Geographic Variation in Human Mitochondrial DNA from Papua New Guinea

Mark Stoneking,*[,1] Lynn B. Jorde,[†] Kuldeep Bhatia[‡] and Allan C. Wilson*

**Division of Biochemistry and Molecular Biology, University of California, Berkeley, California 94720, [†]Department of Human Genetics, University of Utah School of Medicine, Salt Lake City, Utah 84132 and [‡]Papua New Guinea Institute of Medical Research, Goroka, Eastern Highlands Province, Papua New Guinea*

Manuscript received September 16, 1989
Accepted for publication December 5, 1989

ABSTRACT

High resolution mitochondrial DNA (mtDNA) restriction maps, consisting of an average of 370 sites per mtDNA map, were constructed for 119 people from 25 localities in Papua New Guinea (PNG). Comparison of these PNG restriction maps to published maps from Australian, Caucasian, Asian and African mtDNAs reveals that PNG has the lowest amount of mtDNA variation, and that PNG mtDNA lineages originated from Southeast Asia. The statistical significance of geographic structuring of populations with respect to mtDNA was assessed by comparing observed G_{ST} values to a distribution of G_{ST} values generated by random resampling of the data. These analyses show that there is significant structuring of mtDNA variation among worldwide populations, between highland and coastal PNG populations, and even between two highland PNG populations located approximately 200 km apart. However, coastal PNG populations are essentially panmictic, despite being spread over several hundred kilometers. Highland PNG populations also have more mtDNA variability and more mtDNA types represented per founding lineage than coastal PNG populations. All of these observations are consistent with a more ancient, restricted origin of highland PNG populations, internal isolation of highland PNG populations from one another and from coastal populations, and more recent and extensive population movements through coastal PNG. An apparent linguistic effect on PNG mtDNA variation disappeared when geography was taken into account. The high resolution technique for examining mtDNA variation, coupled with extensive geographic sampling within a single defined area, leads to an enhanced understanding of the influence of geography on mtDNA variation in human populations.

MITOCHONDRIAL DNA (mtDNA) has several properties that make it a useful molecule for inferring the genetic structure and evolutionary history of human populations. First, mtDNA is readily purified and analyzed because of its high copy number and localization to a cytoplasmic organelle. Second, the strictly maternal, haploid mode of inheritance of mtDNA (GILES *et al.* 1980) has two important consequences: (1) phylogenetic trees that relate mtDNA types can be interpreted as genealogies reflecting the maternal history of a population or species (WILSON *et al.* 1985; AVISE 1986); and (2) the effective population size for mtDNA genomes is about one-fourth that for nuclear genes (BIRKY, MARUYAMA and FUERST 1983), leading to a higher rate of local differentiation of mtDNA by random genetic drift. Third, mtDNA evolves five to ten times faster than nuclear DNA, with mtDNA mutations appearing to accumulate in a clocklike manner (BROWN, GEORGE and WILSON 1979; BROWN *et al.* 1982; WILSON *et al.* 1985, 1989). This rapid evolution of mtDNA sequences means that local population differentiation will be accelerated, further enhancing the usefulness of mtDNA in studying closely related populations. Fourth, the availability of the complete sequence from one individual (ANDERSON *et al.* 1981) enables the construction of high resolution restriction maps by the sequence-comparison method: by comparing observed fragment patterns to the fragment pattern expected from the sequence, the location (and often the exact nature) of nucleotide substitutions responsible for restriction site polymorphisms can usually be precisely deduced (CANN, BROWN and WILSON 1984; STONEKING, BHATIA and WILSON 1986a). It thus becomes feasible to map hundreds of cleavage sites per mtDNA, allowing many mtDNA types (even closely related ones) to be distinguished.[2]

The publication costs of this article were partly defrayed by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

[1] Current address: Department of Human Genetics, Cetus Corporation, 1400 53rd Street, Emeryville, California 94608. After August 1, 1990: Department of Anthropology, The Pennsylvania State University, University Park, Pennsylvania 16802.

[2] The reliability of the sequence-comparison method of mapping has been questioned by SINGH, NECKELMANN and WALLACE (1987), who discovered that base substitutions in human mtDNA can occasionally produce small alterations in the electrophoretic mobility of restriction fragments in polyacrylamide gels. While such conformational mutations can lead to incorrect map assignments of restriction site polymorphisms, these errors are unlikely to alter estimates of sequence divergence, mtDNA diversity, or phylogeny, as discussed by VIGILANT, STONEKING and WILSON (1988). In particular, conformational mutations will usually not be mistaken for restriction site polymorphisms; thus the criticisms of SINGH, NECKELMANN and WALLACE (1987) have little bearing on these or previous analyses of mtDNA restriction site variation.

The above properties of mtDNA have motivated global surveys of human mtDNA variation (BROWN 1980; JOHNSON *et al.* 1983; CANN, STONEKING and WILSON 1987), resulting in new and controversial perspectives concerning the evolutionary history of humans. Chief among these is the proposal that all mtDNAs in modern human populations trace back to a single common ancestor who probably lived in Africa some 200,000 yr ago (CANN, STONEKING and WILSON 1987; WILSON *et al.* 1987; STONEKING and CANN 1989). WHITTAM *et al.* (1986) have also used the high resolution restriction maps to show that most of the mtDNA genetic diversity in human populations fits the predictions of the neutral model and that gene flow between different worldwide populations is probably limited.

The small effective population size and rapid substitution rate suggest that mtDNA should reveal new perspectives on the genetic structure and evolutionary history of human populations, not on just a global scale but also on a more localized level; with that aim in mind we have been investigating mtDNA variation in Papua New Guinea (PNG). STONEKING, BHATIA and WILSON (1986a) have previously described mtDNA variation in the Eastern Highlands of PNG and utilized estimates of sequence divergence within and between PNG-specific clusters of mtDNA types to derive an intraspecific calibration of the rate of human mtDNA evolution (STONEKING, BHATIA and WILSON 1986b). Here, we present analyses based on a widespread geographic sampling of mtDNA variation from 25 PNG localities, in which the extensive information on the biological and cultural diversity of PNG (summarized below) provides a valuable background with which to compare the mtDNA results.

The principal geographic features of PNG include a rugged, mountainous interior (the highlands) and lowland tropical rainforests and swamps (the coastal region). The highlands constitute one of the most densely populated areas, with a total population in excess of one million people. Highland villages generally consist of fewer than 300 people, located in river valleys at elevations from 1000 to 2500 m between impassable mountain ranges that reach 3000–4000 m in elevation, a situation leading to external and internal isolation of PNG populations (ALLEN 1983). Larger settlements (up to 1000 people) exist on the river systems and estuaries of the coastal region.

The single most indicative feature of this extreme isolation is the enormous linguistic diversity in PNG. Three principal language groups comprising over 1500 languages (about one third of the known languages of the world) are found in the South Pacific (FOLEY 1986): Australoid, consisting of about 200 languages confined entirely to Australia; Melanesian or Austronesian, consisting of over 600 languages distributed from Southeast Asia to Polynesia; and Papuan or non-Austronesian, consisting of over 750 distinct languages and confined almost entirely to New Guinea and adjacent islands. All of the highland and most of the coastal PNG populations speak non-Austronesian (NAN) languages, while Austronesian (AN) speakers are limited to a few coastal areas and the offshore islands.

This extraordinary linguistic situation has prompted numerous investigations into the genetic, biological and cultural diversity of PNG populations (KIRK 1980, 1982a, b; BHATIA *et al.* 1981, 1988, 1989; FROEHLICH and GILES 1981; MOURANT *et al.* 1982; WOOD *et al.* 1982; WURM 1982, 1983; HOPE, GOLSON and ALLEN 1983; SERJEANTSON, KIRK and BOOTH 1983; BHATIA, GOROGO and KOKI 1984; CRANE *et al.* 1985; OHTSUKA *et al.* 1985; SERJEANTSON 1985; WOOD, SMOUSE and LONG 1985; FOLEY 1986). The resulting wealth of information provides a valuable background with which to compare the results of our in-depth study of mtDNA variation in PNG. We report: (1) external and internal isolation of PNG populations with respect to mtDNA; (2) a probable Southeast Asian origin of PNG mtDNA lineages, with no particularly close phylogenetic relationship between Australian and PNG mtDNA lineages; and (3) an unusual relationship between mtDNA genetic distance and geographic distance separating PNG populations that probably reflects ancestral colonization routes through PNG.

MATERIALS AND METHODS

Samples: mtDNAs from 119 individuals from Papua New Guinea were purified from placental tissue by differential ultracentrifugation through cesium chloride density gradients (BROWN, GEORGE and WILSON 1979; STONEKING 1986). Purified mtDNAs were mapped with 12 restriction enzymes (APPENDIX) by the sequence-comparison method (CANN, BROWN and WILSON 1984; STONEKING, BHATIA and WILSON 1986a). The geographic origin, language group, mtDNA type and mtDNA clan (defined below) of each individual from Papua New Guinea are presented in Table 1; a map of the geographic locations is in Figure 1. Restriction maps for 55 of these PNG mtDNAs were previously published by STONEKING, BHATIA and WILSON (1986b) and CANN, STONEKING and WILSON (1987). Additional mtDNA restriction maps included in some analyses came from four other populations (African, Asian, Caucasian and Australian) and have been described by CANN, STONEKING and WILSON (1987).

mtDNA genetic variation within populations: For each pair of mtDNA restriction maps, maximum-likelihood estimates of the amount of sequence divergence (number of nucleotide substitutions per site) were obtained by the method of NEI and TAJIMA (1983). This method uses the ratio of shared sites to total sites for two restriction maps, and the average length of the restriction enzyme recognition sequences, to compute an initial estimate of the probability that the two mtDNAs differ at a given nucleotide position. Starting with this initial estimate, equation 28 of NEI and TAJIMA (1983) was solved iteratively and the sequence divergence calculated from equation 21. A computer program supplied by J. C. STEPHENS was modified to perform these calculations.

TABLE 1

Origin, language group, mtDNA type, and mtDNA clan of 119 Papua New Guineans

Individual	Origin	Language group	mtDNA type	mtDNA clan	Individual	Origin	Language group	mtDNA type	mtDNA clan
1	1	NAN	152	18	61	2	NAN	27	5
2	1	NAN	89	10	62	2	NAN	98	10
3	1	NAN	39	5	63	13	AN	96	10
4	1	NAN	26	4	64	3	NAN	27	5
5	1	NAN	101	10	65	24	AN	111	10
6	1	NAN	105	10	66	24	AN	130	16
7	1	NAN	89	10	67	24	AN	11	1
8	1	NAN	27	5	68	24	AN	130	16
9	1	NAN	109	10	69	18	—	69	9
10	1	NAN	27	5	70	18	AN	119	11
11	1	NAN	100	10	71	18	AN	129	16
12	1	NAN	65	8	72	18	AN	93	10
13	1	NAN	27	5	73	18	AN	69	9
14	1	NAN	99	10	74	18	AN	119	11
15	1	NAN	107	10	75	21	NAN	119	11
16	1	NAN	24	2	76	21	NAN	106	10
17	1	NAN	34	5	77	21	NAN	27	5
18	1	NAN	107	10	78	21	NAN	43	6
19	1	NAN	103	10	79	20	NAN	68	9
20	1	NAN	107	10	80	20	NAN	130	16
21	1	NAN	107	10	81	21	NAN	43	6
22	1	NAN	107	10	82	18	NAN	110	10
23	1	NAN	89	10	83	15	NAN	130	16
24	1	NAN	107	10	84	17	AN	119	11
25	1	NAN	114	10	85	19	AN	67	9
26	2	NAN	40	5	86	17	AN	119	11
27	2	NAN	37	5	87	16	AN	27	5
28	2	NAN	108	10	88	17	AN	127	15
29	2	NAN	27	5	89	19	NAN	130	16
30	2	NAN	115	10	90	17	AN	119	11
31	2	NAN	115	10	91	16	AN	119	11
32	2	NAN	27	5	92	16	AN	119	11
33	2	NAN	102	10	93	17	AN	150	17
34	2	NAN	45	7	94	19	NAN	130	16
35	2	NAN	113	10	95	7	AN	33	5
36	2	NAN	116	10	96	22	AN	120	12
37	2	NAN	27	5	97	25	AN	94	10
38	2	NAN	26	4	98	12	AN	32	5
39	2	NAN	27	5	99	6	AN	27	5
40	2	NAN	116	10	100	23	AN	35	5
41	2	NAN	27	5	101	4	AN	130	16
42	2	NAN	46	7	102	5	NAN	27	5
43	2	NAN	97	10	103	6	—	31	5
44	2	NAN	46	7	104	6	—	95	10
45	2	NAN	115	10	105	11	AN	119	11
46	2	NAN	29	5	106	13	AN	92	10
47	2	NAN	27	5	107	9	NAN	27	5
48	2	NAN	27	5	108	6	AN	26	4
49	2	NAN	112	10	109	10	AN	130	16
50	2	NAN	27	5	110	9	NAN	25	3
51	2	NAN	115	10	111	5	NAN	104	10
52	2	NAN	88	10	112	9	AN	119	11
53	2	NAN	41	5	113	6	NAN	30	5
54	2	NAN	38	5	114	3	NAN	27	5
55	2	NAN	27	5	115	13	AN	114	10
56	2	NAN	27	5	116	6	—	36	5
57	2	NAN	28	5	117	8	AN	121	13
58	2	NAN	42	5	118	5	NAN	27	5
59	2	NAN	90	10	119	14	AN	122	14
60	2	NAN	91	10					

Origin numbers correspond to the map in Figure 1. Language groups are Austronesian (AN), non-Austronesian (NAN), or unknown (—). mtDNA types were determined by restriction site polymorphisms (APPENDIX) and mtDNA clans were determined by phylogenetic analysis as described in the text; mtDNA type numbers and clan numbers correspond to those in Figure 3.

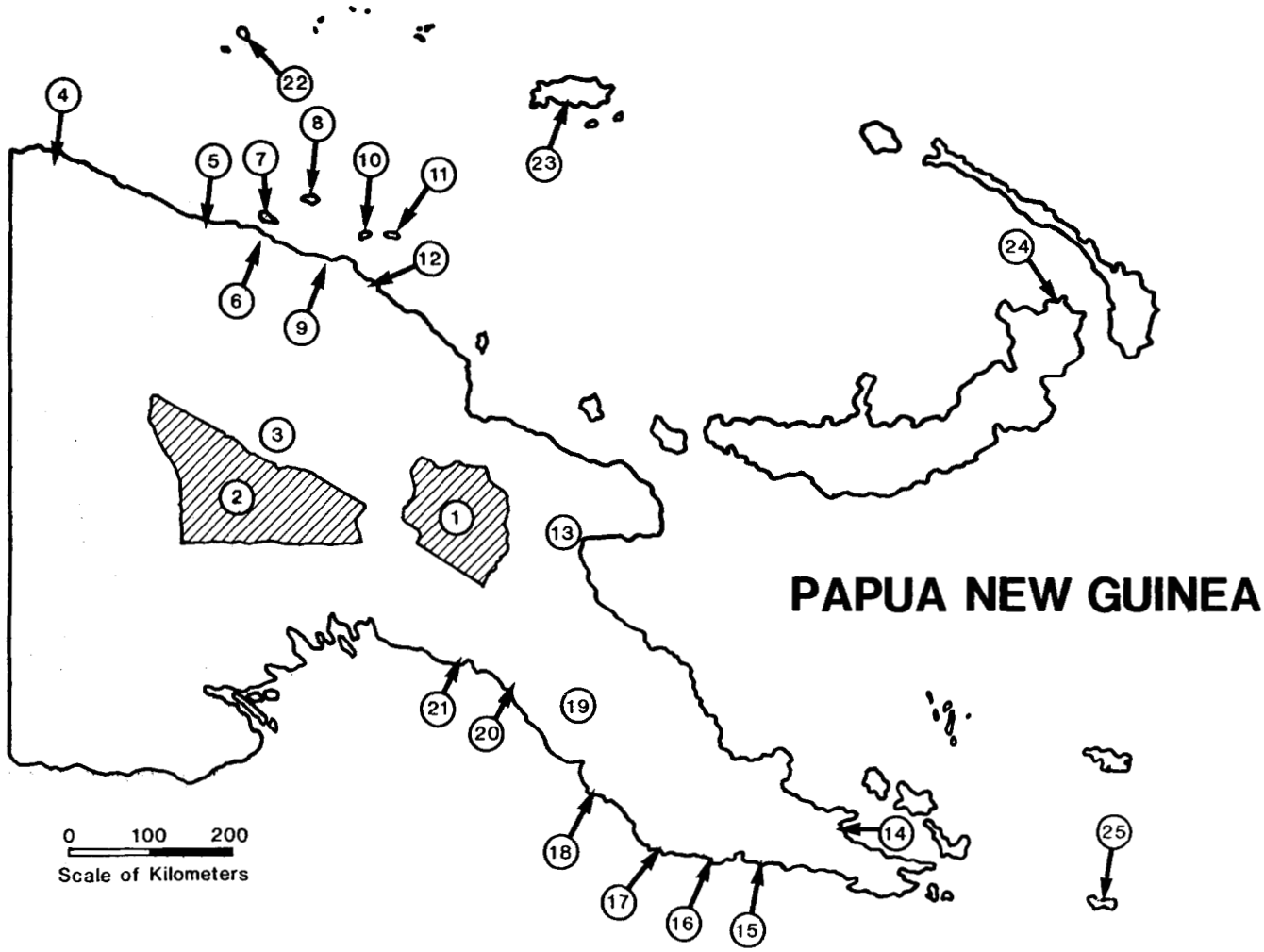

FIGURE 1.—Map of Papua New Guinea showing the sampling localities included in this study. Circled numbers are keyed to Table 1. Shaded areas indicate that many villages were sampled within each region.

Estimates of mean "heterozygosity" and the associated variance were obtained using equations 11 and 21 of ENGELS (1981). While there are no heterozygotes for mtDNA restriction site polymorphisms, ENGELS' definition of mean heterozygosity is nonetheless a meaningful and useful measure of mtDNA variability, for the following reasons: (1) it corrects for an ascertainment bias that arises when using restriction enzymes to measure variability; (2) it is free of assumptions concerning the evolutionary process or the genetic structure of the population; and (3) the variance of the estimate of heterozygosity can be easily determined, allowing evaluation of the statistical significance of differences in levels of heterozygosity between populations. By contrast, the variance of the mean sequence divergence for a population is extremely difficult to calculate, because the estimate of mean sequence divergence is obtained by averaging all pairwise comparisons between individuals and hence these observations are not statistically independent (TAKAHATA and NEI 1985; M. NEI, personal communication).

mtDNA genetic variation between populations: The proportion of the total mtDNA variation attributable to between-population differences was measured using the G_{ST} statistic of NEI (1973). G_{ST} is a convenient measure of population differentiation because it reaches equilibrium quickly (CROW and AOKI 1984) and is relatively insensitive to the number of population subdivisions (TAKAHATA and PALUMBI 1985). G_{ST} values based on mtDNA types were calculated according to equations 7.39, 7.40, and 8.27 of NEI (1987) by treating the entire mtDNA molecule as a single locus, with the number of mtDNA types equivalent to the number of alleles. G_{ST} values based on mtDNA restriction maps, treating each restriction site as a separate locus with two alleles (presence or absence of the site), were calculated according to TAKAHATA and PALUMBI (1985); their approach is designed specifically for mtDNA restriction site polymorphisms and takes into account the ascertainment bias noted by ENGELS (1981).

Since the variance of the G_{ST} statistic is unknown, the approximate statistical significance of observed G_{ST} values was estimated by a variation of the bootstrap procedure. Individuals were sampled without replacement and randomly assigned to populations such that equivalent sample sizes were maintained; this procedure was then repeated 1000 times to generate a distribution of G_{ST} values, permitting an *ad hoc* assessment of the statistical significance of an observed G_{ST} value (see RESULTS).

Three different measures of mtDNA genetic distance between populations were used. A measure of genetic distance based on mean sequence divergence (NEI and TAJIMA 1983) is

$$D_{NT} = d_{xy} - (d_x + d_y)/2 \quad (1)$$

where d_x and d_y are the mean sequence divergence within

TABLE 2

mtDNA variation in five human populations

Population	Sample size	Polymorphic sites	mtDNA types	Sequence divergence (%)	Heterozygosity (± 1 SD)
African	20	71	20	0.47	0.47 ± 0.03
Asian	34	78	34	0.35	0.35 ± 0.02
Caucasian	47	80	43	0.23	0.27 ± 0.02
Australian	21	48	20	0.25	0.27 ± 0.02
PNG	119	85	65	0.21	0.22 ± 0.01
Worldwide	241	230	182	0.27	0.32 ± 0.01

populations x and y, d_{xy} is the observed mean sequence divergence between these two populations, and D_{NT} is thus mean sequence divergence corrected for intrapopulation variation (*cf.* WILSON *et al.* 1985). A measure of genetic distance based on the probability of allelic identity of mtDNA restriction sites (TAKAHATA and PALUMBI 1985) is

$$D_{TP} = -\log_e(J/I) \quad (2)$$

where J is the probability of allelic identity between two populations and I is the geometric mean of the probability of allelic identity within each population. The third measure of genetic distance is derived from the genetic covariance or coefficient of kinship (HARPENDING and JENKINS 1973), defined as

$$r_{ij} = (P_i - P)(P_j - P)/P(1 - P)$$

where P_i and P_j are the frequencies of a particular restriction site polymorphism in populations i and j and P is the mean frequency of the polymorphism in the combined population. The r_{ij} values are averaged over all polymorphic sites, resulting in the following measure of genetic distance:

$$D_{IJ} = r_{ii} + r_{jj} - 2r_{ij} \quad (3)$$

where r_{ii} and r_{jj} are the genetic variances within populations i and j and r_{ij} is the genetic covariance between populations i and j.

Correlation between distance measures: The statistical significance of correlation coefficients derived from comparing various distance measures (genetic, geographic and linguistic) cannot be easily assessed, because the observations consist of all possible pairwise comparisons of the set of populations and hence are not independent. We therefore used the MANTEL (1967) test of matrix correspondence (SMOUSE, LONG and SOKAL 1986) to determine the statistical significance of correlation coefficients involving distance matrices. To assess the significance of a correlation coefficient derived by comparing two distance matrices, Monte Carlo randomization is used to construct a null distribution by holding one matrix constant while randomly permuting the rows and columns of the other matrix. The statistical significance is then determined by comparing the observed correlation coefficient to the randomly generated distribution of correlation coefficients.

Phylogenetic analysis: Maximum-parsimony trees depicting the genealogical relationships of the mtDNA types defined by restriction site polymorphisms were constructed using the computer program PAUP (SWOFFORD 1985). The network produced by PAUP was rooted by the midpoint method and hypothetical ancestral nodes were positioned to reflect the average sequence divergence (NEI and TAJIMA 1983) between the descendant mtDNA types.

Maximum-likelihood trees relating populations were generated using the CONTML program in the computer package PHYLIP (FELSENSTEIN 1987). Fitch-Margoliash trees were computed from the three measures of genetic distance described above by the PHYLIP program FITCH.

RESULTS

Restriction site polymorphisms: The 12 restriction enzymes detect an average of nearly 370 sites per human mtDNA, which represents approximately 9% of the human mtDNA genome. There were 85 polymorphic sites in the sample of 119 PNG mtDNAs and these defined 65 mtDNA types (APPENDIX), where an mtDNA type consists of a distinct combination of polymorphic sites. The mtDNA type of each New Guinean appears in Table 1. No individual appeared to possess more than one mtDNA type, and none of these 65 PNG mtDNA types has been found in any other survey of human mtDNA variation that utilized the same (or largely the same) set of restriction enzymes (HORAI and MATSUNAGA 1986; CANN, STONEKING and WILSON 1987).

mtDNA variation in worldwide populations: The sample size, number of polymorphic restriction sites, number of mtDNA types, mean sequence divergence, and mean heterozygosity are given in Table 2 for each of the five populations and for the total sample. In general the estimates for mean heterozygosity are nearly the same as (although somewhat larger than) the mean sequence divergence estimates, in accordance with theoretical expectations if each restriction site polymorphism is due to change at a single nucleotide position (ENGELS 1981). Using the standard errors of the mean heterozygosities to construct approximate 95% confidence intervals, we find that: (1) Africans are significantly more variable than any other population; (2) Asians are significantly more variable than the remaining populations; and (3) Caucasians and Australians do not differ from each other but are significantly more variable than the PNG population. PNG is thus characterized by the lowest levels of mtDNA variability.

A G_{ST} value of 0.31 was calculated from the average probability of allelic identity of mtDNA restriction maps (TAKAHATA and PALUMBI 1985) within and between these five populations (Table 3). This value means that roughly 30% of the total variance in

TABLE 3

G_{ST} values for human mtDNA, removing each population in turn

Population removed	P_I	P_J	H_0	H_T	G_{ST}
None	0.895	0.836	0.105	0.152	0.31
African	0.898	0.838	0.102	0.147	0.31
Asian	0.893	0.828	0.107	0.156	0.31
Caucasian	0.897	0.836	0.103	0.149	0.31
Australian	0.888	0.827	0.112	0.158	0.29
PNG	0.900	0.850	0.100	0.137	0.27

P_I is the average probability of allelic identity within each population and P_J is the average probability of allelic identity between each pair of populations; H_0 and H_T refer to the average variability within individual populations and within the total population, respectively. All values were calculated according to the equations in TAKAHATA and PALUMBI (1985).

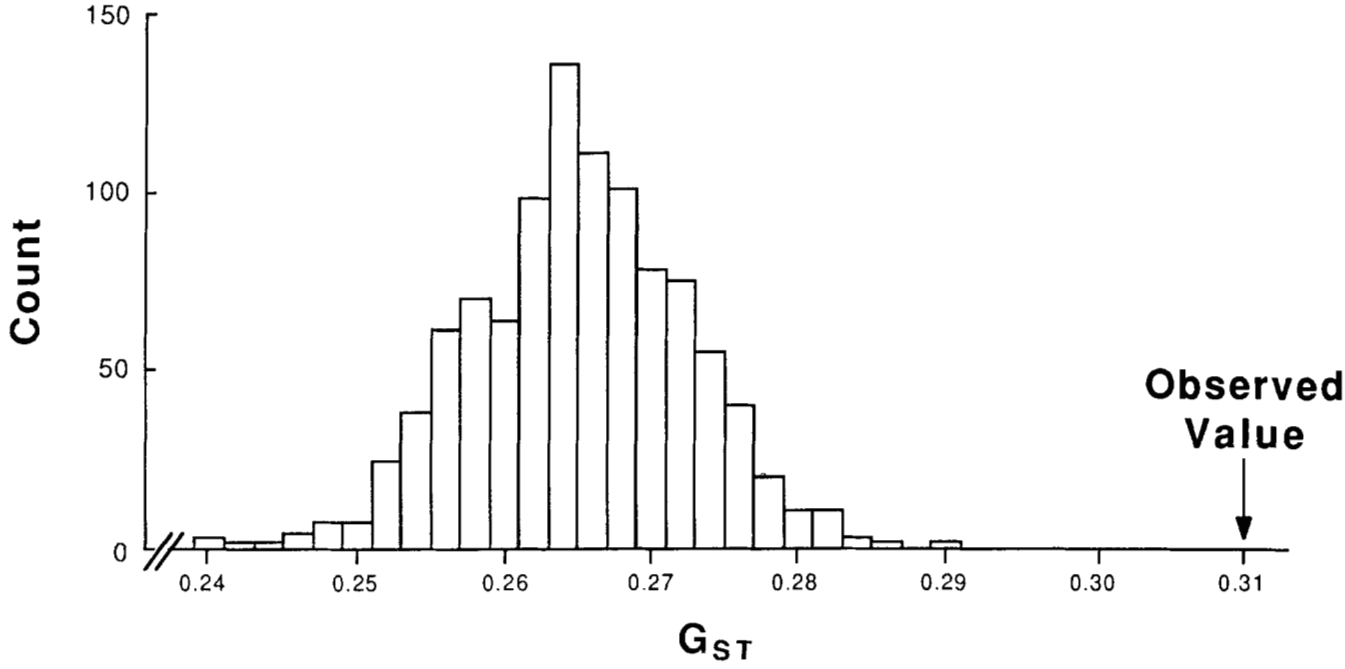

FIGURE 2.—Histogram of G_{ST} values obtained by random resampling of the mtDNA data. Individuals were assigned at random without replacement to five groups with sample sizes identical to the five worldwide populations in the study. G_{ST} values were calculated by the method of TAKAHATA and PALUMBI (1985) and the above procedure was repeated 1000 times to generate the distribution of G_{ST} values. The observed value, 0.31, was greater than any of the randomly-generated values and hence indicates geographic structuring of mtDNA in worldwide populations at an approximate significance level of $P < 0.001$.

mtDNA restriction maps distinguishes human populations and about 70% occurs within populations.

Each population was removed in turn from the analysis and the G_{ST} value recalculated, to investigate the relative contribution of each population to the G_{ST} value (Table 3). Removing Africans, Asians or Caucasians did not alter the G_{ST} value, whereas removing Australians decreased it by 6.5% and removing PNG decreased it by 13%. Thus, PNG and, to a lesser extent, Australia are isolated with respect to the other populations.

To assess the statistical significance of the observed G_{ST} value, individuals from the worldwide sample of 241 mtDNAs were assigned at random without replacement to five populations with sample sizes corresponding to the five surveyed populations and the G_{ST} value was calculated; this procedure was repeated 1000 times, generating a distribution of G_{ST} values (Figure 2). The mean of these randomly-generated G_{ST} values was 0.26 and none was as large as the observed value of 0.31. Since by this method the probability of observing a G_{ST} value as high as 0.31 is less than one in 1000, we conclude that there is statistically significant geographic structuring of mtDNA restriction site polymorphisms with respect to these five populations.

Further evidence for the significant geographic structure of human populations based on mtDNA comes from the distribution of mtDNA types. Individuals from different populations always had different mtDNA types, and individuals with identical mtDNA types always came from the same population.

Tree of mtDNA types: A phylogenetic tree relating

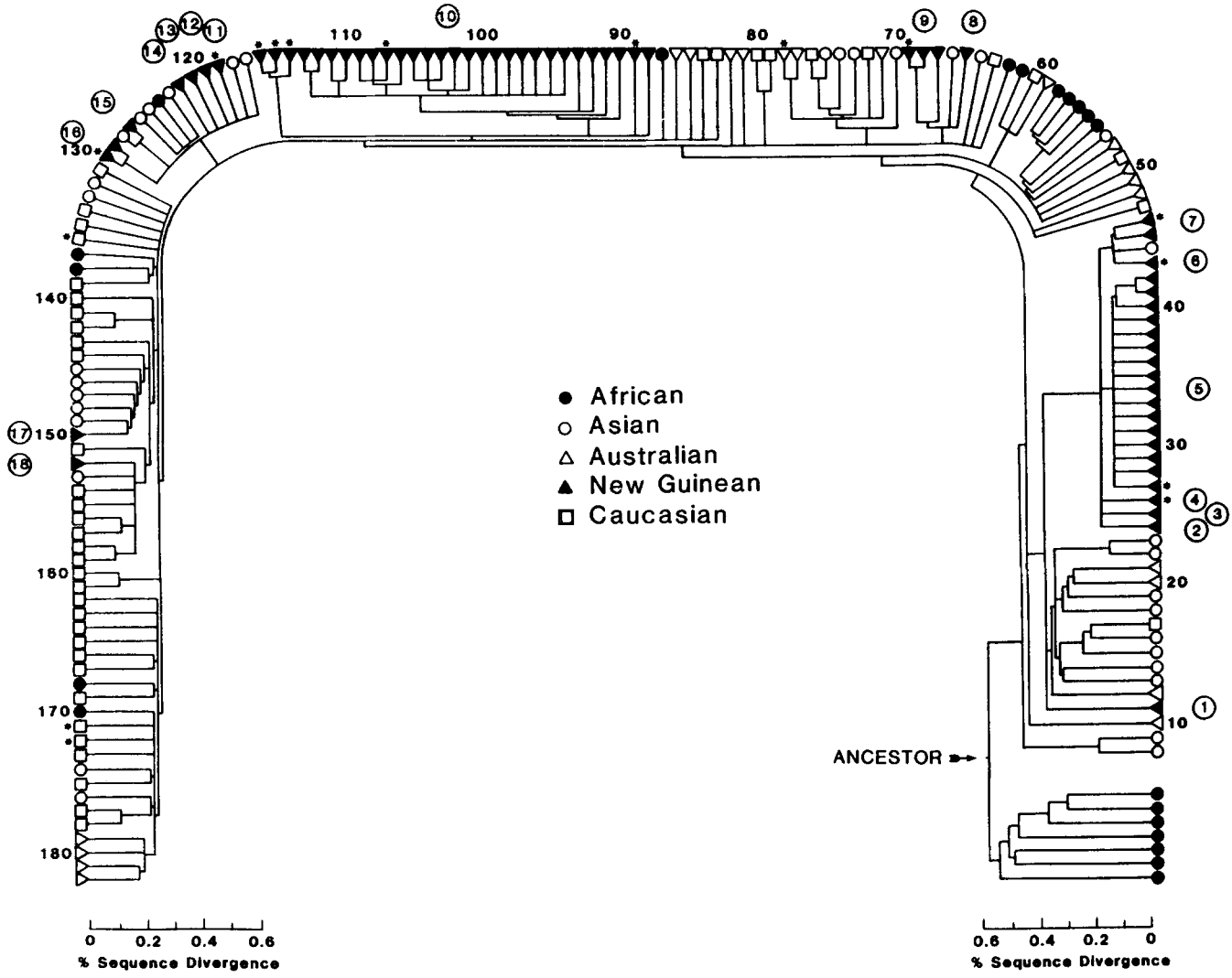

FIGURE 3.—Phylogenetic tree relating 182 mtDNA types found in 241 individuals from five worldwide populations. The branching order was obtained from 109 phylogenetically informative restriction site polymorphisms using the PAUP computer program (SWOFFORD 1985). This parsimony tree, which was rooted by the midpoint method, has a length of 363 changes at these informative sites and a consistency index of 0.3. Although no shorter trees were found, for a data set of this size it is impossible to guarantee that shorter trees do not exist, and the number of trees of equal (or nearly equal) length is probably large. Inferred ancestral nodes are positioned approximately with respect to the scale of sequence divergence. Asterisks indicate mtDNA types found in more than one individual, while circled numbers identify the 18 New Guinean mtDNA clans.

the 182 mtDNA types found in our worldwide sample of 241 humans appears in Figure 3. This tree, like its predecessors that were based on fewer individuals (STONEKING, BHATIA and WILSON 1986b; CANN, STONEKING and WILSON 1987; WRISCHNIK *et al.* 1987), possesses two main features. First, the tree contains two primary branches, one consisting solely of African mtDNAs and the other consisting of all of the other mtDNAs, including some Africans. This is one line of evidence that led to the hypothesis of an African origin for human mtDNA (CANN, STONEKING and WILSON 1987; WILSON *et al.* 1987; STONEKING and CANN 1989).

The second feature of the tree in Figure 3 is that mtDNAs from each population do not form a single clade but rather appear scattered across the tree. For example, PNG mtDNAs appear at 18 separate locations in the tree. We have used the term "clan" to refer to a group of mtDNA types, all from one geographic region, whose nearest relative is from a different geographic region (WILSON *et al.* 1987). As indicated in Figure 3, there are 18 PNG mtDNA clans. We further assume that each clan represents the descendants of a different mtDNA lineage that colonized a particular geographic region; the number of mtDNA clans is thus an estimate of the minimum number of females that colonized that region. For example, the above clan analysis indicates that a minimum of 18 females colonized PNG. We have used this concept of an mtDNA clan to derive estimates of the rate of human mtDNA evolution (STONEKING, BHATIA and WILSON 1986b; WILSON *et al.* 1987;

STONEKING and CANN 1989) and the rate of dispersal of human populations (STONEKING and WILSON 1989); other insights that derive from the concept of mtDNA clans are discussed below.

Geographic origin of founding lineages: The phylogenetic tree of mtDNA types (Figure 3) can be used to examine the geographic origin of the mtDNA lineages that colonized PNG. The geographic distribution of the nearest relative(s) of the 18 PNG clans is shown in Figure 4. For every clan the most closely related mtDNA type(s) from outside PNG are either exclusively Asian or include Asian mtDNA types. By contrast, Australian, African, and Caucasian mtDNA types are the nearest relative (together with Asian mtDNA types) of only two, six and three PNG clans, respectively. A χ^2 analysis reveals that this distribution of nearest relatives differs significantly from the expected distribution based on the sample sizes of these populations ($P < 0.001$). This statistical support for an Asian origin of PNG clans is retained even when the 18 PNG clans in Figure 3 are grouped into larger aggregations of closely related clans (data not shown). Furthermore, five of the six cases in which an African mtDNA type is a nearest relative of a PNG clan involve type 124 (Figure 3). This African-American mtDNA type carries an Asian-specific deletion (WRISCHNIK *et al.* 1987) and is thus probably of maternal Amerindian origin, since the deletion occurs in Amerindians (M. STONEKING, S. PÄÄBO, L. VIGILANT, C. E. BECK, C. ORREGO and A. C. WILSON, unpublished results). Asia is thus overwhelmingly indicated as the source of PNG mtDNA lineages.

mtDNA variation in PNG populations: The 119 PNG individuals came from 25 different localities (Figure 1), which were grouped into the following five geographic subdivisions or "populations" in order to increase sample sizes to statistically meaningful levels yet still retain geographically relevant divisions: Eastern Highlands (EH), locality 1; Southern Highlands (SH), localities 2 and 3; North Coast (NC), localities 4–6, 9, 12 and 13; South Coast (SC), localities 15–21; and offshore islands (IS), localities 7, 8, 10, 11, 14 and 22–25. The sample sizes, number of mtDNA types and clans, ratio of mtDNA types to clans, and mean heterozygosity for these five PNG populations are given in Table 4 (estimates of mean sequence divergence were identical or nearly so to the estimates of mean heterozygosity and therefore are not given). The two highland populations are the most variable, with the NC next most variable and the SC and IS populations having the least variability. Dividing the total sample into a Highland population (consisting of the EH and SH groups) and a Coast population (consisting of the NC, SC and IS groups) reveals that the Highland population is significantly more variable than the Coast population.

The greater amount of mtDNA variability in the Highlands is not simply due to more mtDNA types being present in the Highlands, as the ratio of sample size to mtDNA types is essentially identical in the Highland and Coast populations (Table 4). However, the ratio of mtDNA types to mtDNA clans is over twice as high in the Highland population as in the Coast population, because the Highland population has only half as many mtDNA clans represented. On average, fewer founding maternal lineages are present in the Highlands, but each founding lineage has differentiated into many more types than is true for the Coast population. This is consistent with an earlier, more restricted origin of the Highland population (see DISCUSSION).

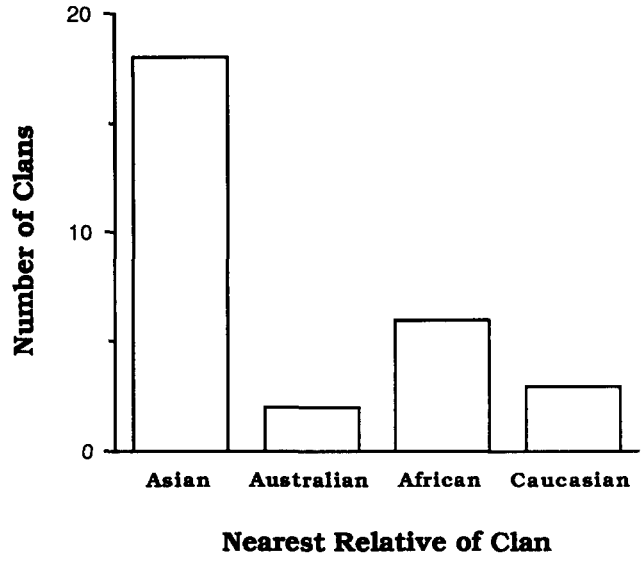

FIGURE 4.—Histogram of the number of occurrences of each geographic population as a nearest relative of a PNG mtDNA clan.

mtDNA differentiation between PNG populations: The G_{ST} value for the five PNG populations is 0.28 for mtDNA sites and 0.08 for types (Table 5), indicating that about 72% of the total variance in mtDNA restriction maps and 92% of the total variance in mtDNA types in PNG occurs within these five populations. Random resampling of the data reveals that these G_{ST} values are statistically significant ($P < 0.001$), indicating that there is geographic structure in PNG populations with respect to mtDNA. Removal of each population in turn from the G_{ST} analysis does not reveal any significant patterns (data not shown). However, G_{ST} values for the Highland-Coast comparison and for the EH-SH comparison are also statistically significant, but the G_{ST} values for a comparison of the three coastal populations are not (Table 5). Thus while the highland and coastal populations, and even the two highland populations, are strongly differentiated with respect to mtDNA, the three coastal populations essentially comprise a single panmictic population.

Three measures of mtDNA genetic distance between each pair of PNG populations are in Table 6;

TABLE 4

mtDNA variation in PNG populations and language groups

Population	(*a*) Sample size	Heterozygosity (± 1 SD)	(*b*) mtDNA types	(*c*) mtDNA clans	Ratio *a/b*	Ratio *b/c*
EH	25	0.24 ± 0.02	16	6	1.6	2.7
SH	39	0.21 ± 0.01	22	4	1.8	5.5
SC	26	0.17 ± 0.02	13	8	2.0	1.6
NC	17	0.19 ± 0.02	14	6	1.2	2.3
IS	12	0.17 ± 0.02	10	8	1.2	1.2
Highland	64	0.24 ± 0.01	36	7	1.8	5.1
Coast	55	0.19 ± 0.01	32	14	1.7	2.3
NAN	81	0.23 ± 0.01	45	12	1.8	3.7
AN	34	0.15 ± 0.01	22	12	1.5	1.8

TABLE 5

G_{ST} values for PNG populations and language groups

Comparison	mtDNA sites		mtDNA types	
	Random	Observed	Random	Observed
EH-SH-NC-SC-IS	0.24	0.28***	0.04	0.08***
Highland-Coast	0.16	0.21***	0.01	0.03***
EH-SH	0.14	0.19**	0.02	0.04**
IS-SC-NC	0.20	0.21	0.04	0.06
NAN-AN	0.16	0.23***	0.01	0.04***
Highland-Coast (NAN only)	0.16	0.22***	0.02	0.03*
NAN-AN (Coast only)	0.15	0.14	0.02	0.05

G_{ST} values were calculated for mtDNA restriction site polymorphisms and mtDNA types as discussed in the text. Random G_{ST} values represent the mean of 1000 random resamplings of the data for each comparison. *, ** and *** indicate that fewer than 50, 10 and 1, respectively, of the 1000 randomly generated G_{ST} values exceeded the observed value, and thus roughly correspond to $P < 0.05$, $P < 0.01$ and $P < 0.001$.

branching diagrams constructed from these genetic distances are in Figure 5. While these branching diagrams will be referred to as trees, it should be emphasized that they are not true phylogenies (*i.e.*, indicating only shared ancestry). Rather, they serve as useful illustrations of overall similarity between PNG populations, with such similarity reflecting both shared ancestry and migration.

The trees in Figure 5 agree in placing the IS and SC populations together but they differ in the placement of the NC population. The NC population groups with the coastal populations in the tree based on TAKAHATA-PALUMBI (D_{TP}) distances and with the highland populations in the other trees, actually grouping more closely with the SH population than does the EH population in the trees based on NEI-TAJIMA (D_{NT}) and HARPENDING-JENKINS (D_{HJ}) distances. This discrepancy may reflect differences in the assumptions underlying each distance measure, as discussed below.

mtDNA variation and geographic distance: Having demonstrated significant differences in the distribution of mtDNA polymorphisms and types between geographic areas of PNG, we can investigate the relationship between geographic distance separating PNG populations and mtDNA genetic distance. A plot of genetic distance versus geographic distance between each pair of PNG populations appears in Figure 6. The correlation for a linear relationship between genetic distance and geographic distance was not statistically significant by the MANTEL matrix regression approach ($R^2 = 0.006$, $P = 0.86$); however, the correlation for a quadratic model that included the square of geographic distance was significant ($R^2 = 0.663, P < 0.01$). Correlations for linear and quadratic models that included linguistic distance either alone or in conjunction with geographic distance were not significant (data not shown).

Inspection of the plot in Figure 6 reveals the reason for this significant quadratic relationship, namely populations that are either close together or far apart have a relatively low genetic distance and populations intermediate in geographic distance have a relatively high genetic distance. In particular, the two populations with the smallest genetic distances overall (Table 6), the SC and IS populations, are also separated by the largest geographic distance. The other population pairs with relatively low genetic distances include the three comparisons involving the EH, SH and NC populations. Although the number of comparisons is small, this quadratic relationship between genetic and geographic distance may actually reflect the historical settlement pattern of PNG (see DISCUSSION).

Distribution of identical mtDNA types: Twelve of

TABLE 6

Genetic and geographic distances between PNG populations

Populations	D_{NT}	D_{TP}	D_{HJ}	Geographic distance (km)
EH-SH	0.00019	0.03049	0.15952	220
EH-NC	0.00012	0.02552	0.13427	240
EH-SC	0.00032	0.02249	0.19152	360
EH-IS	0.00023	0.02480	0.20556	440
SH-NC	−0.00003	0.03126	0.10267	200
SH-SC	0.00033	0.03408	0.18076	560
SH-IS	0.00027	0.02881	0.17392	360
NC-SC	0.00022	0.01826	0.15153	620
NC-IS	0.00016	0.01540	0.14873	160
SC-IS	−0.00004	0.01326	0.11654	800

D_{NT}, D_{TP} and D_{HJ} were calculated according to Equations 1, 2, and 3, respectively.

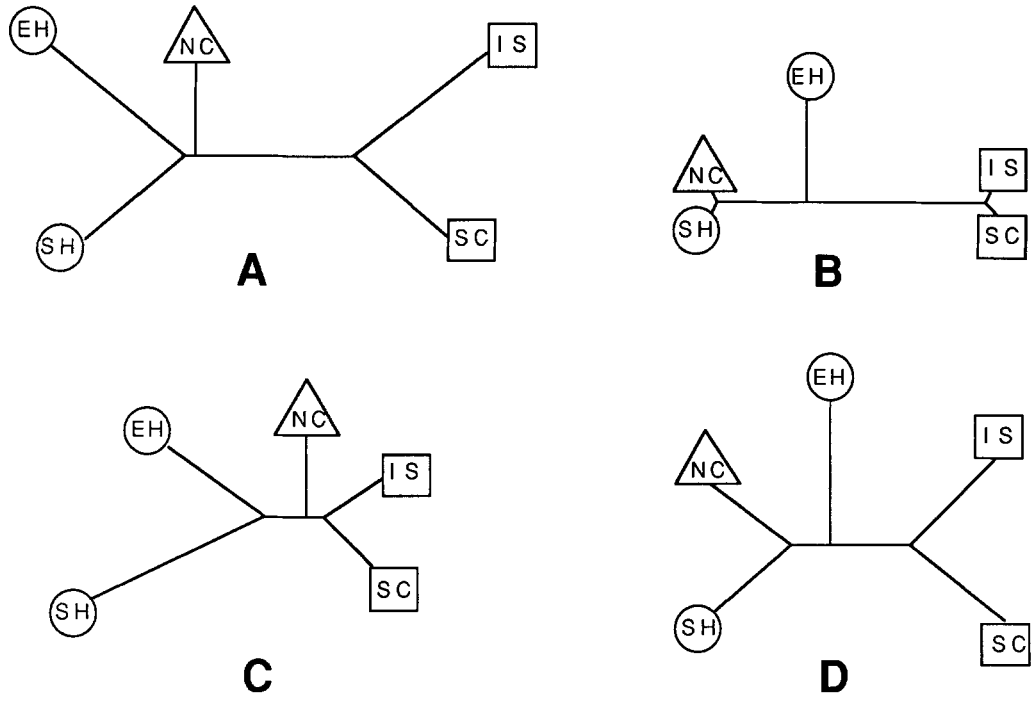

FIGURE 5.—Four unrooted, branching networks (*i.e.*, distance trees) illustrating patterns of similarity between five PNG populations. These distance trees were constructed using the PHYLIP package of computer programs (FELSENSTEIN 1987). A, Maximum-likelihood tree; B, Fitch-Margoliash (FM) tree of D_{NT} distances from Table 6 (with 0.0005 added to each value to avoid negative numbers); C, FM tree of D_{TP} distances from Table 6; D, FM tree of D_{HJ} distances from Table 6. The populations include two from the highlands (EH and SH, circles), two from the coast (IS and SC, squares), and the NC population (triangle), which groups with the highland populations in some trees and with the coastal populations in other trees.

the 65 PNG mtDNA types occurred more than once; the distribution of these 12 types within the five PNG populations is shown in Figure 7. The probability of allelic identity, based on mtDNA types, is 0.1238 within the five PNG populations, 0.0012 among the five PNG populations, and 0.0 between PNG and other worldwide populations (*i.e.*, none of the PNG mtDNA types have been found in any other population). The probability of identity thus decreases by two orders of magnitude when going from comparisons within PNG populations to comparisons between PNG populations.

As a means of further investigating the influence of geographic distance on mtDNA, we compared the distribution of geographic distances separating pairs of individuals with identical mtDNA types to the distribution of geographic distances separating all pairs of PNG individuals (Figure 8). These two distributions are significantly different ($\chi^2 = 183.85$, d.f. = 10, $P < 0.001$), with a much higher proportion of individuals with identical mtDNA types living within 100 km of each other. The average geographic distance between individuals with identical mtDNA types is 239 km, much less than the overall average distance of 383 km between PNG individuals in this study.

mtDNA variation and language groups: The lan-

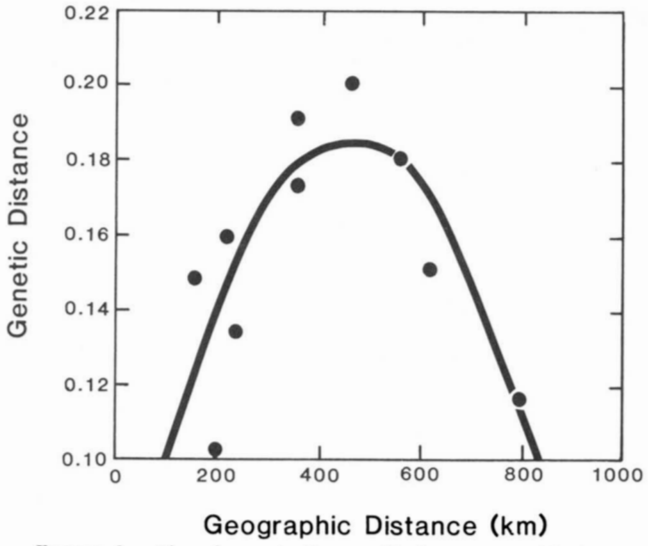

FIGURE 6.—Plot of genetic distance (D_{HJ}) *vs.* geographic distance for five PNG populations.

guage spoken by the mothers of 115 of the 119 PNG individuals was known and could be classified as either Austronesian (AN) or non-Austronesian (NAN); the sample sizes, number of mtDNA types and clans, ratio of mtDNA types to clans, and mean variability for these two groups are in Table 4. The NAN speakers are significantly more variable than the AN speakers and have more mtDNA types represented per clan. The G_{ST} value for the comparison of these two language groups is 0.23 for mtDNA restriction sites and 0.04 for mtDNA types (Table 5). Based on random resampling, both of these values are statistically significant ($P < 0.001$), indicating that there is significant differentiation between these two language groups with respect to mtDNA.

However, a caveat to this last finding is that the analysis is confounded by a geographic effect: all of the AN speakers were from the coast, whereas 64 of the NAN speakers were from the highlands and only 17 were from the coast. Since we have already demonstrated that the highland and coastal populations are significantly differentiated with respect to mtDNA, it is necessary to account for this geographic distinction. Two additional analyses in Table 5 demonstrate that the significant differences observed between NAN and AN speakers with respect to mtDNA are in fact due to this geographic distinction. First, if one considers only the NAN speakers, the G_{ST} values for the comparison of highland NAN speakers to coastal NAN speakers are still statistically significant, indicating that the geographic distinction between highland and coastal PNG mtDNAs persists within NAN speakers. Second, if one considers only the coastal mtDNAs, the G_{ST} values for the comparison of NAN speakers to AN speakers are no longer statistically significant. Thus geography, not language, appears to be the primary factor influencing the distribution of mtDNA variation in PNG.

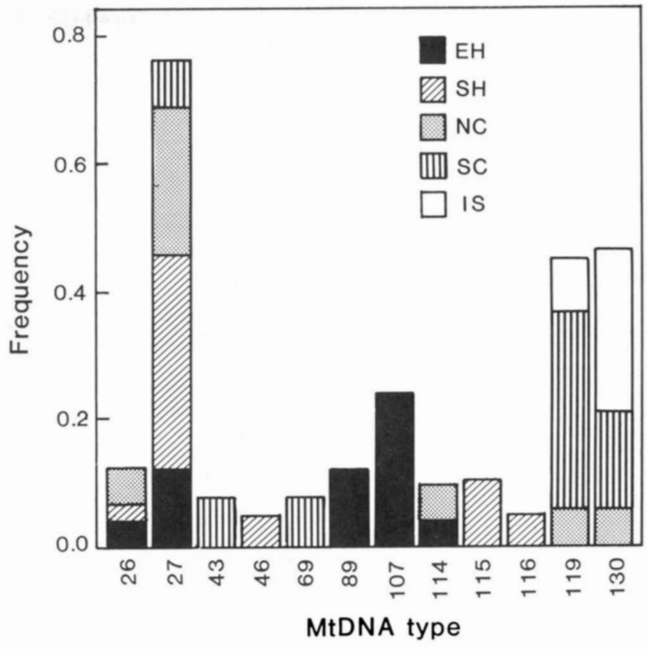

FIGURE 7.—Frequencies of mtDNA types that occurred in more than one individual. Types are numbered according to the tree in Figure 3. The height of each shaded bar indicates the frequency of that type in that particular population.

DISCUSSION

Significant geographic structure in worldwide populations: The analysis of mtDNA variation in the worldwide sample from five geographic localities indicated statistically significant structuring of these five populations with respect to mtDNA; the observed G_{ST} value of 0.31 indicates that about a third of the total variation is attributable to between-population variation. From a subset of these data WHITTAM *et al.* (1986) calculated a G_{ST} value of 0.063. The difference between their estimate and our estimate is probably due to the different definitions of loci and alleles; the TAKAHATA and PALUMBI (1985) method we used essentially treats each cleavage site as a locus with two alleles, while WHITTAM *et al.* (1986) used each functional region of the mtDNA genome as a locus, defining alleles on the basis of distinct combinations of cleavage sites mapping within each region.

The G_{ST} value for nuclear genes is about 0.10 (NEI and ROYCHOUDHURY 1982), which is considerably smaller than the mtDNA G_{ST} value of 0.31. The greater between-population differentiation displayed by mtDNA is in accord with expectations as it presumably reflects the faster evolution and smaller effective population size of mtDNA (due to the maternal, haploid inheritance). TAKAHATA and PALUMBI (1985) describe how the G_{ST} value for mtDNA can be used to derive an estimate of the minimum number of female migrants exchanged on average between populations per generation ($N_e m_e$), while the G_{ST} value for nuclear DNA gives a similar estimate of the total number of migrants (male plus female). From the G_{ST} values given

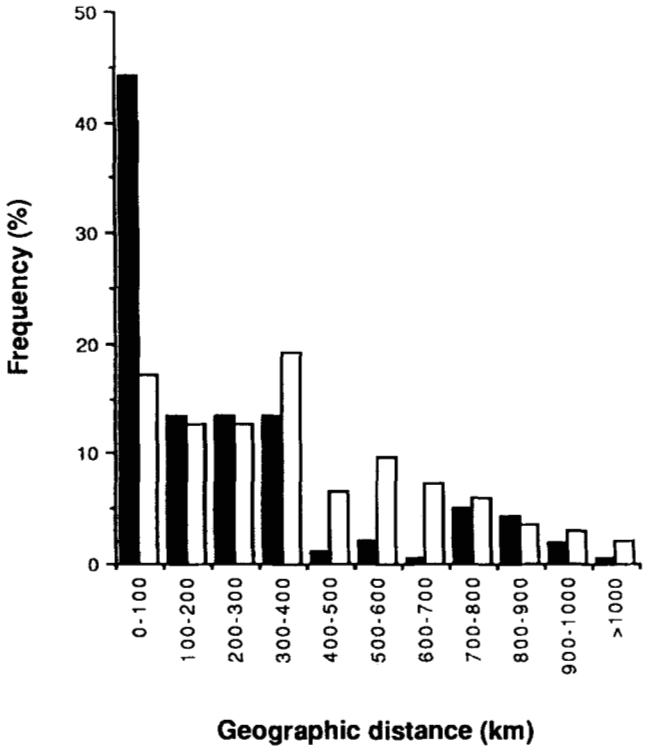

FIGURE 8.—Distribution of geographic distances between pairs of New Guineans with identical mtDNA types (solid bars) and between all pairs of New Guineans (hollow bars).

above, $N_e m_e$ is 1.11 for mtDNA and 2.25 for nuclear DNA. Similarly, an estimate of the effective population size of human females is about 6,000 (WILSON *et al.* 1985), which is about half the estimate of 10,000 from nuclear DNA for males and females (NEI and GRAUR 1984). It should be emphasized that these are only rough estimates, since the assumptions used to derive them (*e.g.*, equilibrium) are no doubt violated by human populations. Furthermore, differences in the methods used to sample genes and calculate G_{ST} values can have a large influence on the results, as when our results are compared to those of WHITTAM *et al.* (1986). A high-resolution sampling of both mtDNA and nuclear DNA genes from the same worldwide populations, with G_{ST} values calculated by similar methods, is needed to verify the above findings.

Lack of association between Australian and PNG mtDNA lineages: The phylogenetic analysis of PNG mtDNA lineages points to a Southeast Asian origin and does not support a particularly close phylogenetic relationship between PNG and Australian mtDNA lineages. This lack of a strong Australia-PNG association is surprising, since at the time of the first colonization of the area by humans, approximately 40,000–50,000 yr ago (WHITE and O'CONNELL 1982; GROUBE *et al.* 1986), Australia and New Guinea were a single land mass, Sahul, and were separated by rising sea levels only 8,000 yr ago (WHITE and O'CONNELL 1982). This presumably would not be sufficient time to erase the association between Australian and PNG populations.

Other evidence bearing on the relatedness of Australian and PNG populations does not give a clear-cut picture. The linguistic evidence for any Australian-PNG association is tenuous at best (FOLEY 1986). Analyses of blood group, serum protein, and red cell enzyme variation do tend to associate Australia and PNG (KEATS 1977; KIRK 1979; NEI and ROYCHOUDHURY 1982; KAMBOH and KIRK 1983a, b, 1984; KAMBOH and KIRWOOD 1984; KAMBOH, RANFORD and KIRK 1984; NEI 1985; CAVALLI-SFORZA *et al.*, 1988; CHEN *et al.* 1990). However, the genetic distance separating Australian and PNG populations tends to be quite large, comparable to the average genetic distance separating Asian and Caucasian populations (*e.g.*, NEI 1985; CAVALLI-SFORZA *et al.* 1988). In particular, CAVALLI-SFORZA *et al.* (1988) found that a bootstrap analysis supported the association of Australia and PNG only about 50% of the time. Furthermore, HLA haplotypes show virtually no association between Australia and PNG (SERJEANTSON, RYAN and THOMPSON 1982; CRANE *et al.* 1985; SERJEANTSON 1985).

NEI (1985) suggests that the large genetic distance between Australia and PNG is caused by inbreeding and drift due to isolation and reduced population sizes, and that HLA has been subject to selection, thereby distorting ancestral relationships. While drift and inbreeding (and selection on HLA) have undoubtedly played a role, the mtDNA results point to a different explanation: additional migrations (primarily through PNG) in the 8,000 yr since the physical separation of Australia and New Guinea have decreased the genetic association between Australia and PNG.

It is known from linguistic and archaeological evidence that major population movements occurred through both the highland and coastal regions of PNG over the past 10,000 yr (WURM *et al.* 1975; KIRK 1980, 1982a, b; WURM 1982, 1983). The last major migration of NAN speakers through the highlands occurred about 5,000 yr ago, while there were two major migrations of AN speakers through the coastal regions 5,000 yr ago and 3,500 yr ago (SERJEANTSON, KIRK and BOOTH 1983; WURM 1983; FOLEY 1986). These migrations may have considerably altered the genetic structure of PNG populations, accounting for the lack of a close genetic relationship between Australia and PNG mtDNA lineages.

Indeed, a 9-bp deletion in a small noncoding region of human mtDNA may provide a marker for the most recent population movement along the coast of PNG. Phylogenetic analysis indicates that this deletion arose only once and it appears to occur only in populations of Asian origin (WRISCHNIK *et al.* 1987; STONEKING and WILSON 1989). The deletion occurs at high frequency in coastal PNG but is absent from highland PNG and from Australia (HERTZBERG *et al.* 1989; STONEKING and WILSON 1989). It also occurs in In-

donesia (M. STONEKING, A. SOFRO and A. C. WILSON, unpublished results) and is nearly fixed in Polynesia (HERTZBERG *et al.* 1989). Thus, the distribution of the deletion mirrors the presumed path of the last AN migration from Southeast Asia through Indonesia, along the coast of New Guinea, and then through Melanesia into Polynesia (BELLWOOD 1987). The lack of the mtDNA deletion in both highland PNG and Australia indicates a stronger genetic association between these two regions, as suggested previously (KIRK 1980, 1982a, b). Studies of mtDNA variation in any highland PNG populations that are derived from the primary colonization of Sahul and isolated from later migrations might reveal even closer genetic associations between Australia and PNG.

Geographic structuring of coastal and highland populations: The analysis of mtDNA variation within PNG populations showed that highland and coastal populations are strongly differentiated with respect to mtDNA. Highland populations have significantly more mtDNA variability than coastal populations and each highland mtDNA lineage is represented on average by nearly twice as many mtDNA types as each coastal mtDNA lineage (Table 4). These results probably reflect the older colonization of the highlands and more frequent recolonization of the coast, for the following reason: assuming no additional migration following colonization, mtDNA lineages will be lost through time due to stochastic processes (AVISE, NEIGEL and ARNOLD 1984; STONEKING and WILSON 1989), while new mtDNA types will be gained via mutation. Thus, the number of mtDNA clans will decrease through time while the ratio of mtDNA types to clans will increase through time. Additional migrations through the coastal region (discussed above) will increase the number of clans but not the ratio of types to clans.

Highland populations are sufficiently isolated that even the Eastern and Southern Highland populations, located about 200 km apart, are significantly differentiated with respect to mtDNA. By contrast, coastal populations located over a range in excess of 1000 km are not significantly differentiated with respect to mtDNA, presumably because there has not been enough time and/or because migration between coastal populations is sufficient to prohibit such differentiation.

These results are in excellent agreement with studies of blood group, serum protein, red cell enzyme, and HLA variation in PNG populations. It is well-known that PNG populations tend to exhibit large between-population differences (GILES, OGAN and STEINBERG 1965; SIMMONS *et al.* 1972; BOOTH 1974; WIESENFELD and GADJUSEK 1976; KEATS 1977; KIRK 1980, 1982a, b; MOURANT *et al.* 1982; WOOD *et al.* 1982; SERJEANTSON, RYAN and THOMPSON 1982; SERJEANTSON, KIRK and BOOTH 1983; CRANE *et al.* 1985; SERJEANTSON 1985). The most extreme differences occur between highland and coastal populations, with F_{ST} values (similar to G_{ST} values) ranging from 0.02 to 0.04 (summarized in JORDE 1980).

A caveat to the overall agreement between the mtDNA results and the above studies of variation in the products of nuclear genes concerns differences in methodology. A more direct comparison is provided by the study of restriction site polymorphism in nuclear DNA by HILL (1986), who provided allele frequencies at five unlinked loci for five highland and three coastal populations. From his data we calculate G_{ST} values of 0.03 for the highland populations and 0.01 for the coastal populations. Thus, as with the mtDNA results, highland populations are more strongly differentiated from one another than coastal populations are differentiated from one another. Furthermore, PNG populations show more differentiation with respect to mtDNA than nuclear DNA, as was the case previously with G_{ST} analyses based on the global survey of human populations.

The genetic distance analyses of the PNG populations (summarized in the branching diagrams in Figure 5) were not completely concordant, contrary to the results commonly obtained when various distance measures are applied to the same data set (JORDE 1985). The discrepancy basically concerns the placement of the NC population: the TAKAHATA-PALUMBI distances place this population with the other two coastal populations (SC and IS), whereas the NEI-TAJIMA and HARPENDING-JENKINS distances, and the maximum-likelihood analysis, place the NC population with the two highland populations (EH and SH). One possible explanation concerns the assumptions underlying each of the methods. The D_{NT}, D_{HJ} and maximum-likelihood analyses all basically assume that populations are evolving independently by mutation and drift; only the D_{TP} analysis explicitly assumes that migration is occurring as well. The assumptions of the latter method would thus appear to be more realistic for the present situation.

Unusual relationship between genetic and geographic distance: The analysis of genetic distance with respect to geographic distance between PNG populations showed an unusual quadratic relationship with populations located either close together or far apart being more similar genetically than populations located at intermediate distances (Figure 6). It is important to keep in mind that this peculiar pattern might simply be due to chance, since the number of populations compared is relatively small. Also, the Euclidean measure of geographic distance probably does not accurately reflect the actual travel distances between PNG populations. However, the pattern shown in Figure 6 is consistent with the genetic distance analyses in that the two distinctly coastal populations (IS and SC) are among the most closely related populations, despite being located several hundred kilometers apart. It appears that this reflects the pat-

tern of migrations to and across PNG. The earlier colonization of the highlands permitted more differentiation between highland populations. Later, separate migrations to the coastal regions resulted in greater differentiation between highland and coastal populations but little or no differentiation between various coastal populations.

In addition to geography, we also investigated the contribution of linguistic differences to mtDNA differentiation between PNG populations, and found that the apparently significant differences between NAN mtDNAs and AN mtDNAs disappeared when geography was considered. Previous studies of nuclear DNA markers originally claimed significant allele frequency distributions between NAN and AN groups (GILES, OGAN and STEINBERG 1965; BOOTH 1974; SCHANFIELD, GILES and GERSHOWITZ 1975); more thorough studies involving more loci and taking into account geography have failed to find any consistent differences between NAN and AN groups (BOYCE *et al.* 1978; KIRK 1980; SERJEANTSON, KIRK and BOOTH 1983). The mtDNA results are thus in excellent agreement with the nuclear DNA results.

Geography has a profound influence on the distribution of mtDNA types in PNG. The probability of allelic identity based on mtDNA types was two orders of magnitude larger within PNG populations than it was between PNG populations. Furthermore, the average geographic distance separating PNG individuals with identical mtDNA types was significantly less than the average distance between any two individuals in this study. As shown in Figure 8, 44% of the pairwise comparisons of geographic distance were less than 100 km for individuals with identical mtDNA types, versus just 17% for the entire sample of individuals.

Previously, the only observation that could be made concerning the effect of geography on the distribution of mtDNA types was that individuals from different continents never shared identical mtDNA types (CANN, STONEKING and WILSON 1987; STONEKING and WILSON 1989). Even this relatively simple observation required high resolution restriction maps that allowed closely related mtDNA types to be distinguished; the previously employed low resolution techniques had found mtDNA types to be shared between continents (*e.g.*, JOHNSON *et al.* 1983). The limited geographic sampling of the above studies also restricted our understanding of the influence of geography on mtDNA variation within continents. Our extensive geographic sampling of 119 individuals from PNG has enhanced our appreciation of the influence of geography. As future studies obtain even more thorough geographic sampling, using higher resolution techniques to examine mtDNA variation (*e.g.*, VIGILANT *et al.* 1989), we expect to attain a deeper understanding of the relationship between geography and patterns of mtDNA variability.

We thank N. DAS, K. EDWARDS and R. L. KIRK for assistance with the collection of tissues, S. KURTZ for technical assistance, K. HESTIR for statistical advice, J. C. STEPHENS and D. SWOFFORD for computational assistance, and M. P. ALPERS, R. L. CANN, W. R. ENGELS, U. GYLLENSTEN, H. OCHMAN, E. M. PRAGER, A. ROGERS, V. M. SARICH, G. THOMSON and L. VIGILANT for useful discussion. This research received support from National Science Foundation and National Institutes of Health grants to A.C.W., and from National Institutes of Health grant GM-39245 and National Science Foundation grant BNS-8720330 to L.B.J. M.S. was supported by a National Institutes of Health postdoctoral fellowship.

LITERATURE CITED

ALLEN, B., 1983 Human geography of Papua New Guinea. J. Hum. Evol. **12:** 3–23.

ANDERSON, S., A. T. BANKIER, B. G. BARRELL, M. H. L. DE BRUIJN, A. R. COULSON, J. DROUIN, I. C. EPERON, D. P. NIERLICH, B. A. ROE, F. SANGER, P. H. SCHREIER, A. J. H. SMITH, R. STADEN and I. G. YOUNG, 1981 Sequence and organization of the human mitochondrial genome. Nature **290:** 457–465.

AVISE, J. C., 1986 Mitochondrial DNA and the evolutionary genetics of higher animals. Philos. Trans. R. Soc. Lond. B **312:** 325–342.

AVISE, J. C., J. E. NEIGEL and J. ARNOLD, 1984 Demographic influences on mitochondrial DNA lineage survivorship in animal populations. J. Mol. Evol. **20:** 99–105.

BELLWOOD, P., 1987 *The Polynesians, Prehistory of an Island People.* Thames and Hudson, London.

BHATIA, K., M. GOROGO and G. KOKI, 1984 HLA-A,B,C and DR antigens in Asaro speakers of Papua New Guinea. Hum. Immunol. **9:** 189–200.

BHATIA, K., N. M. BLAKE, S. W. SERJEANTSON and R. L. KIRK, 1981 Frequency of private electrophoretic variants and indirect estimates of the mutation rate in Papua New Guinea. Am. J. Hum. Genet. **33:** 112–122.

BHATIA, K., M. PRASAD, G. BARNISH and G. KOKI, 1988 Antigen and haplotype frequencies at three human leucocyte antigen loci (HLA-A, -B and -C) in the Pawaia of Papua New Guinea. Am. J. Phys. Anthropol. **75:** 329–340.

BHATIA, K., C. JENKINS, M. PRASAD, G. KOKI and J. LOMBANGE, 1989 Immunogenetic studies of two recently contacted populations from Papua New Guinea. Hum. Biol. **61:** 45–64.

BIRKY, C. W., T. MARUYAMA and P. FUERST, 1983 An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts, and some results. Genetics **103:** 513–527.

BOOTH, P. B., 1974 Genetic distances between certain New Guinea populations studied under the International Biological Programme. Philos. Trans. R. Soc. Lond. B **268:** 257–267.

BOYCE, A. J., G. A. HARRISON, C. M. PLATT, R. W. HORNABROOK, S. W. SERJEANTSON, R. L. KIRK and P. B. BOOTH, 1978 Migration and genetic diversity in an island population: Karkar, Papua New Guinea. Proc. R. Soc. Lond. B **202:** 269–295.

BROWN, W. M., 1980 Polymorphism in mitochondrial DNA of humans as revealed by restriction endonuclease analysis. Proc. Natl. Acad. Sci. USA **77:** 3605–3609.

BROWN, W. M., M. GEORGE and A. C. WILSON, 1979 Rapid evolution of animal mitochondrial DNA. Proc. Natl. Acad. Sci. USA **76:** 1967–1971.

BROWN, W. M., E. M. PRAGER, A. WANG and A. C. WILSON, 1982 Mitochondrial DNA sequences of primates: tempo and mode of evolution. J. Mol. Evol. **18:** 225–239.

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

CANN, R. L., M. STONEKING and A. C. WILSON, 1987

Mitochondrial DNA and human evolution. Nature **325:** 31–36.

CAVALLI-SFORZA, L. L., A. PIAZZA, P. MENOZZI and J. MOUNTAIN, 1988 Reconstruction of human evolution: bringing together genetic, archaeological, and linguistic data. Proc. Natl. Acad. Sci. USA **85:** 6002–6006.

CHEN, L. Z., S. EASTEAL, P. G. BOARD, K. M. SUMMERS, K. BHATIA and R. L. KIRK, 1990 Albumin-Vitamin D binding protein haplotypes in Asian-Pacific populations. Hum. Genet. (in press).

CRANE, G., K. BHATIA, M. HONEYMAN, T. DORAN, N. MESSEL, G. HAKOS, D. TARLINTON, D. B. AMOS and H. BASHIR, 1985 HLA studies of Highland and Coastal New Guineans. Hum. Immunol. **12:** 247–260.

CROW, J. F., and K. AOKI, 1984 Group selection for a polygenic behavioral trait: estimating the degree of population subdivision. Proc. Natl. Acad. Sci. USA **81:** 6073–6077.

ENGELS, W. R., 1981 Estimating genetic divergence and genetic variability with restriction endonucleases. Proc. Natl. Acad. Sci. USA **78:** 6329–6333.

FELSENSTEIN, J., 1987 *PHYLIP 3.1 Manual.* University of California Herbarium, Berkeley.

FOLEY, W. A., 1986 *The Papuan Languages of New Guinea.* Cambridge University Press, Cambridge.

FROEHLICH, J. W., and E. GILES, 1981 A multivariate approach to fingerprint variation in Papua New Guinea: perspectives on the evolutionary stability of dermatoglyphic markers. Am. J. Phys. Anthropol. **54:** 93–106.

GILES, E., E. OGAN and A. G. STEINBERG, 1965 Gamma-globulin factors (Gm and Inv) in New Guinea: anthropological significance. Science **150:** 1158–1160.

GILES, R. E., H. BLANC, H. M. CANN and D. C. WALLACE, 1980 Maternal inheritance of human mitochondrial DNA. Proc. Natl. Acad. Sci. USA **77:** 6715–6719.

GROUBE, L., J. CHAPPELL, J. MUKE and D. PRICE, 1986 A 40,000 year-old human occupation site at Huon Peninsula, Papua New Guinea. Nature **324:** 453–455.

HARPENDING, H., and T. JENKINS, 1973 Genetic distance among southern African populations, pp. 177–199 in *Methods and Theories of Anthropological Genetics*, edited by M. H. CRAWFORD and P. L. WORKMAN. University of New Mexico Press, Albuquerque.

HERTZBERG, M., K. N. P. MICKLESON, S. W. SERJEANTSON, J. F. PRIOR and R. J. TRENT, 1989 An Asian-specific 9-bp deletion of mitochondrial DNA is frequently found in Polynesians. Am. J. Hum. Genet. **44:** 504–510.

HILL, A. V. S., 1986 The population genetics of alpha-thalassemia and the malaria hypothesis. Cold Spring Harbor Symp. Quant. Biol. **51:** 489–498.

HOPE, G. S., J. GOLSON and J. ALLEN, 1983 Paleoecology and prehistory in New Guinea. J. Hum. Evol. **12:** 37–60.

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

JOHNSON, M. J., D. C. WALLACE, S. D. FERRIS, M. C. RATTAZZI and L. L. CAVALLI-SFORZA, 1983 Radiation of human mitochondria DNA types analyzed by restriction endonuclease cleavage patterns. J. Mol. Evol. **19:** 255–271.

JORDE, L. B., 1980 The genetic structure of subdivided human populations, pp. 133–208 in *Current Developments in Anthropological Genetics, Vol. 1: Theory and Methods*, edited by J. H. MIELKE and M. H. CRAWFORD. Plenum Press, New York.

JORDE, L. B., 1985 Human genetic distance studies: present status and future prospects. Annu. Rev. Anthropol. **14:** 343–373.

KAMBOH, M. I., and R. L. KIRK, 1983a Alpha-1-antitrypsin (PI) types in Asian, Pacific, and aboriginal Australian populations. Dis. Markers **1:** 33–42.

KAMBOH, M. I., and R. L. KIRK, 1983b Distribution of transferrin (Tf) subtypes in Asian, Pacific and Australian aboriginal populations: evidence for the existence of a new subtype Tf(C6). Hum. Hered. **33:** 237–243.

KAMBOH, M. I., and R. L. KIRK, 1984 Genetic studies of PGM1 subtypes: population data from the Asian-Pacific area. Ann. Hum. Biol. **11:** 211–219.

KAMBOH, M. I., and C. KIRWOOD, 1984 Genetic polymorphism of the thyroxin-binding globulin (TBG) in the Pacific area. Am. J. Hum. Genet. **36:** 646–654.

KAMBOH, M. I., P. R. RANFORD and R. L. KIRK, 1984 Population genetics of the vitamin D binding protein (GC) subtypes in the Asian-Pacific area: description of new alleles at the GC locus. Hum. Genet. **67:** 378–384.

KEATS, B., 1977 Genetic structure of the indigenous populations in Australia and New Guinea. J. Hum. Evol. **6:** 319–339.

KIRK, R. L., 1979 Genetic differentiation in Australia and the Western Pacific and its bearing on the origin of the first Americans, pp. 211–237 in *The First Americans: Origins, Affinities, and Adaptations*, edited by W. S. LAUGHLIN. Gustav Fischer, New York.

KIRK, R. L., 1980 Language, genes and people in the Pacific, pp. 113–137 in *Population Structure and Genetic Disorders*, edited by A. N. ERIKSSON, H. FORSIUS, H. R. NEVANLINNA, P. L. WORKMAN and P. K. NORIO. Academic Press, New York.

KIRK, R. L., 1982a Linguistic, ecological, and genetic differentiation in New Guinea and the Western Pacific, pp. 229–253 in *Current Developments in Anthropological Genetics, Vol. 2: Ecology and Population Structure*, edited by J. H. MIELKE and M. H. CRAWFORD. Plenum Press, New York.

KIRK, R. L., 1982b Microevolution and migration in the Pacific, pp. 215–225 in *Human Genetics, Part A: The Unfolding Genome*, edited by B. BONNÉ-TAMIR, T. COHEN and R. M. GOODMAN. Alan R. Liss, New York.

MANTEL, N. A., 1967 The detection of disease clustering and a generalized regression approach. Cancer Res. **27:** 209–220.

MOURANT, A. E., D. TILLS, A. C. KOPEC, A. WARLOW, P. TEESDALE, P. B. BOOTH and R. W. HORNABROOK, 1982 Red cell antigen, serum protein and red cell enzyme polymorphisms in Eastern Highlanders of New Guinea. Hum. Hered. **32:** 374–384.

NEI, M., 1973 Analysis of gene diversity in subdivided populations. Proc. Natl. Acad. Sci. USA **70:** 3321–3323.

NEI, M., 1985 Human evolution at the molecular level, pp. 41–64 in *Population Genetics and Molecular Evolution*, edited by T. OHTA and K. AOKI. Springer-Verlag, Berlin.

NEI, M.,1987 *Molecular Evolutionary Genetics.* Columbia University Press, New York.

NEI, M., and D. GRAUR, 1984 Extent of protein polymorphism and the neutral mutation theory. Evol. Biol. **17:** 73–118.

NEI, M., and A. K. ROYCHOUDHURY, 1982 Genetic relationship and evolution of human races. Evol. Biol. **14:** 1–59.

NEI, M., and F. TAJIMA, 1983 Maximum likelihood estimation of the number of nucleotide substitutions from restriction sites data. Genetics **105:** 207–217.

OHTSUKA, R., T. KAWABE, T. INAOKA, T. AKIMICHI and T. SUZUKI, 1985 Inter- and intra-population migration of the Gidra in lowland Papua: a population-ecological analysis. Hum. Biol. **57:** 33–45.

SCHANFIELD, M. S., E. GILES and H. GERSHOWITZ, 1975 Genetic studies in the Markham Valley, northeastern Papua New Guinea: gamma globulin (Gm and Inv), group specific component (Gc) and ceruloplasmin (Cp) typing. Am. J. Phys. Anthropol. **42:** 1–8.

SERJEANTSON, S. W., 1985 Migration and admixture in the Pacific: insights provided by human leucocyte antigens, pp. 133–145 in *Out of Asia—Peopling the Americas and the Pacific*, edited by R. L. KIRK and E. SZATHMARY. Journal of Pacific History, Australian National University, Canberra.

SERJEANTSON, S. W., R. L. KIRK and P. B. BOOTH, 1983 Linguistic and genetic differentiation in New Guinea. J. Hum. Evol. **12:**

77–92.

SERJEANTSON, S. W., D. P. RYAN and A. R. THOMPSON, 1982 The colonization of the Pacific: the story according to human leukocyte antigens. Am. J. Hum. Genet. **34:** 904–918.

SIMMONS, R. T., J. J. GRAYDON, D. C. GADJUSEK, M. P. ALPERS and R. W. HORNABROOK, 1972 Genetic studies in relation to Kuru. II. Blood-group genetic patterns in Kuru patients and populations of the Eastern Highlands of New Guinea. Am. J. Hum. Genet. **24:** 539–571.

SINGH, G., N. NECKELMANN and D. C. WALLACE, 1987 Conformational mutations in human mitochondrial DNA. Nature **329:** 270–272.

SMOUSE, P. E., J. C. LONG and R. R. SOKAL, 1986 Multiple regression and correlation extensions of the Mantel test of matrix correspondence. Syst. Zool. **35:** 627–632.

STONEKING, M., 1986 Human mitochondrial DNA evolution in Papua New Guinea. Ph.D. thesis, University of California, Berkeley.

STONEKING, M., K. BHATIA and A. C. WILSON, 1986a Mitochondrial DNA variation in Eastern Highlanders of Papua New Guinea, pp. 87–100 in *Genetic Variation and Its Maintenance*, edited by D. F. ROBERTS and G. F. DE STEFANO. Cambridge University Press, Cambridge.

STONEKING, M., K. BHATIA and A. C. WILSON, 1986b Rate of sequence divergence estimated from restriction maps of mitochondrial DNAs from Papua New Guinea. Cold Spring Harbor Symp. Quant. Biol. **51:** 433–439.

STONEKING, M., and R. L. CANN, 1989 African origin of human mitochondrial DNA, pp. 17–30 in *The Human Revolution, Behavioural and Biological Perspectives on the Origins of Modern Humans*, edited by P. MELLARS and C. STRINGER. Edinburgh University Press, Edinburgh.

STONEKING, M., and A. C. WILSON, 1989 Mitochondrial DNA, pp. 215–245 in *The Colonization of the Pacific, A Genetic Trail*, edited by A. V. S. HILL and S. W. SERJEANTSON. Oxford University Press, Oxford.

SWOFFORD, D. L., 1985 *Phylogenetic Analysis Using Parsimony (PAUP), Version 2.4.* Illinois Natural History Survey, Champaign.

TAKAHATA, N., and M. NEI, 1985 Gene genealogy and variance of interpopulational nucleotide differences. Genetics **110:** 325–344.

TAKAHATA, N., and S. R. PALUMBI, 1985 Extranuclear differentiation and gene flow in the finite island model. Genetics **109:** 441–457.

VIGILANT, L., M. STONEKING and A. C. WILSON, 1988 Conformational mutation in human mtDNA detected by direct sequencing of enzymatically amplified DNA. Nucleic Acids Res. **16:** 5945–5955.

VIGILANT, L., R. PENNINGTON, H. HARPENDING, T. D. KOCHER and A. C. WILSON, 1989 Mitochondrial DNA sequences in single hairs from a southern African population. Proc. Natl. Acad. Sci. USA **86:** 9350–9354.

WHITTAM, T. S., A. G. CLARK, M. STONEKING, R. L. CANN and A. C. WILSON, 1986 Allelic variation in human mitochondrial genes based on patterns of restriction site polymorphism. Proc. Natl. Acad. Sci. USA **83:** 9611–9615.

WHITE, J. P., and J. F. O'CONNELL, 1982 *A Prehistory of Australia, New Guinea, and Sahul.* Academic Press, New York.

WIESENFELD, S. L., and D. C. GADJUSEK, 1976 Genetic structure and heterozygosity in the Kuru region, Eastern Highlands of New Guinea. Am. J. Phys. Anthropol. **45:** 177–190.

WILSON, A. C., R. L. CANN, S. M. CARR, M. GEORGE, U. B. GYLLENSTEN, K. M. HELM-BYCHOWSKI, R. G. HIGUCHI, S. R. PALUMBI, E. M. PRAGER, R. D. SAGE and M. STONEKING, 1985 Mitochondrial DNA and two perspectives on evolutionary genetics. Biol. J. Linn. Soc. **26:** 375–400.

WILSON, A. C., M. STONEKING, R. L. CANN, E. M. PRAGER, S. D. FERRIS, L. A. WRISCHNIK and R. G. HIGUCHI, 1987 Mitochondrial clans and the age of our common mother, pp. 158–164 in *Human Genetics, Proceedings of the Seventh International Congress, Berlin 1986*, edited by F. VOGEL and K. SPERLING. Springer-Verlag, Berlin.

WILSON, A. C., E. A. ZIMMER, E. M. PRAGER and T. D. KOCHER, 1989 Restriction mapping in the molecular systematics of mammals: a retrospective salute, pp. 407–419 in *The Hierarchy of Life*, edited by B. FERNHOLM, K. BREMER and H. JÖRNVALL. Elsevier Science B. V. (Biomedical Division), Amsterdam.

WOOD, J. W., P. E. SMOUSE and J. C. LONG, 1985 Sex-specific dispersal patterns in two human populations of highland New Guinea. Am. Nat. **125:** 747–768.

WOOD, J. W., P. L. JOHNSON, R. L. KIRK, K. MCLOUGHLIN, N. M. BLAKE and F. A. MATHESON, 1982 The genetic demography of the Gainj of Papua New Guinea. I. Local differentiation of blood group, red cell enzyme, and serum protein allele frequencies. Am. J. Phys. Anthropol. **57:** 15–25.

WRISCHNIK, L. A., R. G. HIGUCHI, M. STONEKING, H. A. ERLICH, N. ARNHEIM and A. C. WILSON, 1987 Length mutations in human mitochondrial DNA: direct sequencing of enzymatically amplified DNA. Nucleic Acids Res. **15:** 529–542.

WURM, S. A., 1982 *Papuan Languages of Oceania.* Gunter Narr, Tübingen.

WURM, S. A., 1983 Linguistic prehistory in the New Guinea area. J. Hum. Evol. **12:** 25–35.

WURM, S. A., D. C. LAYCOCK, C. L. VOORHOEVE and T. E. DUTTON, 1975 Papuan linguistic history and past language migrations in the New Guinea area, pp. 935–960 in *New Guinea Area Languages and Language Study, Vol. 1: Papuan Languages and the New Guinea Linguistic Scene*, edited by S. A. WURM. Pacific Linguistics, Australian National University, Canberra.

Communicating editor: A. G. CLARK

APPENDIX

A list of the state of each polymorphic restriction site in each mtDNA type is given in Figure 9. mtDNA types are numbered according to the tree in Figure 3. A "1" indicates that the site is present and a "0" indicates that the site is absent, except in the case of Region V (the small noncoding region between the COII and lysine tRNA genes), where "1" indicates one copy of the 9-bp repeat and "2" indicates two copies of the repeat. The nomenclature for designating sites follows that of CANN, STONEKING and WILSON (1987): sites are numbered according to the published sequence (ANDERSON *et al.* 1981), with italicized numbers indicating the gain of a site absent from the published sequence and nonitalicized numbers indicating the loss of a site present in the published sequence. The 12 restriction enzymes are designated by the following single-letter code: *Alu*I, a; *Ava*II, b; *Dde*I, c; *Fnu*DII, d; *Hae*III, e; *Hha*I, f; *Hin*fI, g; *Hin*cII, h; *Hpa*II, i; *Mbo*I, j; *Rsa*I, k; *Taq*I, l. Italicized sites separated by commas indicate alternative placements of inferred site gains. Sites separated by a slash indicate simultaneous gain of a site for one enzyme and loss of a site for a different enzyme, due to a single inferred nucleotide substitution; these sites are counted once in the analyses as the gain of the new restriction site.

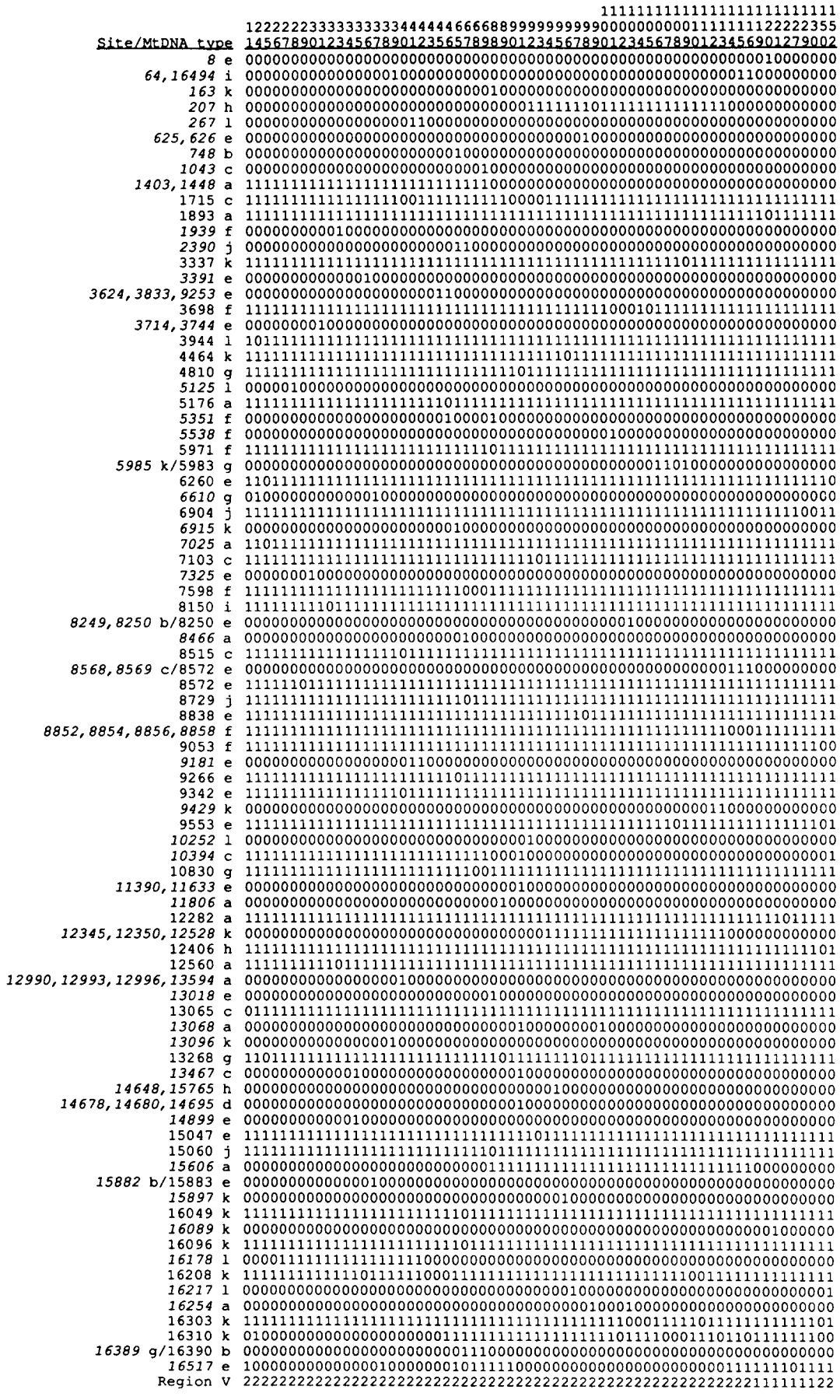

FIGURE 9.—Polymorphic restriction sites.