Am. J. Hum. Genet. 65:808–828, 1999

Peopling of Sahul: mtDNA Variation in Aboriginal Australian and Papua New Guinean Populations

Alan J. Redd[1,*] and Mark Stoneking[1,2]

[1]Department of Anthropology, The Pennsylvania State University, University Park, Pennsylvania and [2]Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

Abstract

We examined genetic affinities of Aboriginal Australian and New Guinean populations by using nucleotide variation in the two hypervariable segments of the mtDNA control region (CR). A total of 318 individuals from highland Papua New Guinea (PNG), coastal PNG, and Aboriginal Australian populations were typed with a panel of 29 sequence-specific oligonucleotide (SSO) probes. The SSO-probe panel included five new probes that were used to type an additional 1,037 individuals from several Asian populations. The SSO-type data guided the selection of 78 individuals from Australia and east Indonesia for CR sequencing. A gene tree of these CR sequences, combined with published sequences from worldwide populations, contains two previously identified highland PNG clusters that do not include any Aboriginal Australians; the highland PNG clusters have coalescent time estimates of ~80,000 and 122,000 years ago, suggesting ancient isolation and genetic drift. SSO-type data indicate that 84% of the sample of PNG highlander mtDNA belong to these two clusters. In contrast, the Aboriginal Australian sequences are intermingled throughout the tree and cluster with sequences from multiple populations. Phylogenetic and multidimensional-scaling analyses of CR sequences and SSO types split PNG highland and Aboriginal Australian populations and link Aboriginal Australian populations with populations from the subcontinent of India. These mtDNA results do not support a close relationship between Aboriginal Australian and PNG populations but instead suggest multiple migrations in the peopling of Sahul.

Received November 24, 1998; accepted for publication July 7, 1999; electronically published August 9, 1999.

Address for correspondence and reprints: Dr. Alan J. Redd, Laboratory of Molecular Systematics and Evolution, Department of Ecology and Evolutionary Biology, Biological Sciences West 239, University of Arizona, Tucson, AZ 85721. E-mail: aredd@dakotacom.net

* Present affiliation: Laboratory of Molecular Systematics and Evolution, Department of Ecology and Evolutionary Biology, University of Arizona, Tucson.

0002-9297/1999/6503-0028$02.00

Introduction

Estimates of the dates and the process of human colonization of Sahul, the single Pleistocene landmass joining Australia, New Guinea, and Tasmania, suggest that this event was ancient and rapid. Radiocarbon dating from dozens of sites (Jones 1995; O'Connell and Allen 1998) and luminescence dating from fewer sites (e.g., Groube et al. 1986) both indicate that humans occupied the far reaches of Sahul by ~35,000–40,000 years ago. However, there is growing evidence that humans reached Sahul much earlier. Luminescence dating from two sites in northern Australia suggests that human colonization of Sahul occurred as much as 53,000–60,000 years ago (Roberts et al. 1990; Roberts and Jones 1994). Paleovegetation change in Australia provides evidence that human immigrants arrived ~60,000–65,000 years ago (Johnson et al. 1999; Miller et al. 1999). Furthermore, a human skeleton from Lake Mungo recently was dated, yielding an age of 62,000 years ± 6,000 years (Thorne et al. 1999). After initial colonization, the pace of movement of people into all regions of Sahul may have been quicker than has been estimated on the basis of the archeological record. For example, Birdsell's (1957) simulations of rates of population growth and movement into an unoccupied niche suggest that humans could have dispersed throughout Sahul in as little as 2,000 years. Because Australia and New Guinea, and Australia and Tasmania, remained joined by land shelves during the Pleistocene until sea levels rose—~8,000 and 12,000 years ago, respectively (White and O'Connell 1982; Jones 1995)—early populations of Sahul could have shared genes for ~30,000–57,000 years after initial settlement.

Given this common history in Sahul, the predominant view is that modern Aboriginal Australian and New Guinean highland populations are the descendants of the earliest migration into this region. This common-origin hypothesis specifies New Guinean highlanders because portions of coastal New Guinea were later colonized by Austronesian speakers—who apparently bypassed Australia and did not penetrate the highlands—~5,000 years ago (Bellwood 1989).

However, various lines of evidence bearing on the hypothesis of a common origin shared by Aboriginal Australians and New Guineans do not present a clear picture. For example, two previous studies of mtDNA data are equivocal with respect to the common-origin hypothesis. A phylogenetic analysis based on mtDNA restriction-enzyme data did not show any association between people from Papua New Guinea (PNG) and Aboriginal Australians in a tree that included African, Asian, and European individuals (Stoneking et al. 1990). Stoneking and colleagues suggested that there have been major population movements in PNG over the last 10,000 years that have altered the genetic relationship between Australian and PNG populations. In contrast, recent studies by van Holst Pellekaan et al. (1997, 1998) of mtDNA control region (CR) sequence variation in hypervariable segment 1 (HVS-I) of two Aboriginal Australian populations (Riverine and Desert regions) showed that the Aboriginal Australian populations are genetically least distant from PNG highlanders, when compared with a limited number of mtDNA from African and Asian populations.

MtDNA-sequence data are rich and they are well suited for tracking the relative influence of histories of population fission and/or gene flow. In addition, models have been developed explicitly for understanding demographic histories that can be inferred from mtDNA-sequence data. Although a single gene in the evolutionary history of populations, mtDNA reveals distinct maternal histories that are unclouded by the loss of phylogenetic signal via recombination. Although the Y chromosome also shares these attributes, mtDNA data are expected to show less influence from recent historic male-mediated admixture.

In this study, we examined the genetic relationship of Aboriginal Australian and PNG highland populations by using two mtDNA data sets. First, sequence-specific oligonucleotide (SSO) typing (Stoneking et al. 1991) of HVS-I and HVS-II of the CR was performed in samples of 1,355 individuals from Aboriginal Australian, PNG, and several Asian populations. Second, CR-sequence data from HVS-I and HVS-II were collected from samples of Aboriginal Australians and east Indonesians and were compared with previously published PNG sequences (Stoneking et al. 1992). Several analyses (phylogenetic, multidimensional scaling [MDS], intermatch, and genetic structure) of these mtDNA data did not suggest a common genetic origin of extant Aboriginal Australian and PNG highland populations. Instead, Aboriginal Australians showed close affinities with populations from the subcontinent of India, whereas PNG highlanders showed more-ancient divergence from all populations with hints of ancient ties to African populations.

Samples and Methods

Samples

Figure 1 shows the locations of the population samples in this study, as well as of those gathered from the literature. All samples in this study were obtained and handled in accordance with the guidelines of the Office of Regulatory Compliance of the Pennsylvania State University.

Australia.—Samples were obtained from two areas, northwest Australia and Arnhem Land. Purified DNA samples from 105 individuals from northwest Australia (Great Sandy Desert and Kimberley regions) derive from five locations: Balgo, Christmas Creek, Derby, Looma, and Turkey Creek. Samples from 95 individuals from Arnhem Land were provided in the form of blood stains on cloth, from which DNA was extracted with the IsoQuick (MicroProbe) kit, following the manufacturer's directions. One CR sequence from an Aboriginal Australian was previously published (Vigilant et al. 1991).

PNG.—Purified DNA from 119 samples from several PNG regions, including the eastern highlands, southern highlands, north coast, south coast, and offshore islands, have been described elsewhere (Stoneking et al. 1990). PNG sequences from 63 individuals were previously published (Vigilant et al. 1991; Stoneking et al. 1992; Redd et al. 1995).

East Indonesia.—DNA samples from 146 individuals from two islands in the Moluccas (Hiri and Ternate) and four islands in the Nusa Tenggaras (Alor, Flores, Roti, and Timor) were described elsewhere by Redd et al. (1995). East Indonesian CR sequences from 38 individuals were previously published (Stoneking et al. 1992; Redd et al. 1995).

Additional CR sequences examined in this paper include 248 Africans (Vigilant et al. 1991; Graven et al. 1995), 23 Asians (Vigilant et al. 1991), 102 southern Indians (Mountain et al. 1995), 115 Europeans (Vigilant et al. 1991; Piercy et al. 1993), and 25 Polynesians (Vigilant et al. 1991; Redd et al. 1995), as well as four chimpanzee sequences (Foran et al. 1988; Kocher and Wilson 1991).

9-bp Deletion

A 9-bp deletion in the COII/tRNA^{Lys} intergenic region of the mtDNA is a useful marker in studies of the dispersal of Austronesian speakers and is closely associated with the evolutionary history of Polynesians (Melton et al. 1995; Redd et al. 1995; Sykes et al. 1995). The 200 Australian samples were screened for the presence or absence of the 9-bp deletion, with use of published methods (Redd et al. 1995). Similar typings were previously

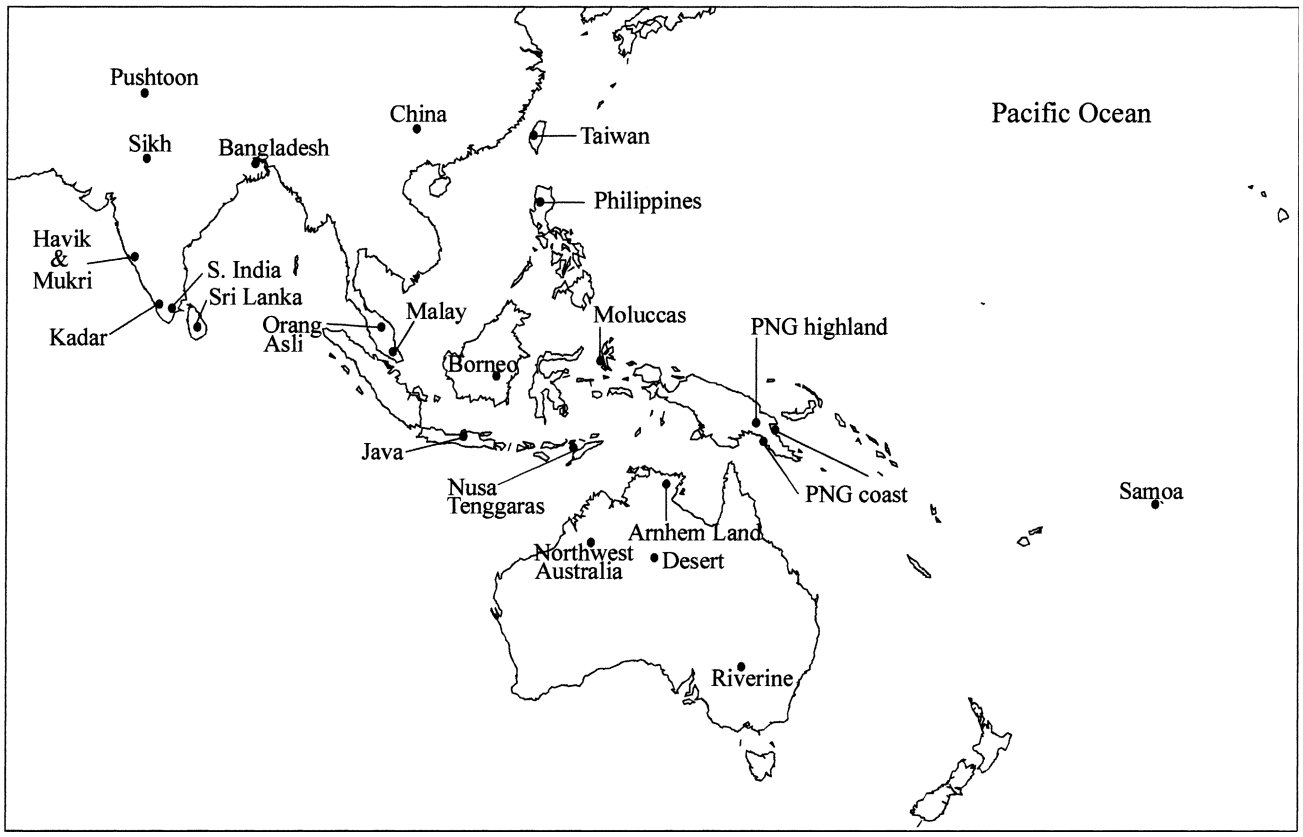


Figure 1 Map of approximate locations of population samples

conducted in the PNG (Stoneking and Wilson 1989) and east Indonesian (Redd et al. 1995) samples.

SSO Typing

In the present study, 318 individuals from Australia and PNG (see table 1) were typed for mtDNA HVS-I and HVS-II variation, with use of 29 SSO probes. The PCR conditions, DNA-typing methods, and 24 of the probes are described elsewhere (Stoneking et al. 1991; Melton et al. 1995). Five new SSO probes (IF1, IF2, IG1, IG2, and IIB4; see table 2) were designed to detect sites corresponding to CR sequences of two previously identified PNG clusters, PNG2 and PNG3, on the basis of phylogenetic analysis of CR sequences (Stoneking et al. 1992). Probes IF2 and IIB4 detect sites that correspond to the PNG3 cluster, whereas probe IG2 detects a site that corresponds to the PNG2 cluster. An additional 1,037 Asian samples (from east Indonesia, Borneo, Java, Philippines, Taiwan, Malaysia, Orang Asli, China, Pushtoon, Bangladesh, Sikh, and southern India/Sri Lanka; see table 1 and fig. 1), which had been previously typed with 24 SSO probes (Melton et al. 1995), were typed with these five new probes.

CR Sequencing

Samples of 25, 28, and 25 individuals, from northwest Australia, Arnhem Land, and east Indonesia, respectively, were selected for CR sequencing for both HVS-I and HVS-II regions. DNA sequencing of PCR products with biotinylated primers was performed as described by Redd et al. (1995).

Data Analysis

SSO-Type Analyses

The combined SSO-probe panel includes 29 probe variants across 11 subregions of the CR. HVS-I includes 7 (IA [1, 2, and 3]; IF [1 and 2]; IB [1, 2, and 3]; IE [1, 2, and 3]; IC [1, 2, and 3]; IG [1 and 2]; and ID [1 and 2]), and HVS-II includes four (IIA [1 and 2]; IIB [1, 2, 3, and 4]; IIC [1, 2, and 3]; and IID [1 and 2]). Thus, the SSO type for a single individual consists of an array of eleven variants (e.g., 3, 2, 2, 1, 3, 1, 1, 2, 2, 1, and 1), one for each subregion. In instances when none of the SSO probes annealed to a particular subregion, a "0" was assigned for the purpose of analysis, regardless of the underlying nucleotide variation. The reliability of

Table 1

Populations and Individuals Examined for mtDNA Variation

Population	No. of Individuals	9-bp Deletion[a] (%)	29 SSO[b] Probes	CR[c] Sequences
Australia:				
Arnhem Land	95	0	95	28
Northwest Australia	105	0	105	25
PNG:				
Highland	70	0	70	30
Coastal	48	42	48	33
East Indonesia:				
Moluccas	49	16	49	17
Nusa Tenggaras	94	24	94	46
Polynesia	25	100	0	25
Asia:				
Java	98	26	98	0
Borneo	95	24	95	0
Philippines	176	40	176	0
Taiwan	82	42	82	0
Malay	81	26	81	0
Orang Asli	30	37	30	0
China	103	22	103	0
Pushtoon	76	0	76	0
Bangladesh	31	0	31	0
Sikh	47	0	47	0
Southern Indian	75	8	75	0
Totals	1,380		1,355	204

[a] Sources for data are as follows: Australia, this study; PNG, Stoneking and Wilson (1989); east Indonesia and Polynesia, Redd et al. (1995); and Asia, Melton et al. (1995).

[b] Sources for data are as follows: Australia and PNG (this study, 29 probes), east Indonesia, Asia (this study, 5 probes; Melton et al. 1995, 24 probes).

[c] Sources for data are as follows: Australia, this study; east Indonesia, this study, Stoneking et al. (1992), and Redd et al. (1995); Polynesia, Redd et al. (1995) and Vigilant et al. (1991); and PNG, Vigilant et al. (1991); Stoneking et al. (1992) and Redd et al. (1995).

these SSO data was checked by comparison with 179 CR sequences from Australia, PNG, and east lndonesia; 99.78% of the SSO-typing results were concordant with the DNA-sequence results. Unbiased estimates of haplotype diversity and the associated variance were calculated with the ARLEQUIN package (ARLEQUIN).

A genetic-distance matrix was calculated between unique SSO types, on the basis of the number of mutational steps between pairs of SSO types, with the underlying molecular information. The distance from the 0 variant to all other variants was set to one mutational step. The genetic-distance matrix between all SSO types and their frequency in each population were used to calculate a variety of statistics and population distances with ARLEQUIN. Population pairwise F_{ST} distances were computed for all pairs of populations, and their significance was tested with a nonparametric permutation test that involves permuting haplotypes between populations and recalculating the distances (ARLEQUIN). In addition, an exact test of population differentiation was carried out with the construction of contingency tables consisting of the number of populations by the number of haplotypes, and the exploration of all possible states of the table was done with a Markov chain (Raymond and Rousset 1995). A matrix of Slatkin-linearized F_{ST} values (Slatkin 1995; ARLEQUIN) was used for construction of a population tree with the neighbor-joining (NJ) method (Saitou and Nei 1987), as implemented in MEGA (Kumar et al. 1993). The matrix of Slatkin-linearized F_{ST} values was used as input for an MDS analysis (Kruskal 1964) with maximum-likelihood estimation (MULTISCALE).

Sequence Analysis

CR sequences from 692 individuals with 754 nucleotides of the combined HVS-I (positions 16,024–16,400) and HVS-II (positions 31–407) were aligned manually with the human-reference sequence (Anderson et al. 1981). Observed insertion polymorphisms were removed. To minimize the impact of missing data on the analysis, sequences that were missing >5% of the sites were excluded, as well as some sequences (n = 8) that were missing sites in polymorphic regions. The final sample of 430 individuals consists of CR sequences from 44 Africans (six !Kung, 10 Biaka Pygmies, three Mbuti Pygmies, 11 Yorubans, eight Mandenka, and six African Americans), 107 Europeans (93 British and Welch and 14 northern Europeans), 81 southern Indians (42 Havik, 34 Mukri, and five Kadar), 19 Asians (14 Chinese, one Filipino, one Japanese, one Taiwanese, one Vietnamese, and one Amerindian), 54 Aboriginal Australians (28 from Arnhem Land, 25 from northwest Australia, and one from Western Australia), 46 PNG (31 from PNG coast and 15 from PNG highland), 59 east Indonesians (16 Moluccans and 43 Nusa Tenggarans), and 20 Polynesians (19 American Samoans and one Tongan). This sample of 430 sequences shares 561 nucleotide sites.

A gene tree was produced, with the NJ method (Saitou and Nei 1987), from a matrix of the proportion of nucleotide differences with use of the MEBoot program (provided by K. Tamura). A p-distance was selected because it has a low coefficient of variance, and the successful performance of a genetic distance is greatly influenced by variance (Tajima and Takezaki 1994). The gene-tree topology was nearly identical under different measures of genetic distances. The NJ tree was further searched to find the minimum evolution (ME) tree (Rzhetsky and Nei 1992, 1993). The reliability of the interior branches of the ME tree was tested with the confidence probability (CP) test (Rzhetsky and Nei 1992, 1993), with use of 1,000 replicates. Furthermore, the clusters in the ME tree were examined with the bootstrap test (Felsenstein 1985), with use of 500 replicates.

To account for bias introduced by rate variation

Table 2

Additional SSO Probes

Nucleotide Position[a] and Marker	Polymorphic Sites[b]	Hybridization Temperature (°C)	Probe Sequence (5′–3′)	Strand[c]
16,138–16,155:				
IF1	16144 T, 16148 C	50	AATACTTGACCACCTGTA	L
IF2	16144 C, 16148 T	50	AATACTCGACTACCTGTA	L
16,344–16,361:				
IG1	16357, T	50	CGAGAAGGGATTTGACTG	H
IG2	16357, C	55	CAGTCAAATCCCTCCTCG	L
141–158:				
IIB4	143, A	50	CTACCTCATCCTATTATT	L

NOTE.—Probes added to those of Stoneking et al. (1991) and Melton et al. (1995).

[a] Nt positions flanked by the SSO probes, as appear in the human reference (Anderson et al. 1981).

[b] Nt position and state on the L strand, detected by the SSO probe.

[c] "H" = heavy strand and "L" = light strand (Anderson et al. 1981).

among sites in the CR, we estimated the gamma parameter, *a,* by counting the number of substitutions for each site (Kocher and Wilson 1991; Tamura and Nei 1993; Wakeley 1993) for the NJ tree topology, with GMAES (Takezaki N).

The height of the branch points of phylogenetic clusters of interest were obtained by the construction of a linearized tree under the assumption of a molecular clock (Takezaki et al. 1995), with use of the LINTRE package (Takezaki N). LINTRE includes a two-cluster test that involves examining the statistical equality of the average substitution rates for two clusters of interest. For the two-cluster test, we first constructed a NJ tree of the phylogenetic cluster and included an outgroup(s) using the Tamura-Nei model of substitution (Tamura and Nei 1993) and a gamma-parameter estimate. Rate constancy was then forced on the NJ topologies (Takezaki et al. 1995), thus providing estimates of the height of the node and their standard errors (SE), which were used to estimate the coalescence times of phylogenetic clusters. The gamma-corrected Tamura-Nei distance was selected because it is, biologically, the most realistic and it increases linearly with time.

The average-nucleotide diversity within and between populations (d_X and d_{XY}, respectively; Nei and Jin 1989) was computed with the SENDBS program (Takezaki N). We constructed a population tree from d_{XY} by using a matrix of the proportion of nucleotide differences. We also used SENDBS to compute Felsenstein's (1985) bootstrap probabilities of the NJ tree, using 1,000 replicates. We computed estimates of population divergence with SENDBS by using net divergence, d_A (Nei and Jin 1989), under the Tamura-Nei model of substitution (Tamura and Nei 1993) with a gamma correction. The matrix of d_{XY} distances (Tamura and Nei model with a gamma correction) was used as input for MDS analysis (Kruskal 1964) using maximum-likelihood estimation with the MULTISCALE program.

Pairwise-difference distributions were used to examine demographic history and population separations (Rogers and Harpending 1992; Harpending et al. 1993). Mismatch and intermatch distributions were constructed with the MISMATCH and IWAVE programs, respectively. Harpending's raggedness statistic, *r,* and Tajima's neutrality statistic, *D,* can be used to differentiate between expanding and constant population size (Tajima 1989; Harpending 1994; Aris-Brosou and Excoffier 1996); both statistics were calculated with the ARLEQUIN package. Estimates of one of the parameters, τ, from the Rogers and Harpending (1992) model of sudden population growth, were used to characterize mismatch distributions; τ was computed with the MISMATCH package (Rogers 1995). Estimates of past population expansions and separations are measured as $\tau = 2\mu t$ units of mutational time, in which μ is the total mutation rate over all sites in the sample and t is time in generations (Rogers 1995).

Estimation of 95% confidence intervals (CI) of the times of population expansion and the coalescence of phylogenetic clusters were computed according to Redd et al. (1995) with their formula (1). This method incorporates the variance in the estimates of the mutation rate and τ (or genetic distance). We estimated the SE of τ by simulating population expansions, using the estimates of τ, sample size, and initial population size. We calculated 95% CI ($\pm$ 2 $\times$ SE), assuming a substitution rate of $8.85 \times 10^{-8} \pm 0.9$ /site/year (Horai et al. 1995; Bonatto et al. 1997).

Results

9-bp Deletion

The 9-bp deletion was absent in the 200 Aboriginal Australians tested (table 1). The frequency of the deletion is <1% in the 656 Aboriginal Australians typed for

this marker (Hertzberg et al. 1989; Stoneking and Wilson 1989; Betty et al. 1996; van Holst Pellekaan et al. 1997), including those typed in this study.

SSO Types

Table 3 shows the haplotype diversity of the SSO types (unique arrays of 11 variants) per population and the frequencies of the new SSO-probe variants. The SSO probes detected high levels of variation within populations; the haplotype-diversity values ranged from 0.92 to 0.99. All of the new probe variants (IF2, IIB4, and IG2) were found in individuals from the PNG coast and the PNG highlands but they were significantly more frequent in the PNG highlands ($P < .01$; χ^2 test). The IF2 variant was restricted to individuals from PNG (23%–40%) and east Indonesia (5%–6%). The IIB4 variant was detected in individuals from PNG (8%–16%) and it was present at very low frequencies in northwest Australian (1%) and Asian populations (< 1% on average). The IG2 variant was present in moderate frequencies in individuals from PNG (15%–36%) lower frequencies in individuals from east Indonesia (3%–8%), and was rare in Aboriginal Australians (2%) and Asians (< 1% on average).

A total of 504 SSO types were detected among the 1,355 individuals. Frequencies of sharing among the SSO types was ⩾5% in at least one population in only 35 SSO types, and was ⩾5% in at least two populations in seven SSO types. Twenty-two SSO types were shared among PNG highland, Aboriginal Australian, east Indonesian, and southern Indian populations: one type (in four individuals) was shared by Aboriginal Australians and PNG highlanders; two types (in 26 individuals) were shared by PNG highlanders and east Indonesians; four types (in 19 individuals) were shared by Aboriginal Australians and east Indonesians; nine types (in 37 individuals) were shared by Aboriginal Australians and southern Indians; and four types (in 23 individuals) were shared by east Indonesians and southern Indians. In addition, one type (in seven individuals) was shared among Aboriginal Australian, east Indonesian, and southern Indian populations; and one type (in seven individuals) was shared among all four populations.

PNG Clusters Inferred from SSO Type Data

After examination of the SSO types, samples from PNG populations were assigned to one of the clusters (PNG1, PNG2, and PNG3) that were defined previously from phylogenetic analysis (Stoneking et al. 1992). Characteristics of the PNG clusters and populations are given in table 4. Several patterns are apparent, consistent with results from previous analyses of PNG mtDNA restriction-enzyme typing (Stoneking et al. 1990) and CR sequencing (Stoneking et al. 1992). PNG1 types (9-bp deletion lineage; Redd et al. 1995) are found in populations from coastal regions only, whereas PNG2 and PNG3 types are more frequent in populations from the highland areas but are also present in populations from

Table 3

mtDNA SSO-Type Diversity and New Variant Frequencies

Population	No. of Individuals	No. of Unique Types	Diversity[a] (h ± SE)	IF2[b] (in %)	IIB4[b] (in %)	IG2[b] (in %)
Australia:						
Arnhem Land	95	51	.970 ± .009	0	0	2.1
Northwest Australia	105	51	.981 ± .004	0	1.0	1.9
PNG:						
Highlands	70	33	.948 ± .014	4.0	15.7	35.7
Coastal	48	24	.943 ± .017	22.9	8.3	14.6
East Indonesia:[c]						
Moluccas	49	29	.967 ± .013	6.1	0	8.2
Nusa Tenggaras	94	54	.974 ± .008	5.3	0	3.2
Asia:[c]						
Java	98	68	.981 ± .007	0	0	1.0
Borneo	95	61	.979 ± .007	0	1.1	1.1
Philippines	176	73	.955 ± .008	0	0	.6
Taiwan	82	39	.967 ± .008	0	0	0
Malay	81	53	.979 ± .007	0	0	2.5
Orang Asli	30	15	.924 ± .028	0	0	0
China	103	73	.992 ± .003	0	0	1.9
Pushtoon	76	56	.990 ± .004	0	0	0
Bangladesh	31	30	.998 ± .009	0	0	0
Sikh	47	35	.981 ± .010	0	4.3	0
Southern Indian	75	66	.996 ± .003	0	1.3	1.3

[a] See Nei (1987) page 180.
[b] New SSO probes (see table 2).
[c] Previously typed with 24 SSO probes (Melton et al. 1995).

Table 4

Characteristics of PNG Clusters per Populations, as Defined by SSO-Type Data

Characteristic	PNG1 Cluster	PNG2 Cluster	PNG3 Cluster	PNG Other	Total
No. of Individuals	21	33	48	16	118
No. of Types	10	11	17	14	52
Residence (in %):					
Highland	.0	37.1	47.1	15.7	70
Coastal	43.8	14.6	31.3	10.4	48
Language (in %):[a]					
NAN	5.9	31.8	48.2	14.1	85
AN	53.3	20.0	16.7	10.0	30
9-bp deletion (in %):					
Nondeleted	.0	34.0	49.5	16.5	97
Deleted	100.0	.0	.0	.0	21
Geography (in %)[b]:					
EH	.0	58.6	31.0	10.3	29
SH	.0	22.0	58.5	19.5	41
NC	11.8	17.6	58.8	11.8	17
SC	57.9	10.5	15.8	15.8	19
IS	66.7	16.7	16.7	.0	12
Population (in %):					
Aboriginal Australian	.0	.5	.0		200
East Indonesian	25.9	4.2	4.2		143

NOTE.—PNG phylogenetic clusters (PNG 1–3) are as defined by phylogenetic analysis (Stoneking et al. 1992) and inferred from SSO-type data.

[a] "NAN" = non-Austronesian and AN = Austronesian. Language affiliation not known for three samples.

[b] "EH" = eastern Highlands, "SH" = southern Highlands, "NC" = north coast, "SC" = south coast, and "IS" = offshore islands (Stoneking et al. 1990).

the coast. Specific trends are also apparent when residence (highland versus coast) and language (Austronesian [AN] vs. Non-Austronesian [NAN] speakers) are compared, with respect to the PNG clusters. PNG2 and PNG3 are significantly more frequent among populations from the highlands ($P < .01$; χ^2 test) and NAN speakers ($P < .01$; χ^2 test) compared with the PNG coastal populations and AN speakers, respectively. In addition, the frequencies of the PNG clusters vary by geographic region. PNG1 sequences are found most frequently among island and south coast regions, although sample sizes preclude reliable χ^2 tests. PNG2 and PNG3, which together account for 84% of the highland sample, differ significantly in frequency within the highlands, with PNG2 most frequent in populations from the eastern highlands ($P < .01$; χ^2 test) and PNG3 most frequent in populations from the southern highlands ($P < .01$; χ^2 test). Table 4 also shows SSO types from east Indonesia and Australia that were assigned to the PNG clusters. A single individual from Australia has an SSO type that belongs to PNG2, whereas there are many SSO types from east Indonesia that belong to PNG1 (26%) and fewer that belong to PNG2 (4%) or PNG3 (4%).

CR Sequences

The SSO-type data from the populations guided the selection of a representative sample (with consideration given to the range and frequency of types) of mtDNA from Australian and east Indonesian populations for CR sequencing. (The composition of the 15 CR sequences from the PNG highlands is also representative [PNG2 = 33%, PNG3 = 47%, and PNG other = 20%] of the 70 PNG highland individuals shown in table 4.) The sample of individuals selected for CR sequencing also included six mtDNA with shared SSO types between populations: one between Aboriginal Australian and PNG highland populations; two between east Indonesian and PNG highland populations; two between Aboriginal Australian and east Indonesian populations; and one among Aboriginal Australian, PNG highland, and east Indonesian populations. Furthermore, the sample included one Aboriginal Australian who had the IIB4 site.

Sequence data were obtained for 78 individuals, including 25 from northwest Australia, 28 from Arnhem Land, 8 from the Moluccas, and 17 from the Nusa Tenggaras. Comparisons of these sequences with previously published sequences from PNG and east Indonesia (Vigilant et al. 1991; Stoneking et al. 1992; Redd et al. 1995), excluding mtDNA with the 9-bp deletion, revealed 100 unique sequence types (fig. 2*A*and2B). Of the 145 variable sites, nearly twice as many ($n = 91$) occur in HVS-I compared with HVS-II ($n = 54$). Transitions alone occurred at 127 sites, transversions alone at 12 sites, and both transitions and transversions at 6 sites. Sequence-type sharing within populations was observed for five sequences (types 101, 102, 106, 247, and 275;fig. 2*A* and *B*). However, type sharing between populations was not observed among the sequences infigure 2. This is in sharp contrast to mtDNA with the 9-bp deletion, in which identical CR sequences are shared across populations from Polynesia, the PNG coast, east Indonesia, and Asia (Redd et al. 1995).

Phylogenetic Results

Gene Tree

Figure 3 shows an ME tree, made on the basis of 597 sites for 308 unique sequences, constructed from a matrix of the proportion of nucleotide differences and rooted with four chimpanzee sequences (Foran et al. 1988; Kocher and Wilson 1991). The ME tree roots first among African sequences. One type (299), from east Indonesia, clustered among the African sequences. Four Aboriginal Australian sequences (types 273–276) and one PNG coast sequence (type 277) clustered near the basal portion of the tree, adjacent to a large African cluster. The ME tree contains the three PNG clusters, largely intact, that were first identified by Stoneking et

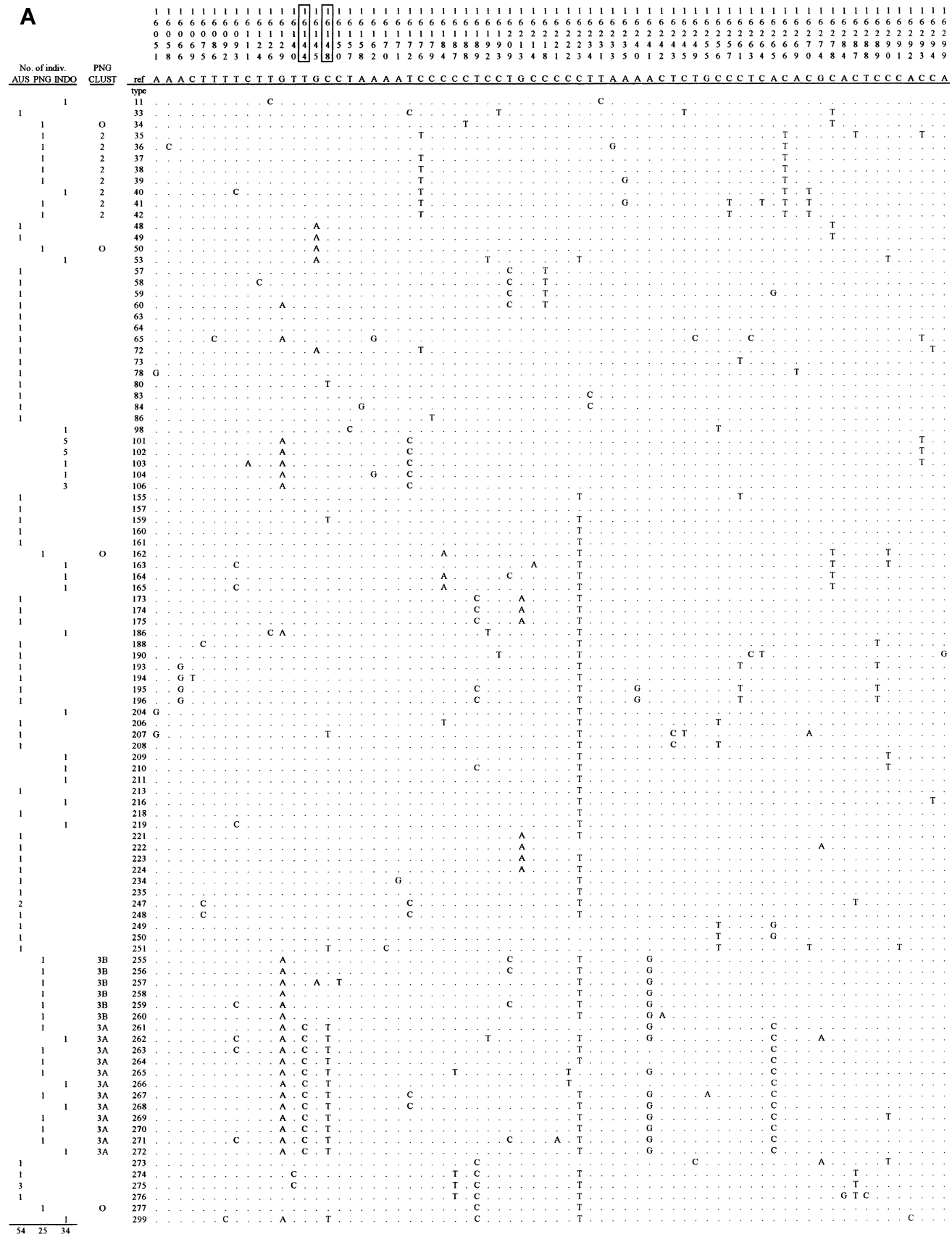
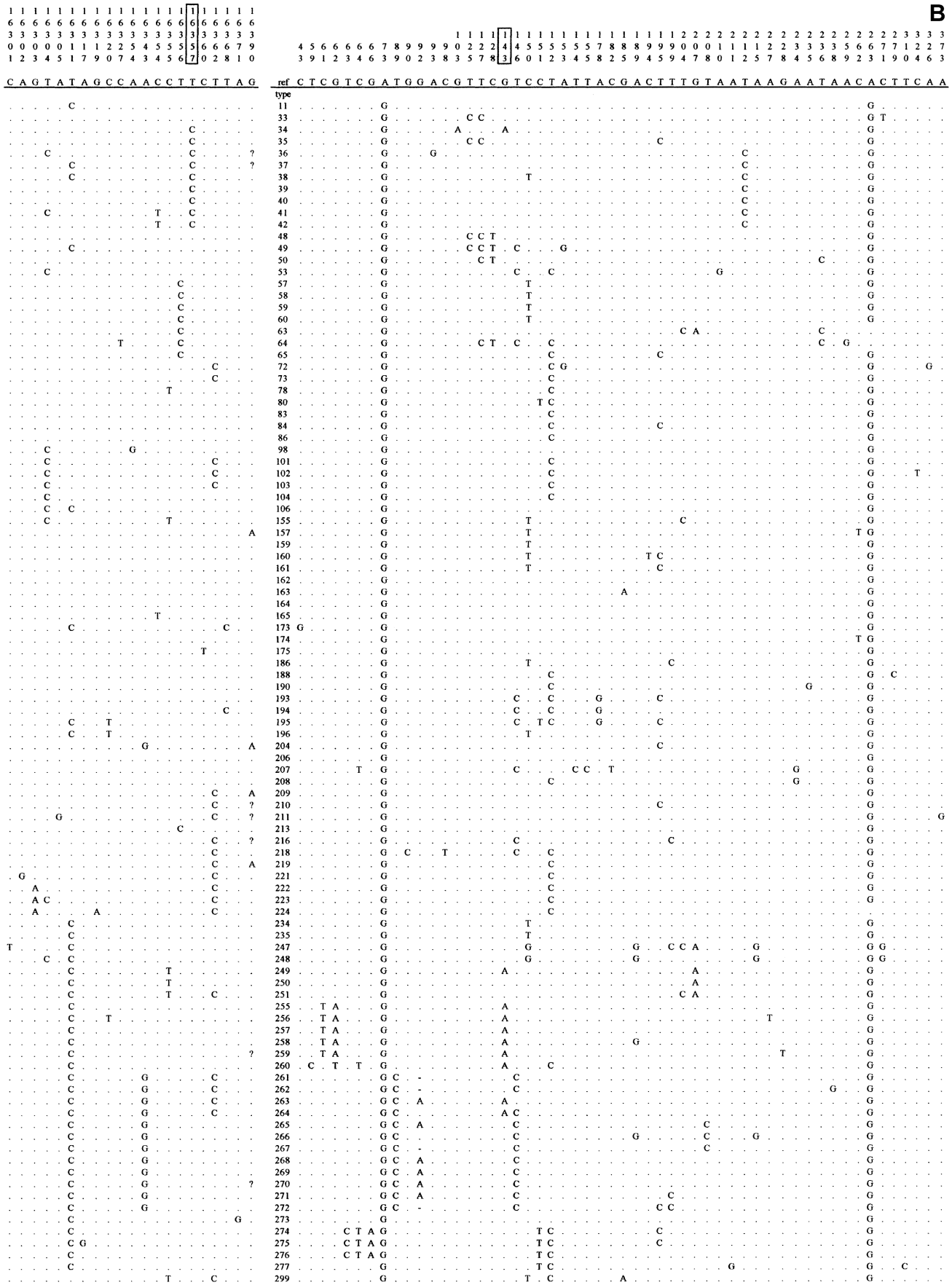


Figure 2 CR-nucleotide differences among Aboriginal Australian, PNG, and east Indonesian mtDNA. Positions 16,051–16,390 correspond to HVS-I , and positions 43–373 correspond to HVS-II in the published human reference (Anderson et al. 1981). Boxed numbers indicate PNG highland sites probed with SSO probes. The numbers of individuals sharing a CR type are shown in the left columns, in which "AUS" and



"INDO" denote Aboriginal Australian and east Indonesian, respectively. "PNG CLUST" denotes phylogenetic clusters PNG2, PNG*3A*, PNG*3B*, and PNG O (other) (Stoneking et al. 1992). CR sequences from mtDNA with the 9-bp deletion were not included in this figure, since they have been shown elsewhere by Redd et al. (1995).

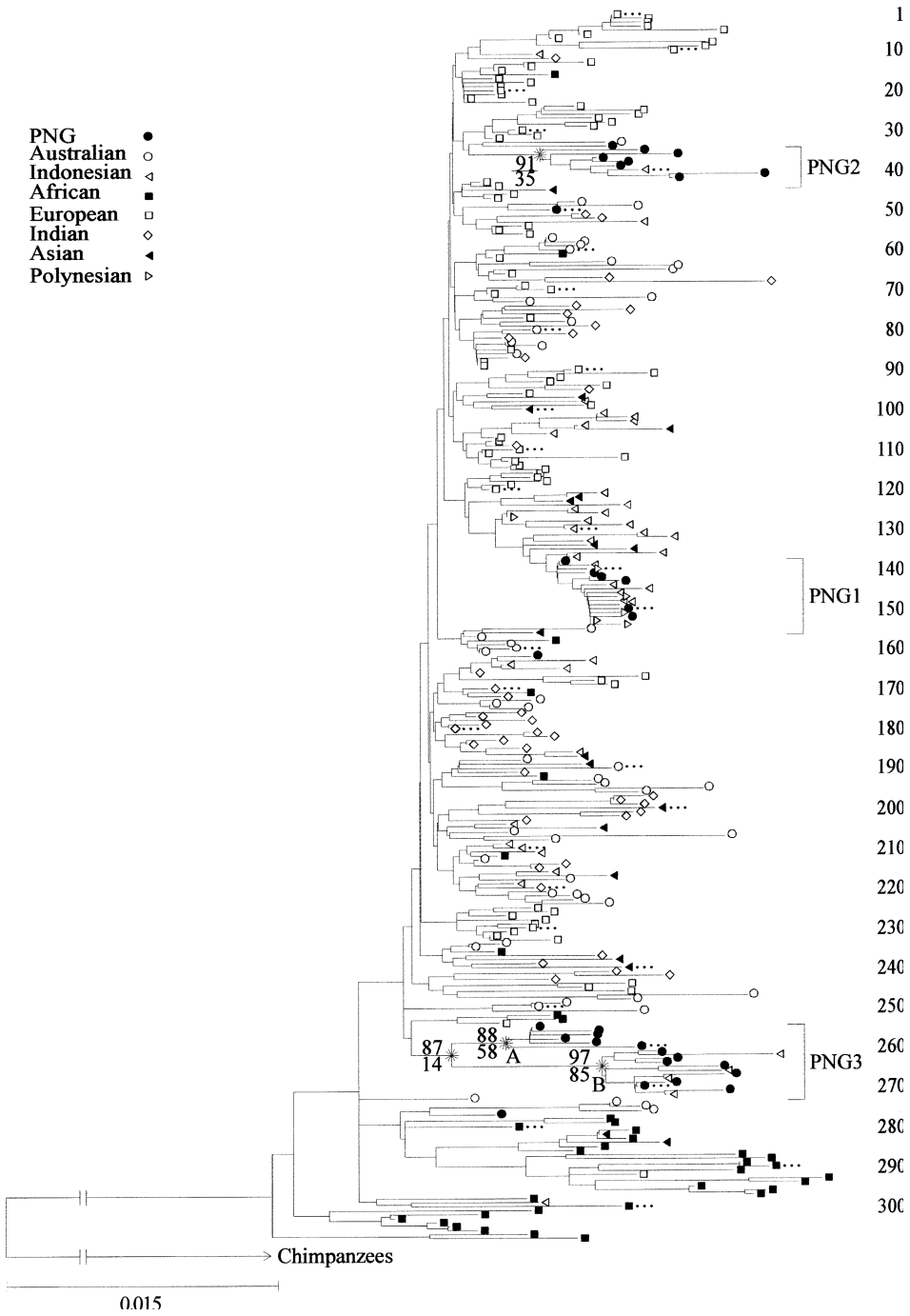


Figure 3 ME tree of 308 CR sequences rooted with four chimpanzee sequences. An NJ tree was first constructed from a distance matrix of the proportion of nucleotide differences, after which the ME tree was determined (Rzhetsky and Nei 1992, 1993). The numbers in the right column (1–300) indicate sequence-type numbers in increments of ten and correspond to the three dots after the geometric symbols. The PNG clusters (1, 2, and 3) are indicated with brackets. The letters "A" and "B" denote PNG3 subclusters. Asterisks (*) denote the PNG nodes, with confidence probabilities shown above, and the bootstrap probabilities shown below. Genetic distance is shown on the scale bar below the tree. Outgroup rooting with the Neanderthal sequence (Krings et al. 1997, 1999) produced nearly identical patterns of clustering.

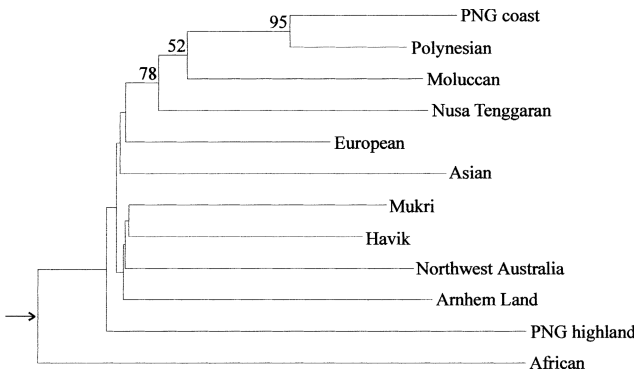


Figure 4 NJ population tree on the basis of CR sequences. The tree was constructed, with use of d_{XY} distances, from a matrix of the proportion of nucleotide differences. The arrow indicates the root, on the basis of chimpanzee outgroup and midpoint rooting. The numbers indicate the bootstrap probabilities (>50%).

al. (1992). The PNG3 cluster is located near the African cluster and includes four east Indonesians (types 262, 266, 268, and 272) in the PNG3B subcluster, whereas PNG3A remained monophyletic. The PNG2 cluster lost one PNG highlander (type 50) and includes one east Indonesian (type 40). All mtDNA with the Asian form of the 9-bp deletion (types 121–154) but one (type 156 [Japanese]) clustered together, consistent with previous findings (Redd et al. 1995). CP values for the highland PNG clusters and subclusters were all >87%, whereas bootstrap values were lower. Four PNG sequences did not fall into any of the PNG clusters, with three of them (types 34, 50, and 277) clustering with Australian sequences—albeit distantly (types 34, 50, and 277 differ by 11, 3, and 9 sites, respectively, from their nearest Australian neighbors)—and the other (type 162) clustering with east Indonesian sequences. In contrast, the Aboriginal Australian sequences formed several smaller clusters with their nearest neighbors from southern India (five occurrences), Africa (five occurrences), Europe (five occurrences), Asia (four occurrences), highland PNG (two occurrences), and coastal PNG (one occurrence). The east Indonesian sequences were also widely distributed across the tree, and their clusters had nearest neighbors from Asia (10 occurrences), PNG (eight occurrences), Polynesia (three occurrences), southern India (three occurrences), and Africa (two occurrences).

Coalescent Estimates

The rough age of the highland PNG clusters can be estimated from the genetic distance to the node of each cluster, with the assumption of a constant rate of mtDNA CR evolution. The estimated-sequence distance to the nodes of the PNG highland clusters (fig. 3) are as follows: PNG2, 0.71% ± 0.10%; PNG3A, 0.61% ± 0.10%; PNG3B, 0.45% ± 0.04%; and PNG3, 1.08% ± 0.04%. The assumption of a rate of mtDNA CR-sequence evolution of $8.85 \times 10^{-8} \pm 0.9$/site/year (Horai et al. 1995; Bonatto et al. 1997) results in a mean estimate of 80,226 years for PNG2 (95% CI = 66,289–94,163 years); 68,927 years for PNG3A (95% CI = 55,630–82,224 years); 50,847 years for PNG3B (95% CI = 43,980–57,715 years); and 122,034 years for PNG3 (95% CI = 108,826–135,242 years).

Population Tree and MDS Plot of CR Sequences

Figure 4 shows an NJ tree for populations, constructed from a matrix of the proportion of nucleotide differences, with d_{XY} distances. Midpoint and chimpanzee-outgroup rooting both place the root between Africa and the other populations. The next major separation in the tree occurs between PNG highland populations and the non-African populations. The Aboriginal Australian populations group with the southern Indian populations, although the bootstrap probabilities are <50%. The cluster that received more support (bootstrap values >50%) from the bootstrap test included the east Indonesian, PNG coastal, and Polynesian populations. A cophenetic-distance matrix (Rohlf and Sokal 1981), produced from the tree topology, correlates with the original distance matrix at value of $r = 0.84$ (Mantel 1967); the fit between the original distances and the one-dimensional tree is, thus, moderately close.

The MDS plot (fig. 5) of the CR sequences displayed patterns of population clustering that were very similar to the phylogenetic tree (fig. 4). Most of the populations grouped loosely in the center of the plot, whereas the African and PNG highland populations were each separated from all other populations. Both the phylogenetic tree and the MDS plot separated PNG highlanders and

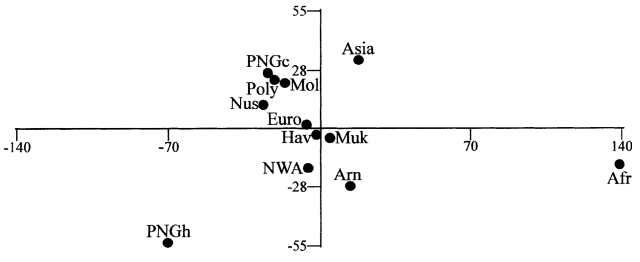


Figure 5 MDS population-plot of CR sequences. Abbreviations: "Afr" = African, "PNGh" = PNG highlander, "PNGc" = PNG coastal, "Poly" = Polynesian, "Mol" = Moluccas, "Nusa" = Nusa Tenggaras, "Asia" = Asian, "Euro" = European, "NWA" = northwest Australia, "Arn" = Arnhem Land, "Hav" = Havik, and "Muk" = Mukri. MDS was performed on a genetic-distance matrix with the assumption of the Tamura-Nei model of substitution with a gamma correction.

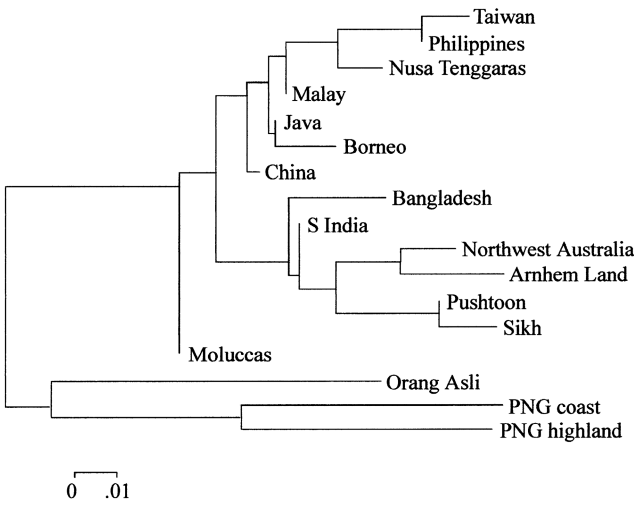


Figure 6 NJ population tree on the basis of SSO-type data. The tree was constructed from a matrix of Slatkin-linearized F_{ST} values that incorporates the genetic distance between 504 SSO-types, detected among 1,355 individuals, and their frequency in 17 populations. Genetic distance is indicated on the scale bar below the tree.

Aboriginal Australians and placed Aboriginal Australians closest to populations from the Indian subcontinent. However, the correlation between the original genetic distances and the two-dimensional plot was much closer ($r = 0.96$).

SSO-Type Population Tree and MDS Plot

The above analyses of mtDNA CR sequences indicate that Aboriginal Australian and PNG highland populations do not group together. In fact, Aboriginal Australian populations group with Indian populations, whereas the highland PNG population is separate from all other populations. To determine whether these patterns would persist in a larger sample of Asian populations, these analyses were repeated with the SSO-type data. Figure 6 shows an NJ tree for 17 populations, which was constructed from the SSO-type data. These data incorporate the genetic distances between 504 SSO types detected among 1,355 individuals. The SSO-type population tree shows a major division between the PNG highland, PNG coastal, and Orang Asli populations and the remaining populations. These remaining populations can be divided into two subclusters, one that includes the Aboriginal Australians, Bangladeshi, and southern and northern Indians and the Pushtoons; and a second subcluster that includes the Chinese and Southeast Asians. The Moluccan sample is located at a basal position to the second subcluster. A cophenetic-distance matrix, produced from the tree topology, correlates with the original distance matrix at a low level ($r = 0.73$).

Figure 7 shows an MDS plot of the SSO-type data. The two major clusters in the plot include a PNG cluster and a large cluster of the remaining populations. This large cluster can be divided into two subclusters, one that includes the Southeast Asian and southern Chinese populations, and another cluster that includes the Aboriginal Australian and Indian subcontinent populations. The Orang Asli population is separated from all other populations in the MDS plot. The correlation between the original genetic distances and the two-dimensional plot was close ($r = 0.97$)

The patterns of population clustering, determined from the CR-sequence data and the SSO-type data, were similar in most respects. Both data sets indicate that the highland PNG and Aboriginal Australians do not cluster together and that the Aboriginal Australians group with populations from the Indian subcontinent. However, the CR-sequence data separate highland and coastal PNG populations, whereas the SSO-type data place the coastal and highland PNG population samples closer together. SSO typing of the complete PNG coastal sample detected mtDNA from PNG1 (43.8% of the sample), from the highland-cluster PNG2 (14.6% of the sample), and from PNG3 (31.3% of the sample) (table 4). In contrast, the sample of CR sequences from the PNG coast contains more PNG1 sequences (68%) and fewer PNG2 (7%) and PNG3 (23%) sequences, because Redd et al. (1995) focused on the evolutionary history of the 9-bp deletion

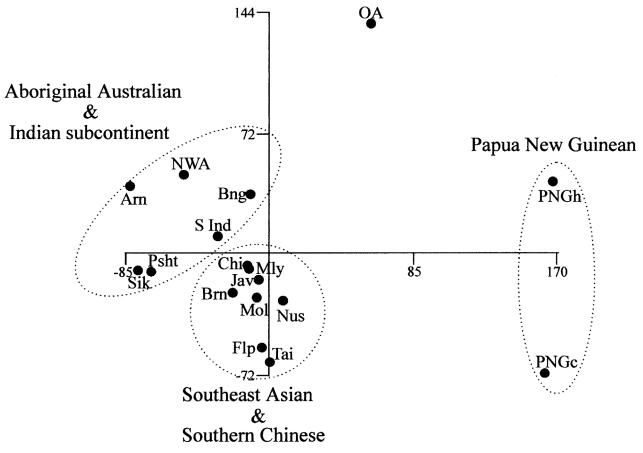


Figure 7 MDS population-plot of SSO-type data. Abbreviations: "PNGh" = PNG highlander, "PNGc" = PNG coastal, "OA" = Orang Aslian, "NWA" = northwest Australian, "Arn" = Arnhem Lander, "Bng" = Bangladeshian, "S Ind" = southern Indian, "Chi" = Chinese, "Psht" = Pushtoon, "Jav" = Javan, "Mly" = Malaysian, "Brn" = Bornean, "Flp" = Filipino, "Tai" = Taiwanese, "Mol" = Moluccasan, "Nus" = Nusa Tenggarasan. MDS was performed on a matrix of Slatkin-linearized F_{ST} values that incorporates the genetic distance between 504 SSO types, detected among 1,355 individuals, and their frequency in 17 populations.

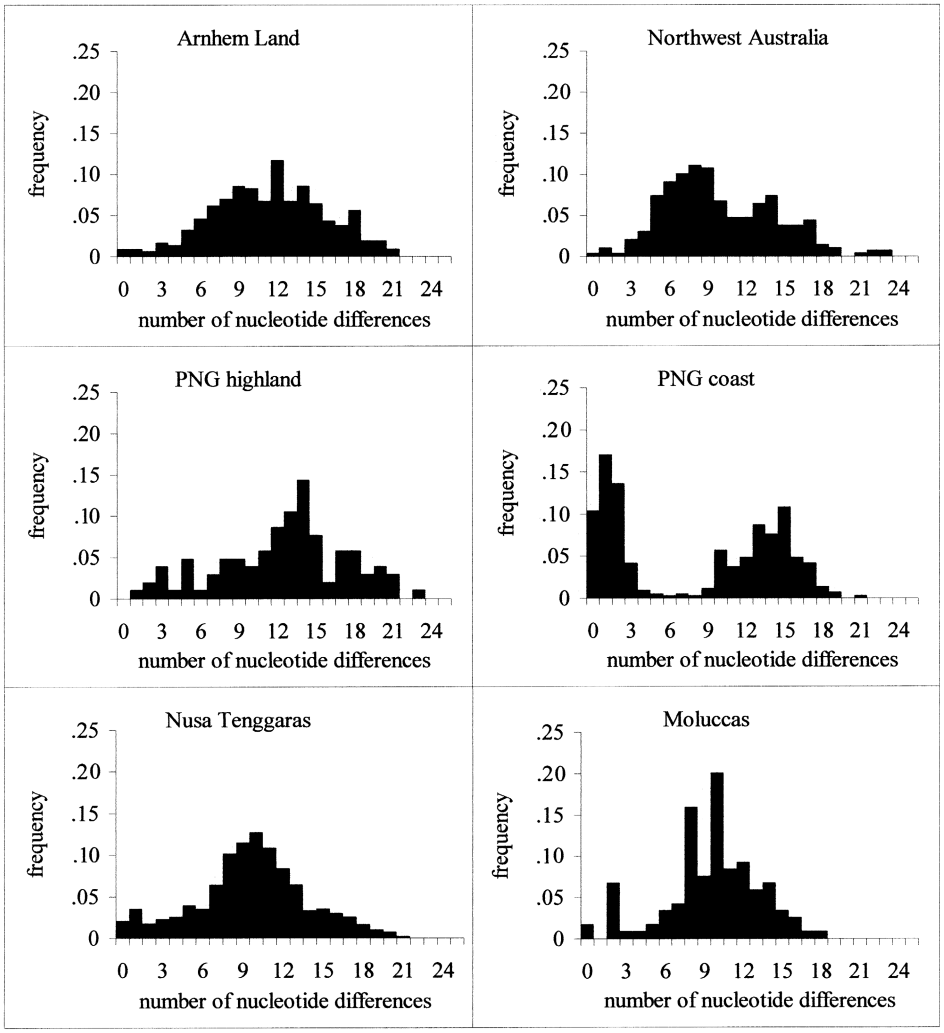


Figure 8 Mismatch distributions of CR sequences within populations from Arnhem Land, northwest Australia, PNG highland, PNG coast, Nusa Tenggaras, and the Moluccas. The numbers of nucleotide differences between all pairs of sequences are indicated along the x-axis, and the frequency of pairs is indicated along the y-axis.

lineage. In fact, when the PNG coastal sample includes only SSO types that were deduced from the sequence data, the trees and MDS plots of the SSO-type data separate highland and coastal PNG samples (data not shown).

Population-Differentiation Tests in SSO-Type Data

The permutation test of pairwise F_{ST} values between populations indicates that nearly all of the populations are differentiated from each other ($P < .05$) with the following exceptions: Moluccas versus Java ($P = .0504$), Moluccas versus Malay ($P = .44$), Java versus Malay ($P = .06$), Malay versus Chinese ($P = .29$), and Pushtoon versus Sikh ($P = .15$). The exact test of population differentiation also revealed significant differences between the majority of populations with the following exceptions: Bangladesh versus southern India ($P = .31$), Borneo versus Java ($P = .16$), Java versus Malay ($P = .36$), and Pushtoon versus Sikh ($P = .30$).

Pairwise Difference Distributions

Mismatch distributions within PNG (coast and highland), Australia (northwest Australia and Arnhem Land), and east Indonesia (Moluccas and Nusa Tenggaras) are shown in figure 8, and their summary statistics are shown in table 5. The mismatch distributions are derived from the same 672 nucleotide sites for all populations. The PNG coast mismatch distribution was clearly bimodal and displayed a very high variance. This result suggests the occurrence of two distinct migrations (Wakeley 1996) into this region, the first being the ancient migration that gave rise to Papuan speakers, and

Table 5

Mismatch Distributions and Their Summary Statistics

Population and Geographic Region	No. of Individuals	Mean	Variance	No. of Segregating Sites	Harpending's r[a]	Tajima's D[b]	τ[c]
Aboriginal Australian:	54	10.93	18.68	97	.004	−1.712[d]	8.14
Arnhem Land	28	11.42	18.70	75	.010	−1.561	8.73
Northwest Australia	25	9.91	18.62	60	.008	−1.461	6.96
PNG:	46	10.56	31.72	70	.009	−1.196	5.96
Highland	15	12.48	23.77	46	.019	−.507	9.12
Coastal	31	7.94	41.24	48	.025	−1.254	2.17
East Indonesians:	59	9.63	16.02	76	.006	−1.419	7.10
Moluccas	16	9.61	13.24	41	.061	−.934	7.70
Nusa Tenggaras	43	9.61	17.37	67	.006	−1.356	6.83

[a] Harpending's (1994) raggedness statistic.
[b] Tajima's (1989) statistic.
[c] $\tau = 2\mu t$ (in which μ = the mutation rate) Rogers (1995).
[d] $P < .05$

the second a more recent migration by Austronesian speakers. When the mismatch distribution of mtDNA with the 9-bp deletion from the PNG coast is plotted separately it corresponds to the leftmost peak (Redd et al. 1995). The Moluccan sample displays the most uneven distribution when Harpending's raggedness statistic is considered. The observed raggedness value of 0.061 in the Moluccan population is within the range of values (0.05–0.75) that suggest constant population size, according to Harpending et al. (1993), although the sample size is rather small.

Although all populations exhibit a negative value for Tajima's *D* statistic, suggesting population expansions, this value reaches statistical significance only in the Aboriginal Australian sample. If we assume that the observed values of Harpending's raggedness statistic roughly indicate population expansions, then the PNG highland mismatch distribution has the largest value for τ, 9.1 ± 2.3, whereas the Aboriginal Australian distribution (not shown but unimodal) and the east Indonesian distribution (not shown but unimodal) have slightly smaller τ values of 8.1 ± 1.8 and 7.1 ± 1.6, respectively. These values correspond to estimated expansion times of ~76,507 years for the PNG highland sample (95% CI = 55,663–97,350 years); ~68,099 years for the Aboriginal Australian sample (95% CI = 51,457–84,742 years); and 59,692 years for the east Indonesian sample (95% CI = 44,934–74,450 years); with the assumption of a substitution rate of $8.85 \times 10^{-8} \pm 0.9$/site/year (Horai et al. 1995; Bonatto et al. 1997).

Population Divergence

Intermatch distributions from comparisons between Africans and all other populations (PNG highland, PNG coastal, Aboriginal Australian, east Indonesian, southern Indian, European, Asian, and Polynesian) are all very similar (fig. 9A). In contrast, the intermatch comparisons between PNG highlanders and all other populations (fig. 9B) show that the intermatch distribution between PNG highlanders and Africans (τ = 15.1 ± 2.1) leads all other intermatch distributions. In addition, the intermatch distribution between PNG highlanders and Aboriginal Australians (τ = 12.5 ± 0.9) mirrors the intermatch distributions between PNG highlanders and the remaining non-African populations. Finally, the intermatch comparisons between Aboriginal Australians and all other populations (fig. 9C) show that the most leading intermatch distribution occurs between Aboriginal Australians and Africans (τ = 12.7 ± 2.3). The second-most leading intermatch distribution occurs between Aboriginal Australians and PNG highlanders (τ = 12.5 ± 0.9), whereas the remaining intermatch distributions between Aboriginal Australians and all remaining non-African populations are quite similar. However, the intermatch distribution between Aboriginal Australians and southern Indians is furthest to the left (τ = 7.96 ± 1.4).

Similar patterns of population divergence are found when one subtracts the ancestral polymorphism and considers only net genetic divergence (d_A) between populations (fig. 10). Estimates made on the basis of the Tamura-Nei model of substitution that include a correction for rate variation (a = 0.19) indicate that the d_A values between the PNG highlanders and all other populations is high, relative to d_A values between Aboriginal Australians and all other populations. Aboriginal Australians show recent divergence with southern Indian populations. In fact, the net divergence between PNG highlanders and Aboriginal Australian populations (d_A = 0.36% ± 0.11%) is 12 times greater than the net divergence between Aboriginal Australians and southern

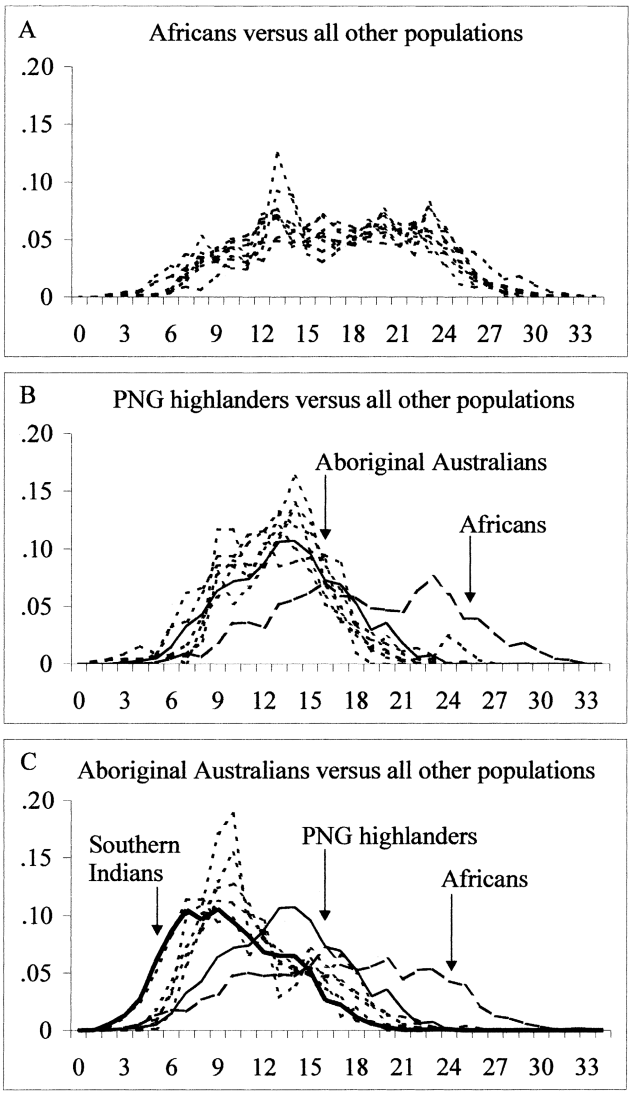


Figure 9 Intermatch distributions of CR sequences among several populations. Panel *A* plots the intermatch distributions between Africans and all other populations (European, Asian, east Indonesian, Polynesian, Aboriginal Australian, PNG highlander, PNG coastal, and southern Indian). Panel *B* plots the intermatch distributions between PNG highlanders and all other populations. Panel *C* plots the intermatch distributions between Aboriginal Australians and all other populations. The numbers of nucleotide differences, between all pairs of sequences among populations, are indicated along the x-axis, and the frequency of pairs is indicated along the y-axis.

Indians (d_A = 0.03% ± 0.03%). These estimates of net divergence correspond to a date of separation between PNG highlanders and Australian Aborginals of 40,678 years (95% CI = 34,128–47,228 years) and a date of separation of Aboriginal Australians from southern Indians of only 3,390 years (95% CI = 1,686–5,093 years).

The population divergence between the Polynesian (and PNG coastal populations) and all other populations is relatively larger, with regard to d_A distances, than with intermatch distributions and is likely the result of bottlenecks. Average-sequence variation within populations indicates that the Africans harbor the most genetic diversity, followed by PNG highlanders, whereas the Polynesians show the least genetic diversity (fig. 10).

Discussion

The analyses of the mtDNA SSO-type and CR-sequence data from this study highlight the distinctiveness of PNG highlanders in comparison with worldwide populations, including Aboriginal Australians, and suggest a possible connection between Aboriginal Australians and populations from the Indian subcontinent. The new SSO probes (IF2, IG2, and IIB4) were detected in statistically higher frequencies in highland PNG populations than in coastal PNG populations and were present in east Indonesians (3%–8%) but were nearly absent in Aboriginal Australians (2%). Examination of SSO types indicates that PNG2 and PNG3 mtDNA are most often found among PNG highlanders (84%), are much less frequent among east Indonesians (8%), and are rare or absent among Aboriginal Australians and all other Asian populations examined. This pattern is consistent with the distribution of Papuan languages in New Guinea and east Indonesia (see below). In addition, the SSO-type variation is structured within PNG, with PNG2 sequences being more frequent in populations from the eastern highlands and PNG3 sequences being more common in populations from the southern highlands. Thus, the SSO-type data provide evidence for substantial internal and external isolation in populations from the PNG highlands.

The ME gene tree constructed from CR sequences displayed very little connection between PNG highlanders and Aboriginal Australians, despite our efforts to detect common mtDNA between them by using the SSO types. East Indonesians were found in all three PNG clusters, whereas Aboriginal Australians were not found in any of the PNG clusters. The highland PNG clusters (PNG3 and PNG2) received high support from the CP test, whereas the PNG3B cluster also received high support from the bootstrap test. This finding is noteworthy because the majority of non-PNG clusters in the ME tree received very low bootstrap probabilities, meaning that these PNG clusters are unique. The PNG3 cluster was located adjacent to the large African cluster, and they share some substitutions, such as 16129A, which is absent or at very low frequency in Aboriginal Australian populations. mtDNA from the PNG2 and PNG3 clusters comprised 84% of the PNG highland sample, 54% of the PNG coast sample, and 8% of the east Indonesian sample. The coalescence times for these two clusters are

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1)	African	**312.7** (51.6)	25.3	22.0	21.1	16.2	15.9	19.8	26.0	31.1	20.7	22.0	49.2
(2)	European	68.3	**123.2** (19.1)	6.5	11.2	4.7	4.9	8.0	8.9	20.6	5.9	4.3	38.3
(3)	Havik	56.8	8.4	**150.9** (25.9)	3.3	2.7	1.6	2.1	9.9	23.8	6.9	6.5	44.2
(4)	Mukri	52.1	14.9	2.1	**165.8** (39.1)	4.0	3.4	3.8	40.6	25.4	7.9	9.3	44.4
(5)	Asian	45.6	6.0	1.8	-0.1	**228.0** (36.3)	2.7	2.2	9.5	17.2	4.0	3.4	31.8
(6)	Arnhem Land	47.5	8.6	2.4	3.3	1.4	**207.3** (35.1)	2.8	12.5	22.8	7.3	4.7	41.3
(7)	Northwest Australia	52.1	12.6	3.7	2.4	2.8	1.5	**190.8** (33.6)	9.4	20.0	6.1	7.0	38.2
(8)	PNG highland	80.4	33.2	33.4	39.6	31.5	36.9	34.8	**241.6** (51.7)	25.2	11.5	10.2	47.9
(9)	PNG coast	89.9	40.3	51.4	56.7	31.8	50.5	43.6	54.3	**147.4** (38.1)	10.0	12.4	10.8
(10)	Moluccas	54.0	13.1	13.1	13.8	1.0	14.2	12.1	31.4	7.5	**188.3** (38.7)	2.5	16.7
(11)	Nusa Tenggaras	58.8	10.1	15.6	14.1	5.9	13.8	18.0	32.0	18.1	-2.7	**190.4** (36.3)	21.0
(12)	Polynesian	136.7	75.8	95.7	98.6	66.4	90.1	82.8	116.4	3.9	34.8	46.6	**31.9** (14.8)

Figure 10 CR-nucleotide diversity within and between populations. Average-nucleotide differences are shown within (d_X along diagonal, with their SE in parentheses) and between populations (d_A below diagonal, with their SE above diagonal) (Nei and Jin 1989), with the assumption of the Tamura and Nei (1993) model of substitution with a gamma parameter of 0.19. Net differences (d_A) between Aboriginal Australian and southern Indian populations, and between Aboriginal Australian and PNG highlander populations, are boxed. All values were multiplied by 10,000.

80,226 years for PNG2 and 122,034 years for PNG3. In contrast, mtDNA from Aboriginal Australian populations were intermingled across the gene tree and were absent from the PNG2 and PNG3 clusters. These patterns suggest that the effects of genetic drift were more important in PNG populations than in Aboriginal Australian populations, and they are consistent with an early migration from Africa to PNG (Sahul), with a much greater degree of isolation in populations from the highlands than in populations from the coast and east Indonesia, as well as in populations from Australia.

The intermatch distributions also provide evidence for large evolutionary differentiation between PNG highlanders and Aboriginal Australians. The largest intermatch separation occurred between PNG highlanders and Africans, whereas another level of separation occurred between PNG highlanders and all other populations including Aboriginal Australians. The peak ($\tau = 12.5 \pm 0.9$) of the intermatch distribution between PNG highlanders and Aboriginal Australians leads the peaks of their mismatch distributions ($\tau = 9.1 \pm 2.3$ and 8.1 ± 1.8, respectively; excess = 3.87). Simulations indicate that leading intermatch distributions occur when ancestral-population separations precede daughter-population expansions and when there are very low migration rates (Harpending et al. 1993). Thus, it appears, from the intermatch distributions, that the Aboriginal Australian and southern Indian populations derive from the same ancestral population, whereas the highland PNG population derives from a completely different ancestral population.

Both the NJ tree and the MDS plot of the mtDNA-sequence data, as well as the tree and MDS plot of the SSO-type data, group the Aboriginal Australian populations with populations from the Indian subcontinent. It is noteworthy that the CR-sequence and SSO-type data derive from different Indian populations: the sequence data are from the Havik and the Mukri, whereas the SSO-type data include Southern Indian, Sri Lankan, Sikh, and Pushtoon populations. Thus, the connection between Australia and the Indian subcontinent suggested here is not dependent on specific Indian populations. In addition, an MDS plot that includes CR sequences (HVS-I only) from two additional groups of Aboriginal Australians, the Desert and the Riverine (van Holst Pellekaan et al. 1997), links all the Aboriginal Australian populations with Indian subcontinent populations, although the Riverine population clusters a bit more distantly from the others (data not shown). Furthermore, the net separation between Aboriginal Australian populations and southern Indian populations appears to be much more recent than the separation between Aboriginal Australian populations and PNG highland populations. The precision of the estimated divergence times

should be considered somewhat cautiously, since these estimates are associated with large uncertainties. However, the patterns are consistent with a separate origin (or ancient separation) for PNG highlanders and Australian Aboriginals and with recent genetic affinities between southern Indian populations and Aboriginal Australian populations.

These findings are somewhat consistent with Birdsell's trihybrid model for the peopling of Sahul, a model that is based on morphological variation (Birdsell 1967, 1993). Birdsell hypothesized that Oceanic "Negritos" first populated Sahul, but that two later migrations replaced most of them in Australia but not in the Cairns area of northeast Queensland or in Tasmania and New Guinea. According to this model, the second migration of populations, with affinities to the Ainu of Japan, dispersed throughout Australia, whereas the third migration of populations, with affinities to tribal populations of India, entered northern Australia around the Gulf of Carpentaria. The gene tree in the present study shows that the PNG3 cluster shares sites with African sequences, a finding that may be consistent with Birdsell's first-migration hypothesis. Our results also suggest that there may have been a migration(s) from an Indian source that reached Australia but not PNG. However, our results do not support two distinct source populations for the subsequent peopling of Australia, because Aboriginal Australian populations cluster together with southern Indian populations. That is, we did not detect a major distinction between populations derived from the putative second migration (northwest Australia [Desert and Riverine]) and the third migration (Arnhem Land), although our analyses did not include Ainu populations. Furthermore, our findings link Hindu caste populations from southern India, rather than exclusively tribal populations, with Aboriginal Australians. This may indicate that substantial mixing of mtDNA has occurred among southern Indian populations (Mountain et al. 1995). To summarize, our data indicate that the PNG highlanders contain distinct and divergent mtDNA, with evidence of ancient African ties, that were rare or absent in Aboriginal Australians and suggest a possible recent connection between Aboriginal Australian populations and populations from the Indian subcontinent.

Anthropologists have long suggested that Aboriginal Australians and southern Indians have affinities (Birdsell 1967; Brown 1997), although corroborative evidence has been weak (Kirk and Thorne 1976; White and O'Connell 1982). The archaeological record in Australia has been noted for its dramatic increase in the density and range of stone-tool technology, including backed blades, across large areas of the continent, and the concomitant first appearance of the dingo ~4,500 years ago (Bellwood 1989). The similarities of the backed-blade technology and dingo morphology in India and Australia have been cited as support for Indian-Australian connections (Glover and Presland 1985; Gollan 1985). A recent multivariate analysis of dingo morphology supports a relationship between Indian and Arabian wolves and Australian dingoes (Corbett 1995). However, Corbett (1995) thinks that dingoes were brought to Australia via the Austronesian expansion and that most of the primitive dogs of the Pacific were descended from southeast Asian dingoes, perhaps from Thailand.

Linguistic diversity in the Pacific does not suggest a common origin for PNG highlanders and Aboriginal Australians. New Guinean and Aboriginal Australian languages are very distinct. Papuan languages encompass one of the most diverse linguistic regions in the world, with >700 different languages spoken by <4 million speakers, nearly all in New Guinea (Wurm 1982). The evolutionary relationships among many of the Papuan languages have not been discerned, thus the term "non-Austronesian" rather than "Papuan" is often used to emphasize this fact (Foley 1986). Pockets of Papuan-speaking peoples are found on some islands in east Indonesia, such as Timor and Halmahera, and extend from eastern PNG to the Solomon Islands chain and to the Santa Cruz Islands (Bellwood 1989). This linguistic distribution is consistent with the mtDNA genetic links (PNG2 and PNG3 CR sequences) between east Indonesians (Hiri, Ternate, Flores, Alor, Roti, and Timor) and PNG.

Australian languages are confined to Australia and the western Torres Strait Islands (Dixon 1980). Today, ~150 Australian languages are spoken by Aboriginal people, although it is thought that ~200 distinct languages were spoken by Aboriginal Australian people before European contact (Dixon 1980). All Australian languages are thought to derive from the same language family. However, two linguistic divisions are recognized, namely Pama-Nyungan, which includes a relatively homogeneous group of languages that are found throughout most of Australia, and non-Pama-Nyungan, which includes a larger number of languages that exhibit more diversity and are found in Arnhmen Land and north-central Australia. A variety of hypotheses have been suggested for Australian language affiliations with other languages, including Malayo-Polynesian, Papuan, Dravidian, Indo-European, and Amerindian. Dixon (1980) says that the Dravidian connection is the only one that deserves any consideration since there are some similarities between Dravidian and Australian languages, but the evidence does not meet standard criteria for a formal genetic relationship. Linguistic relationships are difficult to trace beyond ~8,000 years of separate evolution (Nichols 1994), and, since population divergence in Sahul likely precedes this date, the separate origin or ancient divergence between Papuan- and Australian-speak-

ing populations may be unresolvable with linguistic evidence.

Studies of morphological variation in crania seem to support the common-origin hypothesis for the peopling of Sahul. One large study by Pietrusewsky (1994) includes >2,500 male samples from 53 human groups from Polynesia, Micronesia, Melanesia, Australia, southeast Asia, east Asia, and north Asia. Phylogenetic and multivariate analyses place Aboriginal Australians and New Guineans together in an "Australo-Melanesian" group. A similar suggestion of an Australian-New Guinean link is supported by other studies of cranial variation in worldwide populations (Howells 1989; Hanihara 1996).

The common-origin hypothesis is rather equivocal when genetic evidence is considered. Polymorphisms in the α-globin system have shown distinctive variation in PNG highland populations compared with northern Aboriginal Australian populations (Yenchitosomanus et al. 1985, 1986; Tsintsof et al. 1990), whereas central Australians shared similar haplotypes with PNG highland populations (Roberts-Thomson et al. 1996). β-globin polymorphisms indicated similarities with northern Aboriginal Australians and island Melanesians rather than with PNG highlanders (Chen et al. 1992). Human leukocyte antigen (HLA)-A and HLA-B genes have indicated large genetic differences between Aboriginal Australians and PNG highlanders (Serjeantson 1985), whereas more-recent DNA-based research on HLA class I and II loci support the common-origin hypothesis (Serjeantson 1989; Gao and Serjeantson 1991; Yoshida et al. 1995).

Studies of classic polymorphisms have also led to conflicting conclusions regarding Australian and New Guinean affinities. One study that included 21 populations from Australia and New Guinea for nine classic loci did not find any genetic similarities between Aboriginal Australians and New Guinea highlanders (Keats 1977). Furthermore, two studies of nuclear polymorphisms that included large numbers of populations and loci showed discrepant results. Phylogenetic analyses of classic polymorphisms (Nei and Roychoudhury 1993) in 29 loci (121 alleles) in worldwide populations placed the Aboriginal Australian (Elcho Island, northern, central, and western populations) and New Guinea highlanders (north-central, western, eastern, and central-district highlands) together in a cluster (see also Cavalli-Sforza et al. 1994), but large genetic distances exist between them, which are considerably larger than those that exist between other Asian populations. In contrast, Kirk (1989) compared European, Asian (Japanese and Chinese), African, New Guinea highland (Fore and Goroka), and Aboriginal Australian (central) populations at 72 classic loci and found that the Aboriginal Australian populations were more closely related to Asians than to New Guinea highlanders.

One study (Roychoudhury 1984) specifically addressed the hypothesis of genetic connections between tribal Indian and Aboriginal Australian populations, using 10 classic polymorphisms. The study included populations from India (Toda, Irula, and Kurumba), Sri Lanka (Veddah), Malaysia (Senoi), Australia (northern territory), and New Guinea (western). A phylogenetic tree grouped the Indian tribes closer to one another than to the Australian population, and the Australian and New Guinean populations clustered together. Nevertheless, the distance matrix indicates that the Australian population is equidistant from the Veddahs and the New Guineans. In fact, when the matrix is examined in two dimensions with principal-coordinates analysis, the Australians are closest to the Veddahs (data not shown).

Another study of a large number of polymorphsims (~75 RFLP loci) in small population samples did detect some affinities between Aboriginal Australian and PNG highlanders (Mountain and Cavalli-Sforza 1997). In a phylogenetic tree that included twelve populations from Africa, Asia (Chinese, Japanese, and Cambodian), Europe, and the Pacific (Australian Aboriginals, highland and coastal PNG, and Melanesians) most of the Aboriginal Australians grouped with PNG highlanders. However, a large number of the Aboriginal Australians had haplotypes that appeared to be the result of mixed ancestry (Mountain and Cavalli-Sforza 1997).

Finally, analyses of data from Alu-insertion polymorphisms provide evidence for a separate origin of Aboriginal Australian and PNG populations. A principal-component analysis (PCA) of four Alu loci placed PNG populations close to the ancestral root, whereas the northwest Australian population clustered with west Asian populations that included southern Indian populations (Harpending et al. 1996). Harpending and colleagues suggested that the PNG populations were involved in one of the earliest migrations from Africa. Similar results were found in a PCA of the same Alu data extended to eight loci (Stoneking et al. 1997).

Although these conflicting results leave many questions open for further research, there seems to be growing evidence (Keats 1977; Roychoudhury 1984; Yenchitosomanus et al. 1985, 1986; Kirk 1989; Stoneking et al. 1990; Tsintsof et al. 1990; Chen et al. 1992; Harpending et al. 1996; Stoneking et al. 1997; the present study) that the common-origin hypothesis of Aboriginal Australian and New Guinean highlanders needs closer examination. The studies of HLA (Serjeantson 1989; Gao and Serjeantson 1991; Yoshida et al. 1995), β-globin (Roberts-Thomson et al. 1996), classic polymorphisms (Nei and Roychoudhury 1993), and autosomal DNA (Mountain and Cavalli-Sforza 1997) that support a common origin have not included southern Indians in their comparisons (Nei and Roychoudhury 1993 included a northern Indian population in their study).

Many authors suggest that the genetic distinctions between PNG highlanders and Aboriginal Australians are partially the result of recent migrations that have altered the genetic similarity of PNG highland and Aboriginal Australian populations (Yenchitosomanus et al. 1986; Stoneking et al. 1990; Chen et al. 1992; Roberts-Thomson et al. 1996). Other studies lump Aboriginal Australians and New Guineans together in their analyses. The evidence from the present study suggests that the assumption of a common origin of Aboriginal Australians and PNG highlanders should be tested with studies that include more loci and more populations from the Pacific, Asia, and the Indian subcontinent.

Acknowledgments

We are particularly grateful to the indigenous peoples of the Pacific and Asia who participated in this study. In addition, we thank K. Bhatia, N. Kretchmer, J. Kuhl, and A. S. M. Sofro for providing blood and DNA samples. We appreciate the generous computational assistance from N. Takezaki and S. Sherry. We thank A. G. Clark, L. Excoffier, H. Harpending, S. Kumar, C. R. Rao, A. Rogers, and G. Tiwari for their very helpful suggestions and comments, and T. Melton and R. Melton for assistance with software. We thank K. Tamura for providing MEBoot version 1.0. This project was supported by an National Science Foundation grant (to M.S.).

Electronic-Database Information

Accession numbers and URLs for data in this article are as follows:

ARLEQUIN, http://anthropologie.unige.ch/arlequin (for ARLEQUIN [S. Schneider, J.-M. Kueffer, D. Roessli, L. Excoffier])

GenBank, http://www.ncbi.nlm.nih.gov/Web/Genbank/index.html (for the sequences reported in this paper [accession numbers AF176125–AF176205])

GMAES, LINTRE, SENDBS, ftp.nig.ac.jp/pub/Bio (for GMAES, LINTRE, and SENDBS software packages [N. Takezaki] via FTP)

IWAVE, kimura.anthro.utah.edu/pub (for IWAVE software package [S.T. Sherry] via FTP)

MISMATCH, anthro.utah.edu/pub/rogers (for MISMATCH software package [A.R. Rogers] via FTP)

MULTISCALE, ego.psych.mcgill.ca/pub/ramsay/multiscl (for MULTISCALE software package [J.O. Ramsay, McGill University Psychology] via FTP)

References

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

Aris-Brosou S, Excoffier L (1996) The impact of population expansion and mutation rate heterogeneity on DNA sequence polymorphism. Mol Biol Evol 13:494–504

Bellwood PS (1989) The colonization of the Pacific: some current hypotheses. In: Hill AVS, Serjeantson SW (eds) The colonization of the Pacific: a genetic trail. Oxford University Press, New York, pp 1–59

Betty DJ, Chin-Atkins AN, Croft L, Sraml M, Easteal S (1996) Multiple independent origins of the COII/tRNA^{Lys} intergenic 9-bp mtDNA deletion in Aboriginal Australians. Am J Hum Genet 58:428–433

Birdsell JB (1993) Microevolutionary patterns in Aboriginal Australia: a gradiant analysis of clines. Oxford University Press, New York, pp 22–23

——— (1967) Preliminary data on the trihybrid origin of the Australian Aborigines. Archeology and Physical Anthropology in Oceania 2:100–155

——— (1957) Some population problems involving Pleistocene man. Cold Spring Harbor Symposium on Quantitative Biology 22:47–69

Bonatto SL, Salzano FM (1997) Diversity and age of the four major mtDNA haplogroups, and their implications for the peopling of the New World. Am J Hum Genet 61: 1413–1423

Brown P (1997) Australian paleoanthropology. In: Spencer F (ed) History of physical anthropology. Vol 1. Garland, New York, pp 138–145

Cavalli-Sforza LL, Menozzi P, Piazza A (1994) The history and geography of human genes. Princeton University Press, Princeton, pp 77–79

Chen LZ, Easteal S, Board PG, Kirk RL (1992) Genetic affinities of Oceanic populations based on RFLP and haplotype analysis of genetic loci on three chromosomes. Hum Biol 64:1–15

Corbett L (1995) The dingo in Australia and Asia. Cornell University Press, New York, 9–17

Dixon RMW (1980) The languages of Australia. Cambridge University Press, New York, 1, 19–20, 234–236

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791

Foley WA (1986) The Papuan languages of New Guinea. Cambridge University Press, New York, pp 1–14

Foran DR, Hixson JE, Brown WM (1988) Comparisons of ape and human sequences that regulate mitochondrial DNA transcription and D-loop DNA synthesis. Nucleic Acids Res 16:5841–5861

Gao X, Serjeantson SW (1991) Diversity in HLA-DR4-related DR, DQ haplotypes in Australia, Oceania, and China. Hum Immunol 32:269–276

Glover IC, Presland G (1985) Microliths in Indonesian flaked stone industries. In: Misra VN, Bellwood P (eds) Recent advances in Indo-Pacific prehistory. Oxford and IBH, New Delhi, pp 185–195

Gollan K (1985) Prehistoric dogs in Australia: an Indian origin? In: Misra VN, Bellwood P (eds) Recent advances in Indo-Pacific prehistory. Oxford and IBH, New Delhi, pp 439–443

Graven L, Passarino G, Semino O, Boursot P, Santachiara-Benerecetti S, Langaney A, Excoffier L (1995) Evolutionary correlation between control region sequence and restriction

polymorphisms in the mitochondrial genome of a large Senegalese Mandenka sample. Mol Biol Evol 12:334–345
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
Hanihara T (1996) Comparison of craniofacial features of major human groups. Am J Phys Anthrop 99:389–412
Harpending H (1994) Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. Hum Biol 66:591–600
Harpending H, Relethford J, Sherry ST (1996) Methods and models for understanding human diversity. In: Boyce AJ, Mascie-Taylor CGN (eds) Molecular biology and human diversity. Cambridge University Press, New York, pp 283–299
Harpending HC, Sherry ST, Rogers AR, Stoneking M (1993) The genetic structure of ancient human populations. Curr Anthropol 34:483–496
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. Am J Hum Genet 44:504–510
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. Proc Natl Acad Sci USA 92:532–536
Howells WW (1989) Skull shapes and the map: craniometric analyses in the dispersion of modern Homo. Harvard University Press, Cambridge, 37–79
Johnson BJ, Miller GH, Fogel ML, Magee JW, Gagan MK, Chivas AR (1999) 65,000 years of vegetation change in central Australia and the Australian summer monsoon. Science 284:1150–1152
Jones R (1995) Tasmanian archaeology: establishing the sequences. Annu Rev Anthropol 24:423–446
Keats B (1977) Genetic structure of the indigenous populations in Australia and New Guinea. J Hum Evol 6:319–339
Kirk RL (1989) Population genetic studies in the Pacific: red cell antigen, serum protein, and enzyme systems. In: Hill AVS, Serjeantson SW (eds) The colonization of the Pacific: a genetic trail. Oxford University Press, New York, pp 60–119
Kirk RL, Thorne AG (eds) (1976) The origin of the Australians. Humanities, New Jersey, 277–346, 379–410
Kocher TD, Wilson AC (1991) Sequence evolution of mitochondrial DNA in humans and chimpanzees: control region and a protein-coding region. In: Osawa S, Honjo T (eds) Evolution of life: fossils, molecules and culture. Springer, New York, pp 391–413
Krings M, Geisert H, Schmitz RW, Krainitzki H, Pääbo S (1999) DNA sequence of the mitochondrial hypervariable region II from the Neandertal type specimen. Proc Natl Acad Sci USA 96:5581–5585
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
Kruskal JB (1964) Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika 29:1–27
Kumar S, Tamura K, Nei M (1993) MEGA: molecular evolutionary genetics analysis, version 1.01. The Pennsylvania State University, University Park
Mantel NA (1967) The detection of disease clustering and a generalized regression approach. Cancer Res 27:209–220
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. Am J Hum Genet 57:403–414
Miller GH, Magee JW, Johnson BJ, Fogel ML, Spooner NA, McCulloch MT, Ayliffe LK (1999) Pleistocene extinction of *Genyornis newtoni:* human impact on Australian megafauna. Science 283:205–208
Moutain JL, Cavalli-Sforza LL (1997) Multilocus genotypes, a tree of individuals, and human evolutionary history. Am J Hum Genet 61:705–718
Mountain JL, Hebert JM, Bhattacharyya S, Underhill PA, Ottolenghi C, Gadgil M, Cavalli-Sforza LL (1995) Demographic history of India and mtDNA-sequence diversity. Am J Hum Genet 56:979–992
Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York
Nei M, Jin L (1989) Variances of the average numbers of nucleotide substitutions within and between populations. Mol Biol Evol 6:290–300
Nei M, Roychoudhury AK (1993) Evolutionary relationships of human populations on a global scale. Mol Biol Evol 10: 927–943
Nichols J (1994) The spread of language around the Pacific rim. Evol Anthropol 3:206–215
O'Cornnell JF, Allen J (1998) When did humans first arrive in greater Australia and why is it important to know? Evol Anthropol 6:132–146
Piercy R, Sullivan KM, Benson N, Gill P (1993) The application of mitochondrial DNA typing to the study of white Caucasian genetic identification. Int J Legal Med 106:85–90
Pietrusewsky M (1994) Pacific-Asian relationships: a physical anthropological perspective. Oceanic Linguistics 33: 407–429
Raymond M, Rousset F (1995) An exact test for population differentiation. Evolution 49:1280–1283
Redd AJ, Takezaki N, Sherry ST, McGarvey ST, Sofro ASM, Stoneking M (1995) Evolutionary history of the COII/t-RNA^{Lys} intergenic 9 base pair deletion in human mitochondrial DNAs from the Pacific. Mol Biol Evol 12:604–615
Roberts RG, Jones R (1994) Luminescence dating of sediments: new light on the human colonisation of Australia. Australian Aboriginal Studies 2:2–17
Roberts RG, Jones R, Smith MA (1990) Thermoluminescence dating of a 50,000-year-old human occupation site in northern Australia. Nature 345:153–156
Roberts-Thomson JM, Martinson JJ, Norwich JT, Harding RM, Clegg JB, Boettcher B (1996) An ancient common origin of Aboriginal Australians and New Guinean highlanders is supported by α-globin haplotype analysis. Am J Hum Genet 58:1017–1024
Rogers AR (1995) Genetic evidence for a Pleistocene population explosion. Evolution 49:608–615
Rogers AR, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. Mol Biol Evol 9:552–569

Rohlf FJ, Sokal RR (1981) Comparing numerical taxonomic studies. Syst Zool 30:459–490

Roychoudhury AK (1984) Genetic relationship between Indian tribes and Australian Aboriginals. Hum Hered 34:314–320

Rzhetsky A, Nei M (1992) A simple method for estimating and testing minimum-evolution trees. Mol Biol Evol 9: 945–967

——— (1993) Theoretical foundation of the minimum-evolution method of phylogenetic inference. Mol Biol Evol 10: 1073–1095

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing 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. Oxford University Press, New York, pp 120–173

——— (1985) Migration and admixture in the Pacific: insights provided by human leukocyte antigens. In: Kirk RL, Szathmary E (eds) Out of Asia: peopling the Americas and the Pacific. Australian National University Press, Canberra, pp 133–145

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. Genetics 139:457–462

Stoneking M, Fontius JJ, Clifford SL, Soodyall H, Arcot SS, Saha N, Jenkins T, et al (1997) Alu insertion polymorphisms and human evolution: evidence for a larger population size in Africa. Genome Res 7:1061–1071

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. Philos Trans R Soc Lond B 337:167–175

Stoneking M, Wilson AC (1989) Mitochondrial DNA. In: Hill AVS, Serjeantson SW (eds) The colonization of the Pacific: a genetic trail. Oxford University Press, New York, pp 215–245

Sykes B, Leiboff A, Low-Beer J, Tetzner S, Richards M (1995) The origins of the Polynesians: an interpretation from mitochondrial lineage analysis. Am J Hum Genet 57: 1463–1475

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

Tajima M, Takezaki N (1994) Estimation of evolutionary distance for reconstructing molecular phylogenetic trees. Mol Biol Evol 11:278–286

Takezaki N, Rzhetsky A, Nei M (1995) Phylogenetie test of the molecular clock and linearized trees. Mol Biol Evol 12: 823–833

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10: 512–526

Thorne A, Grün R, Mortimer G, Spooner NA, Simpson JJ, McCulloch M, Taylor L, et al (1999) Australia's oldest human remains: age of the Lake Mungo 3 skeleton. J Hum Evol 36:591–612

Tsintsof AS, Hertzberg MS, Prior JF, Mickleson KNP, Trent RJ (1990) α-globin gene markers identify genetic differences between Australian Aborigines and Melanesians. Am J Hum Genet 46:138–143

van Holst Pellekaan S, Frommer M, Sved J, Boettcher B (1998) Mitochondrial control-region sequence variation in Aboriginal Australians. Am J Hum Genet 62:435–449

——— (1997) Mitochondrial D-loop diversity in Australian riverine and Australian desert Aborigines. Electrophoresis 18:1538–1543

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

Wakeley J (1996) Distinguishing migration from isolation using the variance of pairwise differences. Theoretical Population Biology 49:369–386

——— (1993) Substitution rate variation among sites in hypervariable region 1 of human mitochondrial DNA. J Mol Evol 37:613–623

White JP, O'Connell JF (1982) A prehistory of Australia, New Guinea and Sahul. Academic Press, New York, pp 76–77, 97–98

Wurm SA (1982) Papuan languages of Oceania. Gunter Narr, Tubingen, pp 1–2

Yenchitosomanus P, Summers KM, Bhatia KK, Board PG (1986) A single α-globin gene deletion in Australian Aborigines. Aust J Exp Biol Med Sci 64:297–306

Yenchitsomanus P, Summers KM, Bhatia KK, Cattani J, Board PG (1985) Extremely high frequencies of α-globin gene deletion in Madang and on Karkar Island, Papua New Guinea. Am J Hum Genet 37:778–784

Yoshida M, Ohtsuka R, Nakazawa M, Juji T, and Tokunaga K (1995) HLA-DRB1 frequencies of non-Austronesian-speaking Gidra in South New Guinea and their genetic affinities with Oceanian populations. Am J Phys Anthropol 96:177–181