Am. J. Hum. Genet. 57:403–414, 1995

Polynesian Genetic Affinities with Southeast Asian Populations as Identified by mtDNA Analysis

Terry Melton,[1] Raymond Peterson,[1,2] Alan J. Redd,[1] N. Saha,[3] A. S. M. Sofro,[4] Jeremy Martinson,[5] and Mark Stoneking[1]

[1]Department of Anthropology, The Pennsylvania State University, University Park; [2]Laboratory of Neurogenetics, NIAAA, National Institutes of Health, Bethesda; [3]Department of Pediatrics, National University Hospital, Singapore; [4]IUC Biotechnology and Faculty of Medicine, Gadjah Mada University, Yogyakarta, Indonesia; [5]Institute of Molecular Medicine, University of Oxford, Oxford

Summary

Polynesian genetic affinities to populations of Asia were studied using mtDNA markers. A total of 1,037 individuals from 12 populations were screened for a 9-bp deletion in the intergenic region between the COII and $tRNA^{Lys}$ genes that approaches fixation in Polynesians. Sequence-specific oligonucleotide probes that identify specific mtDNA control region nucleotide substitutions were used to describe variation in individuals with the 9-bp deletion. The 9-bp deletion was not observed in northern Indians, Bangladeshis, or Pakistanis but was seen at low to moderate frequencies in the nine other Southeast Asian populations. Three substitutions in the control region at positions 16217, 16247, and 16261 have previously been observed at high frequency in Polynesian mtDNAs; this "Polynesian motif" was observed in 20% of east Indonesians with the 9-bp deletion but was observed in only one additional individual. mtDNA types related to the Polynesian motif are highest in frequency in the corridor from Taiwan south through the Philippines and east Indonesia, and the highest diversity for these types is in Taiwan. These results are consistent with linguistic evidence of a Taiwanese origin for the proto-Polynesian expansion, which spread throughout Oceania by way of Indonesia.

Introduction

The peopling of the South Pacific has been much examined in recent years, with linguistic, archaeological, and anthropological evidence increasingly being supplemented by genetic data. Two major prehistoric migratory sequences are thought to have occurred in this area. A first, ancient wave of hunter-gatherers moved from Indonesia into Australia and New Guinea 35,000–40,000 years before the present (B.P.) (Bellwood 1978; White and O'Connell 1982; Groube et al. 1983; Turner 1987; Wickler and Spriggs 1988). A more recent wave of Austronesian-language speakers began ~6,000–5,000 years B.P. in the settling of Taiwan from southeast China and ended as recently as A.D. 800 with the settlement of Easter Island and New Zealand by agriculturalists. This later expansion links proto-Polynesians with the Lapita pottery culture and Austronesian languages, which spread by migrations of Southeast Asians across vast stretches of open water (Bellwood 1978, 1989). Skilled navigation culminated in the eventual settlement of Polynesia, that easternmost group of islands which includes Samoa, Hawaii, Easter Island, the Marquesas, and the Society Islands (Spriggs 1984).

Terrell (1986) has suggested an alternative model in which Polynesians are more closely derived from Melanesian populations, that is, Papuan or Australoid groups derived from the earlier expansion ~40,000 years ago. In this interpretation of archaeological data, Polynesians who eventually settled the easternmost expanses of Oceania were the products of complex long-term associations between these biologically diverse Melanesians already in Oceania and later arrivals of Southeast Asians. This model implies that Polynesians have a sizeable genetic heritage from Melanesian populations, but data on genetic loci indicate that Polynesians are probably more closely related to Southeast Asians than to Melanesians (see Hill and Serjeantson 1989). There does appear to be at least a minor contribution (<10%) of Melanesian mtDNA to contemporary Polynesians (Lum et al. 1994), and while the data on nuclear loci can be interpreted to indicate that some admixture with Melanesians occurred during the migration of proto-Polynesians (O'Shaughnessy et al. 1990), it is unlikely that proto-Polynesians originated exclusively from Melanesian populations. Although the general genetic affinities of modern Polynesians thus seem to be established with little question, genetic evidence has not been used to more closely determine the source of the proto-Polynesian expansion. Since a language phylogeny traces Austronesian languages to Formosan (Taiwanese)

Received December 13, 1994; accepted for publication May 11, 1995.

Address for correspondence and reprints: Terry Melton, Department of Anthropology, The Pennsylvania State University, University Park, PA 16802.

0002-9297/95/5702-0024$02.00

sources (Blust 1988; Bellwood 1991), populations between Taiwan and Polynesia may hold genetic clues to the origins of Polynesian peoples. The purpose of this study is to survey Asian populations for mtDNA markers commonly found in Polynesians, in an attempt to identify the most likely source population for Polynesians.

An intergenic COII/tRNA^{Lys} mtDNA 9-bp deletion has been a useful marker in Asian populations (Horai and Matsunaga 1986; Wrischnik et al. 1987; Ballinger et al. 1992; Harihara et al. 1992; Passarino et al. 1993; Lum et al. 1994; Redd et al. 1995). It is found in moderate frequencies throughout Southeast Asia (Ballinger et al. 1992) and seems to be fixed or nearly fixed in Polynesia (Hertzberg et al. 1989; Hagelberg and Clegg 1993; Lum et al. 1994). This 9-bp deletion is largely absent in Melanesian populations—for example, aboriginal groups of Australia and highland Papua New Guinea (PNG)—while it is present in coastal populations of PNG that are thought to be more recent arrivals to the island (Hertzberg et al. 1989; Stoneking et al. 1990). While the frequency of this deletion has been reported for many populations throughout Asia, the frequency alone does not reveal either the source of the deletion or the origin of Polynesians.

However, patterns of variation in hypervariable segments of the noncoding mtDNA control region can provide insights into the evolutionary history of Polynesian mtDNAs (Hagelberg and Clegg 1993; Lum et al. 1994; Redd et al. 1995). In particular, Hagelberg and Clegg (1993) identified an apparently unique pattern of nucleotide substitutions in the control region of Polynesians. Three transitions, at nucleotide positions 16217, 16247, and 16261 (Anderson et al. 1981) have been seen together in modern Polynesians at high frequency (80%–100%) as well as in ancient DNA from sites in the Chatham Islands, Society Islands, Hawaii, and Easter Island dating from ≥400 years B.P. (Hagelberg et al. 1994; Lum et al. 1994). We have termed this trio of substitutions the "Polynesian motif." These three changes are temporally embedded within each other and the 9-bp deletion—that is, the nucleotide change at 16217 occurred on the background of the 9-bp deletion; the nucleotide change at 16261 occurred on the background of the nucleotide change at 16217; and the nucleotide change at 16247 occurred on the background of the change at 16261 (Redd et al. 1995).

We have capitalized on the presence of the easily screened-for 9-bp deletion and these three control region substitutions in mtDNA, to more closely examine the origins of Polynesians. We report here the frequency of the 9-bp deletion in 1,037 Asians from 12 populations. We also used sequence-specific oligonucleotide (SSO) probes to screen for the Polynesian motif as well as for substitutions at 12 additional sites known to be polymorphic in worldwide populations (Stoneking et al. 1991). We then compared the mtDNA SSO-type variation in these 12 populations to Samoans and coastal Papua New Guineans with SSO types inferred from control region sequences (Stoneking et al. 1992; Redd et al. 1995), to trace Polynesian origins.

Subjects and Methods

Subjects

In this study, 1,037 individuals were screened for the 9-bp deletion and typed for sequence variants at 15 nucleotide positions in nine regions across the mtDNA control region. All samples were purified genomic DNA obtained from maternally unrelated individuals. The 12 populations studied are from the following areas:

Borneo.—95 samples from the Barito River area (southeast Borneo).

Bangladesh.—31 samples collected from immigrants (male workers) to Singapore.

China.—103 samples from southern (Han) Chinese immigrants to Singapore.

Philippines.—60 samples from immigrants (female workers) to Singapore, 97 samples from the northern island of Ilocano (provided by Roche Laboratories), and 19 samples from Filipino U.S. military personnel.

Southern India.—77 samples from immigrants to Singapore from southern India and Sri Lanka, primarily of Dravidian origin.

Northern India.—47 samples from Sikhs who immigrated to Singapore.

East Indonesia.—94 samples from the Nusa Tenggaras (Alor, 21; Flores, 22; Roti, 26; Timor, 25), and 49 samples from the Moluccas (Hiri, 26; Ternate, 23).

Java.—98 samples from rural areas of central Java, near Semarang.

Malaysia.—81 samples from native Malays, and 30 samples from the Orang Asli (Semai Senoi), an aboriginal group from the peninsular center.

Pakistan.—76 samples from Pushtoons of Peshwar in northwest Pakistan.

Taiwan.—82 samples from four aboriginal groups (Ami, 22; Atayal, 20; Bunum, 19; Paiwan, 21).

In addition, for some analyses, SSO types were included that were inferred from mtDNA control region sequences of 24 Samoan and 23 coastal PNG mtDNAs, all with the 9-bp deletion (Stoneking et al. 1992; Redd et al. 1995).

9-bp Deletion Screening

Each sample was screened for the mtDNA 9-bp deletion after PCR amplification of the intergenic region between the COII and lysine tRNA genes. Amplifications were performed in a 25-μl volume containing 5

pmol of primers A and B from Wrischnik et al. (1987), a 200 μM concentration of each deoxynucleotide triphosphate (dNTP), 25 μg of bovine serum albumin (BSA), ~0.1 μg genomic DNA, 1 mM of $MgCl_2$, and 0.5 U of *Taq* DNA polymerase in reaction buffer supplied by the manufacturer (Promega). Thirty cycles of denaturation at 94°C for 1 min, annealing at 56°C for 1 min, and extension at 74°C for 1 min were carried out. Negative controls were run simultaneously to detect reagent contamination. Ten microliters of PCR product were electrophoresed through a 2% Metaphor agarose (USB) gel for 1 h to distinguish the 112-bp fragment expected from mtDNAs with the deletion from the 121-bp fragment expected from mtDNAs without the deletion.

SSO Typing

A 1024-bp fragment of the mtDNA control region was amplified from each sample in a 100-μl volume. Reaction conditions including molar concentrations of $MgCl_2$ and reaction buffer were the same as above, except that 20 pmol of primers L15996 and H408 of Vigilant et al. (1989) were used with 2 U of *Taq* DNA polymerase, 100 μg of BSA, and 0.1 μg of genomic DNA.

Five microliters of the mtDNA control region PCR product were denatured in 50 μl of 0.4 M NaOH, 25 mM EDTA and dotted on a nylon membrane (Biodyne) in a 96-well format. Membranes were prepared in quadruplicate, and each membrane included a panel of control samples, encompassing all known SSO sequence variants, to optimize hybridization and washing conditions. Membranes were prewarmed in 5× SSPE (20× SSPE = 3 M NaCl; 0.2 M $NaH_2PO_4 \cdot H_2O$; 0.02 M EDTA, pH 7.4), 0.5% SDS for 10 min at the optimal probe-annealing temperature, then were hybridized for 20 min after adding biotinylated probe to a final concentration of 1 pmol/ml in 5× SSPE, 0.5% SDS. The membranes were washed in 2× SSPE, 0.1% SDS for ≤40 min at the optimal probe-annealing temperature. Membranes were placed in 30 ml of 2× SSPE, 0.5% SDS containing 90 μl of a streptavidin-horseradish peroxidase (SA-HRP) conjugate (Roche) and washed for 10 min at 42°C, then washed for 15 min in 2× SSPE, 0.1% SDS. Chemiluminescent detection of bound SA-HRP was carried out with ECL detection reagents (Amersham), according to the manufacturer's directions. The membranes were then wrapped in plastic and exposed to x-ray film for 30 s to 2 h.

Twenty-one of the probes were described in detail by Stoneking et al. (1991). Modified versions, of four of them were used; these have undergone addition or removal of nucleotides from the 5′ or 3′ ends to increase specificity. Probes IIE1 and IIE2 of Stoneking et al. (1991) were not used in this study, because of problems with cross-hybridization, presumably because these probes detect variation in the length of a run of consecutive cytosine residues, and not single nucleotide substitutions. Probe IB3 identified the first substitution in the Polynesian motif at nucleotide position 16217. A new set of probes (IE1, IE2, and IE3) was designed for this study from Samoan and PNG sequence data (Stoneking et al. 1992; Redd et al. 1995) to identify the remaining two sites in the Polynesian motif, at positions 16247 and 16261. Table 1 gives the sequences and related information for the four probes modified from Stoneking et al. (1991) plus the three new probes for identification of the Polynesian motif.

Statistical Analysis

Individual results of the SSO typing were arranged into a mtDNA SSO type based on the probe variants across nine mtDNA control region sites. For example, the mtDNA type 1-1-2-1-2-1-1-1-0 indicates that probe variant IA1 annealed in the IA region, that IB1 annealed in the IB region, etc. IE probe variants are reported in the third position, since sites 16247 and 16261 are between IB and IC in the control region. A zero indicates that a blank result for the above example was obtained for the IID variant. A blank result for IID occurs either when a substitution in a nearby site prevents probe annealing or when a nucleotide other than A or G is present at position 247 (the IID probe-specific site). While blanks in different individuals for a probe region could reflect different substitutions, for the purpose of analysis, blanks are considered to be the same variant. Reliability of SSO typing was confirmed by comparing SSO types from 37 Indonesian individuals with their control region sequences. The overall error rate for individual probe results was found to be <1%.

An unbiased estimate of diversity (h) was calculated to quantify the amount of mtDNA variation present in each population:

$$h = \frac{(1 - \Sigma x^2)n}{n - 1}, \quad (1)$$

where n is the sample size and x is the frequency of each mtDNA type (Tajima 1989). The variance for this estimate of diversity was calculated as

$$V(h) = \frac{2[\Sigma x^3 - (\Sigma x^2)^2]}{n} \quad (2)$$

(Nei 1987). Because they were chosen to highlight diversity, the SSO probes used for this study could be said to bias diversity estimates; however, these estimates are valid for comparisons between the populations included in this study since the same set of probes was used for all comparisons.

Table 1

SSO Probes Differing From or Supplementing Those of Stoneking et al. (1991)

Region[a] and Variant	Polymorphic Nucleotides[b]	Hybridization Temperature (°C)	Probe Sequence (5′ to 3′)	Strand[c]
16116–16138:				
IA3[d]	16126 T, 16129 A	50	ATGAATATTGTACAGTACCATAA	L
16244–16264:				
IE1	16247 A, 16261 C	55	GCAACTCCAAAGCCACCCCTC	L
IE2	16247 A, 16261 T	65	CGTTGAGGTTTCGGTGGAGAG	H
IE3	16247 G, 16261 T	55	CGTCGAGGTTTCGGTGGAGAG	H
16298–16319:				
IC2[d]	16304 C, 16311 T	50	TAACAGCACATAGTACATAAAG	L
IC3[d]	16304 T, 16311 C	50	TAACAGTACATAGCACATAAAG	L
68–81:				
IIA2[d]	73 G	50	ATCGCGTGCACACC	H

[a] Nucleotides encompassed by the SSO probes, numbered according to Anderson et al. (1981).
[b] Nucleotide position and state on the L strand detected by the SSO probe.
[c] H-probe sequence corresponding to heavy strand of Anderson et al. (1981) published sequence; L-probe sequence corresponding to light strand.
[d] Modified from Stoneking et al. 1991.

The probability of identity between populations i and j was calculated as

$$p = \sum_{i,j}^{n} x_i x_j, \quad (3)$$

where x_i and x_j are the frequencies of a mtDNA type in populations i and j, summed over the n mtDNA types in the two populations.

A phylogenetic analysis was done to estimate relatedness between populations using the neighbor-joining (NJ) method (Saitou and Nei 1987) with genetic distance $d = 1 - p$. Bootstrap tests and confidence probability tests were judged to be inappropriate for this data set because of the small number of nucleotide sites, so a four-cluster analysis (Rzhetsky et al. 1995) was applied to test stability of the branching pattern of the NJ tree.

Results

9-bp Deletion

Table 2 and figure 1 show the frequency of the 9-bp deletion in the populations included in this study. The 9-bp deletion was not observed in the Pakistani, northern Indian (Sikh), or Bangladeshi populations. It was observed in 6 of 77 southern Indians; the remaining eight groups, all Southeast Asian, had frequencies ranging from 21% to 41%. The frequency of the deletion is also reported for all of the population subgroups in table 2 and is highest in the Paiwan of Taiwan, an aboriginal group. The Taiwanese and Filipino groups have the 9-bp deletion at frequencies of ~40%, whereas those groups to the south display slightly lower frequencies. Overall, of the 1,037 individuals included in this study, 243 carry the 9-bp deletion.

mtDNA Types

In 243 individuals with the 9-bp deletion we observed 70 different mtDNA SSO types (see the appendix, table A1). One type, 1-3-1-1-1-2-1-1-1, was observed 48 times, while three other types were observed 26, 25, and 11 times each. Fifteen types occurred between three and nine times apiece, and the remaining 51 types occurred one or two times each. The type that occurs 25 times is closely related to the Polynesian motif and has the site changes at positions 16217 and 16261, but not the change at 16247.

The total number of mtDNA types, estimated SSO-type diversity, and associated standard error (SE) for each population are also shown in table 2. The highest diversity was observed in the southern Indian group, where the six individuals with the deletion each had a different mtDNA type. Diversity values for most of the other Southeast Asian groups were also quite high, ranging between .883 and .957. The 23 coastal Papua New Guineans and 24 Samoans with the 9-bp deletion (Redd et al. 1995) have inferred SSO-type diversity estimates of .589 and .431 respectively, showing the reduced mtDNA diversity in Polynesian populations. The Orang Asli, an isolated peninsular aboriginal Malay group, is the only group in this study with lower diversity. Population subgroups in general show reduced diversity; for example, the Ami, Atayal, Bunum, and Paiwan of Taiwan have diversity estimates between .778 and .844.

Table 2

Frequency of the 9-bp Deletion and Estimates of mtDNA-Type Diversity in Asian Populations with the Deletion

Population	N	Frequency of Deletion (%)	No. with Deletion	No. of Types with Deletion	$h \pm$ SE
Bangladeshi	31	.0	...	...	...
Bornean	95	24.2	23	15	.941 ± .020
Southern Chinese	103	22.3	23	15	.941 ± .020
Miscellaneous	79	44.3	35	23	.950 ± .159
Ilocano	97	36.1	35	10	.792 ± .356
Total Filipino	176	39.8	70	27	.878 ± .020
South Indian	75	8.0	6	6	1.000 ± .000
North Indian (Sikh)	47	.0	...	...	...
Total Indian	122	4.9	6	6	1.000 ± .000
Nusa Tenggaras	94	23.4	22	13	.931 ± .021
Moluccas	49	16.3	8	6	.929 ± .031
Total East Indonesian	143	21.0	30	15	.936 ± .014
Javanese	98	25.5	25	14	.883 ± .035
Malay	81	25.9	21	15	.957 ± .014
Orang Asli	30	36.7	11	3	.345 ± .120
Pakistani	76	.0	...	...	...
Ami	22	45.5	10	4	.778 ± .045
Atayal	20	25.0	5	3	.800 ± .051
Bunum	19	47.4	9	5	.806 ± .070
Paiwan	21	47.6	10	6	.844 ± .061
Total Taiwanese	82	41.5	34	16	.925 ± .015

Table 3 shows the numbers of shared mtDNA types with the 9-bp deletion and the average probability of identity between each pair of populations. Of the 11 Orang Asli individuals with the 9-bp deletion, nine have the same type, leading to a relatively high probability of identity with other groups. The probability of identity with Samoans and coastal Papua New Guineans is highest for east Indonesians, with values of 10.0% and 9.7% respectively, and drops to <5% for the other groups.

To further investigate population relationships based on SSO types associated with the 9-bp deletion, we constructed an NJ tree from genetic distances based on the probability of identity (fig. 2). This tree indicates that Samoan and coastal PNG SSO types cluster most closely with east Indonesian types, followed by Taiwanese, Filipino, Orang Asli and Malay types. Javanese types and Bornean types cluster, while Chinese and southern Indians fall outside other groupings. Four-cluster analysis (Rzhetsky et al. 1995) was used to test the stability of each internal branch in the NJ tree against alternative trees. This analysis indicates that the NJ tree in figure 2 is preferred over all alternative trees, with the exception that a tree grouping the Malay population with the Philippine and Orang Asli populations is equally consistent with the data.

Polynesian Motif

The sample of individuals with the 9-bp deletion was examined in detail with respect to mtDNA types. Figure 3 reports population frequencies of the Polynesian motif and related mtDNA types, that is, types where there are changes at nt 16217 (T→C, detected by the IB3 probe variant); nt 16261 (C→T, detected by probe variants IE2 and IE3); and nt 16247 (A→G, detected by probe variant IE3). The Polynesian motif (CGT) was observed in 20% of east Indonesians with the 9-bp deletion, but elsewhere in only one Malay, and probably one Filipino; this last individual carries the site changes at 16261 and 16247 but has a blank result for probe IB. The substitution at 16217 therefore probably exists in this individual, but an additional site change in the IB probe region resulted in the blank typing. The Polynesian motif was never observed in the 794 individuals without the 9-bp deletion (data not shown).

The intermediate state (CAT) was seen at highest frequency in Taiwan and was also present at moderate frequencies in east Indonesia and the Philippines. The ancestral pattern CAC is widely distributed throughout all populations at moderate to high frequencies, but is highest in the Orang Asli, all of whom carry this pattern in conjunction with the 9-bp deletion.

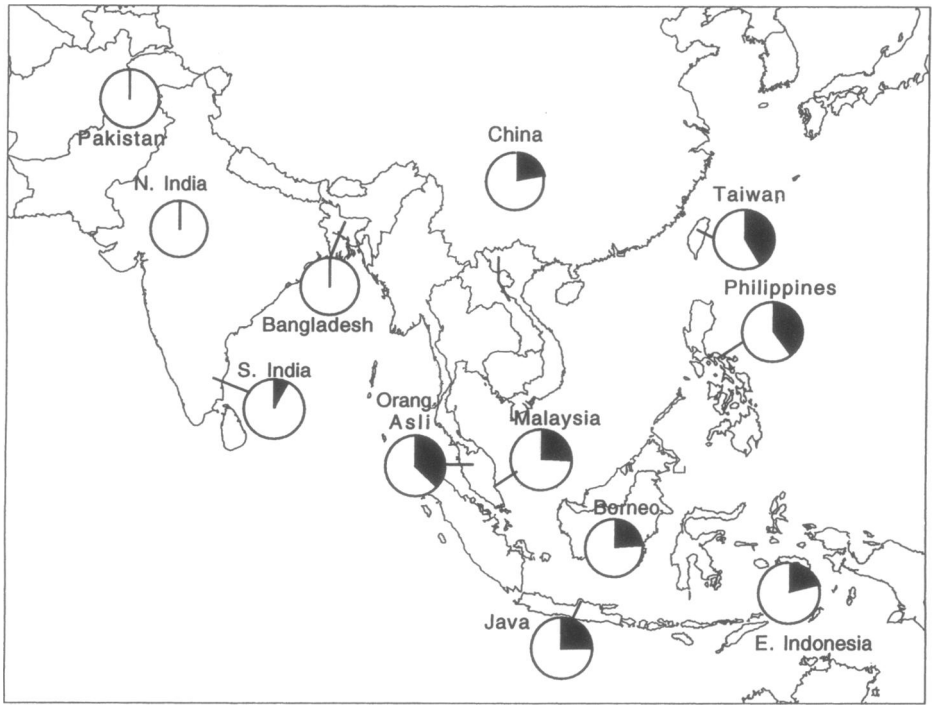

Figure 1 Locations of the 12 populations included in this study. The frequency of the 9-bp deletion is indicated by the black portion of each circle.

We also examined the distribution of specific mtDNA types that either carried the Polynesian motif or were closely related. Table 4 shows the frequency in the Southeast Asian populations of mtDNA types with the 9-bp deletion that occur more than once and are most closely related to Samoan and coastal PNG mtDNA types. All types shown have (1) the 9-bp deletion, (2) the site change at 16217, and (3) either the site change at 16261 or changes at both 16261 and 16247. These related types occur most often in Taiwan, east Indonesia, and the Philippines and are rare or absent in Borneo, China, Java, and Malaysia.

Discussion

The near fixation of the 9-bp deletion in Polynesia (Hertzberg et al. 1989; Hagelberg and Clegg 1993; Ha-

Table 3

Numbers of Shared mtDNA Types (above the Diagonal) and Probability of Identity[a] (below the Diagonal) between Populations for Individuals with the 9-bp Deletion

Population	Bornean	Chinese	Filipino	South Indian	East Indonesian	Java	Malay	Orang Asli	Taiwan	Samoan	PNG
Bornean		3	5	3	4	8	4	1	5	1	1
Chinese	.061		5	2	3	4	5	1	5	0	0
Filipino	.042	.035		3	5	6	7	1	7	1	1
South Indian	.029	.036	.007		0	3	3	0	2	0	0
East Indonesian	.039	.016	.082	.000		3	5	1	5	4	2
Javanese	.104	.090	.064	.020	.051		6	1	5	0	0
Malay	.050	.056	.062	.024	.013	.080		1	6	2	2
Orang Asli	.071	.071	.245	.000	.136	.131	.117		1	0	0
Taiwanese	.023	.015	.061	.024	.052	.024	.032	.048		1	1
Samoan	.005	.000	.020	.000	.100	.000	.042	.000	.026		2
PNG	.009	.000	.034	.000	.097	.000	.039	.000	.045	.484	

[a] Calculated according to equation (3).

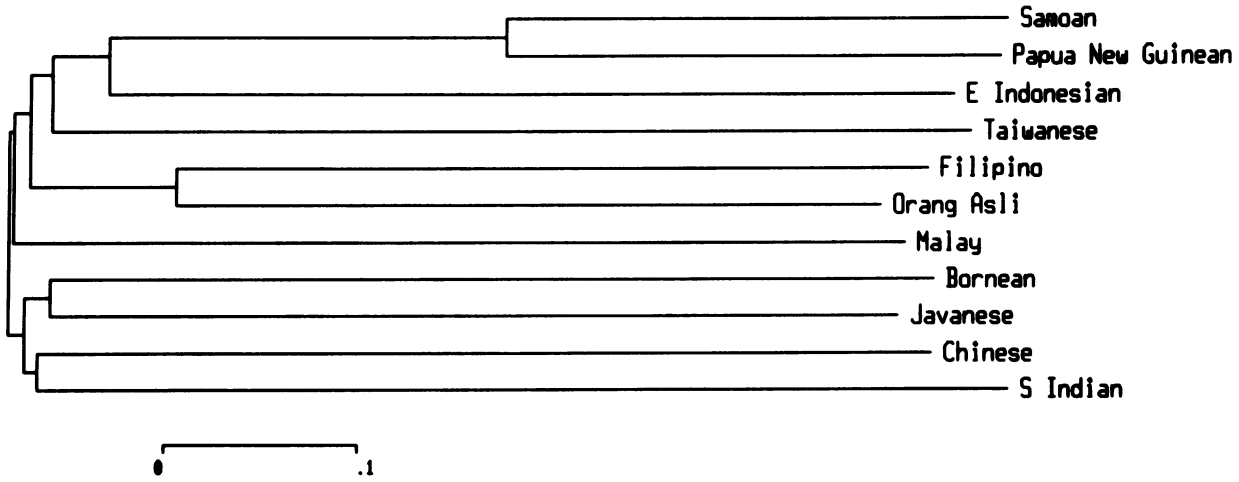

Figure 2 Unrooted NJ tree for populations, based on mtDNA SSO types with the 9-bp deletion. The genetic distance between each pair of populations was computed from the probability of identity (table 4) as described in Subjects and Methods.

gelberg et al. 1994; Lum et al. 1994) indicates that there have been strong population bottlenecks and/or founder events in that region, with a concomitant reduction in genetic diversity. In accordance with previous studies, we found that the 9-bp deletion is ubiquitous throughout Southeast Asia, so it alone is not informative for tracing Polynesian history. Ballinger et al. (1992) found that 40% of Taiwanese Han carry the 9-bp deletion marker, a figure in agreement with our results. Their frequency of 19% of the marker in Sabah (northwest Borneo) aborigines approximates our value of 24% found in southeast Borneo, and their Malay estimate of 14% is somewhat lower than our value of 26%. Overall, their frequencies of the deletion in coastal Southeast Asian populations were lower than those seen in the groups sampled here, which is further argument for founder events having established higher frequencies in island populations. Harihara et al. (1992) found that Negritos of Luzon in the Philippines have a frequency of 92%, much higher than that of the Philippine populations reported here, none of which are Negrito.

Outside of Southeast Asia, the deletion was not observed in Bangladeshi, northern Indian, or Pakistani populations. Passarino et al. (1993) did not observe the deletion in Hindus from Chitwan and New Delhi, but noted that 8% of Tharus from southern Nepal carry it. The absence of the deletion in this study in north India and Pakistan is consistent with the genetic distinctiveness of northwest and Southeast Asia (Nei and Roychoudhury 1993). The absence of the deletion in Bangladesh is somewhat surprising but may be due to the small sample size.

It is interesting that the six southern Indians with the 9-bp deletion found in this study share their mtDNA types most closely with those from China and Borneo, suggesting that migration from these regions west to India and Sri Lanka may be a possibility. An alternative explanation is that there may have been an independent origin of the deletion in India. Redd et al. (in press) noted that there was an independent origin of the deletion in Africa, and multiple origins of the deletion in Asia have been suggested (Ballinger et al. 1992; Redd et al. 1995). Control region sequence analysis would help to resolve this question of multiple origins.

Estimates of mtDNA type diversity associated with the 9-bp deletion ought to be informative for tracing ancestry, as in general the ancestral (source) population is expected to have the greatest diversity (Stoneking 1993). Although diversity is highest in the south Indian group and quite low in Samoa and coastal PNG, diversity estimates for most of the geographically intermediate populations studied here are very similar. Redd et al. (1995) have shown that the Asian 9-bp deletion probably arose $\geq$58,000 years B.P.; it is likely that the genetic record of the spread of this rather ancient marker has been erased by the many migrations throughout Southeast Asia.

We observed that the Polynesian motif, this trio of nucleotide changes in the control region at positions 16217, 16247, and 16261 (CGT), occurred exclusively on the background of the 9-bp deletion. This motif, seen in 79.2% of Samoans and 73.9% of coastal Papua New Guineans, was observed in 20% of east Indonesians with the 9-bp deletion. These east Indonesians were from the islands of Alor, Flores, Hiri, Ternate, and Timor. Remarkably, it was not observed elsewhere in Southeast Asia (including Borneo and Java in Indonesia), except in 1 of 81 Malays, and probably 1 of 176 Filipinos.

The above results can be interpreted to indicate that the Polynesian motif arose in east Indonesia, although an origin in Malaysia or the Philippines cannot be ruled

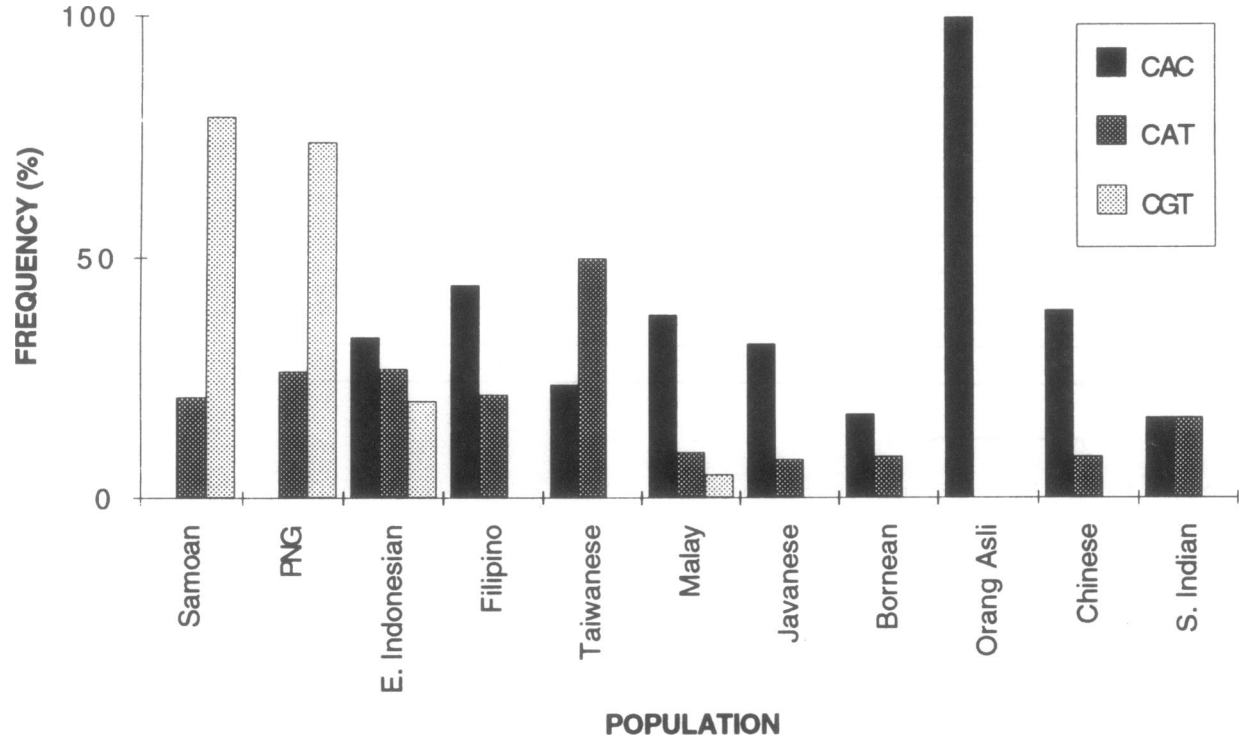

Figure 3 Frequencies of three patterns of nucleotide substitution occurring with the 9-bp deletion in Pacific and Southeast Asian populations. CGT is the Polynesian motif. Frequencies are the percent of individuals with the 9-bp deletion who carry a particular nucleotide substitution pattern—that is, CAC, CAT, or CGT (the Polynesian motif)—with the order of mutations being CAC→CAT→CGT.

out. Alternatively, the presence of this combination of substitutions in east Indonesia might reflect back migration that occurred from Polynesia and/or coastal New Guinea to this region. The amount of diversity associated with the Polynesian motif might distinguish between these competing explanations. For the six Indonesian individuals from the Moluccas and the Nusa Tenggaras who carry the Polynesian motif, mtDNA type diversity is .733. In coastal PNG, the sample of 17 individuals with the motif has an mtDNA-type diversity of .309, while the 19 Samoans with the motif have an mtDNA-type diversity of .105. Although we have used only 24 Samoans as a reference Polynesian sample, their mtDNA variation is representative of that of other Polynesian populations (Lum et al. 1994). For individuals with the Polynesian motif, sample sizes are small, but it appears from the diversity values that this combination of substitutions did originate in east Indonesia; mtDNA sequence analysis further supports this conclusion (Redd et al. 1995).

Table 4

Frequencies of the Polynesian Motif and Related mtDNA Types

mtDNA Type[a]	Samoan	PNG	East Indonesian	Taiwanese	Malay	Filipino	Bornean	Chinese	Javanese
132312201	.0	.0	6.3	.0	.0	.0	.0	.0	.0
132312221	.0	.0	.0	.0	.0	.0	4.3	.0	4.0
333112211[b]	.0	.0	6.3	.0	.0	.0	.0	.0	.0
132112201	.0	.0	.0	2.9	.0	4.3	.0	.0	.0
133112211[b]	75.0	60.9	10.0	.0	4.8	.0	.0	.0	.0
132312211	8.0	.0	3.3	11.8	.0	.0	.0	.0	.0
132112111	.0	.0	.0	11.8	4.8	1.4	.0	8.7	.0
132112211	12.5	21.7	16.7	20.6	4.8	15.7	4.3	.0	.0

[a] Types occurring more than once, arranged from least to most common.
[b] Types with the Polynesian motif.

While the CGT (Polynesian) motif is almost completely restricted to Polynesia, coastal PNG, and east Indonesia, the intermediate motif, CAT, is useful for tracing population origins as well. Because of the temporal embeddedness of these three sites, there is a trail that can be uncovered by examining both the frequency of this pattern in Southeast Asia and mitochondrial types related to it. We observed that the CAT pattern, whose origin predates that of the CGT pattern, occurs in high frequency in three aboriginal groups of Taiwan, and overall the frequency of this pattern is highest in Taiwan. Furthermore, the mtDNA diversity in Taiwan for individuals with this pattern is .757. The frequency of the CAT pattern is lower in all other groups; Filipinos and east Indonesians with the deletion have it at frequencies of 21.4% and 26.7%, with diversity estimates of .448 and .607, respectively; in the other groups it is rare or absent. These results suggest that the intermediate CAT motif arose in Taiwan and spread south through the Philippines and Indonesia.

Tracing back further, the pattern CAC, whose origin predates that of the CAT motif but still exists on the background of the 9-bp deletion, was observed at frequencies ranging from 0% in Polynesia and the Ami of Taiwan to 100% in the Orang Asli. This pattern, like the 9-bp deletion, is seen in substantial frequencies throughout Southeast Asia and is likely to be ancient compared with the patterns CAT and CGT and to have spread extensively during early population expansions.

The following scenario for the history of the 9-bp deletion in Southeast Asia, and for Polynesian origins, is compatible with these data. The 9-bp deletion occurred ~60,000 years B.P. (Redd et al. 1995), followed soon after by the nucleotide substitution at 16217, and mtDNAs carrying these markers spread extensively throughout Southeast Asia. The source for the 9-bp deletion is unknown but is likely to be in a region extending from China to Southeast Asia (Ballinger et al. 1992). The change at nucleotide position 16261 occurred next, probably in Taiwan, and spread outward beginning ~6,000 years B.P. throughout the Philippines and Indonesia. Finally, the mutation at position 16247 occurred, probably in east Indonesia, and rose to high frequency in coastal PNG. This would represent a significant founder event, especially since the coastal PNG groups with the Polynesian motif are from diverse locations (Stoneking et al. 1990). This large island was most likely a major staging ground for the ensuing settlement of Polynesia. Only coastal populations with well-developed navigational skills moved eastward, accounting for the rarity of highland PNG "Melanesian" mtDNA types in Polynesia (Lum et al. 1994).

The phylogenetic analysis shown by the NJ tree in figure 2 supports this scenario: Samoans and Papua New Guineans cluster most closely with east Indonesians, followed by Taiwanese and Filipinos. However, all of the internal branches of the tree are relatively short (with the exception of the branch linking Samoans with Papua New Guineans), indicating a fairly rapid spread of the 9-bp deletion throughout Southeast Asia.

This scenario of Polynesian origins is consistent with the overall genetic and linguistic evidence accumulated to date. Nuclear markers have also linked Polynesians and Southeast Asians: an unusual triple ζ-globin gene rearrangement has been observed in both Southeast Asia and Polynesia, and α-globin haplotypes Ia, IIa, IIC, and IIe and HLA-DR5 associated DR.DQ haplotypes are also shared (Winichagoon et al. 1982; Trent et al. 1986; Hertzberg et al. 1988; Serjeantson 1989; O'Shaughnessy et al. 1990). Linguistic evidence indicates that the most ancient split in the Austronesian language family occurred between the Malayo-Polynesian and Formosan subgroups in Taiwan and the Philippines (Blust 1988; Bellwood 1991). The Malayo-Polynesian subgroup then differentiated into western and central-eastern clusters in the Moluccas and Lesser Sunda (Nusa Tenggara) islands, and the central-eastern language groups went on to spread throughout Oceania (Blust 1988; Bellwood 1991).

The prehistory of languages in Indonesia fits quite well with the scenario of Polynesian origins suggested by this mtDNA analysis. In particular, we note the correspondence between (1) the linguistic evidence placing the source of the Austronesian language expansion in Taiwan and the mtDNA evidence suggesting that the intermediate CAT motif arose in Taiwan; and (2) the linguistic evidence placing the primary source of Polynesian languages in eastern Indonesia and the mtDNA evidence suggesting that the Polynesian motif arose in eastern Indonesia. One caveat is that the motifs we have described here may in fact have been more widespread in the past but that recent migrations have erased their record in contemporary populations. Further investigations of ancient remains, such as those carried out by Hagelberg and Clegg (1993), would be useful to explore this possibility.

In this study, we have demonstrated the utility of SSO probes for rapid screening of well-defined, informative mtDNA nucleotide substitutions. This method, which obviates the need for sequencing entire hypervariable segments of the mtDNA control region in large numbers of individuals, can provide valuable information about population origins and history.

Acknowledgments

We thank Sean Walsh and Roche Molecular Systems for providing some probes; Himla Soodyall and Steve Sherry for assistance with sequence data and probe design; Anne Bu-

chanan and Robert Melton for help with data management; and Sudhir Kumar for help with statistical analysis. The collection of east Indonesian samples was supported by a grant from the Wenner-Gren Foundation to M.S. This project was supported by NSF grant BNS 90-20567 and NIJ grant 92-IJ-CX-K040 to M.S.

Appendix

Table A1

Count of mtDNA SSO Types Observed in 243 Individuals with the 9-bp Deletion from Southeast Asia

mtDNA Type	1	2	3	4	5	6	7	8	9	Total
031112111	0	0	0	1	0	0	0	0	0	1
101112010	0	0	1	0	0	0	0	0	0	1
101112021	0	0	0	0	0	1	0	1	0	2
101112101	0	1	0	0	0	0	0	0	0	1
101112111	0	1	0	1	0	0	0	0	0	2
102112111	0	0	0	1	0	0	0	1	0	2
102112211	0	0	0	1	0	0	0	0	0	1
102212211	0	1	0	0	0	0	0	0	0	1
103112211	0	0	0	1	0	0	0	0	0	1
111111011	0	0	0	1	0	0	0	0	0	1
111112011	1	0	0	0	1	0	0	1	0	3
111112021	1	0	0	0	0	0	0	0	0	1
111112101	0	0	0	0	0	0	0	0	1	1
111112111	5	5	2	2	8	3	0	0	1	26
111112121	1	0	0	0	0	0	0	0	0	1
111112201	0	0	0	0	1	0	0	0	0	1
111112211	2	0	0	1	1	1	0	1	0	6
111112301	0	0	1	0	0	0	0	0	0	1
111112311	0	0	0	7	1	1	0	0	2	11
111122111	0	0	0	0	0	0	0	0	1	1
111122211	0	0	0	0	0	1	0	0	0	1
111212011	1	0	0	0	0	0	0	0	0	1
111312211	1	0	0	0	0	0	0	0	0	1
112112111	0	0	0	0	1	0	0	0	0	1
112112311	3	0	0	0	2	0	0	0	0	5
112312311	1	0	0	0	0	0	0	0	0	1
121112100	0	0	0	1	0	0	0	0	0	1
121112111	0	0	0	0	0	0	0	0	1	1
121112201	0	0	0	1	0	0	0	0	0	1
121122001	0	0	0	1	0	0	0	0	0	1
121122011	1	0	0	0	0	0	0	0	0	1
121122111	0	1	0	0	0	1	0	0	1	3
122112211	0	0	1	2	0	1	0	0	0	4
130112211	0	0	1	0	0	0	0	0	0	1
130112311	0	0	0	1	0	0	0	0	0	1
130122011	0	0	0	1	0	0	0	0	0	1
131012111	0	1	0	1	0	0	0	0	0	2
131012121	0	1	0	0	0	0	0	0	0	1
131102111	0	0	0	0	0	0	0	0	4	4
131112001	0	0	1	0	0	0	0	0	0	1
131112011	1	3	0	0	1	0	0	1	1	7
131112021	1	0	2	3	1	0	0	0	1	8
131112101	0	0	0	0	0	0	1	0	0	1
131112110	0	0	0	1	0	0	0	0	0	1
131112111	2	2	5	21	4	3	9	0	2	48
131112211	0	0	0	2	1	1	0	0	0	4
131112221	0	0	0	1	0	0	0	0	0	1
131112311	0	1	2	0	0	0	0	0	0	3
131212111	0	0	0	0	0	0	1	0	0	1
131312211	0	0	0	1	0	0	0	0	0	1
132112111	0	2	0	1	0	1	0	1	4	9

(*continued*)

Table A1 (continued)

mtDNA Type	1	2	3	4	5	6	7	8	9	Total
132112201	0	0	0	3	0	0	0	0	1	4
132112211	1	0	5	11	0	1	0	0	7	25
132112221	0	0	0	0	1	0	0	0	0	1
132212211	0	0	0	0	0	0	0	0	1	1
132312201	0	0	2	0	0	0	0	0	0	2
132312211	0	0	1	0	0	0	0	0	4	5
132312221	1	0	0	0	1	0	0	0	0	2
133112011	0	0	1	0	0	0	0	0	0	1
133112211	0	0	3	0	0	1	0	0	0	4
221112111	0	0	0	0	0	1	0	0	0	1
311112001	0	1	0	0	0	0	0	0	0	1
311112111	0	0	0	1	0	0	0	0	0	1
311122111	0	0	0	1	0	0	0	0	0	1
311212010	0	1	0	0	0	0	0	0	0	1
321122311	0	1	0	0	0	0	0	0	0	1
321122331	0	0	0	0	0	0	0	0	2	2
331112021	0	1	0	0	1	3	0	0	0	5
331212021	0	0	0	0	0	1	0	0	0	1
333112211	0	0	2	0	0	0	0	0	0	2
All	23	23	30	70	25	21	11	6	34	243

NOTE.—1 = Borneo; 2 = China; 3 = east Indonesia; 4 = Philippines; 5 = Java; 6 = Malaysia; 7 = Orang Asli; 8 = south India; and 9 = Taiwan.

References

Anderson S, Bankier AT, Barrell BG, DeBruijn MHL, Coulson AR, Drouin J, Eperon IC, et al (1981) Sequence and organization of the human mitochondrial genome. Nature 290:457–465

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

Bellwood P (1978) Man's conquest of the Pacific: the prehistory of southeast Asia and Oceania. Oxford University Press, New York

——— (1989) The colonization of the Pacific: some current hypotheses. In: Hill AVS, Serjeantson SW (eds) The colonization of the Pacific: a genetic trail. Clarendon, Oxford, pp 1–59

——— (1991) The Austronesian dispersal and the origin of languages. Sci Am 265:88–93

Blust R (1988) The Austronesian homeland: a linguistic perspective. Asian Perspect 26:45–67

Groube L, Chappell J, Muke J, Price D (1983) A 40,000-year-old human occupation site at Huon Peninsula, Papua New Guinea. Nature 324:453–455

Hagelberg E, Clegg JB (1993) Genetic polymorphism in prehistoric Pacific islanders determined by analysis of ancient bone DNA. Proc R Soc Lond B 252:163–170

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

Harihara S, Hirai M, Suutou Y, Shimizu K, Omoto K (1992) Frequency of a 9-bp deletion in the mitochondrial DNA among Asian populations. Hum Biol 64:161–166

Hertzberg MS, Mickleson KNP, Serjeantson SW, Prior JF, Trent RJ (1989) An Asian-specific 9-bp deletion of mitochondrial DNA is frequently found in Polynesians. Am J Hum Genet 44:504–510

Hertzberg MS, Mickleson KNP, Trent RJ (1988) α-Globin gene haplotypes in Polynesians: their relationships to population groups and gene rearrangements. Am J Hum Genet 43:971–977

Hill AVS, Serjeantson SW (eds) (1989) The colonization of the Pacific: a genetic trail. Clarendon, Oxford

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

Lum JK, Rickards O, Ching C, Cann RL (1994) Polynesian mitochondrial DNAs reveal three deep maternal lineage clusters. Hum Biol 66:567–590

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Nei M, Roychoudhury AK (1993) Evolutionary relationships of human populations on a global scale. Mol Bio Evol 10:927–943

O'Shaughnessy DF, Hill AVS, Bowden DK, Weatherall DJ, Clegg JB (1990) Globin genes in Micronesia: origins and affinities of Pacific island peoples. Am J Hum Genet 46:144–155

Passarino G, Semino O, Modiano G, Santachiara-Benerecetti AS (1993) COII-tRNA^{Lys} intergenic 9-bp deletion and other mtDNA markers clearly reveal that the Tharus (southern Nepal) have Oriental affinities. Am J Hum Genet 53:609–618

Redd A, Takezaki N, Sherry S, McGarvey S, Sofro ASM, Stoneking M (1995) Evolutionary history of the COII/

tRNA^{Lys} intergenic 9-bp deletion in human mitochondrial DNAs from the Pacific. Mol Biol Evol 12:604–615

Rzhetsky A, Kumar S, Nei M (1995) Four-cluster analysis: a simple method to test phylogenetic hypotheses. Mol Biol Evol 12:163–167

Saitou N, Nei M (1987) The neighbor-joining method: a new method for constructing phylogenetic trees. Mol Biol Evol 4:406–425

Serjeantson SW (1989) HLA genes and antigens. In: Hill AVS, Serjeantson SW (eds) The colonization of the Pacific: a genetic trail. Clarendon, Oxford, pp 120–173

Spriggs M (1984) The Lapita cultural complex. In: Szathmary RL, Szathmary S (eds) Out of Asia. Australian National University Press, Canberra, pp 185–202

Stoneking M (1993) DNA and recent human evolution. Evolutionary Anthropol 2:60–73

Stoneking M, Hedgecock D, Higuchi RG, Vigilant L, Erlich HA (1991) Population variation of human mtDNA control region sequences detected by enzymatic amplification and sequence-specific oligonucleotide probes. Am J Hum Genet 48:370–382

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

Stoneking M, Sherry ST, Redd AJ, Vigilant L (1992) New approaches to dating suggest a recent age for the human mtDNA ancestor. Phil Trans R Soc Lond B 337:167–175

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

Terrell J (1986) Prehistory in the Pacific Islands. Cambridge University Press, Cambridge

Trent RJ, Mickleson KWP, Wilkinson T, Yakas J, Dixon MW, Hill PJ, Kronenberg H (1986) Globin genes in Polynesians have many rearrangements including a recently described γγγ/. Am J Hum Genet 39:350–360

Turner, CG II (1987) Late Pleistocene and Holocene population history of East Asia based on dental variation. Am J Phys Anthropol 68:149–155

Vigilant L, Pennington R, Harpending H, Kocher TD, Wilson AC (1989) Mitochondrial DNA sequences in single hairs from a southern African population. Proc Natl Acad Sci USA 86:9350–9354

White JP, O'Connell JF (1982) A prehistory of Australia, New Guinea, and Sahul. Academic Press, London

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

Winichagoon P, Higgs DR, Goodbourn SEY, Lamb J, Clegg JB, Weatherall DJ (1982) Multiple rearrangements of the human embryonic zeta globin genes. Nucleic Acids Res 10:5853–5868

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 Res 15:529–542