	0.780
Total Senoi[1]	**52**	**0.803**
Semelai	60	0.851
Temuan	32	0.889
Total Aboriginal Malay[2]	**94**	**0.899**
Medan	42	0.959
Pekanbaru	52	0.952
Padang	24	0.939
Palembang	28	0.940
Bangka	34	0.951
Total Sumatra	**180**	**0.982**
Java – Tengger	**36**	**0.904**
Banjarmasin	89	0.979
Kota Kinabalu	68	0.968
Total Borneo	**157**	**0.984**
Manadao	89	0.944
Palu	38	0.943
Ujung Padang	46	0.955
Toraja	64	0.939
Total Sulawesi	**237**	**0.961**
Bali – Denpasar	**65**	**0.976**
Lombok – Mataram	**44**	**0.963**
Sumba – Waingapu	**50**	**0.956**
Alor	**45**	**0.959**
Ambon	**43**	**0.951**

[1] Includes 1 Semai Senoi

[2] Includes 2 Jakun Aboriginal Malay

5.2 Haplogroup Ages

The ages of most haplogroups/subhaplogroups were calculated using the statistic ρ (Forster *et al.* 1996) and are shown in table 7 where they are placed in order according to age. As can be seen in the table, the vast majority of haplogroups date to the Pleistocene.

Table 7 – Ages of haplogroups cited in the text

Age range	Haplogroup	Age (years)	Standard Error (years)
Over 38,000 years	M21	57,000	6,700
	R9b	50,700	20,100
	N21	43,000	25,000
	N9a	38,100	11,200
30,000 – 38,000 years	Q	37,500	7,800
	B4c	37,500	15,600
	B5b	35,000	11,800
	F3b	34,000	13,300
	B4	33,600	8,600
	M22	31,700	20,600
	P – 176-266 cluster	30,300	12,000
22,000 – 27,000 years	B4b	27,000	6,100
	F1a1	26,100	13,700
	M7c1	25,700	14,500
	B4a	25,200	6,800
	M7c1a	24,500	9,700
	D5	24,000	9,500
	E1	23,900	10,000
	B4* - 147 cluster	22,700	16,000
	B5b – 111 cluster	22,600	8,600
12,000 - 18,000 years	B5a	17,300	3,900
	B4c – 335 cluster	17,100	4,800
	M7* - ISE Asian branch	16,800	6,300
	R9b – 192-288 cluster	16,500	10,200
	M7b3	15,400	5,900
	D5 – 311 cluster	12,600	5,600
	R22	12,500	5,200
5,000 – 10,000 years	F1a1a	9,200	2,800
	E1a	8,700	2,700
	D5 – 092-148 cluster	8,000	3,300
	E1b	7,300	3,200
	B4a1	7,300	2,700
	M7c1c	6,000	1,600
	M7c1b	5,500	3,200
	N9a1 – 294 cluster	5,500	2,600
Under 5,000 years	F1a – 294 cluster	4,300	1,900
	Y2	3,600	1,400

5.3 Haplogroup Diversity

The diversity and ages of some of the main Southeast Asian haplogroups was also calculated using ρ, see table 8.

As seen in table 8, haplogroups B4a, B4c1, B5b, F1a1a and N9a are most diverse in China. Conversely, haplogroup B5a is most diverse in Thailand while D5 is most diverse in Indonesia and E1a is most diverse amongst Taiwanese aboriginals. Haplogroup M7c1c exhibits approximately the same levels of diversity in Borneo and amongst Taiwanese aboriginals

5.4 Principal Component Analysis

5.4.1 Principal Component Analysis of Island Southeast Asian Data

Principal component analysis is a useful way of representing high-dimensional data in two or three dimensions by means of projection, and of estimating the amount of variance which is represented by each principal component. These are extracted one after another and each contains as much of the remaining variation as possible.

A principal component analysis was carried out on haplogroup frequencies from all the populations from Island Southeast Asia, as well as the Chinese data of Kivisild *et al.* (2002), Yao *et al.* (2000), Yao *et al.* (2002a), Yao *et al.* (2002b), Yao *et al.* (2002c), the Taiwanese data of Tajima *et al.* (2003) and Melton *et al.* (1998) which has also been elaborated on in this study, and unpublished data from the Philippines (Martin Richards, personal communication). As shown in figure 31, the first principal component accounts for 20.2% of the variation and separates the Chinese groups from all the other populations; the groups from Northern China are the most extreme outliers.

Table 8 – Diversity of some of the main Southeast Asian haplogroups

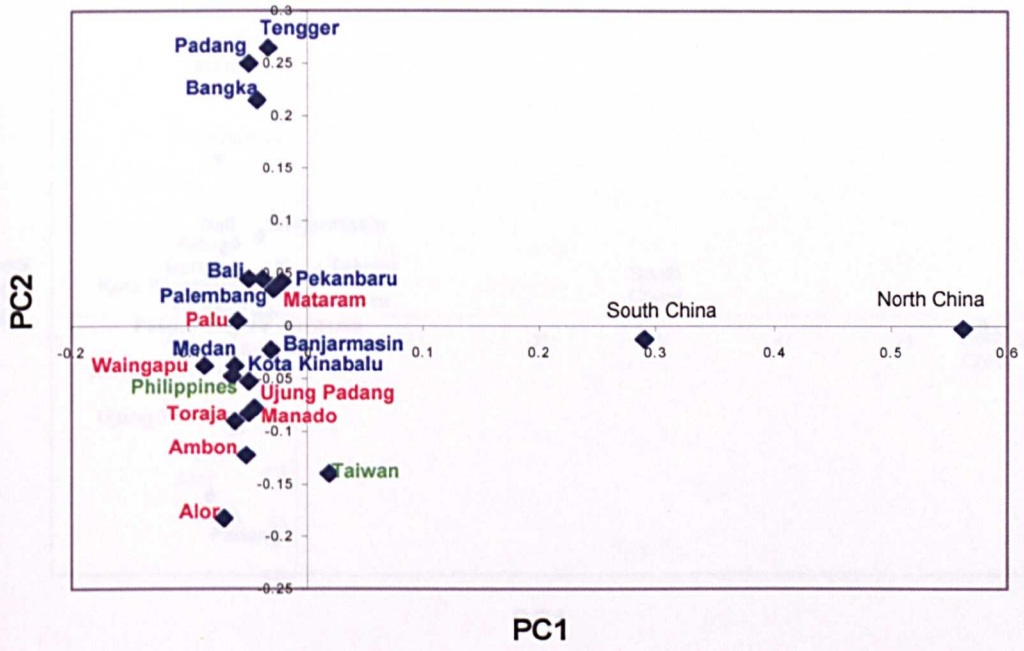
	Sample size (n)	ρ	Age	σ	Age SE
B4a – South China	30	2.38	48,000	0.72	14,500
B4a – Aboriginal Taiwanese	46	1.34	27,000	0.51	10,300
B4a – Thailand	13	2.31	46,600	0.67	13,500
B4a - Papua New Guinea	23	1.16	23,400	0.74	14,900
B4a – Indonesia	55	0.73	14,800	0.64	6,900
B4a1 – Indonesia	26	0.28	5,700	0.13	2,600
B4a1 – Papua New Guinea	22	0.29	5,900	0.12	2,400
B4c1 – South China	7	2.29	46,200	0.83	16,700
B4c1 – Aboriginal Taiwanese	11	0.82	16,500	0.47	9,500
B4c1 – Indonesia	29	1.50	30,300	0.91	18,400
B5a – South China	26	0.65	13,100	0.24	4,800
B5a – Aboriginal Taiwanese	16	0.50	10,100	0.29	5,900
B5a – Malay Peninsula	12	0.60	12,100	0.28	5,700
B5a – Thailand	27	1.11	22,400	0.33	6,700
B5a – Indonesia	34	0.47	9,500	0.16	3,200
B5b – China	10	2.10	42,400	0.79	15,900
B5b – Indonesia	15	0.47	9,500	0.27	5,500
F1a1a – South China	20	0.60	12,100	0.22	4,400
F1a1a – Thailand	22	0.14	2,800	0.08	1,600
F1a1a – Malay Peninsula	37	0.45	9,100	0.22	4,400
F1a1a – Indonesia	28	0.36	7,300	0.13	2,600
N9a – China	16	1.81	36,500	0.54	10,900
N9a – Aboriginal Taiwanese	5	1.60	32,300	0.69	13,900
N9a – Malay Peninsula	19	0.21	4,200	0.13	2,600
N9a – Indonesia	10	1.10	22,200	0.69	13,900
D5 – China	20	0.80	16,100	0.30	6,100
D5 – Aboriginal Taiwanese	8	0.88	17,800	0.41	8,300
D5 – Indonesia	26	1.62	32,700	0.84	17,000
E1a – Aboriginal Taiwanese	15	1.00	20,200	0.45	9,100
E1a – Philippines	6	0.67	13,500	0.33	6,700
E1a – Indonesia	71	0.37	7,500	0.15	3,000
E1a – Borneo	14	0.79	15,900	0.58	11,700
E1a – Sulawesi	41	0.24	4,800	0.10	2,000
M7c1c – Aboriginal Taiwanese	11	0.91	18,400	0.53	10,700
M7c1c – Indonesia	75	0.39	7,900	0.11	2,200
M7c1c – Borneo	12	0.92	18,600	0.46	9,300
M7c1c – Sumatra	16	0.31	6,300	0.17	3,400
M7c1c – Sulawesi	26	0.27	5,500	0.15	2,000

ρ is the average number of sites which differ between a set of sequences and their common ancestor.

Age = ρ / mutation rate

σ = standard error

Figure 31 – PC1 and PC2 of Island Southeast Asian, Taiwanese and Chinese data.

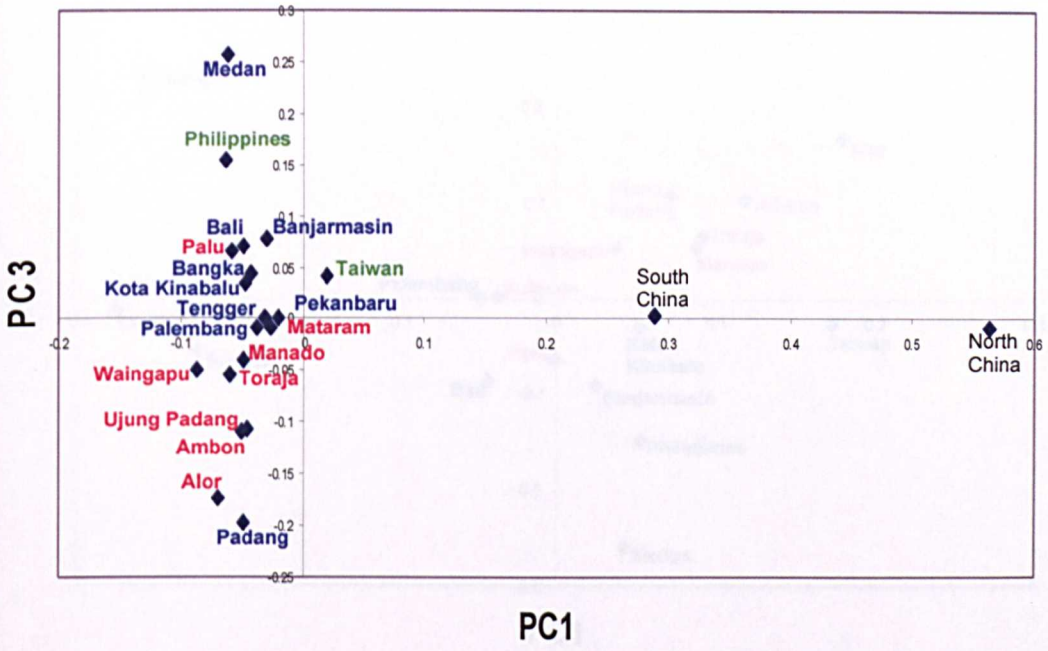


Populations in blue are found west of the Wallace line, those in red are east of the Wallace line and the Taiwanese and Filipinos are in green. PC1 = 20.2%, PC2 = 12.8%

PC2 (12.8%) separates the populations from Tengger, Padang and Bangka from all the other groups. In PC2, the Chinese fell within the variation found in most of the Island Southeast Asian populations. This was also the case in PC3, see figure 32. PC3 (9.4%) is a broadly west - east axis, with the exception of the population from Padang which is found at the opposite extreme to most of the other Western populations, possibly due to its high levels of haplogroup B*.

As the Chinese groups were strong outliers in PC1 they were excluded from further analyses to gain greater clarity for the other data.

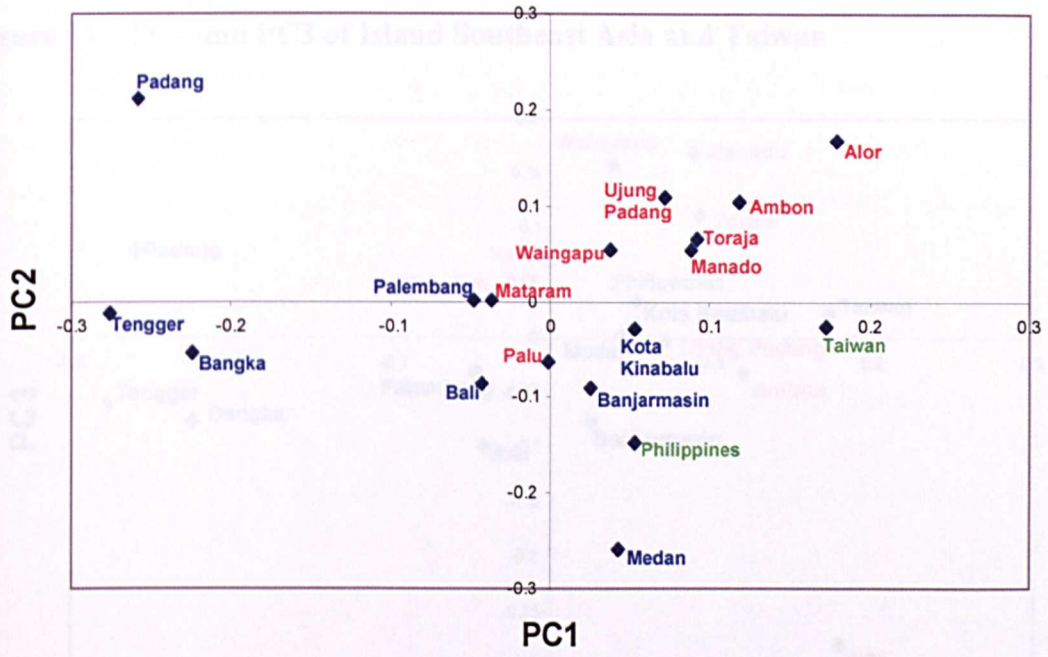
Figure 32 – PC1 and PC3 of Island Southeast Asian, Taiwanese and Chinese data.



Colour coding as in figure 31. PC1 = 20.2%, PC3 = 9.4%

The second analysis (figure 33) shows the first principal component (16.6%) to be a Sumatra/Java – Taiwan/Alor axis. In particular, the populations from Tengger, Bangka and Padang are again separated from all other groups. For each cycle of principal component analysis performed, a second analysis was carried out to plot the contribution of each haplogroup to each principal component. In this case, this demonstrated that the separation of the Sumatran and Javanese populations is due to the high levels of haplogroup B* in Padang and the putative haplogroup M10 clade in Tengger and Bangka. The less dramatic separation at the other pole seems to be due to the high levels of haplogroup F* in Taiwan and haplogroup Q in Alor.

Figure 33 – PC1 and PC2 of Island Southeast Asia and Taiwan.



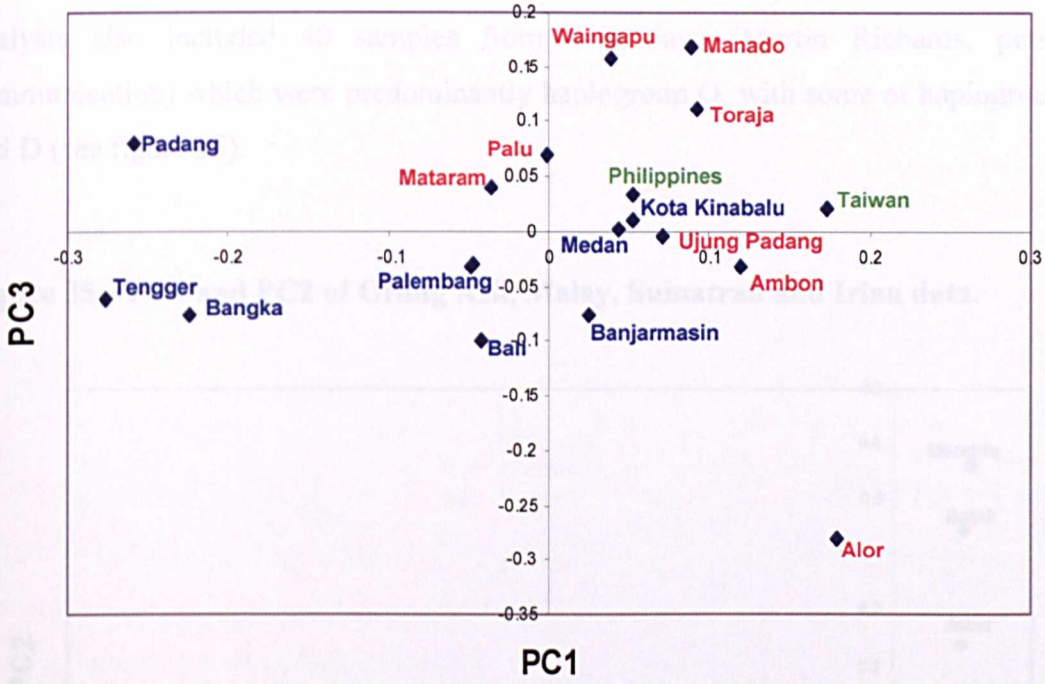
Colour coding as in figure 31. PC1 = 16.6%. PC2 = 11.8%

The second principal component (11.8%) is roughly east-west, with Medan at one extreme and Alor at the other; again Padang is the main exception. Medan is separated due to its relatively high levels of haplogroup Y2, while Alor is present at the other extreme because it contains elevated amounts of haplogroup Q.

The east – west patterning becomes more obvious when both principal components are considered together. Almost all the eastern populations are grouped together in one corner, the only exceptions being Mataram and Palu. In the case of the Mataram population this is perhaps not surprising as it is found so close to the Wallace line, and therefore the Western populations. The Palu population is unusual due to the much reduced level of haplogroup B types found there; it is also one of the few Eastern populations to contain any of haplogroups N9a1 and Y2 which could be due to recent arrivals from the west or north. In this east – west patterning, the Taiwanese groups can be seen to be slightly separate. However, the Filipino group is definitely found in the western part of the distribution.

As can be seen in figure 34, the east – west patterning is less obvious when PC1 and PC3 are plotted together.

Figure 34 – PC1 and PC3 of Island Southeast Asia and Taiwan



Colour coding as in figure 31. PC1 = 16.6%, PC3 = 10.1%

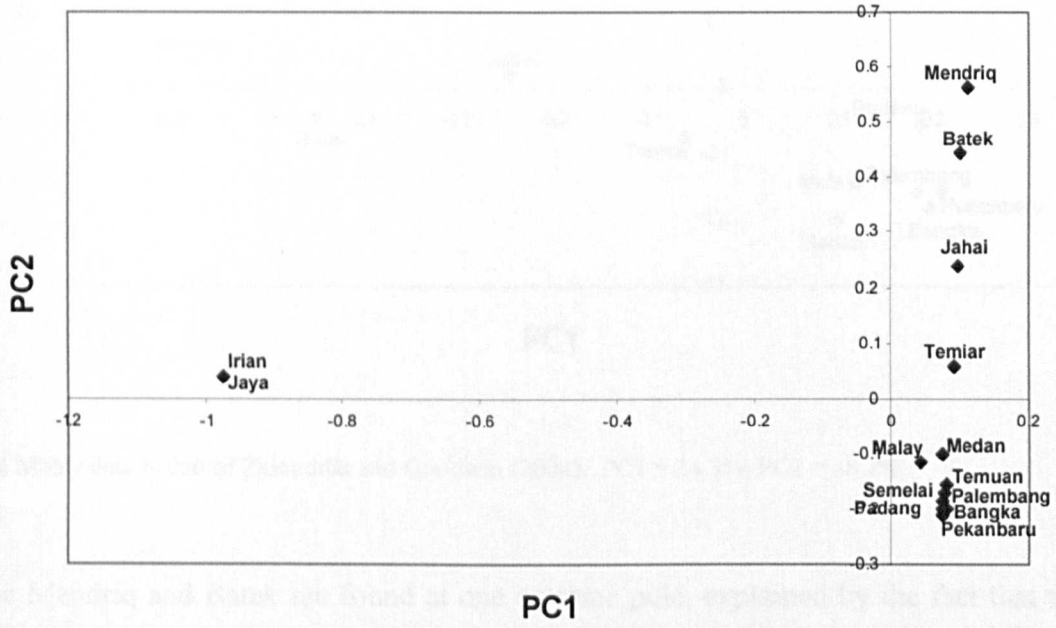
PC3 (10.1%) separates the group from Alor from all the other populations, again due to its high levels of haplogroup Q. Again the people of Mataram and Palu group with the Western populations; as, in this analysis, do the people of Waingapu.

Despite the relative lack of resolution seen in the last analysis, a definite east-west pattern can be seen across Island Southeast Asia. This suggests geographic structuring and that the population history of Island Southeast Asia cannot be simply explained by a population replacement.

5.4.2 Principal Component Analysis of Orang Asli Data

A principal component analysis was carried out on all of the Orang Asli ethnic groups for which there was sufficient data, as well as the Sumatran populations. The first analysis also included 40 samples from Irian Jaya (Martin Richards, personal communication) which were predominantly haplogroup Q, with some of haplogroups P and D (see figure 35).

Figure 35 – PC1 and PC2 of Orang Asli, Malay, Sumatran and Irian data.

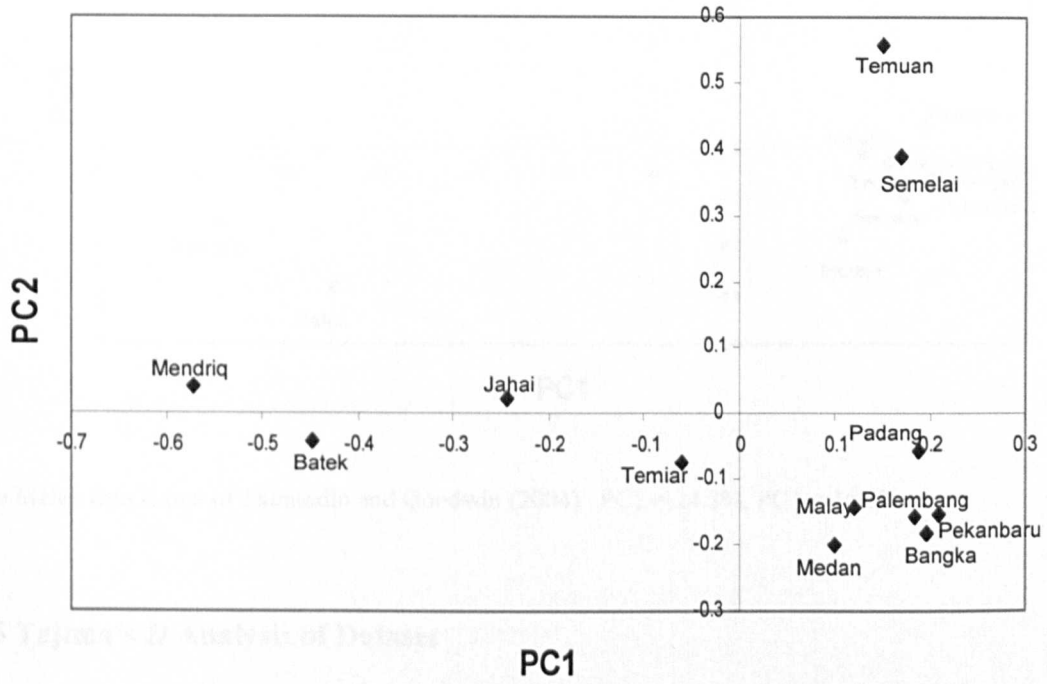


The Malay data is that of Zainuddin and Goodwin (2004). PC1 = 23.8%, PC2 = 18.8%

In PC2, which accounted for 18.8% of the variation, the Irianese fell between the Senoi and the group from Medan, with the Semang at one extreme and the Sumatrans (with the exception of those from Medan) and Aboriginal Malay at the other. However, they were a strong outlier for PC1 (23.8%), and so were excluded from further analyses to gain greater resolution within Southeast Asia.

The second analysis (figure 36) shows PC1 (representing 24.3% of the variation) to be a Semang - Aboriginal Malay/Sumatra axis with the Senoi in the middle.

Figure 36 – PC1 and PC2 of Orang Asli, Malay and Sumatran data

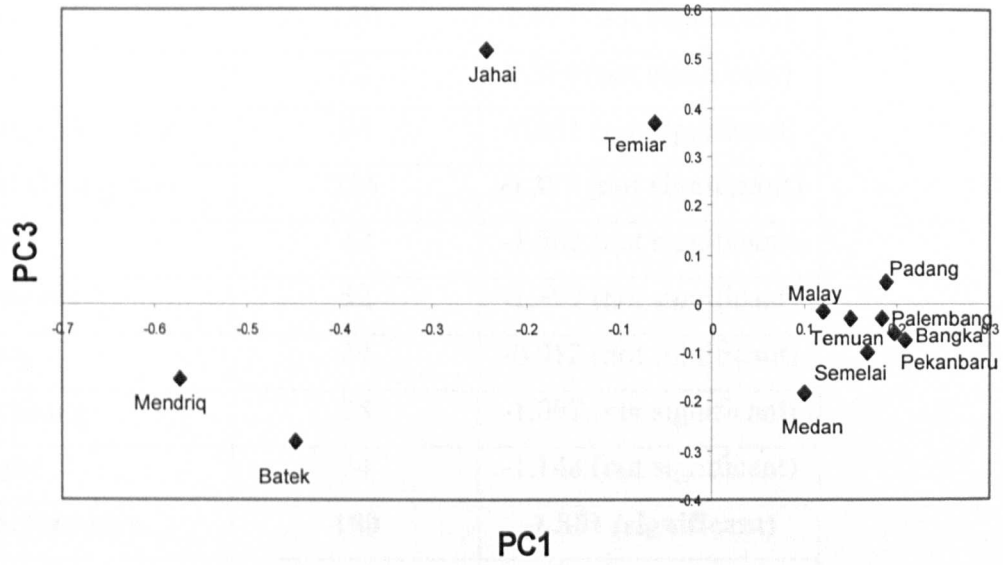


The Malay data is that of Zainuddin and Goodwin (2004). PC1 = 24.3%, PC2 = 18.2%

The Mendriq and Batek are found at one extreme pole, explained by the fact that they share haplogroup M21a at high frequencies. The Jahai group more closely with the Temiar as they share haplogroup R21, while the Aboriginal Malays are found at the opposite pole with the Sumatran and Malay groups. PC2 (18.2%) separates the Temuan and Semelai from all other populations due to their high levels of haplogroups R9b and N21.

PC3 (16.6%) separates out the Jahai and the Temiar from all the other groups, again due to their sharing of haplogroups R21 and F1a1a (figure 37).

Figure 37 – PC1 and PC3 of Orang Asli, Malay and Sumatran data



The Malay data is that of Zainuddin and Goodwin (2004). PC1 = 24.3%, PC3 = 16.6%

5.5 Tajima's *D* Analysis of Dataset

Tajima's *D* is a statistic which is used to test for demographic effects or selection processes. It compares two estimates of θ (the expected level of diversity in a population under neutral evolution): S (based on the number of segregating sites) and π (nucleotide diversity). Under neutrality, these different estimates of θ should be equal and therefore Tajima's *D* should equal zero. If Tajima's *D* is found to be significantly positive, this indicates balancing selection or population subdivision; a significantly negative result signifies positive selection or population growth (Tajima 1989). The results of this test on the current dataset are shown in table 9.

Table 9 – Tajima's *D* analysis of dataset

	Sample size (n)	Tajima's *D*
Semang	110	1.677 (not significant)
Senoi	52	0.310 (not significant)
Aboriginal Malay	94	-0.631 (not significant)
Total Orang Asli	**256**	**-0.327 (not significant)**
Medan	42	-1.582 (not significant)
Pekanbaru	52	-1.397 (not significant)
Padang	24	-0.912 (not significant)
Palembang	28	-1.097 (not significant)
Bangka	34	-1.148 (not significant)
Total Sumatra	**180**	**-1.801 (significant)**
Java – Tengger	**36**	**-0.619 (not significant)**
Banjarmasin	89	-1.644 (not significant)
Kota Kinabalu	68	-1.703 (not significant)
Total Borneo	**157**	**-1.746 (not significant)**
Manadao	89	-1.879 (significant)
Palu	38	-1.779 (not significant)
Ujung Padang	46	-1.453 (not significant)
Toraja	64	-0.973 (not significant)
Total Sulawesi	**237**	**-1.903 (significant)**
Bali – Denpasar	**65**	-1.679 (not significant)
Lombok – Mataram	**44**	-1.402 (not significant)
Sumba – Waingapu	**50**	-1.461 (not significant)
Alor	**45**	-0.919 (not significant)
Ambon	**43**	-0.937 (not significant)
Total Island Southeast Asia	**857**	**-1.955 (significant)**

As seen in table 9, the results for all populations apart from the Semang and Senoi were negative, although only one of the individual populations was significantly so.

However, the result for the whole of Island Southeast Asia and also Sulawesi and Sumatra were also significantly negative. This is most likely to be due to a population expansion.

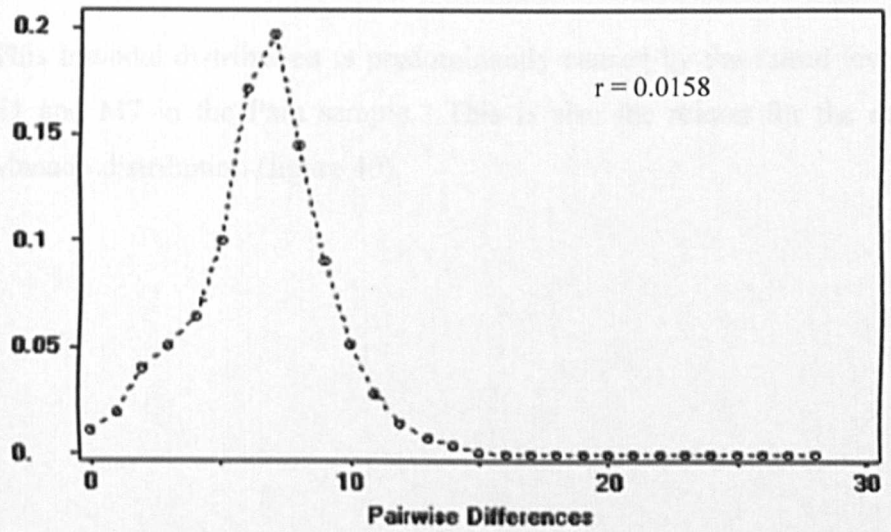
5.6 Mismatch Distributions

5.6.1 Mismatch Distributions of Island Southeast Asian Data

Mismatch distributions, or distributions of pairwise differences, are another way of measuring diversity and visualising demographic events. The shape of these distributions can be used to infer events in the population's history. For example, a smooth, bell-shaped distribution indicates a population expansion while a ragged, multimodal distribution indicates constant population size (Rogers and Harpending 1992).

As seen in figure 38, the mismatch distribution for Island Southeast Asia as a whole is bell-shaped which, as stated above, is indicative of a population expansion.

Figure 38 – Mismatch distribution for Island Southeast Asia



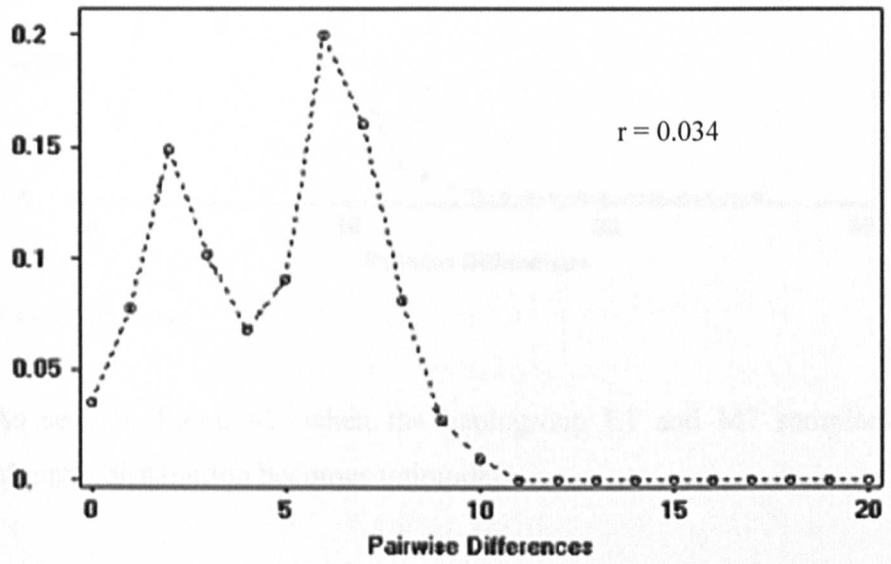
Y axis = frequency

The mismatch distributions for Bali, Banjarmasin, Bangka, Kota Kinabalu, Mataram, Padang, Pekanbaru, Ujung Padang and Waingapu (as well as those for Sumatra and

Borneo as a whole) were found to be almost identical to that for Island Southeast Asia thus also indicating population expansions (data not shown).

However, the mismatch distributions for Palu and Manadao were found to be bimodal (see figures 39 and 40).

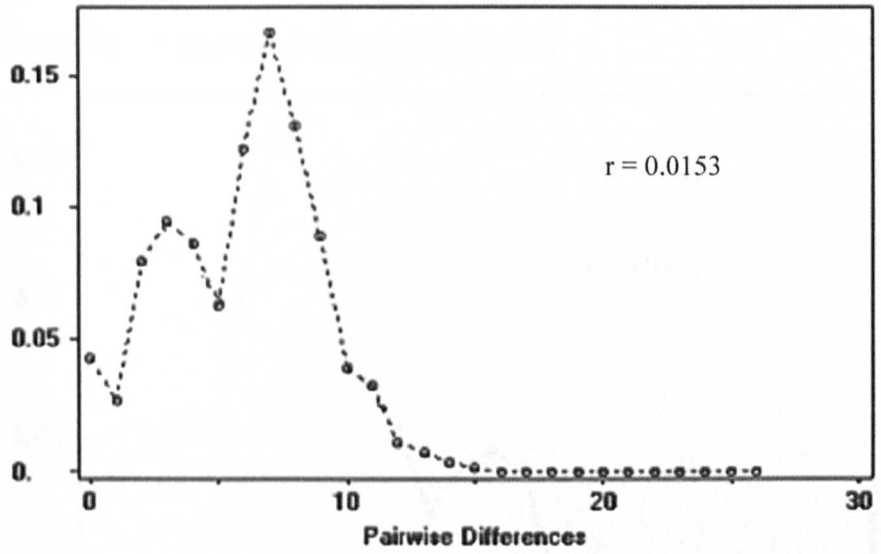
Figure 39 – Mismatch distribution for Palu



Y axis = frequency

This bimodal distribution is predominantly caused by the raised levels of haplogroups E1 and M7 in the Palu sample. This is also the reason for the second peak in the Manado distribution (figure 40).

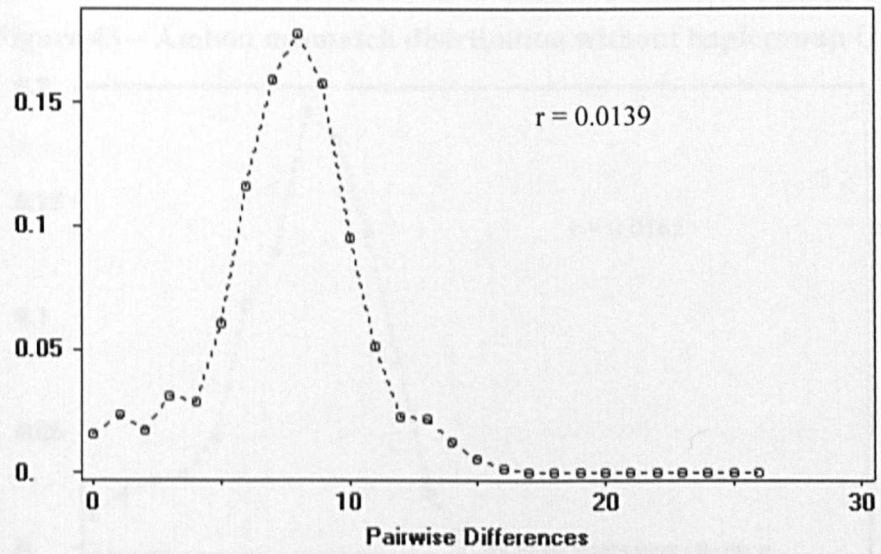
Figure 40 – Mismatch distribution for Manadao



Y axis = frequency

As seen in figure 41, when the haplogroup E1 and M7 samples are removed, the Manado distribution becomes unimodal.

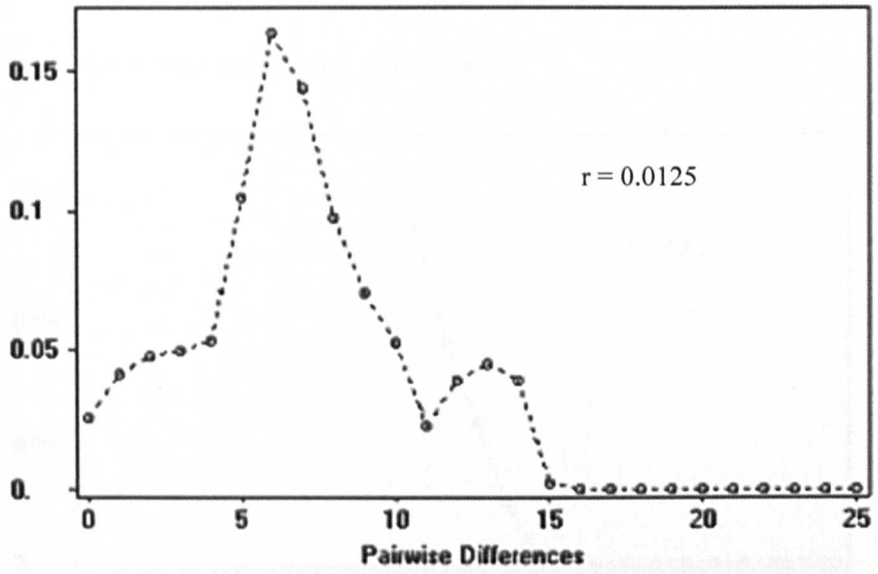
Figure 41 – Manado mismatch distribution without haplogroup E1 and M7 samples



Y axis = frequency

The mismatch distribution for Ambon also has a second, subsidiary, peak (see figure 42).

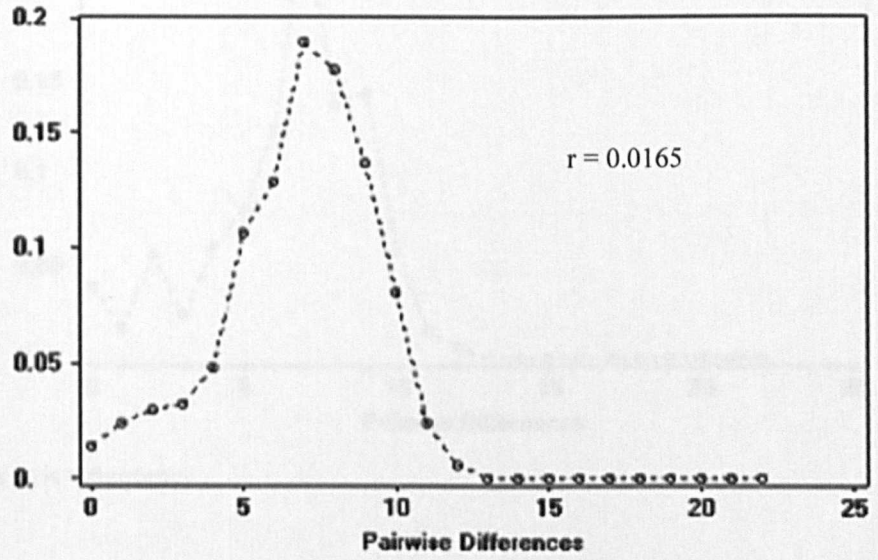
Figure 42 – Ambon mismatch distribution



Y axis = frequency

In this case, this is caused by the high frequency of the Melanesian haplogroup Q in the Ambon sample. As seen in figure 43, the distribution becomes unimodal when the haplogroup Q samples are removed.

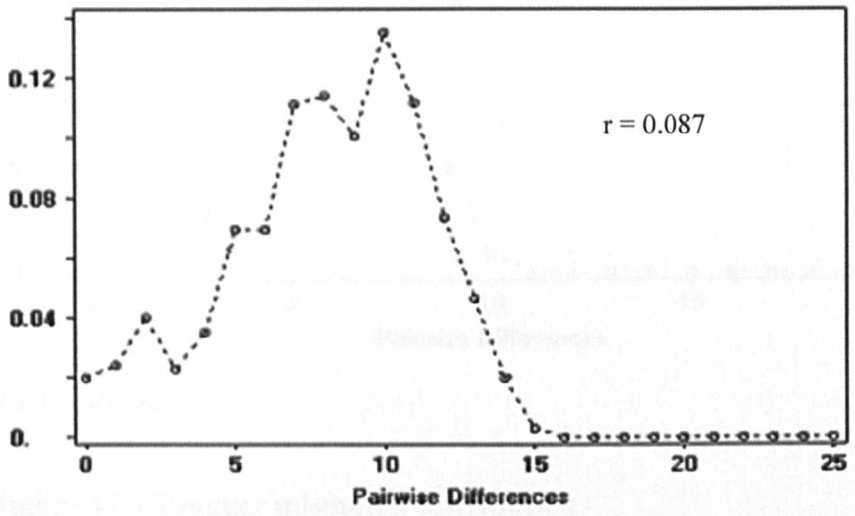
Figure 43 – Ambon mismatch distribution without haplogroup Q samples



Y axis = frequency

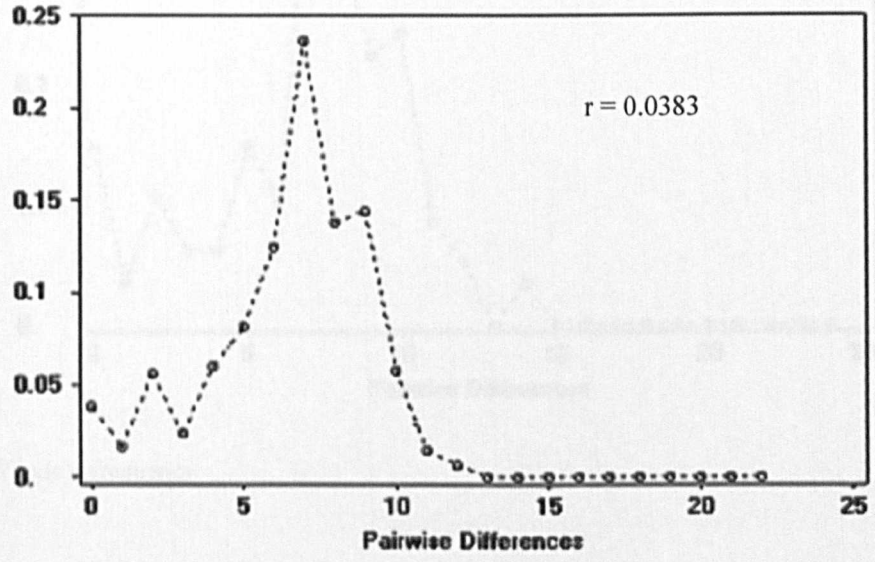
Furthermore, the mismatch distributions for Alor, Palembang, Medan, Tengger, Toraja and Sulawesi as a whole were found to be substantially more ragged than the distribution for all of Island Southeast Asia (see figures 44-49)

Figure 44 – Alor mismatch distribution



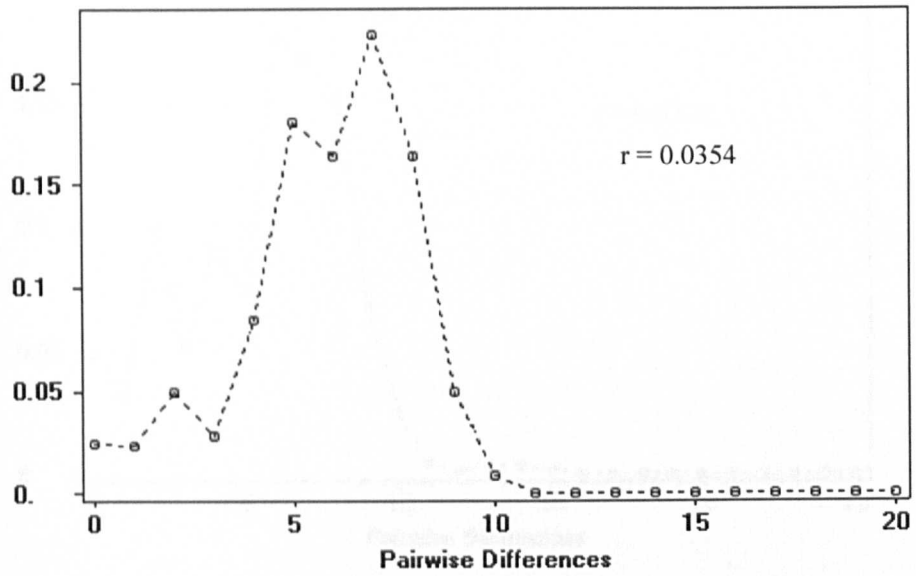
Y axis = frequency

Figure 45 – Palembang mismatch distribution



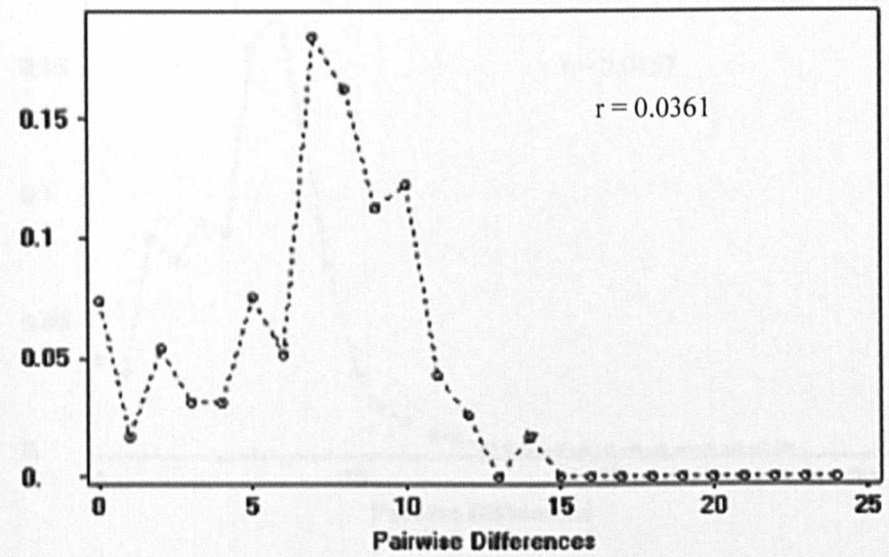
Y axis = frequency

Figure 46 – Medan mismatch distribution



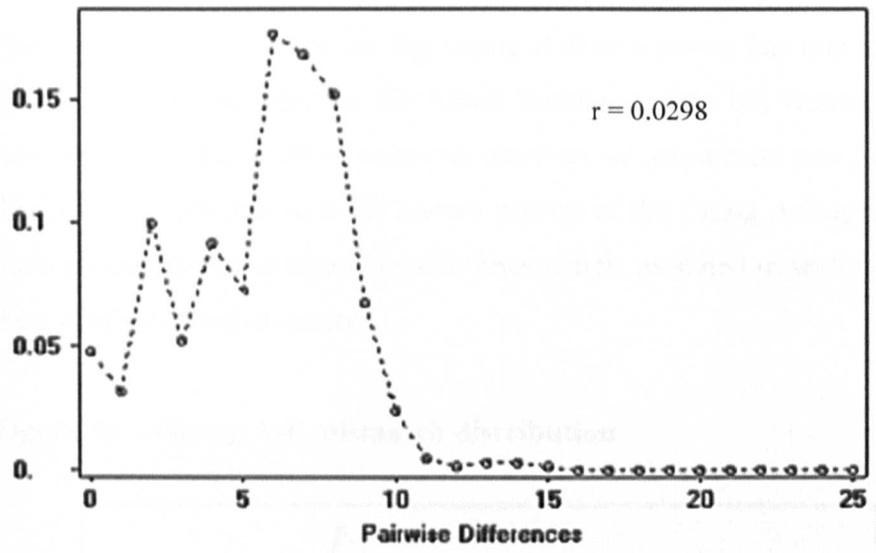
Y axis = frequency

Figure 47 – Tengger mismatch distribution



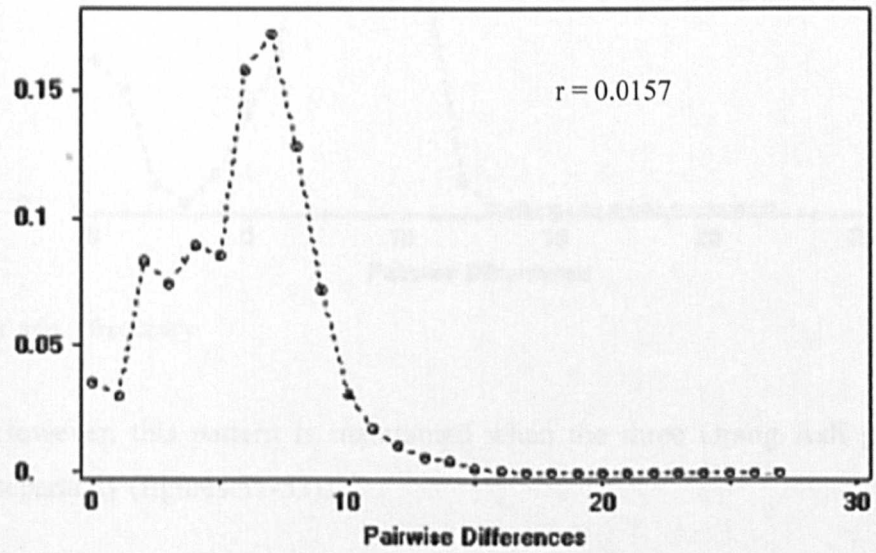
Y axis = frequency

Figure 48 – Toraja mismatch distribution



Y axis = frequency

Figure 49 – Sulawesi mismatch distribution



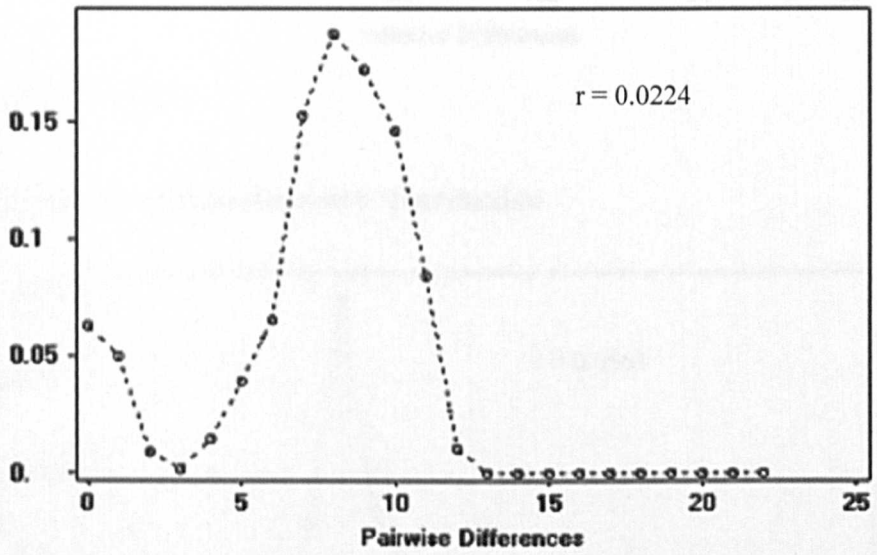
Y axis = frequency

The varying patterns of the mismatch distributions across Island Southeast Asia are possible evidence of differences in the demographic histories of the populations. The bimodal and ragged distributions seem to be particularly common in Sulawesi and could be the result of two separate expansions or the fusion of two disparate populations.

5.6.2 Mismatch Distributions of the Orang Asli Data

The mismatch distribution for the Orang Asli as a whole has one main peak similar to that seen in the distribution for Island Southeast Asia but it also has a second peak which could be the result of a second, more recent, expansion (see figure 50). However, this is not compatible with the known history of the Orang Asli groups which has seen them much reduced in size in recent times which, as stated in section 5.1, can be seen in their much reduced diversity.

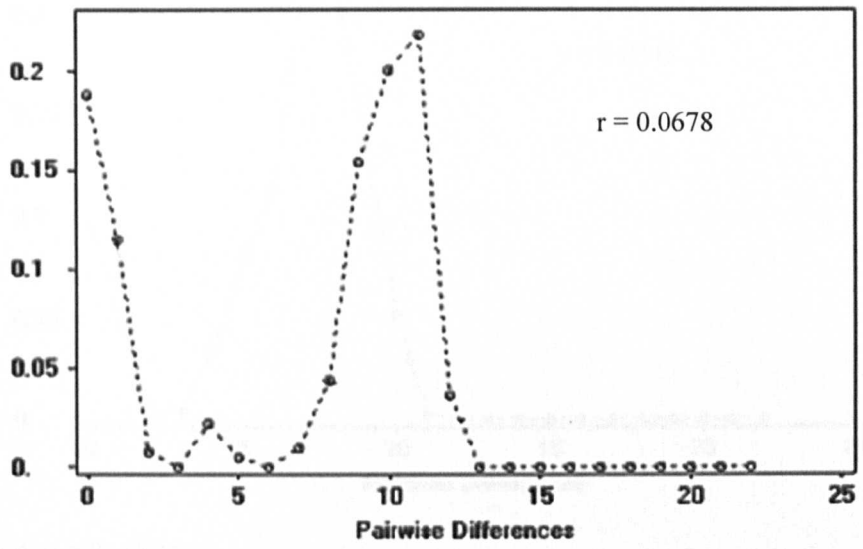
Figure 50 – Orang Asli mismatch distribution



Y axis = frequency

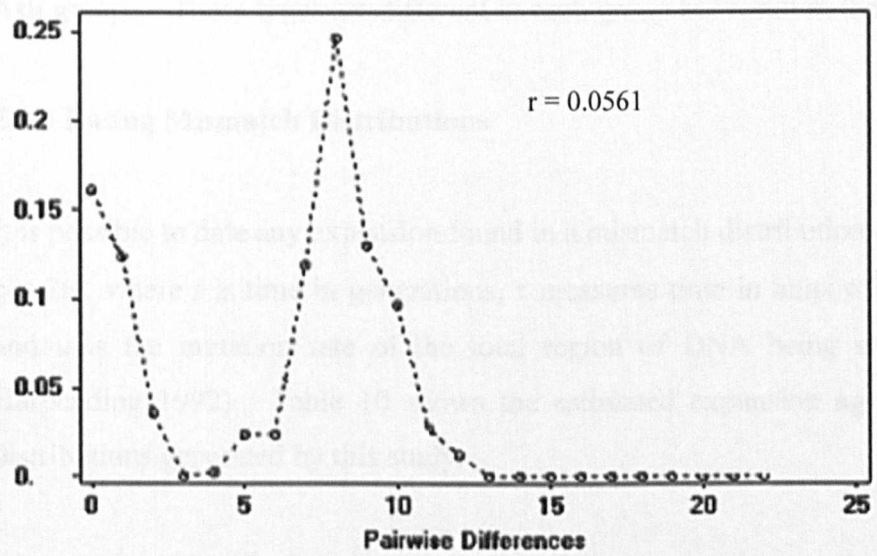
However, this pattern is maintained when the three Orang Asli groups are analysed separately (figures 51-53).

Figure 51 – Semang mismatch distribution



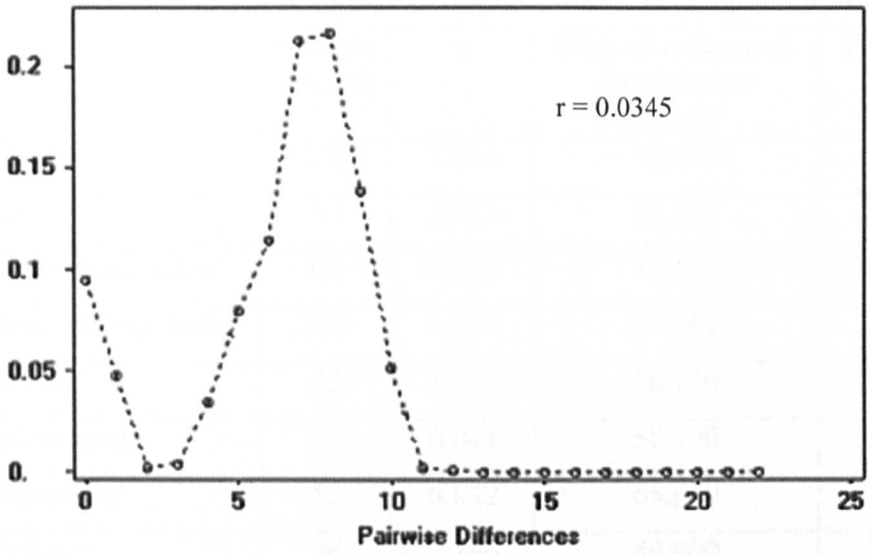
Y axis = frequency

Figure 52 – Senoi mismatch distribution



Y axis = frequency

Figure 53 – Aboriginal Malay mismatch distribution



Y axis = frequency

This pattern is most likely to be due to the fact that, because of the extensive drift which has occurred, two distinct sequence types have come to predominate in all three Orang Asli groups. These types are different in each group but result in the same effect.

5.6.3 Dating Mismatch Distributions

It is possible to date any expansion found in a mismatch distribution using the statistic τ. $\tau = 2ut$, where t is time in generations, τ measures time in units of $1 / 2u$ generations, and u is the mutation rate of the total region of DNA being studied (Rogers and Harpending 1992). Table 10 shows the estimated expansion ages of the mismatch distributions generated by this study.

As seen in table 10, the expansions seen in most populations date to approximately 60,000 years. This agrees with dates found from mismatch distributions in other studies of demographic patterns in human mtDNA (e.g. Rogers and Harpending 1992; Yao *et al.* 2002b); and also with the dates postulated for an out-of-Africa expansion (Mountain *et al.* 1995; Watson *et al.* 1997). Some of the populations in this study gave more recent expansion dates; however, these were the populations with bimodal or more ragged distributions and so the results should be treated with caution.

Table 10 – Dates of population expansions seen in Southeast Asia

Population	Sample size (n)	τ	Date of mismatch distribution (years)[1]	Date of mismatch distribution (years)[2]
Semang	110	3.186	30,600	28,400
Senoi	52	2.922	28,100	26,100
Aboriginal Malay	94	5.002	48,100	44,700
Total Orang Asli	256	6.028	57,961	53,821
Medan	42	5.838	56,100	52,100
Palembang	28	6.041	58,100	53,900
Pekanbaru	52	6.832	65,600	61,000
Padang	24	5.196	49,900	46,000
Bangka	34	6.561	63,100	58,600
Total Sumatra	180	6.567	63,100	58,700
Java - Tengger	36	4.929	47,400	44,000
Banjarmasin	89	6.738	64,800	60,200
Kota Kinabalu	68	6.135	59,000	54,800
Total Borneo	157	6.579	63,300	58,700
Ujung Padang	46	4.942	47,500	44,100
Toraja	64	4.294	41,300	38,300
Palu	38	3.536	34,000	31,600
Manado	89	4.231	40,700	37,800
Total Sulawesi	237	4.399	42,300	39,300
Bali - Denpasar	65	6.824	65,600	60,200
Lombok - Mataram	44	6.242	60,000	55,700
Sumba - Waingapu	50	6.553	63,000	58,500
Alor	45	6.389	61,400	57,000
Ambon	43	4.648	44,700	41,500
Total Island Southeast Asia	857	6.587	63,300	58,800

[1]Using divergence rate of 33%/nucleotide/Myr (Ward *et al.* 1991)

[2]Using divergence rate of 36%/nucleotide/Myr (Forster *et al.* 1996)

5.7 Analysis of Molecular Variance of Island Southeast Asia and the Orang Asli

Analysis of molecular variance (or AMOVA) enables population diversity to be calculated for different hierarchic levels: among groups, among populations within groups, and within populations (Excoffier *et al.* 1992). It is therefore possible to use it to group populations into different combinations to see which, if any, have significant differences in variation.

In all of the analyses performed on the current dataset (shown in table 11), the genetic variation within each group was always much higher than the variation between groups. When all 17 Island Southeast Asian populations were grouped together, only 1.31% of the variation was found to be between the populations; however, this value was still highly significant ($P < 0.00001$).

Significant differences were also found when the populations were split into two groups according to their position relative to the Wallace line. The variation between the Western populations (Pekanbaru, Medan, Padang, Palembang, Bangka, Tengger, Banjarmasin, Kota Kinabalu and Bali) and the Eastern populations (Mataram, Ujung Padang, Toraja, Palu, Manado, Waingapu, Ambon and Alor) was 0.17% ($P = 0.01857$). This difference was even more significant when a central group was separated out. The variation between the Western (Pekanbaru, Medan, Padang, Palembang and Bangka), Central (Tengger, Banjarmasin, Kota Kinabalu, Bali and Mataram) and Eastern groups (Ujung Padang, Toraja, Palu, Manado, Waingapu, Ambon and Alor) was 0.26% ($P = 0.00391$).

However, no significant difference was found when the groups were separated according to language. Most of the populations studied speak Western Malayo-Polynesian languages; however, those from Alor and Waingapu speak Central Malayo-Polynesian languages and the Ambonese speak a Malay-based Creole. No significant difference was found when the populations were divided along the above lines (language analysis 1 in table 11), the variation between groups was 0.07% ($P = 0.29396$). Equally, no significant difference was found when the Ambonese were

included with the Western Malayo-Polynesian groups (on the basis that Malay is a Western Malayo-Polynesian language, shown as language analysis 2 in table 11), the inter-group variation was then found to be 0.06% (P = 0.28928).

When the Orang Asli groups were analysed by themselves, 13.01% of the variation was found to be between the groups (P < 0.00001). Pairwise F_{ST} values showed the greatest difference to be between the Aboriginal Malay and the Semang (F_{ST} = 0.13702); however, the values for all combinations of the three groups were significantly different. When the Orang Asli groups were compared to those from Island Southeast Asia, 2.95% of the variation was found to be between the two groups (P = 0.01156).

Pairwise F_{ST} values showed most combinations of Island Southeast Asian groups to be significantly different from each other; however, the F_{ST} values were not as high as those for the Orang Asli. In particular, the populations from Kota Kinabalu, Mataram and Tengger are significantly different from all other groups. The population from Tengger seems to be the most different, possibly due to its high levels of haplogroup M10 which is absent in most other populations. Other particularly different combinations are: Ambon and Palembang (F_{ST} = 0.026), Ambon and Waingapu (F_{ST} = 0.020), Manado and Medan (F_{ST} = 0.024), Manado and Pekanbaru (F_{ST} = 0.020), Medan and Palembang (F_{ST} = 0.024), and Palembang and Pekanbaru (F_{ST} = 0.021). This emphasises the distinct differences which can be found across island locations, in this case Sumatra.

Table 11 – Results of analysis of molecular variance of Island Southeast Asian populations and the Orang Asli

	% of Variation	Significance
All Island Southeast Asian populations		
Among populations	1.31	Significant (P < 0.00001)
Within populations	98.69	
East vs. West		
Among groups	0.17	Significant (P = 0.01857)
Among populations within groups	1.22	Significant (P < 0.00001)
Within populations	98.61	Significant (P < 0.00001)
East vs. West vs. Central		
Among groups	0.26	Significant (P = 0.00391)
Among populations within groups	1.13	Significant (P < 0.00001)
Within populations	98.61	Significant (P < 0.00001)
Language analysis 1		
Among groups	0.07	Not significant (P = 0.29396)
Among populations within groups	1.29	Significant (P < 0.00001)
Within populations	98.64	Significant (P < 0.00001)
Language analysis 2		
Among groups	0.06	Not significant (P = 0.28928)
Among populations within groups	1.3	Significant (P < 0.00001)
Within populations	98.64	Significant (P < 0.00001)
All Orang Asli populations		
Among populations	13.01	Significant (P < 0.00001)
Within populations	86.99	
Orang Asli vs. Island Southeast Asia		
Among groups	2.95	Significant (P = 0.01156)
Among populations within groups	3.12	Significant (P < 0.00001)
Within populations	93.93	Significant (P < 0.00001)

6. Results – Phylogeography

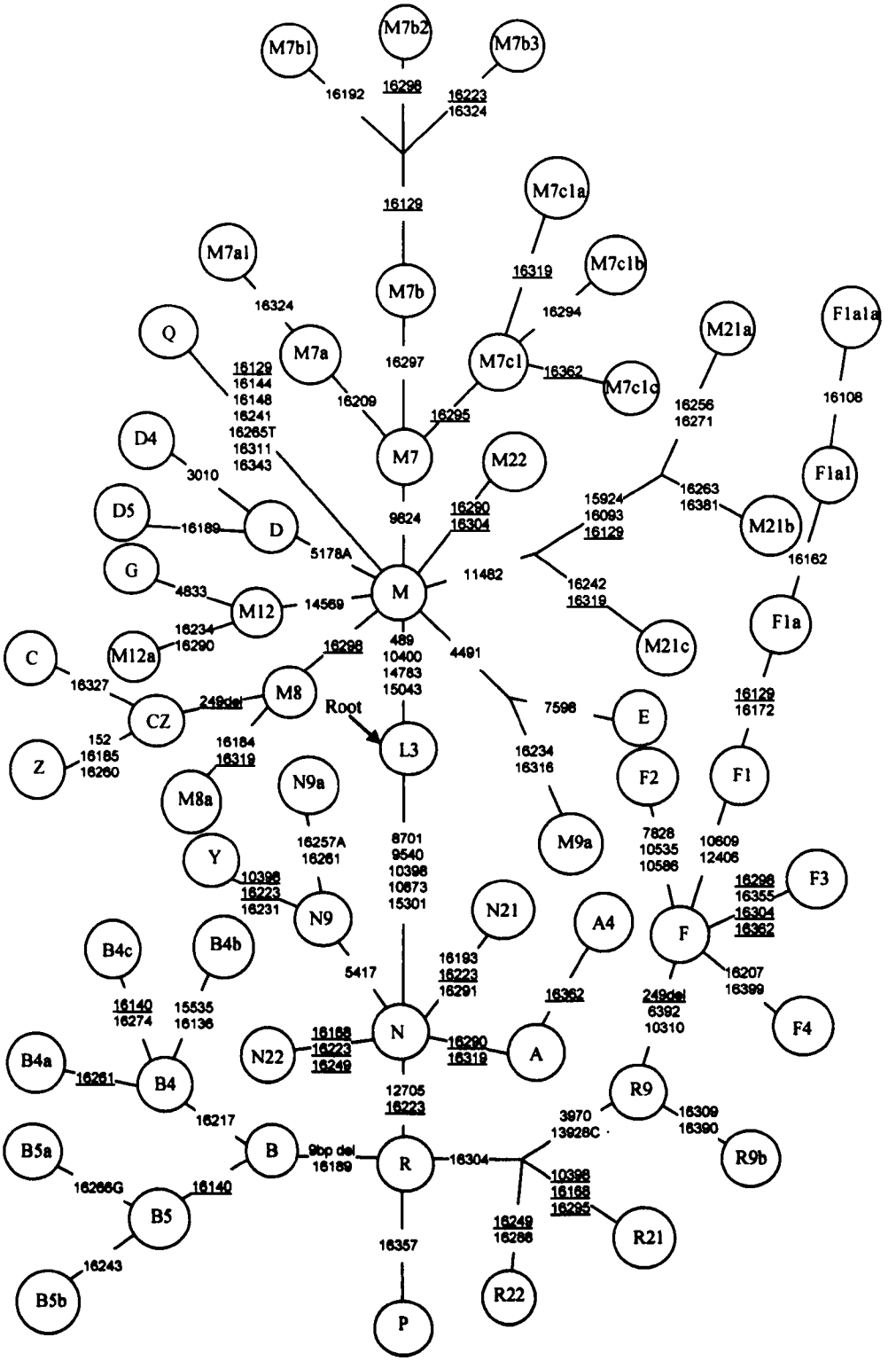
6. Results - Phylogeography

As described in section 1.2 the root of all Eurasian mtDNA haplogroups is found in the African haplogroup L3 (Watson *et al.* 1997) which is defined by the lack of a *Hpa*I site at np 3592 relative to the Cambridge Reference sequence (CRS [Anderson *et al*, 1981]) and which expanded from East Africa (possibly Ethiopia [Kivisild *et al.* 2004]) approximately 60,000 years ago (Mountain *et al.* 1995; Watson *et al.* 1997). The non-African subclusters of L3 themselves form 'macrohaplogroups' M (defined by a transition at np 10400 and the resultant gain of an *Alu*I site at np 10397) and N (defined by a transition at np 10873 and the subsequent loss of a *Mnl*I site at np 10871) (Torroni *et al.* 1994; Quintana-Murci *et al.* 1999). One of the branches of N forms another macrohaplogroup, termed R, which includes most of the main European and Asian haplogroups (Macaulay *et al.* 1999).

The mtDNA phylogeny has grown ever more complex over the past decade or so with the discovery of more and more haplogroups and subhaplogroups. The branches of macrohaplogroup M which are of most relevance to this study are as follows: haplogroup C (Torroni *et al.* 1992), haplogroup D (Torroni *et al.* 1992), haplogroup E (Torroni *et al.* 1994), haplogroup G (Torroni *et al.* 1994), haplogroup M7 (Kivisild *et al.* 2002), haplogroup Q (Forster *et al.* 2001) and haplogroup Z (Schurr *et al.* 1999). The branches of macrohaplogroup N which are of most relevance are: haplogroup A (Torroni *et al.* 1992), haplogroup N9 (Kivisild *et al.* 2002) and haplogroup Y (Schurr *et al.*1999). The most relevant branches of macrohaplogroup R are: haplogroup B (Torroni *et al.* 1992), haplogroup F (Torroni *et al.* 1994) and haplogroup P (Forster *et al.* 2001).

Several new haplogroups are defined for the first time in this study: M21 (including subgroups M21a, M21b and M21c), M22, N21, N22, N9a1, R9b, R21, R22, F3a and E1b. They are discussed further in later sections of this chapter. The phylogenetic relationships between the main known East Eurasian haplogroups and those found in Southeast Asia are shown in figure 54.

Figure 54 – Skeleton tree of Eurasian and Orang Asli haplogroups



6.1 Complete Sequencing

A number of samples (8 Orang Asli and one *Melayu* Malay) which were partly sequenced (i.e. HVS-I and HVS-II) in this study were completely sequenced by Antonio Torroni (University of Pavia) and colleagues; the results of this were used by Vincent Macaulay (University of Glasgow) to create a maximum likelihood tree of the Asian phylogeny. Four other Asian mtDNAs (Cambodian, Chinese Han, Bougainville Nan and Chinese Tujia) and one Ugandan were also completely sequenced. The control region variants for these samples are shown in table 12.

Table 12 – Control region variants in mtDNAs submitted to complete sequencing

ID	Source	Haplo-group	Variants in 16024–16569	Variants in 1–576 (263 315+C in addition)
1	Cambodia	F1a1a	108 129 162 172 304 311 519	73 249d 309+C 522d 523d
2	China	D4b	223 224 291 319 362 519	73 194 309+C 489 522d 523d
3	Bougainville	B4a	182d 183d 189 217 247 261 311 519	73 146 309+CC 522d 523d
4	China	A	209 214 223 290 319 362	64 73 152 235 309+C 522d 523d
5	Batek	M21a	93 129 223 256 271 362	73 152 309+C 489
6	Jahai	M21b	93 129 223 263 381 519	73 315+C 489
7	Temuan	N22	75 168 223 249	73 150 291+TA
8	Semelai	R9b	86 170 223 288 304 309 390	73 143 152 183 309+C 522d 523d 573+C
9	Semelai	N21	193 291 519	73 150 195 337d
10	Semelai	M21c	223 242 319 519	73 200 204 309+CC 489
11	Melayu	M9	136 217 223 319 381	73 94 173 204 482 489
12	Temuan	M22	93 184 223 290 304 519	73 489
13	Batek	R21	168 295 304	73 146 152 195 199 249
14	Uganda	L0f	129 169 172 173 187 189 223 230 239 278 311 327 368 519	143 146 152 185 189 247

Seventeen sequences from the literature (Andrews *et al.* 1999; Shin *et al.* 2000; Ingman *et al.* 2000; Maca-Meyer *et al.* 2001; Herrnstadt *et al.* 2002; Mishmar *et al.* 2003)

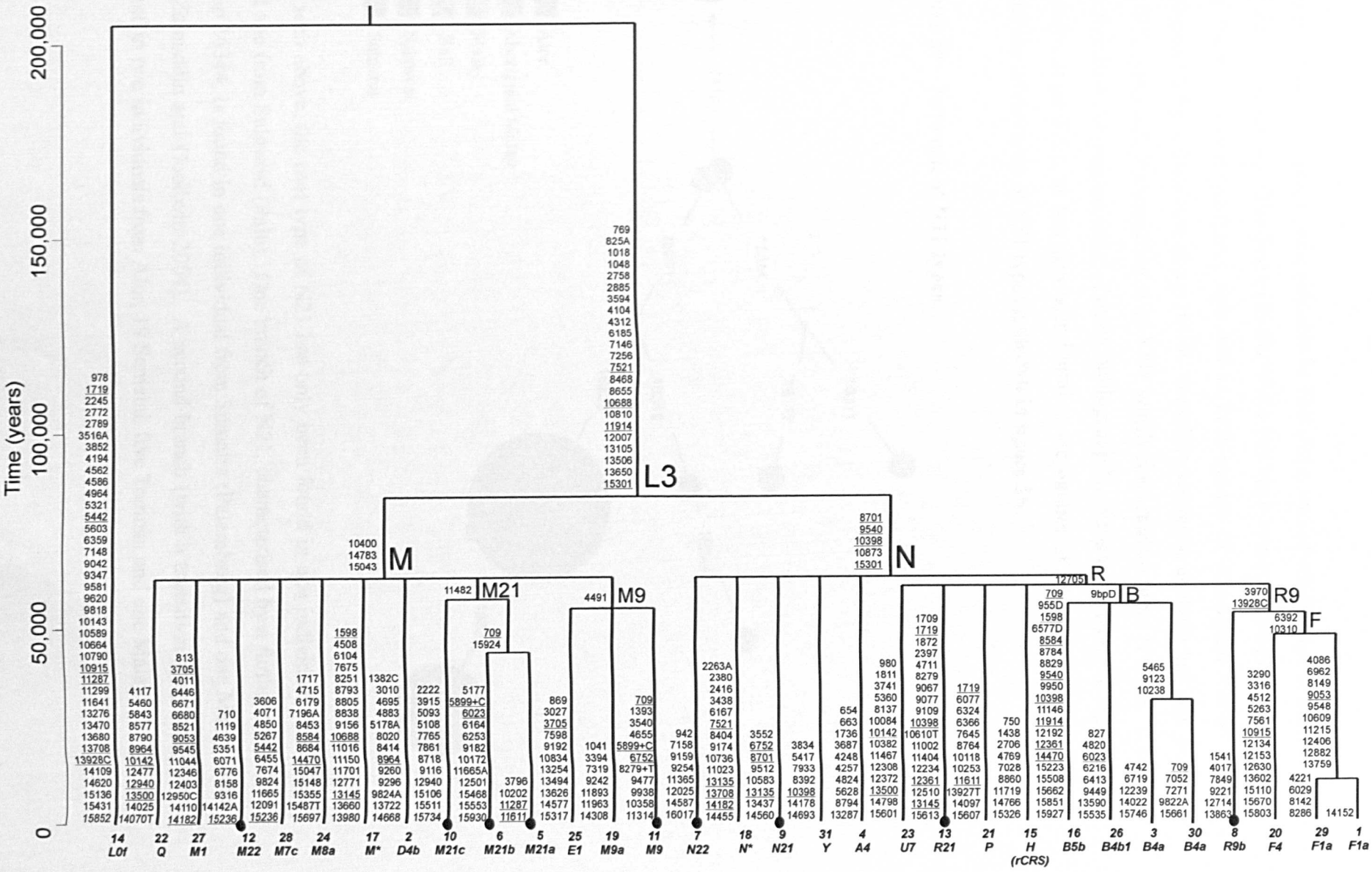
which represent deep-rooting lineages within the Asian phylogeny were also added to the above samples to create the maximum likelihood tree seen in figure 55. Two chimpanzee sequences (Horai *et al.* 1995) were used to root the whole tree. Ages of some of the major haplogroups, and the MRCA were calculated from this phylogeny, see table 13. This used a coding region mutation rate estimate of 1.26×10^{-8} which was calibrated on the basis of a human-chimpanzee split at 6.5 million years.

Table 13 - Divergence of relevant haplogroups in the mtDNA coding region phylogeny

Haplogroup	TMRCA [one SE] (substitutions per site)	TMRCA [one SE] (years)
L (all humans)	0.002578 [0.000278]	204,700 [22,100]
L3	0.001052 [0.000106]	83,500 [8,400]
M	0.000792 [0.000066]	62,900 [5,200]
M9	0.000691 [0.000091]	54,900 [7,200]
M21	0.000718 [0.000084]	57,000 [6,700]
M21a'b	0.000553 [0.000105]	43,900 [8,300]
N	0.000791 [0.000063]	62,800 [5,000]
R	0.000760 [0.000060]	60,300 [4,800]
R9	0.000669 [0.000073]	53,100 [5,800]
F	0.000600 [0.000077]	47,600 [6,100]
F1a	0.000135 [0.000057]	10,700 [4,500]
B	0.000704 [0.000071]	55,900 [5,600]
B4a	0.000398 [0.000106]	31,600 [8,400]

As shown above, the MRCA was dated to ~200,000 years which is similar to the estimate of Mishmar *et al.* (2003) but somewhat older than that of Ingman *et al.* (2000). Haplogroup L3 was dated to ~83,000 years with M and N both having ages of ~63,000 years. Haplogroup R appears to have diverged from N soon afterwards at ~60,000 years. This suggests an 'out of Africa' event between ~60,000 and ~100,000 years ago (Macaulay *et al.* 2005). The ages of all haplogroups mentioned in the text were summarised in table 7 in section 5.2 where they are placed according to their age. The vast majority of haplogroups date to the Pleistocene thus suggesting a greater age for much of the Island Southeast Asian population than is often cited.

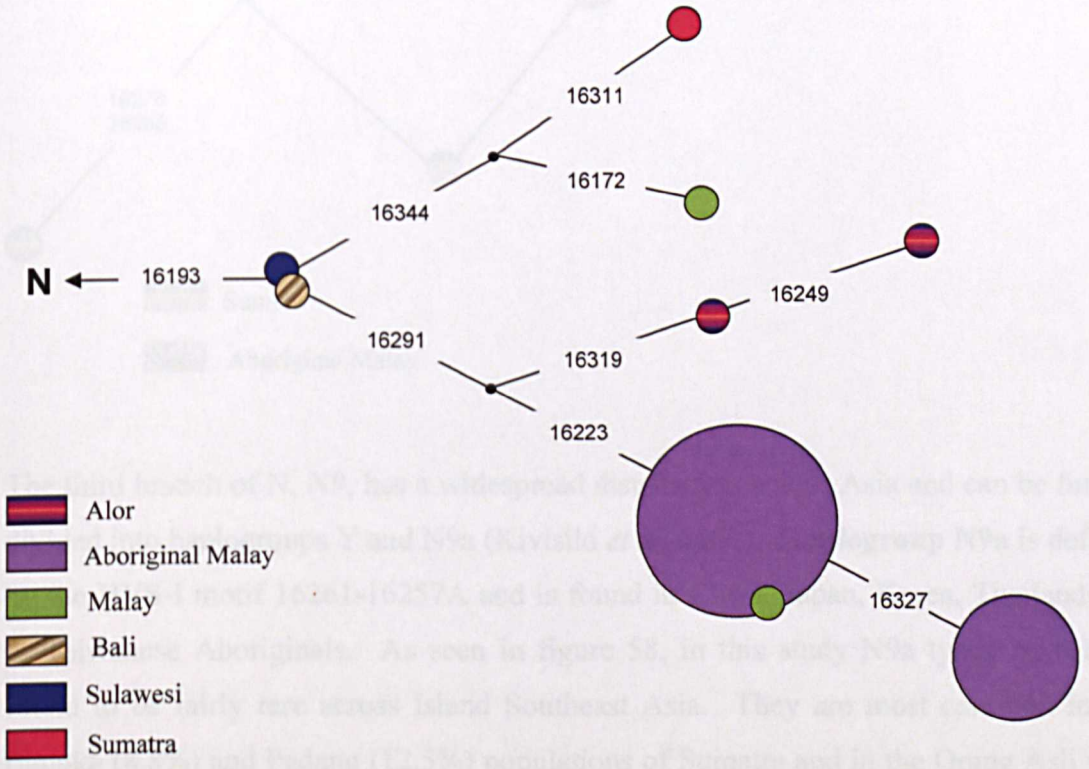
Figure 55 – Maximum likelihood tree of the Asian phylogeny



6.2 Macrohaplogroup N

Macrohaplogroup N has 5 main branches in East Asia, two of which were found for the first time in this study. The first of these, which has been termed **haplogroup N21**, has been found in Island Southeast Asia and in both groups of Aboriginal Malay. It is characterised by a transition at np 16193; complete sequence analysis has also shown that N21 types lack the transition at np 8701 which characterises other N types. It could therefore be a one-step ancestor of macrohaplogroup N, but is more likely to represent a reversion at np 8701, as has also been found in the dataset of Fuku *et al* (2002). The branching relationship of N21 types is shown in figure 56.

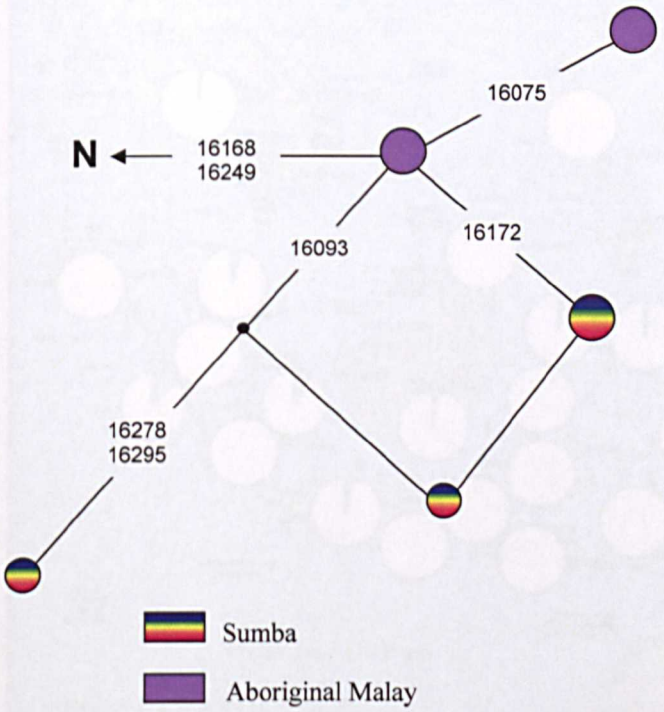
Figure 56 – Network of N21 types



As seen above, the root type of N21 has only been found in one individual from Bali and one from Sulawesi (Palu). One branch of N21, characterised by a further transition at np 16344, is found in one individual from Sumatra (Palembang) and one Malay (data of Zainuddin and Goodwin 2004). A second branch (with a transition at np 16291) is found in two individuals from Alor, 19 Semelai, five Temuan and one Malay.

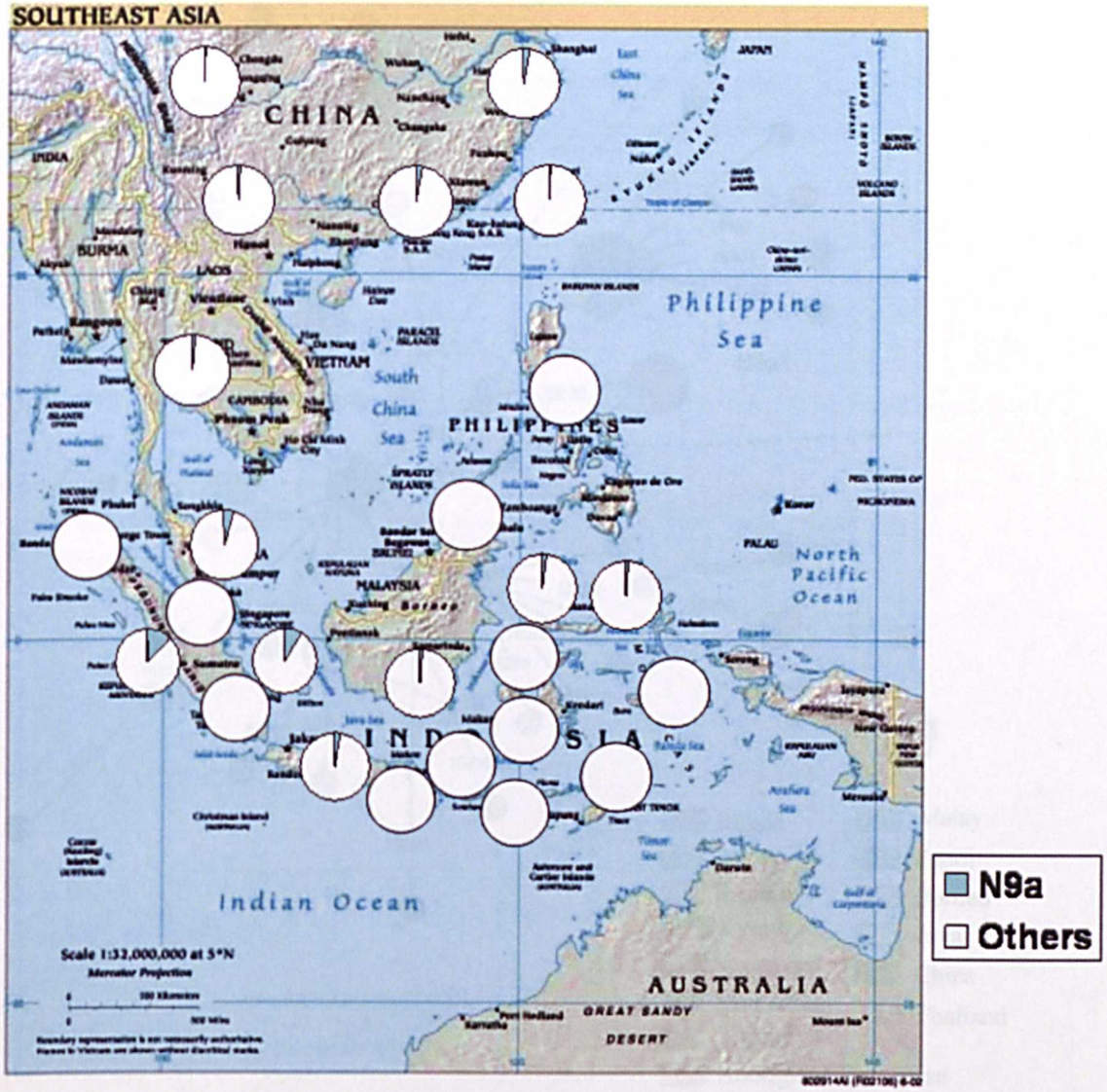
The second new branch of N found by this study has been termed **haplogroup N22** and is defined by the HVS-I motif 16168-16223-16249. It is even less common than N21 with the root type being only found in two Temuan. Derived types are found in two more Temuan and in four individuals from Sumba (figure 57).

Figure 57 – Network of N22 types



The third branch of N, **N9**, has a widespread distribution across Asia and can be further divided into haplogroups Y and N9a (Kivisild *et al.* 2002). **Haplogroup N9a** is defined by the HVS-I motif 16261-16257A and is found in China, Japan, Korea, Thailand and in Taiwanese Aboriginals. As seen in figure 58, in this study N9a types have been found to be fairly rare across Island Southeast Asia. They are most common in the Bangka (8.8%) and Padang (12.5%) populations of Sumatra and in the Orang Asli with only isolated examples found in Java, Borneo and Sulawesi.

Figure 58 – Map showing the distribution of haplogroup N9a in Southeast Asia

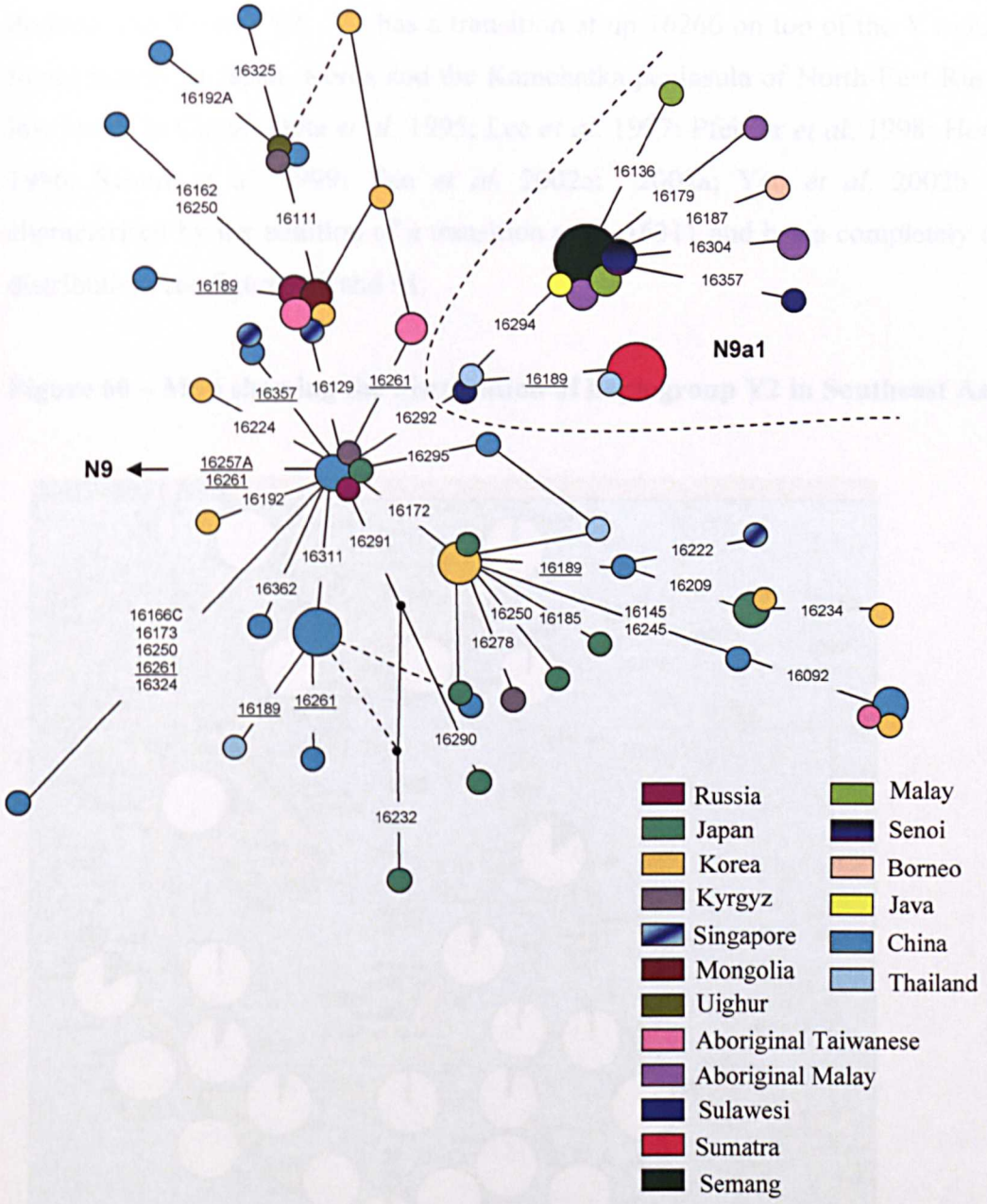


All maps courtesy of the CIA World Factbook.

Haplogroup frequency data can be found in Appendix IV.

A network of N9a types is shown in figure 59. As seen here, the root type of N9a is found in China, Japan, Russia and in the Kyrgyz people of Central Asia; more diverse types are common in China, Japan and Korea. N9a as a whole dates to 38,100 years (SE 11,200 years). All the N9a mtDNA types found in Island Southeast Asia and the Orang Asli belong to a subclade (termed N9a1) which is defined by a transition at np 16292.

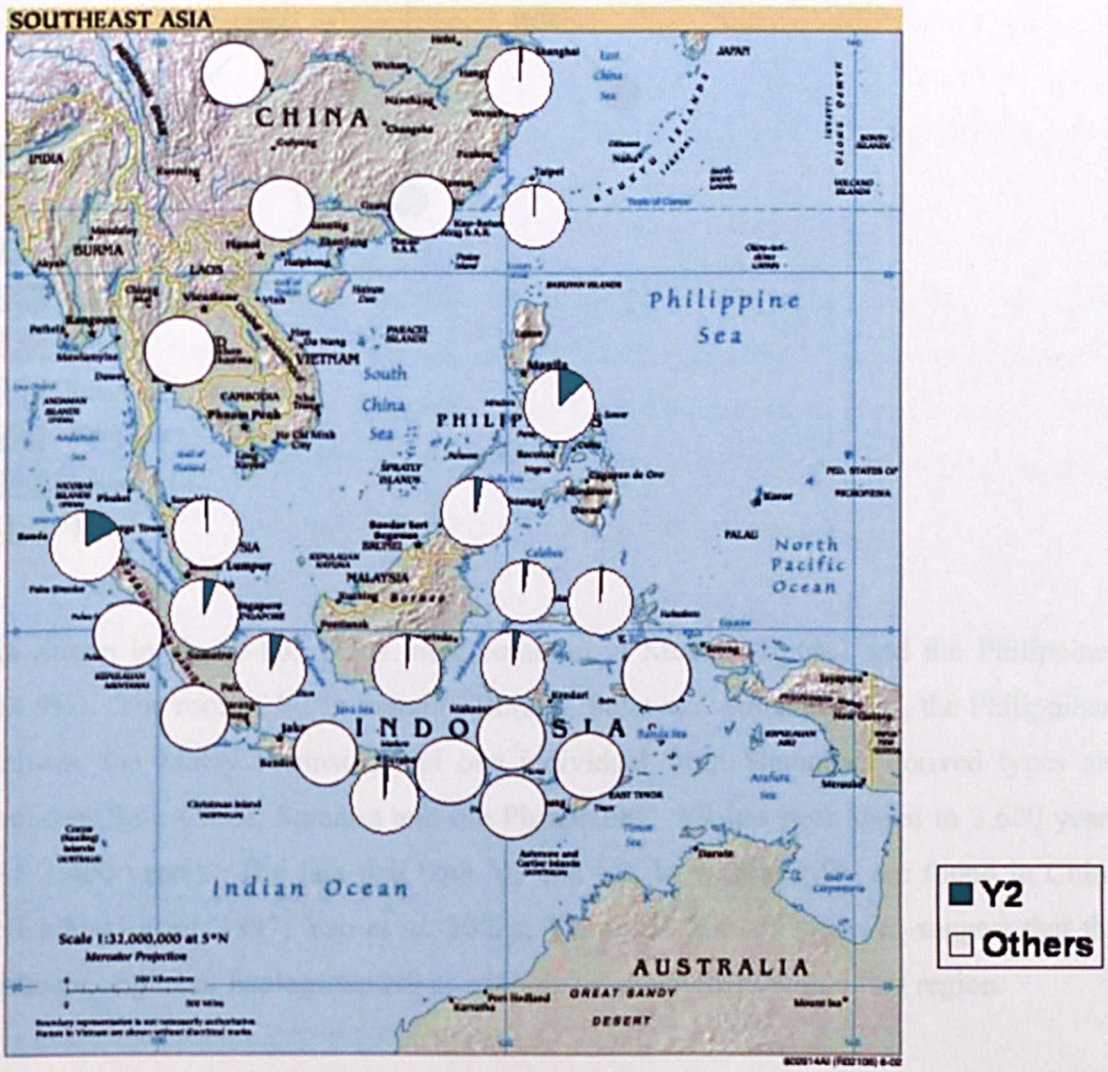
Figure 59 – Network of N9a and N9a1 types



The root type of N9a1 is found in Thailand and Sulawesi. All the N9a1 types found in Sumatra (3 from Bangka and 3 from Padang) and one type from Thailand form a branch of this subclade which also has a transition at np 16189. The main branch of N9a1, with the addition of a transition at np 16294, has been dated to 5,500 years (SE 2,600 years) and is found in the Orang Asli, Sulawesi, Java and Borneo.

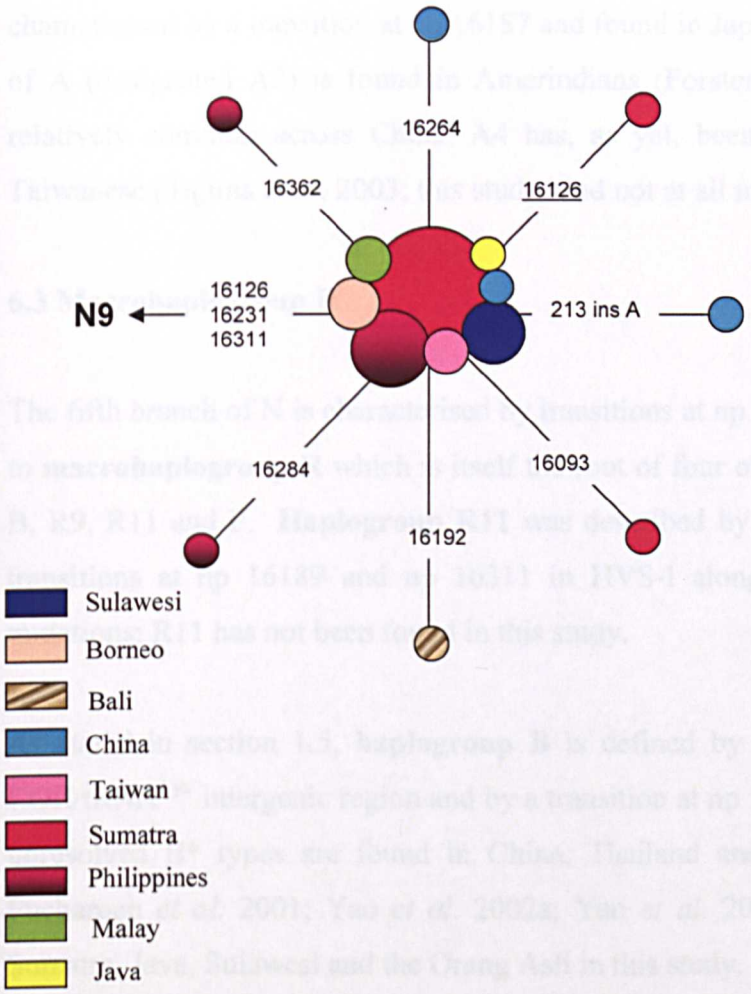
The second branch of N9 is **haplogroup Y** (HVS-I motif 16126-16231) which itself is divided into Y1 and Y2. Y1 has a transition at np 16266 on top of the Y motif and is found mainly in Japan, Korea and the Kamchatka peninsula of North-East Russia with low levels in China (Oota *et al.* 1995; Lee *et al.* 1997; Pfeiffer *et al.* 1998; Horai *et al.* 1996; Schurr *et al.* 1999; Yao *et al.* 2002a; 2002a; Yao *et al.* 2002b. Y2 is characterised by the addition of a transition at np 16311 and has a completely different distribution, see figures 60 and 61.

Figure 60 – Map showing the distribution of haplogroup Y2 in Southeast Asia.



Haplogroup frequency data can be found in Appendix IV

Figure 61 – Network of Y2 types



As shown in figure 60, Y2 is most common in Medan (16.6%) and the Philippines (14.9%). The root of Y2 is found in Borneo, Sulawesi, Sumatra, Java, the Philippines, Taiwan, the Malay Peninsula and one individual from Shanghai; derived types are found in Bali, China, Sumatra and the Philippines. Y2 has been dated to 3,600 years (SE 1,400 years). The fact that both Y1 and Y2, as well as N9a, are found in China (Nishimaki *et al.* 1997; Yao *et al.* 2002a; Yao *et al.* 2002b) seems to suggest that the ultimate origins of haplogroup N9 as a whole lie somewhere within that region.

The fourth branch of N, which is characterised by transitions at np 16290 and np 16319 in HVS-I and by a number of mutations in the coding region (see Kivisild *et al.* 2002), is designated **haplogroup A** and is common across China, Japan and Korea (Lee *et al.* 1997; Nishimaki *et al.* 1999; Yao *et al.* 2002a; Yao *et al.* 2002b). Two branches of A

have been found in East Asia; **A4** which is characterised by a transition at np 16362 and is found predominantly in the Yunnan region of southern China, and **A5** which is characterised by a transition at np 16187 and found in Japan and Korea; a further branch of A (designated A2) is found in Amerindians (Forster *et al.* 1996). Despite being relatively common across China, A4 has, as yet, been found in only 5 Aboriginal Taiwanese (Tajima *et al.* 2003; this study) and not at all in Island Southeast Asia.

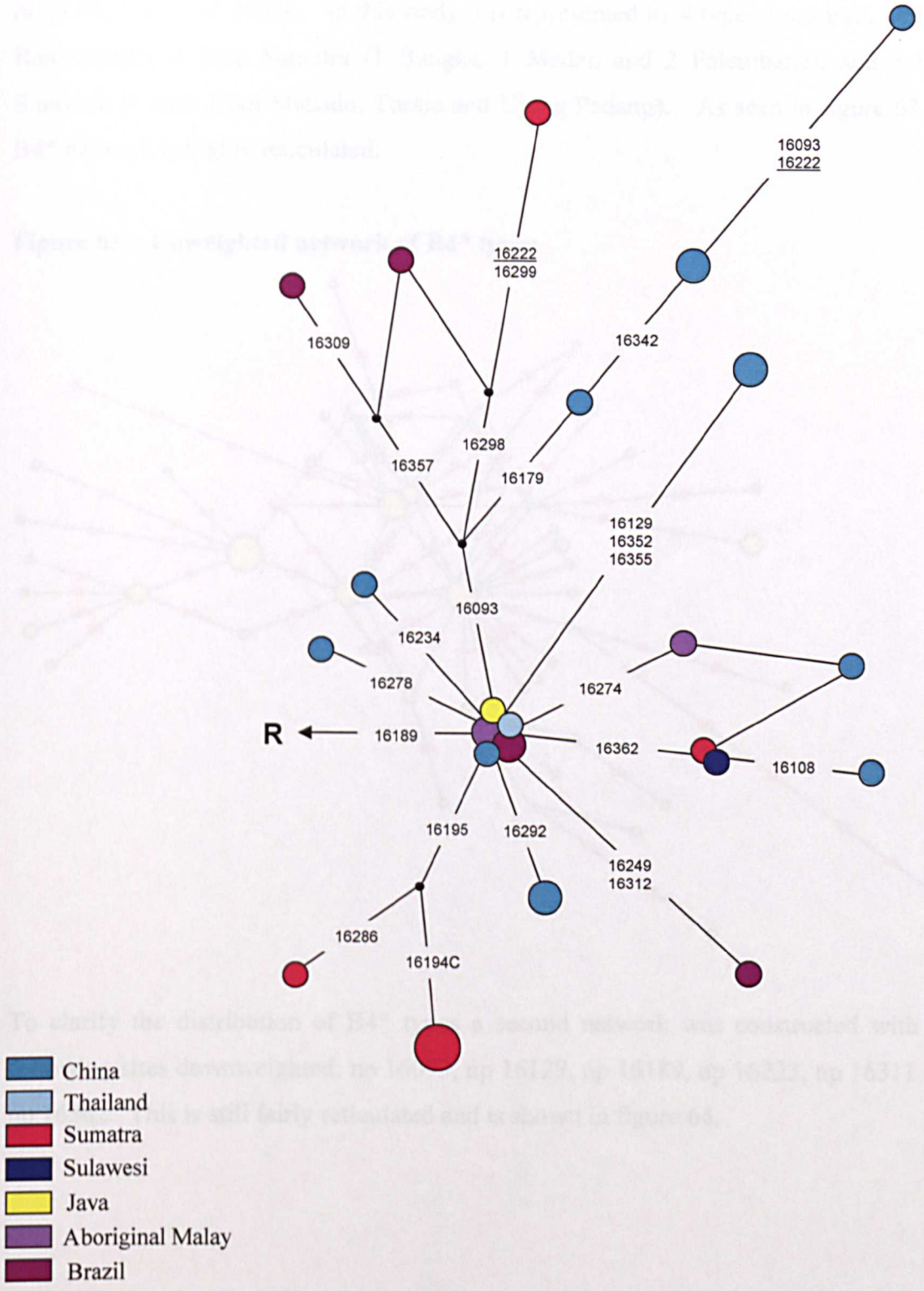
6.3 Macrohaplogroup R

The fifth branch of N is characterised by transitions at np 12705 and np 16223 and leads to **macrohaplogroup R** which is itself the root of four of the main Asian haplogroups: B, R9, R11 and P. **Haplogroup R11** was described by Kong *et al.* (2003) as having transitions at np 16189 and np 16311 in HVS-I along with several coding region mutations; R11 has not been found in this study.

As stated in section 1.5, **haplogroup B** is defined by a 9 base-pair deletion in the COII/tRNA^{Lys} intergenic region and by a transition at np 16189. As shown in figure 62, unresolved B* types are found in China, Thailand and Brazil (Santos *et al.* 1996; Fucharoen *et al.* 2001; Yao *et al.* 2002a; Yao *et al.* 2002b) and have been found in Sumatra, Java, Sulawesi and the Orang Asli in this study.

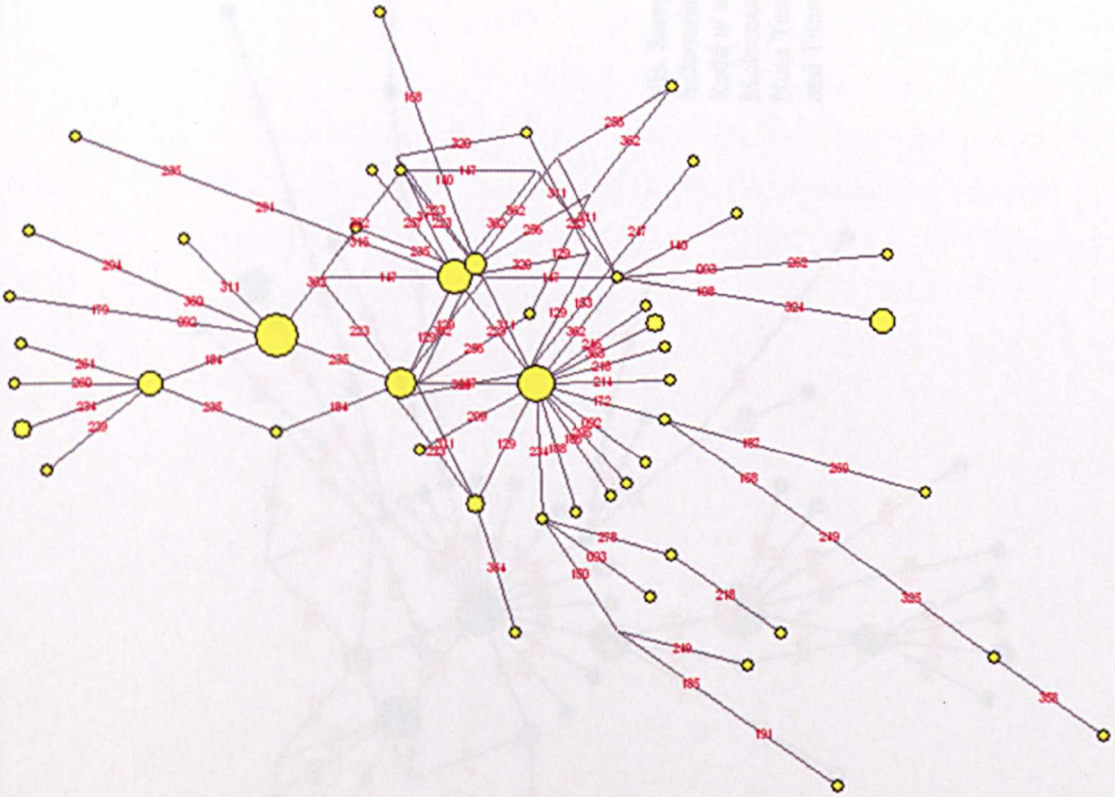
Of the B* types found in this study, five Sumatrans form a cluster defined by a transition at np 16195. One of these, from Padang, has an additional transition at np 16286 while the other four (3 from Padang and 1 from Pekanbaru) have a transversion from A to C at np 16194.

Figure 62 – Network of B* types



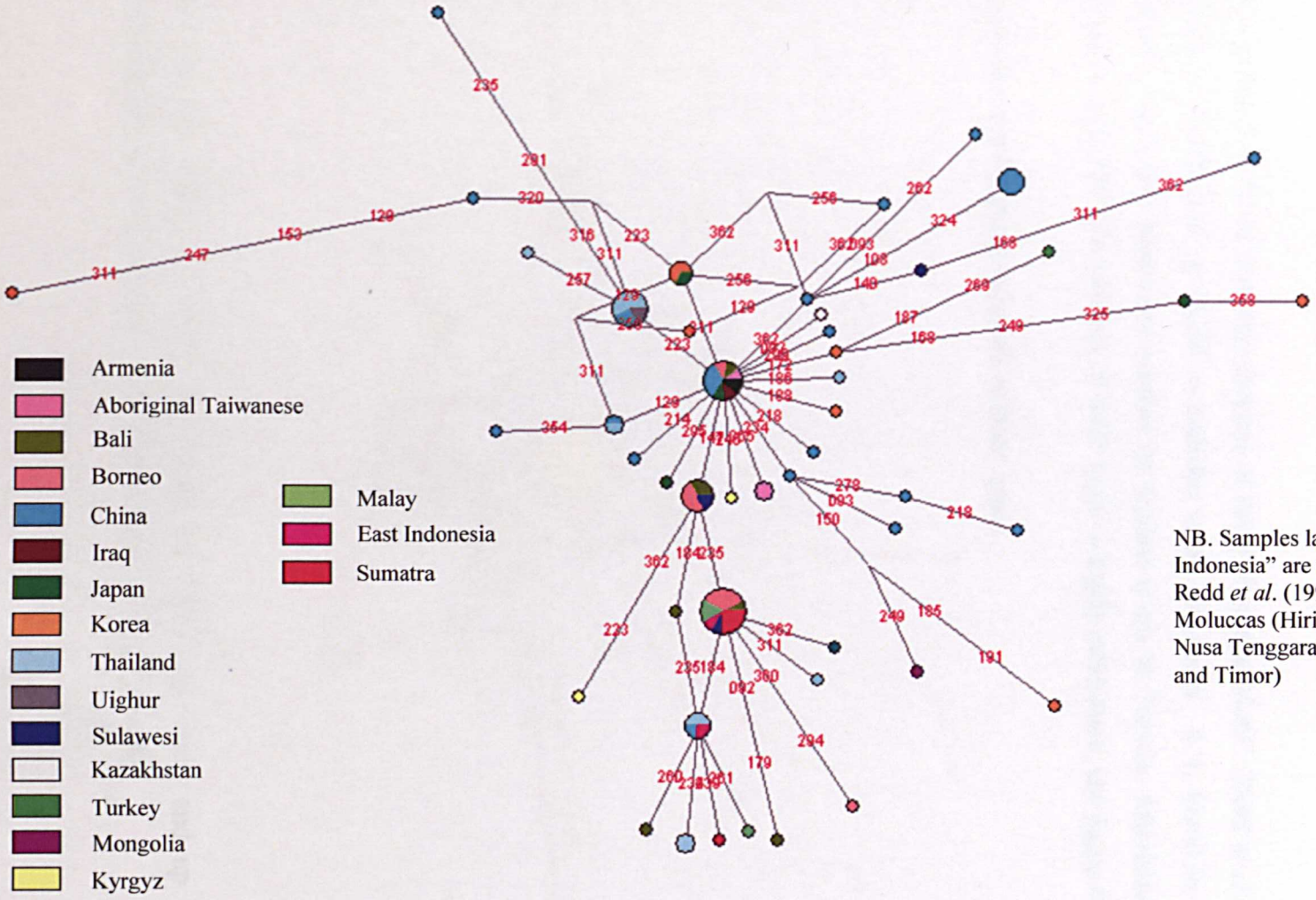
B4 is further defined by a transition at np 16217 and is widespread across East Asia and into the Pacific (e.g. Lum *et al.* 1998; Pfeiffer *et al.* 1998; Kivisild *et al.* 2002; Yao *et al.* 2002; Yao *et al.* 2002b. In this study it is represented by 4 types from Bali, 10 from Banjarmasin, 4 from Sumatra (1 Bangka, 1 Medan and 2 Palembang), and 3 from Sulawesi (1 each from Manado, Toraja and Ujung Padang). As seen in figure 63, the B4* network is highly reticulated.

Figure 63 – Unweighted network of B4* types



To clarify the distribution of B4* types a second network was constructed with the following sites downweighted: np 16093, np 16129, np 16189, np 16223, np 16311 and np 16362. This is still fairly reticulated and is shown in figure 64.

Figure 64 – Weighted network of B4* types

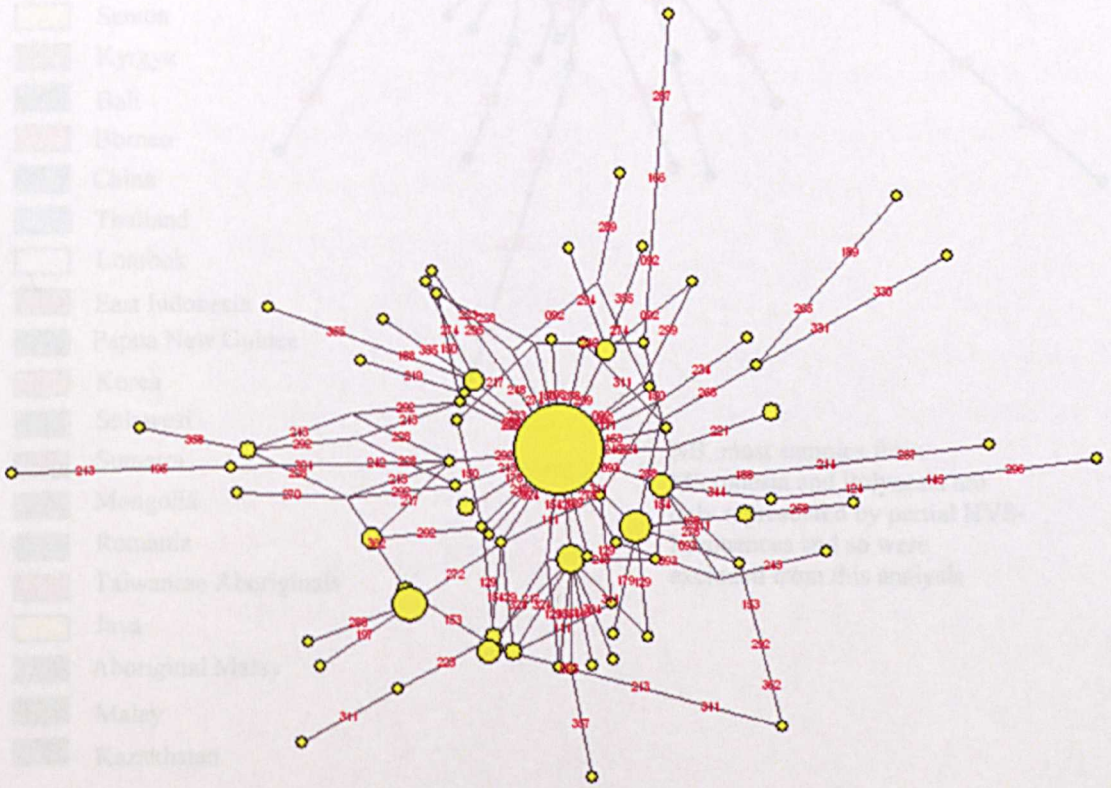


NB. Samples labelled "East Indonesia" are from the data of Redd *et al.* (1995) and are from the Moluccas (Hiri and Ternate) and Nusa Tenggara (Alor, Flores, Roti and Timor)

As seen above, B4* types are most common in China (especially Yunnan province) and are also common in Korea and Thailand. B4 dates to 33,600 years (SE 8,600 years). The majority of the B4* types from Island Southeast Asia form a subclade with a further transition at np 16147 which is also found in a handful of samples from China, Thailand, Malaysia, Japan and the Moluccas (Redd *et al.* 1995; Seo *et al.* 1998; Fucharoen *et al.* 2001; Yao *et al.* 2002b; Zainuddin and Goodwin 2004), and which dates to 22,700 years (SE 16,000 years)

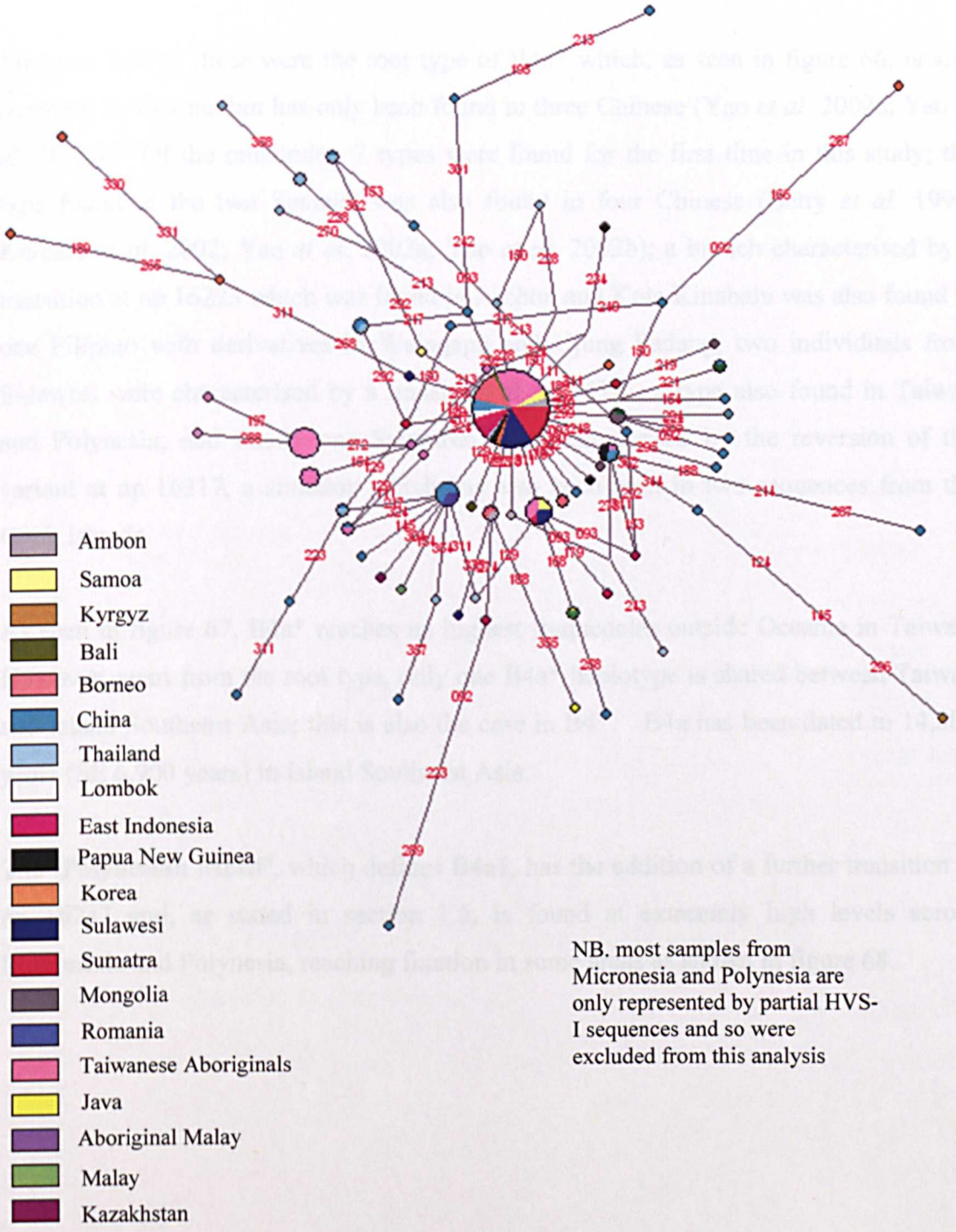
Haplogroup B4a is the one-step ancestor of the 'Polynesian Motif' (Redd *et al.* 1995) and has a transition at np 16261 in addition to the B4 motif. It is found in China, Thailand, Japan and Korea but reaches its highest levels in Taiwan, Micronesia and Polynesia. Like B4*, the network of B4a* types is highly reticulated, see figure 65.

Figure 65 – Unweighted network of B4a* types



Even with np 16093, np 16129, np 16189, np 16223, np 16311 and np 16362 downweighted, the B4a* network is still very reticulated, see figure 66.

Figure 66 – Weighted network of B4a* types



In this study, B4a* types were found in the following: 4 samples from Ambon, 1 from Bali, 7 from Banjarmasin, 5 from Kota Kinabalu, 3 from Manado, 2 from Mataram, 2 from Padang, 8 from Pekanbaru, 1 from Palembang, 4 from Toraja, 5 from Ujung

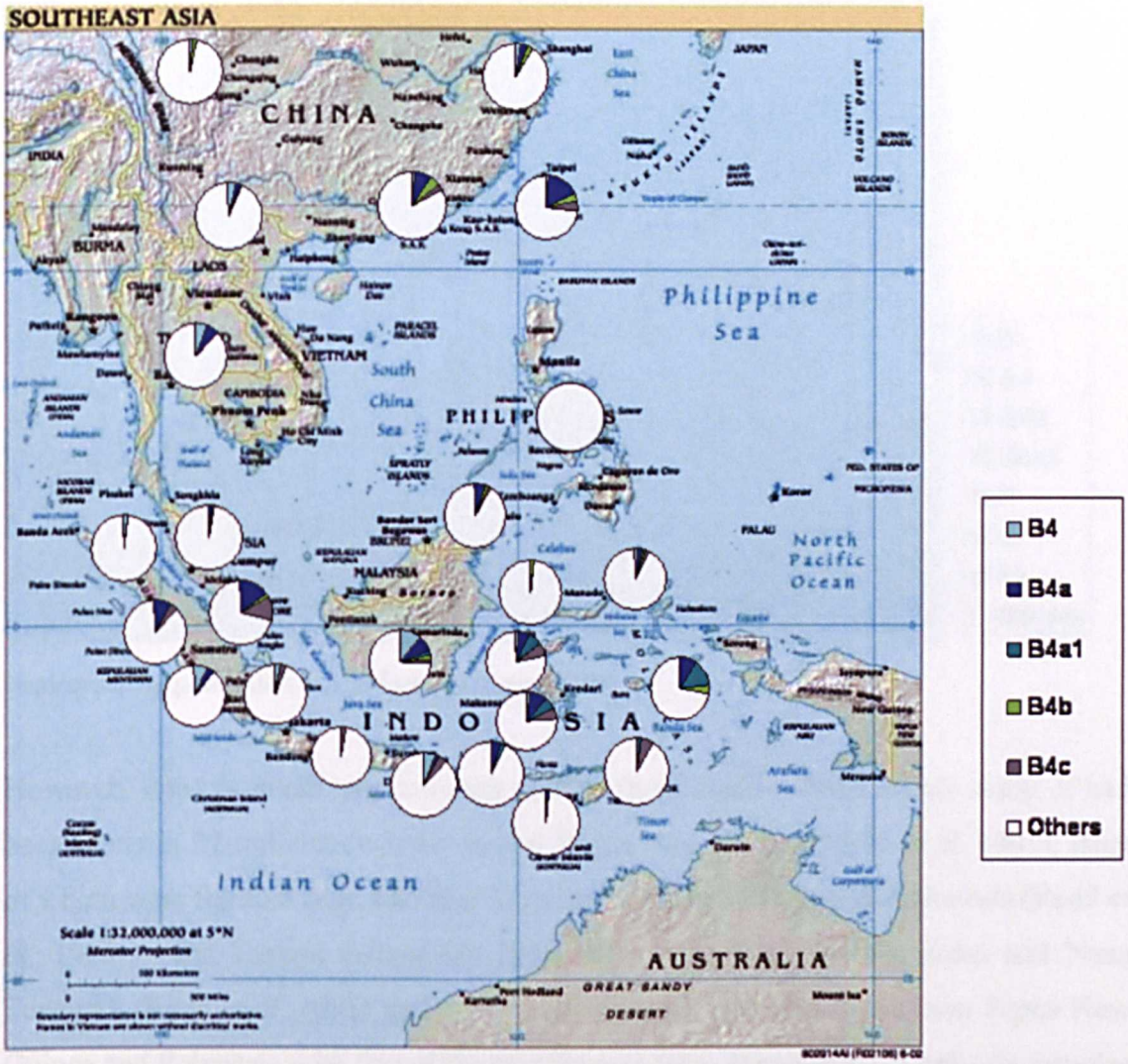
Padang and 1 from Waingapu. The highest frequency of B4a* (15.4%) was seen in Pekanbaru. B4a* was also found in two Semelai.

Just over half of these were the root type of B4a* which, as seen in figure 66, is also common in Taiwan, but has only been found in three Chinese (Yao *et al.* 2002a; Yao *et al.* 2002b). Of the remainder, 7 types were found for the first time in this study; the type found in the two Semelai was also found in four Chinese (Betty *et al.* 1996; Kivisild *et al.* 2002; Yao *et al.* 2002a; Yao *et al.* 2002b); a branch characterised by a transition at np 16223 which was found in Ambon and Kota Kinabalu was also found in one Filipino with derivatives in Waingapu and Ujung Padang; two individuals from Sulawesi were characterised by a transition at np 16311, a type also found in Taiwan and Polynesia; and finally one Sumatran was characterised by the reversion of the variant at np 16217, a situation which has also been seen in two sequences from the Cook Islands.

As seen in figure 67, B4a* reaches its highest frequencies outside Oceania in Taiwan. However, apart from the root type, only one B4a* haplotype is shared between Taiwan and Island Southeast Asia; this is also the case in B4*. B4a has been dated to 14,800 years (SE 6,900 years) in Island Southeast Asia.

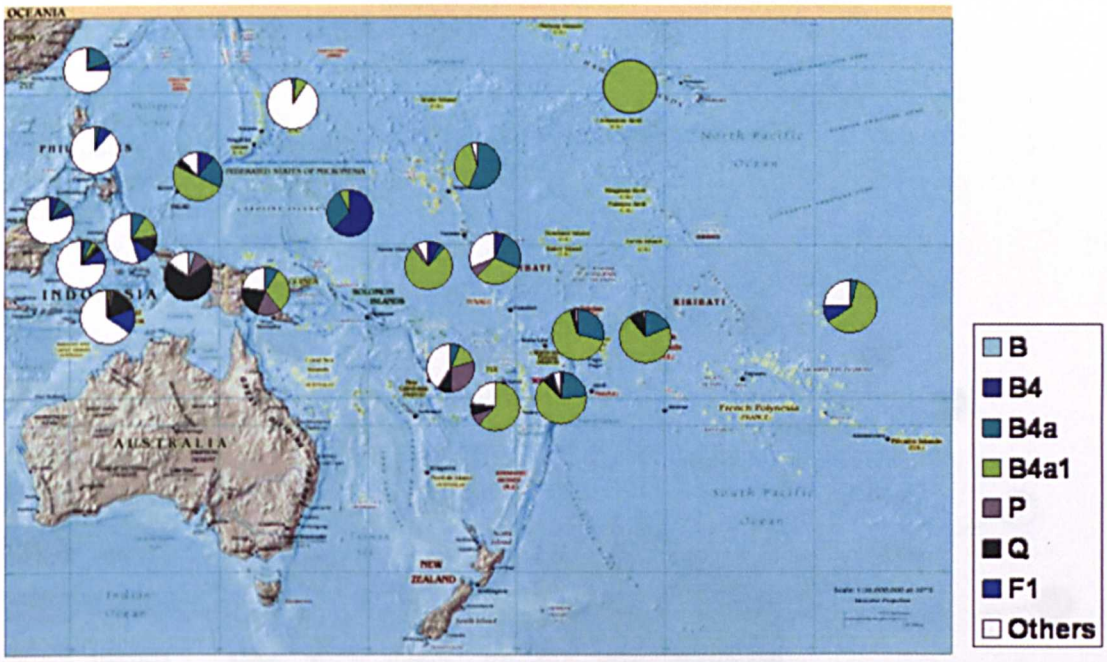
The **'Polynesian Motif'**, which defines **B4a1**, has the addition of a further transition at np 16247 and, as stated in section 1.5, is found at extremely high levels across Micronesia and Polynesia, reaching fixation in some areas as shown in figure 68.

Figure 67 – Map showing the distribution of B4 haplogroups in Southeast Asia



Haplogroup frequency data can be found in Appendix IV

Figure 68 – Map showing the distribution of the main Oceanic haplogroups



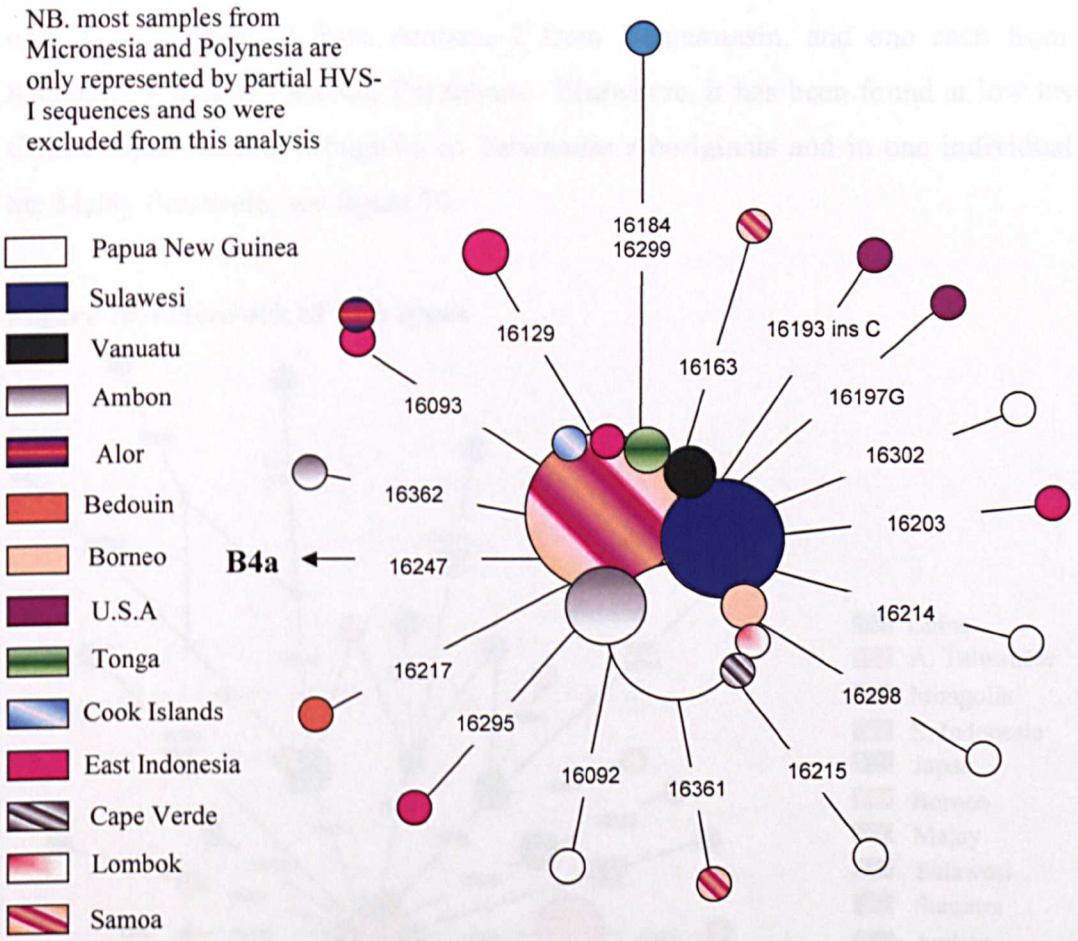
Haplogroup frequency data can be found in Appendix IV

However, B4a1 is much less common outside these areas. Prior to this study, it had been found in 22 individuals from coastal Papua New Guinea (Redd *et al.* 1995), most of which were the root type, and also in six individuals from Eastern Indonesia (Redd *et al.* 1995). The Eastern Indonesian individuals were from the Moluccas and Nusa Tenggara (Redd *et al.* 1995) and were much more diverse than those from Papua New Guinea and Polynesia with five different sequence types represented in only six samples, only one of which was the root type. Other than this, B4a1 has only been found in one individual from Liaoning province in Northeastern China, two African Americans, one individual from Cape Verde and possibly one from a Bedouin individual (Di Rienzo *et al.* 1991; Handt *et al.* 1998; Brehm *et al.* 2002; Yao *et al.* 2002a).

In this study, despite the extensive sampling across Island Southeast Asia, only 19 individuals were found to belong to haplogroup B4a1. These were found in the following locations: 1 from Alor (representing 2% of the population), 6 from Ambon (14%), 2 from Banjarmasin (2%), 1 from Lombok (2%), 1 from Manado (1%), 5 from the Toraja (8%) and 3 from Ujung Padang (6.5%). Therefore, B4a1 is not found

further west than Southeastern Borneo and Lombok and is most common in the Moluccas.

Figure 69 – Network of B4a1 types

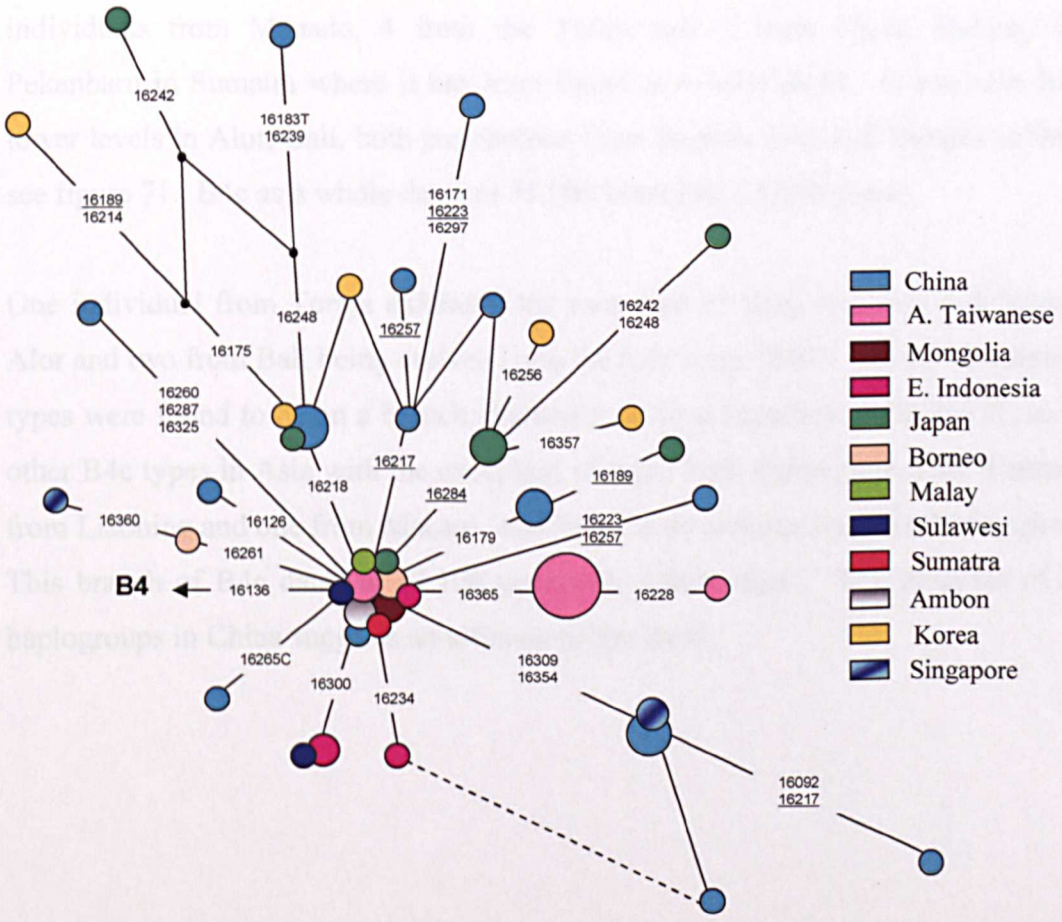


In contrast to the high diversity found in Eastern Indonesia by Redd *et al.* (1995), all but two of these were the root type. The root type is particularly common in Sulawesi where it is found in 9 individuals; in contrast, no derived types are found in Sulawesi which suggests that these could be the result of a recent migration. The derived types found in Island Southeast Asia were both one-step derivatives. One of these was found in Ambon and had an additional transition at np 16362. The second derived type found in Island Southeast Asia had an additional transition at np 16093 and was found in Alor, this type was also found in the data of Redd *et al.* (1995). A network of B4a1 types is shown in figure 69. B4a1 dates to 5,700 years (SE 2,700 years) in Island Southeast

Asia. However, when the samples from Sulawesi are removed the age increases to 8,800 years (SE 4,200 years)

There are two other branches of B4. **Haplogroup B4b** is characterised by a transition at np 16136 in HVS-I and is much less common than B4a. In this study, it was found in only 8 individuals: 2 from Ambon, 2 from Banjarmasin, and one each from Kota Kinabalu, Manado, Palu and Pekanbaru. Elsewhere, it has been found at low levels in China, Japan, Korea, Mongolia, in Taiwanese Aboriginals and in one individual from the Malay Peninsula, see figure 70.

Figure 70 – Network of B4b types



Of the B4b mtDNAs found in Island Southeast Asia, 6 were the root type with the other two being one-step derivatives. Of these, one found in Banjarmasin also had a transition at np 16261 (a further one-step derivative of this type has been found in Singapore); the other was found in Palu and also contained a transition at np 16300, this

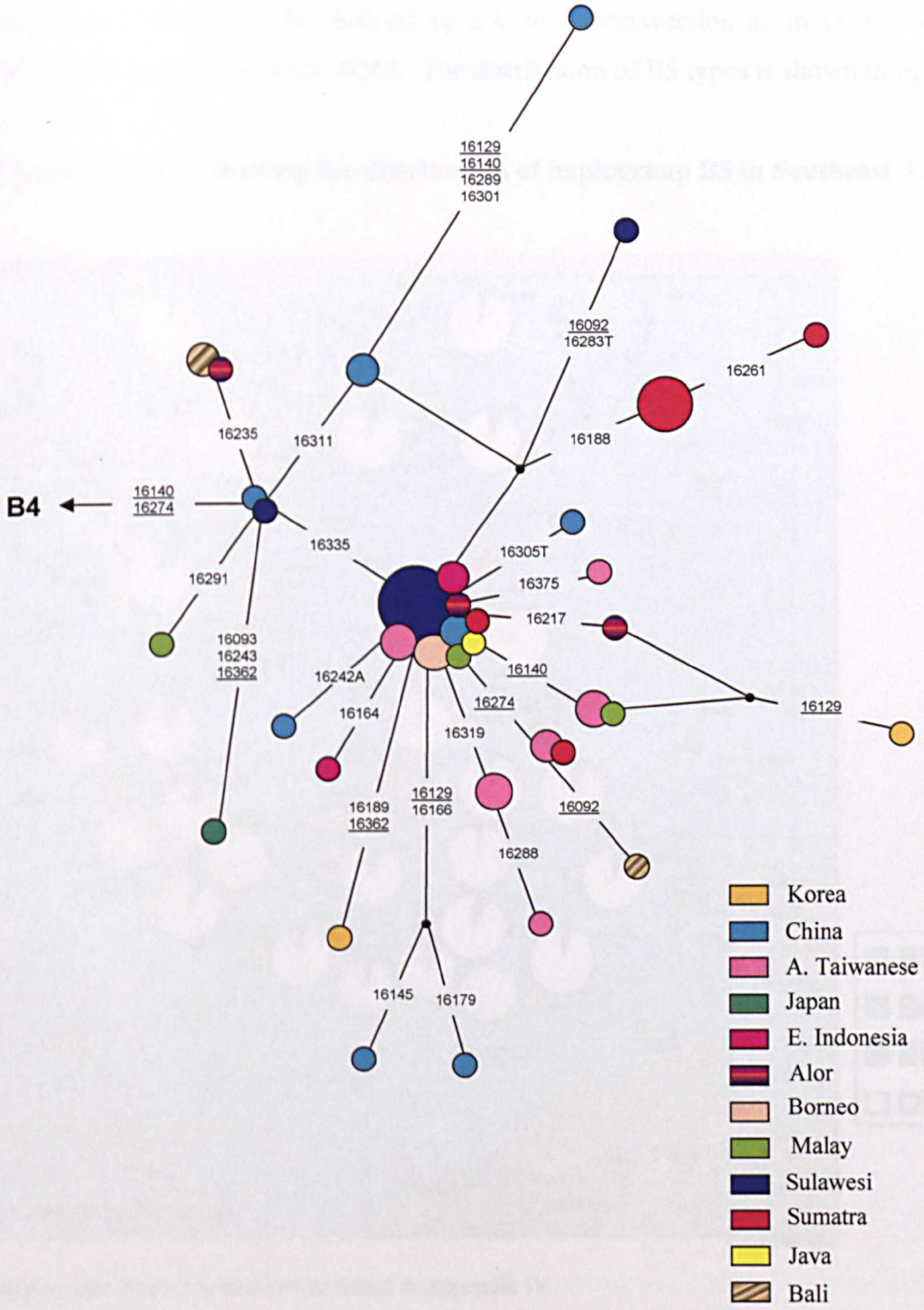
type was found in three of the Eastern Indonesian individuals studied by Redd *et al.* (1995). B4b dates to 27,000 years (SE 6,100 years) overall and to 8,400 years (SE 5,600 years) in Island Southeast Asia

.

Haplogroup B4c is characterised by transitions at np 16140 and np 16274 in HVS-I and, again, is less common than B4a. Prior to this study, it had only been found at relatively low levels in China, Taiwan, the Malay Peninsula, Eastern Indonesia, and in one individual from Japan (Redd *et al.* 1995; Seo *et al.*1998; Kivisild *et al.* 2002; Yao *et al.* 2002a; Tajima *et al.* 2003; Zainuddin and Goodwin 2004). Just under half of all known B4c mtDNAs have been found in Island Southeast Asia during the course of this study. Specifically, B4c is most common in Sulawesi (where it has been found in 2 individuals from Manado, 4 from the Toraja and 3 from Ujung Padang) and in Pekanbaru in Sumatra where it has been found in 6 individuals. It was also found at lower levels in Alor, Bali, both populations from Borneo, Java and Bangka in Sumatra, see figure 71. B4c as a whole dates to 35,700 years (SE 15,600 years).

One individual from Toraja exhibited the root type of B4c, with one individual from Alor and two from Bali being derived from the root at np 16235. All the remaining B4c types were found to be on a branch characterised by a transition at np 16335, as are all other B4c types in Asia with the exception of three from China (one from Yunnan, one from Liaoning and one from Macau), one from Japan and one from the Malay peninsula. This branch of B4c dates to 17,100 years (SE 4,800 years). The presence of all B4 haplogroups in China suggests an ultimate origin there.

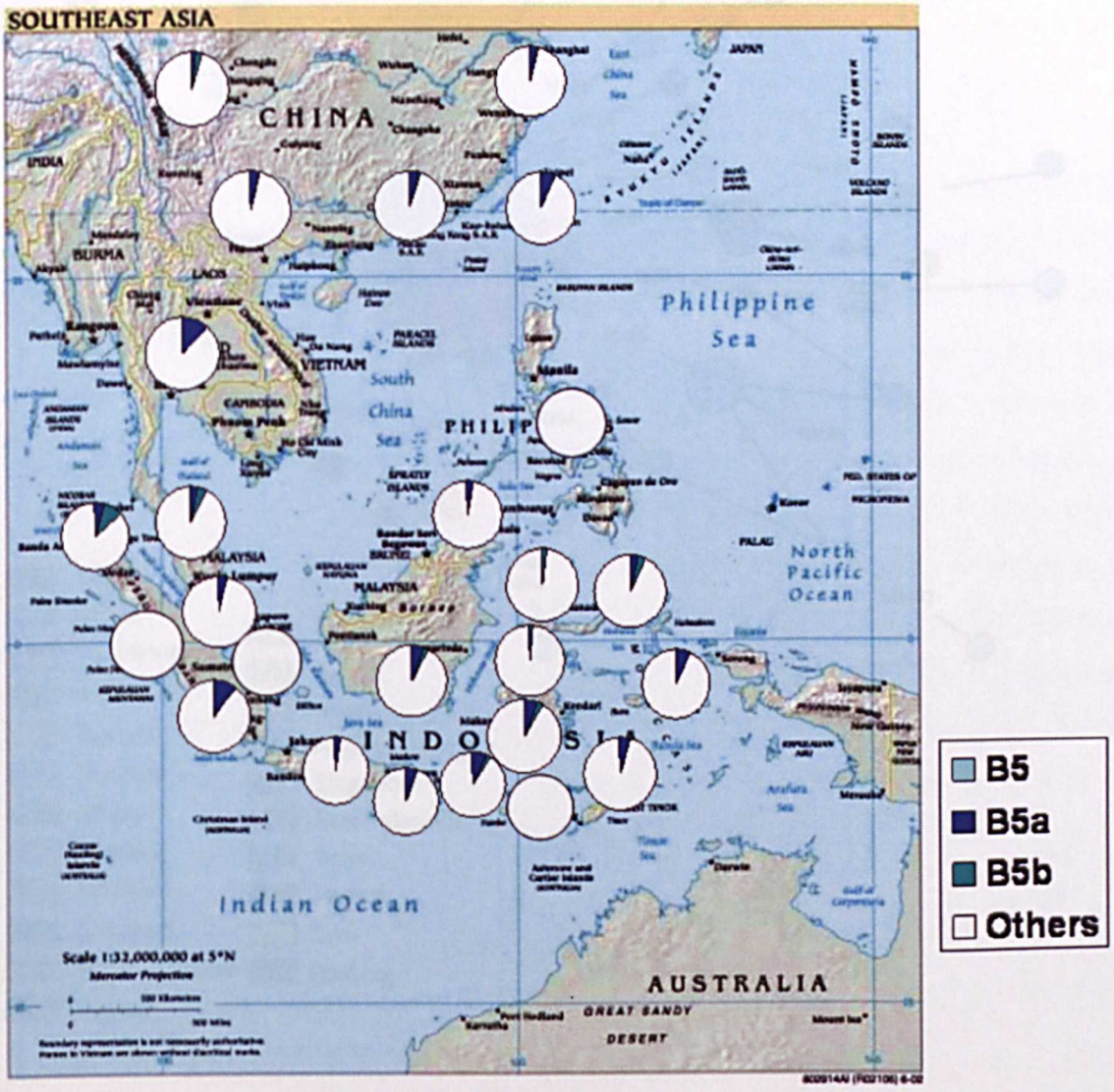
Figure 71 – Network of B4c types



A second branch major of B, **B5**, is defined by a transition at np 16140 in HVS-I and another at np 10398 in the coding region. As yet, only three putative B5* types have

been found, one in the Nicobar Islands, one in the Yami group of Taiwan and one in Java (unpublished data of Peter Forster, University of Cambridge). All others belong to one of two branches: **B5a**, defined by a C to A transversion at np 16266, and **B5b**, defined by a transition at np 16243. The distribution of B5 types is shown in figure 72.

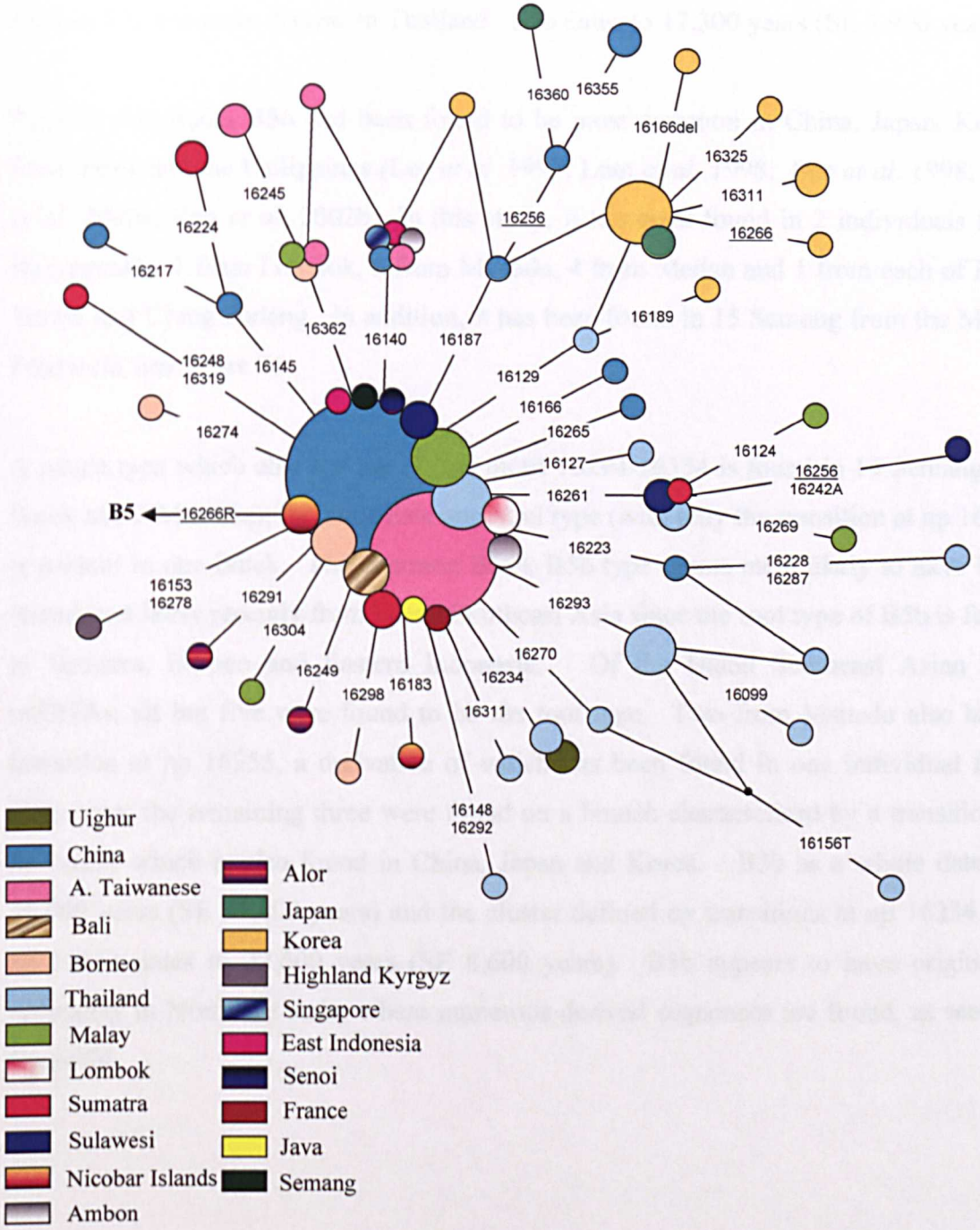
Figure 72 – Map showing the distribution of haplogroup B5 in Southeast Asia



Haplogroup frequency data can be found in Appendix IV

B5a is the more common of the two and is most prevalent in China, Taiwan and Thailand. It has also been found in Japan, Korea, the Nicobar Islands and the Malay Peninsula. A network of B5a types is shown in figure 73.

Figure 73 – Network of B5a types



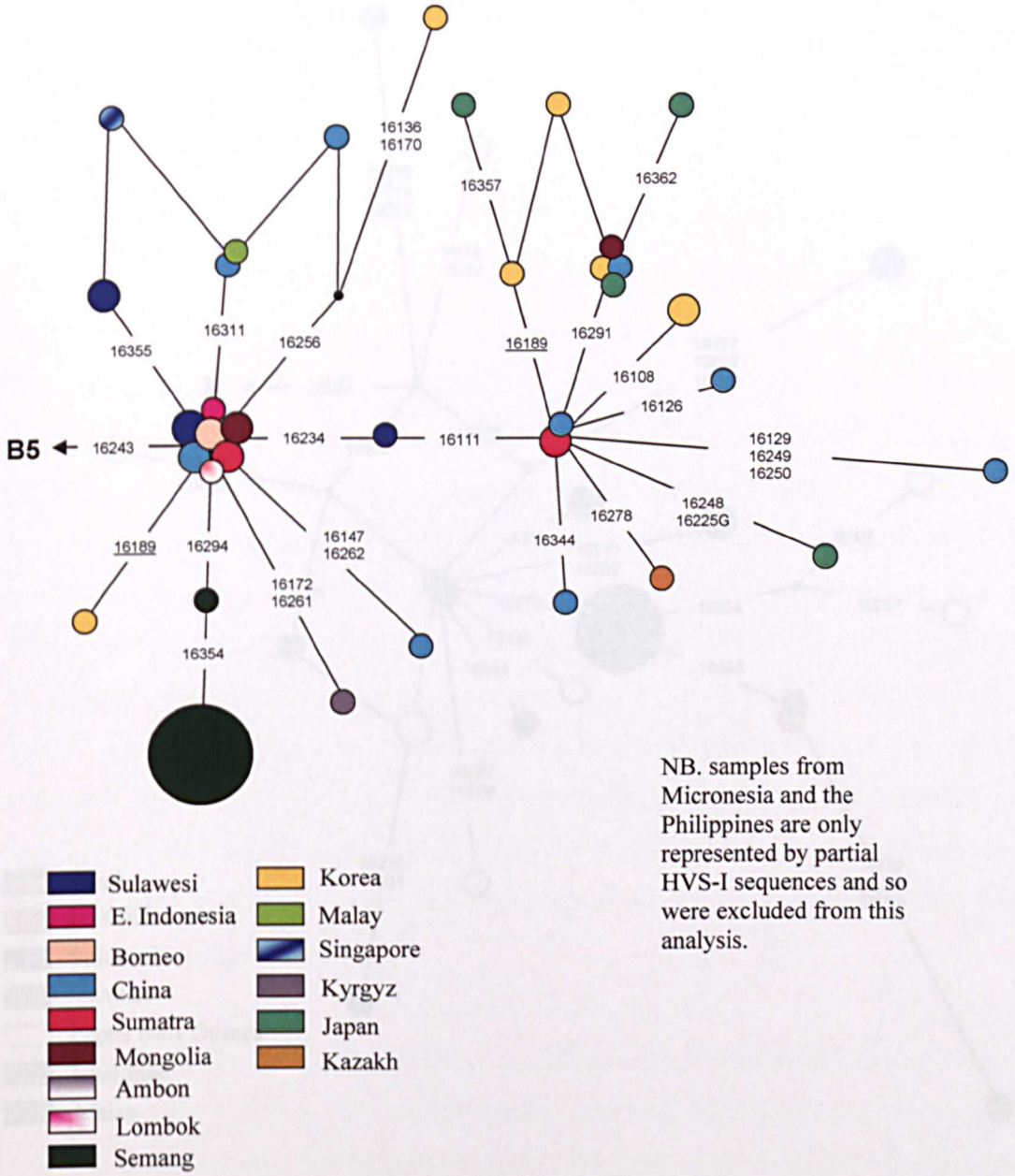
In this study, B5a has been found to be fairly common in Island Southeast Asia. It has been found in: 2 individuals from Alor, 3 from Ambon, 4 from Bali, 3 from Lombok, 6 from Banjarmasin, 2 from Kota Kinabalu, 1 from Java, 3 from Manado, 3 from Ujung Padang, 2 from Medan, 3 from Palembang and 2 from Pekanbaru. It has also been found in one Semang and one Senoi from the Malay Peninsula. The root of B5a is most

common in China and Taiwan, suggesting an origin there. However, as shown in section 5.3, it is more diverse in Thailand. B5a dates to 17,300 years (SE 3,900 years).

Prior to this study, **B5b** had been found to be most common in China, Japan, Korea, Micronesia and the Philippines (Lee *et al*.1997; Lum *et al.* 1998; Seo *et al.* 1998; Yao *et al.* 2002a; Yao *et al.* 2002b. In this study, it has been found in 2 individuals from Banjarmasin, 1 from Lombok, 3 from Manado, 4 from Medan and 1 from each of Palu, Toraja and Ujung Padang. In addition, it has been found in 15 Semang from the Malay Peninsula, see figure 74.

A single type which also has the HVS I motif 16294-16354 is found in 14 Semang (12 Batek and 2 Mendriq); its immediate ancestral type (with only the transition at np 16294) is present in one Batek. The common Batek B5b type seems most likely to have been introduced fairly recently from Island Southeast Asia since the root type of B5b is found in Sumatra, Borneo and Eastern Indonesia. Of the Island Southeast Asian B5b mtDNAs, all but five were found to be the root type. Two from Manado also had a transition at np 16355, a derivative of which has been found in one individual from Singapore; the remaining three were found on a branch characterised by a transition at np 16234 which is also found in China, Japan and Korea. B5b as a whole dates to 35,000 years (SE 11,800 years) and the cluster defined by transitions at np 16234 and np 16111 dates to 22,600 years (SE 8,600 years). B5b appears to have originated ultimately in Northeast Asia, where numerous derived sequences are found, as seen in figure 74.

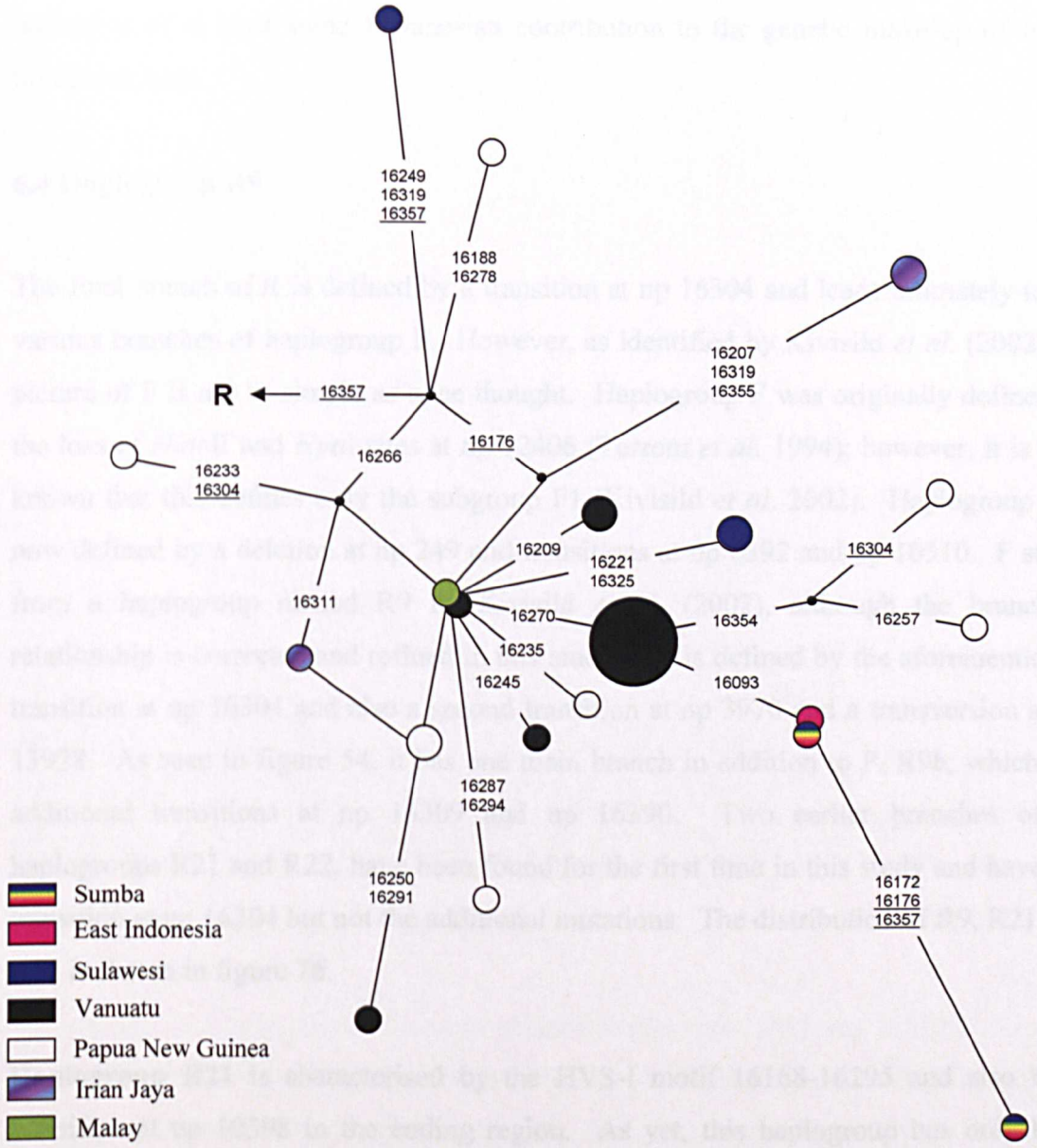
Figure 74 – Network of B5b types



NB. samples from Micronesia and the Philippines are only represented by partial HVS-I sequences and so were excluded from this analysis.

The second branch of R is characterised by a transition at np 16357 and is designated **haplogroup P**. It is a predominantly Melanesian and Micronesian haplogroup with more diversity being found in the former suggesting an origin there. The distribution of haplogroup P is shown in figure 75.

Figure 75 – Network of P types



Samples from Micronesia are only represented by partial HVS-I sequences and so were excluded from this analysis.

As seen in figure 75, haplogroup P was also found in one Indonesian individual by Redd *et al.* (1995) and has been found in a further five in this study. All but one of these (three of whom were from Manado and two from Sumba) belonged to a branch of P characterised by transitions at np 16176 and np 16266 which is relatively common in Papua New Guinea and Vanuatu but which is not found in Micronesia. This branch has

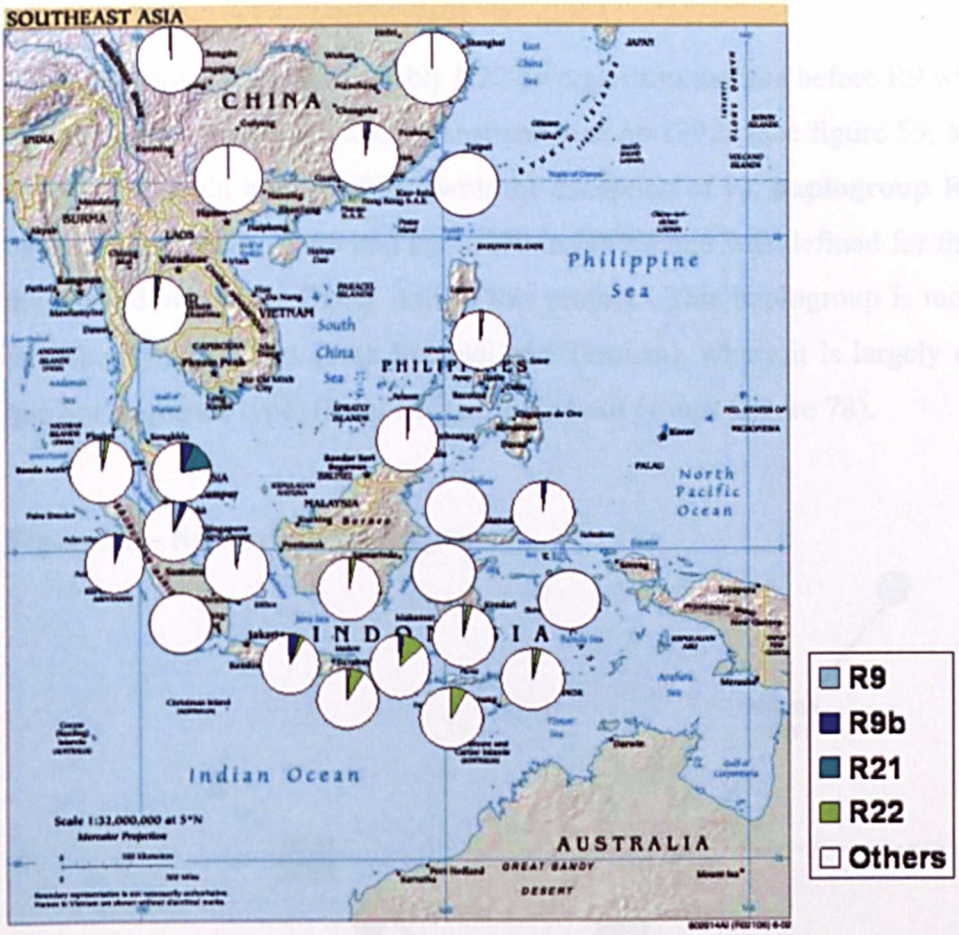
been dated to 30,300 years (SE 12,000 years). It seems almost certain that P is an indigenous Melanesian haplogroup so its presence in the sample studied here is indicative of at least some Melanesian contribution to the genetic make-up of Island Southeast Asia.

6.4 Haplogroup R9

The final branch of R is defined by a transition at np 16304 and leads ultimately to the various branches of haplogroup F. However, as identified by Kivisild *et al.* (2002) the picture of F is not as simple as once thought. Haplogroup F was originally defined by the loss of *Hinc*II and *Hpa*I sites at np 12406 (Torroni *et al.* 1994); however, it is now known that this defines only the subgroup F1 (Kivisild *et al.* 2002). Haplogroup F is now defined by a deletion at np 249 and transitions at np 6392 and np 10310. F stems from a haplogroup named R9 by Kivisild *et al.* (2002), although the branching relationship is corrected and refined in this study. R9 is defined by the aforementioned transition at np 16304 and also a second transition at np 3970 and a transversion at np 13928. As seen in figure 54, it has one main branch in addition to F, R9b, which has additional transitions at np 16309 and np 16390. Two earlier branches of R, haplogroups R21 and R22, have been found for the first time in this study and have the transition at np 16304 but not the additional mutations. The distribution of R9, R21 and R22 is shown in figure 76.

Haplogroup R21 is characterised by the HVS-I motif 16168-16295 and also by a transition at np 10398 in the coding region. As yet, this haplogroup has only been found in the Jahai and Mendriq Semang and the Temiar Senoi, most of whom share a single sequence type. Two instances of this haplogroup can also be found in the Malay data of Zainuddin and Goodwin (2004). There are no obvious neighbouring types, suggesting that this clade may be indigenous to the Semang/Senoi, and may represent a component of deep Pleistocene ancestry within the Malay Peninsula.

Figure 76 – Map showing the distribution of haplogroups R9, R21 and R22 in Southeast Asia



Haplogroup frequency data can be found in Appendix IV

The second new haplogroup has been termed **haplogroup R22** and is defined by the HVS-I motif 16249-16288. Some types also have a transition at np 16390, suggesting that they may be related to R9b types, but the branching relationship is unclear at present. R22 has been found in 5 individuals from Bali, 5 from Lombok, 4 from Sumba, 2 from Banjarmasin and 1 each from Alor, Medan, Ujung Padang, Kota Kinabalu and Java. R22 has also been found in three individuals from the Nicobar Islands (Prasad *et al.* 2001) and in two individuals from Thailand (Yao *et al.* 2002b). The branching relationship of R22 types is shown in figure 77.

The origins of R22 are unclear. The most derived types are seen in Thailand and the Nicobars; however, it is far more common in Bali and Nusa Tenggara and the root type

is only found in Lombok and Alor. This suggests that it could be a possible indigenous marker for that area. R22 dates to 12,500 years (SE 5,200 years).

As stated above, R21 and possibly R22 diverge from the tree before R9 which is defined by a transition at np 3970 and a transversion at np 13928 (see figure 55; Macaulay *et al.* 2005). The main branch of R9 (with the exception of F), **haplogroup R9b**, is defined by transitions at np 16309 and np 16390 in HVS-I and was defined for the first time by the work done on the Orang Asli in this project. This haplogroup is most common in the Aboriginal Malays (both Semelai and Temuan), where it is largely represented by just one sequence type, found frequently in both groups (figure 78).

Figure 77 – Network of R22 types

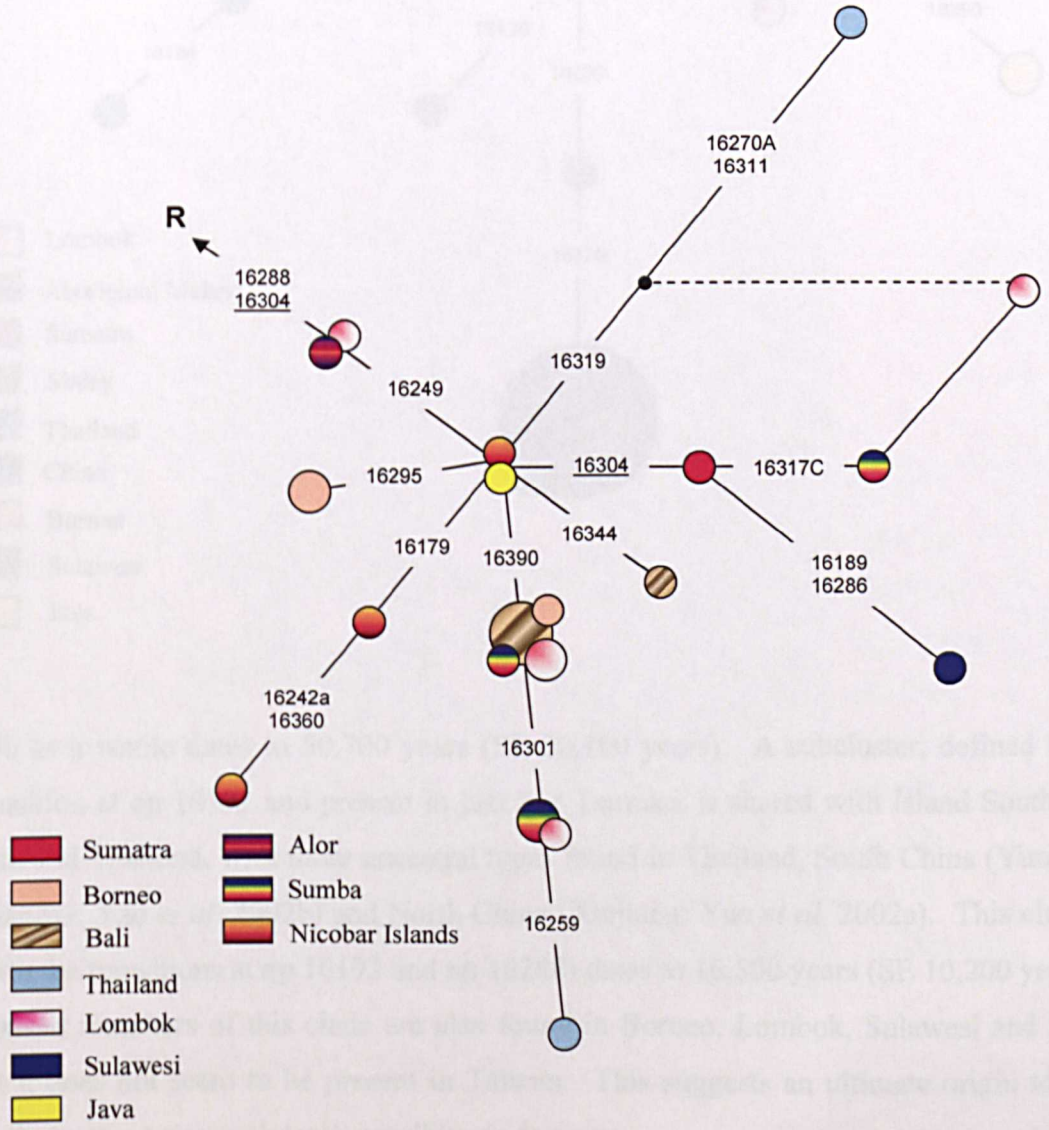
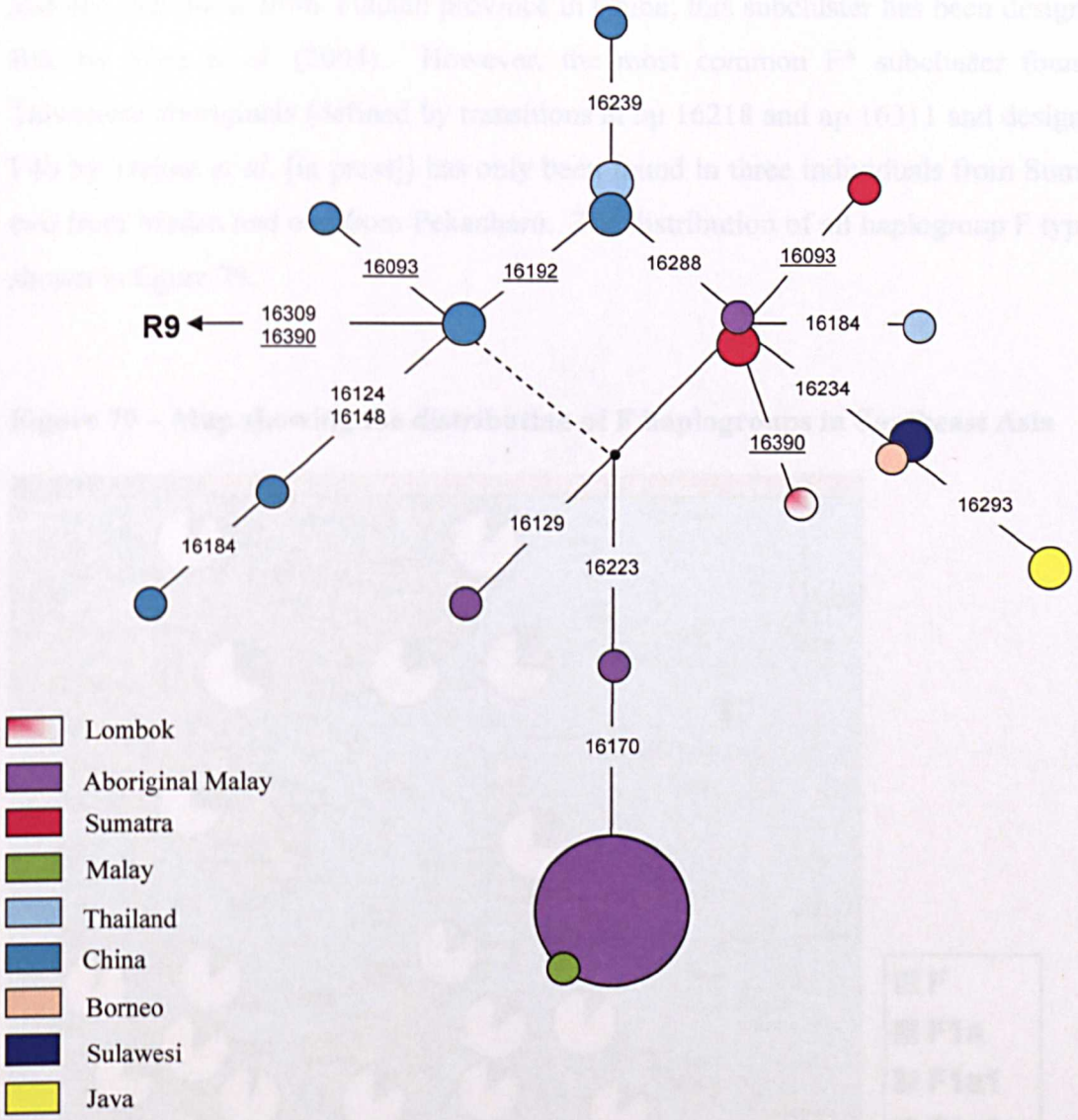


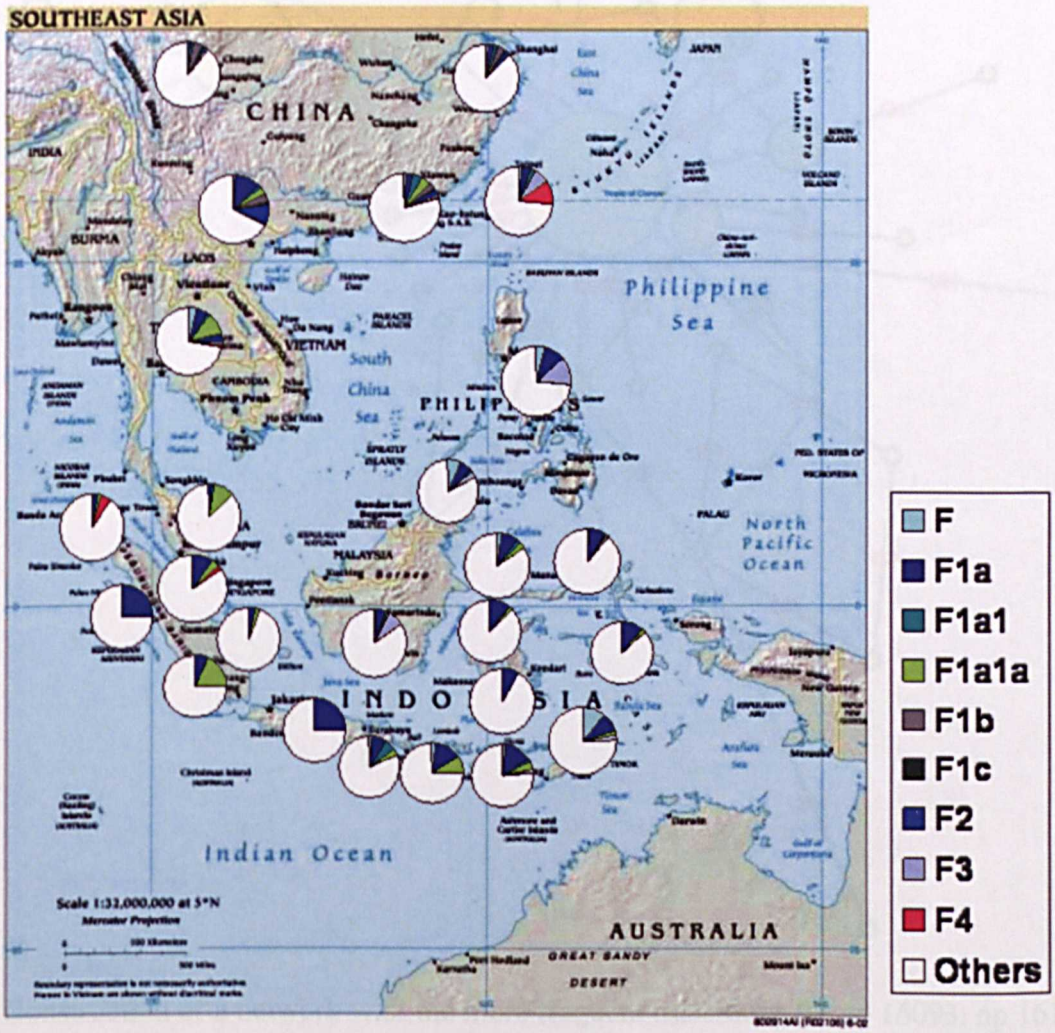
Figure 78 – Network of R9b types



R9b as a whole dates to 50,700 years (SE 20,100 years). A subcluster, defined by a transition at np 16192 and present in just one Temuan, is shared with Island Southeast Asia and Thailand, with more ancestral types found in Thailand, South China (Yunnan, Guangxi: Yao *et al.* 2002b) and North China (Xinjiang: Yao *et al.* 2002a). This cluster (with the transitions at np 16192 and np 16288) dates to 16,500 years (SE 10,200 years). Derived members of this clade are also found in Borneo, Lombok, Sulawesi and Java but it does not seem to be present in Taiwan. This suggests an ultimate origin to the north, on the Asian mainland, possibly via Sumatra.

The main **F*** subcluster found in Island Southeast Asia is defined by the addition of transitions at np 16157 and np 16256 and is found in Alor, Kota Kinabalu, Manado, Palu, Medan and Sumba. It has also been found in the Philippines, the Tsou of Taiwan and one individual from Yunnan province in China; this subcluster has been designated R9c by Wen *et al.* (2004). However, the most common F* subcluster found in Taiwanese aboriginals (defined by transitions at np 16218 and np 16311 and designated F4b by Trejaut *et al.* [in press]) has only been found in three individuals from Sumatra, two from Medan and one from Pekanbaru. The distribution of all haplogroup F types is shown in figure 79.

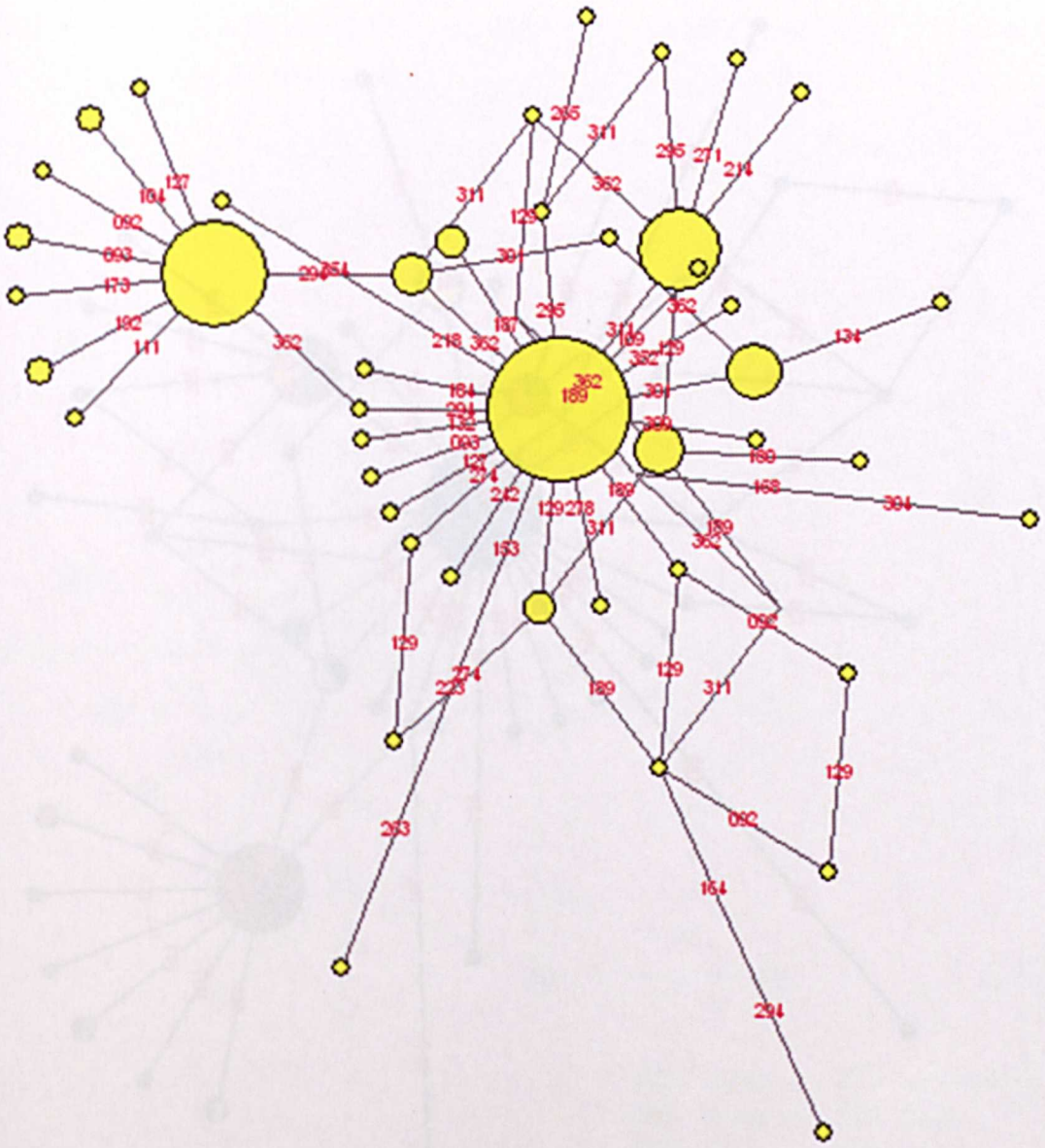
Figure 79 – Map showing the distribution of F haplogroups in Southeast Asia



Haplogroup frequency data can be found in Appendix IV

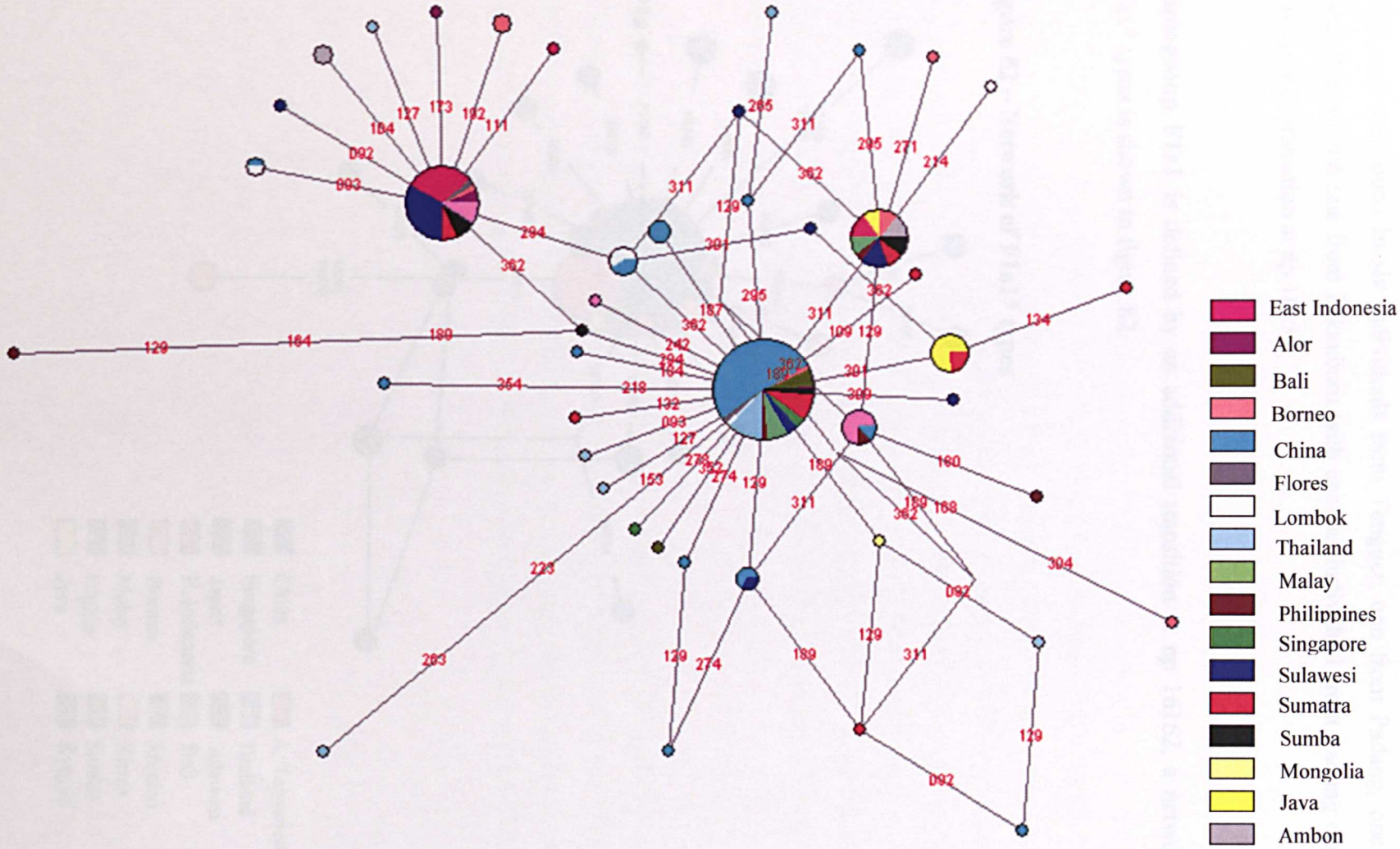
As stated above, **haplogroup F1** is defined by transitions at np 10609 and np 12406, **haplogroup F1a** is further defined by transitions at np 16129 and np 16172. The network of F1a* types is very reticulated and is shown in figure 80.

Figure 80 – Unweighted network of F1a* types



Construction of a network with the more frequent transitions (at np 16093, np 16129, np 16189, np 16311 and np 16362) downweighted (figure 81) shows that the root type of F1a is most common in China (mainly Yunnan province).

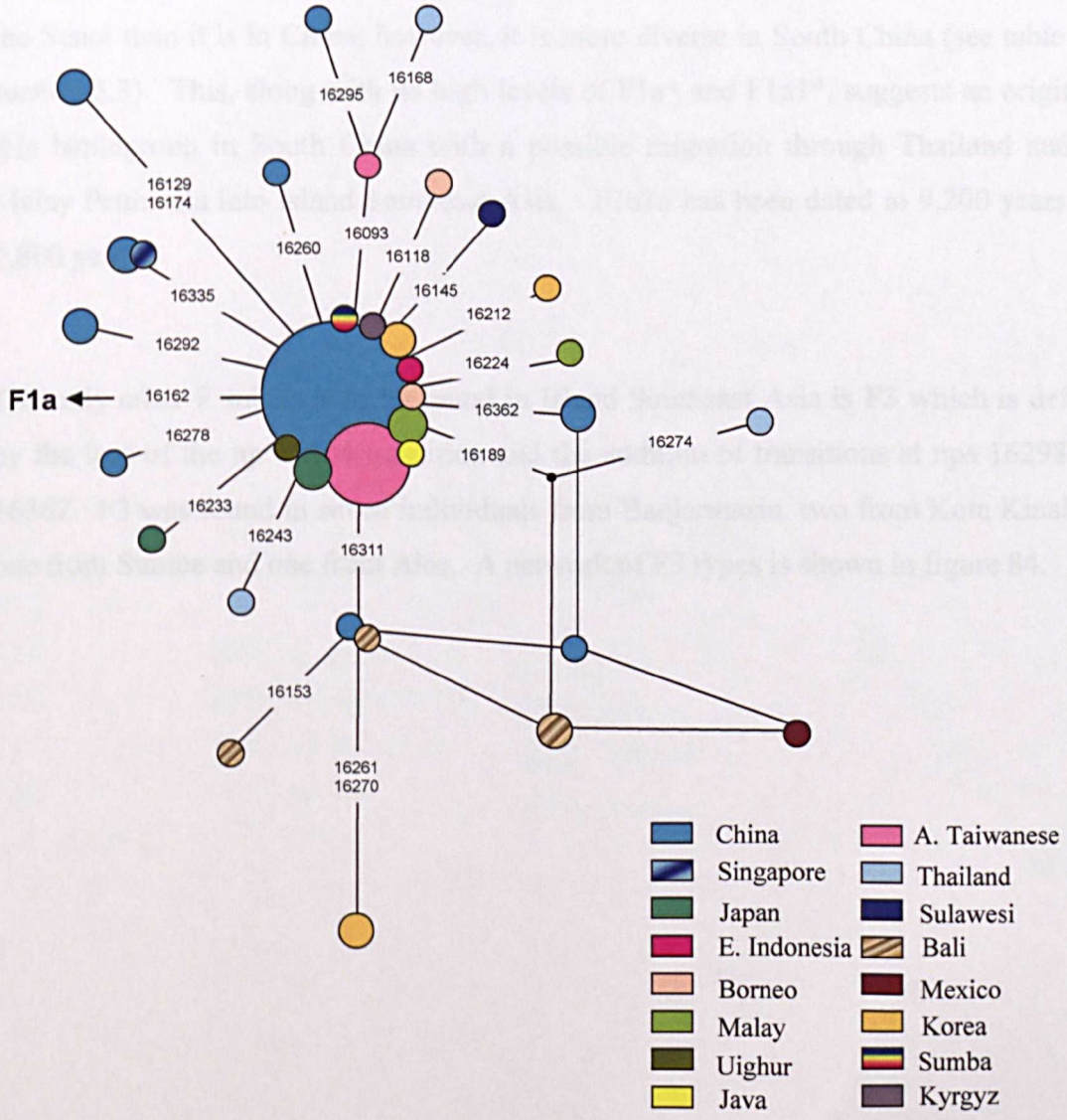
Figure 81 – Weighted network of F1a* types



As seen in figure 81, most Island Southeast Asian types are found on a branch characterised by a transition at np 16294 which dates to 4,300 years (SE 1,900 years). A second branch specific to Island Southeast Asia is characterised by a transition at np 16301 and was found in six individuals from Tengger, one from Padang, one from Ujung Padang and one from Pekanbaru with another individual from Padang carrying an additional transition at np 16134.

Haplogroup F1a1 is defined by an additional transition at np 16162, a network of F1a1* types is shown in figure 82.

Figure 82 – Network of F1a1* types



As seen in figure 82, F1a1* is also most common in South China; the root type is also relatively common amongst Taiwanese Aboriginals. F1a1* is less common in Island Southeast Asia than F1a*, the root type is only found in four individuals (including one from the data of Redd *et al.* [1995] and one from the unpublished data of Peter Forster [personal communication]). Derivatives are, however, found in Borneo, Sulawesi and Bali. F1a1 dates to 26,000 years (SE 13,700 years).

Haplogroup F1a1a is further defined by a transition at np 16108, a network of F1a1a types is shown in figure 83. The root type of F1a1a is most common in Thailand, China and the Senoi; it is also relatively common across Island Southeast Asia, particularly in Sumatra. Derivative types are also found in Sumatra, Bali, Sumba, Borneo and the Orang Asli (both Senoi and Aboriginal Malay). F1a1a is more common in Thailand and the Senoi than it is in China; however, it is more diverse in South China (see table 8 in section 5.3). This, along with its high levels of F1a* and F1a1*, suggests an origin for this haplogroup in South China with a possible migration through Thailand and the Malay Peninsula into Island Southeast Asia. F1a1a has been dated to 9,200 years (SE 2,800 years).

The only other F subclade to be found in Island Southeast Asia is **F3** which is defined by the loss of the np 16304 transition and the addition of transitions at nps 16298 and 16362. F3 was found in seven individuals from Banjarmasin, two from Kota Kinabalu, one from Sumba and one from Alor. A network of F3 types is shown in figure 84.

Figure 83 – Network of F1a1a types

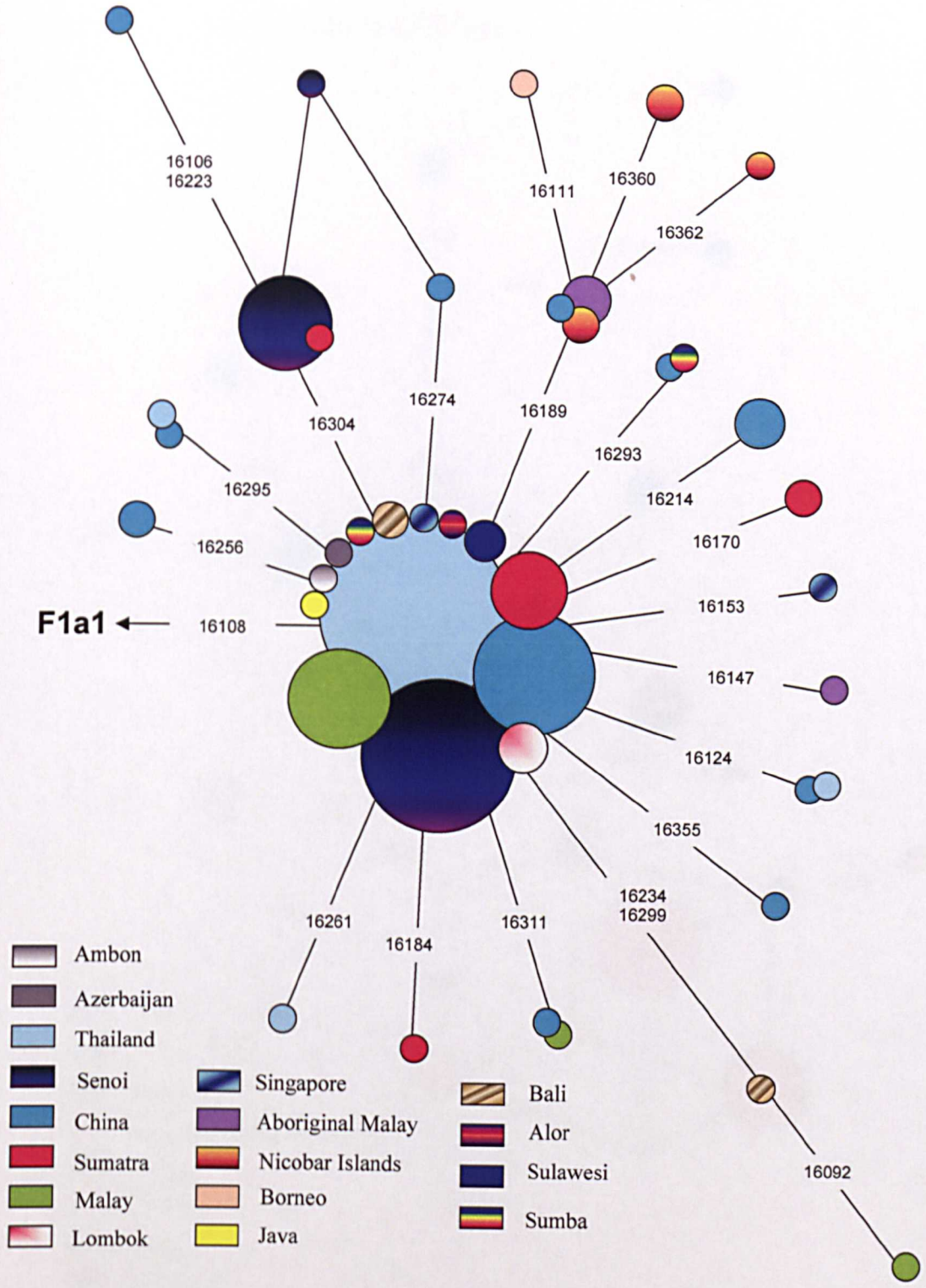
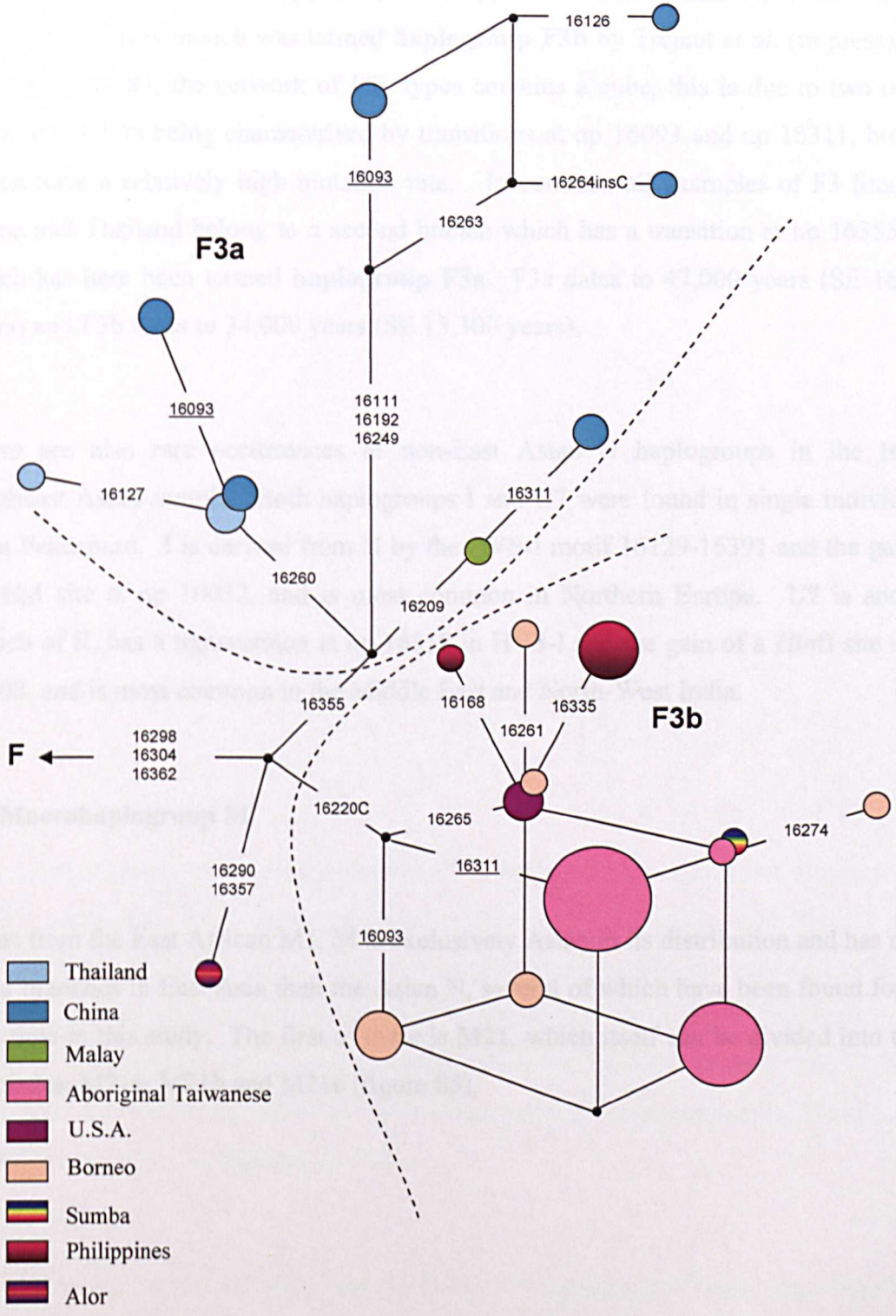


Figure 84 – Network of F3a and F3b types



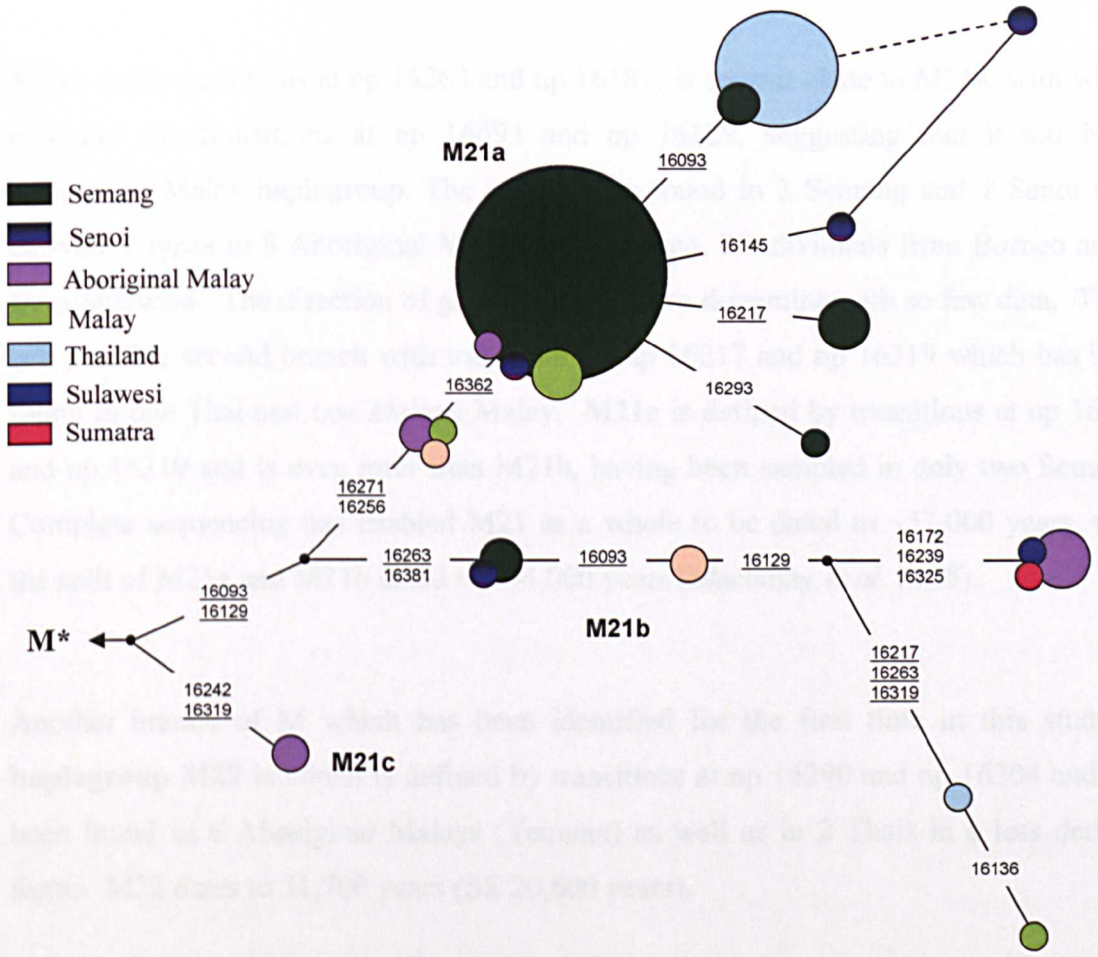
All but one of the sequence types from Island Southeast Asia were found to have a transversion from A to C at np 16220 which has only been found elsewhere in Taiwan (Tajima *et al.* 2003), the Philippines (this study) and in three African Americans (Handt *et al.* 1998). This branch was termed **haplogroup F3b** by Trejaut *et al.* (in press). As seen in figure 84, the network of F3b types contains a cube, this is due to two of the branches of F3b being characterised by transitions at np 16093 and np 16311, both of which have a relatively high mutation rate. In contrast, all examples of F3 found in China and Thailand belong to a second branch which has a transition at np 16355 and which has here been termed **haplogroup F3a**. F3a dates to 47,000 years (SE 16,000 years) and F3b dates to 34,000 years (SE 13,300 years).

There are also rare occurrences of non-East Asian N haplogroups in the Island Southeast Asian sample. Both haplogroups I and U7 were found in single individuals from Pekanbaru. **I** is derived from N by the HVS-I motif 16129-16391 and the gain of an *Alu*I site at np 10032, and is most common in Northern Europe. **U7** is another branch of R, has a transversion at np 16318 in HVS-I and the gain of a *Hin*fI site at np 12308, and is most common in the Middle East and North-West India.

6.5 Macrohaplogroup M

Apart from the East African M1, M is exclusively Asian in its distribution and has more basal branches in East Asia than the Asian N, several of which have been found for the first time in this study. The first of these is **M21**, which itself can be divided into three subclades: M21a, M21b and M21c (figure 85).

Figure 85 – Network of M21 types



The most common of these subclades, **M21a**, is defined by transitions at nps 16093, 16129, 16256, and 16271 and is common in the Orang Asli (reaching its highest levels at 84% in the Mendriq), the only other non-Malaysian people known to possess it at high frequencies are the Sakai Semang of southern Thailand (data of Fucharoen *et al.* 2001) which suggests that it is an indigenous Semang haplogroup; note that the M21a types found in the Fucharoen *et al.* (2001) data were mischaracterised by Tanaka *et al.* (2004) as belonging to haplogroup D4a.

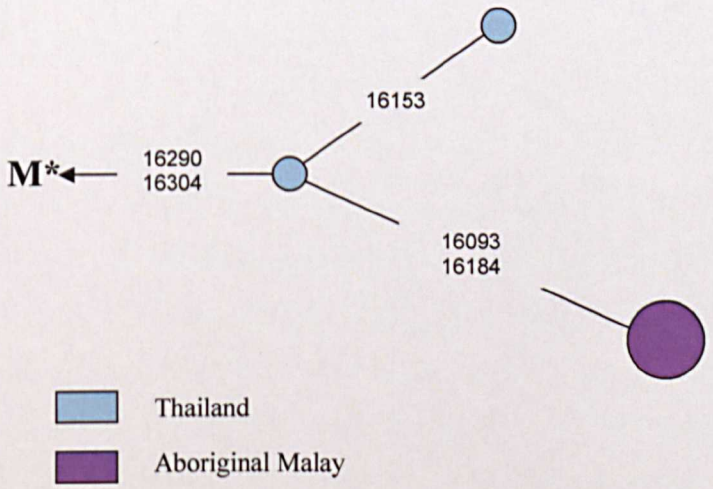
The root type of M21a has only been found in two Aboriginal Malays, one Malay (data of Zainuddin and Goodwin 2004) and one individual from Banjarmasin, most likely suggesting gene flow from the Semang or Senoi into these groups. The most common type (which also has a transition at np 16362) was found in 38 Semang, 4 Malays (data

of Zainuddin and Goodwin 2004), 1 Senoi and 1 Aboriginal Malay. Derivatives of the most prevalent type are found in 16 Thais, 8 Semang and 2 Senoi.

M21b (with transitions at np 16263 and np 16381) is a sister clade to M21a, with which it shares the transitions at np 16093 and np 16129, suggesting that it too is an indigenous Malay haplogroup. The root type is found in 3 Semang and 1 Senoi with derivative types in 5 Aboriginal Malays, 1 Sumatran, 2 individuals from Borneo and 1 from Sulawesi. The direction of gene flow is hard to determine with so few data. There is a possible second branch with transitions at np 16217 and np 16319 which has been found in one Thai and one *Melayu* Malay. **M21c** is defined by transitions at np 16242 and np 16319 and is even rarer than M21b, having been sampled in only two Semelai. Complete sequencing has enabled M21 as a whole to be dated to ~57,000 years, with the split of M21a and M21b dated to ~44,000 years (Macaulay *et al.* 2005).

Another branch of M which has been identified for the first time in this study is **haplogroup M22** is which is defined by transitions at np 16290 and np 16304 and has been found in 6 Aboriginal Malays (Temuan) as well as in 2 Thais in a less derived form. M22 dates to 31,700 years (SE 20,600 years).

Figure 86 – Network of M22 types



One of the most common Asian haplogroups is **M7** which is found throughout China, Korea, Japan, Island Southeast Asia and Micronesia. M7 is characterised by transitions at np 6455 and np 9824, the latter being recognised by the gain of a *Hin*fI site. **M7*** types are relatively uncommon, having only been found in China and Thailand prior to this study. In this study they have been found in one individual from Bali, 1 from Padang, 3 from Medan, 1 from Pekanbaru, 3 from Palu, 1 from Ambon, 1 from Kota Kinabalu and 1 from Waingapu. A network of M7* types is shown in figure 87.

As seen in figure 87 all but one of the Island Southeast Asian M7* samples are found on one branch which is characterised by a transition at np 16362. The only other sample found in this cluster is one from Thailand (Fucharoen *et al.* 2001). This cluster dates to 16,800 years (SE 6,300 years). All M7* types found in China are found on another two main branches along with another sample from Thailand and the one found in Kota Kinabalu.

M7 can be divided into three further branches: M7a, M7b and M7c (Kivisild *et al.* 2002). **M7a** is a Northeast Asian haplogroup which is defined by a transition at np 16209 and has been found exclusively in Japan and Korea with only one exception in the Philippines. However, M7b and M7c have much more widespread distributions (see figure 88).

Figure 87 – Network of M7* types

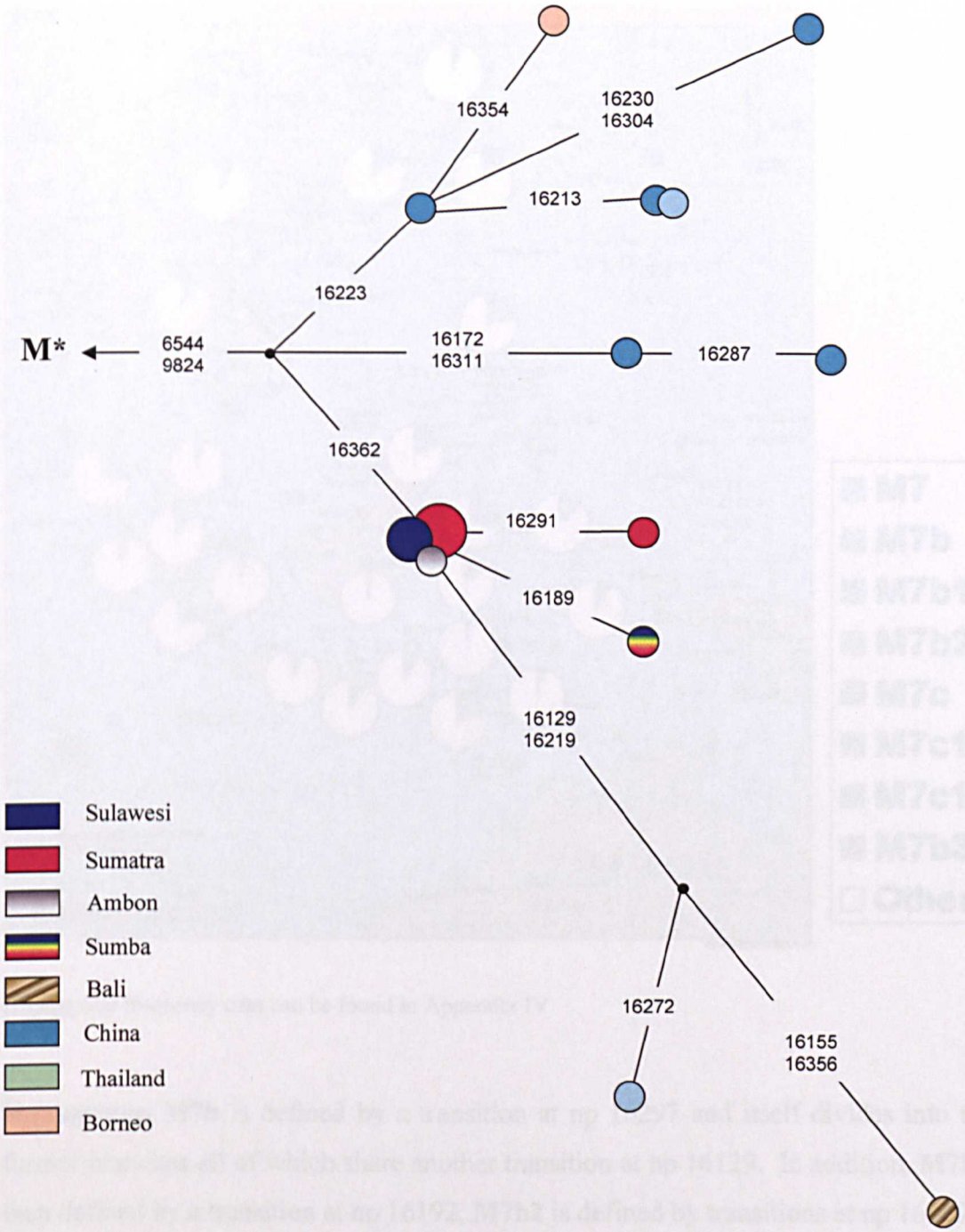
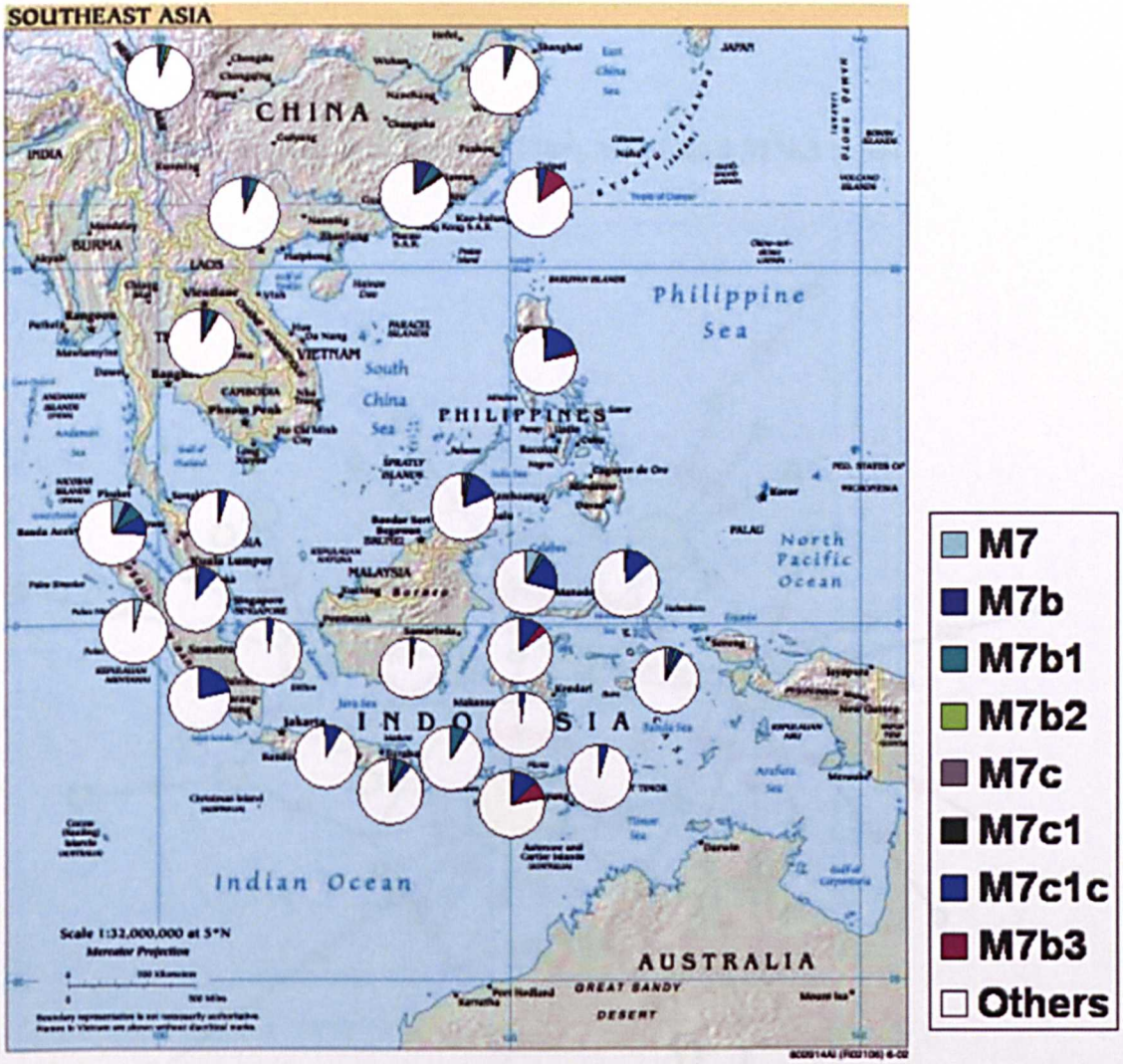


Figure 88 – Map showing the distribution of M7 haplogroups in Southeast Asia

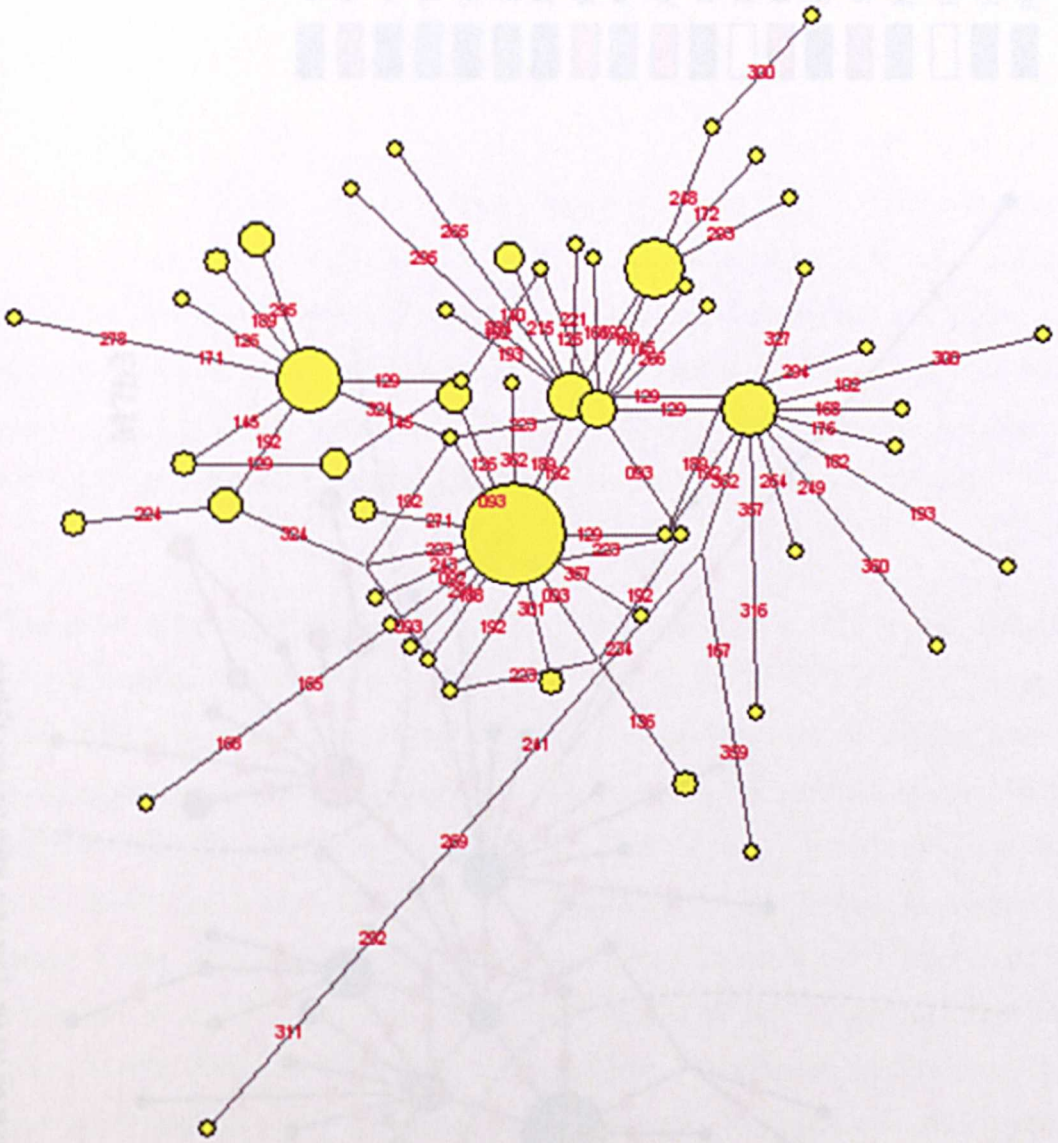


Haplogroup frequency data can be found in Appendix IV

Haplogroup M7b is defined by a transition at np 16297 and itself divides into three further branches all of which share another transition at np 16129. In addition, **M7b1** is then defined by a transition at np 16192, **M7b2** is defined by transitions at np 16189 and np 16298 (Kivisild *et al.* 2002) and **M7b3** is defined by the loss of the transition at np 16223 and the addition of a transition at np 16324 (Yao *et al.* 2004). M7b2 is, like M7a, an almost exclusively Northeast Asian haplogroup with almost all examples being found in Japan and Korea (Horai *et al.* 1996; Lee *et al.* 1997; Lum *et al.* 1998; Seo *et al.* 1998; Nishimaki *et al.* 1999), the only exceptions being five individuals from across China (Nishimaki *et al.* 1999; Yao *et al.* 2002a) . However, M7b*, M7b1 and M7b3 are

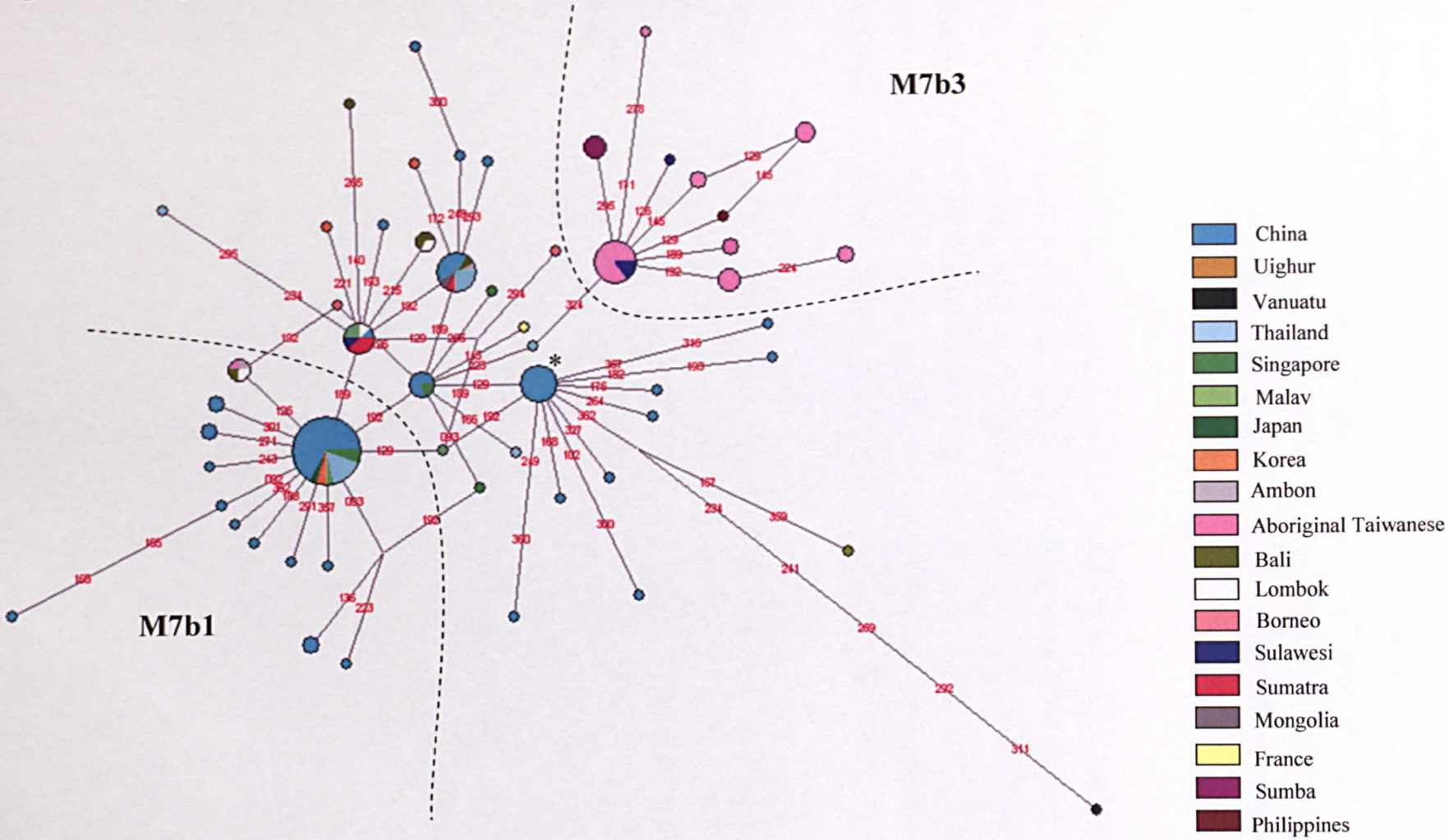
more widespread. The network of these M7b types is highly reticulated, as seen in figure 89.

Figure 89 – Unweighted network of M7b*, M7b1 and M7b3 types



The network is still fairly reticulated even with np 16093, np 16129, np 16189, np 16223, np 16311 and np 16362 downweighted, see figure 90.

Figure 90 – Weighted network of M7b*, M7b1 and M7b3 types



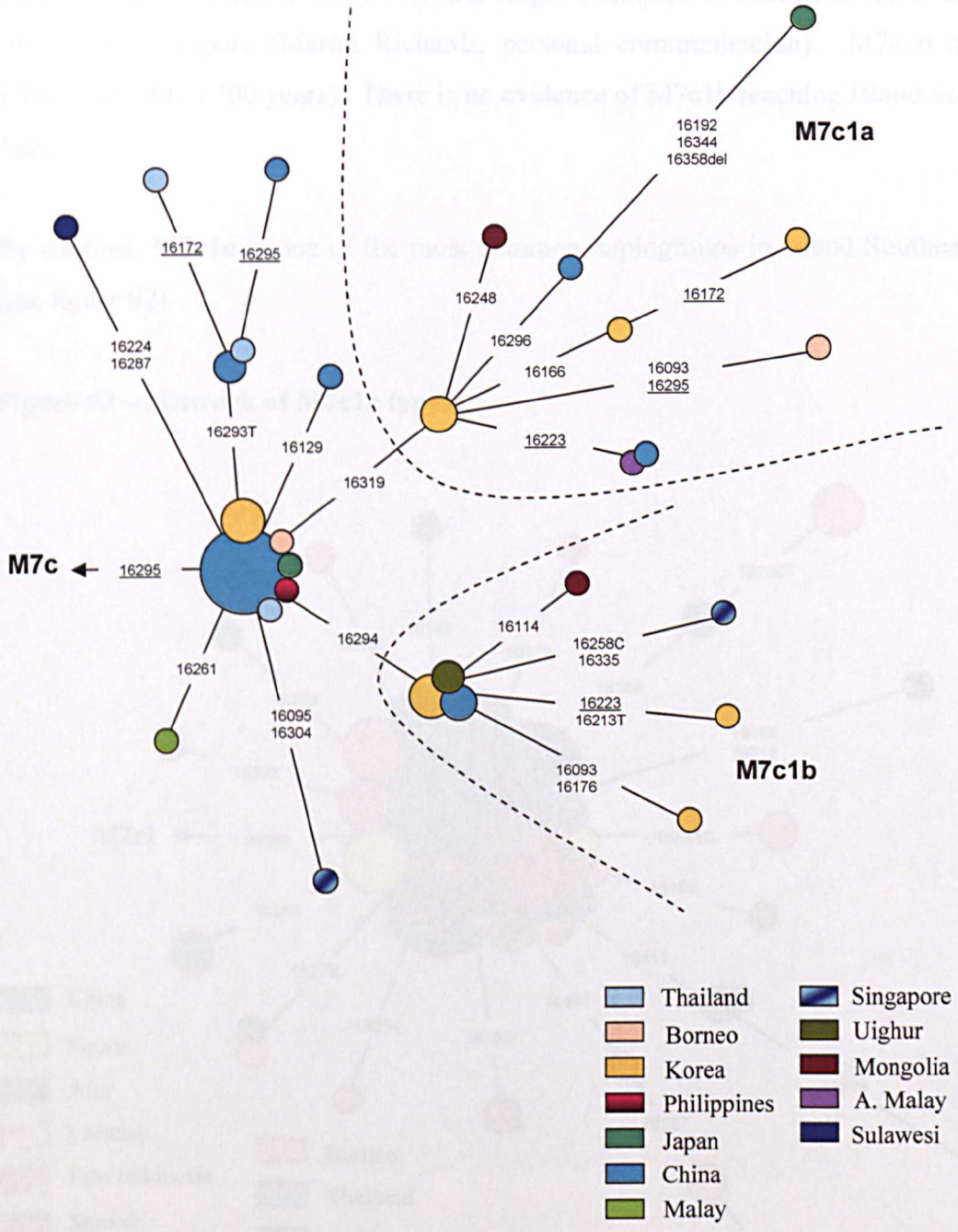
* denotes the root type of M7b

M7b is most common in China (Kivisild *et al.* 2002; Yao *et al.* 2002a; Yao *et al.* 2002b), suggesting a root there. It is also relatively common in Taiwanese aboriginals and has been found in Singapore, the Philippines, Thailand and in one individual from Vanuatu. This study has shown M7b to be found at low levels across Island Southeast Asia; it is most common in Sumba, where the same derived M7b3 type was found in four individuals. M7b3 is mainly found elsewhere in Taiwanese aboriginals but has also been found in three individuals from Toraja and one from the Philippines. M7b3 dates to 15,400 years (SE 5,900 years).

M7b* and M7b1 types are also found in Island Southeast Asia. M7b* types are found in single individuals from Ambon, Bali, Banjarmasin, Kota Kinabalu and Medan. M7b1 reaches its highest frequency in Island Southeast Asia in Bali and Medan where it is found in three individuals. However, the root type is not found in Island Southeast Asia. M7b1 is found in China, Thailand, the *Melayu* Malays, Korea and Japan. The root of M7b1 is most common in southern China suggesting an origin there.

The most widespread subclade of M7 is **haplogroup M7c**, and in particular **M7c1**. M7c is characterized by transitions at np 146 and np 199 in HVS-II and M7c1 is further defined by a transition at np 16295 in HVS-I. M7c1 then is further divided into **M7c1a** which also has a transition at np 16319, **M7c1b** which has a transition at np 16294 and **M7c1c** which has a transition at np 16362. M7c1* is most common in China (Kivisild *et al.* 2002; Yao *et al.* 2002a; Yao *et al.* 2002b) but is also found, at lower levels, in Japan, Korea, the Philippines, Singapore, the Malay Peninsula and Thailand (Horai *et al.* 1996; Lee *et al.* 1997; Seo *et al.* 1998; Fucharoen *et al.* 2001; Zainuddin and Goodwin 2004; Martin Richards, personal communication). As shown in figure 91, the only derivative example of M7c1* found in this study was from Manado, this type has not been found elsewhere. However, the root type of M7c1 has previously been found in one individual from Kota Kinabalu (Sykes *et al.* 1995). M7c1 dates to 25,700 years (SE 14,500 years).

Figure 91 – Network of M7c1*, M7c1a and M7c1b types

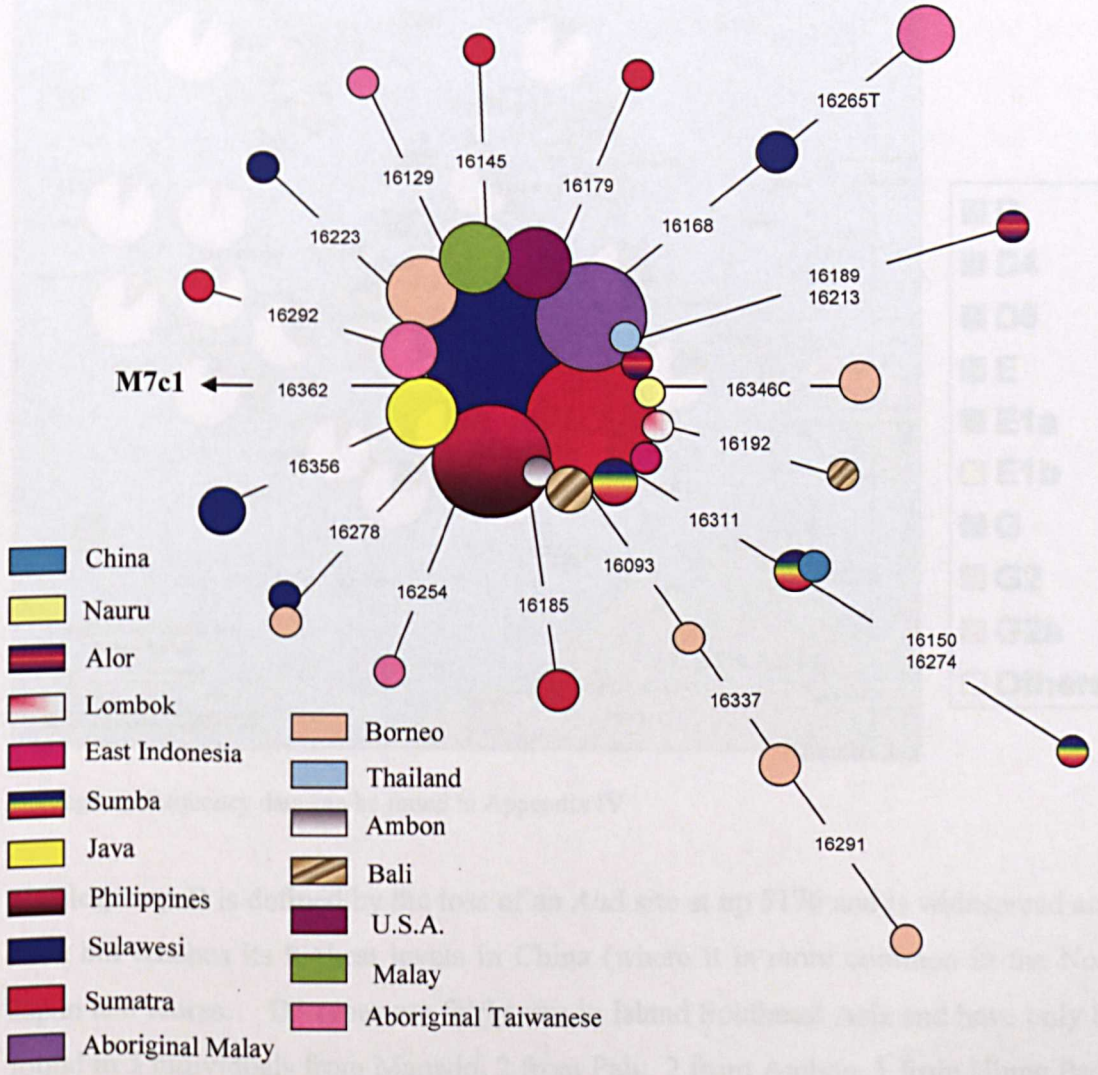


As shown in figure 91, **M7c1a** is quite rare and reaches its highest levels in Korea (Horai *et al.* 1996; Lee *et al.* 1997; Pfeiffer *et al.* 1998). In this study, only two derived examples of M7c1a have been found – one in Borneo which is not found elsewhere and one in an Aboriginal Malay (Semelai) which is also found in an individual from Wuhan

in China (Yao *et al.* 2002a). M7c1a dates to 24,500 years (SE 9,700 years). **M7c1b** is even rarer and is found mainly in Korea (Lee *et al.* 1997) with two examples in the Uighur people (Comas *et al.* 1998) and single examples in Mongolia (Kolman *et al.* 1996) and Singapore (Martin Richards, personal communication). M7c1b dates to 5,500 years (SE 3,200 years). There is no evidence of M7c1b reaching Island Southeast Asia.

By contrast, **M7c1c** is one of the most common haplogroups in Island Southeast Asia (see figure 92).

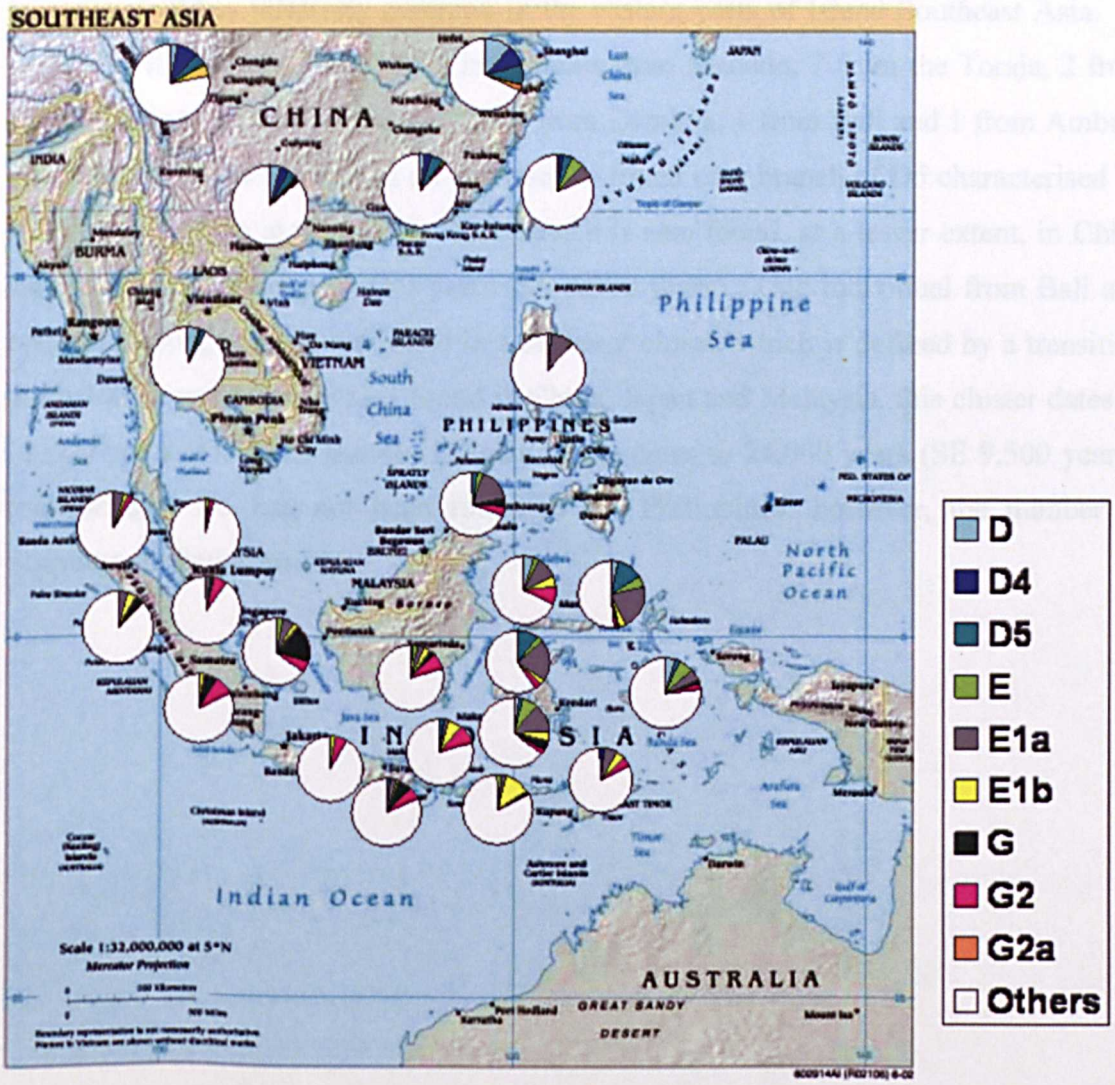
Figure 92 – Network of M7c1c types



Samples from Micronesia are only represented by partial HVS-I sequences and so were excluded from this analysis.

The other major M clades found in Island Southeast Asia are **haplogroups D, E and G**. The distribution of these haplogroups is shown in figure 93.

Figure 93 – Map showing the distribution of haplogroups D, E and G in Southeast Asia



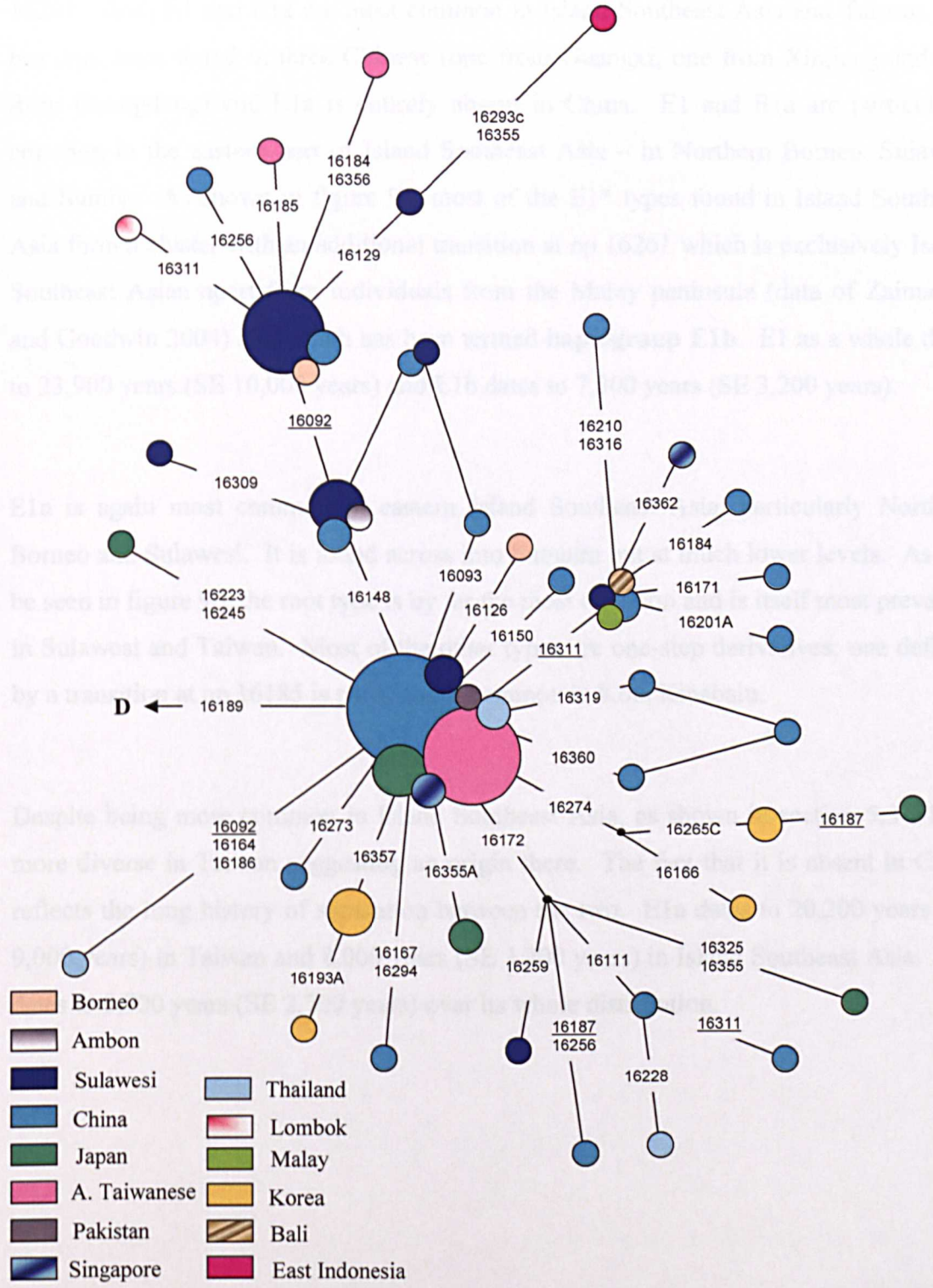
Haplogroup frequency data can be found in Appendix IV

Haplogroup D is defined by the loss of an *Alu*I site at np 5176 and is widespread across Asia but reaches its highest levels in China (where it is more common in the North), Japan and Korea. D* types are fairly rare in Island Southeast Asia and have only been found in 3 individuals from Manado, 2 from Palu, 2 from Ambon, 1 from Ujung Padang and 1 from Banjarmasin. There are two main branches of D in East Asia – D4 and D5.

D4 is characterised by three transitions in the coding region (at nps 3010, 8414 and 14668) and splits into **D4a**, characterised by a further transition at np 16129, and **D4b**, which also has a transition at np 16319. Only one derived D4a type has been found in this study, which belonged to one individual from Alor and which has also been found in one individual from the indigenous Saisat group of Taiwan.

In contrast, **D5** is relatively common in the eastern parts of Island Southeast Asia. In this study it has been found in 11 individuals from Manado, 7 from the Toraja, 2 from Ujung Padang, 2 from Kota Kinabalu, 1 from Lombok, 1 from Bali and 1 from Ambon. As shown in figure 94, almost all of these are found on a branch of D5 characterised by a further transition at np 16148. This branch is also found, at a lesser extent, in China and Taiwan and dates to 8,000 years (SE 3,300 years). One individual from Bali and one from Ujung Padang are found in a different cluster which is defined by a transition at np 16311 and which is also found in China, Japan and Malaysia, this cluster dates to 12,600 years (SE 5,600 years). D5 as a whole dates to 24,000 years (SE 9,500 years). Interestingly, D5 has not been found in the Philippines; however, the number of Filipinos studied is smaller.

Figure 94– Network of D5 types



Haplogroup E is a branch of haplogroup M9 which is defined by a transition at np 4491; E is further characterised by another transition at np 7598 which leads to the loss

of a *Hha*I site. In addition, **E1** has a transition at np 16390 and **E1a** has another at np 16291. Both E1 and E1a are most common in Island Southeast Asia and Taiwan. E1 has only been found in three Chinese (one from Guangxi, one from Xinjiang and one from Guangdong) and E1a is entirely absent in China. E1 and E1a are particularly common in the eastern part of Island Southeast Asia – in Northern Borneo, Sulawesi and Sumba. As shown in figure 95, most of the E1* types found in Island Southeast Asia form a cluster with an additional transition at np 16261 which is exclusively Island Southeast Asian apart from individuals from the Malay peninsula (data of Zainuddin and Goodwin 2004) and which has been termed **haplogroup E1b**. E1 as a whole dates to 23,900 years (SE 10,000 years) and E1b dates to 7,300 years (SE 3,200 years).

E1a is again most common in eastern Island Southeast Asia, particularly Northern Borneo and Sulawesi. It is found across into Sumatra but at much lower levels. As can be seen in figure 96, the root type is by far the most common and is itself most prevalent in Sulawesi and Taiwan. Most of the other types are one-step derivatives; one defined by a transition at np 16185 is particularly common in Kota Kinabalu.

Despite being more common in Island Southeast Asia, as shown in section 5.3 E1a is more diverse in Taiwan suggesting an origin there. The fact that it is absent in China reflects the long history of separation between the two. E1a dates to 20,200 years (SE 9,000 years) in Taiwan and 8,000 years (SE 3,200 years) in Island Southeast Asia. E1a dates to 8,700 years (SE 2,700 years) over its whole distribution.

Figure 95 – Network of E1* and E1b types

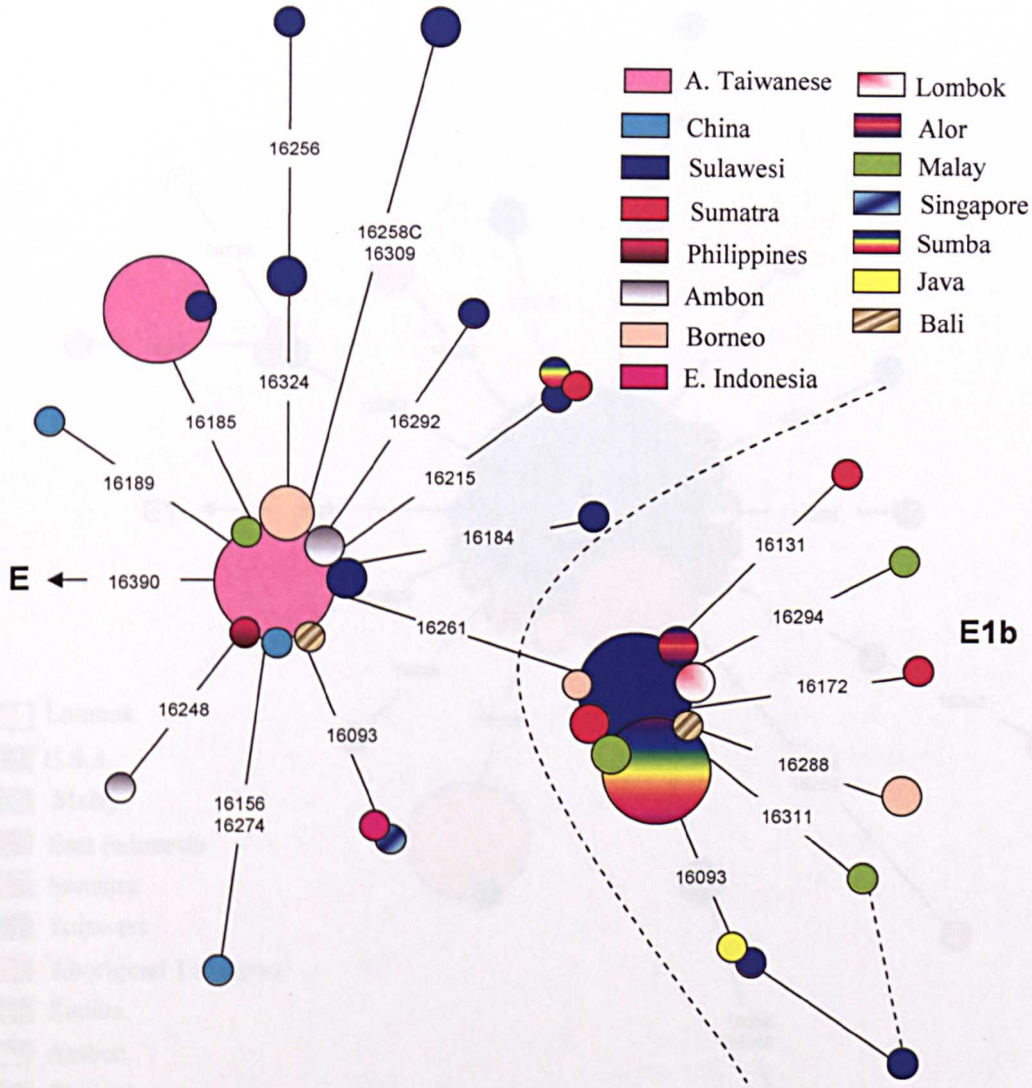
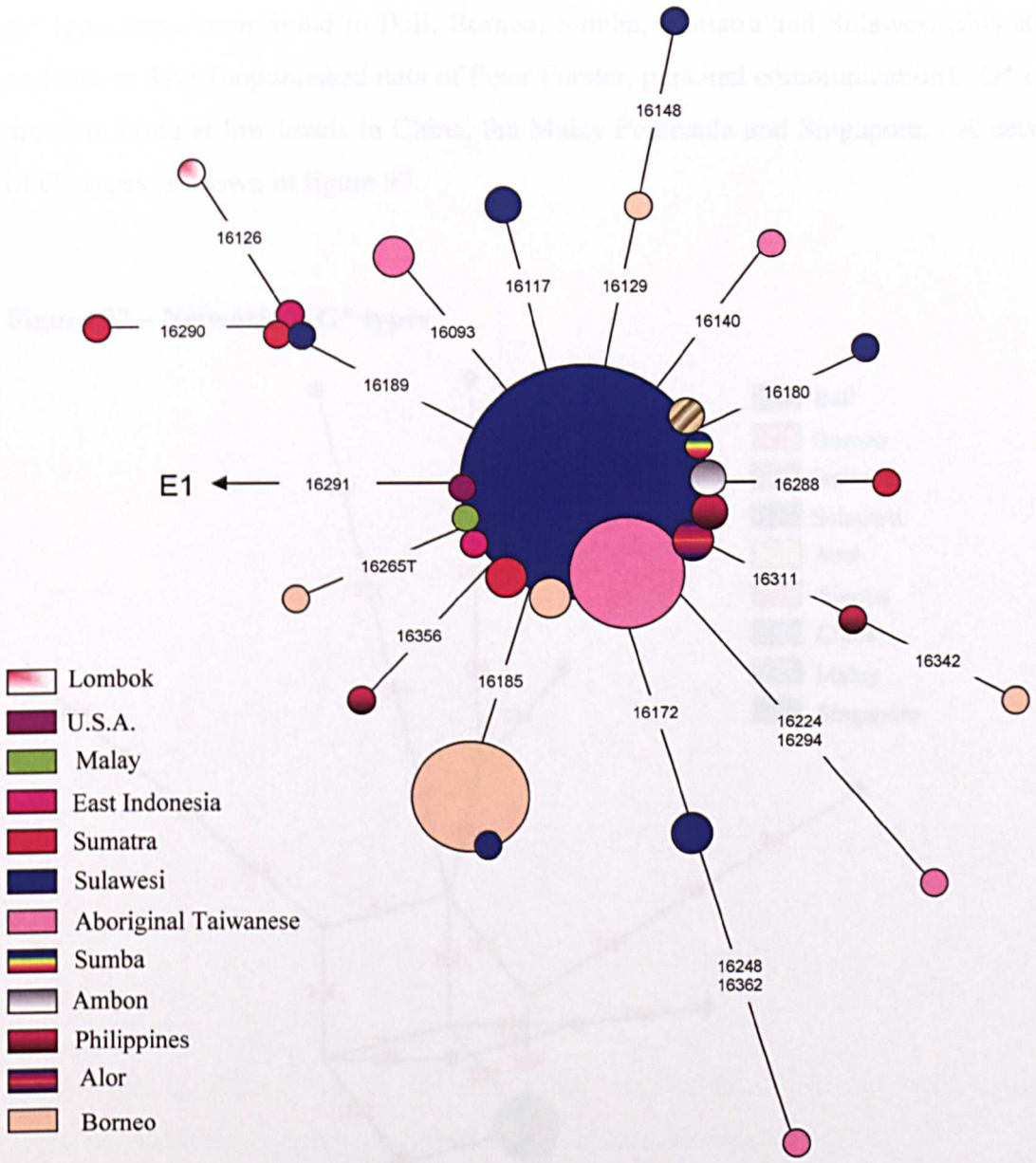


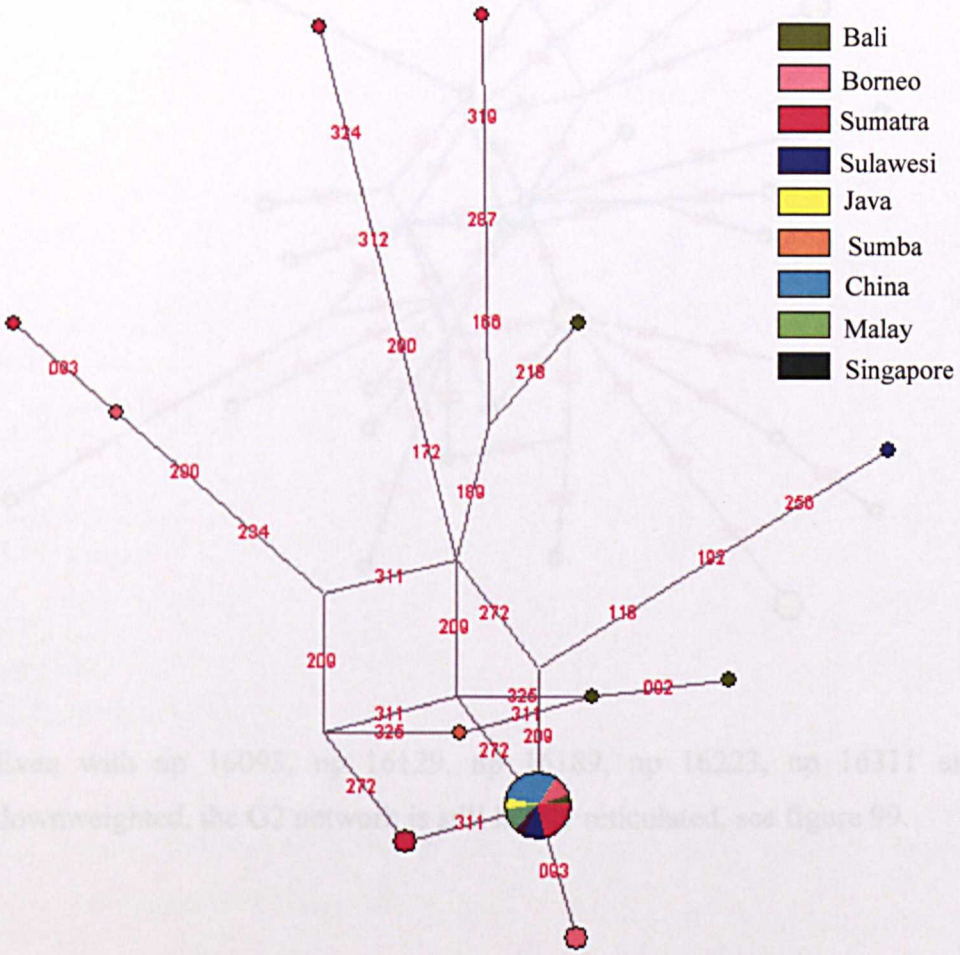
Figure 96 – Network of E1a types



Haplogroup G is defined by the gain of a *Hha*I site at np 4831 and has three main branches **G1** (defined by transitions at np 8200, np 15323 and np 15497), **G2** (defined by transitions at np 5601 and np 13563) and **G3** (defined by a transition at np 16274) (Kong *et al.* 2003). Some of the haplogroup G types found in this study most closely resemble haplogroup G1 types found in the Kamchatka peninsula of North-East Russia (Schurr *et al.* 1999); however, this would be a surprising result as G1 types have not been found in any of the intervening regions. Therefore, until complete sequencing has been used to resolve this issue, these samples will be treated as G* types.

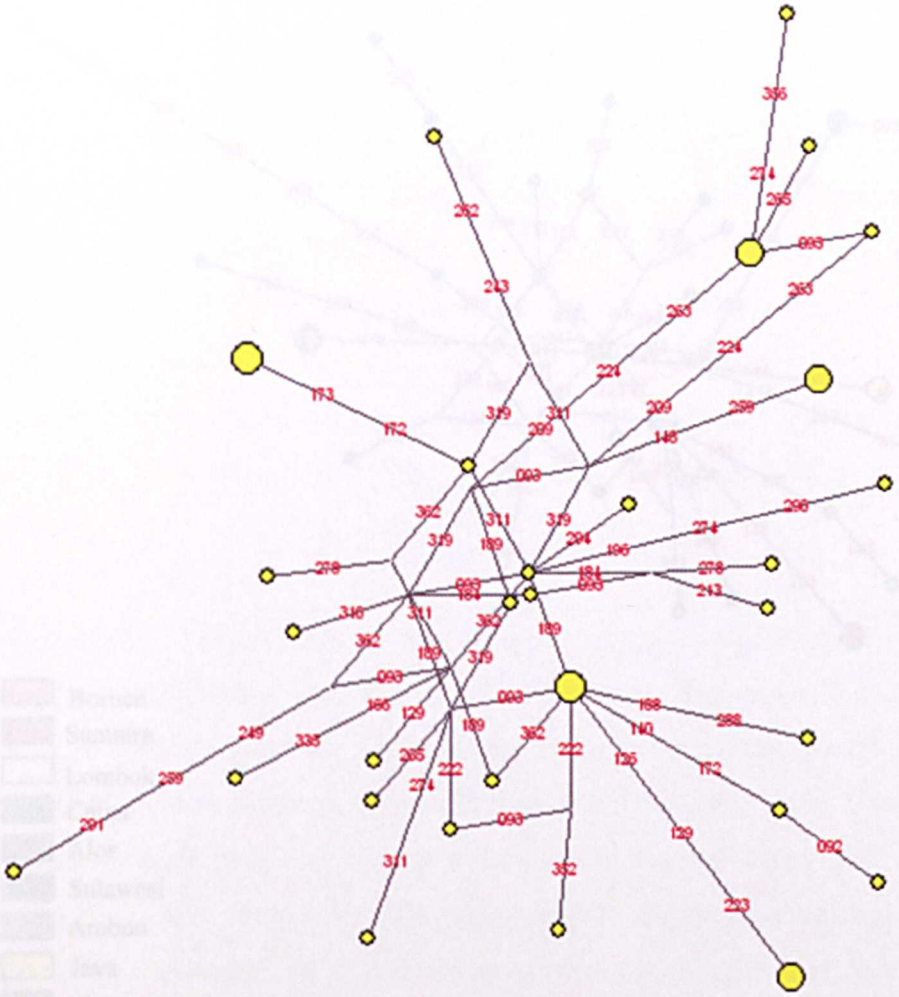
G* types have been found in Bali, Borneo, Sumba, Sumatra and Sulawesi (this study) and also in Java (unpublished data of Peter Forster, personal communication). G* types are also found at low levels in China, the Malay Peninsula and Singapore. A network of G* types is shown in figure 97.

Figure 97 – Network of G* types



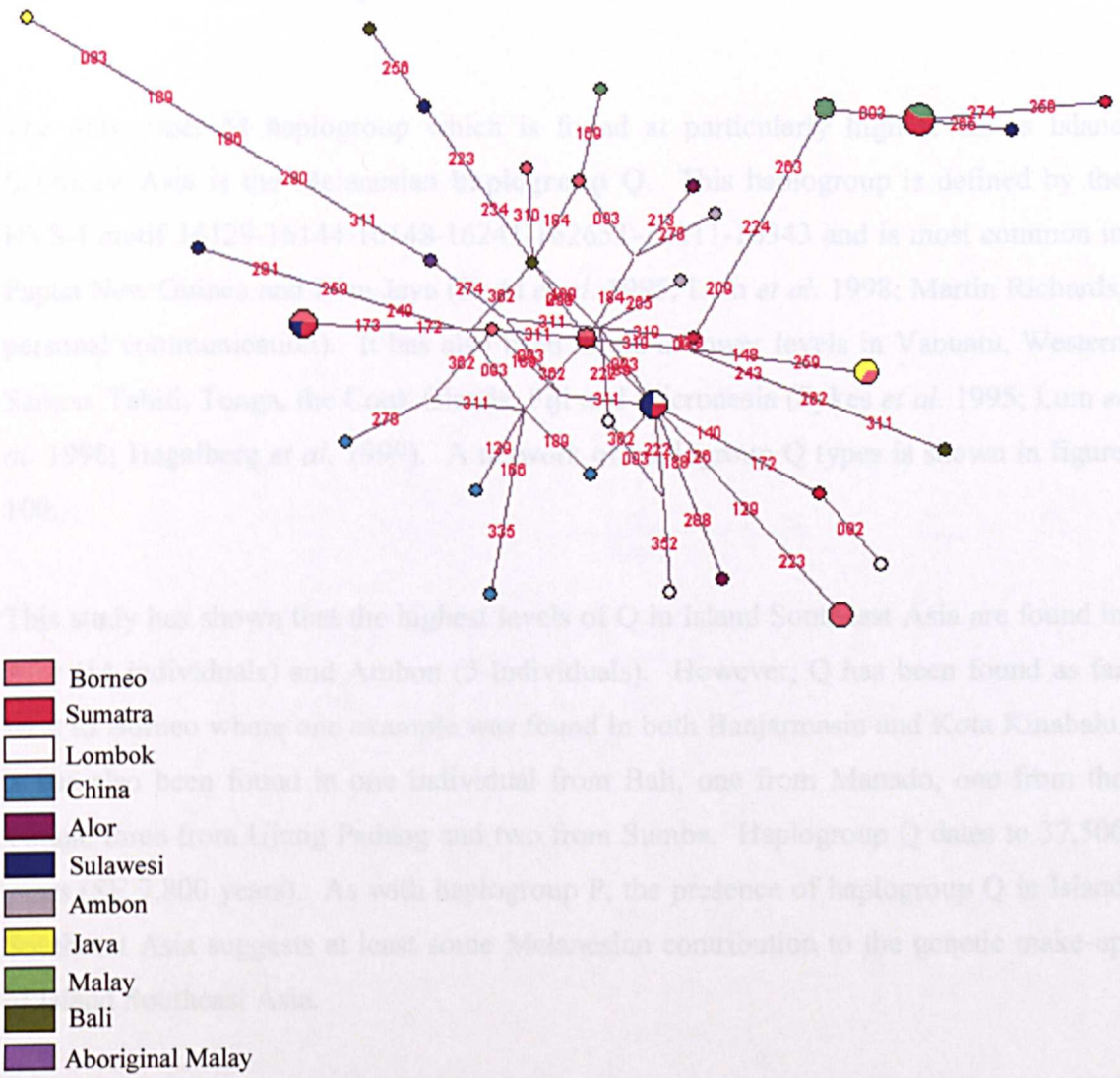
Haplotypes from the subclade **G2** are also found in Island Southeast Asia. G2 is also found, albeit at low levels, in China and the Malay Peninsula and has also been found in one Aboriginal Malay (Temuan). As seen in figure 98, the network of G2 types is highly reticulated.

Figure 98 – Unweighted network of G2 types



Even with np 16093, np 16129, np 16189, np 16223, np 16311 and np 16362 downweighted, the G2 network is still highly reticulated, see figure 99.

Figure 99 – Weighted network of G2 types



G2 is most common in Sumatra and Southern Borneo; however, more derived G2a types are common in China (not shown).

Haplogroup G was originally believed to be a principal branch of macrohaplogroup M; however, it has now been subsumed by **haplogroup M12** which is characterised by a transition at np 14569 and which includes all the branches of haplogroup G and another branch known as **M12a** (Kong *et al.* 2003). M12a is characterised by transitions at np 16234 and np 16290 and has been found at low levels in South China, Thailand,

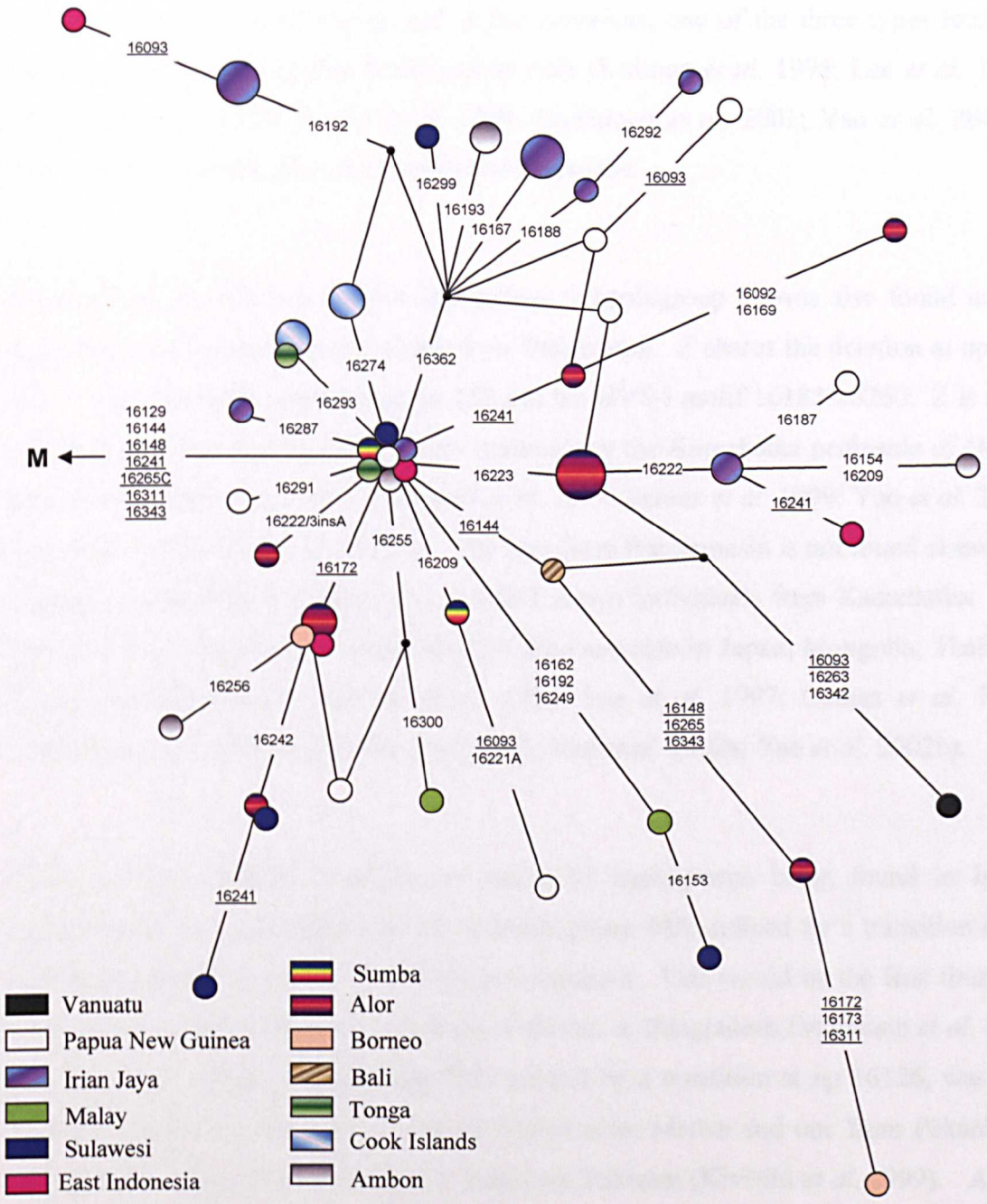
Vietnam and the Malay Peninsula. In this study it has been found in one individual from Bali and one from Banjarmasin.

The only other M haplogroup which is found at particularly high levels in Island Southeast Asia is the Melanesian **haplogroup Q**. This haplogroup is defined by the HVS-I motif 16129-16144-16148-16241-16265T-16311-16343 and is most common in Papua New Guinea and Irian Jaya (Redd *et al.* 1995; Lum *et al.* 1998; Martin Richards, personal communication). It has also been found at lower levels in Vanuatu, Western Samoa, Tahiti, Tonga, the Cook Islands, Fiji and Micronesia (Sykes *et al.* 1995; Lum *et al.* 1998; Hagelberg *et al.* 1999). A network of haplogroup Q types is shown in figure 100.

This study has shown that the highest levels of Q in Island Southeast Asia are found in Alor (14 individuals) and Ambon (5 individuals). However, Q has been found as far west as Borneo where one example was found in both Banjarmasin and Kota Kinabalu. It has also been found in one individual from Bali, one from Manado, one from the Toraja, three from Ujung Padang and two from Sumba. Haplogroup Q dates to 37,500 years (SE 7,800 years). As with haplogroup P, the presence of haplogroup Q in Island Southeast Asia suggests at least some Melanesian contribution to the genetic make-up of Island Southeast Asia.

There are also possible representatives of **haplogroup M10** in Island Southeast Asia, found in seven individuals from Java and five from Bangka. All these individuals have transitions at np 16223, np 16311 and np 16357 which are shared with most other known M10 types and around half of them match an M10 sample from Guangxi in China (Yao *et al.* 2002b). Apart from these, M10 has only been found in 8 other individuals from China (Yao *et al.* 2002a; Yao *et al.*2002b). Interestingly, all but one of the possible M10 types from Java and Bangka have lost the *Alu*I site at np 10397.

Figure 100 – Network of Q types



Most samples from Polynesia and Micronesia are only represented by partial HVS-I sequences and so were excluded from this analysis.

There are also a small number of minor M haplogroups found in the Island Southeast Asian sample. For example, **haplogroup C** (characterised by a deletion at np 249 and a transition at np 16327) was found in two individuals from Kota Kinabalu, and one from

each of Lombok, Java and Ujung Padang. Haplogroup C is most common in China (where it is found from Xinjiang in the north to Yunnan and Guangxi in the south), the Kamchatka peninsula of Russia, and in the Americas; one of the three types found in Island Southeast Asia is also found across Asia (Kolman *et al.* 1996; Lee *et al.* 1997; Nishimaki *et al.* 1999; Schurr *et al.* 1999; Fucharoen *et al.* 2001; Yao *et al.* 2002b), whereas the other two have not been found elsewhere.

Haplogroup Z, which is a sister haplogroup to haplogroup C, was also found in one individual from Banjarmasin and one from Palembang. Z shares the deletion at np 249 with C but also has a transition at np 152 and the HVS-I motif 16185-16260. Z is most common in China and Northern Asia – particularly the Kamchatka peninsula of North-East Russia (Lee *et al.* 1997; Nishimaki *et al.* 1999; Schurr *et al.* 1999: Yao *et al.* 2000; Yao *et al.* 2002a; Yao *et al.* 2002b). The type from Banjarmasin is not found elsewhere although a one-step derivative of it is found in two individuals from Kamchatka. The type found in Palembang is found across China and also in Japan, Mongolia, Thailand, Korea and Kazakhstan (Kolman *et al.* 1996; Lee *et al.* 1997; Comas *et al.* 1998; Nishimaki *et al.* 1999; Fucharoen *et al.* 2001; Yao *et al.* 2002a; Yao *et al.* 2002b).

There are also possible instances of Indian M haplogroups being found in Island Southeast Asia. A possible example of **haplogroup M5**, defined by a transition at np 16129, was found in one individual from Pekanbaru. This would be the first time this haplogroup has been found outside India, Pakistan or Bangladesh (Mountain *et al.* 1995; Kivisild *et al.* 1999). **Haplogroup M3**, defined by a transition at np 16126, was also found in Sumatra – this time in one individual from Medan and one from Pekanbaru. Again, M3 is only found elsewhere in India and Pakistan (Kivisild *et al.* 1999). Also a possible example of **haplogroup M2**, defined by a transition at np 16319, was found in another individual from Medan, once again M2 is only found elsewhere in India (Mountain *et al.* 1995). These could be the result of historical migrations from the Indian subcontinent.

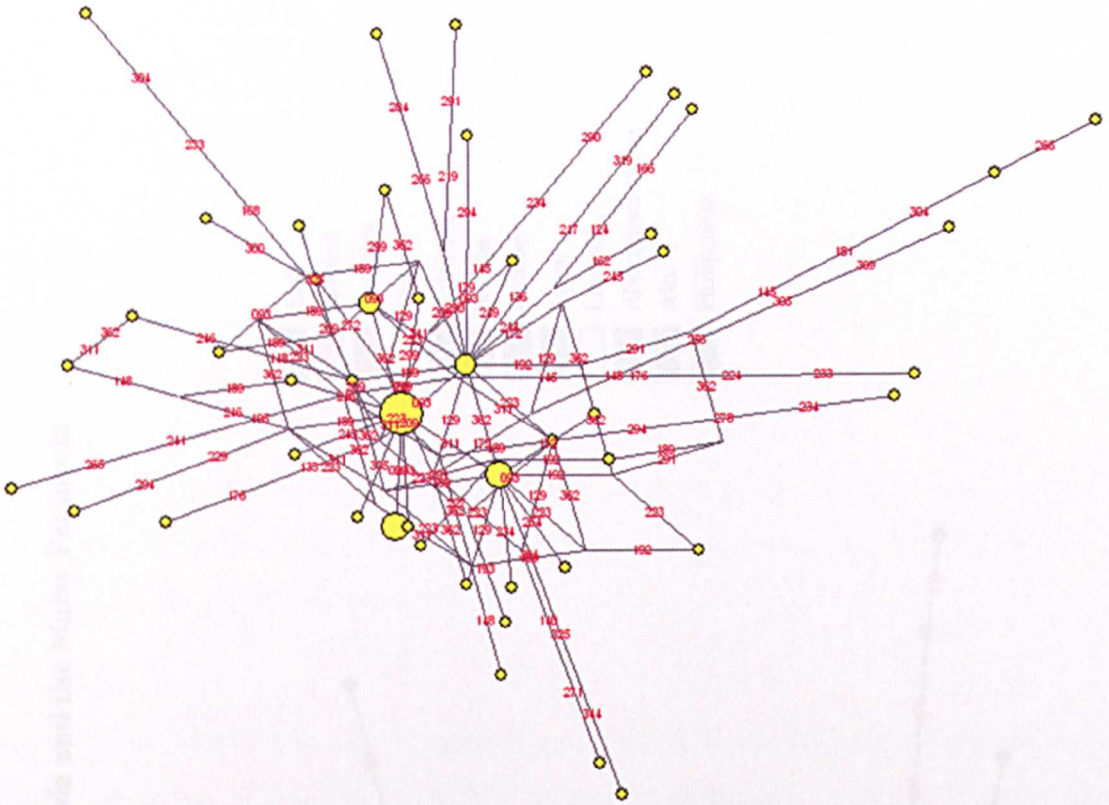
6.6 Unresolved M*, N* and R* types

The remaining samples are classified as either M*, N* or R* as they possess the relevant restriction sites but cannot be further categorized into one of the subclades of M, N or R. Of the three, the M* types are more common and make up 6% of the sample as a whole. M* types are most common in Java (17%) but are also relatively frequent in Bangka (15%), Bali (14%), Mataram (14%), Palu (11%), Medan (10%) and Banjarmasin (10%).

Of the M* types, most are found only once in the dataset, the main exceptions to this have a combination of transitions at the following nucleotide positions: 16093, 16223, 16311 and 16362. As these positions mutate comparatively frequently it is impossible to say for certain that these M* types are related without employing complete sequencing. However, the four mutations are found together in only four individuals from Java and one from Banjarmasin (elsewhere they have only been found on an M* background in one Taiwanese Han individual) which suggests they form a separate clade.

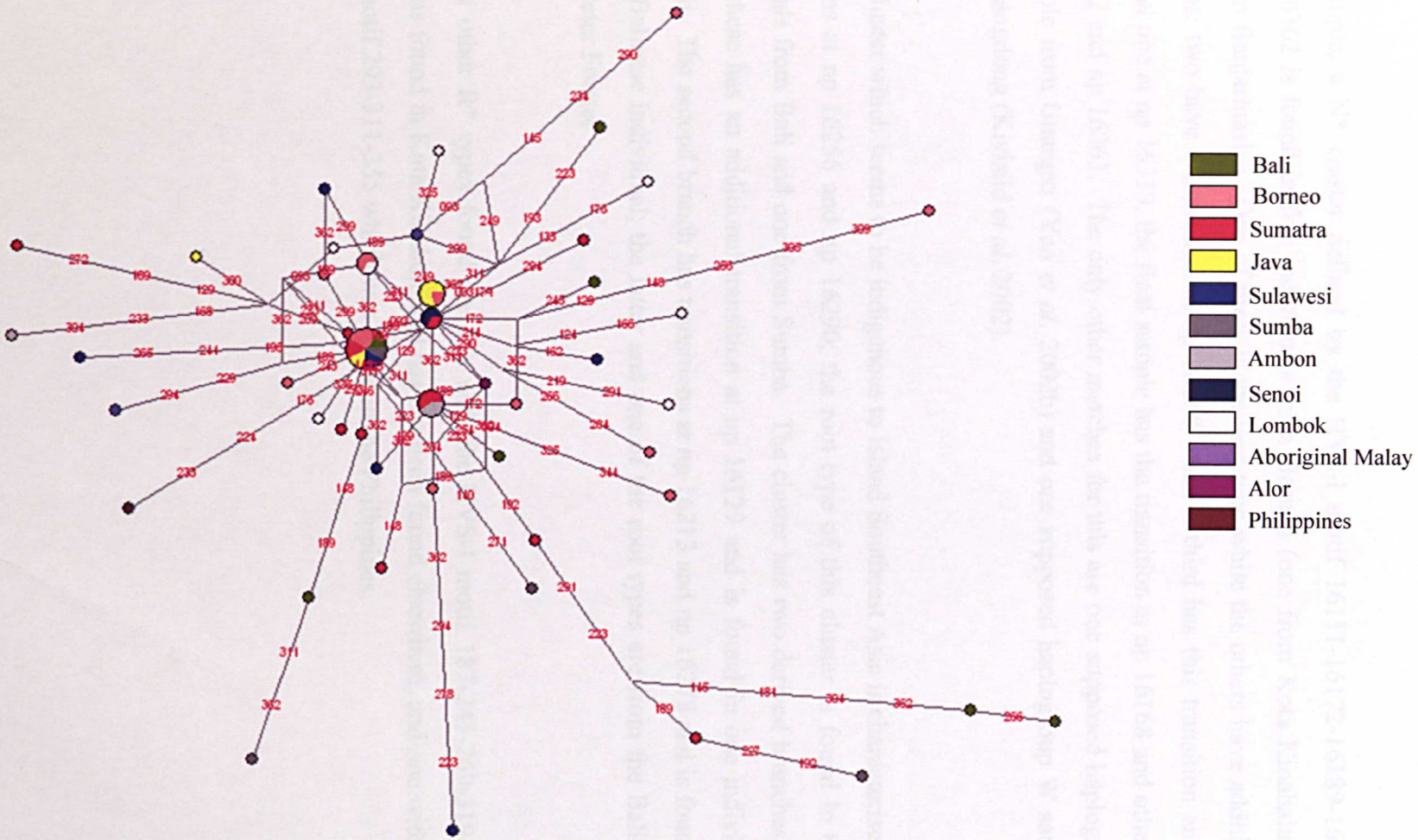
There is also a rare possible clade with the HVS-I motif 16145-16181-16192-16223-16291-16304 which has been found in two individuals from Bali, it has also been found in one Malay (data of Zainuddin and Goodwin 2004). The network of M* types is extremely reticulated, see figure 101.

Figure 101 – Unweighted network of M* types from Island Southeast Asia and the Malay Peninsula



Even with the more frequent transitions (at np 16093, np 16129, np 16189, np 16311 and np 16362) downweighted, the M* network is still highly reticulated (see figure 102).

Figure 102 – Weighted network of M* types from Island Southeast Asia and the Malay Peninsula



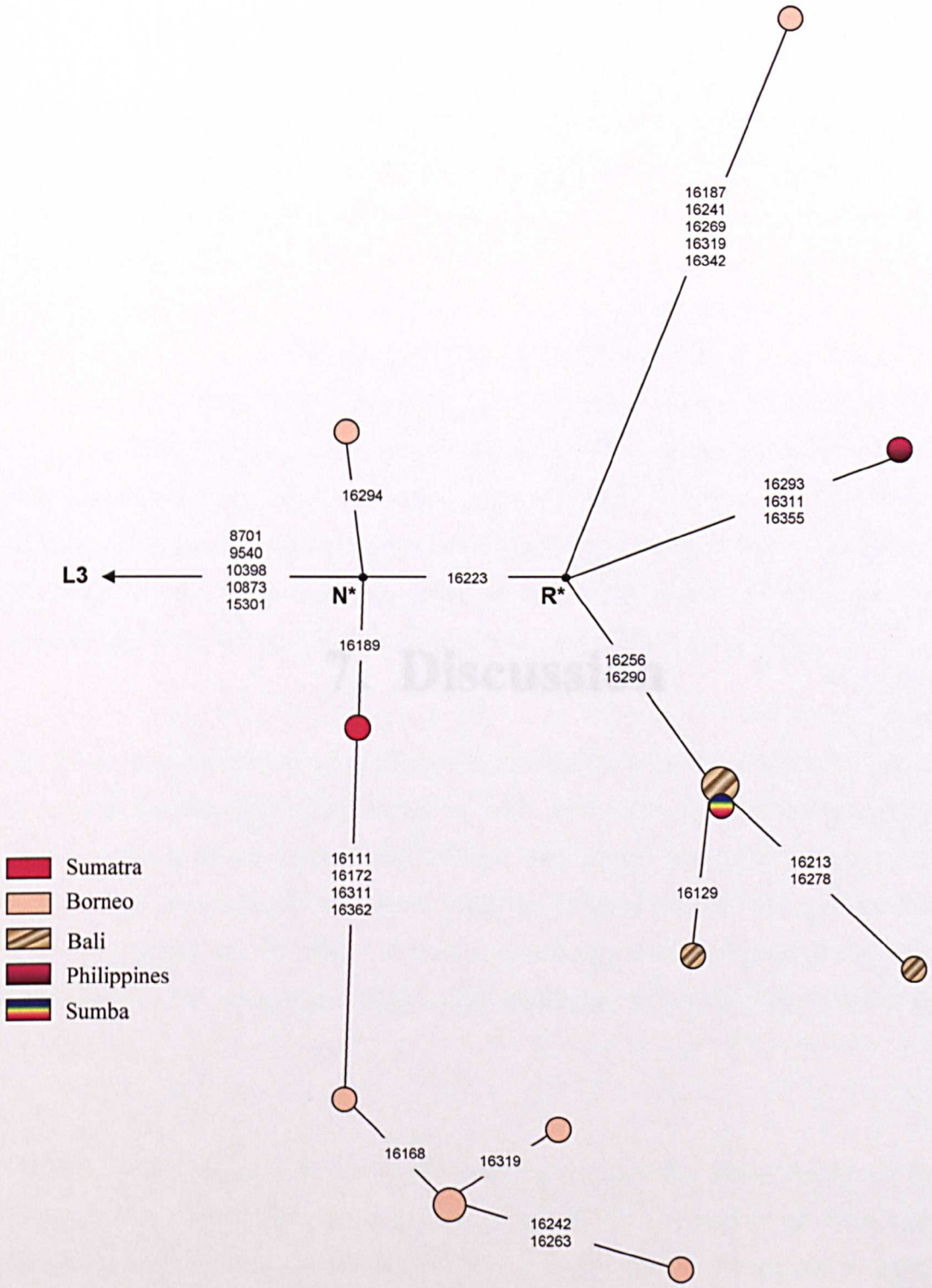
N* and R* types are less frequent in the dataset than M* types, making up only 1% and 0.6% of the dataset respectively. They are most frequent in Bali where they make up 4.5% of the population. Again there appear to be a couple of rare clusters that are specific to Island Southeast Asia, see figure 103.

For example, a N* cluster defined by the HVS-I motif 16111-16172-16189-16223-16311-16362 is found in five individuals from Borneo (one from Kota Kinabalu and four from Banjarmasin). One of these is the root type while the others have additional mutations; two have a transition at np 16168 while a third has this transition and an additional one at np 16319, the final sample has the transition at np 16168 and others at np 16242 and np 16263. The only other matches for this are one supposed haplogroup D5 sample from Guangxi (Yao *et al.* 2002b) and one supposed haplogroup W sample from Guangdong (Kivisild *et al.* 2002).

An R* cluster which seems to be indigenous to Island Southeast Asia is characterised by transitions at np 16256 and np 16290; the root type of this cluster is found in three individuals from Bali and one from Sumba. The cluster has two derived branches, the first of these has an additional transition at np 16129 and is found in one individual from Bali. The second branch has transitions at np 16213 and np 16278 and is found in another Balinese individual; the latter and one of the root types are from the Balinese data of Peter Forster.

The only other R* types found are one with the HVS-I motif 187-241-269-319-342 which was found in Kota Kinabalu and has not been found elsewhere, and one with the HVS-I motif 293-311-355 which was found in the Philippines.

Figure 103 – Network of N* and R* types in Island Southeast Asia



7. Discussion

7. Discussion

7.1 Discussion of Results of Orang Asli Study

All three groups of Orang Asli (especially the Semang and, to a lesser extent, the Senoi) have been subject to relatively high levels of genetic drift, as shown by the non-starlike topologies of the haplogroup networks and the elevation of a small number of sequence types to very high frequencies (see Appendix I). This is not unexpected given their extremely small group sizes. In particular the Semang groups as a whole consist only of around 2,000 individuals (Bellwood 1993) and perhaps as few as 1,300 (Carey 1976). The Senoi and Aboriginal Malays are more extensive and consist of approximately 25,500 and 19,000 individuals respectively (Carey 1976). It also appears that the Orang Asli have been isolated for a substantial amount of time. This is evident in the number of 'new' haplogroups which have been found in the course of this investigation e.g. M21, M22, N21, N22 and R21, most of which are found predominantly, if not exclusively, in the Orang Asli (see Appendix I, figures 56-57 and 85-86).

The three groups are strikingly different in their haplogroup composition. The Mendriq and Batek Semang are rather similar to each other, although interestingly the Jahai Semang appear closer to the Temiar Senoi than to the two other sampled Semang groups. The two sampled Aboriginal Malay groups, the Semelai and the Temuan, are also rather similar to each other - at least as similar as are the Sumatran groups sampled, who resemble the Aboriginal Malays much more closely than they do the other Orang Asli groups.

There is strong evidence for an indigenous origin within the Malay Peninsula for the Semang. The predominant Semang haplogroup, M21a, is found in all three sampled Malay Semang groups, but elsewhere in only a few Senoi, Aboriginal Malays, and *Melayu* Malays, in one individual from Borneo and in the Semang of Southern Thailand (Fucharoen *et al.* 2001). Distinct sequence types (including the root type in the Semelai only) occur in all of the Orang Asli groups, suggesting that the drift has been relatively

recent and that fragments of the pre-existing diversity have survived in all of the Orang Asli groups. The sister clade M21b is much rarer but is also found in all three groups as well as in Borneo, Sumatra and Sulawesi and is possibly related to sequences found in Thailand (see figure 85). Complete sequencing has shown that M21 as a whole dates to ~57,000 years, with the split of M21a and M21b dated to ~44,000 years (Macaulay *et al.* 2005).

The second major Semang haplogroup is R21 which appears predominantly as a signature of the Jahai Semang (62%) and the Temiar Senoi (37%). It is present in all of the sampled Semang groups, the ancestral type is found only in one Batek, and there are no closely related lineages found anywhere else in the world (see Appendices I and III). R21 may therefore, like M21a, represent an ancient Semang clade that has been retained in the Temiar Senoi. The distribution contrasts with that of B5b in the Orang Asli, which is restricted largely to the Batek Semang, suggesting a much more recent intrusion, probably from Island Southeast Asia (see figure 74).

The Semang therefore appear to represent the descendants of the indigenous peoples of the Peninsula, who have experienced some minor subsequent gene flow from outside, probably in the recent past. However, the three Semang groups are somewhat different to each other in their haplogroup distributions, and the Jahai in particular resemble the Temiar rather more than they do the other Semang groups. None of them resemble the Andamanese (who have mainly Asian Subcontinental mtDNAs: Endicott *et al.* 2003), suggesting that the term 'Negrito' should not be used to imply an over-simplified genetic history of these peoples. The lack of the main Semang mtDNA clades in other populations is consistent with the theory that the origins of these people lie deep in the Pleistocene (Bellwood 1993; Bellwood 1997). This is supported by the early dates calculated for these haplogroups and the fact that they lie near the roots of macrohaplogroups M and R.

Only 2% of the Temiar Senoi sample was found to belong to haplogroup B, which is in sharp contrast to the 37% found in the Semai Senoi by Melton *et al.* (1995). This

suggests that, like the Semang, the different Senoi groups might actually have quite different genetic compositions. Aside from R21, referred to above, the most common haplogroup in the Senoi is F1a1a, which was found in 43% of individuals and which occurs predominantly elsewhere in Thailand and China, although it is found at lower levels across Island Southeast Asia (see figures 79 and 83). F1a1a is most diverse in South China and dates to 9,200 years (see tables 7 – 8). This suggests that almost half of the maternal lineages of the Senoi can be traced back to an origin in Indo-China at some point within the last 9,000 years or so. This is consistent with the view of Bellwood (1993) that the Neolithic was brought into peninsular Malaysia by Ban Kao groups from Southern and Central Thailand who interbred with some groups ancestral to the Semang to create the ancestral mix of the Senoi.

N9a1 may also represent Neolithic input into the ancestors of the Senoi, as its root type has been found in Thailand and Vietnam (see figure 59). However, it is also found in Sulawesi and the clade as a whole is also found in the Semang, the Aboriginal Malays and other locations in Island Southeast Asia.

The two Aboriginal Malay groups, the Semelai and Temuan, have very different haplogroup distributions from both the Semang and the Senoi, and mainly carry the novel haplogroups R9b (26%) and N21 (25%). As seen in their heterozygosity values (see table 6), they show higher levels of diversity than either the Semang or the Senoi, and show some similarities to Island Southeast Asians, with whom they cluster in the PC plot (see figures 36 – 37). The Temuan are also closer linguistically to Island Southeast Asians as they speak an Austronesian language; the Semelai, however, speak a branch of Southern Aslian. R9b is present in both the Semelai and Temuan and has drifted onto just one main type, with a few more ancestral types to suggest a former higher diversity (see figure 78). Related types are found in Sumatra, Borneo, Sulawesi, Java, Lombok, Thailand and China. The Chinese data (Yao *et al.* 2002a; Yao *et al* 2002b) indicate a small but diverse collection of types, all but one of which are from South China (Yunnan and Guangxi), suggesting a likely origin there; however, its route into the Malay Peninsula remains unclear due to reticulation in the HVS-I network; this may be resolved by complete sequencing. A southern route from China through

Thailand and the Malay Peninsula to Island Southeast Asia is possible. However, the presence of Island Southeast Asian samples deeper in the tree suggests that R9b may have been introduced to the Malay Peninsula from Sumatra. Unfortunately it is difficult to be certain due to the relative scarcity of R9b types and the amount of drift which has occurred in the Aboriginal Malays.

N21 is also found in both the Aboriginal Malays and in Island Southeast Asia. The root type of N21 is only found in single individuals from Bali and Sulawesi. However, individuals from the Malay Peninsula are found on both branches of N21 (see figure 56). One branch has been found in Sumatra and in the Malay data of Zainuddin and Goodwin (2004) while the other is found in Alor and is raised to high levels in the Aboriginal Malays. It seems likely that N21 is a rare, indigenous Island Southeast Asian haplogroup which was introduced into the Malay peninsula from Sumatra and which has been raised to high levels in the Aboriginal Malays due to drift.

The only other haplogroup found in Aboriginal Malays but in no other Orang Asli group is M7c1c, which is found only in the Aslian-speaking Semelai. As M7c1c is otherwise almost exclusively Austronesian (see figure 92), this suggests either gene flow from an Austronesian-speaking group, or that the Semelai originally spoke an Austronesian language and have only adopted an Aslian language fairly recently. Other haplogroups found in the Aboriginal Malays include M21a, M21b, F1a1a and N9a1, probably indicating gene flow from the Semang and/or Senoi. Two Semelai belonged to B4a* which was found in 36.7% of the Semai Senoi samples studied by Melton *et al.* (1995), and so could also be due to gene flow from the Senoi. However, this haplogroup is widespread over much of Southeast and East Asia (see figure 66). N22 and M22 were also found but are restricted to the Temuan. N22 has only been found elsewhere in Sumba; the origins of this clade are unclear (see figure 57). M22 is also found in Thailand, and probably represents introgression from the north (see figure 86).

The origins of the Aboriginal Malays are the least clear of the three Orang Asli groups. However, they do contain a definite Island Southeast Asian component. There is,

furthermore, a small component deriving from the Semang/Senoi. However, there is only a very small potential Taiwanese component, represented by M7c1c, which makes up 8.5% of the Aboriginal Malay sample. The principal component analysis indicates that the Aboriginal Malays fall closest to the Sumatrans in the first two components (amounting to 43% of the variation), but the direction of exchange is unclear (see figure 36). Again, substantial differences can be seen between the two Aboriginal Malay groups, with N22, N9a1 and M22 mainly found in the Temuan and M7c1a and B4a* restricted to the Semelai. This suggests that, as with the Semang and the Senoi, the tri-partite classifications are an oversimplification. Thus, while the crude classification does have some biological meaning, there is also substantial internal diversity within the three groups.

In some respects, the mtDNA results from Malaysia support the view of Orang Asli history suggested by Bellwood (1993). The Semang may represent descendants of the first modern human migrants to the peninsula and have retained ancestral links with the Senoi. The three Semang groups studied are, however, rather different and the Jahai in particular are very different on the maternal side from the Batek and Mendriq, most likely as a result of intermarriage with the Temiar. The Senoi themselves seem to be a product of immigrants from Indo-China and the pre-existing ancestors of the Semang. The immigrants may have introduced the Austro-Asiatic language family to the Peninsula, along with swidden agriculture, about 4,000 years ago. The origins of the Aboriginal Malays are unclear but definitely include some Island Southeast Asian influence. Like the other groups of Orang Asli, they too appear to have a complex origin, and there is no clear evidence in the mtDNA data of an ancestry for these people in Taiwan.

These results would appear to rule out the more extreme 'local evolutionary' models for Orang Asli ethnogenesis. However, a more complex model such as that proposed by Rayner and Bulbeck (2001) stands up well in the light of the mtDNA data. Such a model would incorporate some local evolution for all three groups, from at least the early Holocene onwards, but also allow for some immigration—from the north, affecting the gene pool of the Senoi, and from the east, affecting the Aboriginal Malays.

Interestingly though, complicating factors such as the close links between the Jahai and the Temiar do not seem to have been recognised by any of the models proposed in the past.

7.2 Discussion of Results of Island Southeast Asian Study

In contrast to the Orang Asli groups of peninsular Malaysia, the various populations of Island Southeast Asia have a high level of genetic diversity. This can be seen in the high levels of heterozygosity maintained in all groups (see table 6). Even the least diverse Island Southeast Asian group (that from Tengger in Java) is more diverse than the most variable Orang Asli group (see table 6). This diversity is maintained across the Island Southeast Asian sample when taken as a whole. The sample contains 48 different haplogroups or subhaplogroups; of these 48, only 4 (M*, M7c1c, E1a and F1a*) are found in more than 5% of the samples. Nineteen haplogroups are each found in less than 1% of the population. This high level of diversity suggests that the Island Southeast Asian groups have been of relatively large size over long periods of time and so have not been as susceptible to genetic drift as the Orang Asli groups. Eleven haplogroups (D4a, F1b, G3, I, M2, M3, M5, M7c1*, M7c1a, U7 and Z) are represented in the dataset by only one or two individuals; these may provide evidence of relatively recent gene flow from other locations. This seems particularly likely in the cases of haplogroups I (which is mainly Northern European), U7 (found in the Middle East and the Indian subcontinent), M2, M3 and M5 (all most common in the Indian subcontinent).

7.2.1 Geographic Structuring of Island Southeast Asia

As discussed in the background to this investigation, it has often been suggested that modern Island Southeast Asians are the descendents of Neolithic immigrants from Taiwan who largely replaced the indigenous hunter-gatherer populations due to their technological 'superiority' (Diamond 1988; Bellwood 1997). If this was the case, then it might be expected that the modern populations of Island Southeast Asia would be relatively similar to each other as they would be descended from the same Taiwanese founder types. However, a number of analyses conducted in the course of this investigation suggest that the picture is not that simple.

For example, the principal component analysis carried out on the data showed east-west patterning in principal components one and two (which accounted for 28.4% of the variation; see figure 33). In this analysis, most of the Eastern groups can be seen to cluster together, the only exceptions being the groups from Mataram and Palu. In the case of Mataram this is not particularly surprising as it lies almost on the boundary between west and east (which I have divided at the Wallace line). Palu is more unusual as it definitely lies within the Eastern part of Island Southeast Asia. However, it does have some Western characteristics in that it is one of the few locations in the east to contain any of haplogroups Y2, N9a1 or G2. It is probably for this reason that it is found closer to the Western populations in the principal component analysis. This east-west patterning suggests that geographic structuring occurs across Island Southeast Asia.

This geographic structuring was confirmed by carrying out an AMOVA test in which the Island Southeast Asian populations were found to be significantly different from each other (see table 11). Furthermore, when the groups were divided into Eastern and Western groups (according to their position relative to the Wallace line) significant differences were found between the two. The difference was even more significant when a central group (representing Java, Borneo, Bali and Lombok) was also separated out. This suggests that the Eastern and Western populations have, at least to some extent, different demographic histories and have perhaps been influenced by different external groups.

However, no significant differences were found when the groups were separated according to language (see table 11). This could suggest that the linguistic and genetic histories of the groups do not correlate. However, all but three of the groups studied speak Western Malayo-Polynesian languages; this bias may have influenced the results. Perhaps if more Central or Eastern Malayo-Polynesian speaking groups had been included in the study, a more significant difference may have been found.

Pairwise F_{ST} values confirmed that most individual Island Southeast Asian populations are significantly different from each other (see section 5.7). They also demonstrated that significant differences are even found across individual islands. For example, in

Sumatra the populations from Medan, Palembang and Pekanbaru are all significantly different to each other.

The sheer amount of variation and geographic differentiation present in Island Southeast Asia suggests that invoking an influx of Taiwanese immigrants who induced an almost complete population replacement may be too simplistic an argument. If this did occur, the incoming group must have been relatively large to maintain so much variation within it, yet the differences found between populations would seem to imply a history of small populations and genetic drift or founder effects.

7.2.2 Does an Austronesian Signature Exist in Island Southeast Asia?

As discussed in section 1.5, it has previously been claimed that haplogroup B4a1 represents an Austronesian 'signature' due to its high frequency in Polynesia and the presence of ancestral types in Taiwanese aboriginals (e.g. Redd *et al* 1995; Melton *et al.* 1995; Sykes *et al.* 1995). However, the people of Polynesia are unusual due to their relatively recent ancestry and the numerous founder events which have occurred during their history. Because of this, certain haplogroups, particularly B4a1, are raised to extremely high frequencies in Polynesian groups, which may not be representative of the rest of the Austronesian-speaking world.

This is confirmed by the fact that B4a1 has been found to be relatively rare in this study, making up only 2.1% of the population as a whole and reaching a high of 14% in Ambon (see figures 67 and 69). Furthermore, B4a1 is absent across much of Island Southeast Asia and is not found further west than Southeastern Borneo. If any single haplogroup can be argued to be an Austronesian 'signature', M7c1c seems like a much more plausible candidate. M7c1c has been found in all locations studied in this investigation and is the third most common haplogroup in Indonesia making up 8.1% of the whole sample. It is also found in Taiwanese aboriginals, the Philippines, Micronesia and in one individual from Fiji (see figure 92).

M7c1c has been found in a small number of non-Austronesian speakers: 1 from the Guangxi province of Southern China, 8 Semelai, 1 Thai and 5 African Americans (see

figure 92). However, all but one of these belong to the root type and so can probably be ascribed to recent gene flow or perhaps language shift in the case of the Semelai. The only exception is the individual from Guangxi. The latter is a one-step derivative of the root type and is shared with two individuals from Sumba. A third Sumbanese individual is derived from this type by another two transitions suggesting that this branch originated in Island Southeast Asia. M7c1c is most diverse in Taiwan and Northern Borneo which suggests an origin in that region. Furthermore, its starlike phylogeny indicates that it has undergone a population expansion into East and West Indonesia. This, along with the fact that M7c1* is found in China and its date of ~6,000 years (SE 1,600 years) means that M7c1c could have been part of an 'out of Taiwan' event. However, it could also have been part of a mid-Holocene dispersal centred on Borneo.

Several other haplogroups have similar distributions and expansion ages. For example, haplogroup E1 (including E1a and E1b) is almost entirely restricted to Austronesian-speaking areas; it has only been found elsewhere in three individuals from China, one from Singapore and in one African American (see figures 95 – 96). E1 is also relatively common in Island Southeast Asia, being found in just over 14% of the population. Like M7c1c, E1a is most diverse in Taiwan and Borneo (see table 8); however, a number of diverse E1* types are found in Sulawesi. E1 as a whole dates to 23,900 years (SE 10,000 years), E1a dates to 8,700 years (SE 2,700 years) and E1b dates to 7,300 years (SE 3,200 years), see table 7. The age of E1, and its almost complete absence in any other locations, suggests that it is indigenous to Island Southeast Asia and Taiwan. The lack of E1 types in China emphasises the long history of separation between China and Taiwan, although ultimately a mainland origin is indicated by the distribution of its ancestral haplogroup, M9.

Haplogroup F3b makes up 1.7% of the dataset and is also only found in Taiwan, the Philippines and Borneo (with the exception of one individual from Sumba and three African Americans) and dates to 34,000 years (SE 13,300 years), see figure 84 and table 7. Again, this haplogroup seems to have originated somewhere around that area. M7b3 is another rare haplogroup which is restricted to Taiwan and Island Southeast

Asia, in this case the Philippines, Sulawesi and Sumba (see figure 90). M7b3 dates to 15,400 years (SE 5,900 years), see table 7.

As discussed above, B4a1 has a quite different geographic distribution to M7c1c, E1, F3b and M7b3 as it is relatively rare in Island Southeast Asia, yet reaches extremely high frequencies in Polynesia and Micronesia. Within Island Southeast Asia it is most common in Sulawesi; however, the only haplotype found in Sulawesi is the root type, these may therefore be the result of a recent expansion into that area (see figure 69). B4a1 seems likely to have arisen in the eastern area of Indonesia which is where it is most diverse. When the types found in Sulawesi are removed, B4a1 dates to 8,800 years (SE 4,200 years) in Island Southeast Asia. Therefore, despite the differences in the geographic distribution of B4a1 and those of the haplogroups discussed above, it also seems to have originated too early to have been part of an 'out of Taiwan' event as it is traditionally visualised.

All the haplogroups discussed above are found predominantly in Austronesian-speaking peoples, yet (with the possible exception of M7c1c) they seem to be too old to fit in with the 'out of Taiwan' theory proposed by Bellwood (1997) amongst others. Bellwood (1997) suggests dates of 3,500-4,500 years ago for Taiwanese immigrants to arrive in Eastern Island Southeast Asia, and he only sees an expansion out from this central area occurring after 3,500 years ago. This, therefore, leaves a large gap between the theoretical dates and the dates obtained for the 'Austronesian' haplogroups. Many of the dates obtained in this project have large confidence intervals (see table 7). However, the dates do seem to support each other. If only one of these 'Austronesian' haplogroups dated too early for a potential 'out of Taiwan' event then it could be seen as questionable. However, as a number of them appear to be too old then the argument is strengthened. As it is, this data seems more supportive of an argument such as that of Solheim *et al.* (in press; personal communication) or Meacham (1984-1985) who envisaged a network of trade and movement across Island Southeast Asia through the Philippines and into Taiwan.

7.2.3 Rare Indigenous Haplogroups in Island Southeast Asia

There are rare instances of other indigenous haplogroups in the dataset. For example, there is a rare branch of R which is characterised by transitions at np 16256 and np 16290 and which is only found in five individuals from Bali and one from Sumba (see figure 103). There is no evidence for this cluster being found anywhere outside Island Southeast Asia which suggests it is indigenous to the area.

An indigenous Island Southeast Asian origin for haplogroup N21 also seems most likely on the current evidence. The root type is only found in Bali and Sulawesi and individuals from Island Southeast Asia are found on both branches (see figure 56). It seems probable that N21 was passed into the Aboriginal Malays from Sumatra. One of the branches of N21 is found in one Sumatran and one Malay, this branch may also have originally been present in the Aboriginal Malays but subsequently been lost to drift.

Haplogroup R22 may also have originated in Island Southeast Asia although this is more difficult to determine. It is most common in Bali, Lombok and Sumba, and the root type is only found in Lombok and Alor. However, some of the most derived types are found in Thailand and the Nicobar Islands (see figure 77). Unfortunately the direction of gene flow is impossible to determine from the current dataset.

7.2.4 Is There any Evidence of a Recent 'Out of Taiwan' Event?

Unfortunately, like the Orang Asli, the Aboriginal groups of Taiwan have undergone large amounts of genetic drift which makes it difficult to assess their earlier diversity. As discussed above, it is possible that M7c1c entered Island Southeast Asia via Taiwan. Are any other haplogroups plausible 'out of Taiwan' candidates?

Haplogroup B4a* makes up 4.8% of the Island Southeast Asian sample. It is found in most areas but is most common in Sumatra. B4a seems to have its ultimate origins in China as many diverse types are found there; however, it is most common in Taiwan and is also found in the Philippines (see figure 66). It is difficult to be certain because

of the high level of reticulation seen in the network of B4a* types, but a passage through Taiwan into Island Southeast Asia does seem most likely on the current evidence.

Haplogroup B4c (which makes up 3% of the Island Southeast Asian sample) is another possible candidate that could have been carried into Island Southeast Asia from Taiwan. The root type of B4c is only found in China and Sulawesi; however, the branch characterised by a transition at np 16335 (which is most common in Island Southeast Asia) is found in Taiwan (see figure 71). This branch is also more diverse in Taiwan than in Island Southeast Asia, suggesting that it too could have been part of a migration from Taiwan.

Haplogroup B5a may also represent some Taiwanese input into Island Southeast Asia. B5a makes up 4% of the Island Southeast Asian sample, but is most common in China and Taiwan (see figure 73). There are also shared types between Island Southeast Asia and Taiwan. B5a is not found in the Philippines; however, the amount of data available from the Philippines is much smaller than that available for the other areas in question so this may simply be due to insufficient sampling.

However, despite their potential Taiwanese ancestry, B4a*, B4c and B5a all seem to be too old to be part of a traditional 'out of Taiwan' event. B4a dates to ~25,000 years, B4c dates to ~13,000 years in Island Southeast Asia and B5a dates to ~9,500 years in Island Southeast Asia (see tables 7 – 8). Therefore, like E1, F3b and perhaps M7c1c, they seem to have been present in Island Southeast Asia for too long to have been part of a relatively recent expansion of farmers.

The only other haplogroups which it seems plausible to ascribe to such a migration event are D5 and Y2. The root type of D5 is most common in China and Taiwan and is also found in a few individuals from Island Southeast Asia. However, most Island Southeast Asian samples lie on a branch which is characterised by transitions at np 16148 and np 16092 and which dates to ~4,000 years in Island Southeast Asia (see figure 94). The root type of this branch is not found in Taiwan, but two derived types

are found there suggesting the root type may have been lost due to drift. D5 is not found in the Philippines but once again this may be due to insufficient sampling.

Haplogroup Y2 may illustrate another aspect of a Taiwanese migration event. Bellwood (1997) has suggested that once the Taiwanese immigrants reached the Philippines and Sulawesi, the migration would then have separated into two distinct directions - one east across the Moluccas and into island Melanesia, and the other west into Borneo and Sumatra. The distribution of Y2 fits quite well to this western branch. It is found in Taiwan, the Philippines, Sulawesi, Sumatra, Bali and Java but in none of the eastern groups studied (see figures 60 – 61). It also dates to 3,600 years (see table 7) which fits well with Bellwood's dates for this western branch.

7.2.5 Evidence for a Melanesian Influence in Island Southeast Asia

As discussed in the background to this investigation, the main Melanesian haplogroups are known as P and Q and are thought to represent the indigenous Pleistocene inhabitants of Melanesia (Forster *et al.* 2001). Haplogroup Q has also been found in Polynesia and has been cited as evidence of at least some Melanesian contribution to the genetic make-up of modern Polynesians (Lum *et al.* 1994; Sykes *et al.* 1995).

In this study, both 'Melanesian' haplogroups have also been found in Island Southeast Asia. However, they are both relatively rare, making up ~4% of the sample as a whole. Haplogroup P is particularly rare and has only been found in three individuals from Manado in Sulawesi and two individuals from Sumba (see figure 75). Haplogroup Q, in contrast, makes up 3.1% of the sample as a whole and is found from Ambon to Borneo. It is, however, most common in the easternmost locations studied, and is particularly common in Alor where it makes up 29% of the sample (see figure 97).

This suggests that there has been a definite Melanesian contribution to the populations of Island Southeast Asia. This seems to have been largest in the east, nearest to New Guinea. That the contribution is highest in Alor should not be particularly surprising, as a number of Alorese groups speak Papuan languages, thus suggesting a linguistic connection to the indigenous groups of Melanesia (Pawley 2003; www.ethnologue.com).

It is unclear whether this represents an ancient or recent Melanesian influence in Island Southeast Asia. It is possible that it demonstrates a connection between the peoples of Melanesia and Eastern Island Southeast Asia which dates back to the Pleistocene. However, it has been suggested that the Papuan-speakers in central and eastern Timor, Alor, Pantar, Morotai and northern Halmahera may be the result of migrations within the last ~4,000 years (Pawley 2003). The Melanesian contribution found in this dataset could therefore be the result of such migrations.

7.2.6 Evidence for an Indo-Chinese Influence in Island Southeast Asia

There is also evidence of an Indo-Chinese influence in Island Southeast Asia; this is most obvious in the form of haplogroup F1a1a. As discussed in section 7.1, F1a1a is most diverse in South China and dates to 9,200 years. However, it is most common in Thailand and in the Senoi and is consistent with a Neolithic link between Thailand and the indigenous groups of peninsular Malaysia.

However, F1a1a is also found at lower levels across Island Southeast Asia and makes up 2.6% of the sample as a whole. It is not found in Taiwan, the Philippines or Northern Borneo so there is no evidence of a Taiwanese link between China and Island Southeast Asia (see figure 83). It is most common in Palembang where it makes up 17% of the sample (see figure 79). Palembang was the centre of the Srivijaya kingdom which reached its height between the 7th and 13th Centuries and which controlled the trade routes through the straits of Malacca. It is thought that the Malay languages may have been introduced into the peninsula by such maritime empires in Sumatra (Adelaar 2004). It is therefore tempting to think that the high levels of F1a1a in Palembang could have been introduced from the Malay Peninsula via these trade routes.

It is possible that haplogroup N9a1 could also have been introduced into Island Southeast Asia from Indo-China. The root type of N9a1 is only found in one individual from Palu, one from Thailand and two from Vietnam (however, the latter are only represented by partial HVS-I sequences). N9a1 is most common in the Malay Peninsula where it is found in all three groups of Orang Asli as well as in *Melayu* Malays (see

figure 59). It is rarer than F1a1a, being found in only 1.1% of the Island Southeast Asian sample, but is still most common in Sumatra. Its scarcity makes it difficult to be certain, but an Indo-Chinese origin for N9a1 does seem most likely on the current evidence.

It is also possible that haplogroup B4* may have been introduced to Island Southeast Asia from Indo-China. Almost all examples of B4* found to date in Island Southeast Asia belong to a branch which is characterised by transitions at np 16147 and np 16235. Outside Island Southeast Asia, this branch has only been found in three individuals from the Malay Peninsula, five from Thailand, one from South China, one from Japan and one from Central Asia (see figure 64). The absence of this branch in Taiwan and the Philippines suggests that a Western entry point into Island Southeast Asia is more likely. If F1a1a, N9a1 and B4* are all considered to be Indo-Chinese haplogroups, this means that around 6% of the Island Southeast Asian gene pool is derived from sources in Indo-China.

7.2.7 Evidence for an Orang Asli Influence in Island Southeast Asia

There is some evidence of gene flow from the Orang Asli into Island Southeast Asia. As discussed in section 7.1, the newly discovered haplogroups M21a, M21b and M21c seem to represent an indigenous presence in the Malay Peninsula. Both M21a and M21b are found in Island Southeast Asia, albeit at a very low level. The root type of M21a has been found in one individual from Banjarmasin and M21b has been found in two further individuals from Banjarmasin, one from Medan and one from Manado (see figure 85).

Haplogroup R9b is also found in Island Southeast Asia and the Orang Asli, in this case the Aboriginal Malays (see figure 78). As discussed in section 7.1, it is difficult to clarify the history of this haplogroup due to its relative rarity and the amount of drift which has occurred in the Aboriginal Malays. Its ultimate origins seem to lie in China but from there it is difficult to tell whether it travelled through Thailand and the Malay Peninsula into Island Southeast Asia, or whether it was passed to the Aboriginal Malays

from Sumatra. However, there is definitely no evidence of it being present in Taiwan , the Philippines or anywhere east of Lombok.

The origins of the novel haplogroup N22 are also obscure. It is very rare, being found in only 8 individuals: 4 from Sumba and 4 Aboriginal Malays (all Temuan). The root type is only found in two Aboriginal Malays, one branch is only found in two more Aboriginal Malays while a second branch is found only in Sumba (see figure 57). It is impossible to discern the direction of movement from the present data; however, there is no evidence of a Taiwanese origin.

7.2.8 Evidence for Indian and European Influence in Island Southeast Asia

Despite the long period of European influence in Island Southeast Asia, only one example of a European haplogroup has been found in this investigation. Specifically, one individual from Pekanbaru in Sumatra was found to belong to haplogroup I. This haplogroup is predominantly found in Northern Europe. The lack of any other evidence for European admixture suggests that the long period of colonialism did not substantially alter the indigenous gene pool.

There is also some evidence of a small genetic contribution from the Indian subcontinent to Island Southeast Asia. Another individual from Pekanbaru belonged to haplogroup U7 which is mainly found in India and the Middle East. The specific haplotype seen in Pekanbaru matches one found in the Andhra Pradesh region of Eastern India. There are also another four possible Indian haplotypes in Sumatra: one belonging to haplogroup M2, two belonging to haplogroup M3 and one belonging to haplogroup M5. This is the first time these haplogroups have been found outside the Indian subcontinent. There is archaeological evidence of Indian trade contact with Island Southeast Asia from ~2,000 BP (Ardika and Bellwood 1991) and there is also evidence from ancient DNA of Indian traders being present in Bali at around the same time (Lansing *et al.* 2004). No Indian haplogroups have been found in the Balinese sample included in this study, but these rare haplotypes found in Sumatra are probably a result of such contact.

7.2.9 Conclusions

One of the clearest results of this investigation is that the history of Island Southeast Asia is much too complex to be explained by any simple model. The sheer amount of variation present suggests that a simple migration and replacement model is far too crude to explain the data. However, the data do suggest that some migratory events have taken place, but not from only one direction, and that they have simply added more variation to that already present.

One of the most striking things about the dataset is the fact that the Aboriginal Taiwanese seem to have much more in common with individuals from Island Southeast Asia rather than China, as can be seen in chapter 6. There is little evidence of a link from China through Taiwan to Island Southeast Asia as would be predicted by the 'out of Taiwan' model of Bellwood (1997). The only haplogroups which seem to fit in any way with this model are D5, Y2 and possibly M7c1c.

As discussed in sections 7.2.2 and 7.2.4, most of the other haplogroups which have any links to Taiwan seem to be too old to be linked to a traditional Neolithic 'out of South China via Taiwan' event. Most of the haplogroups which are predominantly, if not exclusively, found in Austronesian speakers (such as M7c1c, E1, F3b and M7b3) are most diverse in Borneo and Taiwan suggesting an origin somewhere in that region. Of these, E1 definitely does not seem to have any recent roots in China, emphasising the level of separation between China and Taiwan. B4a1 also seems to have originated in Island Southeast Asia, albeit further towards the east, and again seems to be too old to be explained via the conventional 'out of Taiwan' hypothesis. Other indigenous haplogroups such as N21 and the rare branches of R, plus possibly R22, R9b, and N22 have also persisted into modern times. Indigenous Y chromosome haplogroups were also found to be common in Island Southeast Asia by Su *et al.* (2000) and Kayser *et al.* (2000).

There is some evidence of a possible migration from Taiwan which could be represented by haplogroups M7c1c, D5 and Y2; however, this only accounts for ~13%

of the current dataset. This is somewhat similar to the results found for the Y chromosome by Capelli *et al.* (2001) who found that only ~19.5% of their Island Southeast Asian sample could be accounted for by potential Taiwanese haplogroups (haplogroup O3). However, Kayser *et al.* (2003) suggest that haplogroup O1 is also 'Austronesian' which would increase the frequency of potential Taiwanese haplogroups. Nevertheless, my results suggest that if any Neolithic migration did occur, it would seem that it was demographically minor and that the immigrants integrated into the resident population rather than replacing it. Influences from Thailand, the Malay Peninsula and Melanesia have also been found in this investigation, all of which decrease further the likelihood that a simple 'out of Taiwan' migration can be used to explain the prehistory of Island Southeast Asia.

Instead the current dataset seems to support the work of Solheim *et al.* (in press; personal communication) who have proposed greater connections between Island Southeast Asia and Taiwan than between Taiwan and China. The results may also suggest large-scale dispersals after flooding events which would be associated with the end of the ice age in Southeast Asia, this has previously been suggested by Oppenheimer (1998). The dataset also supports the idea of a 'voyaging corridor' within Island Southeast Asia (Terrell and Welsch 1987; Irwin 1992) which would have allowed more contact between the islands than is often allowed for. This may be seen in the distribution of haplogroup B4a1 which extends west to Sulawesi and Borneo, east towards New Guinea and the Pacific, and also across to Madagascar. A further implication of this is that it would tend to suggest that the Austronesian language family more likely arose outside China; possibly in the area around North Borneo, the Philippines and Taiwan. This is perhaps in conflict with the prevailing wisdom but has also been suggested by Solheim *et al.* (in press; personal communication) and Meacham (1984-1985), both of whom proposed that the high levels of linguistic diversity found in Taiwan could have been caused by isolation rather than the languages developing there.

However, it is possible that the Austronesian languages could have been spread from Taiwan by a small elite. This would explain the fact that the most diverse branches of Austronesian are only found in Taiwan and also the lack of an obvious major Neolithic

contribution to the genetic makeup of Island Southeast Asia. If this did occur then there must have been a strong mechanism operating in favour of language replacement as the amount of variation found in Island Southeast Asia, most of which seems to date to the Pleistocene, suggests that most of the modern inhabitants are descended from people who had been living in the area long before the occurrence of any potential 'farming' migration.

7.3 Implications of this Study for the Origins of Modern Humans

The complete sequencing work which was done as an extension to this project enabled the most recent common ancestor of modern humans to be dated to ~200,000 years. This is similar to the estimate of Mishmar *et al.* (2003) but somewhat older than that of Ingman *et al.* (2000) and provides further support to the 'out of Africa' theory as opposed to that of multiregionalism.

The route which these early humans took out of Africa has been a matter of some debate. Certain groups have suggested a relatively late expansion (~45,000 years) via the Levant (Cordaux and Stoneking 2003; Prugnolle *et al.* 2005) while other groups have suggested an earlier (~60,000 years), southern route through East Africa and along the coast towards Southeast Asia (Quintana-Murci *et al.* 1999) and yet other groups have suggested two separate out of Africa events: one early southern route which occurred 59,000-69,000 years ago and a second later route through the Levant at around 39,000-52,000 years ago (Maca-Meyer *et al.* 2001). The answer to this question can be clarified by the work done on this project.

Most of the mismatch distributions constructed for the current dataset gave expansion dates of ~60,000 years. This was roughly in agreement with the dates obtained from complete sequencing for macrohaplogroups M (~63,000 years), N (also ~63,000 years) and R (~60,000 years). These dates all suggest an early out of Africa event which must have occurred sometime prior to ~60,000 BP

As discussed both above and in the background section, there is much evidence (both archaeological and now genetic) that at least the Semang, if not all Orang Asli groups in

part, are descended from the original Pleistocene inhabitants of the Malay Peninsula. A number of novel haplogroups have been discovered in the Orang Asli; however, these diverge from the same set of founder types (M, N and R) as the rest of Eurasia. These new haplogroups all diverge from close to the roots of these macrohaplogroups which suggests they are of considerable antiquity. This is at least partly confirmed by the dating of M21 to ~57,000 years by complete sequencing.

The fact that the Orang Asli groups descend from the same few founder types as all other non-African populations studied to date strongly suggests that only a single 'out of Africa' event took place. That M, N and R all date to ~60,000 years implies that they were all part of this single process which took place at some point around that time. Furthermore, both the position of the newly discovered Orang Asli haplogroups so close to the root of the tree, and the results of a founder analysis on Eurasian and Australasian mtDNAs which gave arrival dates of ~66,000 years in India, ~64,500 years in China and ~63,000 years in Australasia (Macaulay *et al.* 2005) confirms that the out of Africa dispersal must have continued extremely rapidly along a southern, coastal route to Southeast Asia.

7.4 Future Work

This study has filled in many of the gaps which were previously present in the mtDNA coverage of Southeast Asia; however, a number of areas still remain to be explored. For example, the only samples from Borneo which were available for this study were from individuals living in two of the large coastal cities (Kota Kinabalu in Sabah and Banjarmasin in Kalimantan). Therefore, the whole of interior Borneo has not yet been studied. This could prove to be an extremely fruitful area for study as a large amount of variation has been found in the two locations which were included in the current dataset. If more locations could be included in any further work then they could help identify any prehistoric migrations which may have been centred on Borneo.

It would also be useful to study more individuals from various locations within the Philippines. This area is crucial for examining the 'out of Taiwan' hypothesis; however, for this study less than 50 samples were available from the whole country. It could also

be interesting to study more individuals from Java. Only 36 Tenggerese were included in this study and they are probably not particularly representative of the island as a whole. It might be helpful to be able to compare the mtDNAs of these individuals to some obtained from the ethnic Javanese who make up the majority of the island's population.

To increase the accuracy of the distribution and age estimates of certain haplogroups, in particular B4a1, it may also be worthwhile to study more locations in Eastern Indonesia and coastal Papua New Guinea. Despite the extensive sampling of Island Southeast Asia in this study, it is still unclear where haplogroup B4a1 originated, this may be clarified by increased sampling of Eastern Indonesia. The island of Flores in Eastern Indonesia may be especially interesting in light of the discovery of a new hominin species, *Homo floresiensis*, which has been made there recently (Brown *et al.* 2004; Morwood *et al.* 2004). It may also be particularly constructive to study the island of Halmahera in the Moloccas as it has been suggested from studies of both rat and pig mtDNA that it was in that area that the Lapita culture originated (Matisoo-Smith and Robins 2004; Larson *et al.* 2005).

It would also be helpful to employ complete mtDNA sequencing to clarify some of the issues raised by this study. For example, due to ambiguity in the HVS-I network, it is currently unclear whether haplogroup R9b entered Island Southeast Asia from the Malay Peninsula or whether it was introduced into the Malay Peninsula from Sumatra. This could be clarified by complete sequencing. Complete sequencing may also help refine the network of haplogroup R22 and show how it relates to haplogroup R9b. Complete sequencing is also needed to clarify the position of the G* types found in this study.

It would also be valuable to sequence the complete mtDNA genome of some of the M*, N* and R* types found in this study. This would be particularly helpful in the case of the M* types as they make up a relatively large portion of the dataset and it is currently difficult to see how they relate to each other due to the extremely high levels of reticulation seen within the network.

Finally, complete sequencing could also help to refine some of the dating done in this study. Increased precision could be particularly useful for haplogroups B4a1, M7c1c, E1, F3b, D5 and Y2 as it could help to further test the traditional models of Island Southeast Asian ancestry and to work out what percentage, if any, of the modern populations of Island Southeast Asia can be described as being descendents of a putative Neolithic 'out of Taiwan' migration.

References

References

Adelaar KA (2004) Where does Malay come from? Twenty years of discussions about homeland, migrations and classifications. *Bijdragen tot de Taal-, Land-en Volkenkunde* **160**:1-30

Allen J, Gosden C, Jones R, White JP (1988) Pleistocene dates for the human occupation of New Ireland, North Melanesia. *Nature* **331**:707-709

Allen J (1996) The Pre-Austronesian settlement of Island Melanesia: implications for Lapita archaeology. In: Prehistoric Settlement of the Pacific (ed. WH Goodenough) pp 11-27. American Philosophical Society, Philadelphia.

Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG (1981) Sequence and organization of the human mitochondrial genome. *Nature* **290**:457-465

Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nature Genetics* **23**:147

Ankel-Simons F, Cummins JM (1996) Misconceptions about mitochondria and mammalian fertilization: implications for theories on human evolution. *Proceedings of the National Academy of Sciences USA* **93**:13859-13863

Ardika IW, Bellwood P (1991) Sembiran: The beginnings of Indian contact with Bali. *Antiquity* **65**:221-232

Ayala FJ (1995) The myth of Eve: molecular biology and human origins. *Science* **270**:1930-1936

Ballinger SW, Schurr TG, Torroni A, Gan YY, Hodge JA, Hassan K, Chen K-H, Wallace DC (1992) Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient mongoloid migrations. *Genetics* **130**:139-152

Bandelt H-J, Forster P, Sykes BC, Richards M (1995) Mitochondrial portraits of human populations using median networks. *Genetics* **141**:743-753

Bandelt H-J, Lahermo P, Richards M, Macaulay V (2001) Detecting errors in mtDNA data by phylogenetic analysis. *International Journal of Legal Medicine* **115**:64-69

Bandelt H-J, Quintana-Murci L, Salas A, Macaulay V (2002) The fingerprint of phantom mutations in mitochondrial DNA data. *American Journal of Human Genetics* **71**:1150–1160

Barker G, Barton H, Beavitt P, Bird M, Daly P, Doherty C, Gilbertson D, Hunt C, Krigbaum J, Lewis H, Manser J, McLaren S, Paz V, Piper P, Pyatt B, Rabett R, Reynolds T, Rose J, Rushworth G, Stephens M (2002) Prehistoric foragers and farmers in South-East Asia: renewed investigations at Niah Cave, Sarawak. *Proceedings of the Prehistoric Society* **68**:147-164

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 Hawai'i Press, Honolulu

Bellwood P (2003) Farmers, foragers, languages, genes: the genesis of agricultural societies. In: Examining the farming/language dispersal hypothesis (eds. P Bellwood, C Renfrew) pp 17-28. McDonald Institute for Archaeological Research, Cambridge

Bellwood P (2005) First farmers, the origins of agricultural societies. Blackwell Publishing, Oxford

Benedict G (1975) Austro-Thai language and culture. HRAF press, New Haven

Benjamin G (1985) In the long term: Three themes in Malaysian cultural ecology. In: Cultural values and human ecology in Southeast Asia (eds. KL Hutterer, AT Rambo and G Lovelace). Michigan University Press, Ann Arbor, Michigan

Benjamin G (1996) Issues in the ethnology of Padang. In: Pembangunan Arkeologi Pelancongan Negeri Pahang (eds. A Rahman and AH Khairuddin). Perbadanan Muzium Negeri Kelantan, Kota Bharu

Betty DJ, Chin-Atkins AN, Croft L, Straml M, Esteal S (1996) Multiple independent origins of the COII/tRNALys intergenic 9-bp mtDNA deletion in aboriginal Australians. *American Journal of Human Genetics* **58**:428-433

Blust R (1995) The prehistory of the Austronesian-speaking peoples: a view from language. *Journal of World Prehistory* **9**:453-510

Blust R (1996a) Austronesian culture history: the window of language. In: Prehistoric Settlement of the Pacific (ed. WH Goodenough) pp 28-35. American Philosophical Society, Philadelphia.

Blust R (1996b) Beyond the Austronesian homeland: The Austric hypothesis and its implications for archaeology. In: Prehistoric Settlement of the Pacific (ed. WH Goodenough) pp 117-140. American Philosophical Society, Philadelphia.

Blust R (1999) Subgrouping, circularity and extinction: some issues in Austronesian comparative linguistics. In: Selected papers from the Eighth International Conference on Austronesian linguistics (eds. E Zeitoun and PJ Li). Academia Sinica, Taipei, Taiwan

Bowler JM, Johnston H, Olley JM, Prescott JR, Roberts RG, Shawcross W, Spooner NA (2003) New ages for human occupation and climatic change at Lake Mungo, Australia. *Nature* **421**:837-840

Brehm A, Pereira L, Bandelt H-J, Prata MJ, Amorim A (2002) Mitochondrial portrait of the Cabo Verde archipelago: the Senegambian outpost of Atlantic slave trade. *Annals of Human Genetics* **60**:49-60

Brown P, Sutikna T, Morwood MJ, Soejono RP, Jatmiko, Saptomo EW, Due RA (2004) A new small-bodied hominin from the Late Pleistocene of Flores, Indonesia. *Nature* **431**:1055-1061

Brown WM, George M, Wilson AC (1979) Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences USA* **76**:1967-1971

Brown WM (1980) Polymorphism in mitochondrial DNA of humans as revealed by restriction endonuclease analysis. *Proceedings of the National Academy of Sciences USA* **77**:3605-3609

Bulbeck D (1996) Holocene biological evolution of the Malay Peninsula aborigines (Orang Asli). *Perspectives in Human Biology* **2**:37-61

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

Bulbeck D, Pasqua M, Di Lello A (2001) Culture history of the Toalean of South Sulawesi, Indonesia. *Asian Perspectives* **39**:71-108

Bulbeck D (2004) Indigenous traditions and exogenous influences in the early history of Peninsular Malaysia. In: Southeast Asia from prehistory to history (eds I Glover and P Bellwood). Routledge Curzon, London and New York.

Cann RL, Stoneking M, Wilson AC (1987) Mitochondrial DNA and human evolution. *Nature* **325**:31-36

Capelli C, Wilson JF, Richards M, Stumpf MPH, Gratrix F, Oppenheimer S, Underhill P, Pascali VL, Ko T-M, Goldstein DB (2001) A predominantly indigenous paternal heritage for the Austronesian-speaking peoples of insular Southeast Asia and Oceania. *American Journal of Human Genetics* **68**:432-443

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

Chen YS, Torroni A, Excoffier L, Santachiara-Benerecetti AS, Wallace DC (1995) Analysis of mtDNA variation in Africa reveals the most ancient of all human continent-specific haplogroups. *American Journal of Human Genetics* **57**:133-149

Chromas website:
http://www.technelysium.com.au/chromas.html
Accessed 07/07/04

CIA "The World Factbook" website:
http://www.cia.gov/cia/publications/factbook/docs/refmaps.html
Accessed 17/10/04

Clayton DA (1982) Replication of animal mitochondrial DNA. *Cell* **28**:693-705

Comas D, Calafell F, Mateu E, Pérez-Lezaun A, Bosch E, Martínez-Arias R, Clarimon J, Facchini F, Fiori G, Luiselli D, Pettener D, Bertranpetit J (1998) Trading genes along the Silk Road: mtDNA sequences and the origin of central Asian populations. *American Journal of Human Genetics* **63**:1824-1838

Cordaux R, Stoneking M (2003) South Asia, the Andamanese, and the genetic evidence for an "early" human dispersal out of Africa. *American Journal of Human Genetics* **72**:1586-1590

Diamond JM (1988) Express train to Polynesia. *Nature* **336**:307-308

Ding YC, Wooding S, Harpending HC, Chi H-C, Li H-P, Fu Y-X, Pang J-F, Yao Y-G, Xiang Yu J-G, Moyzis R, Zhang Y-P (2000) Population structure and history in East Asia. *Proceedings of the National Academy of Sciences USA* **97**:14003-14006

Di Rienzo A, Wilson AC (1991) Branching pattern in the evolutionary tree for human mitochondrial DNA. *Proceedings of the National Academy of Sciences USA* **88**:1597-1601

DnaSP website:
http://www.ub.es/dnasp/
Accessed 03/11/04

Dyen I (1962) The lexicostatistical classification of the Malayopolynesian languages. *Language* **38**:38-46

Endicott P, Gilbert TMP, Stringer C, Lalueza-Fox C, Willerslev E, Hansen AJ, Cooper A (2003) The genetic origins of the Andaman islanders. *American Journal of Human Genetics* **72**:178–184

Ethnologue website:
http://www.ethnologue.com
Accessed 20/03/05

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**:479-491

Finnilä S, Lehtonen MS, Majamaa K (2001) Phylogenetic network for European mtDNA. *American Journal of Human Genetics* **68**:1475-1484

Fix AG (1995) Malayan paleosociology. Implications for the patterns of genetic variation among the Orang Asli. *American Anthropologist* **97**:313-323

Fix AG (2000) Genes, language, and ethnic groups: reconstructing Orang Asli prehistory. *Indo-Pacific Prehistory Association Bulletin* **19**:11-16

Foley WA (1986) The Papuan languages of New Guinea. Cambridge University Press, Cambridge

Forster P, Harding R, Torroni A, Bandelt H-J (1996) Origin and evolution of Native American mtDNA variation: a reappraisal. *American Journal of Human Genetics* **59**:935-945

Forster P, Torroni A, Renfrew C, Röhl A (2001) Phylogenetic star contraction applied to Asian and Papuan mtDNA evolution. *Molecular Biology and Evolution* **18**:1864-1881

Fucharoen G, Fucharoen S, Horai S (2001) Mitochondrial DNA polymorphisms in Thailand. *Journal of Human Genetics* **46**:115-125

Fuku N, Oshida Y, Takeyasu T, Guo L–J, Kurata M, Yamada Y, Sato Y, Tanaka M (2002) Mitochondrial ATPase subunit 6 and cytochrome *b* gene polymorphisms in young obese adults. *Biochemical and Biophysical Research Communications* **290**:1199–1205

Gajra B, Candlish JK, Heng CK, Mak, JW, Saha N (1997) Genotype associations among seven apolipoprotein B polymorphisms in a population of Orang Asli of western Malaysia. *Human Biology* **69**:629-640

Gray RD, Jordan FM (2000) Language trees support the express-train sequence of Austronesian expansion. *Nature* **405**:1052-1055

Green RC (1991) The Lapita cultural complex: current evidence and proposed models. *Bulletin of the Indo-Pacific Prehistory Association* **11**:295-305

Greenhill SJ, Gray RD (In Press) Testing population dispersal hypotheses: Pacific settlement, phylogenetic trees and Austronesian languages. In: The Evolution of Cultural Diversity: Phylogenetic Approaches (eds. R Mace, C Holden, S Shennan) UCL Press

Groube L, Chappell J, Muke J, Price D (1986) A 40,000 year old human occupation site at Huon peninsula, Papua New Guinea. *Nature* **324**:453-455

Groves CP (1981) Ancestors for the pigs: taxonomy and phylogeny of the genus *Sus*. Technical Bulletin 3, Department of Prehistory, Australian National University, Canberra

Gyllensten U, Wharton D, Josefsson A, Wilson AC (1991) Paternal inheritance of mitochondrial DNA in mice. *Nature* **352**:255-257

Hagelberg E, Clegg JB (1993) Genetic polymorphisms in prehistoric Pacific islanders determined by analysis of ancient bone DNA. *Proceedings of the Royal Society of London Series B* **252**:163-170

Hagelberg E, Quevedo S, Turbon D, Clegg JB (1994) DNA from ancient Easter Islanders. *Nature* **369**:25-26

Hagelberg E, Goldman N, Lió P, Whelan S, Schiefenhövel W, Clegg JB, Bowden DK (1999) Evidence for mitochondrial DNA recombination in a human population of island Melanesia. *Proceedings of the Royal Society of London Series B* **266**:485-492

Hagelberg E, Goldman N, Lió P, Whelan S, Schiefenhövel W, Clegg JB, Bowden DK (2000) Evidence for mitochondrial DNA recombination in a human population of island Melanesia: correction. *Proceedings of the Royal Society of London Series B* **267**:1595-1596

Handt O, Meyer S, von Haeseler A (1998) Compilation of human mtDNA control region sequences. *Nucleic Acids Research* **26**:126–129

Harihara S, Mirai M, Suutou Y, Shimizu K, Omoto K (1992) Frequency of a 9-bp deletion in the mitochondrial DNA among Asian populations. *Human Biology* **64**:161-166

Herrnstadt C, Elson JL, Fahy E, Preston G, Turnbull DM, Anderson C, Ghosh SS, Olefsky JM, Beal MF, Davis RE, Howell N (2002) Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences of the major African, Asian and European haplogroups. *American Journal of Human Genetics* **70**:1152-1171

Hertzberg M, Mickleson KNP, Serjeantson SW, Prior JF, Trent RJ (1989) An Asian-specific 9-bp deletion of mitochondrial DNA is frequently found in Polynesians. *American Journal of Human Genetics* **44**:504-510

Heyerdahl T (1950) Kontiki: across the Pacific by raft. Rand McNally, Chicago

Hill AVS, Bowden DK, Trent RJ, Higgs DR, Oppenheimer SJ, Thein SL, Mickleson KNP, Weatherall DJ, Clegg JB (1985) Melanesians and Polynesians share a unique α-thalassemia mutation. *American Journal of Human Genetics* **37**:571-580

Horai S, Hayasaka K (1990) Intraspecific nucleotide sequence differences in the major noncoding region of human mitochondrial DNA. *American Journal of Human Genetics* **46**:828-842

Horai S, Hayasaka K, Kondo R, Tsugane K, Takahata N (1995) Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. *Proceedings of the National Academy of Sciences USA* **92**:532–536

Horai S, Murayama K, Hayasaka K, Matsubayashi S, Hattori Y, Fucharoen G, Harihara S, Park KS, Omoto K, Pan I-H (1996) mtDNA polymorphism in East Asian populations, with special reference to the peopling of Japan. *American Journal of Human Genetics* **59**:579-590

Ingman M, Kaessmann H, Pääbo S, Gyllensten U (2000) Mitochondrial genome variation and the origin of modern humans. *Nature* **408**:708-713

Irwin G (1992) The prehistoric expansion and colonisation of the Pacific. Cambridge University Press, Cambridge

Kayser M, Brauer S, Weiss G, Underhill PA, Roewer L, Schiefenhövel W, Stoneking M (2000) Melanesian origin of Polynesian Y chromosomes. *Current Biology* **10**:1237-1246

Kayser M, Brauer S, Weiss G, Schiefenhövel W, Underhill P, Shen P, Oefner P, Tommaseo-Ponzetta M, Stoneking M (2003) Reduced Y-chromosome, but not mitochondrial DNA, diversity in Human Populations from West New Guinea. *American Journal of Human Genetics* **72**:281-302

Kivisild T, Bamshad MJ, Kaldma K, Metspalu M, Metspalu E, Reidla M, Laos S, Parik J, Watkins WS, Dixon ME, Papiha SS, Mastana SS, Mir MR, Ferak V, Villems R (1999) Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages. *Current Biology* **9**:1331-1334

Kivisild T, Tolk H-V, Parik J, Wang Y, Papiha SS, Bandelt H-J, Villems R (2002) The emerging limbs and twigs of the East Asian mtDNA tree. *Molecular Biology and Evolution* **19**:1737-1751

Kivisild T, Reidla M, Metspalu E, Rosa A, Brehm A, Pennarun E, Parik J, Geberhiwot T, Usanga E, Villems R (2004) Ethiopian mitochondrial DNA heritage: tracking gene flow across and around the Gate of Tears. *American Journal of Human Genetics* **75**:752-770

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

Kolman CJ, Tuross N (2000) Ancient DNA analysis of human populations. *American Journal of Physical Anthropology* **111**:5-23

Kong Q-P, Yao Y-G, Sun C, Bandelt H-J, Zhu C-L, Zhang Y-P (2003) Phylogeny of East Asian mitochondrial DNA lineages inferred from complete sequences. *American Journal of Human Genetics* **73**:671-676

Krings M, Stone A, Schmitz RW, Krainitzki H, Stoneking M, Pääbo S (1997) Neandertal DNA sequences and the origin of modern humans. *Cell* **90**:19-30

Krings M, Geisert H, Schmitz RW, Krainitzki H, Pääbo S (1999a) DNA sequence of the mitochondrial hypervariable region II from the Neandertal type specimen. *Proceedings of the National Academy of Sciences USA* **96**:5581-5585

Krings M, Salem AH, Bauer K, Geisert H, Malek AK, Chaix L, Simon C, Welsby D, Di Rienzo A, Utermann G, Sajantila A, Pääbo S, Stoneking M (1999b) mtDNA Analysis of Nile River Valley Populations: A Genetic Corridor or a Barrier to Migration? *American Journal of Human Genetics* **64**:1166-1176

Lansing JS, Redd AJ, Karafet TM, Watkins J, Ardika IW, Surata SPK, Schoenfelder JS, Campbell M, Merriwether AM, Hammer MF (2004) An Indian trader in ancient Bali? *Antiquity* **78**:287-293

Larson G, Dobney K, Albarella U, Fang M, Matisoo-Smith E, Robins J, Lowden S, Finlayson H, Brand T, Willerslev E, Rowley-Conwy P, Andersson L, Cooper A (2005) Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science* **307**:1618-1621

Lee SD, Shin CH, Kim KB, Lee YS, Lee JB (1997) Sequence variation of mitochondrial DNA control region in Koreans. *Forensic Science International* **87**:99-116

LeBar FM (1972) Ethnic groups of insular Southeast Asia Volume 1: Indonesia, Andaman Islands and Madagascar. Human Relations Area Files Press, New Haven.

Lum JK, Rickards O, Ching C and Cann RL (1994) Polynesian mitochondrial DNAs reveal three deep lineage clusters. *Human Biology* **66**:567-590.

Lum JK, Cann RL, Martinson JJ, Jorde LB (1998) Mitochondrial and nuclear genetics relationships among Pacific island and Asian populations. *American Journal of Human Genetics* **63**:613-624

Lum JK, Cann RL (2000) mtDNA lineage analyses: origins and migrations of Micronesians and Polynesians. *American Journal of Physical Anthropology* **113**:151-168

Maca–Meyer N, González AM, Larruga JM, Flores C, Cabrera VM (2001) Major genomic mitochondrial lineages delineate early human expansions. *BMC Genetics* **2**:13

Macaulay V, Richards M, Hickey E, Vega E, Cruciani F, Guida V, Scozzari R, Bonné-Tamir B, Sykes B, Torroni A (1999) The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. *American Journal of Human Genetics* **64**:232-249

Macaulay V, Hill C, Achilli A, Rengo C, Clarke D, Meehan W, Blackburn J, Semino O, Scozzari R, Cruciani F, Taha A, Shaari NK, Raja JM, Ismail P, Zainuddin Z, Goodwin W, Bulbeck D, Bandelt H-J. Oppenheimer S, Torroni A, Richards M (2005) Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science* **308**:1034-1036

Margulis L (1981) Symbiosis in cell evolution: life and its environment on the early Earth. W.H. Freeman & Co., New York

Matisoo-Smith E, Robins JH (2004) Origins and dispersals of Pacific peoples: evidence from mtDNA phylogenies of the Pacific rat. *Proceedings of the National Academy of Sciences USA* **101**:9167-9172

Meacham W (1984-1985) On the improbability of Austronesian origins in South China. *Asian Perspectives* **26**:89-105

Melton T, Peterson R, Redd AJ, Saha N, Sofro ASM, Martinson J, Stoneking M (1995) Polynesian genetic affinities with Southeast Asian populations as identified by mtDNA analysis. *American Journal of Human Genetics* **57**:403-414

Melton T, Clifford S, Martinson J, Batzer M, Stoneking M (1998) Genetic evidence for the proto-Austronesian homeland in Asia: mtDNA and nuclear DNA variation in Taiwanese aboriginal tribes. *American Journal of Human Genetics* **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, Wallace DC (2003) Natural selection shaped regional mtDNA variation in humans. *Proceedings of the National Academy of Sciences USA* **100**:171–176

Morwood MJ, Soejono RP, Roberts RG, Sutikna T, Turnet CSM, Westaway KE, Rink WJ, Zhao J-X, van den Bergh GD, Due RA, Hobbs DR, Moore MW, Bird MI, Fifield LK (2004) Archaeology and age of a new hominin from Flores in eastern Indonesia. *Nature* **431**:1087-1091

Mountain JL, Herbert JM, Bhattacharyya S, Underhill PA, Ottolenghi C, Gadgil M, Cavalli-Sforza LL (1995) Demographic history of India and mtDNA-sequence diversity. *American Journal of Human Genetics* **56**:979-992

Network website:
http://www.fluxus-engineering.com/sharenet.htm
Accessed 08/01/05

Nichols J (1998) The origin and dispersal of languages: linguistic evidence. In: The Origin and Diversification of Language (eds. NG Jablonski and LC Aiello) pp 127-170. Memoirs of the California Academy of Sciences 24

Nishimaki Y, Sato K, Fang L, Ma M, Hasekura H, Boettcher B (1999) Sequence polymorphism in the HV1 region in Japanese and Chinese. *Legal Medicine* **1**:238-249

Old JM, Clegg JB, Weatherall DJ, Booth PB (1978) Haemoglobin J Tongariki is associated with α thalassemia. *Nature* **273**:319-320

Oota H, Saitou N, Matsushita T, Ueda S (1995) A genetic study of 2,000-year-old human remains from Japan using mitochondrial DNA sequences. *American Journal of Physical Anthropology* **98**:133-145

Oota H, Kurosaki K, Pookajorn S, Ishida T, Ueda S (2001) Genetic study of the Palaeolithic and Neolithic Southeast Asians. *Human Biology* **73**:225-231

Oppenheimer S (1998) Eden in the east: the drowned continent of Southeast Asia. Weidenfeld & Nicolson, London

Ovchinnikov IV, Götherstöm A, Romanova GP, Kharitonov VM, Lidén K, Goodwin W (2000) Molecular analysis of Neanderthal DNA from the northern Caucasus. *Nature* **404**:490-493

Parson W, Parsons TJ, Scheithauer R, Holland MM (1998) Population data for 101 Austrian Caucasian mitochondrial DNA d-loop sequences: Application of mtDNA sequence analysis to a forensic case. *International Journal of Legal Medicine* **111**:124-132

Pawley A (2003) The Austronesian dispersal: languages, technology and people. In: Examining the farming/language dispersal hypothesis (eds. P Bellwood and C Renfrew) pp 251-273. McDonald Institute for Archaeological Research, Cambridge,

Paz V (2003) Island Southeast Asia: spread or friction zone? In: Examining the farming/language dispersal hypothesis (eds. P Bellwood and C Renfrew) pp 275-285. McDonald Institute for Archaeological Research, Cambridge

Penny D, Steel M, Waddell PJ, Hendy MD (1995) Improved analyses of human mtDNA sequences support a resent African origin for *Homo Sapiens. Molecular Biology and Evolution* **12**:863-882

Pfeiffer H, Steighner R, Fisher R, Mörnstad H, Yoon CL, Holland MM (1998) Mitochondrial DNA extraction and typing from isolated dentin - experimental evaluation in a Korean population. *International Journal of Legal Medicine* **111**:309-313

Pietrusewsky M (1997) The people of Ban Chiang: an early Bronze Age site in Northeast Thailand. *Indo-Pacific Prehistory Association Bulletin* **16**, The Chiang Mai papers: 119-147

Ponce de León MS, Zollikofer CPE (2001) Neanderthal cranial ontogeny and its implications for late hominid diversity. *Nature* **412**:534-538

Prasad BVR, Ricker CE, Watkins WS, Dixon ME, Rao BB, Naidu JM, Jorde LB, Bamshad M (2001) Mitochondrial DNA variation in Nicobarese Islanders. *Human Biology* **73**:715-725

Prugnolle F, Manica A, Balloux F (2005) Geography predicts neutral genetic diversity of human populations. *Current Biology* **15**:R159-R160

Qian YP, Chu Z-T, Dai Q, Wei C-D, Chu YJ, Tajima A, Horai S (2001) Mitochondrial DNA polymorphism in Yunnan nationalities in China. *Journal of Human Genetics* **46**:211-220

Quintana-Murci L, Semino O, Bandelt H-J, Passarino G, McElreavey K, Santachiara-Benerecetti AS (1999) Genetic evidence for an early exit of *Homo sapiens sapiens* from Africa through eastern Africa. *Nature Genetics* **23**:437-441

Rambo AT (1988) Why are the Semang? In: Ethnic diversity and the control of natural resources in Southeast Asia (eds. AT Rambo, K Gillogly and KL Hutterer) pp19-35. Michigan papers on South and Southeast Asia, Ann Arbor

Rayner D, Bulbeck D (2001) Dental morphology of the "Orang Asli" aborigines of the Malay Peninsula. In: Causes and effects of human variation (ed. M Henneberg). Australasian Society for Human Biology: Adelaide.

Redd AJ, Takezaki N, Sherry ST, McGarvey ST, Sofro ASM, Stoneking M (1995) Evolutionary history of the COII/tRNA(Lys) intergenic 9-base-pair deletion in human mitochondrial DNAs from the Pacific. *Molecular Biology and Evolution* **12**:604-615

Renfrew C (1987) Archaeology and Language: The Puzzle of Indo-European Origins. Cambridge University Press, Cambridge.

Richards M, Corte-Real H, Forster P, Macaulay V, Wilkinson-Herbots H, Demaine A, Papiha S, Hedges R, Bandelt HJ, Sykes B (1996) Paleolithic and Neolithic lineages in the European mitochondrial gene pool. *American Journal of Human Genetics* **59**:185-203

Richards M, Oppenheimer S, Sykes B (1998) mtDNA suggests Polynesian origins in Eastern Indonesia. *American Journal of Human Genetics* **63**:1234-1236

Rogers AR, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution* **9**:552-569

Ruhlen M (1987) A guide to the world's languages. Volume 1: Classification. Edward Arnold, London

Saha N, Mak JW, Tay JSH, Liu Y, Tan JAMA, Low PS, Singh M (1995) Population genetic study among the Orang Asli (Semai Senoi) of Malaysia: Malayan aborigines. *Human Biology* **67**:37-57

Saillard J, Forster P, Lynnerup N, Bandelt H-J, Nørby S (2000) mtDNA variation among Greenland Eskimos: the edge of the Beringian Expansion. *American Journal of Human Genetics* **67**:718-726

Santos SE, Ribeiro-Dos-Santos AK, Meyer D, Zago MA (1996) Multiple founder haplotypes of mitochondrial DNA in Amerindians revealed by RFLP and sequencing. *Annals of Human Genetics* **60**:305-319

Schneider, S.,Roessli, D., and Excoffier, L. (2000) Arlequin: A software for population genetics data analysis. Ver 2.000. Genetics and Biometry Lab, Dept. of Anthropology, University of Geneva.

Schurr TG, Ballinger SW, Gan YY, Hodge JA, Merriwether DA, Lawrence DN, Knowler WC, Weiss KM, Wallace DC (1990) Amerindian mitochondrial DNAs have rare Asian mutations at high frequencies, suggesting they are derived from four primary maternal lineages. *American Journal of Human Genetics* **46**:613-623

Schurr TG, Sukernik RI, Starikovskaya YB, Wallace DC (1999) Mitochondrial DNA variation in Koryaks and Itel'men: population replacement in the Okhotsk Sea – Bering Sea region during the Neolithic. *American Journal of Physical Anthropology* **108**:1-39

Seo Y, Stradmann-Bellinghausen B, Rittner C, Takahama K, Schneider PM (1998) Sequence polymorphism of mitochondrial DNA control region in Japanese. *Forensic Science International* **97**:155-164

Shin WS, Tanaka M, Suzuki J, Hemmi C, Toyo–oka T (2000) A novel homoplasmic mutation in mtDNA with a single evolutionary origin as a risk factor for cardiomyopathy. *American Journal of Human Genetics* **67**:1617–1620

Shitara H, Hayashi J-I, Takahama S, Kaneda H, Yonekawa H (1998) Maternal inheritance of mouse mtDNA in interspecific hybrids: segregation of the leaked paternal mtDNA followed by the prevention of subsequent paternal leakage. *Genetics* **148**:851-857

Solheim WGII (1980) Searching for the origins of the Orang Asli. *Federation Museums Journal* **25**:61-75

Solheim WGII, Bulbeck D, Flavel A (in press). The Nusantao Trading and Communication Network. Diliman, Quezon City: Archaeological Studies Program, University of the Philippines

Soodyall H, Jenkins T, Stoneking M (1995) 'Polynesian' mtDNA in the Malagasy. *Nature Genetics* **10**:377-378

Spriggs M (2003) Chronology of the Neolithic transition in Island Southeast Asia and the western pacific: a view from 2003. *The Review of Archaeology* **24**:57-80

Stoneking M, Jorde LB, Bhatia K, Wilson AC (1990) Geographic variation in human mitochondrial DNA from Papua New Guinea. *Genetics* **124**:717-733

Stringer CB, Andrews P (1988) Genetic and fossil evidence for the origin of modern humans. *Science* **239**:1263-1268

Stringer C (1999) Has Australia backdated the human revolution? *Antiquity* **73**:876-879

Su B, Xiao J, Underhill P, Deka R, Zhang W, Akey J, Huang W, Shen D, Lu D, Luo J, Chu J, Tan J, Shen P, Davis R, Cavalli-Sforza L, Chakraborty R, Xiong M, Du R, Oefner P, Chen Z, Jin L (1999) Y-chromosome evidence for a northward migration of modern humans into eastern Asia during the last ice age. *American Journal of Human Genetics* **65**:1718-1724

Su B, Jin L, Underhill P, Martinson J, Saha N, McGarvey ST, Shriver MD, Chu J, Oefner P, Chakraborty R, Deka R (2000) Polynesian origins: insights from the Y chromosome. *Proceedings of the National Academy of Sciences USA* **97**:8225-8228

Sykes B, Leiboff A, Low-Beer J, Tetzner S, Richards M (1995) The Origins of the Polynesians: an interpretation from mitochondrial lineage analysis. *American Journal of Human Genetics* **57**:1463-1475

Szabó K, O'Connor S (2004) Migration and complexity in Holocene Island Southeast Asia. *World Archaeology* **36**:621-628

Tajima A, Sun C-S, Pan I-H, Ishida T, Saitou N, Horai S (2003) Mitochondrial DNA polymorphisms in nine aboriginal groups of Taiwan: implications for the population history of aboriginal Taiwanese. *Human Genetics* **113**:24-33

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**:585-595

Tanaka M, Cabrera VM, González AM, Larruga JM, Takeyasu T, Fuku N, Guo L-J, Hirose R, Fujita Y, Kurata M, Shinoda K, Umetsu K, Yamada Y, Oshida Y, Sato Y, Hattori N, Mizuno Y, Arai Y, Hirose N, Ohta S, Ogawa O, Tanaka Y, Kawamori R, Shamoto-Nagai M, Maruyama W, Shimokata H, Suzuki R, Shimodaira H (2004) Mitochondrial genome variation in Eastern Asia and the peopling of Japan. *Genome Research* **14**:1832-1850

Templeton AR (1993) The "Eve" hypotheses: a genetic critique and reanalysis. *American Anthropologist* **95**:51-72

Terrell J (1981) Linguistics and the peopling of the Pacific islands. *Journal of the Polynesian Society* **90**:225-258

Terrell JE (2004) The 'sleeping giant' hypothesis and New Guinea's place in the prehistory of Greater Near Oceania. *World Archaeology* **36**:601-609

Terrell JE, Welsch RL (1997) Lapita and the temporal geography of prehistory. *Antiquity* **71**:548-572

Tommaseo-Ponzetta M, Attimonelli M, De Robertis M, Tanzariello F, Saccone C (2002) Mitochondrial DNA variability of West New Guinea populations. *American Journal of Physical Anthropology* **117**:49-67

Torroni A, Schurr TG, Yang C-C, Szathmary EJE, Williams RC, Schanfield MS, Troup GA, Knowler WC, Lawrence DN, Weiss KM, Wallace DC (1992) Native American mitochondrial DNA analysis indicates that the Amerind and Nadene populations were founded by two independent migrations. *Genetics* **130**:153-162

Torroni A, Miller JA, Moore LG, Zamudio S, Zhuang J, Droma T, Wallace DC (1994) Mitochondrial DNA analysis in Tibet: implications for the origin of the Tibetan population and its adaptation to high altitude. *American Journal of Physical Anthropology* **93**:189-199

Torroni A, Huoponen K, Francalacci P, Petrozzi M, Morelli L, Scozzari R, Obinu D, Savontaus M-L, Wallace DC (1996) Classification of European mtDNAs from an analysis of three European populations. *Genetics* **144**:1835-1850

Torroni A, Bandelt H-J, Macaulay V, Richards M, Cruciani F, Rengo C, Martinez-Cabrera V, Villems R, Kivisild T, Metspalu E, Parik J, Tolk H-V. Tambets K, Forster P, Karger B, Francaletti P, Rudan P, Janicijevic B, Rickards O, Savontaus M-L, Huoponen K, Laitinen V, Koivumäki S, Sykes B, Hickey E, Novelletto A, Moral P, Sellitto D, Coppa A, Al-Zaheri N, Santachiara-Benerecetti AS, Semino O, Scozzari R (2001) A signal, from human mtDNA, of postglacial recolonization in Europe. *American Journal of Human Genetics* **69**:844-852

Trejaut JA, Kivisild T, Loo JH, Lee CL, He CL, Xi JR, Li ZY, Lin M (In press) Traces of archaic mitochondrial lineages persist in Austronesian speaking Formosan populations. Unambiguous linkage between Aboriginal Taiwanese and Polynesians to the exclusion of Mainland Asians.

Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC (1991) African populations and the evolution of human mitochondrial DNA. *Science* **253**:1503-1507

Ward RH, Frazier BL, Dew-Jager K, Pääbo S (1991) Extensive mitochondrial diversity within a single Amerindian tribe. *Proceedings of the National Academy of Sciences USA* **88**:8720-8724

Watson E, Forster P, Richards M, Bandelt H-J (1997) Mitochondrial footprints of human expansions in Africa. *American Journal of Human Genetics* **61**:691-704

Wen B, Li H, Gao S, Mao X, Gao Y, Li F, Zhang F, He Y, Dong Y, Zhang Y, Huang W, Jin J, Xiao C, Lu D, Chakraborty R, Su B, Deka R, Jin L (2004) Genetic structure of Hmong-Mien speaking populations in East Asia as revealed by mtDNA lineages. *Molecular Biology and Evolution* **22**:725-734

Wickler S, Spriggs M (1988) Pleistocene human occupation of the Solomon Islands, Melanesia. *Antiquity* **62**:703-706

Wolpoff M, Thorne A (1991) The case against Eve. *New Scientist* **130**:37-41

Wrischnik LA, Higuchi RG, Stoneking M, Erlich HA, Arnheim N, Wilson AC (1987) Length mutations in human mitochondrial DNA: direct sequencing of enzymatically amplified DNA. *Nucleic Acids Research* **15**:529-542

Wurm SA (1983) Linguistic prehistory in the New Guinea area. *Journal of Human Evolution* **12**:25-35

Y Chromosome Consortium (2002) A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Research* **12**:339-348

Yao Y-G, Lu X-M, Luo H-R, Li W-H, Zhang Y-P (2000) Gene admixture in the silk road of China - evidence from mtDNA and melanocortin 1 receptor polymorphism. *Genes and Genetic Systems* **75**:173-178

Yao Y-G, Kong Q-P, Bandelt H-J, Kivisild T, Zhang Y-P (2002a) Phylogeographic differentiation of mitochondrial DNA in Han Chinese. *American Journal of Human Genetics* **70**:635-651

Yao Y-G, Nie L, Harpending H, Fu Y-X, Yuan Z-G, Zhang Y-P (2002b) Genetic relationship of Chinese ethnic populations revealed by mtDNA sequence diversity. *American Journal of Physical Anthropology* **118**:63-76

Yao Y-G, Zhang Y-P (2002c) Phylogeographic analysis of mtDNA variation in four ethnic populations from Yunnan Province: new data and a reappraisal. *Journal of Human Genetics* **47**:311-318

Yao Y-G, Kong Q-P, Wang C-Y, Zhu C-L, Zhang Y-P (2004) Different matrilineal contributions to genetic structure of ethnic groups in the silk road region in China. *Molecular Biology and Evolution* **21**:2265-2280

Zainuddin Z, Goodwin W (2004) Mitochondrial DNA profiling of modern Malay and Orang Asli populations in peninsular Malaysia. *International Congress Series* **1261**:428-430

Appendix I – Results of HVS – I Sequencing and Restriction Fragment Length Polymorphism Tests

Series	No	HVS-I Variants	Haplogroup	HVS-I Start	HVS-I End	RFLP Variants
ALO	1	16193 16223 16249 16291 16319 (16098N 16159N)	N21	16030	16500	-10397AluI, -10394DdeI, -15606AluI
ALO	4	16140 16189 16274 16335	B4c	16060	16465	+9bpdel
ALO	5	16129 16144 16148 16241 16265C 16311 16343	Q	16001	16386	+10397AluI, +10394DdeI
ALO	6	16129 16144 16148 16222/3insA 16241 16265C 16311 16343	Q	16024	16503	+10397AluI
ALO	7	16129 16172 16304 (16159N)	F1a	16030	16500	
ALO	8	16157 16256 16304 16335 (16159N)	F	16030	16500	
ALO	10	16140 16189 16266A 16291 (16159N 16215N)	B5a	16055	16470	+9bpdel
ALO	17	16189 16213 16223 16295 16362 (16159N)	M7c1c	16030	16470	+10397AluI, +10394DdeI, -9bpdel, +5176AluI, +9824HinfI, +7598HhaI
ALO	27	16108 16129 16162 16172 16304	F1a1a	16030	16500	
ALO	29	16223 16295 16362	M7c1c	16060	16500	+5176AluI, +7598HhaI
ALO	36	16086 16223 16291 16362 16390 16465	E1a	16050	16500	+10397AluI, +10394DdeI, +5176AluI, -7598HhaI
ALO	40	16157 16256 16304 16335	F	16045	16500	
ALO	41	16129 16172 16294 16304 16362	F1a	16050	16500	
ALO	44	16157 16256 16304 16335 (16054N)	F	16050	16500	
ALO	45	16140 16189 16249 16266A	B5a	16070	16470	+9bpdel
ALO	51	16129 16172 16294 16304 16362	F1a	16050	16500	
ALO	53	16093 16189 16217 16247 16261	B4a1	16027	16392	+9bpdel
ALO	54	16129 16144 16148 16172 16223 16241 16242 16265C 16311 16343	Q	16050	16500	+10397AluI
ALO	57	0	M	16060	16500	+10397AluI, +10394DdeI, +5176AluI,+7598HhaI, -9824HinfI
ALO	58	16288 16304 (16054N 16168N)	R22	16050	16500	
ALO	63	16129 16209 16223 16233 16259 16274 16290 16304	R9	16060	16500	
ALO	64	16157 16256 16335 (16054N)	F	16050	16500	

Series	No	HVS-I Variants	Haplogroup	HVS-I Start	HVS-I End	RFLP Variants
ALO	65	16129 16144 16148 16241 16265C 16311 16343	Q	16011	16500	+10397AluI
ALO	68	16129 16144 16148 16241 16265C 16311 16343	Q	16030	16500	
ALO	71	16129 16144 16148 16172 16223 16241 16265C 16311 16343 (16159N)	Q	16030	16500	
ALO	72	16157 16256 16335 (16159N)	F	16060	16500	
ALO	75	16129 16144 16148 16223 16265C 16311 16343 (16098N 16159N)	Q	16030	16500	+10397AluI
ALO	78	16129 16144 16148 16172 16223 16241 16265C 16311 16343	Q	16011	16500	+10397AluI
ALO	79	16129 16223 16241 16311	Q	16011	16500	+10397AluI, +10394DdeI, +5176AluI, +7598HhaI, -9824HinfI
ALO	98	16129 16144 16148 16241 16265C 16311 16343	Q	16030	16500	
ALO	100	16140 16189 16217 16235 16274	B4c	16060	16460	+9bpdel
ALO	103	16129 16172 16173 16294 16304 16362	F1a	16030	16500	
ALO	107	16140 16189 16217 16274 16335	B4c	16060	16460	+9bpdel
ALO	114	16223 16291 16362 16390	E1a	16060	16500	+10397AluI, +10394DdeI, +5176AluI
ALO	115	16193 16223 16291 16319	N21	16070	16500	-10397AluI, -10394DdeI, -15606AluI
ALO	119	16092 16129 16144 16148 16169 16223 16265C 16311 16343	Q	16060	16500	+10397AluI
ALO	127	16129 16144 16148 16241 16265C 16311 16343	Q	16060	16500	
ALO	133	16184A 16213 16223 16278	G2	16060	16500	+10397AluI, +10394DdeI, +5176AluI, +7598HhaI, -9824HinfI, +4831HhaI
ALO	142	16066 16290 16298 16357 16362	F3	16030	16500	-10397AluI, -10394DdeI
ALO	147	16129 16223 16274 16311 16317 16362	D4a	16060	16500	+10397AluI, +10394DdeI, -5176AluI
ALO	149	16188 16189 16223 16278 16288	G2	16030	16500	-9bpdel, +10397AluI, +10394DdeI, +5176AluI, +7598HhaI, -9824HinfI, +4831HhaI
ALO	152	16223 16291 16362 16390	E1a	16030	16500	+10397AluI, +5176AluI, -7598HhaI
ALO	158	16223 16261 16362 16390 (16259N)	E1b	16060	16500	+10397AluI, +5176AluI, -7598HhaI

Series	No	HVS-I Variants	Haplogroup	HVS-I Start	HVS-I End	RFLP Variants
ALO	160	16129 16144 16148 16172 16223 16241 16265C 16311 16343	Q	16060	16500	
ALO	161	16223 16261 16362 16390	E1b	16060	16500	+10397AluI, -7598HhaI
AMB	3	16223 16362	M7	16015	16500	+10397AluI, +10394DdeI, +7598HhaI, +9824HinfI
AMB	4	16108 16129 16162 16172 16304	F1a1a	16030	16500	
AMB	6	16129 16144 16148 16172 16223 16241 16256 16265C 16311 16343	Q	16025	16500	+10397AluI, +10394DdeI
AMB	9	16189 16266A	B5a	16030	16465	+9bpdel
AMB	10	16140 16189 16266A	B5a	16040	16465	+9bpdel
AMB	12	16189 16217 16247 16261 16362	B4a1	16030	16470	+9bpdel
AMB	14	16223 16362 16390	D	16015	16500	+10397AluI, +10394DdeI, +7598HhaI, -5176AluI
AMB	17	16136 16189 16217	B4b1	16045	16470	+9bpdel
AMB	20	16189 16217 16261	B4a	16030	16470	+9bpdel
AMB	22	16223 16291 16362 16390	E1a	16030	16500	+10397AluI, +5176AluI
AMB	26	16104 16129 16172 16294 16304 16362	F1a	16030	16500	
AMB	28	16104 16129 16172 16294 16304 16362 (16054N)	F1a	16030	16500	
AMB	36	16189 16217 16247 16261	B4a1	16015	16480	+9bpdel
AMB	39	16189 16217 16247 16261	B4a1	16012	16394	
AMB	52	16168 16189 16209 16223 16233 16304	M	16050	16480	+10397AluI, +5176AluI, +7598HhaI, -9824HinfI
AMB	56	16051 16223 16362 16390	E1	16025	16500	+10397AluI, +10394DdeI, +5176AluI, -9824HinfI, -7598HhaI
AMB	57	16223 16248 16362 16390	E1	16015	16500	+10397AluI, +5176AluI, -9824HinfI, -7598HhaI
AMB	63	16148 16189 16223 16362	D5	16050	16480	-10397AluI, -10394DdeI, -9bpdel, -5176AluI
AMB	65	16295 16362	D	16050	16460	+10397AluI, +10394DdeI, -5176AluI
AMB	66	16051 16223 16362 16390	E1	16030	16500	+5176AluI
AMB	67	16093 16189 16223 16265 16278	G2	16045	16480	+10397AluI, +5176AluI, -9824HinfI
AMB	68	16223 16362 16390	M	16015	16500	+10397AluI, +5176AluI, +7598HhaI, -9824HinfI
AMB	71	16223 16362 16390	M	16055	16500	+5176AluI, +7598HhaI
AMB	72	16184A 16223	G2a?	16045	16500	+10397AluI, +10394DdeI, +5176AluI, +7598HhaI, -9824HinfI

Series	No	HVS-I Variants	Haplogroup	HVS-I Start	HVS-I End	RFLP Variants
AMB	75	16129 16144 16148 16154 16209 16222 16241 16265C 16311 16343	Q	16075	16500	+10397AluI
AMB	81	16136 16189 16217	B4b1	16050	16480	
AMB	84	16189 16217 16261 16278 (16156N)	B4a	16050	16480	+9bpdel
AMB	88	16189 16217 16247 16261	B4a1	16033	16408	+9bpdel
AMB	89	16189 16217 16247 16261	B4a1	16037	16385	
AMB	94	16129 16189 16223 16297	M7b	16015	16420	+10397AluI, +10394DdeI, +5176AluI, +7598HhaI, +9824HinfI
AMB	96	16129 16172 16304 16311	F1a	16030	16500	
AMB	97	16129 16144 16148 16223 16241 16265C 16311 16343	Q	16050	16500	
AMB	98	16189 16217 16261	B4a	16030	16460	
AMB	102	16140 16189 16266A	B5a	16050	16475	
AMB	103	16129 16144 16148 16193 16223 16241 16265C 16311 16343 16362	Q	16050	16500	
AMB	105	16189 16217 16223 16261	B4a	16020	16475	+9bpdel
AMB	108	16189 16217 16247 16261	B4a1	16026	16424	
AMB	109	16129 16172 16304 16311	F1a	16020	16500	
AMB	110	16129 16172 16304 16311	F1a	16060	16500	
AMB	111	16129 16144 16148 16193 16223 16241 16265C 16311 16343 16362	Q	16020	16500	+10397AluI
AMB	112	16126 16129 16192 16223 16297	M7b1	16050	16480	+10397AluI, +10394DdeI, +5176AluI, +7598HhaI, +9824HinfI
AMB	113	16223 16291 16362 16390	E1a	16045	16500	+10397AluI, +10394DdeI, +5176AluI
AMB	118	16223 16295 16362	M7c1c	16030	16500	+10397AluI, +5176AluI, +7598HhaI, +9824HinfI
BAL	2	16148 16189 16223 16246T 16311 16362	M	16050	16500	-9bpdel, +10397AluI, +10394DdeI, +5176AluI, +7598HhaI, -9824HinfI
BAL	3	16129 16162 16172 16189 16304 16311 (16252N)	F1a1	16020	16430	
BAL	4	16108 16189 16217 16261	B4a	16030	16430	+9bpdel
BAL	5	16140 16189 16266A	B5a	16010	16440	+9bpdel
BAL	6	16093 16209 16223 16325	M	16000	16500	+10397AluI, +10394DdeI, +5176AluI, +7598HhaI, -9824HinfI

Series	No	HVS-I Variants	Haplogroup	HVS-I Start	HVS-I End	RFLP Variants
BAL	7	16129 16189 16218 16223 (16280N 16284N)	G	16040	16400	-9bpdel, +10397AluI, +5176AluI, +7598HhaI, -9824HinfI, +4831HhaI
BAL	8	16126 16129 16192 16223 16297	M7b1	16000	16500	+10397AluI, +10394DdeI, +5176AluI, +7598HhaI, +9824HinfI
BAL	10	16145 16181 16192 16223 16291 16304	M	16000	16500	+10397AluI, +10394DdeI, +5176AluI, +7598HhaI, -9824HinfI
BAL	12	16189 16223 16311 16362	D5	16030	16430	-9bpdel, +10397AluI, -5176AluI
BAL	13	16256 16290 16465 (16362N)	R	16030	16500	-10397AluI ,-10394DdeI, +5176AluI, -15606AluI
BAL	14	16108 16129 16162 16172 16304	F1a1a	16000	16500	
BAL	15	16223 16291 16362 16390	E1a	16000	16500	+10397AluI, -7598HhaI
BAL	16	16145 16181 16192 16223 16266 16291 16304	M	16000	16500	+10397AluI, +5176AluI, +7598HhaI, -9824HinfI
BAL	17	16249 16288 16304 16390	R22	16030	16500	
BAL	18	16086 16129 16148 16223 16241 16265C 16311 16343	Q	16010	16400	+10397AluI
BAL	19	16086 16223 16243 16262 16278 16311 16319	G2	16000	16500	+10397AluI, +10394DdeI, +5176AluI, +7598HhaI, -9824HinfI, +4831HhaI
BAL	21	16223 16254 16362	M	16000	16500	+10397AluI, +10394DdeI, +5176AluI, +7598HhaI, -9824HinfI
BAL	22	16129 16153 16162 16172 16304 16311 16399	F1a1	16000	16500	
BAL	25	16129 16172 16304 16352	F1a	16000	16500	
BAL	26	16129 16162 16172 16304 16311 16399	F1a1	16010	16500	
BAL	27	16037 16129 16172 16304	F1a	16000	16500	
BAL	28	16129 16162 16172 16189 16304 16311	F1a1	16030	16400	
BAL	29	16249 16288 16304 16344	R22	16000	16500	
BAL	30	16234 16256 16278 16294	G2	16000	16500	+10397AluI, +10394DdeI, +5176AluI, +7598HhaI, -9824HinfI
BAL	32	16147 16189 16217	B4	16000	16440	+9bpdel
BAL	33	16129 16155 16219 16223 16356 16362	M7	16000	16500	+10397AluI, +10394DdeI, +5176AluI, +7598HhaI, +9824HinfI
BAL	34	16256 16290 16465	R	16000	16500	-10397AluI, -10394DdeI, -15606AluI
BAL	35	16129 16189 16192 16215 16223 16297	M7b1	16000	16500	+10397AluI, +7598HhaI, +9824HinfI

Series	No	HVS-I Variants	Haplogroup	HVS-I Start	HVS-I End	RFLP Variants
BAL	38	16129 16140 16189 16192 16223 16265 16297	M7b1	16010	16410	-9bpdel, +10397AluI, +5176AluI, +9824HinfI
BAL	39	16147 16189 16217 16235	B4	16000	16500	+9bpdel
BAL	41	16092 16140 16189 16217 16335	B4c	16000	16400	+9bp
BAL	43	16140 16189 16217 16235 16274	B4c	16030	16420	-10397AluI, +9bpdel
BAL	44	16108 16129 16162 16172 16234 16299 16304	F1a1a	16010	16500	
BAL	45	16249 16288 16304 16390	R22	16000	16500	
BAL	45	16249 16288 16304 16390	R22	16050	16500	
BAL	47	16124 16189 16278 16292 16362	G2	16010	16410	-9bpdel, +10397AluI, +10394DdeI, +5176AluI, -9824HinfI
BAL	48	16037 16129 16172 16304	F1a	16000	16500	
BAL	49	16249 16288 16304 16390 (16477N)	R22	16000	16500	
BAL	50	16189 16217	B4	16050	16420	+9bpdel -10397AluI
BAL	51	16193 16223	N21	16000	16500	-10397AluI, -10394DdeI, -15606AluI
BAL	52	16093 16193	M	16000	16500	+10397AluI, +10394DdeI, +5176AluI, -9824HinfI, +7598HhaI
BAL	53	16223 16311 16362	M	16000	16500	+10397AluI, +5176AluI, -9824HinfI, +7598HhaI
BAL	54	16223 16234 16261 16290	M12a	16020	16500	+10397AluI, +10394DdeI, +5176AluI, -9824HinfI
BAL	56	16092 16147 16179 16189 16217 16235	B4	16000	16410	+9bp
BAL	58	16124 16189 16209 16293C 16304 16362	R9	16000	16410	-9bpdel, +5176AluI
BAL	60	16172 16223 16245A	M	16000	16500	+10397AluI, +10394DdeI, +5176AluI, -9824HinfI
BAL	61	16140 16189 16266A	B5a	16030	16440	
BAL	62	16223 16311 16362	M	16000	16500	+10397AluI, +5176AluI, -9824HinfI
BAL	63	16223 16278 16294	G2	16000	16500	+10397AluI, +10394DdeI, +5176AluI, +7598HhaI, -9824HinfI, +4831HhaI
BAL	64	16223 16291 16362 16390	E1a	16000	16430	+10397AluI, +5176AluI
BAL	68	16140 16189 16266A (16358N)	B5a	16010	16400	
BAL	72	16129 16209 16223 16325	G	16000	16500	+10397AluI, +5176AluI, -9824HinfI, +7598HhaI
BAL	73	16126 16192 16231 16311	Y2	16000	16500	-10397AluI, +10394DdeI
BAL	75	16129 16209 16223 16272	G	16000	16500	+10397AluI, +10394DdeI, +5176AluI, +7598HhaI, -9824HinfI, +4831HhaI

Series	No	HVS-I Variants	Haplogroup	HVS-I Start	HVS-I End	RFLP Variants
BAL	76	16129 16189 16223 16297	M7b	16010	16410	-9bpdel, +5176AluI, +10397AluI, +9824HinfI, +10394DdeI
BAL	77	16092 16129 16209 16223 16325 (16477N)	G	16000	16500	+10397AluI, +10394DdeI, +5176AluI, -9824HinfI, +4831HhaI
BAL	78	16223 16295 16362	M7c1c	16000	16500	+10397AluI, +5176AluI, +9824HinfI
BAL	80	16129 16256 16290 16465	R	16000	16500	-10397AluI, -10394DdeI, -15606AluI
BAL	81	16192 16223 16295 16362	M7c1c	16000	16500	+10397AluI, +5176AluI, +7598HhaI, +9824HinfI
BAL	82	16037 16129 16172 16304	F1a	16000	16500	
BAL	83	16304 16362	F	16000	16500	
BAL	84	16140 16189 16217 16235 16274	B4c	16030	16410	
BAL	85	16140 16189 16266A	B5a	16010	16420	+9bpdel
BAL	86	16223 16261 16362 16390 (16477N)	E1b	16010	16500	+10397AluI, +5176AluI, -7598HhaI
BAL	87	16129 16172 16223 16291 16305	F1a	16000	16500	
BAN	1	16129 16185 16189 16223 16260 16298	Z	16001	16500	+10397AluI +10394DdeI +5176AluI -9bpdel, +7598HhaI
BAN	2	16093 16220C 16298 16362	F3b	16024	16500	-10397AluI -10394DdeI
BAN	3	16189 16217 16261 (16324N)	B4a	16020	16410	
BAN	4	16129 16172 16189 16223 16297	M7b	16024	16420	+10397 AluI +10394 DdeI +5176AluI +9824HinfI
BAN	5	16108 16111 16129 16162 16172 16189 16304	F1a1a	16020	16417	-9bpdel
BAN	6	16140 16189 16243	B5b	16020	16410	
BAN	7	16140 16189 16217 16274 16335 (16341N)	B4c	16033	16400	+9bpdel
BAN	8	16129 16209 16223 16272	G	16021	16500	+10397AluI +5176AluI -9824HinfI, +7598HhaI
BAN	9	16223 16261 16362 16390	E1b	16016	16500	+10397AluI +5176AluI -9824HinfI, -7598HhaI
BAN	10	16172 16173 16223 16278 16311	G2	16016	16500	+10397AluI +5176AluI -9824HinfI, +7598HhaI
BAN	11	16223 16261 16288 16362 16390 (16416N)	E1b	16043	16500	+10397AluI +5176AluI -9824HinfI
BAN	12	16223 16261 16288 16362 16390	E1b	16027	16500	+10397AluI +5176AluI -9824HinfI
BAN	13	16136 16189 16217 16261	B4b1	16018	16426	
BAN	14	16045 16223 16311 16362 (16124N 16199N 16342N 16398N 16412N 16440N)	M	16043	16460	+10397AluI +5176AluI -9824HinI
BAN	15	16111 16168 16172 16189 16223 16311 16362 (16120N)	N	16045	16450	-10397 AluI -10394 DdeI +5176AluI, -9bpdel, -15606AluI

Series	No	HVS-I Variants	Haplogroup	HVS-I Start	HVS-I End	RFLP Variants
BAN	16	16093 161845 16223 16278	G2	16012	16500	+10397AluI +5176AluI -9824HinfI, +7598HhaI
BAN	17	16129 16144 16148 16172 16223 16241 16265C 16311 16343	Q	16016	16500	+10397 AluI +10394 DdeI +5176AluI -9824HinfI
BAN	18	16093 16220C 16298 16362	F3b	16016	16500	-10397AluI -10394DdeI
BAN	21	16140 16189 16266A	B5a	16027	16430	
BAN	23	16223 16311	M	16016	16500	+10397AluI +5176AluI -9824HinfI, +7598HhaI
BAN	25	16189 16217 16261	B4a	16016	16409	-10397AluI -10394DdeI
BAN	26	16189 16217 16247 16261	B4a1	16028	16450	-10397AluI -10394DdeI, +9bpdel
BAN	27	16093 16220C 16265 16298 16362	F3b	16016	16500	-10397AluI -10394DdeI
BAN	28	16172 16173 16223 16278 16311	G2	16000	16500	+10397 AluI +10394 DdeI +5176AluI -9824HinfI, +7598HhaI
BAN	30	16093 16129 16223 16256 16271 (16045N)	M21a	16019	16500	+10397AluI +5176AluI
BAN	32	16223 162652 16291 16362 16390	E1a	16019	16500	+10397AluI +5176AluI -9824HinfI
BAN	34	16168 16172 16189 16311 16362	F1a	16040	16398	-10397AluI -10394DdeI -9bpdel
BAN	35	16223 16311 16362	M	16010	16500	+5176AluI -9824HinfI, +7598HhaI
BAN	36	16223 16299 16311 16362	M	16019	16500	+10397AluI +5176AluI -9824HinfI, +7598HhaI
BAN	37	16172 16362	D	16027	16500	+10397AluI -5176AluI
BAN	38	16192 16234 16288 16304 16309 16390	R9b	16030	16500	-10397AluI -10394DdeI
BAN	39	16178 16189 16217 16261	B4a	16020	16455	+9bpdel
BAN	40	16140 16189 16266A 16274	B5a	16010	16430	
BAN	41	16189 16217 16247 16261 (16045N)	B4a1	16020	16390	
BAN	42	16093 16223 16319	M7c1a	16016	16500	+10397AluI +5176AluI +9824HinfI
BAN	43	16223 16291 16362 16390	E1a	16016	16500	+10397AluI +5176AluI -9824HinfI, -7598HhaI
BAN	44	16147 16189 16217 16235 16294 16360	B4	16016	16410	-10397AluI -10394DdeI, +9bpdel
BAN	46	16051 16223 16362 16390 (16045N)	M	16026	16500	+10397AluI +5176AluI -9824HinfI, +7598HhaI
BAN	47	16086 16148 16223 16259 16278 16319 16399 (16045N)	G2	16019	16500	+10397AluI +5176AluI -9824HinfI
BAN	48	16093 16220C 16265 16298 16362	F3b	16024	16500	-10397 AluI -10394 DdeI
BAN	49	16126 16231 16311	Y2	16016	16500	-10397AluI +10394DdeI
BAN	50	16147 16189 16217 16235 16294G	B4	16000	16440	-10397 AluI -10394 DdeI
BAN	51	16220C 16261 16265 16298 16362	F3b	16020	16500	-10397AluI -10394DdeI
BAN	52	16178 16189 16217 16261 (16045N)	B4a	16019	16420	+9bpdel

Series	No	HVS-I Variants	Haplogroup	HVS-I Start	HVS-I End	RFLP Variants
BAN	53	16093 16220C 16298 16362	F3b	16016	16500	-10397 AluI -10394 DdeI
BAN	54	16129 16223 16263 16381	M21b	16000	16500	+10397 AluI +10394 DdeI +5176AluI -9824HinfI
BAN	55	16249 16288 16295 16304	R22	16016	16500	-10397AluI -10394DdeI
BAN	56	16016 16140 16189 16266A 16298	B5a	16010	16430	+9bpdel
BAN	57	16093 16223 16311 16362	M	16024	16500	+10397AluI +5176AluI -9824HinfI, +7598HhaI
BAN	58	16129 16172 16304 16311	F1a	16016	16500	-10397AluI -10394DdeI
BAN	59	16147 16189 16217 16235	B4	16016	16450	
BAN	60	16140 16189 16304	F?	16020	16410	-10397 AluI -10394 DdeI -9bpdel
BAN	61	16147 16189 16217	B4	16020	16455	+9bpdel
BAN	62	16129 16223 16263 16381 (16045N 16086N)	M21b	16018	16500	+10397AluI +5176AluI -9824HinfI
BAN	63	16129 16172 16294 16304 16362	F1a	16026	16500	-10397AluI -10394DdeI
BAN	64	16185 16223 16291 16362 16390 (16045N)	E1a	16016	16500	+10397AluI +5176AluI -9824HinfI
BAN	65	16223 16295 16362	M7c1c	16026	16500	+10397AluI +5176AluI +9824HinfI
BAN	66	16140 16189 16266A	B5a	16033	16420	
BAN	67	16111 16168 16172 16189 16223 16311 16362	N	16030	16430	-10397 AluI -10394 DdeI, -15606AluI
BAN	68	16223 16278 16311	G2	16016	16500	+10397AluI +5176AluI -9824HinfI, +7598HhaI, +4831HhaI
BAN	69	16249 16288 16295 16304 (16045N)	R22	16016	16500	-10397AluI -10394DdeI
BAN	70	16140 16189 16266A	B5a	16020	16400	
BAN	71	16189 16217 16261 (16064N)	B4a	16030	16410	
BAN	72	16086 16187 16223 16257A 16261 16292 16294	N9a1	16016	16500	-10397AluI -10394DdeI -9bpdel, -15606AluI
BAN	74	16147 16189 16217 16235	B4	16016	16410	
BAN	75	16189 16217 16261	B4a	16020	16400	
BAN	76	16136 16189 16217	B4b1	16020	16400	
BAN	77	16129 16223 16234 16290 16311 (16045N)	G	16019	16500	+10397AluI +5176AluI -9824HinfI, +7598HhaI, +4831HhaI
BAN	78	16086 16147 16189 16217	B4	16019	16410	
BAN	79	16129 16162 16172 16304	F1a1	16016	16500	-10397AluI -10394DdeI

Series	No	HVS-I Variants	Haplogroup	HVS-I Start	HVS-I End	RFLP Variants
BAN	80	16051 16093 16145 16223 16234 16249 16290 16399	M12a	16013	16500	+10397AluI +5176AluI -9824HinfI, +7598HhaI
BAN	81	16111 16168 16172 16189 16223 16311 16319 16362 (16045N)	N	16016	16400	-10397 AluI -10394DdeI -9bpdel, -15606AluI
BAN	82	16093 16220C 16298 16362	F3b	16024	16500	
BAN	84	16147 16189 16217 16235	B4	16020	16435	
BAN	85	16147 16189 16217	B4	16015	16400	+9bpdel
BAN	85	16189 16217	B4	16019	16405	
BAN	86	16223 16311 16362	M	16010	16500	+10397 AluI +10394 DdeI +5176AluI -9824HinfI, +7598HhaI
BAN	87	16093 16223 16278 16310 (16045N)	G2	16026	16500	+10397AluI +5176AluI -9824HinfI, +7598HhaI
BAN	88	16223 16266 16284 16290	M	16016	16500	+10397AluI +5176AluI -9824HinfI, +7598HhaI
BAN	89	16140 16189 16243	B5b	16016	16415	
BAN	90	16129 16148 16172 16223 16256 16305 16309	M	16010	16490	+10397 AluI +5176AluI -9824HinfI, +10394DdeI
BAN	91	16189 16217 16261	B4a	16016	16410	
BAN	92	16051 16223 16362 16390 (16045N)	E1	16016	16500	+10397AluI +5176AluI -9824HinfI
BAN	94	16111 16172 16189 16223 16311 16362	N	16015	16410	-10397 AluI -9bpdel, -15606AluI
BAN	95	16140 16189 16217 16274 16335 (16045N)	B4c	16013	16410	-10397 AluI +9bpdel
BAN	97	16129 16172 16271 16304 16311	F1a	16015	16500	-10397AluI -10394DdeI
BAN	98	16051 16223 16362 16390	E1	16015	16500	+10397AluI +5176AluI -9824HinfI
BAN	100	16147 16189 16217 16235	B4	16010	16430	
BAN	106	16223 16278	G2	16016	16500	+10397AluI +5176AluI -9824HinfI, +4831HhaI
BGK	5	16189 16223 (16051N 16261N)	M	16010	16500	+10397AluI, -9bpdel, +10394DdeI, +5176AluI, -9824HinfI, +7598HhaI
BGK	14	16209 16223 16234 16261 16290 16304 (16168N)	R9	16020	16500	
BGK	18	16223 16263 16274 16311 16343 16357	M10	16015	16500	-10397AluI, +10394DdeI, -9824HinfI
BGK	25	16223 16263 16274 16311 16343 16357 (16273N)	M10	16015	16500	-10397AluI, +10394DdeI, +5176AluI, -9824HinfI, +7598HhaI, -15606AluI
BGK	26	16093 16189 16223 16278 16319	G2	16015	16410	-9bpdel, +5176AluI, +7598HhaI, -15606AluI, +10397AluI, +10394DdeI

Series	No	HVS-I Variants	Haplogroup	HVS-I Start	HVS-I End	RFLP Variants
BGK	28	16223 16291 16362 16390	E1a	16015	16500	+10397AluI, +5176AluI
BGK	37	16223 16362 16390	M	16015	16500	+10397AluI, +10394DdeI
BGK	38	16086 16129 16209 16223 16272	G	16015	16500	+7598HhaI, +10397AluI, +10394DdeI, +4831HhaI
BGK	40	16126 16231 16311	Y2	16000	16500	-10397AluI, -9bpdel, +5176AluI
BGK	41	16086 16129 16209 16223 16272	G	16050	16500	+10397AluI, +10394DdeI, +7598HhaI
BGK	43	16051 16215 16223 16362 16390 16399	E1	16045	16500	+10397AluI, +10394DdeI, +5176AluI, -9824HinfI, -7598HhaI
BGK	44	16189 16223 16257A 16261 16292	N9a1	16050	16410	-9bpdel, -10397AluI, -10394DdeI, +5176AluI
BGK	47	16093 16129 16223 16234 16290 16311 16384	G	16015	16500	+10397AluI, +10394DdeI, +5176AluI, -9824HinfI, +7598HhaI, +4831HhaI
BGK	49	16223 16291 16362 16390	E1a	16065	16500	+10397AluI, +5176AluI, -7598HhaI, +10394DdeI
BGK	53	16223 16295 16362	M7c1c	16060	16500	+10397AluI, +5176AluI, +9824HinfI
BGK	60	16129 16172 16304	F1a	16045	16500	
BGK	66	16172 16223 16261 16362 16390	E1b	16020	16500	+10397AluI, -7598HhaI
BGK	68	16189 16223 16257A 16261 16292	N9a1	16030	16410	-9bpdel
BGK	70	16093 16209 16223 16224 16263 16274 16278 16319 16356	G2	16015	16500	+10397AluI, +5176AluI, -9824HinfI, +7598HhaI
BGK	78	16223 16263 16274 16311 16343 16357	M10	16020	16500	-10397AluI, +10394DdeI, +5176AluI, +7598HhaI, -15606AluI
BGK	79	16086 16129 16209 16223 16272 16311	G	16045	16500	+10397AluI, +10394DdeI, +5176AluI, +7598HhaI, -9824HinfI, +4831HhaI
BGK	86	16140 16189 16217 16335	B4c	16020	16420	+9bpdel
BGK	89	16051 16189 16362	B	16045	16410	-10397AluI, +9bpdel
BGK	90	16126 16231 16311	Y2	16020	16500	
BGK	92	16189 16223 16257A 16261 16292 (16156N)	N9a1	16050	16420	-10397AluI, -10394DdeI
BGK	93	16147 16189 16217 16235	B4	16045	16460	+9bpdel
BGK	99	16086 16129 16209 16223 16272 16311	G	16020	16500	+10397AluI, +5176AluI, +7598HhaI
BGK	100	16223 16311 16362	M	16045	16500	+10397AluI, +5176AluI, +7598HhaI
BGK	101	16129 16166C 16189 16223 16287 16319	G	16035	16455	+10397AluI, +10394DdeI, +5176AluI, +7598HhaI, -9824HinfI, +4831HhaI
BGK	103	16223 16293 16311 16362	M	16015	16500	+10397AluI, +5176AluI, +7598HhaI, -9824HinfI

Series	No	HVS-I Variants	Haplogroup	HVS-I Start	HVS-I End	RFLP Variants
BGK	104	16189 16213 16223 16271 16311	M10	16020	16475	+10397AluI, +10394DdeI, +5176AluI, +7598HhaI, -9824HinfI
BGK	106	16223 16311 16362	M	16010	16405	+10397AluI, +5176AluI, +7598HhaI
BGK	109	16108 16129 16162 16172 16304	F1a1a	16065	16500	
BGK	111	16223 16263 16274 16311 16343 16357	M10	16060	16500	-10397AluI, +10394DdeI, +5176AluI, +7598HhaI, -15606AluI
BUN	1	16218 16304 16311	F4b	16080	16500	
BUN	5	16218 16304 16311	F4b	16050	16497	
BUN	7	16218 16304 16311	F4b	16010	16497	
BUN	10	16086 16185 16223 16362 16390	E1	16050	16500	+10397AluI, +5176AluI, -7598HhaI
BUN	15	16218 16304 16311	F4b	16010	16500	
BUN	16	16218 16304 16311	F4b	16015	16500	
BUN	18	16140 16189 16266G	B5a	16050	16480	+9bpdel
BUN	20	16086 16129 16297 16324	M7b3	16040	16500	+10397AluI
FIL	3	16157 16169 16256 16304 16311 16335	F	16015	16500	
FIL	7	16126 16231 16311	Y2	16030	16500	
FIL	8	16126 16231 16311	Y2	16020	16500	
FIL	9	16223 16291 16362 16390	E1a	16020	16500	+10397AluI, +10394DdeI
FIL	13	16223 16269 16271 16311	M10	16020	16500	+10397AluI, +7598HhaI, +10394DdeI, -9824HinfI
FIL	14	16293 16311 16355	R	16030	16500	-10397AluI, +7598HhaI, -10394DdeI, -9824HinfI
FIL	17	16129 16172 16304 16311	F1a	16020	16500	
FIL	19	16223 16291 16311 16362 16390	E1a	16030	16500	+10397AluI, -7598HhaI, +10394DdeI
FIL	20	16145 16176 16224 16233 16311	M	16020	16500	+7598HhaI, -9824HinfI
FIL	31	16220c 16265 16298 16335 16362	F3b	16030	16500	
FIL	35	16220c 16265 16298 16335 16362	F3b	16020	16500	
FIL	37	16220c 16265 16298 16335 16362	F3b	16040	16500	
FIL	38?	16220c 16265 16298 16335 16362	F3b	16050	16500	
FIL	39	16220c 16265 16298 16335 16362	F3b	16050	16500	
FIL	42	16051 16223 16362 16390	E1	16040	16500	+10397AluI, +10394DdeI
FIL	44	16126 16231 16284 16311	Y2	16030	16500	
FIL	62	16223 16295 16362 (16156N 16159N)	M7c1c	16050	16500	+10397AluI, +9824HinfI
FIL	63	16223 16295 16362	M7c1c	16050	16500	+10397AluI

Series	No	HVS-I Variants	Haplogroup	HVS-I Start	HVS-I End	RFLP Variants
FIL	66	16223 16295 16362	M7c1c	16060	16500	
KK	6	16175 16223 16298 16311 16327 (16110N)	C	16050	16500	+10397AluI, +10394DdeI, +5176AluI, +7598HhaI
KK	15	16304 16362	F	16030	16500	
KK	17	16223 16243 16311 16362	M	16015	16500	+10397AluI, +10394DdeI, +5176AluI, +7598HhaI, -9824HinfI
KK	18	16093 16223 16295 16337 16362	M7c1c	16015	16500	+10397AluI, +7598HhaI, +9824HinfI
KK	21	16185 16223 16291 16362 16390	E1a	16015	16500	+10397AluI, +10394DdeI, +5176AluI, -7598HhaI, -9824HinfI
KK	22	16185 16223 16291 16362 16390	E1a	16020	16500	
KK	23	16189 16192 16223 16294G 16297	M7b1	16030	16470	+10397AluI, +5176AluI, +7598HhaI, +9824HinfI
KK	25	16126 16129 16189 16278	G2	16015	16480	-9bpdel, +10394AluI, -9824HinfI, +10394DdeI, +4831HhaI
KK	26	16129 16223 16291 16362 16390	E1a	16025	16500	+10397AluI, +5176AluI, -7598HhaI, -9824HinfI
KK	27	16187 16241 16269 16319 16342	R	16015	16500	-10397AluI, -10394DdeI, -9bpdel, +5176AluI, -15606AluI
KK	28	16354	M7	16015	16500	+10397AluI, +10394DdeI, +5176AluI, +7598HhaI, +9824HinfI
KK	29	16126 16189 16223 16362	D5	16020	16480	-10397AluI, -9bpdel, -5176AluI
KK	30	16172 16223 16362 16390	M	16015	16500	+10397AluI, +5176AluI, +7598HhaI, -9824HinfI
KK	31	16111 16168 16172 16189 16223 16242 16263 16311 16362	N	16025	16470	-10397AluI, -9bpdel, -10394DdeI, -15606AluI
KK	32	16129 16172 16192 16294 16304 16362	F1a	16015	16500	
KK	45	16126 16231 16311	Y2	16060	16500	-10397AluI, +10394DdeI
KK	46	16093 16223 16295 16362	M7c1c	16050	16500	+10397AluI, +9824HinfI
KK	48	16223 16295 16346C 16362 (16110N)	M7c1c	16080	16500	+10397AluI, +9824HinfI
KK	49	16223 16295 16362 (16110N)	M7c1c	16050	16500	+9824HinfI
KK	55	16223 16291 16362 16390 (16179N)	E1a	16050	16500	+10397AluI, +10394DdeI, -9824HinfI
KK	57	16185 16223 16291 16362 16390	E1a	16050	16500	+5176AluI, -9824HinfI, -7598HhaI
KK	60	16129 16172 16304 16311	F1a	16045	16500	-9824HinfI
KK	61	16223 16291 16311 16342 16362 16390	E1a	16015	16500	+10397AluI, -9824HinfI, -7598HhaI
KK	63	16223 16298 16327 (16110N)	C	16050	16500	+10397AluI, +5176AluI, +7598HhaI, -9824HinfI
KK	64	16223 16295 16362	M7c1c	16060	16500	+9824HinfI

Series	No	HVS-I Variants	Haplogroup	HVS-I Start	HVS-I End	RFLP Variants
KK	65	16220C 16265 16298 16362	F3b	16060	16500	
KK	68	16249 16288 16304 16390	R22	16050	16500	
KK	69	16157 16256 16304 16335	F	16050	16500	
KK	70	16129 16209 16223 16272 (16342N)	G	16030	16500	+10397AluI, +10394DdeI, +5176AluI, +7598HhaI, -9824HinfI
KK	71	16126 16231 16311	Y2	16025	16500	-10397AluI
KK	131	16223 16278 16295 16362	M7c1c	16010	16500	+10397AluI, +10394DdeI, +5176AluI, +7598HhaI, +9824HinfI
KK	132	16093 16129 16209 16223 16272	G	16000	16500	+10397AluI, +5176AluI, -9824HinfI, +4831HhaI
KK	133	16093 16223 16295 16337 16362	M7c1c	16000	16500	+10397AluI, +9824HinfI
KK	134	16185 16223 16291 16362 16390 (16064N)	E1a	16030	16470	+10397AluI, +5176AluI, -9824HinfI
KK	135	16223 16294 (16311N 16362N het?)	N	16000	16500	-10397AluI, +5176AluI, -15606AluI
KK	136	16126 16129 16223 16297 (16477N)	M7b	16000	16500	+10397AluI, +10394DdeI, +5176AluI, +7598HhaI, +9824HInfI
KK	137	16051 16223 16362 16390	E1	16020	16500	+10397AluI, +5176AluI, -9824HinfI
KK	138	16185 16223 16291 16362 16390	E1a	16000	16500	+10397AluI, +5176AluI, -7598HhaI, -9824HinfI
KK	139	16185 16223 16291 16362 16390	E1a	16000	16500	
KK	140	16140 16189 16266A 16362	B5a	16035	16400	+9bpdel
KK	141	16189 16217 16261	B4a	16040	16400	+9bpdel
KK	142	16223 16291 16362 16390	E1a	16030	16500	
KK	143	16220C 16265 16274 16298 16311 16362	F3b	16030	16500	-10397AluI, -10394DdeI
KK	144	16189 16217 16223 16261	B4a	16010	16420	+9bpdel
KK	145	16140 16189 16266A	B5a	16000	16455	+9bpdel
KK	147	16118 16129 16162 16172 16304	F1a1	16020	16500	
KK	148	16223 16311	M	16000	16500	+10397AluI, -9824HinfI
KK	149	16126 16129 16189 16278 (16042N)	G2	16030	16410	-+10397AluI, +5176AluI, -9824HinfI
KK	150	16185 16223 16291 16362 16390	E1a	16030	16500	
KK	151	16126 16129 16189 16278	G2	16000	16400	+10397AluI, +10394DdeI, +5176AluI, +7598HhaI, -9824HinfI
KK	154	16066 16129 16172 16173 16223 16241	Q	16000	16500	+10397AluI, +5176AluI, +7598HhaI, -9824HinfI
KK	155	16140 16189 16217 16274 16335	B4c	16000	16400	+9bpdel
KK	157	16093 16129 16209 16223 16272	G	16000	16500	+10397AluI, +5176AluI, +7598HhaI, -9824HinfI

Series	No	HVS-I Variants	Haplogroup	HVS-I Start	HVS-I End	RFLP Variants
KK	158	16304	F	16000	16500	
KK	161	16092 16148 16189 16223 16362	D5	16010	16430	-10397AluI, -10394DdeI, -9bpdel, -5176AluI
KK	162	16189 16217 16261	B4a	16010	16430	+9bpdel
KK	163	16189 16217 16261	B4a	16010	16430	
KK	164	16086 16136 16189 16217	B4b1	16000	16460	+9bpdel
KK	167	16223 16234 16362	M	16000	16500	+10397AluI, +5176AluI, +7598HhaI, -9824HinfI
KK	168	16157 16256 16304 16335	F	16030	16500	
KK	169	16223 16295 16362	M7c1c	16000	16500	+10397AluI, +5176AluI, +7598HhaI, +9824HinfI
KK	171	16051 16223 16362 16390	E1	16010	16500	+10397AluI, +5176AluI, -7598HhaI
KK	172	16093 16223 16291 16295 16337 16362	M7c1c	16000	16500	+10397AluI, +5176AluI, +7598HhaI, +9824HinfI
KK	173	16129 16172 16304	F1a	16000	16500	
KK	174	16223 16295 16346C 16362	M7c1c	16000	16500	+10397AluI, +10394DdeI, +5176AluI, +7598HhaI, +9824HinfI
KK	176	16129 16172 16192 16294 16304 16362	F1a	16000	16500	
KK	178	16223 16304 16325 16344 16362 16381	M	16020	16500	+10397AluI, +10394DdeI, +7598HhaI, -9824HinfI
KK	179	16157 16256 16304 16335	F	16000	16500	
MED	3	16209 16223 16233 16274 16304	R9	16014	16497	-10397 AluI -10394 DdeI
MED	4	16126 16147 16153 16223	M3	16026	16500	+10397 AluI -9824 HinfI
MED	7	16140 16189 16243	B5b	16036	16398	-10397 AluI
MED	8	16170 16218 16304 16311	F4b	16012	16500	-10397 AluI -10394 DdeI
MED	19	16126 16231 16311	Y2	16038	16500	-10397 AluI -10394 DdeI
MED	21	16111 16140 16189 16234 16243	B5b	16038	16398	-10397 AluI
MED	47	16129 16189 16192 16223 16297	M7b1	16026	16500	+10397 AluI +9824 HinfI
MED	50	16111 16140 16189 16234 16243 16399	B5b	16011	16422	-10397 AluI
MED	51	16140 16189 16266A	B5a	16011	16432	-10397 AluI
MED	52	16223 16362	M	16026	16500	+10397 Alu I -9824 HinfI +5176 AluI
MED	53	16108 16129 16162 16172 16304	F1a1a	16026	16500	-10397 AluI -10394 DdeI
MED	54	16223 16362	M7	16026	16500	+10397 AluI +10394 DdeI +9824 HinfI
MED	56	16223 16362	M7	16026	16500	+10397 AluI +5176 AluI +9824 HinfI
MED	57	16131 16223 16261 16362 16390	E1b	16026	16500	+10397 AluI -9824 HinfI +5176 AluI -7598 HhaI
MED	59	16129 16189 16192 16223 16297	M7b1	16026	16500	+10397 AluI +10394 DdeI +9824 HinfI
MED	60	16189 16223 16290 16291 16362 16390	E1a	16075	16434	+10397 AluI -9824 HinfI +5176 AluI -7598 HhaI

Series	No	HVS-I Variants	Haplogroup	HVS-I Start	HVS-I End	RFLP Variants
MED	62	16223 16362	M7	16037	16500	+10397 AluI +5176 AluI +9824 HinfI
MED	64	16189 16192 16223 16291 16362	M	16033	16497	+10397 AluI -9824 HinfI +5176 AluI
MED	65	16129 16189 16192 16223 16297	M7b1	16053	16497	+10397 AluI +9824 HinfI
MED	67	16249 16288	R22	16026	16500	-10397 AluI -10394 DdeI -12308 HinfI
MED	75	16223 16292 16295 16362	M7c1c	16026	16500	+10397 AluI +9824 HinfI
MED	103	16223	M	16011	16500	+10397 AluI -9824 HinfI +5176 AluI
MED	104	16126 16231 16311	Y2	16011	16500	-10397 AluI +10394 DdeI
MED	105	16069 16231 16311	Y2	16012	16500	-10397 AluI +10397 DdeI +5176 AluI
MED	107	16147 16189 16217 16235	B4	16038	16398	-10397 AluI
MED	108	16185 16223 16295 16362	M7c1c	16024	16500	+10397 AluI +9824 HinfI
MED	109	16223 16288 16291 16362 16390	E1a	16026	16497	+10397 Alu -7598 HhaI
MED	110	16129 16189 16223 16297	M7b	16011	16412	+10397 AluI +9824 HinfI
MED	114	16126 16231 16311 (16042N)	Y2	16036	16500	-10397 AluI +10394 DdeI
MED	117	16223 16278	G2	16033	16500	+10397 AluI -9824 HinfI +5176 AluI
MED	119	16218 16304 16311	F4b	16026	16500	-10397 AluI -10394 DdeI
MED	121	16129 16172 16304	F1a	16026	16500	-10397 AluI -10394 DdeI
MED	131	16140 16189 16243	B5b	16019	16410	
MED	135	16126 16231 16311	Y2	16015	16371	-10397 AluI +10394 DdeI
MED	142	16093 16126 16231 16311	Y2	16024	16500	–10397 AluI
MED	144	16192 16362	M	16024	16500	+10397 AluI +7025 AluI -9824 HinfI +5176 AluI
MED	147	16145 16223 16295 16362	M7c1c	16024	16500	+10397 AluI -5176 AluI +9824 HinfI
MED	148	16172 16223 16239 16263 16325 16381	M21b	16026	16500	+10397 AluI -9824 HinfI
MED	150	16140 16189 16266A	B5a	16011	16423	-10397 AluI
MED	153	16223 16295 16362	M7c1c	16044	16500	+10397 AluI +9824 HinfI
MED	155	16093 16223 16231 16319	M2	16024	16500	+10397 AluI -9824 HinfI +5176 AluI
MED	183	16126 16231 16311	Y2	16037	16500	-10397 AluI +10394 DdeI
MND	1	16092 16148 16189 16223 16362	D5	16000	16460	-10397AluI, -10394DdeI, -9bpdel, -5176AluI
MND	2	16157 16256 16304 16335	F	16000	16500	
MND	3	16129 16172 16294 16304 16362 (16384N)	F1a	16000	16400	
MND	4	16223 16295 16362	M7c1c	16000	16465	+9824HinfI
MND	5	16148 16189 16223 16362	D5	16020	16430	-10397AluI, -9bpdel, -10394DdeI, -5176AluI
MND	6	16192 16234 16288 16304 16309 16390	R9b	16000	16500	

Series	No	HVS-I Variants	Haplogroup	HVS-I Start	HVS-I End	RFLP Variants
MND	7	16140 16189 16242A 16256 16261 16266A	B5a	16015	16460	+9bpdel
MND	8	16051 16223 16362 16390	E1	16000	16500	-7598HhaI
MND	9	16176 16221 16266 16325 16357	P	16000	16500	-10397AluI, -10394DdeI, -9bpdel , -10394DdeI, +15606AluI
MND	10	16189 16223 16291 16362 16390	E1a	16000	16440	-7598HhaI
MND	11	16249 16319 16390 (16120N)	P	16080	16500	-10397AluI, +10394DdeI, +5176AluI, +7598HhaI, -9824HinfI, +15606AluI
MND	12	16093 16209 16223 16224 16263 16265 16278 16319	G2	16000	16500	+10397AluI, +10394DdeI, +7598HhaI, -9824HinfI, +4831HhaI
MND	13	16223 16291 16362 16390	E1a	16000	16500	-7598HhaI
MND	14	16051 16184 16223 16362 16390	E1	16000	16500	-7598HhaI
MND	15	16092 16140 16189 16217 16274 16283T 16311 16335	B4c	16000	16460	+9bpdel
MND	16	16172 16223 16291 16362 16390	E1a	16000	16500	+10397AluI, +5176AluI, -7598HhaI
MND	17	16148 16189 16223 16362	D5	16020	16460	-10397AluI, -10394DdeI, -9bpdel, -5176AluI
MND	18	16140 16189 16243 16355	B5b	16040	16460	+9bpdel
MND	19	16092 16148 16189 16223 16362 (16120N)	D5	16040	16460	-10397AluI, -10394DdeI, -9bpdel, -5176AluI
MND	20	16188 16223 16274 16311 16362	D	16050	16500	+10397AluI, +10394DdeI, -5176AluI
MND	21	16189 16217 16247 16261	B4a1	16030	16440	
MND	22	16129 16144 16148 16172 16223 16241 16242 16265C 16311 16343	Q	16000	16500	
MND	23	16223 16257A 16261 16292 16294 16357	N9a1	16000	16500	
MND	24	16092 16148 16189 16223 16362	D5	16020	16435	-10397AluI, -10394DdeI, -5176AluI
MND	25	16051 16223 16362 16390	E1	16000	16500	+5176AluI, -7598HhaI
MND	26	16108 16129 16162 16172 16304	F1a1a	16010	16500	
MND	27	16092 16129 16148 16189 16223 16362	D5	16020	16460	-10397AluI, -10394DdeI, -5176AluI
MND	29	16086 16136 16189 16217	B4b1	16010	16435	+9bpdel
MND	30	16223 16261 16362 16390	E1b	16000	16500	+5176AluI, +10397AluI, -7598HhaI, +10394DdeI
MND	33	16223 16295 16362	M7c1c	16000	16500	+9824HinfI
MND	34	16129 16172 16304	F1a	16000	16500	
MND	36	16189 16217 16261 16311	B4a	16010	16430	+9bpdel
MND	37	16189 16217 16261 16288	B4a	16010	16415	+9bpdel

Series	No	HVS-I Variants	Haplogroup	HVS-I Start	HVS-I End	RFLP Variants
MND	38	16223 16295 16362	M7c1c	16030	16500	+9824HinfI
MND	39	16223 16295 16356 16362	M7c1c	16010	16500	+5176AluI, +7598HhaI, +9824HinfI
MND	40	16172 16223 16291 16362 16390	E1a	16010	16500	-7598HhaI
MND	41	16223 16291 16362 16390	E1a	16000	16500	+5176AluI, -7598HhaI
MND	42	16223 16291 16362 16390	E1a	16000	16500	
MND	43	16223 16295 16362	M7c1c	16000	16500	+9824HinfI
MND	44	16129 16172 16294 16304 16362	F1a	16000	16500	
MND	45	16051 16223 16292 16362 16390	E1	16000	16500	+5176AluI, -7598HhaI
MND	48	16192 16223 16274 16362 (16278N)	D	16000	16500	-5176AluI
MND	49	16093 16148 16189 16223 16362	D5	16000	16430	-10397AluI, -10394DdeI, -9bpdel, -5176AluI
MND	50	16223 16291 16362 16390	E1a	16000	16500	+5176AluI, -7598HhaI
MND	51	16223 16291 16362 16390	E1a	16000	16500	-7598HhaI
MND	52	16223 16261 16362 16390	E1b	16000	16500	+5176AluI, +10397AluI, -9824HinfI, +10394DdeI
MND	53	16223 16291 16362 16390	E1a	16000	16500	-7598HhaI
MND	54	16189 16284 16304	F1b?	16000	16460	-10397AluI, -10394DdeI, +5176AluI
MND	55	16223 16291 16362 16390	E1a	16010	16500	-9824HinfI
MND	57	16223 16291 16362 16390	E1a	16000	16500	-9824HinfI
MND	58	16223 16291 16362 16390	E1a	16030	16500	
MND	59	16129 16142 16166 16223 16255 16274 16294 16327A	G3?	16000	16500	+10397AluI, +10394DdeI, +5176AluI, +7598HhaI, -9824HinfI, +4831HhaI
MND	61	16129 16172 16304 16311	F1a	16030	16500	
MND	62	16140 16189 16217 16274 16335	B4c	16010	16435	+9bpdel
MND	63	16223 16291 16362 16390	E1a	16000	16500	-7598HhaI
MND	64	16126 16231 16311	Y2	16000	16500	-10397AluI, +10394DdeI
MND	65	16172 16223 16239 16263 16325 16381	M21b	16010	16500	+10397AluI, +10394DdeI
MND	66	16223 16291 16362 16390	E1a	16000	16500	
MND	67	16172 16223 16291 16362 16390	E1a	16000	16500	+5176AluI, -7598HhaI
MND	68	16223 16295 16356 16362	M7c1c	16000	16500	+7598HhaI, +9824HinfI
MND	69	16189 16217 16261	B4a	16010	16430	
MND	71	16223 16291 16362 16390	E1a	16000	16500	
MND	72	16172 16189 16223 16259 16362 (16164N)	D5	16010	16410	-10397AluI, -10394DdeI, -9bpdel, -5176AluI
MND	73	16192 16234 16288 16304 16309 16390	R9b	16050	16470	

Series	No	HVS-I Variants	Haplogroup	HVS-I Start	HVS-I End	RFLP Variants
MND	75	16140 16189 16243 16355	B5b	16000	16430	+9bpdel
MND	76	16148 16189 16223 16362	D5	16010	16435	-10397AluI, -10394DdeI, -9bpdel, -5176AluI
MND	77	16223 16295 16362	M7c1c	16000	16500	+10397AluI, +9824HinfI
MND	78	16147 16189 16217	B4	16010	16420	
MND	79	16118 16129 16192 16223 16256 16272	G	16040	16500	+5176AluI, +10397AluI, +7598HhaI, -9824HinfI, +10394DdeI, -15606AluI, +4831HhaI
MND	80	16148 16189 16223 16362	D5	16010	16410	-10397AluI, -10394DdeI, -9bpdel, -5176AluI
MND	81	16223 16291 16362 16390	E1a	16000	16500	
MND	82	16223 16295 16362	M7c1c	16000	16500	
MND	83	16223 16286 16362	D	16000	16500	+10394DdeI, -5176AluI
MND	84	16223 16224 16287 16295	M7c1	16010	16500	+5176AluI, +7598HhaI, +9824HInfI, -15606AluI
MND	87	16223 16291 16362 16390	E1a	16010	16500	
MND	88	16223 16295 16356 16362	M7c1c	16000	16500	+9824HinfI
MND	89	16223 16295 16362	M7c1c	16000	16500	+9824HinfI
MND	90	16140 16189 16266A	B5a	16000	16455	+9bpdel
MND	91	16140 16189 16266A	B5a	16000	16460	
MND	92	16223 16291 16362 16390	E1a	16000	16500	
MND	93	16223 16261 16362 16390	E1b	16000	16500	+10397AluI, +5176AluI, -9824HinfI, +10394DdeI, -7598HhaI
MND	94	16051 16223 16258C 16309 16362 16390	E1	16000	16500	+10397AluI, +10394DdeI, +5176AluI, -9824HinfI, -7598HhaI
MND	95	16148 16189 16223 16309 16362	D5	16000	16470	-9bpdel, -5176AluI
MND	96	16129 16172 16294 16304 16362	F1a	16000	16500	
MND	97	16176 16221 16266 16325 16357	P	16000	16500	+15606AluI
MND	98	16140 16189 16234 16243	B5b	16000	16420	+9bpdel
MND	99	16129 16172 16304 16311	F1a	16000	16500	
MND	100	16129 16172 16304 16311	F1a	16000	16500	
MND	101	16223 16295 16362	M7c1c	16000	16500	+7598HhaI
MTR	1	16223 16261 16362 16390	E1b	16040	16500	+10397AluI, +5176AluI, -9824HinfI`
MTR	3	16126 16189 16223 16291 16362 16390	E1a	16030	16480	+10397AluI, +5176AluI, -9824HinfI, -7598HhaI
MTR	5	16129 16189 16192 16223 16297	M7b1	16050	16500	+10397AluI, +10394DdeI, +5176AluI, +9824HinfI
MTR	6	16223 16261 16362 16390	E1b	16015	16500	+10397AluI, +5176AluI, -7598HhaI

Series	No	HVS-I Variants	Haplogroup	HVS-I Start	HVS-I End	RFLP Variants
MTR	7	16223 16261 16362 16390	E1b	16020	16500	
MTR	8	16189 16217 16247 16261	B4a1	16030	16400	+9bpdel
MTR	9	16223 16311 16335 16362	M	16030	16500	+10397AluI, -9824HinfI, +7598HhaI
MTR	11	16093 16189 16222 16223 16278 (16252N)	G2	16040	16430	+10397AluI, +10394DdeI, +5176AluI, +7598HhaI, +4831HhaI
MTR	18	16124 16166del 16214 16223	M	16030	16500	+10397AluI, +10394DdeI, +5176AluI, +7598HhaI
MTR	19	16129 16172 16214 16304 16311	F1a	16030	16500	
MTR	20	16129 16172 16304 16362 16381	F1a	16030	16500	
MTR	34	16288 16304	R22	16050	16500	
MTR	50	16108 16129 16162 16172 16304	F1a1a	16050	16500	
MTR	57	16126 16129 16192 16223 16297	M7b1	16045	16500	+10397AluI, +5176AluI, +7598HhaI
MTR	61	16108 16129 16162 16172 16304	F1a1a	16030	16500	
MTR	62	16189 16217 16261 (16156N)	B4a	16045	16470	+9bpdel
MTR	64	16219 16223 16290 16291	M	16030	16500	+10397AluI, +10394DdeI, +5176AluI, +7598HhaI
MTR	65	16140 16189 16243	B5b	16040	16470	+9bpdel
MTR	67	16189 16223 16278	G2	16030	16410	+10397AluI, +5176AluI, -9824HinfI
MTR	70	16140 16189 16266A	B5a	16030	16470	+9bpdel
MTR	71	16129 16172 16304 16362	F1a	16045	16500	
MTR	74	16129 16304 16362 16359 16390	F	16025	16500	
MTR	76	16249 16288 16304 16390	R22	16030	16500	
MTR	82	16192 16288 16304 16309	R9b	16030	16500	
MTR	83	16051 16223 16298 16327	C	16030	16500	+10397AluI, +10394DdeI, -9824HinfI, +7598HhaI
MTR	84	16223 16311	M	16020	16500	+10397AluI, -9824HinfI, +7598HhaI
MTR	85	16093 16129 16172 16294 16304 16362	F1a	16025	16500	
MTR	87	16129 16189 16192 16215 16223 16297	M7b1	16025	16500	+10397AluI, -9bpdel, +10394DdeI
MTR	90	16092 16148 16189 16223 16311 16362	D5	16030	16470	-9bpdel
MTR	91	16189 16217 16261	B4a	16045	16460	+9bpdel
MTR	93	16189 16222 16223 16278 16352	G2	16030	16470	-9bpdel, +7598HhaI, +10397AluI, +10394DdeI, +4831HhaI
MTR	95	16249 16288 16301 16304 16390	R22	16025	16500	
MTR	97	16108 16129 16162 16172 16304 16391	F1a1a	16030	16500	
MTR	99	16223 16295 16362	M7c1c	16015	16500	+10397AluI, +9824HinfI

Series	No	HVS-I Variants	Haplogroup	HVS-I Start	HVS-I End	RFLP Variants
MTR	100	16093 16189 16209 16223 (16318N)	M	16045	16470	+10397AluI, +5176AluI, +7598HhaI, -9824HinfI
MTR	103	16092 16140 16172 16189 16223 16278	G2	16015	16475	-9bpdel, +7598HhaI, +4831HhaI
MTR	105	16093 16133 16176 16223	M	16015	16500	+10397AluI, +5176AluI, +7598HhaI, -9824HinfI
MTR	112	16129 16172 16304 16362 16400	F1a	16060	16500	
MTR	118	16140 16189 16266A	B5a	16015	16480	+9bpdel
MTR	121	16249 16288 16304 16390 (16064N)	R22	16030	16500	
MTR	123	16129 16172 16304	F1a	16030	16500	
MTR	124	16140 16189 16266A	B5a	16015	16390	+9bpdel
MTR	125	16108 16129 16162 16172 16304	F1a1a	16060	16500	
MTR	147	16249 16288 16317C 16319	R22	16000	16500	-15606AluI
ORA	1	16140 16189 16243 16294 16354	B5b	16012	16422	–10397 AluI
ORA	2	16140 16189 16243 16294 16354	B5b	16020	16412	–10397 AluI
ORA	3	16140 16189 16243 16294 16354	B5b	16020	16414	–10397 AluI
ORA	4	16129 16223 16256 16271 16362	M21a	16013	16497	
ORA	5	16129 16223 16256 16271 16362	M21a	16025	16497	+10397AluI
ORA	6B	16140 16189 16243 16294 16354	B5b	16012	16422	–10397 AluI -10394 DdeI -4830HhaI
ORA	6A	16140 16189 16243 16294 16354	B5b	16012	16433	–10397 AluI -10394 DdeI +9bpdel
ORA	7B	16140 16189 16243 16294	B5b	16012	16422	–10397 AluI -10394 DdeI
ORA	7A	16108 16129 16162 16172	F1a1a	16012	16497	–10397 AluI
ORA	8A	16140 16189 16243 16294 16354	B5b	16012	16422	–10397 AluI
ORA	8B	16093 16129 16223 16256 16271 16362	M21a	16012	16497	+10397 AluI +10394 DdeI
ORA	9A	16093 16129 16217 16223 16256 16271 16362	M21a	16012	16497	+10397 AluI
ORA	9B	16093 16129 16223 16256 16271 16362	M21a	16012	16497	+10397 AluI +10394 DdeI
ORA	10B	16140 16189 16243 16294 16354	B5b	16012	16435	–10397 AluI -10394 DdeI
ORA	10A	16168 16295 16304	R21	16012	16497	-10397AluI -10394 DdeI
ORA	11A	16140 16189 16243 16294 16354	B5b	16012	16422	-10397
ORA	11B	16093 16129 16223 16256 16271 16362	M21a	16012	16497	+10397 AluI +10394 DdeI
ORA	12A	16093 16129 16223 16256 16271 16362	M21a	16012	16497	+10397 AluI +10394 DdeI
ORA	12B	16093 16129 16223 16256 16271 16362	M21a	16012	16497	+10397 AluI +10394 DdeI
ORA	13A	16093 16129 16223 16256 16271 16362	M21a	16012	16497	+10397 AluI +10394 DdeI
ORA	13B	16093 16129 16223 16256 16271 16362	M21a	16012	16497	+10397 AluI +10394 DdeI

Series	No	HVS-I Variants	Haplogroup	HVS-I Start	HVS-I End	RFLP Variants
ORA	14B	16140 16189 16243 16294 16354	B5b	16012	16430	-10397 AluI
ORA	14A	16108 16129 16162 16172	F1a1a	16012	16497	–10397 AluI
ORA	15A	16093 16129 16223 16256 16271 16362	M21a	16012	16497	+10397 AluI +10394 DdeI
ORA	16A	16140 16189 16243 16294 16354	B5b	16012	16422	-10397 AluI
ORA	17A	16140 16189 16243 16294 16354	B5b	16012	16431	-10397 AluI
ORA	17B	16108 16129 16162 16172 16304	F1a1a	16012	16497	–10397 AluI -10394 DdeI
ORA	18A	16093 16129 16223 16256 16271 16362	M21a	16012	16497	+10397 AluI +10394 DdeI
ORA	18B	16086 16168 16295 16296 16304	R21	16012	16497	-10397 AluI -10394 DdeI
ORA	19A	16140 16189 16243 16294 16354	B5b	16012	16431	-10397 AluI
ORA	19B	16086 16168 16295 16296 16304	R21	16012	16497	-10397 AluI -10394 DdeI
ORA	20B	16108 16129 16162 16172 16304	F1a1a	16012	16497	–10397 AluI -10394 DdeI
ORA	20A	16093 16129 16223 16256 16271 16362	M21a	16012	16497	+10397 AluI +10394 DdeI
ORA	21A	16093 16129 16223 16256 16271 16362	M21a	16012	16497	+10397 AluI +10394 DdeI
ORA	21B	16086 16168 16295 16296 16304	R21	16012	16497	-10397 AluI -10394 DdeI
ORA	22A	16108 16129 16162 16172 16304	F1a1a	16012	16497	–10397 AluI -10394 DdeI
ORA	22B	16168 16209 16295 16296 16304	R21	16012	16497	-10397 AluI -10394 DdeI
ORA	23B	16108 16129 16162 16172 16304	F1a1a	16012	16497	–10397 AluI -10394 DdeI
ORA	23A	16086 16168 16295 16296 16304	R21	16012	16497	-10397 AluI -10394 DdeI
ORA	24B	16108 16129 16162 16172 16304	F1a1a	16012	16497	–10397 AluI -10394 DdeI
ORA	24A	16086 16168 16295 16296 16304	R21	16012	16497	-10397 AluI -10394 DdeI
ORA	25A	16108 16129 16162 16172 16304	F1a1a	16012	16497	–10397 AluI
ORA	26A	16086 16168 16295 16296 16304	R21	16012	16497	-10397 AluI -10394 DdeI
ORA	27B	16108 16129 16162 16172 16304	F1a1a	16012	16497	–10397 AluI -10394 DdeI
ORA	27A	16086 16168 16295 16296 16304	R21	16012	16497	-10397 AluI -10394 DdeI
ORA	28B	16108 16129 16162 16172 16304	F1a1a	16012	16497	–10397 AluI -10394 DdeI
ORA	28A	16086 16168 16295 16296 16304	R21	16012	16497	-10397 AluI -10394 DdeI
ORA	29A	16108 16129 16162 16172 16304	F1a1a	16012	16497	–10397 AluI
ORA	29B	16086 16223	M	16012	16497	+10397 AluI +10394 DdeI +9052 HhaI -5351 HhaI -10054 HinfI -10143 AluI +5176 AluI -9824 HinfI +7598 HhaI -4830 HhaI
ORA	30B	16108 16129 16162 16172	F1a1a	16012	16497	–10397 AluI -10394 DdeI
ORA	30A	16108 16129 16162 16172 16304	F1a1a	16012	16497	–10397 AluI

Series	No	HVS-I Variants	Haplogroup	HVS-I Start	HVS-I End	RFLP Variants
ORA	31A	16086 16168 16295 16296 16304	R21	16012	16497	-10397 AluI -10394 DdeI
ORA	31B	16168 16209 16295 16296 16304	R21	16012	16497	-10397 AluI -10394 DdeI
ORA	32A	16086 16168 16295 16296 16304	R21	16012	16497	-10397 AluI -10394 DdeI
ORA	33A	16108 16129 16162 16172 16304	F1a1a	16012	16497	–10397 AluI -10394 DdeI
ORA	33B	16108 16129 16162 16172 16304	F1a1a	16012	16497	–10397 AluI -10394 DdeI
ORA	34A	16108 16129 16162 16172 16304	F1a1a	16012	16497	–10397 AluI -10394 DdeI
ORA	35B	16093 16129 16223 16256 16271 16362	M21a	16012	16497	+10397 AluI +10394 DdeI
ORA	35A	16168 16209 16295 16296 16304	R21	16012	16497	-10397 AluI
ORA	36B	16108 16129 16162 16172	F1a1a	16012	16497	–10397 AluI
ORA	36A	16223 16257A 16261 16292 16294	N9a1	16012	16497	-10397 AluI
ORA	37A	16086 16223	M	16017	16497	+10397 AluI +10394 DdeI +9052 HhaI -5351 HhaI -10054 HinfI -10143 AluI +5176 AluI -9824 HinfI +7598 HhaI -4830 HhaI
ORA	38A	16108 16129 16162 16172	F1a1a	16012	16497	–10397 AluI -10394 DdeI
ORA	38B	16108 16129 16162 16172 16304	F1a1a	16012	16497	–10397 AluI
ORA	39A	16108 16129 16162 16172	F1a1a	16012	16497	–10397 AluI
ORA	39B	16093 16129 16223 16256 16271 16362	M21a	16012	16497	+10397 AluI +10394 DdeI
ORA	40B	16093 16129 16223 16256 16271 16362	M21a	16012	16497	+10397 AluI +10394 DdeI
ORA	40A	16168 16209 16295 16296 16304 (16072-16074N)	R21	16012	16497	-10397 AluI
ORA	41A	16093 16129 16223 16256 16271 16362	M21a	16012	16497	+10397 AluI +10394 DdeI
ORA	41B	16093 16129 16223 16256 16271 16362	M21a	16012	16497	+10397 AluI +10394 DdeI
ORA	42B	16140 16189 16243 16294 16354	B5b	16012	16422	-10397 AluI -10394 DdeI
ORA	42A	16168 16295 16296 16304	R21	16012	16497	-10397 AluI +10394 DdeI
ORA	43B	16093 16129 16223 16256 16271 16362	M21a	16012	16497	+10397 AluI +10394 DdeI -9bpdel
ORA	43A	16168 16295 16296 16304	R21	16012	16497	-10397 AluI +10394 DdeI
ORA	44A	16093 16129 16217 16223 16256 16271 16362	M21a	16012	16497	
ORA	44B	16168 16295 16296 16304	R21	16012	16497	-10397 AluI -10394 DdeI
ORA	45B	16093 16129 16223 16256 16271 16362	M21a	16012	16497	+10397 AluI +10394 DdeI
ORA	45A	16168 16295 16296 16304	R21	16012	16497	-10397 AluI
ORA	46B	16140 16189 16243 16294 16354	B5b	16012	16424	-10397 AluI -10394 DdeI

Series	No	HVS-I Variants	Haplogroup	HVS-I Start	HVS-I End	RFLP Variants
ORA	46A	16093 16129 16223 16256 16271 16362	M21a	16012	16497	
ORA	47B	16093 16129 16223 16263 16381	M21b	16012	16497	+10397 AluI +10394 DdeI
ORA	47A	16168 16295 16296 16304	R21	16012	16497	-10397 AluI +10394 DdeI
ORA	48A	16093 16129 16223 16256 16271 16362	M21a	16012	16497	
ORA	48B	16093 16129 16223 16256 16271 16362	M21a	16012	16497	+10397 AluI +10394 DdeI
ORA	49A	16093 16129 16223 16256 16271 16293 16362	M21a	16047	16497	+10397 AluI +10394 DdeI
ORA	49B	16093 16129 16223 16256 16271 16362	M21a	16012	16497	+10397 AluI
ORA	50B	16140 16189 16266A	B5a	16012	16422	–10397 AluI
ORA	51B	16093 16129 16217 16223 16256 16271 16362	M21a	16012	16497	
ORA	51A	16093 16129 16223 16256 16271 16362	M21a	16012	16497	+10397 AluI +10394 DdeI
ORA	52B	16093 16129 16223 16256 16271 16362	M21a	16012	16497	+10397 AluI
ORA	52A	16086 16168 16295 16296 16304	R21	16012	16497	-10397 AluI -10394 DdeI
ORA	53B	16093 16129 16217 16223 16256 16271 16362	M21a	16012	16497	+10397 AluI +10394 DdeI
ORA	53A	16168 16295 16296 16304	R21	16012	16497	-10397 AluI
ORA	55B	16093 16129 16217 16223 16256 16271 16362	M21a	16012	16497	+10397 AluI +10394 DdeI
ORA	56B	16093 16129 16223 16256 16271 16362	M21a	16012	16497	+10397 AluI +10394 DdeI
ORA	56A	16093 16129 16223 16256 16271 16362	M21a	16012	16497	
ORA	58A	16093 16129 16145 16223 16256 16271 16362	M21a	16012	16497	+10397 AluI +10394 DdeI
ORA	58B	16086 16168 16295 16296 16304	R21	16012	16497	-10397 AluI
ORA	59B	16189 16223 16229 16294 16311 16362	M	16012	16422	+10397 AluI +10394 DdeI +9052 HhaI -5351 HhaI -10054 HinfI -10143 AluI +5176 AluI -9824 HinfI +7598 HhaI -4830 HhaI
ORA	60B	16140 16189 16266A	B5a	16012	16422	–10397 AluI
ORA	60A	16093 16129 16223 16256 16271 16362	M21a	16012	16497	+10397 AluI +10394 DdeI
ORA	61A	16093 16129 16223 16256 16271 16362	M21a	16012	16497	+10397 AluI +10394 DdeI
ORA	61B	16086 16168 16295 16296 16304	R21	16012	16497	-10397 AluI
ORA	62B	16093 16129 16223 16256 16271 16362	M21a	16012	16497	

Series	No	HVS-I Variants	Haplogroup	HVS-I Start	HVS-I End	RFLP Variants
ORA	62A	16093 16129 16223 16256 16271 16362	M21a	16012	16497	+10397 AluI +10394 DdeI
ORA	63B	16093 16129 16217 16223 16256 16271 16362	M21a	16030	16497	+10397 AluI +10394 DdeI
ORA	63A	16093 16129 16223 16256 16271 16362	M21a	16012	16497	+10397 AluI +10394 DdeI
ORA	64B	16093 16129 16223 16263 16381	M21b	16012	16497	+10397 AluI +10394 DdeI
ORA	65A	16093 16129 16223 16256 16271 16362	M21a	16012	16497	+10397 AluI +10394 DdeI
ORA	65B	16093 16129 16223 16256 16271 16362	M21a	16012	16497	+10397 AluI +10394 DdeI
ORA	66B	16093 16129 16223 16256 16271 16362	M21a	16022	16497	+10397 AluI +10394 DdeI
ORA	67A	16086 16168 16295 16296 16304	R21	16012	16497	-10397 AluI
ORA	67B	16086 16168 16295 16296 16304	R21	16022	16497	-10397 AluI
ORA	68A	16086 16168 16295 16296 16304	R21	16012	16497	
ORA	68B	16086 16168 16295 16296 16304	R21	16012	16497	-10397 AluI
ORA	69A	16086 16168 16295 16296 16304	R21	16012	16497	
ORA	69B	16086 16168 16295 16296 16304	R21	16012	16497	-10397 AluI
ORA	70A	16093 16129 16223 16256 16271 16362	M21a	16012	16497	
ORA	70B	16086 16168 16295 16296 16304	R21	16012	16497	-10397 AluI
ORA	71B	16129 16145 16223 16256 16271 16362	M21a	16031	16497	+10397 AluI +10394 DdeI
ORA	71A	16086 16168 16295 16296 16304	R21	16012	16497	
ORA	72B	16093 16129 16223 16256 16271 16362	M21a	16012	16497	+10397AluI
ORA	73A	16093 16129 16223 16256 16271 16362	M21a	16012	16497	
ORA	73B	16168 16295 16296 16304	R21	16013	16497	-10397 AluI -10394 DdeI
ORA	75B	16223 16257A 16261 16292 16294	N9a1	16004	16497	-10397 AluI
ORA	76B	16093 16129 16223 16256 16271 16362	M21a	16012	16497	+10397 AluI +10394 DdeI
ORA	76A	16168 16295 16296 16304	R21	16012	16497	-10394 DdeI
ORA	77A	16168 16295 16296 16304	R21	16012	16497	–10397 AluI
ORA	77B	16168 16295 16296 16304	R21	16012	16497	-10397 AluI +10394 DdeI
ORA	78A	16093 16129 16223 16263 16381	M21b	16012	16497	+10397 AluI +10394 DdeI
ORA	78B	16168 16295 16296 16304	R21	16012	16497	-10397 AluI
ORA	79A	16093 16129 16223 16263 16381	M21b	16012	16497	+10397 AluI +10394 DdeI
ORA	80B	16223 16257A 16261 16292 16294	N9a1	16012	16497	-10397 AluI
ORA	81A	16168 16295 16296 16304	R21	16012	16497	-10397 AluI +10394 DdeI
ORA	82B	16223 16257A 16261 16292 16294	N9a1	16013	16497	-10397 AluI

Series	No	HVS-I Variants	Haplogroup	HVS-I Start	HVS-I End	RFLP Variants
ORA	82A	16086 16168 16295 16296 16304	R21	16012	16497	-10397 AluI
ORA	83B	16168 16295 16296 16304	R21	16004	16497	-10397 AluI
ORA	84B	16086 16168 16295 16296 16304	R21	16012	16497	-10397 AluI
ORA	84A	16168 16295 16296 16304	R21	16013	16497	-10397 AluI
ORA	85A	16086 16168 16295 16296 16304	R21	16012	16497	
ORA	85B	16086 16168 16295 16296 16304	R21	16069	16500	
ORA	86B	16223 16257A 16261 16292 16294	N9a1	16034	16497	-10397 AluI
ORA	87B	16086 16168 16295 16296 16304	R21	16013	16497	-10397 AluI
ORA	87A	16168 16295 16296 16304	R21	16013	16497	-10397 AluI
ORA	88A	16223 16257A 16261 16292 16294	N9a1	16012	16497	
ORA	88B	16223 16257A 16261 16292 16294	N9a1	16020	16497	-10397 AluI
ORA	89A	16223 16257A 16261 16292 16294	N9a1	16012	16497	-10397 AluI
ORA	89B	16168 16295 16296 16304	R21	16022	16497	-10397 AluI
ORA	90B	16129 16223 16256 16271 16362	M21a	16039	16497	+10397 AluI +5176 AluI -9824 HinfI -5351 HhaI +7598 HhaI -4830 HhaI +9052 HhaI -10054 HinfI -10143 AluI
ORA	90A	16223 16257A 16261 16292 16294	N9a1	16012	16497	-10397 AluI
ORA	91A	16086 16168 16295 16296 16304	R21	16012	16497	-10397 AluI
ORA	92A	16168 16295 16296 16304	R21	16012	16497	-10397 AluI
ORA	93A	16093 16129 16223 16256 16271 16362	M21a	16012	16497	+10397 AluI
ORA	94A	16223 16257A 16261 16292 16294	N9a1	16012	16497	-10397 AluI
ORA	95A	16168 16295 16296 16304	R21	16013	16497	-10397 AluI
ORA	96A	16168 16295 16296 16304	R21	16013	16497	-10397 AluI
ORA	97A	16168 16295 16296 16304	R21	16013	16497	-10397 AluI
ORA	98A	16168 16295 16296 16304	R21	16013	16497	-10397 AluI
ORA	99A	16168 16295 16296 16304	R21	16012	16497	-10397 AluI
ORA	100B	16223 16295 16362	M7c1c	16013	16497	+10397AluI +9824 HinfI
ORA	100A	16086 16223 16288 16304 16309 16390	R9b	16025	16497	–10397 AluI
ORA	101B	16193 16291	N21	16038	16497	-10397 AluI
ORA	101A	16086 16170 16223 16288 16304 16309 16390	R9b	16031	16497	-10397 AluI
ORA	102A	16108 16129 16162 16172 16189 16304	F1a1a	16033	16420	–10397 AluI -10394 DdeI

Series	No	HVS-I Variants	Haplogroup	HVS-I Start	HVS-I End	RFLP Variants
ORA	102B	16086 16170 16223 16288 16304 16309 16390	R9b	16012	16497	
ORA	103B	16108 16129 16162 16172 16189 16304	F1a1a	16012	16423	–10397 AluI
ORA	104A	16172 16223 16239 16263 16325 16381	M21b	16019	16497	+10397AluI +5176 AluI -9824 HinfI -5351 HhaI +7598 HhaI -4830 HhaI +9052 HhaI -10054 HinfI -10143 AluI
ORA	104B	16086 16170 16223 16288 16304 16309 16390	R9b	16012	16497	–10397 AluI
ORA	105A	16129 16189 16217 16261 (16216N)	B4a	16031	16411	–10397 AluI
ORA	105B	16193 16291	N21	16012	16497	–10397 AluI
ORA	106A	16223 16295 16362	M7c1c	16027	16497	+10397AluI +9824 HinfI
ORA	106B	16086 16170 16223 16288 16304 16309 16390	R9b	16012	16497	–10397 AluI
ORA	107A	16093 16129 16223 16256 16271	M21a	16012	16497	+10397 AluI
ORA	107B	16086 16170 16223 16288 16304 16309 16390	R9b	16039	16497	-10397 AluI
ORA	108B	16193 16291	N21	16012	16497	–10397 AluI
ORA	108A	16170 16223 16288 16304 16309 16390	R9b	16013	16497	–10397 AluI
ORA	109A	16108 16129 16162 16172 16189 16304	F1a1a	16012	16422	–10397 AluI
ORA	109B	16086 16170 16223 16288 16304 16309 16390	R9b	16012	16497	–10397 AluI
ORA	110A	16129 16189 16217 16261	B4a	16012	16423	–10397 AluI
ORA	110B	16193 16291 16327	N21	16012	16497	–10397 AluI
ORA	111A	16193 16291 16327	N21	16012	16497	-10397 AluI
ORA	112B	16223 16242 16319	M21c	16012	16497	+10397 AluI +5176 AluI -9824 HinfI -5351 HhaI +7598 HhaI -4830 HhaI +9052 HhaI -10054 HinfI -10143 AluI
ORA	112A	16086 16170 16223 16288 16304 16309 16390	R9b	16012	16497	–10397 AluI
ORA	113A	16223 16295 16362	M7c1c	16012	16497	+10397 AluI +5176 AluI +9824 HinfI
ORA	113B	16223 16295 16362	M7c1c	16012	16497	+10397 AluI +5176 AluI +9824 HinfI
ORA	114B	16193 16291 16327	N21	16012	16497	–10397 AluI
ORA	114A	16193 16291 16327	N21	16019	16497	-10397 AluI

Series	No	HVS-I Variants	Haplogroup	HVS-I Start	HVS-I End	RFLP Variants
ORA	115A	16193 16291 16327	N21	16022	16497	-10397 AluI
ORA	116A	16193 16291	N21	16012	16497	-10397 AluI
ORA	117A	16223 16295 16362	M7c1c	16031	16497	+10397AluI +9824 HinfI
ORA	118A	16086 16170 16223 16288 16304 16309 16390	R9b	16012	16497	–10397 AluI
ORA	118B	16086 16170 16223 16288 16304 16309 16390	R9b	16012	16497	–10397 AluI
ORA	119A	16223 16295 16362	M7c1c	16012	16497	+10397AluI +9824 HinfI
ORA	119B	16086 16170 16223 16288 16304 16309 16390	R9b	16012	16497	–10397 AluI
ORA	120A	16086 16170 16223 16288 16304 16309 16390	R9b	16012	16497	–10397 AluI
ORA	121A	16223 16242 16319	M21c	16012	16497	+10397 AluI +9052 HhaI -5351 Hha -10054 HinfI -10143 AluI +5176 AluI -9824 HinfI +7598 HhaI -4830 HhaI
ORA	121B	16193 16291	N21	16012	16497	-10397 AluI
ORA	122A	16223 16295 16362	M7c1c	16012	16497	+10397 AluI +9824 HinfI
ORA	122B	16193 16291	N21	16012	16497	-10397 AluI
ORA	123A	16172 16223 16239 16263 16325 16381	M21b	16012	16497	+10397 AluI
ORA	124B	16093 16189 16223 16274 16278 16311	G	16013	16430	+10397 AluI +4831 HhaI
ORA	124A	16086 16170 16223 16288 16304 16309 16390	R9b	16012	16497	–10397 AluI
ORA	125B	16172 16223 16239 16263 16325 16381	M21b	16032	16497	+10397AluI +5176 AluI -9824 HinfI -5351 HhaI +7598 HhaI -4830 HhaI +9052 HhaI -10054 HinfI -10143 AluI
ORA	125A	16223 16295 16362	M7c1c	16012	16497	+10397AluI +9824 HinfI
ORA	126B	16093 16184 16223 16290 16304	M22	16013	16497	+10397 AluI +9052 HhaI -5351 HhaI -10054 HinfI -10143 AluI +5176 AluI
ORA	126A	16086 16170 16223 16288 16304 16309 16390	R9b	16012	16497	–10397 AluI
ORA	127B	16108 16129 16162 16172 16189 16304	F1a1a	16012	16422	–10397 AluI
ORA	127A	16093 16129 16223 16256 16271	M21a	16012	16497	+10397 AluI
ORA	128B	16193 16291	N21	16013	16497	-10397 AluI

Series	No	HVS-I Variants	Haplogroup	HVS-I Start	HVS-I End	RFLP Variants
ORA	128A	16086 16170 16223 16288 16304 16309 16390	R9b	16012	16497	–10397 AluI
ORA	129B	16086 16170 16223 16288 16304 16309 16390	R9b	16012	16497	-10397 AluI -10394 DdeI
ORA	130B	16086 16170 16223 16288 16304 16309 16390	R9b	16013	16497	–10397 AluI
ORA	131A	16193 16291 16327	N21	16012	16497	-10397 AluI
ORA	131B	16129 16288 16304 16309 16390	R9b	16012	16497	-10397 AluI
ORA	132B	16193 16291	N21	16013	16497	-10397 AluI
ORA	132A	16193 16291 16327	N21	16012	16497	-10397 AluI -9bpdel
ORA	133A	16093 16184 16223 16290 16304	M22	16069	16497	
ORA	133B	16192 16288 16304 16309 16390	R9b	16012	16497	–10397 AluI
ORA	134A	16193 16291	N21	16012	16497	
ORA	134B	16193 16291	N21	16012	16497	-10397 AluI
ORA	135B	16051 16189 16274	B	16012	16412	–10397 AluI +9bpdel
ORA	135A	16017 16168 16223 16249	N22	16012	16497	-9bpdel
ORA	136A	16108 16129 16147 16162 16172 16304	F1a1a	16012	16497	–10397 AluI
ORA	136B	16179 16223 16257A 16261 16292 16294	N9a1	16012	16497	
ORA	137A	16051 16189	B	16019	16401	–10397 AluI +9bpdel
ORA	137B	16193 16291	N21	16012	16497	-10397 AluI
ORA	138A	16093 16129 16223 16256 16271 16362	M21a	16047	16497	+10397 AluI
ORA	138B	16193 16291	N21	16013	16497	-10397 AluI
ORA	139B	16223 16257A 16261 16292 16294 16304	N9a1	16013	16497	-10397 AluI
ORA	140A	16295 16319	M7c1a	16030	16497	+10397 AluI +9824 HinfI -5351 HhaI
ORA	140B	16193 16291	N21	16047	16497	
ORA	141B	16093 16184 16223 16290 16304	M22	16013	16497	+10397 AluI +9052 HhaI -5351 HhaI -10054 HinfI -10143 AluI +5176 AluI
ORA	142B	16168 16223 16249 (16056N 16061N)	N22	16047	16497	-10397 AluI
ORA	143B	16193 16291	N21	16012	16497	
ORA	143A	16223 16257A 16261 16292 16294	N9a1	16032	16497	-10397 AluI
ORA	144B	16086 16170 16223 16288 16304 16309 16390	R9b	16012	16497	–10397 AluI

Series	No	HVS-I Variants	Haplogroup	HVS-I Start	HVS-I End	RFLP Variants
ORA	145B	16193 16291 (16380-16383N)	N21	16012	16497	-10397 AluI
ORA	145A	16223 16257A 16261 16292 16294	N9a1	16060	16497	-10397 AluI
ORA	146A	16093 16184 16223 16290 16304	M22	16012	16497	+10397 AluI
ORA	146B	16086 16170 16223 16288 16304 16309 16390	R9b	16012	16497	–10397 AluI
ORA	147A	16223 16257A 16261 16292 16294 16304	N9a1	16012	16497	-10397 AluI
ORA	147B	16086 16170 16223 16288 16304 16309 16390	R9b	16012	16497	–10397 AluI
ORA	148B	16193 16291	N21	16013	16497	-10397 AluI
ORA	148A	16017 16075 16168 16223 16249	N22	16012	16497	-10397 AluI -9bpdel
ORA	149B	16093 16184 16223 16290 16304	M22	16013	16497	+10397 AluI
ORA	150A	16172 16223 16239 16263 16325 16381	M21b	16012	16497	+10397 AluI
ORA	150B	16170 16223 16288 16304 16309 16390	R9b	16047	16497	–10397 AluI
ORA	151A	16051 16189	B	16012	16422	–10397 AluI +9bpdel
ORA	151B	16136 16217 16223 16319 16381	M21b	16018	16497	+10397 AluI +5176 AluI -9824 HinfI -5351 HhaI +7598 HhaI -4830 HhaI +9052 HhaI -10054 HinfI -10143 AluI
ORA	152B	16193 16291	N21	16012	16497	-10397 AluI -9bpdel
ORA	153B	16193 16291	N21	16032	16497	-10397 AluI
ORA	154A	16086 16170 16223 16288 16304 16309 16390	R9b	16012	16497	–10397 AluI
ORA	156B	16172 16223 16239 16263 16325 16381	M21b	16043	16497	+10397 AluI
ORA	157A	16086 16223 16249A	M	16012	16497	+10397 AluI
ORA	158A	16017 16075 16168 16223 16249	N22	16012	16497	-10397 AluI -9bpdel
ORA	159B	16108 16129 16162 16172 16274	F1a1a	16013	16497	–10397 AluI
ORA	160B	16108 16129 16162 16172 16304	F1a1a	16012	16497	–10397 AluI
ORA	160A	16223 16257A 16261 16292 16294	N9a1	16012	16497	-10397 AluI
ORA	162A	16108 16129 16162 16172 16304	F1a1a	16012	16497	–10397 AluI
PAD	3	16129 16172 16304	F1a	16040	16500	
PAD	14	16172 16189 16223 16249 16290	?	16030	16400	-9bpdel, +5176AluI, -12308HinfI
PAD	15	16172 16189 16304	F1a	16060	16420	-9bpdel
PAD	18	16129 16172 16301 16304 16400	F1a	16030	16500	

Series	No	HVS-I Variants	Haplogroup	HVS-I Start	HVS-I End	RFLP Variants
PAD	19	16189 16195 16286	B	16070	16470	+9bpdel, -10397AluI, -10394DdeI
PAD	20	16223 16362 (16294N)	M	16050	16500	+10397AluI, +10394DdeI, -9824HinfI, +5176AluI
PAD	27	16189 16217 16261 (16054N)	B4a	16030	16410	+9bpdel
PAD	29	16129 16209 16223 16272	G	16030	16500	+10397AluI, +5176AluI, -9824HinfI
PAD	36	16189 16223 (16126N het? 16311N)	N	16015	16475	-9bpdel, -10397AluI, -10394del, +5176AluI, -15606AluI
PAD	38	16129 16172 16304 16311	F1a	16020	16500	
PAD	39	16189 16223 16257A 16261 16292	N9a1	16045	16420	-9bpdel, -10397AluI, +5176AluI
PAD	78	16051 16189 16194C 16195	B	16030	16410	-10397AluI, +9bpdel
PAD	79	16189 16223 16257A 16261 16292	N9a1	16020	16390	-9bpdel
PAD	82	16051 16189 16194C 16195 (16039N 16215N)	B	16015	16410	-10397AluI, +9bpdel
PAD	94	16223 16291 16362 16390	E1a	16020	16500	+10397AluI, +5176AluI, -9824HinfI, -7598HhaI
PAD	100	16223 16362	M7	16050	16500	+5176AluI, +10397AluI, +10394DdeI, +7598HhaI, +9824HinfI
PAD	104	16129 16134 16172 16301 16304 16400	F1a	16030	16500	
PAD	105	16189 16223 16257A 16261 16292	N9a1	16060	16385	-9bpdel, -10397AluI, -10394DdeI, +5176AluI
PAD	108	16140 16188 16189 16217 16274 16311 16335	B4c	16030	16500	+9bpdel
PAD	110	16129 16172 16304	F1a	16030	16500	
PAD	112	16223 16261 16362 16390	E1b	16030	16500	+10397AluI, -9824HinfI, -7598HhaI
PAD	114	16189 16217 16261	B4a	16025	16410	-10397AluI, +9bpdel
PAD	116	16192 16288 16304 16309 16390	R9b	16030	16500	
PAD	117	16189 16194C 16195	B	16060	16470	-10397AluI
PAI	3	16223 16295 16362	M7c1c	16015	16500	+10397AluI, +9824HinfI
PAL	1	16126 16231 16311 (16159N)	Y2	16060	16500	-10397AluI, +10394DdeI
PAL	8	16223 16362 (16064N 16102N)	M7	16030	16500	+10397Alu, +5176AluI, +9824HinfI
PAL	9	16108 16129 16162 16172 16304	F1a1a	16030	16500	
PAL	18	16223 16362	M7	16015	16470	+10397Alu, +10394DdeI, +5176AluI, +9824HinfI
PAL	29	16192 16223 16274 16362	D	16030	16500	+10397Alu, +10394DdeI, -5176AluI
PAL	30	16193 16223	N21	16030	16500	-10397AluI, -10394DdeI, +5176AluI, +7598HhaI
PAL	34	16129 16172 16294 16304 16362 (16159N)	F1a	16040	16500	

Series	No	HVS-I Variants	Haplogroup	HVS-I Start	HVS-I End	RFLP Variants
PAL	36	16172 16173 16223 16278 16311	G2	16060	16500	+10397Alu, +10394DdeI, +5176AluI, -9824HinfI, +7598HhaI, +4831HhaI
PAL	37	16223 16299 16311	M	16010	16500	+10397Alu, +10394DdeI, +5176AluI, -9824HinfI, +7598HhaI
PAL	39	16223 16261 16362 16390	E1b	16000	16500	+10397Alu, +5176AluI, -9824HinfI, -7598HhaI
PAL	42	16223 16295 16362 (16152N 16159N)	M7c1c	16030	16500	+10397Alu, +9824HinfI, +10394DdeI
PAL	43	16223 16295 16362	M7c1c	16050	16450	+10397AluI, +10394DdeI
PAL	52	16140 16189 16243	B5b	16050	16470	+9bpdel
PAL	54	16215 16223 16362 16390	E1	16055	16500	+10397Alu, +10394DdeI, +5176AluI, -9824HinfI, -7598HhaI
PAL	55	16223 16257A 16261 16292 (16081N 16152N 16159N)	N9a1	16060	16500	
PAL	56	16129 16189 16192 16223 16297	M7b1	16070	16400	+9824HinfI
PAL	58	16223 16274 16362	D	16060	16500	+10397Alu, -5176AluI
PAL	59	16196 16223 16274 16278 16290 (16159N)	G2?	16030	16500	+10397AluI, +10394DdeI, +5176AluI, +7598HhaI, -9824HinfI, +4831HhaI
PAL	61	16223 16295 16362	M7c1c	16030	16475	+10397Alu, +9824HinfI
PAL	63	16223 16311 16362	M	16040	16500	+10397AluI, +10394DdeI
PAL	64	16129 16172 16304 16311 16362	F1a	16020	16500	
PAL	74	16223 16291 16362 16390 (16159N)	E1a	16030	16500	
PAL	75	16223 16291 16362 16390	E1a	16030	16500	
PAL	80	16086 16157 16256 16304 16335	F	16000	16500	
PAL	95	16185 16223 16291 16362 16390	E1a	16010	16500	-9824HinfI, -7598HhaI
PAL	99	16093 16223 16261 16311 16362 16390	E1b	16030	16500	+10397Alu, +5176AluI, -9824HinfI, -7598HhaI
PAL	107	16048 16162del 16214 16223	M	16030	16470	+10397AluI, +7598HhaI, +10394DdeI, -9824HinfI, +5176AluI
PAL	108	16093 16223 16249 16259 16278 16291 16362	G2	16030	16500	+10397Alu, +10394DdeI, +5176AluI, -9824HinfI, +7598HhaI, +4831HhaI
PAL	109	16136 16189 16217 16300	B4b1	16030	16400	+9bpdel
PAL	142	16223 16295 16362	M7c1c	16050	16500	+9824HinfI
PAL	145	16223 16362	M7	16050	16500	+10397Alu, +5176AluI, +9824HinfI
PAL	152	16129 16172 16304 16309	F1a	16000	16500	

Series	No	HVS-I Variants	Haplogroup	HVS-I Start	HVS-I End	RFLP Variants
PAL	155	16223 16311 16362	M	16050	16480	+10397Alu, +5176AluI, -9824HinfI, +7598HhaI
PAL	162	16129 16145 16162 16172 16304	F1a1	16040	16500	
PAL	166	16223 16291 16362 16390	E1a	16030	16500	+10397AluI, +10394DdeI, +5176AluI, -9824HinfI, -7598HhaI
PAL	167	16223 16295 16362	M7c1c	16030	16500	
PAL	170	16223 16278 16295 16362	M7c1c	16040	16500	+10397Alu, +10394DdeI, +9824HInfI
PAL	175	16223 16295 16362	M7c1c	16000	16500	
PEK	4	16223 16246T 16311 16362	M	16013	16500	+10397 AluI -9824 HinfI +5176 AluI
PEK	6	16179 16223 16295 16362	M7c1c	16011	16500	+10397 AluI +9824 Hinf
PEK	7	16170 16218 16304 16311	F4b	16007	16500	-10397 AluI -10394 DdeI
PEK	9	16192 16288 16304 16309 16390	R9b	16011	16500	-10397 AluI -10394 DdeI
PEK	14	16136 16189 16217	B4b1	16036	16398	-10397 AluI
PEK	16	16223 16291 16362	M7	16026	16500	+10397 AluI +9824 HinfI +5176 AluI
PEK	18	16092 16209 16223 16224 16263 16278 16319	G2	16062	16500	+10397 AluI -9824 HinfI +5176 AluI
PEK	20	16066 16209 16304 16311 16399	R9	16032	16500	-10397 AluI -10394 DdeI
PEK	21	16148 16189 16362	M	16036	16398	+10397 AluI +10394 DdeI -9824 HinfI +5176 AluI
PEK	23	16140 16188 16189 16217 16274 16311 16335	B4c	16033	16500	-10397 AluI -10394 DdeI
PEK	24	16189 16217 16261	B4a	16013	16436	-10397 AluI
PEK	26	16189 16217 16261 16286	B4a	16078	16397	-10397 AluI
PEK	27	16140 16189 16217 16274 16335	B4c	16026	16394	-10397 AluI
PEK	28	16126 16214A 16223 16271 16278 16298	M3	16011	16500	+10397 AluI -9824 HinfI
PEK	30	16129 16172 16304	F1a	16005	16397	-10397 AluI
PEK	31	16108 16129 16162 16172 16304	F1a1a	16026	16500	-10397 AluI -10394 DdeI
PEK	33	16093 16192 16288 16304 16309 16390	R9b	16026	16500	-10397 AluI -10394 DdeI
PEK	36	16051 16189 16194C 16195	B	16036	16407	-10397 AluI -10394 DdeI +9bpdel -9824 HinfI
PEK	39	16189 16223 16291 16362 16390	E1a	16026	16396	+10397 AluI +5176 AluI -7598 HhaI
PEK	42	16126 16231 16311	Y2	16011	16500	-10397 AluI +10394 DdeI
PEK	47	16129 16223	M5	16014	16500	+10397 AluI -9824 HinfI +5176 AluI
PEK	50	16223 16295 16362	M7c1c	16026	16500	+10397 AluI +9824 HinfI
PEK	59	16108 16129 16162 16172 16304	F1a1a	16036	16500	- 10397 AluI -10394 DdeI

Series	No	HVS-I Variants	Haplogroup	HVS-I Start	HVS-I End	RFLP Variants
PEK	63	16129 16172 16301 16304 16400	F1a	16023	16500	-10397 AluI -10394 DdeI
PEK	74	16069 16207 16309 16318T	U7	16023	16500	-10397 AluI -10394DdeI +12308 HinfI
PEK	92	16129 16209 16223 16272	M	16035	16436	+10397 AluI -9824 HinfI +5176 AluI
PEK	94	16126 16231 16311 (16109N 16330N 16347N 16384N)	Y2	16067	16400	-10397 AluI
PEK	101	16129 16172 16294 16304 16362	F1a	16024	16500	-10397 AluI -10394 DdeI
PEK	104	16223 16295 16362	M7c1c	16025	16500	+10397 AluI +9824 HinfI +5176 AluI
PEK	106	16140 16188 16189 16217 16274 16311 16335	B4c	16033	16500	-10397 AluI -10394 DdeI
PEK	108	16129 16166C16189 16223 16287 16319	G	16024	16396	+10397 AluI +5176 AluI -9824 HinfI
PEK	109	16093 16209 16223 16224 16263 16278 16319	G2	16012	16500	+10397 AluI -9824 HinfI
PEK	110	16093 16209 16223 16224 16263 16278 16319	G2	16024	16500	+10397 AluI -9824 HinfI +5176 AluI
PEK	111	16129 16132 16172 16304	F1a	16020	16398	-10397 AluI
PEK	112	16185 16223 16295 16362	M7c1c	16036	16500	+9824 HinfI
PEK	113	16140 16188 16189 16217 16274 16311 16335	B4c	16025	16500	-10397 AluI -10394 DdeI +9bpdel
PEK	115	16140 16188 16189 16217 16274 16311 16335	B4c	16024	16500	-10397 AluI -10394 DdeI
PEK	116	16140 16145 16189 16224 16266A	B5a	16011	16398	-10397 AluI
PEK	117	16189 16217 16261	B4a	16011	16432	-10397 AluI
PEK	118	16189 16217 16261	B4a	16014	16420	
PEK	119	16140 16145 16189 16224 16266A	B5a	16011	16399	
PEK	121	16129 16172 16294 16304 16362	F1a	16011	16500	-10397 AluI -10394 DdeI
PEK	122	16189 16217 16261	B4a	16029	16396	-10397 AluI
PEK	128	16093 16189 16222 16298 16299 16399	B	16024	16423	-10397 AluI -10394 DdeI +9bpdel
PEK	130	16140 16188 16189 16217 16261 16274 16311 16335	B4c	16026	16500	-10397 AluI -10394 DdeI +9bpdel
PEK	131	16189 16217 16261	B4a	16026	16401	-10397 AluI
PEK	133	16189 16217 16261	B4a	16060	16398	-10397 AluI
PEK	137	16223 16295 16362	M7c1c	16060	16497	+10397 AluI +9824 HinfI

Series	No	HVS-I Variants	Haplogroup	HVS-I Start	HVS-I End	RFLP Variants
PEK	138	16129 16223 16311 16391	I	16011	16500	-10397 AluI +10394 DdeI +10032 AluI
PEK	139	16126 16231 16311	Y2	16011	16500	-10394 AluI +10394 DdeI
PEK	140	16066 16209 16304 16311 16399	R9	16024	16500	-10397 AluI -10394 DdeI
PEK	151	16189 16217 16261	B4a	16033	16500	
PLB	10	16179 16223 16294	M	16000	16400	+10397Alu, -9824HinfI, +7598HhaI, +5176AluI
PLB	14	16223 16295 16362 (16358N)	M7c1c	16000	16400	+10397Alu
PLB	19	16189 16223 16278	G2	16030	16410	+10397Alu, +10394DdeI, +7598HhaI, +4831HhaI
PLB	23	16109C 16129 16172 16304	F1a	16050	16500	
PLB	26	16223 16295 16362	M7c1c	16030	16500	
PLB	37	16223 16295 16362	M7c1c	16000	16500	
PLB	44	16108 16129 16162 16170 16172 16304	F1a1a	16050	16500	
PLB	45	16108 16129 16162 16170 16172 16304	F1a1a	16050	16500	
PLB	47	16086 16172 16173 16223 16278 16311	G2	16015	16500	+10397Alu, +10394DdeI, +7598HhaI, -9824HinfI, +4831HhaI
PLB	54	16189 16261	B4a	16030	16410	+9bpdel
PLB	58	16223 16261 16362 16390	E1b	16025	16500	+10397Alu, -7598HhaI
PLB	63	16086 16129 16209 16223 16272	G	16025	16500	+10397AluI, +10394DdeI, +7598HhaI
PLB	66	16223 16295 16362	M7c1c	16015	16500	+9824HinfI
PLB	67	16147 16184A 16189 16217 16235 16239	B4	16025	16500	+9bpdel
PLB	73	16223 16295 16362	M7c1c	16040	16500	
PLB	78	16108 16129 16162 16172 16304	F1a1a	16000	16500	
PLB	80	16129 16172 16304 16311 (16189N)	F1a	16015	16500	
PLB	83	16193 16223 16311 16344	N21	16045	16500	-10397AluI, -15606AluI
PLB	85	16140 16172 16189 16223 16278 (16039N)	G2	16025	16470	-9bpdel, +4831HhaI
PLB	89	16223 16295 16362	M7c1c	16040	16500	
PLB	97	16140 16189 16248 16266A 16319	B5a	16010	16430	+9bpdel
PLB	99	16140 16189 16261 16266A	B5a	16045	16425	+9bpdel
PLB	100	16129 16172 16223 16234 16290 16312	G	16050	16500	+10397Alu, +5176AluI, -9824HinfI, +7598HhaI, +4831HhaI
PLB	103	16140 16189 16266A	B5a	16050	16430	+9bpdel
PLB	106	16108 16129 16162 16172	F1a1a	16000	16500	

Series	No	HVS-I Variants	Haplogroup	HVS-I Start	HVS-I End	RFLP Variants
PLB	108	16185 16223 16260 16298	Z	16015	16500	+10397Alu, +10394DdeI, +5176AluI, +7598HhaI, -9824HinfI
PLB	110	16147 16189 16217 16335	B4	16030	16400	
PLB	114	16108 16129 16162 16172 16184 16304 16398	F1a1a	16010	16500	
TGR	1	16223 16295 16362	M7c1c	16030	16500	+10397Alu, +5176AluI, +9824HinfI
TGR	11	16093 16223 16261 16362 16390	E1b	16060	16500	+10397Alu, +5176AluI, -7598HhaI, -9824HinfI
TGR	13	16223 16311 16362	M	16040	16500	+10397Alu, +5176AluI, +7598HhaI, -9824HinfI
TGR	18	16223 16311 16327 (16102N 16159N)	C	16050	16500	+5176AluI, +7598HhaI, -9824HinfI
TGR	18	16129 16172 16301 16304 16400	F1a	16045	16500	-10397AluI, -10394DdeI
TGR	21	16223 16311 16362 (16093N)	M	16050	16500	+10397Alu, +5176AluI, +7598HhaI, -9824HinfI
TGR	24	16094 16129 16223 16263 16274 16311 16343 16357	M10	16030	16500	-10397AluI, +10394DdeI, +5176AluI, +7598HhaI, -9824HinfI, -15606AluI
TGR	25	16086 16148 16223 16259 16278 16319 16399	G2	16050	16500	+10397Alu, +10394DdeI, +5176AluI, +7598HhaI, -9824HinfI, +4831HhaI
TGR	26	16093 16223 16311 16362 (16159N)	M	16030	16500	+10397Alu, +5176AluI, +7598HhaI, -9824HinfI
TGR	28	16093 16223 16311 16362	M	16050	16500	+5176AluI, +7598HhaI, -9824HinfI
TGR	29	16093 16223 16311 16362	M	16040	16460	+5176AluI, +7598HhaI
TGR	30	16093 16223 16311 16362	M	16025	16500	+5176AluI
TGR	34	16223 16257A 16261 16292 16294 (16412N)	N9a1	16040	16500	-10397AluI, -10394DdeI
TGR	37	16223 16295 16362 (16416N)	M7c1c	16050	16460	+9824HinfI
TGR	38	16129 16172 16304 16311	F1a	16030	16500	
TGR	43	16129 16172 16301 16304 16400	F1a	16030	16500	
TGR	45	16129 16172 16301 16304 16400 (16412N)	F1a	16040	16500	
TGR	46	16094 16129 16223 16263 16274 16311 16343 16357	M10	16030	16500	-10397AluI, +10394DdeI, +5176AluI, +7598HhaI, -9824HinfI
TGR	47	16129 16172 16304 16311	F1a	16030	16500	
TGR	48	16129 16172 16301 16304 16400	F1a	16030	16500	
TGR	49	16129 16172 16301 16304 16400	F1a	16050	16500	
TGR	54	16189	B	16055	16470	-10397AluI, +9bpdel
TGR	59	16140 16189 16266A	B5a	16050	16470	+9bpdel

Series	No	HVS-I Variants	Haplogroup	HVS-I Start	HVS-I End	RFLP Variants
TGR	68	16086 16148 16223 16259 16278 16319 16399 (16110N)	G2	16050	16480	+10397Alu, +10394Ddel, +5176AluI, +7598HhaI, -9824HinfI
TGR	77	16223 16295 16362	M7c1c	16070	16500	
TGR	82	16129 16172 16301 16304 16400	F1a	16030	16500	
TGR	86	16094 16129 16223 16263 16274 16311 16343 16357 (16064N 16173N)	M10	16030	16500	-10397AluI, +10394Ddel, +5176AluI, +7598HhaI, -9824HinfI, -15606AluI
TGR	87	16192 16234 16288 16293 16304 16309 16390	R9b	16030	16500	-10397AluI
TGR	88	16223 16263 16274 16311 16343 16357	M10	16030	16500	-10397AluI, +10394Ddel, +5176AluI, +7598HhaI, -9824HinfI
TGR	90	16094 16129 16223 16263 16274 16311 16343 16357	M10	16060	16500	-10397AluI, +10394Ddel, +5176AluI, +7598HhaI, -9824HinfI
TGR	91	16192 16234 16288 16293 16304 16309 16390	R9b	16010	16500	-10397AluI, -10394Ddel
TGR	92	16140 16189 16217 16274 16335 (16094N)	B4c	16040	16390	
TGR	93	16094 16129 16223 16263 16274 16311 16343 16357	M10	16080	16460	-10397AluI, +10394Ddel, +5176AluI, +7598HhaI
TGR	94	16094 16129 16223 16263 16274 16311 16343 16357	M10	16060	16500	-10397AluI, +10394Ddel, +7598HhaI
TGR	95	16249 16288 16304	R22	16030	16500	-10397AluI, -10394Ddel
TGR	96	16129 16172 16301 16304 16400	F1a	16040	16500	
TOR	1	16223 16261 16362 16390	E1b	16050	16475	+10397Alu, +5176AluI, -7598HhaI, -9824HinfI
TOR	2	16086 16129 16297 16324	M7b3	16030	16470	+10397Alu, +10394Ddel, +5176AluI, +9824HinfI
TOR	4	16129 16223 16362 16390	M	16030	16500	+10397AluI, +5176AluI, +10394Ddel, +7598HhaI
TOR	5	16189 16217 16247 16261 (16169-16170N)	B4a1	16032	16424	
TOR	6	16223 16291 16362 16390	E1a	16080	16480	+10397Alu, +5176AluI, -9824HinfI, -7598HhaI
TOR	10	16223 16261 16362 16390	E1b	16000	16500	+10397AluI, +5176AluI, -9824HinfI
TOR	13	16092 16148 16189 16223 16362	D5	16030	16410	-10397AluI, -9bpdel, -10394Ddel, -5176AluI, +7598HhaI
TOR	16	16189 16217 16247 16261	B4a1	16033	16410	
TOR	18	16189 16195A 16241 162651 16311 (16199N 16223N)	M	16050	16430	+10397Alu, +10394Ddel, -9824HinfI, +5176AluI
TOR	20	16189 16217 16247 16261	B4a1	16027	16433	

Series	No	HVS-I Variants	Haplogroup	HVS-I Start	HVS-I End	RFLP Variants
TOR	21	16129 16172 16294 16304 16362	F1a	16030	16500	
TOR	22	16223 16291 16362 16390	E1a	16000	16500	
TOR	23	16168 16223 16295 16362 (16399N)	M7c1c	16030	16500	+10397Alu, +9824HinfI
TOR	24	16223 16295 16362	M7c1c	16030	16500	
TOR	26	16223 16291 16362 16390	E1a	16000	16500	+10397AluI, +10394DdeI, -9824HinfI
TOR	27	16086 16129 16297 16324	M7b3	16040	16500	+10397AluI, +10394DdeI, +9824HinfI
TOR	28	16140 16189 16217 16274 16335	B4c	16050	16460	
TOR	31	16189 16217 16261 (16195N 16198N 16199N)	B4a	16065	16440	+9bpdel
TOR	34	16223 16295 16362	M7c1c	16010	16500	
TOR	35	16223 16291 16362 16390	E1a	16010	16500	
TOR	36	16126 16231 16311	Y2	16030	16500	-10397AluI, +10394DdeI
TOR	37	16234 16278 16294	G2	16060	16430	+10397Alu, +10394DdeI, +5176AluI, +7598HhaI, -9824HinfI
TOR	38	16129 16144 16148 16223 16241 16265C 16299 16311 16343 16362	Q	16030	16500	
TOR	39	16092 16148 16189 16223 16362	D5	16010	16415	-10397AluI, -10394DdeI, -9pbdel, -5176AluI
TOR	41	16140 16189 16217 16362	B4	16020	16410	
TOR	42	16223 16295 16362	M7c1c	16030	16500	
TOR	43	16189 16223 16362 (16039N)	D5	16020	16410	-10397AluI, -10394DdeI, -5176AluI
TOR	44	16129 16172 16294 16304 16362	F1a	16030	16500	
TOR	45	16223 16295 16362	M7c1c	16010	16500	
TOR	52	16189 16217 16247 16261	B4a1	16037	16431	
TOR	53	16223 16291 16362 16390	E1a	16030	16470	+5176AluI, +10397AluI, +10394DdeI, -7598HhaI, -9824HinfI
TOR	54	16129 16172 16294 16304 16362	F1a	16030	16470	
TOR	55	16189 16217 16261 16311	B4a	16050	16430	
TOR	56	16223 16324 16362 16390	E1	16000	16500	+5176AluI, +10397AluI, +10394DdeI, -7598HhaI, -9824HinfI
TOR	59	16140 16189 16243	B5b	16030	16410	
TOR	60	16140 16189 16217 16274	B4c	16030	16415	
TOR	63	16126 16231 16311	Y2	16020	16500	

Series	No	HVS-I Variants	Haplogroup	HVS-I Start	HVS-I End	RFLP Variants
TOR	65	16092 16148 16189 16223 16362	D5	16060	16420	-10397AluI, -10394DdeI, -9bpdel, -5176AluI
TOR	68	16223 16291 16362 16390	E1a	16030	16500	-7598HhaI
TOR	70	16189 16217 16261	B4a	16050	16400	
TOR	71	16295 16362	M7c1c	16050	16500	+5176AluI, +9824HinfI
TOR	74	16223 16291 16362 16390	E1a	16040	16500	-7598HhaI
TOR	75	16092 16148 16189 16223 16362	D5	16030	16410	-10397AluI, -10394DdeI, -9bpdel, -5176AluI
TOR	77	16189 16223 16362	D5	16040	16400	-10397AluI, -10394DdeI, -9bpdel, -5176AluI
TOR	78	16189 16217 16247 16261	B4a1	16030	16410	
TOR	80	16140 16189 16217 16274 16335	B4c	16050	16430	
TOR	81	16189 16223 16278	G2	16020	16425	+10397Alu, +10394DdeI, +5176AluI, +7598HhaI, -9824HinfI, +4831HhaI
TOR	82	16140 16189 16217 16274 16335	B4c	16060	16410	
TOR	85	16168 16223 16295 16362 (16078N 16120N)	M7c1c	16050	16500	+10397AluI, +10394DdeI, +9824HinfI
TOR	86	16223 16324 16362 16390	E1	16010	16500	+10397AluI, +10394DdeI, -9824HinfI
TOR	90	16129 16172 16294 16304 16362	F1a	16050	16500	
TOR	91	16117 16223 16291 16362 16390	E1a	16040	16500	+10397AluI, +10394DdeI, -7598HhaI, -9824HinfI
TOR	92	16117 16223 16291 16362 16390	E1a	16010	16500	+10397AluI, +10394DdeI, -7598HhaI
TOR	96	16129 16172 16294 16304 16362	F1a	16080	16400	
TOR	106	16223 16291 16362 16390	E1a	16030	16500	
TOR	108	16108 16129 16162 16172 16304	F1a1a	16010	16500	
TOR	110	16180 16223 16291 16362 16390	E1a	16030	16500	+10397AluI, +10394DdeI, -9824HinfI, -7598HhaI
TOR	111	16086 16126 16129 16297 16324	M7b3	16030	16500	+10397AluI, +10394DdeI, +9824HinfI
TOR	113	16129 16172 16294 16304 16362	F1a	16030	16500	
TOR	114	16189 16223 16362	D5	16050	16400	-5176AluI, -9824HinfI
TOR	116	16223 16291 16362 16390	E1a	16030	16470	-7598HhaI
TOR	122	16129 16172 16294 16304 16362	F1a	16030	16500	
TOR	127?	16189 16217 16261	B4a	16023	16394	
TOR	127	16223 16291 16362 16390	E1a	16050	16500	-7598HhaI
UJP	1	16093 16192 16223 16271 16316 16362	D	16030	16500	+10397AluI, +10394DdeI, -9824HinfI, +7598HhaI, -5176AluI
UJP	2	16189 16217 16261	B4a	16040	16410	+9bpdel

Series	No	HVS-I Variants	Haplogroup	HVS-I Start	HVS-I End	RFLP Variants
UJP	3	16093 16223 16261 16362 16390	E1b	16030	16500	+10397AluI, +5176AluI, -9824HinfI
UJP	4	16189 16217 16247 16261	B4a1	16030	16380	+9bpdel
UJP	6	16129 16172 16304	F1a	16010	16420	
UJP	8	16051 16223 16258C 16309 16362 16390	E1	16030	16500	+10397AluI, +10394DdeI, +5176AluI, -7598HhaI, -9824HinfI
UJP	9	16129 16144 16148 16153 16162 16192 16223 16241 16249 16265C 16311 16343	Q	16000	16410	+10397AluI
UJP	10	16129 16144 16148 16223 16241 16265C 16311 16343	Q	16040	16500	+10397AluI
UJP	12	16223 16298 16327	C	16020	16500	+10397AluI, +10394DdeI, +5176AluI, -9824HinfI
UJP	13	16051 16189 16362	B	16030	16430	-10397AluI, -10394DdeI, +9bpdel
UJP	14	16129 16148 16223 16291 16362 16390	E1a	16050	16500	+10397AluI, +10394DdeI, -7598HhaI, -9824HinfI
UJP	15	16189 16209 16223 16300	M	16030	16370	+10397Alu, -9bpdel, +10394DdeI, +7598HhaI, -9824HinfI, +5176AluI
UJP	16	16223 16291 16362 16390	E1a	16010	16500	+10397AluI, -7598HhaI, -9824HinfI
UJP	17	16140 16189 16217 16274 16335	B4c	16030	16400	+9bpdel
UJP	18	16189 16217 16261	B4a	16050	16400	+9bpdel
UJP	19	16223 16261 16362 16390	E1b	16030	16500	+10397Alu, -7598HhaI
UJP	20	16189 16217 16261	B4a	16030	16410	+9bpdel
UJP	22	16223 16291 16362 16390	E1a	16030	16500	-7598HhaI
UJP	23	16140 16189 16266A	B5a	16040	16430	+9bpdel
UJP	24	16189 16223 16311 16362	D5	16050	16430	+10397AluI, +10394DdeI, +7598HhaI, -9824HinfI, -5176AluI
UJP	25	16092 16129 16172 16294 16304 16362	F1a	16030	16500	
UJP	26	16223 16256 16324 16362 16390	E1	16010	16470	+10397AluI, -7598HhaI
UJP	27	16148 16189 16223 16362	D5	16030	16380	-10397AluI, -9bpdel, -10394DdeI
UJP	28	16223 16291 16362 16390	E1a	16030	16500	
UJP	29	16172 16304	F1a	16030	16500	
UJP	30	16189 16217 16247 16261	B4a1	16030	16410	+9bpdel
UJP	31	16189 16217 16261	B4a	16030	16435	+9bpdel
UJP	32	16223 16291 16362 16390	E1a	16030	16500	-7598HhaI

Series	No	HVS-I Variants	Haplogroup	HVS-I Start	HVS-I End	RFLP Variants
UJP	33	16129 16209 16223 16272	G	16045	16430	+10397Alu, +10394DdeI, +5176AluI, +7598HhaI, -9824HinfI
UJP	34	16140 16189 16261 16266A	B5a	16040	16410	+9bpdel
UJP	35	16140 16189 16243	B5b	16040	16430	+9bpdel
UJP	36	16189 16223 16278	G2	16030	16430	+10397AluI, +10394DdeI, +5176AluI, +7598HhaI, -9824HinfI, +4831HhaI
UJP	39	16140 16189 16261 16266A	B5a	16030	16410	+9bpdel
UJP	40	16129 16209 16223 16272	G	16000	16500	+10397AluI, +10394DdeI, +5176AluI, +7598HhaI, -9824HinfI
UJP	41	16189 16217 16247 16261	B4a1	16040	16430	+9bpdel
UJP	42	16129 16144 16148 16172 16223 16242 16265C 16311 16343	Q	16020	16500	+10397AluI
UJP	43	16093 16168 16187 16288 16304	R21/R9??	16030	16500	-10397AluI, -10394DdeI
UJP	44	16189 16217 16223 16261 16335	B4a	16040	16425	+9bpdel
UJP	45	16140 16189 16217 16274 16335 (16252N)	B4c	16040	16415	+9bpdel
UJP	46	16223 16295 16362	M7c1c	16030	16500	+10397AluI, +9824HinfI
UJP	47	16189 16249 16286 16288	R22	16050	16435	-10397AluI , -10394DdeI, -12308HinfI
UJP	48	16051 16185 16223 16362 16390	E1	16030	16500	+10397AluI, +10394DdeI, +5176AluI, -7598HhaI, -9824HinfI
UJP	49	16223 16291 16362 16390	E1a	16050	16500	
UJP	50	16140 16189 16217 16274 16335	B4c	16040	16435	+9bpdel
UJP	51	16147 16189 16217 16235 (16358N)	B4	16030	16410	+9bpdel
UJP	52	16129 16172 16301 16304 16362 16400	F1a	16030	16500	
WAI	3	16223 16295 16311 16362	M7c1c	16080	16500	+10397AluI, +10394DdeI, +5176AluI, +9824HinfI
WAI	7	16223 16295 16311 16362 (16412N 16434N)	M7c1c	16080	16480	+5176AluI
WAI	9	16086 16129 16295 16297 16324	M7b3	16030	16500	+10397AluI, +5176AluI, +9824HinfI
WAI	12	16129 16172 16294 16304 16362	F1a	16050	16500	
WAI	23	16168 16172 16223 16249	N22	16030	16500	-10397AluI, -10394DdeI, -15606AluI
WAI	24	16223 16311 16362 16400	M	16030	16500	+10397AluI, +7598HhaI, -9824HinfI, +5176AluI
WAI	25	16093 16168 16172 16223 16249	N22	16030	16500	-10397AluI, -10394DdeI, -15606AluI

Series	No	HVS-I Variants	Haplogroup	HVS-I Start	HVS-I End	RFLP Variants
WAI	26	16150 16223 16274 16295 16311 16362	M7c1c	16000	16500	+5176AluI, +10397AluI, +7598HhaI, +10394DdeI, +9824HinfI
WAI	27	16223 16261 16362 16390	E1b	16050	16500	+10397AluI, +5176AluI, -7598HhaI, -9824HinfI
WAI	28	16223 16261 16362 16390	E1b	16030	16500	+5176AluI, +10397AluI, +10394DdeI
WAI	30	16051 16215 16223 16362 16390	E1	16030	16500	+10397AluI, +5176AluI, -7598HhaI, -9824HinfI
WAI	31	16256 16290 16465	R	16000	16500	-10397AluI, -10394DdeI, -15606AluI
WAI	33	16168 16172 16223 16249	N22	16030	16500	-10397AluI, -10394DdeI, -15606AluI
WAI	35	16223 16295 16362	M7c1c	16010	16500	+10397AluI, +9824HinfI
WAI	36	16108 16129 16162 16172 16304	F1a1a	16030	16500	
WAI	37	16189 16223 16227 16291 16362 (16356N 16358N 16360N)	M	16030	16390	+10397AluI, +5176AluI, -7598HhaI, -9824HinfI
WAI	38	16129 16172 16304 16311 (16477N)	F1a	16050	16500	
WAI	39	16249 16288 16301 16304 16390	R22	16050	16500	-10397AluI, -10394DdeI
WAI	40	16129 16209 16223 16311 16325	G	16000	16500	+10397AluI, +10394DdeI, +5176AluI, +7598HhaI, -9824HinfI, +4831HhaI
WAI	41	16129 16172 16294 16304 16362	F1a	16000	16500	
WAI	42	16129 16144 16148 16223 16241 16265C 16311 16343	Q	16040	16500	+10397AluI
WAI	43	16223 16261 16362 16390	E1b	16040	16500	+5176AluI, +10397AluI, +10394DdeI, -7598HhaI, -9824HinfI
WAI	45	16220C 16265 16298 16311 16362	F3b	16060	16480	-10397AluI
WAI	46	16129 16172 16304 16311 (16465N)	F1a	16020	16500	
WAI	47	16093 16172 16266 16270	P	16000	16500	-10397AluI, -10394DdeI, +15606AluI
WAI	48	16086 16129 16295 16297 16324	M7b3	16060	16500	+10397AluI, +5176AluI, +9824HinfI
WAI	50	16129 16172 16304	F1a	16040	16500	
WAI	55	16129 16140 16271	M	16040	16500	+10397AluI, +10394DdeI, +5176AluI, +7598HhaI, -9824HinfI
WAI	56	16189 16223 16362	M7	16000	16460	+10397AluI, +1039DdeI, +5176AluI, +9824HinfI
WAI	57	16223 16295 16362	M7c1c	16010	16500	+5176AluI, +9824HinfI
WAI	58	16086 16129 16295 16297 16324	M7b3	16040	16500	+10397AluI, +5176AluI, +9824HinfI
WAI	59	16223 16261 16362 16390 (16348N)	E1b	16070	16500	+5176AluI, -7598HhaI, -9824HinfI
WAI	60	16223 16295 16362	M7c1c	16030	16500	

Series	No	HVS-I Variants	Haplogroup	HVS-I Start	HVS-I End	RFLP Variants
WAI	62	16223 16261 16362 16390	E1b	16040	16500	+5176AluI, -7598HhaI, -9824HinfI
WAI	64	16086 16129 16295 16297 16324	M7b3	16040	16500	
WAI	65	16129 16162 16172 16304	F1a1	16040	16500	
WAI	66	16249 16288 16301 16304 16390	R22	16080	16500	
WAI	67	16249 16288 16317C	R22	16060	16500	-10397AluI, -10394DdeI
WAI	69	16223 16291 16362 16390	E1a	16040	16460	+10397AluI, +5176AluI, -7598HhaI
WAI	70	16093 16168 16223 16249 16278 16295	N22	16040	16500	-10397AluI, -10394DdeI, -15606AluI
WAI	72	16249 16288 16304 16390	R22	16070	16500	
WAI	76	16108 16129 16162 16172 16293 16304	F1a1a	16000	16500	
WAI	77	16093 16176 16266 16270 16357	P	16070	16470	-10397AluI, -10394DdeI, +15606AluI
WAI	78	16157 16256 16304 16335 (16086N)	F	16060	16500	-10397AluI, -10394DdeI
WAI	81	16223 16261 16362 16390	E1b	16050	16500	+10397Alu, +10394DdeI, +5176AluI, -7598HhaI
WAI	83	16129 16172 16294 16304 16362 (16412N)	F1a	16050	16500	
WAI	84	16129 16172 16294 16304	F1a	16000	16500	
WAI	88	16129 16144 16148 16209 16223 16241 16265C 16311 16343	Q	16050	16500	+10397AluI

Appendix II – Results of HVS-II Sequencing, np 10310 and np 8701 Status

Appendix II – Results of HVS-II Sequencing, np 10310 and np 8701 Status

Series	Number	HVS-II Start	HVS-II End	HVS-II Variants	Other Variants
ALO	8	40	429	073 249del 263	n/t
ALO	40	40	429	073 249del 263	n/t
ALO	44	40	429	073 249del 263 (85N 94N)	n/t
ALO	58	90	429	263	n/t
ALO	63	60	429	73 143 199 263	n/t
ALO	72	100	400	249del 263	n/t
BAL	58	90	429	152 263	n/t
FIL	3	40	429	73 249del 263	n/t
KK	6	40	429	73 146 249del 263	n/t
KK	63	40	429	73 249del 263	n/t
KK	69	50	420	73 249del 263	n/t
MED	3	40	370	73 143 263	10310G
MED	8	40	420	73 249del 263	10310A
MED	113	40	420	73 223 263	10310G
MND	2	130	400	249del 263	n/t
MND	54	50	420	73 146 249del	10310A
MTR	83	50	420	73 145 152 249del 263	n/t
PAL	80	50	420	73 249del 263	10310G
PEK	7	60	120	73 249del 263	n/t
PEK	20	40	420	73 89 90 263	10310G
PEK	107	50	430	73 146 199 263	10310G
PLB	108	96	420	73 249del 263	n/t
UJP	12	96	420	143 146 195 249del 263	n/t
ORA	9B	39	426	73 150 152 263 315insC	n/t
ORA	158B	39	426	73 263 309insC 315insC	n/t
ORA	7A	41	420	73 249del 263	n/t
ORA	112A	60	420	73 143 152 183 263	n/t
ORA	118A	40	430	58 73 143 152 183 263	n/t
ORA	80B	50	420	73 150 263	n/t
ORA	114A	n/t	n/t	n/t	8701G
ORA	115A	n/t	n/t	n/t	8701G
ORA	122B	n/t	n/t	n/t	8701G

n/t = not tested

Appendix III – Table of Southeast Asian Haplotypes

Appendix III – Table of Southeast Asian Haplotypes

Haplogroup	HVS-I Variants	China	Taiwan	Philippines	Thailand	Semang	Senoi	A. Malay	Melayu	Palembang	Pekanbaru	Padang	Bangka	Medan	Tengger	Banjarmasin	Kota Kinabalu	Manado	Palu	Ujung Padang	Toraja	Bali	Lombok	Sumba	Alor	Ambon
?	093 129 219 223 261 278 325	.	.	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
?	104 219 223 287	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
?	111 129 183 189 223 311	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
?	124 148 223	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
?	124 193 223	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
?	127 129 166 217 319 365	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
?	129 223	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
?	148 172 189 223 234 261 290	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
?	148 223 234 355 356 362	.	.	.	3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
?	162 223 259 289 324 355 362	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
?	166 189	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
?	172 189 223 249 290	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.
?	172 189 223 311	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
?	174 223 260 264 271	.	.	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
?	182 193 223	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
?	185 189 324	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
?	188 223 266 271 362	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
?	189 311	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
?	192 223 316 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
?	192 278 325	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
?	214 223	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
?	214 223 274 311	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
?	214A 223 228 256 278	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
?	214A 223 256 278	.	.	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
?	219 223 239 325 355	.	.	.	3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
?	223	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
?	223 266 278 302	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
?	356	.	.	.	1.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
A	093 223 263 290 293C 319	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
A	126 223 234 256 290 319	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
A	126 235 290 319	1	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
A	129 189 223 290 319	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
A	129 213 223 290 319	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
A	129 223 290 311 325	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

Haplogroup	HVS-I Variants	China	Taiwan	Philippines	Thailand	Semang	Senoi	A. Malay	Melayu	Palembang	Pekanbaru	Padang	Bangka	Medan	Tengger	Banjarmasin	Kota Kinabalu	Manado	Palu	Ujung Padang	Toraja	Bali	Lombok	Sumba	Alor	Ambon
A	189 223 290 319	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
A	223 234 290 293C 319	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
A	223 249 319	.	4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
A	223 290 293C 319	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
A	223 290 319	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
A4	092 223 256 290 319 362	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
A4	092 223 319 362	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
A4	093 129 145 223 290 319 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
A4	124 223 290 319 362	6	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
A4	126 223 290 319 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
A4	129 223 274 290 319 362	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
A4	131 222A 223 290 319 362	1	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
A4	223 271 287 319 356 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
A4	223 274 290 319 325 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
A4	223 274 290 319 362	18	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
A4	223 289 290 319 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
A4	223 290 294 319 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
A4	223 290 319 356 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
A4	223 290 319 362	7	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B	093 179 189	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B	093 179 189 342	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B	093 189 222 298 299	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B	108 189 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B	129 189 352 355	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B	179 189 222 342	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B	189	1	.	.	1	.	.	2	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.
B	189 194C 195	.	.	.	.	.	.	.	.	.	1	3	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B	189 195 286	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B	189 234	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B	189 274	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B	189 274 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B	189 278	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B	189 292	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B	189 362	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	1	.	.	.	.	.	.

Haplogroup	HVS-I Variants	China	Taiwan	Philippines	Thailand	Semang	Senoi	A. Malay	Melayu	Palembang	Pekanbaru	Padang	Bangka	Medan	Tengger	Banjarmasin	Kota Kinabalu	Manado	Palu	Ujung Padang	Toraja	Bali	Lombok	Sumba	Alor	Ambon
B4	092 147 179 189 217 235	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.
B4	093 189 217 234	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4	093 189 217 262 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4	108 189 217 324 362	4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4	129 189 217	1	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	1	.	.	.	.
B4	129 189 217 354	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4	140 168A 189 217 311	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4	140 189 217 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.
B4	147 184A 189 217 234 235	.	.	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4	147 184A 189 217 235	1	.	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4	147 184A 189 217 235 239	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4	147 184A 189 217 235 261	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4	147 189 217	.	.	.	.	.	.	.	.	.	.	.	.	.	.	3	.	1	.	.	.	1	.	.	.	.
B4	147 189 217 235	.	.	.	.	.	.	2	1	.	.	.	1	1	.	4	.	.	.	1	.	1	.	.	.	.
B4	147 189 217 235 294 360	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.
B4	147 189 217 235 294G	.	.	.	1	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.
B4	147 189 217 235 311	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4	186 189 217	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4	189 214 217	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4	189 217	3	1	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	1	.	.	.	.
B4	189 217 218	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4	189 217 218 234 278	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4	189 217 223	2	.	.	4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4	189 217 223 235 291 316	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4	189 217 223 257	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4	189 217 223 311 320	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4	189 217 234	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4	189 217 234 278	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4	189 217 256 311 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4	189 217 299	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4	189 217 362	1	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4	189 217 365	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4a	092 189 217 261 299	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

Haplogroup	HVS-I Variants	China	Taiwan	Philippines	Thailand	Semang	Senoi	A. Malay	Melayu	Palembang	Pekanbaru	Padang	Bangka	Medan	Tengger	Banjarmasin	Kota Kinabalu	Manado	Palu	Ujung Padang	Toraja	Bali	Lombok	Sumba	Alor	Ambon
B4a	093 153 189 217 261 292 311 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4a	093 188 189 214 217 261 287	1	.	.	.	.	.	3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4a	093 189 217 243 261 278	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4a	093 189 217 261	1	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.
B4a	093 189 217 261 344	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.
B4a	108 189 217 261	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4a	111 129 189 217 261 324	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4a	111 189 217 261 324	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4a	129 145 189 217 261	1	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4a	129 154 189 217 261	.	.	.	.	.	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4a	129 189 217 261	3	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4a	129 189 217 261 311	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4a	129 189 217 261 324	1	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4a	129 189 217 261 354 357	1	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4a	129 189 223 261	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4a	129 189 223 261 311	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4a	129 189 261	1	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4a	150 189 217 228 261	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4a	150 189 217 240 261	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4a	150 189 261	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4a	153 189 213 217 261 292 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4a	153 189 217 261	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4a	154 189 217 240 261	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4a	154 189 217 261 324	.	4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4a	178 189 217 261	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	.	.	.	.	.	.	.	.	.	.
B4a	188 189 217 223 261 355	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.
B4a	189 197 217 261 272 324	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4a	189 213 217 228 261 292	.	.	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4a	189 213 217 228 261 292 358	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4a	189 213 217 261 270 292	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4a	189 213 217 261 292 362	1	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4a	189 217 221 240 261	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.
B4a	189 217 223 261	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1

Haplogroup	HVS-I Variants	China	Taiwan	Philippines	Thailand	Semang	Senoi	A. Malay	Melayu	Palembang	Pekanbaru	Padang	Bangka	Medan	Tengger	Banjarmasin	Kota Kinabalu	Manado	Palu	Ujung Padang	Toraja	Bali	Lombok	Sumba	Alor	Ambon
B4a	189 217 223 261 335	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.
B4a	189 217 228 261	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4a	189 217 248 261 295	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.
B4a	189 217 261	3	14	.	1	.	.	.	.	.	7	2	.	.	.	5	3	1	.	4	3	.	2	.	.	2
B4a	189 217 261 272 288 324	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4a	189 217 261 272 324	.	10	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4a	189 217 261 278	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1
B4a	189 217 261 286	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4a	189 217 261 288	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.
B4a	189 217 261 292	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4a	189 217 261 299 355	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4a	189 217 261 311	.	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	1	.	.	.	.	.
B4a	189 217 261 324	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4a	189 257 261 360	1	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4a	189 261	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4a	189 261 292	2	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4a1	093 189 217 247 261	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.
B4a1	184 189 217 247 261 299	1	.	.	.	.	.	.	.	.	.	.	.	.	.	2	.	1	.	3	5	.	1	.	.	5
B4a1	189 217 247 261	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4a1	189 217 247 261 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1
B4b	092 136 189 309 354	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4b	126 136 189 217	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4b	136 171 189 223 297	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4b	136 179 189 217	3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4b	136 183T 189 217 218 239 248	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4b	136 189	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4b	136 189 217	1	3	.	.	.	.	.	1	.	1	.	.	.	.	1	1	1	.	.	.	.	.	.	.	2
B4b	136 189 217 218	4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4b	136 189 217 223 257	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4b	136 189 217 228 365	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4b	136 189 217 234 309 354	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4b	136 189 217 260 287 325	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4b	136 189 217 261	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.
B4b	136 189 217 265C	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

Haplogroup	HVS-I Variants	China	Taiwan	Philippines	Thailand	Semang	Senoi	A. Malay	Melayu	Palembang	Pekanbaru	Padang	Bangka	Medan	Tengger	Banjarmasin	Kota Kinabalu	Manado	Palu	Ujung Padang	Toraja	Bali	Lombok	Sumba	Alor	Ambon
B4b	136 189 217 300	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.
B4b	136 189 217 309 354	4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4b	136 189 217 365	.	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4b	136 189 257	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4b	136 189 284	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4c	092 140 189 217 274 283T 311 335	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.
B4c	092 140 189 217 335	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.
B4c	129 140 145 166 189 217 274 335	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4c	129 140 166 179 189 217 274 335	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4c	129 189 217 274 289 301 311	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4c	140 188 189 217 261 274 311 335	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4c	140 188 189 217 274 311 335	.	.	.	.	.	.	.	.	.	4	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4c	140 189 217 228 274 319 335	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4c	140 189 217 235 274	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	.	.	1	.
B4c	140 189 217 242A 274 335	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4c	140 189 217 274	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.
B4c	140 189 217 274 291	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4c	140 189 217 274 305T 335	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4c	140 189 217 274 311	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4c	140 189 217 274 319 335	.	3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4c	140 189 217 274 335	2	3	.	.	.	.	.	1	.	1	.	.	.	1	2	1	1	.	3	3	.	.	.	1	.
B4c	140 189 217 274 335 375	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B4c	140 189 217 335	.	2	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.
B4c	140 189 274 335	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.
B4c	189 217 274 335	.	3	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B5a	099 140 189 266A 293	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B5a	124 140 189 261 266A	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B5a	127 140 189 266A	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B5a	129 140 189 266A	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B5a	140 145 189 217 266A	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

Haplogroup	HVS-I Variants	China	Taiwan	Philippines	Thailand	Semang	Senoi	A. Malay	Melayu	Palembang	Pekanbaru	Padang	Bangka	Medan	Tengger	Banjarmasin	Kota Kinabalu	Manado	Palu	Ujung Padang	Toraja	Bali	Lombok	Sumba	Alor	Ambon
B5a	140 145 189 224 266A	.	.	.	.	.	.	.	.	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B5a	140 145 189 266A	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B5a	140 148 189 266A 292	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B5a	140 156T 189 266A 270 293	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B5a	140 166 189 266A	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B5a	140 187 189 256 266A	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B5a	140 187 189 256 266A 355	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B5a	140 187 189 266A	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B5a	140 189 223 228 266A 287	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.
B5a	140 189 223 266A	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B5a	140 189 223 266A 293	.	.	.	1	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.
B5a	140 189 234 266A	.	.	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B5a	140 189 242A 256 261 266A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B5a	140 189 245 266G 362	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B5a	140 189 248 266A 319	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B5a	140 189 249 266A	.	.	.	.	.	.	.	.	.	.	.	.	.	2	.	.	.	.	.	.	.	.	.	1	.	.
B5a	140 189 261 266A	.	.	.	2	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B5a	140 189 261 266A 269	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B5a	140 189 265 266A	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B5a	140 189 266A 270	.	.	.	1	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.
B5a	140 189 266A 274	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.
B5a	140 189 266A 291	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B5a	140 189 266A 293	.	.	.	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B5a	140 189 266A 298	.	.	.	.	.	.	.	1	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.
B5a	140 189 266A 304	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B5a	140 189 266A 311	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B5a	140 189 266A 362	.	1	.	.	.	.	.	6	1	.	.	.	2	1	3	1	2	.	.	1	.	4	3	.	.	2
B5a	140 189 266R	18	11	.	6	1	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B5a	189 266G 362	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1
B5a	189 266R	1	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B5b	111 126 140 189 234 243	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B5b	111 129 140 189 234 243 249 250	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B5b	111 140 189 234 243	1	.	.	.	.	.	.	.	.	.	.	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.

Haplogroup	HVS-I Variants	China	Taiwan	Philippines	Thailand	Semang	Senoi	A. Malay	Melayu	Palembang	Pekanbaru	Padang	Bangka	Medan	Tengger	Banjarmasin	Kota Kinabalu	Manado	Palu	Ujung Padang	Toraja	Bali	Lombok	Sumba	Alor	Ambon
B5b	111 140 189 234 243 291	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B5b	111 140 189 234 243 344	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B5b	140 147 189 243 262	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B5b	140 189 234 243	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.
B5b	140 189 243	2	.	.	.	.	.	.	.	.	.	2	.	2	.	2	.	.	1	1	1	.	1	.	.	.
B5b	140 189 243 256 311	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B5b	140 189 243 294	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B5b	140 189 243 294 354	.	.	.	.	14	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B5b	140 189 243 311	1	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B5b	140 189 243 355	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	.	.	.	.	.	.	.	.
C	092 129 189 223 298 327 355	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
C	092 189 223 298 327 355	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
C	093 129 223 298 327	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.
C	093 172 223 298 327	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
C	093 172 298	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
C	126 189 223 298 311 327	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
C	129 148 193 223 298 327	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
C	129 148 223 242 298 311 319 327	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
C	129 148 223 298 327	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
C	129 223 298 327	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
C	175 223 298 311 327	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.
C	189 223 261 298 327	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
C	189 223 298 311 327 357	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
C	189 223 298 319 327	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
C	189 223 298 327	6	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
C	189 298 327	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
C	217 223 298 311 327	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
C	223 243 297 298 324 327	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
C	223 298 327	4	.	.	4	.	.	.	.	.	.	.	.	.	.	.	1	.	1	.	.	.	1	.	.	.
C	223 311 327	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.
C	223 327	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D?	092 223 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D?	092 362	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

Haplogroup	HVS-I Variants	China	Taiwan	Philippines	Thailand	Semang	Senoi	A. Malay	Melayu	Palembang	Pekanbaru	Padang	Bangka	Medan	Tengger	Banjarmasin	Kota Kinabalu	Manado	Palu	Ujung Padang	Toraja	Bali	Lombok	Sumba	Alor	Ambon
D	093 192 223 234 271 316 362	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D	093 192 223 271 316 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.
D	111 172 193 223 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D	111 223 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D	114 223 294 318 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D	153 223 362	1	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D	172 193 223 362	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.
D?	172 223 274 291 362	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.
D	172 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.
D	174 223 311 362	3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D	188 223 274 311 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	1	.	.	.	.	.	.	.	.
D	192 223 274 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	1	.	.	.	.	.	.	.	.
D?	223 249 274 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D	223 256 311 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D	223 274 291 362	5	.	.	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.
D	223 274 362	1	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D	223 286 362	.	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.
D?	223 301 362	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.
D	223 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.
D	295 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D4	092 218 223 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D4	093 176 223 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D4	093 223 232 290 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D4	093 223 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D4	093 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D4	095 209 223 294 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D4	095 209 223 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D4	111 187 223 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D4	111 223 261 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D4	148 223 249 301 342 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D4	174 223 311 343 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D4	174 223 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D4	174 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D4	176 223 291A 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

Haplogroup	HVS-I Variants	China	Taiwan	Philippines	Thailand	Semang	Senoi	A. Malay	Melayu	Palembang	Pekanbaru	Padang	Bangka	Medan	Tengger	Banjarmasin	Kota Kinabalu	Manado	Palu	Ujung Padang	Toraja	Bali	Lombok	Sumba	Alor	Ambon
D4	184 223 311 362	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D4	189 223 292 362	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D4	192 223 278 316 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D4	201 223 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D4	209 223 266 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D4	223 224 245 292 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D4	223 245 269 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D4	223 249 261 278 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D4	223 249 362	3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D4	223 291 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.
D4	223 362	11	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D4	362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D4a	111G 129 223 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D4a	129 162 223 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.
D4a	129 193 223 256 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	3	.	.	4	.	.	.	.	.
D4a	129 223 234 249 311 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.
D4a?	129 223 249 278 311 362	1	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D4a	129 223 274 311 317 362	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.
D4a?	129 223 362	3	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D4b?	181 223 311 319	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D4b	184in/del 186 189 223 319 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D4b	223 287 319 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.
D5	092 129 148 189 223 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D5	092 148 184 189 223 356 362	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D5	092 148 185 189 223 362	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D5	092 148 189 223 256 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.
D5	092 148 189 223 311 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	3	.	.	4	.	.	.	.	.
D5	092 148 189 223 362	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D5	092 164 186 189 223 362	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.
D5	093 148 189 223 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D5	093 189 223 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D5	111 172 189 223 228 362	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D5	111 172 189 223 311 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D5	111 172 189 223 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

Haplogroup	HVS-I Variants	China	Taiwan	Philippines	Thailand	Semang	Senoi	A. Malay	Melayu	Palembang	Pekanbaru	Padang	Bangka	Medan	Tengger	Banjarmasin	Kota Kinabalu	Manado	Palu	Ujung Padang	Toraja	Bali	Lombok	Sumba	Alor	Ambon
D5	126 189 223 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.
D5	148 189 223 309 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	1
D5	148 189 223 362	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	4	.	1	.	.	.	.	.	.
D5	150 189 223 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D5	167 189 223 294 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D5	171 189 223 311 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D5	172 187 189 223 256 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.
D5	172 189 223 259 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.
D5	184 189 223 311 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.
D5	189 201A 223 311 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D5	189 210 223 311 316 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D5	189 223 273 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D5	189 223 311 362	2	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	1	.	1	.	.	.	.	.
D5	189 223 319 360 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D5	189 223 319 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D5	189 223 360 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D5	189 223 362	12	10	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	3	.	.	.	.	.
D5a	092 145 164 189 223 266 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D5a	092 164 167 189 266 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D5a	092 172 189 223 266 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D5a	092 172 189 223 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D5a	164 172 189 223 235 266 291	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D5a	164 172 189 223 266 300 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D5a	164 172 189 223 266 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D5a	169 172 189 223 266 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D5a	172 189 223 266 299 319 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D5a	172 189 223 266 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
E?	148 223 261 362	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
E1	156 223 274 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.
E1	184 223 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.
E1.	185 223 362	.	6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
E1	189 223 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
E1	215 223 362	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	1	.	.	.	.	1	.	.	.
E1	223 248 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1

Haplogroup	HVS-I Variants	China	Taiwan	Philippines	Thailand	Semang	Senoi	A. Malay	Melayu	Palembang	Pekanbaru	Padang	Bangka	Medan	Tengger	Banjarmasin	Kota Kinabalu	Manado	Palu	Ujung Padang	Toraja	Bali	Lombok	Sumba	Alor	Ambon
E1	223 256 324 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.
E1	223 258C 309 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	1	.	.	.	.	.	.
E1	223 292 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.
E1	223 324 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	.	.	.	.	2
E1	223 362	1	8	1	.	.	.	.	1	.	.	.	.	.	.	2	2	2	.	.	.	.	.	.	.	.
E1a	093 223 291 362	.	3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
E1a	117 223 291 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	.	.	.	.	.
E1a	126 189 223 291 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.
E1a	129 148 223 291 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.
E1a	129 223 291 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.
E1a	140 223 291 362	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
E1a	172 223 248 291	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.
E1a	172 223 291 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	3	.	.	.	.	.	.	.	.
E1a	180 223 291 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	7	.	.	.	1	.	.	.	.	.
E1a	185 223 291 362	.	.	.	.	.	.	.	.	.	.	.	.	1	.	1	.	15	3	5	10	2	.	1	3	2
E1a	189 223 290 291 362	.	.	.	.	.	.	.	.	.	1	.	.	.	.	2	.	.	1	.	.	.	.	.	.	.
E1a	189 223 291 362	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	1	.	.	1	.	.	.	.	.	.
E1a	223 224 291 294 362	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	3	1	1	2	1	2	6	2	.
E1a	223 265T 291 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.
E1a	223 288 291 362	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.
E1a	223 291 311 342 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.
E1a	223 291 311 362	.	.	1	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	1	.	.	.	.	.	.
E1a	223 291 356 362	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
E1a	223 291 362	.	7	2	.	.	.	.	1	.	.	1	2	.	.	1	2	15	3	5	10	2	.	1	3	2
E1b	093 223 261 311 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.
E1b	093 223 261 362	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	1	.	.	.	.	.	.
E1b	131 223 261 362	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.
E1b	172 223 261 362	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.
E1b	223 261 288 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	.	.	.	.	.	.	.	.	.	.
E1b	223 261 294 362	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
E1b	223 261 311 362	.	.	.	.	.	.	.	1	.	.	.	.	.	.	1	.	3	1	1	2	1	2	6	2	.
E1b	223 261 362	.	.	.	.	.	.	.	2	1	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F	093 304 311 362	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F	118 129 304	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

Haplogroup	HVS-I Variants	China	Taiwan	Philippines	Thailand	Semang	Senoi	A. Malay	Melayu	Palembang	Pekanbaru	Padang	Bangka	Medan	Tengger	Banjarmasin	Kota Kinabalu	Manado	Palu	Ujung Padang	Toraja	Bali	Lombok	Sumba	Alor	Ambon
F	129 304 362 359	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.
F	157 169 256 304 311 335	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F	157 256 304 335	1	4	1	.	.	.	.	.	.	.	.	.	.	.	.	3	1	1	.	.	.	.	1	2	.
F	157 256 304 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.
F	157 256 335	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	.
F	157 304	.	.	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F	228 304 362	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F	256 304 335	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F	304	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	1	.	.	.	.
F	304 362	1	.	.	1	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.
F1a	092 129 172 189 304	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.
F1a	092 129 172 294 304 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1a	092 172 189 304	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.
F1a	093 129 172 294 304 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1a	093 129 172 304	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1a	104 129 172 294 304 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2
F1a	109C 129 172 304	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1a	127 129 172 294 304 362	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1a	127 129 172 304	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1a	129 132 172 304	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1a	129 134 172 301 304	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1a	129 153 172 223 263 304	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.
F1a	129 172 173 294 304 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1a	129 172 184 304	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1a	129 172 187 304	3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	.	.	.	.	.	.	.	.	.
F1a	129 172 192 294 304 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.
F1a	129 172 214 304 311	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1a	129 172 218 304 354	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1a	129 172 218A 304	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1a?	129 172 223 291 305	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.
F1a	129 172 242 304	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1a	129 172 265 295 304	.	.	.	1	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.
F1a	129 172 271 304 311	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1a	129 172 274 304	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

Haplogroup	HVS-I Variants	China	Taiwan	Philippines	Thailand	Semang	Senoi	A. Malay	Melayu	Palembang	Pekanbaru	Padang	Bangka	Medan	Tengger	Banjarmasin	Kota Kinabalu	Manado	Palu	Ujung Padang	Toraja	Bali	Lombok	Sumba	Alor	Ambon
F1a	129 172 294 304	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.
F1a	129 172 294 304 362	.	3	.	.	.	.	.	.	.	2	.	.	.	.	1	.	3	1	.	7	.	.	3	2	.
F1a	129 172 295 304	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1a	129 172 295 304 311	1	.	.	.	.	.	.	.	.	.	.	.	.	6	.	.	.	.	.	.	.	.	.	.	.
F1a	129 172 301 304	.	.	.	.	.	.	.	.	.	1	1	.	.	.	.	.	.	.	1	.	.	.	.	.	.
F1a	129 172 301 304 362	.	.	.	.	.	.	.	4	.	1	2	1	1	.	.	1	1	.	1	.	3	1	1	1	.
F1a	129 172 304	29	.	1	7	.	.	.	.	.	1	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.
F1a	129 172 304 309	.	.	.	.	.	.	.	2	1	.	1	.	.	2	1	1	3	.	.	.	.	.	2	.	3
F1a	129 172 304 311	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.
F1a	129 172 304 311 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.
F1a	129 172 304 352	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	3	.	.	.
F1a	129 172 304 362	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1a	164 172 189 294 304	.	.	1	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.
F1a	168 172 189 311 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1a	172 180 304 311	.	.	1	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1a	172 189 304	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1a	172 274 304	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.
F1a	172 304	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1a	172 304 311	1	6	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1a/M 12	108 129 172 223 234 290	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1a1	093 129 162 168 172 304	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1a1	093 129 162 172 295 304	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1a1	093 129 162 172 304	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1a1	118 129 162 172 304	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	1	.	.	.	.	.	.	.
F1a1	129 145 162 172 304	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.
F1a1	129 153 162 172 304 311	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.
F1a1	129 162 172 189 274 304	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1a1	129 162 172 189 304 311	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	.	.	.	.
F1a1	129 162 172 224 304	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1a1	129 162 172 243 304	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1a1	129 162 172 260 304	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1a1	129 162 172 278 304	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1a1	129 162 172 292 304	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

Haplogroup	HVS-I Variants	China	Taiwan	Philippines	Thailand	Semang	Senoi	A. Malay	Melayu	Palembang	Pekanbaru	Padang	Bangka	Medan	Tengger	Banjarmasin	Kota Kinabalu	Manado	Palu	Ujung Padang	Toraja	Bali	Lombok	Sumba	Alor	Ambon
F1a1	129 162 172 304	13	5	.	.	.	.	.	3	.	.	.	.	.	.	1	.	.	.	.	.	.	.	1	.	.
F1a1	129 162 172 304 311	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.
F1a1	129 162 172 304 311 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1a1	129 162 172 304 335	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1a1	129 162 172 304 362	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1a1	162 172 174 304	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1a1a	092 108 129 162 172 234 299 304	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1a1a	106 108 129 162 172 223	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1a1a	108 111 129 162 172 189 304	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.
F1a1a	108 124 129 162 172 304	1	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1a1a	108 129 147 162 172 304	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1a1a	108 129 162 170 172 304	.	.	.	.	.	.	.	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1a1a	108 129 162 172	.	.	.	.	.	6	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1a1a	108 129 162 172 184 304	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1a1a	108 129 162 172 189 304	1	.	.	.	.	.	4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1a1a	108 129 162 172 214 304	4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.
F1a1a	108 129 162 172 234 299 304	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1a1a	108 129 162 172 256 304	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1a1a	108 129 162 172 261 304	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1a1a	108 129 162 172 274	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1a1a	108 129 162 172 274 304	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1a1a	108 129 162 172 293 304	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.
F1a1a	108 129 162 172 295 304	1	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1a1a	108 129 162 172 304	9	.	.	18	.	16	.	7	1	2	.	1	1	.	.	.	1	1	.	1	1	4	1	1	1
F1a1a	108 129 162 172 304 311	1	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1a1a	108 129 162 172 304 355	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1b	111 189 232A 249 304 311	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1b	126 178 189 304	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1b	126 189 304	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1b	129 145 189 232A 249 304 311 344	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1b	129 189 304	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1b	140 189 266 284 304	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

Haplogroup	HVS-I Variants	China	Taiwan	Philippines	Thailand	Semang	Senoi	A. Malay	Melayu	Palembang	Pekanbaru	Padang	Bangka	Medan	Tengger	Banjarmasin	Kota Kinabalu	Manado	Palu	Ujung Padang	Toraja	Bali	Lombok	Sumba	Alor	Ambon
F1b	172 180 189 304	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1b	189 232A 249 264 304 311	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1b	189 232A 249 266 304	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1b	189 232A 249 304 311	3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1b	189 249 300 304	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1b	189 266 284 304	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1b	189 284 304	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.
F1b	189 304	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1b	189 304 311	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1c	111 129 172 189 304	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1c	111 129 172 304	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1c	111 129 189 304	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1c	111 129 243 266 304	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1c	111 129 266 304	4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1c	111 129 294 304	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F1c	111 129 304	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F2	0	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F2	092A 291 304 359	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F2	129 203 304	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F2?	145 192 223 291 304 362	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F2	167 203 304	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F2	167 203 304 318	9	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F2	192 304	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F2	203 239 291 304	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F2	203 304	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F2	209 304	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F2	261	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F2	299 304	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F2a	092A 170C 189 291 294 304	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F2a	092A 170T 183T 189 291 304	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F2a	092A 170T 189 291 304	3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F2a	092A 291 304	2	.	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F2a	145 192 223 291 304	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F2a	185 258 266A 291 304	4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

Haplogroup	HVS-I Variants	China	Taiwan	Philippines	Thailand	Semang	Senoi	A. Malay	Melayu	Palembang	Pekanbaru	Padang	Bangka	Medan	Tengger	Banjarmasin	Kota Kinabalu	Manado	Palu	Ujung Padang	Toraja	Bali	Lombok	Sumba	Alor	Ambon
F2a	185 266A 291 304	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F2a	266 291 304	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F2a	291 304	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F3	290 298 357 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.
F3a	093 111 126 192 249 263 298 355 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F3a	093 111 192 249 298 355 362	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F3a	093 260 298 355 362	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F3a	111 192 249 263 264insC 298 355 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F3a	127 260 298 355 362	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F3a	209 298 311 355 362	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F3a	209 298 355 362	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F3a	260 298 355 362	2	.	3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F3b	093 220C 265 298 311 362	.	6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F3b	093 220C 265 298 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	.	.	.	.	.	.	.	.	.	.
F3b	093 220C 298 362	.	.	1	.	.	.	.	.	.	.	.	.	.	.	4	.	.	.	.	.	.	.	.	.	.
F3b	168 220C 265 298 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F3b	220C 261 265 298 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.
F3b	220C 265 274 298 311 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.
F3b	220C 265 298 311 362	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.
F3b	220C 265 298 335 362	.	.	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F3b	220C 265 298 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.
F3b	220C 298 311 362	.	10	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F4	126 140 207 304 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F4	129 185 207	.	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F4	207 304 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F4b?	129 218 265 304 311 355	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F4b	129 218 304 311	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F4b	170 218 304 311	.	.	.	.	.	.	.	.	.	1	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.
F4b	218 241 304 311	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F4b	218 304 311	.	29	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.
G	092 129 209 223 325	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.
G	093 129 209 223 272	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	.	.	.	.	.	.	.	.	.

Haplogroup	HVS-I Variants	China	Taiwan	Philippines	Thailand	Semang	Senoi	A. Malay	Melayu	Palembang	Pekanbaru	Padang	Bangka	Medan	Tengger	Banjarmasin	Kota Kinabalu	Manado	Palu	Ujung Padang	Toraja	Bali	Lombok	Sumba	Alor	Ambon
G	093 129 223 234 290 311	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.
G	118 129 192 223 256 272	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.
G	129 166C 189 223 287 319	.	.	.	.	.	.	.	.	.	1	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.
G	129 172 223 324 290 312	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
G	129 189 218 223	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.
G	129 209 223 272	6	.	.	.	.	.	.	2	1	.	1	2	.	.	1	1	.	.	2	.	1	.	.	.	.
G	129 209 223 272 311	.	.	.	.	.	.	.	.	.	.	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.
G	129 209 223 311 325	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.
G	129 209 223 325	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.
G .	129 223 234 290 311	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	1	.	.	.
G2	092 140 172 189 223 278	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
G2	093 169 184A 223 278	.	.	.	.	.	.	.	1	.	.	.	.	.	.	1	.	.	.	.	.	.	1	.	.	.
G2	093 184A 223 278	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
G2	093 189 222 223 278	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.
G2	093 189 223 265 278	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1
G2	093 189 223 274 278 311	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
G2	093 189 223 278 319	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.
G2	093 209 223 224 263 265 278 319	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.
G2	093 209 223 224 263 274 278 319 356	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.
G2	093 209 223 224 263 278 319	.	.	.	.	.	.	.	2	.	3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
G2	093 223 249 259 278 291 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.
G2	093 223 278 310	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.
G2	124 189 278 292 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.
G2	126 129 189 278	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	3	.	.	.	.	.	.	.	.	.
G2	129 223 278 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
G2	140 172 189 223 278	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
G2	148 223 259 278 319	.	.	.	.	.	.	.	.	.	.	.	.	.	2	1	.	.	.	.	.	.	.	.	.	.
G2	166 223 278 335 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
G2	172 173 223 278 311	.	.	.	.	.	.	.	.	1	.	.	.	.	.	2	.	.	1	.	.	.	.	.	.	.
G2	184A 213 223 278	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.
G2?	184A 223	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1
G2	188 189 223 278 288	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.

Haplogroup	HVS-I Variants	China	Taiwan	Philippines	Thailand	Semang	Senoi	A. Malay	Melayu	Palembang	Pekanbaru	Padang	Bangka	Medan	Tengger	Banjarmasin	Kota Kinabalu	Manado	Palu	Ujung Padang	Toraja	Bali	Lombok	Sumba	Alor	Ambon
G2	189 222 223 278 352	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.
G2	189 223 278	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	1	1	.	1	.	.	.
G2	189 223 278 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
G2	196 223 274 278 290	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.
G2	209 223 224 263 278 319	.	.	.	.	.	.	.	2	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.
G2	223 243 262 278 311 319	.	.	.	.	.	.	.	.	.	.	.	.	1	.	1	.	.	.	.	1	.	.	.	.	.
G2	223 278	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.
G2	223 278 294	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
G2	223 278 311	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.
G2	223 311 362	1	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.
G2	234 256 278 294	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.
G2	234 278 294	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
G2a	111 209 223 227 274 278 326 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
G2a	111 223 227 278 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
G2a	189 194 223 227 278 311 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
G2a	189 223 227 256 278 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
G2a	209 223 227 234 278 309 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
G2a	223 227 262 278 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
G2a	223 227 272 278 319 362 365	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
G2a	223 227 278 311 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
G3?	129 142 166 223 255 274 294 327A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.
H?	192	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
H/F/M 7	0	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
HV2	217 240	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
I	129 223 311	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
L2a	223 278 294	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
L2a1	223 274 278 294 309	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M	0	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.
M	093 104 111 223 235 362	3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M	093 104 111 223 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M	093 129 223 311	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

Haplogroup	HVS-I Variants	China	Taiwan	Philippines	Thailand	Semang	Senoi	A. Malay	Melayu	Palembang	Pekanbaru	Padang	Bangka	Medan	Tengger	Banjarmasin	Kota Kinabalu	Manado	Palu	Ujung Padang	Toraja	Bali	Lombok	Sumba	Alor	Ambon
M	093 133 176 223	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.
M	093 145 223 234 249 290	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.
M	093 189 209 223	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.
M	093 193	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.
M	093 209 223 325	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.
M	093 223 256 274	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M	093 223 311 362	.	.	.	.	.	.	.	.	.	.	.	.	.	4	1	.	.	.	.	.	.	.	.	.	.
M	093 223 362	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M	095 129 140 223 265C 271	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M	124 166del 214 223	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.
M?	126 189 223 325	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M	126 223 362	1	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M	129 140 223 265C 271	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.
M?	129 140 271	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M	129 148 172 223 256 305 309	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.
M	129 172 223 256 305 309	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M	129 179 209 223 272	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M	129 223 256	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M	129 223 270 362	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M	129 223 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.
M	130 185 223	.	.	1	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.
M?	134 219 223 225 291	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.
M	145 176 224 233 311	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M	145 181 192 223 266 291 304	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.
M	145 181 192 223 291 304	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.
M	148 189 223 246T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.
M	148 189 223 246T 311 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.
M	148 189 362	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M	149 185 223 362	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M	158 223 234 362	4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.
M	162del 214 223	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M	168 189 209 223 233 304	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.
M	172 173 223	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M	172 174 223 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

Haplogroup	HVS-I Variants	China	Taiwan	Philippines	Thailand	Semang	Senoi	A. Malay	Melayu	Palembang	Pekanbaru	Padang	Bangka	Medan	Tengger	Banjarmasin	Kota Kinabalu	Manado	Palu	Ujung Padang	Toraja	Bali	Lombok	Sumba	Alor	Ambon
M	172 223 245A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.
M	172 223 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.
M	179 223 294	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M	189 192 223 291 362	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	1	.	.	.	.	.
M	189 195A 241 265C 311	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.
M	189 209 223 300	.	.	.	.	.	.	.	.	.	.	.	2	.	.	.	.	.	2	.	.	.	.	1	.	.	.
M	189 223	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.
M	189 223 227 291 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M	189 223 229 294 311 362	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	2
M	189 293C 325 362	1	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.
M	192 362	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M	219 223 290 291	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	1	.	.	.	.	.	1	.	.	.	.
M?	219 223 290 291 295	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M	219 223 362	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.
M	223	2	1	.	.	.	2	.	2	.	.	.	.	1	.	.	1	.	.	.	.	.	1	.	.	.	.
M	223 234 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.
M	223 239 289 356	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M	223 243 311 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.
M	223 246T 311 362	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M	223 249A	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M	223 254 362	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	1	.	.	.	.	.
M	223 266 284 290	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.
M	223 293	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M	223 293 311 362	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	1	.	.	.	.	.	.	.
M	223 299 311	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	1	.	.	.	.	.	.	.	.
M	223 299 311 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.
M	223 304 325 344 362 381	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.
M	223 311	1	.	.	.	.	2	.	2	.	.	.	.	.	.	1	1	.	.	.	.	.	1	.	.	.	.
M	223 311 335 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.
M	223 311 362	1	.	.	4	.	.	.	1	.	.	.	2	.	3	3	.	.	2	.	.	2	.	1	.	.	.
M	223 362	.	.	.	1	.	.	.	.	.	.	1	.	1	.	1	.	.	.	.	.	.	.	.	.	.	2
M10	092 223 311	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M10	093 129 193 223 311 357	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M10	093 129 223 311 357	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

Haplogroup	HVS-I Variants	China	Taiwan	Philippines	Thailand	Semang	Senoi	A. Malay	Melayu	Palembang	Pekanbaru	Padang	Bangka	Medan	Tengger	Banjarmasin	Kota Kinabalu	Manado	Palu	Ujung Padang	Toraja	Bali	Lombok	Sumba	Alor	Ambon
M10	093 193 223 311 357	1	.	.	.	.	.	.	.	.	.	.	.	.	6	.	.	.	.	.	.	.	.	.	.	.
M10	094 129 223 263 274 311 343 357	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M10	129 223 311	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M10?	172 223 234 290 311	1	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M10?	179 223 264 311	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.
M10	189 213 223 271 311	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.
M10?	223 234 300 311	.	.	.	.	.	.	.	.	.	.	.	4	.	1	.	.	.	.	.	.	.	.	.	.	.
M10	223 263 274 311 343 357	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M10	223 269 271 311	1	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M12a	093 129 223 234 286 290 311 362	.	.	.	.	.	.	.	1	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.
M12a	093 129 223 234 290 311	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M12a	129 223 234 290 362	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M12a	223 234 249 261 290 360	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.
M12a	223 234 261 290	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M12a	223 234 287 290 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M12a	223 234 290 325 362	.	.	.	4	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.
M2	093 223 231 319	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M21a	093 129 145 223 256 271 362	.	.	.	.	6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M21a	093 129 217 223 256 271 362	.	.	.	.	.	.	2	1	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.
M21a	093 129 223 256 271	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M21a	093 129 223 256 271 293 362	.	.	.	.	38	1	1	4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M21a	093 129 223 256 271 362	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M21a	129 145 223 256 271 362	.	.	.	.	3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M21a	129 223 256 271 362	.	.	.	16	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M21b	093 129 223 263	.	.	.	.	3	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M21b	129 223 263	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	.	.	.	.	.	.	.	.	.	.
M21b	136 217 223 319	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M21b	172 223 239 263 325	.	.	.	.	.	.	5	.	.	.	.	.	1	.	.	.	1	.	.	.	.	.	.	.	.
M21b	217 223 319	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M21c	223 242 319	.	.	.	.	.	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M22	093 184 223 290 304	.	.	.	.	.	.	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M22	153 223 290 304	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

Haplogroup	HVS-I Variants	China	Taiwan	Philippines	Thailand	Semang	Senoi	A. Malay	Melayu	Palembang	Pekanbaru	Padang	Bangka	Medan	Tengger	Banjarmasin	Kota Kinabalu	Manado	Palu	Ujung Padang	Toraja	Bali	Lombok	Sumba	Alor	Ambon
M22	223 290 304	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M3	126 147 153 223	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.
M3	126 214A 223 271 278 298	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M5	129 223	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M7	0	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	1	.	.
M7	129 155 219 223 356 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M7	129 219 223 272 362	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M7	172 223 287 311	1	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.
M7	172 223 311	1	.	.	.	.	.	.	.	.	.	.	.	3	.	.	.	.	3	.	.	.	.	.	1	.
M7	189 223 362	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	1	.	.	.	.	.	.	.	.	.
M7	213	1	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.
M7	223 291 362	.	.	.	.	.	.	.	.	.	1	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.
M7	223 362	.	.	.	.	.	.	.	.	.	.	.	.	3	.	.	.	.	.	.	.	.	.	.	.	.
M7	230 304	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.
M7	354	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M7a1	209 223 324	.	.	1	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.
M7b	102 223 297 300	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M7b	126 129 223 297	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.
M7b	129 166 223 297	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M7b	129 172 189 223 297	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.
M7b	129 189 223 248 297	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M7b	129 189 223 248 297 300	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M7b	129 189 223 293 297	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M7b	129 189 223 297	5	.	.	2	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	1	.	.	.	.
M7b	129 223 297	4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M7b	168 223 297	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M7b	176 223 297	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M7b	182 193 223 297	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M7b	223 249 297 360	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M7b	223 264 297	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M7b	223 297	7	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M7b	223 297 316 357	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M7b	223 297 327	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

Haplogroup	HVS-I Variants	China	Taiwan	Philippines	Thailand	Semang	Senoi	A. Malay	Melayu	Palembang	Pekanbaru	Padang	Bangka	Medan	Tengger	Banjarmasin	Kota Kinabalu	Manado	Palu	Ujung Padang	Toraja	Bali	Lombok	Sumba	Alor	Ambon
M7b1	092 129 165 168insA 192 223 297	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M7b1	092 129 192 223 297	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M7b1	093 129 136 192 223 297	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M7b1	093 129 192 297	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M7b1	126 129 192 223 297	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	1	.	.	.	1
M7b1	129 140 189 192 223 265 297	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.
M7b1	129 188G 192 223 297	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M7b1	129 189 192 193 223 297	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M7b1	129 189 192 215 223 297	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	1	.	.	.	.
M7b1	129 189 192 223 234 295 297	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M7b1	129 189 192 223 297	1	.	.	.	.	.	.	2	.	.	.	.	3	.	.	.	.	1	.	.	.	1	.	.	.	.
M7b1	129 192 223 243 297	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M7b1	129 192 223 271 297	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M7b1	129 192 223 291 297	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M7b1	129 192 223 297	20	.	.	6	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M7b1	129 192 223 297 301G	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M7b1	129 192 223 297 357	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M7b1	129 192 223 297 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M7b1	189 192 223 294G 297	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.
M7b1	192 223 297	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M7b2	129 189 223 297 298	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M7b2	129 189 223 297 298 325	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M7b3	126 129 297 324	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.
M7b3	129 145 297 324	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M7b3	129 171 278 297 324	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M7b3	129 189 192 297 324	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M7b3	129 189 297 324	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M7b3	129 192 224 297 324	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M7b3	129 192 297 324	.	6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M7b3	129 295 297 324	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	4	.	.	.
M7b3	129 297 324	.	12	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	.	.	.	.	.	.
M7b3	145 297 324	.	3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M7b3	297 324	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

Haplogroup	HVS-I Variants	China	Taiwan	Philippines	Thailand	Semang	Senoi	A. Malay	Melayu	Palembang	Pekanbaru	Padang	Bangka	Medan	Tengger	Banjarmasin	Kota Kinabalu	Manado	Palu	Ujung Padang	Toraja	Bali	Lombok	Sumba	Alor	Ambon
M7c1	129 223 295	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M7c1	172 223 293T 295	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M7c1	223 224 287 295	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.
M7c1	223 261 295	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M7c1	223 293T	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M7c1	223 293T 295	2	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M7c1	223 295	7	.	1	1	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	2	.	.	.	.	.
M7c1a	093 223 319	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M7c1a	223 295 296 319	1	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	1	.
M7c1a	295 319	1	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.
M7c1b	223 294 295	3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M7c1c	093 223 291 295 337 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	1	.	.	.	.	.	.
M7c1c	093 223 295 337 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	.	.	.	.	.	.	.	.	.
M7c1c	093 223 295 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	2	.	.
M7c1c	129 223 295 362	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M7c1c	145 223 295 362	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	3	.	.	.	.	.	.	.
M7c1c	150 223 274 295 311 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.
M7c1c	168 223 265T 295 362	.	4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	.	.	.	.	.
M7c1c	168 223 295 362	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M7c1c	179 223 295 362	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	1	.	.	.	.	.	.	.
M7c1c	185 223 295 362	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.
M7c1c	189 213 223 295 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.
M7c1c	192 223 295 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M7c1c	223 254 295 362	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M7c1c	223 278 295 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	1	.	.	.	.	.	.	.
M7c1c	223 292 295 362	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.
M7c1c	223 295 311 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	.	.
M7c1c	223 295 346C 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	.	.	.	.	.	.	.	.	.
M7c1c	223 295 356 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	3	.	.	.	.	.	.	.	.
M7c1c	223 295 362	.	4	9	1	.	.	8	5	6	3	.	1	1	3	1	3	8	6	1	4	1	1	3	1	1
M7c1c	295 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.
M8a	134 184 223 298 319	3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M8a	184 189 223 298 311 319	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M8a	184 189 223 319	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

Haplogroup	HVS-I Variants	China	Taiwan	Philippines	Thailand	Semang	Senoi	A. Malay	Melayu	Palembang	Pekanbaru	Padang	Bangka	Medan	Tengger	Banjarmasin	Kota Kinabalu	Manado	Palu	Ujung Padang	Toraja	Bali	Lombok	Sumba	Alor	Ambon
M8a	184 209 223 293 298 311 319	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M8a	184 223 260 298 319	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M8a	184 223 278 298 319	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M8a	184 223 293 298 319	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M8a	184 223 293C 298 319	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M8a	184 223 298 319	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M8a	223 298 319	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M9	093 223 234	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M9	093 223 234 362	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M9	145 148 223 234 294 316	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M9.	145 223 234 316	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M9	209 223 234 291 316 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M9	223 234 255 271 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M9	223 234 291 316 362	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M9	223 234 311 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M9	223 234 316 362	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M9a	223 234 291 316	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
N	111 168 172 189 223 242 263 311 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.
N	111 168 172 189 223 311 319 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.
N	111 168 172 189 223 311 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	.	.	.	.	.	.	.	.	.	.
N	111 172 189 223 311 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.
N	172 223 278 291A 298 325 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
N	177 223 263 266 274 311 343	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
N	189 223	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.
N	189 223 355	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
N	223 294	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.
N21	172 193 223 344	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
N21	193 223	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	1	.	.	.	.
N21	193 223 249 291 319	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.
N21	193 223 291 319	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.
N21	193 223 311 344	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
N21	193 291	.	.	.	.	.	.	17	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

Haplogroup	HVS-I Variants	China	Taiwan	Philippines	Thailand	Semang	Senoi	A. Malay	Melayu	Palembang	Pekanbaru	Padang	Bangka	Medan	Tengger	Banjarmasin	Kota Kinabalu	Manado	Palu	Ujung Padang	Toraja	Bali	Lombok	Sumba	Alor	Ambon
N21	193 291 327	.	.	.	.	.	.	7	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
N22	093 168 172 223 249	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.
N22	093 168 223 249 278 295	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.
N22	168 172 223 249	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	.	.
N22	168 223 249	.	.	.	.	.	.	4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
N9a	092 145 172 223 245 257A 261	2	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
N9a	111 129 192A 223 257A 261	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
N9a	111 129 223 257A 261	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
N9a	111 129 223 257A 261 325	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
N9a	129 162 223 250 257A 261	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
N9a	129 189 223 257A 261	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
N9a	129 223 257A 261	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
N9a	145 172 223 245 257A 261	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
N9a	166c 173 223 250 257A 324	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
N9a	172 189 223 257A 261	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
N9a	172 223 257A 261 295	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
N9a	172 223 257A 261 311	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
N9a	189 223 257A 261 311	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
N9a	223 257A	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
N9a	223 257A 261	3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
N9a	223 257A 261 295	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
N9a	223 257A 261 311	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
N9a	223 257A 261 357	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
N9a	223 257A 261 362	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
N9a	223 257A 311	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
N9a1	136 223 257A 261 292 294	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
N9a1	179 223 257A 261 292 294	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
N9a1	187 223 257A 261 292 294	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.
N9a1	189 223 257A 261 292	.	.	.	1	.	.	.	.	.	.	3	3	.	.	.	.	.	.	.	.	.	.	.	.	.
N9a1	223 257A 261 292	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.
N9a1	223 257A 261 292 294	.	.	.	.	8	3	2	2	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.
N9a1	223 257A 261 292 294 304	.	.	.	.	.	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
N9a1	223 257A 261 292 294 357	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.
P	093 172 266 270	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.

Haplogroup	HVS-I Variants	China	Taiwan	Philippines	Thailand	Semang	Senoi	A. Malay	Melayu	Palembang	Pekanbaru	Padang	Bangka	Medan	Tengger	Banjarmasin	Kota Kinabalu	Manado	Palu	Ujung Padang	Toraja	Bali	Lombok	Sumba	Alor	Ambon
P	093 176 266 270 357	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.
P	176 221 266 325 357	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	.	.	.	.	.	.	.	.
P	176 266 357	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
P	249 319	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.
Q	092 129 144 148 169 223 265C 311 343	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.
Q	129 144 148 153 162 192 223 241 249 265C 311 343	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.
Q	129 144 148 154 209 222 241 265c 311 343	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1
Q	129 144 148 162 192 223 241 249 265C 311 343	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Q	129 144 148 172 223 241 242 265C 311 343	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	1	.
Q	129 144 148 172 223 241 256 265C 311 343	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.
Q	129 144 148 172 223 241 265C 311 343	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	3	.
Q	129 144 148 172 223 242 265C 311 343	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.
Q	129 144 148 193 223 241 265C 311 343 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2
Q	129 144 148 209 223 241 265C 311 343	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.
Q	129 144 148 222/3insA 223 241 265C 311 343	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.
Q	129 144 148 223 241 255 265C 300 311 343	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Q	129 144 148 223 241 265C 299 311 343 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.
Q	129 144 148 223 241 265C 311 343	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	1	.	.
Q	129 144 148 223 265C 311 343	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.
Q	129 144 148 241 265C 311 343	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	5	.

Haplogroup	HVS-I Variants	China	Taiwan	Philippines	Thailand	Semang	Senoi	A. Malay	Melayu	Palembang	Pekanbaru	Padang	Bangka	Medan	Tengger	Banjarmasin	Kota Kinabalu	Manado	Palu	Ujung Padang	Toraja	Bali	Lombok	Sumba	Alor	Ambon
Q	129 148 223 241 265C 311 343	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.
Q	129 172 173 223 241	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	1	.
Q	129 223 241 311	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
R	093 157 201	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.
R	129 256 290	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
R	172 192A 207 325	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
R	187 241 269 319 342	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.
R	189 259 311	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
R	189 311 365	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	.	.	.	.
R	256 290	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.
R	293 311 355	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
R/F?	093 129 256 304 357	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	.	.
R1	166 266 304 311	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	1	.	.	.	.	4	2	1	.	.
R21	168 209 295 296 304	.	.	.	.	.	4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.
R21	168 295 296 304	.	.	.	.	34	15	.	1	.	.	.	.	.	.	2	.	.	.	.	.	.	.	1	.	.
R21	168 295 304	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	1	.
R21	295 296 304 354	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.
R21/R 9?	093 168 187 288 304	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	1	.
R22	189 249 286 288	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.
R22	249 259 288 301 304	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
R22	249 270A 288 304 311 319	.	.	.	1	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.
R22	249 288	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	.	.	.	.	.	.	.	.	.	.
R22	249 288 295 304	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	2	.	.
R22	249 288 301 304	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	1	.	.	.	.	4	2	1	.	.
R22	249 288 304	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.
R22	249 288 304 344	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.
R22	249 288 317C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.
R22	249 288 317C 319	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	1	.
R22	288 304	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.
R9	124 189 209 293C 304 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.
R9	129 209 223 233 259 274 290 304	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.
R9?	140 189 304	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

Haplogroup	HVS-I Variants	China	Taiwan	Philippines	Thailand	Semang	Senoi	A. Malay	Melayu	Palembang	Pekanbaru	Padang	Bangka	Medan	Tengger	Banjarmasin	Kota Kinabalu	Manado	Palu	Ujung Padang	Toraja	Bali	Lombok	Sumba	Alor	Ambon
R9'	145 266 304 309 325 356	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
R9	209 223 233 274 304	.	.	1	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.
R9	209 223 234 261 290 304	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.
R9	209 304 311	.	.	.	.	.	.	.	.	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
R9b	093 192 288 304 309	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
R9b	093 304 309	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
R9b	124 148 184 304 309	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
R9b	124 148 304 309	1	.	.	.	.	.	1	.	.	.	.	.	.	2	.	.	.	.	.	.	.	.	.	.	.
R9b	129 288 304 309	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	2	.	.	.	.	.	.	.	.
R9b	170 223 288 304 309	.	.	.	.	.	.	21	1	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.
R9b	184 192 288 304 309	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
R9b	192 234 288 293 304 309	.	.	.	.	.	.	.	.	.	.	1	1	.	.	.	.	.	.	.	.	.	.	.	.	.
R9b	192 234 288 304 309	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
R9b	192 239 304 309	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
R9b	192 288 304 309	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
R9b	192 304 309	2	.	.	2	.	.	.	2	.	3	.	2	5	.	1	2	1	1	.	2	.	.	.	.	.
R9b	223 288 304 309	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
R9b	304 309	2	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.
T1	126 163 186 189 294	1	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.
U7	207 309 318T	.	.	.	.	.	.	.	.	.	1	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.
Y1	126 193 231 266	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.
Y1	126 231 266	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Y1	126 231 266 293	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Y1	126 231 266 325	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Y2	093 126 231 311	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.
Y2	126 192 231 311	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.
Y2	126 213insA 231 311	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Y2	126 231 264 311	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Y2	126 231 284 311	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Y2	126 231 311	1	2	5	.	.	.	.	2	.	3	.	2	5	.	1	2	1	1	.	2	.	.	.	.	.
Y2	126 231 311 362	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Y2	231 311	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.
Z	129 185 189 223 260 298	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.
Z	136 185 223 260 298	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

Ambon	.	.	.
Alor	.	.	.
Sumba	.	.	.
Lombok	.	.	.
Bali	.	.	.
Toraja	.	.	.
Ujung Padang	.	.	.
Palu	.	.	.
Manado	.	.	.
Kota Kinabalu	.	.	.
Banjarmasin	.	.	.
Tengger	.	.	.
Medan	.	.	.
Bangka	.	.	.
Padang	.	.	.
Pekanbaru	.	.	.
Palembang	.	1	.
Melayu	.	.	.
A. Malay	.	.	.
Senoi	.	.	.
Semang	.	.	.
Thailand	.	1	.
Philippines	.	.	.
Taiwan	.	.	.
China	1	9	1
HVS-I Variants	185 189 223 224 260 261 298 302	185 223 260 298	185 223 260 298 302
Haplogroup	Z	Z	Z

Appendix IV – Haplogroup Frequencies

Appendix IV – Haplogroup Frequencies

Haplogroup	NW China[1]	NE China[2]	SW China[3]	SE China[4]	Thailand	Malay Peninsula[5]	Philippines	Medan	Pekanbaru	Palembang	Bangka	Padang	Tengger	Banjarmasin	Kota Kinabalu	Bali	Mataram	Waingapu	Manado	Palu	Toraja	Ujung Padang	Ambon	Alor
N*	0	1	0	2	0	1	0	0	0	0	0	1	0	4	2	0	0	0	0	0	0	0	0	0
N21	0	0	0	0	0	26	0	0	0	1	0	0	0	0	0	1	0	0	0	1	0	0	0	0
N22	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	4	1	0	0	0	0	0
N9a	1	7	3	7	4	19	0	0	0	0	3	3	1	1	0	0	0	0	1	1	0	0	0	0
Y2	0	3	0	0	0	2	7	7	3	0	2	0	0	1	2	1	0	0	1	1	2	0	0	0
R*	0	2	0	1	0	2	1	0	0	0	0	0	0	0	1	3	0	1	0	0	0	0	0	0
P	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	2	3	0	0	0	0	0
B*	1	2	8	1	1	3	0	0	2	0	1	4	1	0	0	0	0	0	0	0	0	1	0	0
B4*	2	8	10	5	13	3	0	1	0	2	1	0	0	10	0	4	0	0	1	0	1	1	0	0
B4a*	0	8	6	21	13	3	0	0	8	1	0	2	0	7	4	1	2	1	3	0	4	5	4	0
B4a1	0	1	0	0	0	0	0	0	0	0	0	0	0	2	0	0	1	0	1	0	5	3	6	1
B4b	3	6	1	12	0	1	0	0	1	0	0	0	0	2	1	0	0	0	1	1	0	0	2	0
B4c	0	1	1	6	0	3	0	0	6	0	1	1	1	2	1	3	0	0	2	0	4	3	0	3
B5*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B5a	2	8	9	10	26	12	0	2	2	3	0	0	1	5	2	4	3	0	3	0	0	3	3	2
B5b	4	2	1	3	0	16	0	4	0	0	0	0	0	2	0	0	1	0	3	1	1	1	0	0
R9*	0	1	0	2	0	1	1	1	2	0	1	0	0	0	0	1	0	0	0	0	0	1	0	1
R9b	1	0	2	5	3	25	0	0	2	0	0	1	2	1	0	0	1	0	2	0	0	0	0	0
R21	0	0	0	0	0	56	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R22	0	0	0	0	2	0	0	1	0	0	0	0	1	2	1	5	5	4	0	0	0	1	0	1
F*	0	4	2	3	7	0	4	2	0	0	0	0	0	1	5	1	1	1	1	1	0	0	0	5
F1a*	2	6	31	8	13	6	9	5	1	5	2	1	6	9	4	4	5	6	7	3	7	4	5	4

Haplogroup	NW China[1]	NE China[2]	SW China[3]	SE China[4]	Thailand	Malay Peninsula[5]	Taiwan	Philippines	Medan	Pekanbaru	Palembang	Bangka	Padang	Tengger	Banjarmasin	Kota Kinabalu	Bali	Mataram	Waingapu	Manado	Palu	Toraja	Ujung Padang	Ambon	Alor
F1a1*	2	5	4	11	3	4	6	0	0	0	0	0	0	0	1	1	4	0	1	0	1	0	0	0	0
F1a1a	0	1	9	11	22	37	0	0	1	2	5	1	0	0	1	0	2	4	2	1	1	1	0	1	1
F1b	7	8	10	6	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
F1c	1	3	4	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F2	2	5	23	8	6	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
F3	2	1	2	3	4	1	17	6	0	0	0	0	0	0	7	2	0	0	0	0	0	0	0	0	1
F4	0	0	0	1	2	0	27	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M*	1	3	6	17	11	19	1	4	4	3	1	5	1	6	9	5	1	6	4	0	4	1	0	0	1
M7*	1	1	1	2	2	0	0	0	3	1	0	0	1	0	0	1	0	0	1	0	3	0	0	1	0
M7b*	1	4	7	17	5	0	0	0	1	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0
M7b1	4	5	10	14	7	4	1	0	3	0	0	0	0	0	0	1	3	3	0	0	1	0	0	1	0
M7b2	2	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M7b3	0	0	0	0	0	0	26	1	0	0	0	0	0	0	0	0	0	0	4	0	0	3	0	0	0
M7c*	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M7c1*	1	2	0	6	2	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
M7c1c	0	0	0	1	1	13	10	9	4	5	6	1	0	3	1	10	2	1	6	11	7	7	1	1	2
M10	1	4	7	3	0	0	0	1	0	0	0	5	0	7	0	0	0	0	0	0	0	0	0	0	0
M12	1	0	1	6	6	3	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0
M21	0	0	0	0	18	71	0	0	1	0	0	0	0	0	3	0	0	0	0	1	0	0	0	0	0
M22	0	0	0	0	2	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D*	4	29	12	6	11	0	1	1	0	0	0	0	0	0	1	0	0	0	0	3	2	0	1	2	0
D4	13	24	10	7	2	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
D5	7	23	9	14	5	1	8	0	0	0	0	0	0	0	0	2	1	1	0	11	0	7	2	1	0

Haplogroup	NW China[1]	NE China[2]	SW China[3]	SE China[4]	Thailand	Malay Peninsula[5]	Taiwan	Philippines	Medan	Pekanbaru	Palembang	Bangka	Padang	Tengger	Banjarmasin	Kota Kinabalu	Bali	Mataram	Waingapu	Manado	Palu	Toraja	Ujung Padang	Ambon	Alor
E*	1	0	0	2	0	1	16	1	0	0	0	1	0	0	2	2	0	0	1	5	1	2	3	3	0
E1a	0	0	0	0	0	1	13	5	2	1	0	2	1	0	3	11	2	1	1	19	4	13	6	2	3
E1b	0	0	0	0	0	4	0	0	1	0	1	1	1	1	3	0	1	3	6	3	2	2	2	0	2
G*	0	0	6	0	0	2	0	0	0	1	2	6	1	0	2	3	4	0	1	1	0	0	2	1	0
G2*	0	1	1	2	0	6	0	0	1	3	3	2	0	2	7	3	4	4	0	1	3	2	1	1	2
G2a	7	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Q	0	0	0	0	0	2	0	0	0	0	0	0	0	0	1	1	1	0	2	1	0	1	3	5	13
Others	71	73	65	31	46	6	5	4	2	4	1	0	1	1	2	2	0	1	0	1	0	0	1	0	0
Total	145	262	262	257	239	365	233	49	42	52	28	34	24	36	89	68	65	44	50	89	38	64	46	43	45

[1] Qinghai and Xinjiang
[2] Shanghai, Wuhan, Liaoning and Qingdao
[3] Yunnan
[4] Guangdong, Guangxi and Macau
[5] Orang Asli and Melayu