Am. J. Hum. Genet. 63:613–624, 1998

Mitochondrial and Nuclear Genetic Relationships among Pacific Island and Asian Populations

J. Koji Lum,[1] Rebecca L. Cann,[2] Jeremy J. Martinson,[3] and Lynn B. Jorde[4]

[1]Institute of Statistical Mathematics, Tokyo; [2]Department of Genetics, John A. Burns School of Medicine, University of Hawaii at Manoa, Honolulu; [3]Department of Human Genetics, University of Pittsburgh, Pittsburgh; and [4]Department of Human Genetics, Eccles Institute of Human Genetics, University of Utah Health Sciences Center, Salt Lake City

Summary

Mitochondrial and autosomal short tandem-repeat (STR) genetic distances among 28 Pacific Island and Asian populations are significantly correlated ($r = .25$, $P < .01$) but describe distinct patterns of relationships. Maternally inherited–mtDNA data suggest that Remote Oceanic Islanders originated in island Southeast Asia. In contrast, biparental STR data reveal substantial genetic affinities between Remote Oceanic Islanders and Near Oceanic populations from highland Papua New Guinea and Australia. The low correlation between maternal and biparental genetic markers from the same individuals may reflect differences in genome-effective population sizes or in sex-biased gene flow. To explore these possibilities, we have examined genetic diversity, gene flow, and correlations among genetic, linguistic, and geographic distances within four sets of populations representing potential geographic and cultural spheres of interaction. G_{ST} estimates (a measure of genetic differentiation inversely proportional to gene flow) from mtDNA sequences vary between 0.13 and 0.39 and are typically five times greater than G_{ST} estimates from STR loci (0.05–0.08). Significant correlations ($r > .5$, $P < .05$) between maternal genetic and linguistic distances are coincident with high mtDNA G_{ST} estimates (>0.38). Thus, genetic and linguistic distances may coevolve, and their correspondence may be preserved under conditions of genetic isolation. A significant correlation ($r = .65$, $P < .01$) between biparental genetic and geographic distances is coincident with a low STR G_{ST} estimate (0.05), indicating that isolation by distance is observed under conditions of high nuclear-gene flow. These results are consistent with an initial settlement of Remote Oceania from island Southeast Asia and with extensive postcolonization male-biased gene flow with Near Oceania.

Received October 8, 1997; accepted for publication June 2, 1998; electronically published June 29, 1998.

Address for correspondence and reprints: Dr. J. Koji Lum, The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106-8569, Japan. E-mail: lum@ism.ac.jp

0002-9297/98/6202-0037$02.00

Introduction

The origin and affinities of Pacific Islanders have been examined from many perspectives, resulting in different groupings and nomenclature. Largely on the basis of geography, the Pacific Islands are commonly divided into three regions: Micronesia, Melanesia, and Polynesia (fig. 1). Archaeological evidence indicates two major periods of Pacific Island settlement. The first wave of settlement is dated to the Pleistocene and included Australia 50,000 years before the present (BP) (Roberts et al. 1990), New Guinea 40,000 years BP (Groube et al. 1986), the Bismarck Archipelago 33,000 years BP (Allen et al. 1988), and the northern Solomon Islands 29,000 years BP (Wickler and Spriggs 1988). The geographic limit of Pleistocene settlement corresponds to both an increase in distances between islands and a general reduction in island size. These factors served as a barrier not only to further human dispersal, but also to the dispersal of many other species of animals and plants. This region of early settlement is correlated with the geographic limit of Papuan-speaking populations in the Pacific. Papuan languages are a loosely defined group of extremely diverse languages, in contrast to the relatively homogenous Austronesian language family discussed below.

The second wave of Pacific colonization is dated to the last 4,000 years BP and includes the remaining, Austronesian-speaking portions of Melanesia and all of Polynesia and Micronesia. Austronesian-speaking populations are also found throughout island Southeast Asia and along the coasts and offshore islands of western Melanesia, particularly along the north coast of New Guinea. The coastal distribution of Austronesian languages within Melanesia has been interpreted as marking the route of the second wave of settlement (Diamond 1988). On the basis of the concordant linguistic and biogeographical boundaries, two alternative groupings of Pacific Islands—Near Oceania and Remote Oceania—have been proposed (Pawley and Green 1973; Green 1991). Thus, both linguistic and archaeological data describe three groups of Pacific Islanders: Near Oceanic Papuan speakers inhabiting regions of great antiquity, Near Oceanic Austronesian speakers in coastal

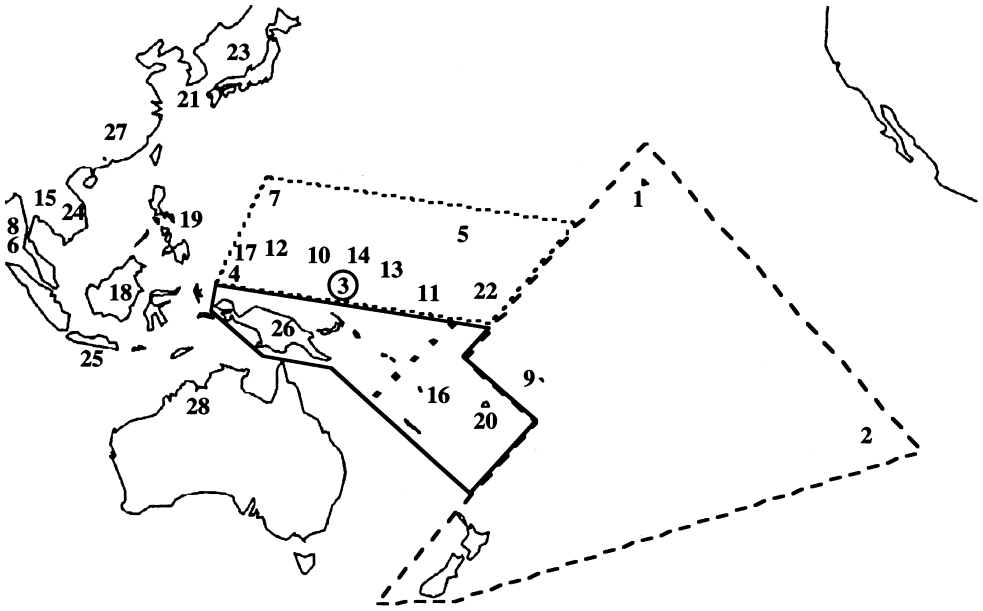

Figure 1 Map showing boundaries of Micronesia (dotted line), Melanesia (solid line), and Polynesia (dashed line) and locations of the populations examined. Remote Oceania consists of Micronesia, Polynesia, and eastern Melanesia, delineated with a dashed line. The circle around Kapingamarangi (3) signifies that this Polynesian atoll is geographically within Micronesia. The populations are numbered in ascending order of mtDNA genetic diversities (see table 1): 1 = Hawaii; 2 = Rapanui; 3 = Kapingamarangi; 4 = Southwest Palau; 5 = Marshalls; 6 = Moken; 7 = Marianas; 8 = Urak Lawoi; 9 = Samoa; 10 = Outer Island Yap; 11 = Nauru; 12 = Yap; 13 = Kosrae; 14 = Pohnpei; 15 = Thailand; 16 = Vanuatu; 17 = Palau; 18 = Borneo; 19 = Philippines; 20 = Fiji; 21 = Korea; 22 = Kiribati; 23 = Japan; 24 = Vietnam; 25 = Java; 26 = Papua New Guinea; 27 = China; and 28 = Australia.

regions, and Remote Oceanic Austronesian speakers in relatively recently settled regions (fig. 1).

The linguistic patterns and correlated settlement dates described above support the "express train" model (Diamond 1988) of Remote Oceanic colonization. This model sees island Southeast Asia as the origin of the Austronesian language family and, thus, of the Remote Oceanic Islanders. mtDNA analyses of Pacific Islanders have identified predominantly Asian haplotypes in Polynesia and Micronesia, with minimal genetic input (5%) from Near Oceania (Hertzberg et al. 1989; Lum et al. 1994; Redd et al. 1995; Sykes et al. 1995; Lum and Cann 1998), in general agreement with the express train model.

A competing hypothesis sees Remote Oceanic settlement as an outgrowth of complex, long-term interactions among Near Oceanic and other western Pacific populations. These long-term contacts are thought to result in culturally and biologically heterogeneous populations in western Melanesia that gave rise to Remote Oceanic colonists, without a second migration. This view argues that Pacific populations cannot be described by a simple branching tree but, rather, should be viewed as an entangled bank (Terrell 1988). The term "entangled bank" is borrowed from Darwin's (1859) work, in which he attempted to emphasize that the complex, subtle forces affecting the fitness of species are difficult to dissect by examining a single trait or behavior. Data from nuclear encoded loci (Serjeantson 1985; O'Shaughnessy et al. 1990; Martinson 1996; Roberts-Thomson 1996) have demonstrated contributions from both Asia and Near Oceania into Remote Oceanic gene pools. These studies also document gene flow from Southeast Asia into Near Oceanic gene pools, especially in areas where Austronesian languages are spoken. Polynesians and Micronesians generally share 70% of their nuclear alleles with Asians and 30% with Near Oceanic, Papuan-speaking Melanesians (Serjeantson 1985; Martinson 1996). Thus, nuclear genetic studies reveal complex patterns of gene flow consistent with the entangled bank model or, alternatively, a "slow train" to Remote Oceania.

The contrasting patterns of genetic relationships described in previous studies may reflect differences in the populations and individuals sampled, differences in the effective population sizes of the genomes sampled, or sex-biased gene flow. To control the former possibility, we have generated and analyzed mtDNA and nuclear short tandem-repeat (STR) data from 745 individuals representing 28 Pacific Island and Asian populations (fig. 1). To explore the latter possibilities, we have examined genetic diversity within and between populations and

correlations among genetic, linguistic, and geographic distances within four sets of populations representing potential geographic and cultural spheres of interaction.

Subjects and Methods

This study was approved by the Committee for Compliance on Experimentation with Human Subjects, University of Hawaii at Manoa; the Human Subjects Research Review Committee, Federated States of Micronesia; and the traditional leaders of Palau (Rubekul Belau) and of Yap State (Council of Pilung). Five plucked hairs were collected from each of 467 consenting participants. In many areas of Micronesia, informed consent was obtained from participants, with the assistance of the local hospital or clinic staff. An additional 128 blood samples from Micronesia and Vanuatu had previously been collected as part of surveys of hemoglobinopathies in Oceania (O'Shaughnessy et al. 1990; Ganczakowski et al. 1995). The remaining 150 samples were obtained from other researchers.

Genomic DNA was extracted from plucked-hair samples by the method of Higuchi (1989) or by the silica-extraction protocol of Boom et al. (1990), abbreviated as follows. Five hairs were washed with ethanol and rinsed with sterile distilled water, and the bulb ends were clipped, placed in a sterile tube containing 200 ml of lysis buffer 6 (Boom et al. 1990), and incubated at 60°C for 2 h. After centrifugation, the supernatant was transferred to a new tube, 20 ml of the silica solution was added, and the tube was vortexed. The sample was incubated for 10 min and then centrifuged, to pellet the silica. The pellet was washed twice with 500 ml of 70% ethanol and once with 500 ml of acetone and was then loosely covered with foil and dried at 56°C. Genomic DNA was eluted from the dried pellet with 150 ml of sterile water.

One milliliter of the elutant was used to PCR amplify, with primers L15996 (5′-CTC CAC CAT TAG CAC CCA AAG C-3′) and H16401 (5′-TGA TTT CAC GGA GGA TGG TG-3′), an ~400-bp fragment of the mtDNA hypervariable I segment of the control region. Single-stranded template was obtained either by a second PCR reaction with asymmetric primer ratios (Vigilant et al. 1989) or by magnetic bead (Dynal) separation of a biotin-labeled strand. Purified single strands were then sequenced directly by use of ^{35}S-labeled dATP (Amersham) and the Sequenase kit (U.S. Biochemical). One hundred ninety bases of sequence (16184–16373) from each individual were compared for these analyses (available from Genbank).

The genomic DNA extracted from plucked hairs was initially insufficient for nuclear–STR loci genotyping. Whole-genome amplification of these extracts was performed with fully degenerate pentadecamers in the primer-extension preamplification reaction (Zhang et al. 1992). Subsequently, these samples were used as templates for multiplexed PCR reactions with fluorescently labeled primers for 7 and 10 STR loci. These reactions were performed in 10-ml volumes including Taqstart (Clonetech) bound *Taq* polymerase (Boehringer). STR alleles were then separated on an ABI 377 automated sequencer and sized by the Genotyper software (Applied Biosystems). The 17 nuclear STR loci examined are unlinked. Genome Database identification numbers for the STR markers are D1S404, D1S407, D3S1545, D3S2322, D4S1527, D4S1530, D5S1347, D5S612, D6S400, D6S942, D7S623, D9S762, D10S525, D12S297, D13S252, D20S161, and D22S417. These loci are a subset of those examined in previous studies (Jorde et al. 1995; Jorde et al. 1997). The frequencies of the 181 alleles observed, within each of the 28 populations, at these 17 loci are available from the authors, in electronic form.

Data were analyzed for four sets of populations representing potential spheres of interaction. Recent experimental (Finney 1994) and simulated (Irwin 1992) voyages have highlighted the feasibility of extensive gene flow throughout the Pacific, questioning long-held assumptions of isolation (Rehg 1995; Terrell et al. 1997). Therefore, we have grouped our populations into geographically and linguistically related subsets, to examine gene flow within potential interaction spheres. These subsets are 6 Oceanic Austronesian-speaking Polynesian and Melanesian populations; 8 Oceanic Austronesian-speaking Micronesian populations; 21 Austronesian-speaking populations; and all 28 populations.

Nucleotide diversity for the mitochondrial sequence was measured as $[n/(n-1)] \sum x_i x_j \pi_{ij}$, where n is the number of individuals, x_i is the frequency of the ith mtDNA haplotype in the population, and π_{ij} is the proportion of nucleotides that differ between the ith and jth mtDNA haplotypes (Nei 1987). The standard error of this estimate was calculated with equation 10.7 of Nei (1987). Allele frequencies for each STR system were estimated directly by gene counting. Heterozygosity for each system was estimated as $1 - \sum x_i^2$, where x_i is the estimated frequency of the ith allele in the system. Standard errors were obtained by use of equation 8.7 of Nei (1987). Among geographically isolated populations, the amount of genetic diversity within groups with lower migration rates is expected to be lower than the amount within groups with higher migration rates (Tajima 1990). The opposite is expected of between-group diversity. The proportion of genetic variance attributable to population subdivision was estimated with the G_{ST} statistic (Wright 1965; Nei 1987). The G_{ST} statistic is a measure of genetic differentiation and is inversely proportional to gene flow.

For the mtDNA sequence data, Kimura's (1980) two-parameter model was used to estimate, with the program

DNADIST (Felsenstein 1993), nucleotide diversity between each pair of individuals. The average diversity within and between populations was then estimated by use of equation 10.21 of Nei (1987). For the nuclear STR loci, we estimated genetic distances between pairs of populations, using the method of Shriver et al. (1995), which weights distances by the number of repeat-unit differences and thus assumes a stepwise-mutation model.

The genetic relationships among populations were depicted by principal-components (PC) analysis (Lalouel 1973). This technique summarized the 28-dimensional distance matrices into 2-dimensional PC maps that retained most of the genetic variance.

Linguistic comparisons were restricted to Austronesian-speaking populations because there is no consensus on the relationships among language families. Twenty-one of the populations speak Austronesian languages. The relationships among the languages spoken by these populations, based on exclusively shared innovations and shared cognate frequencies (Bender 1971; Pawley and Green 1973; Jackson 1983; Jackson 1986; Pawley and Ross 1993), are depicted in figure 2. Yapese is considered an Oceanic language (Ross 1996). Quantitative linguistic distances among these 21 populations were estimated by assigning relative branch lengths to the qualitative phylogeny. Note that the two western Micronesian populations of Palau and the Marianas speak languages only distantly related to each other and to the Oceanic Austronesian languages spoken by other Pacific Islanders.

We constructed a geographic-distance matrix between populations, using great-circle distances calculated from their latitudinal and longitudinal coordinates. Congruence among genetic-, linguistic-, and geographic-distance matrices was evaluated by means of the Mantel permutation method, extended to examine partial correlations of multiple matrices (Mantel 1967; Smouse and Long 1992) by use of Matrix Correlation Analysis version 1.0 (Long 1996). Partial correlation analysis allows one to determine whether a given correlation between two variables results from a correlation of both to a third variable. To determine significance levels, the observed correlations were compared with null distributions generated from 10,000 permutations.

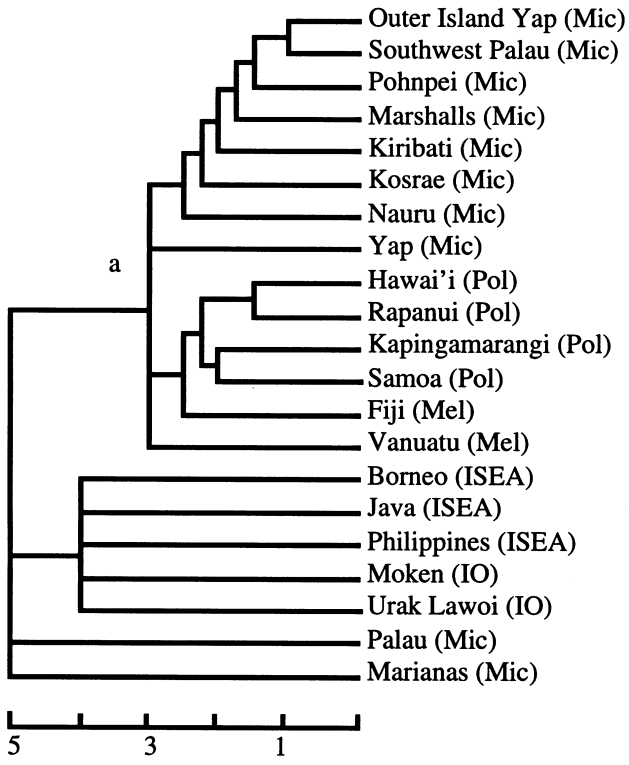

Figure 2 Qualitative linguistic relationships among the 21 Austronesian-speaking populations. The Oceanic branch of the Austronesian language family ("a") contains the eight Micronesian and the six Polynesian and Melanesian populations examined as potential spheres of interaction. The geographic region of each population is shown in parentheses. Mic = Micronesia; Pol = Polynesia; Mel = Melanesia; IO = Indian Ocean; and ISEA = island Southeast Asia (see table 1). Quantified linguistic distances between populations were estimated by assigning relative lengths to this tree. For example, the linguistic distance between the Marianas and any other population is estimated to be 10 units, and the linguistic distance between Vanuatu and any other Oceanic Austronesian-speaking population is estimated to be 6 units.

Results

Gene Diversities within Populations

The gene diversities for each population are given in table 1. The mtDNA and STR diversities are presented in two sets of columns ranked in ascending order. The numbers to the left of the population names correspond to the mtDNA diversities and the geographic locations in figure 1. The mtDNA diversities (0.017 ± 0.009) exhibit more variance than do the STR diversities (0.662 ± 0.040). There are three populations that have mtDNA and STR diversities >1 SD below the mean (<0.008 and <0.622, respectively). All three of these populations are Oceanic Austronesian speakers from Polynesia and Micronesia. The populations from Rapanui and Kapingamarangi are among the most isolated Polynesians, the latter being geographically isolated within Micronesia (see fig. 1). The population from the southwest islands of Palau is ecologically marginal and geographically isolated. Vietnam is the only population with both mtDNA and STR diversities >1 SD above the mean (>0.026 and >0.702, respectively). These results indicate a general loss of both mtDNA and STR genetic diversity in geographically isolated Remote Oceanic populations relative to other populations, consistent with low rates

Table 1

Gene Diversities within Pacific Island and Asian Populations

Population	*n*	Region[a]	Language[b]	*s*[c]	mtDNA Diversity (± SD)	Population	*k*[d]	STR Diversity (± SD)
1 Hawai'i	17	Pol	OCAN	0	.000 ± .000[e]	4 Southwest Palau	3.6	.559 ± .051[e]
2 Rapanui	10	Pol	OCAN	1	.001 ± .001[e]	2 Rapanui	4.2	.572 ± .046[e]
3 Kapingamarangi	27	Pol	OCAN	1	.002 ± .000[e]	28 Australia	4.0	.588 ± .051[e]
4 Southwest Palau	7	Mic	OCAN	2	.006 ± .001[e]	3 Kapingamarangi	5.1	.612 ± .041[e]
5 Marshalls	25	Mic	OCAN	9	.008 ± .002	1 Hawai'i	5.8	.643 ± .043
6 Moken	8	IO	AN	7	.009 ± .007	21 Korea	4.7	.643 ± .038
7 Marianas	48	Mic	AN	13	.010 ± .002	8 Urak Lawoi	4.3	.645 ± .042
8 Urak Lawoi	8	IO	AN	6	.011 ± .002	26 Papua New Guinea	4.5	.647 ± .032
9 Samoa	19	Pol	OCAN	13	.012 ± .004	22 Kiribati	4.6	.647 ± .028
10 Outer Island Yap	126	Mic	OCAN	16	.013 ± .002	6 Moken	4.2	.650 ± .030
11 Nauru	25	Mic	OCAN	13	.013 ± .004	27 China	6.2	.658 ± .046
12 Yap	59	Mic	OCAN	10	.015 ± .001	14 Pohnpei	5.6	.664 ± .035
13 Kosrae	25	Mic	OCAN	11	.018 ± .003	18 Borneo	4.6	.666 ± .037
14 Pohnpei	22	Mic	OCAN	13	.021 ± .002	10 Outer Island Yap	7.0	.670 ± .033
15 Thailand	9	MA	NAN	13	.021 ± .003	11 Nauru	5.2	.671 ± .029
16 Vanuatu	24	Mel	OCAN	18	.022 ± .003	5 Marshalls	5.5	.672 ± .029
17 Palau	115	Mic	AN	25	.023 ± .001	16 Vanuatu	5.7	.677 ± .038
18 Borneo	8	ISEA	AN	13	.023 ± .003	9 Samoa	6.2	.677 ± .030
19 Philippines	22	ISEA	AN	22	.023 ± .002	12 Yap	6.3	.681 ± .029
20 Fiji	13	Mel	OCAN	15	.023 ± .005	15 Thailand	5.1	.684 ± .032
21 Korea	7	MA	NAN	11	.024 ± .005	20 Fiji	5.6	.685 ± .033
22 Kiribati	17	Mic	OCAN	16	.024 ± .004	13 Kosrae	5.5	.685 ± .028
23 Japan	27	MA	NAN	30	.025 ± .002	19 Philippines	5.8	.696 ± .025
24 Vietnam	22	MA	NAN	24	.027 ± .002[f]	25 Java	6.0	.699 ± .034
25 Java	19	ISEA	AN	25	.028 ± .002[f]	23 Japan	6.4	.705 ± .032[f]
26 Papua New Guinea	11	Mel	NAN	16	.028 ± .003[f]	7 Marianas	7.0	.706 ± .028[f]
27 China	20	MA	NAN	28	.029 ± .003[f]	17 Palau	7.6	.707 ± .031[f]
28 Australia	5	Aus	NAN	14	.031 ± .008[f]	24 Vietnam	6.9	.720 ± .032[f]
Total	745				.017 ± .009[g]			.662 ± .040[g]

[a] Pol = Polynesia; Mic = Micronesia; Mel = Melanesia; IO = Indian Ocean; ISEA = island Southeast Asia; MA = mainland Asia; and Aus = Australia.

[b] OCAN = Oceanic Austronesian; AN = Austronesian; and NAN = non-Austronesian.

[c] No. of polymorphic sites.

[d] Average no. of alleles at each locus.

[e] Diversity <1 SD from the mean.

[f] Diversity >1 SD from the mean.

[g] Mean.

of gene flow into these populations (Tajima 1990). The more-extreme reduction of mitochondrial diversity may reflect the lingering effects of a recent population bottleneck and a paucity of female gene flow in the Pacific.

There are a number of Pacific Island populations that have both high STR diversities and average numbers of alleles at each locus (*k*). Palau and the Marianas in particular have some of the highest STR diversities (>0.7) and *k* values (>6.9) and are geographically intermediate between other Pacific Islanders and Asian populations (fig. 1). Similarly, Fiji, Yap, Samoa, and Vanuatu have relatively high STR diversities and appear geographically between other Remote Oceanic Islanders and Near Oceania (fig. 1).

PC Maps

Figure 3 displays the first two eigenvectors extracted from the 28-dimensional mtDNA-distance matrix. Together, these two dimensions account for 72% of the variance of the original matrix. The first dimension (*X*-axis) accounts for 63% of the genetic variance and separates most Remote Oceanic Pacific Islanders from other populations. The second dimension (*Y*-axis) accounts for 9% of the genetic variance and primarily distinguishes the Moken from all other populations.

Fourteen of the Remote Oceanic Pacific Island populations, including 13 of the 14 Micronesian and Polynesian populations examined, cluster together and essentially form a straight line out to the peripheral Polynesian populations of Hawaii and Rapanui. The population with the highest affinity to these populations is the Philippines, followed by China, Vietnam, Borneo, and Java. Thus, all Polynesians and most Micronesians studied cluster together and are most closely related to populations from island and Southeast Asia, consistent with previous mtDNA studies (Hertzberg et al. 1989;

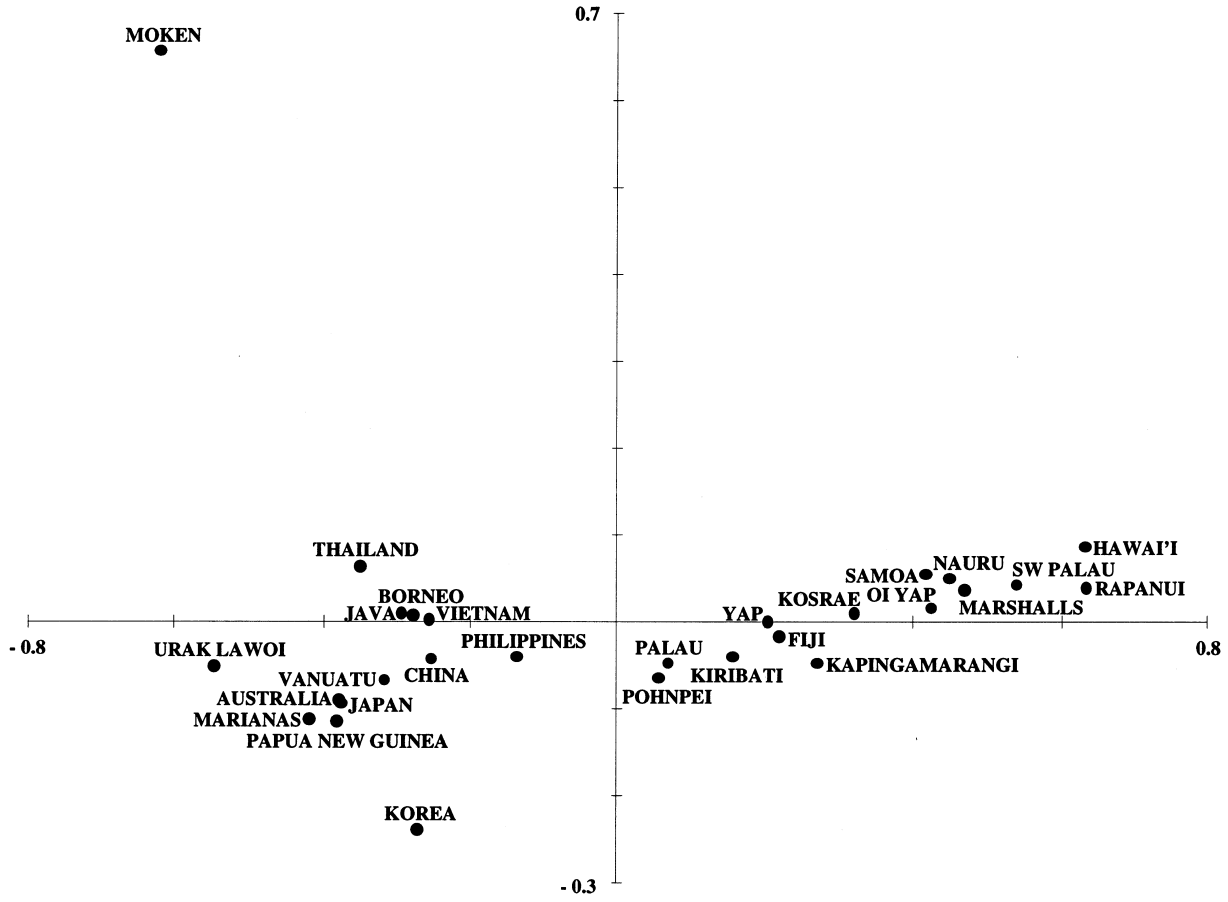

Figure 3 PC map based on mtDNA genetic distances. The first (*X*-axis) and second (*Y*-axis) eigenvectors summarize 63% and 9% of the genetic variance, respectively. OI Yap = Outer Island Yap; and SW Palau = Southwest Palau.

Lum et al. 1994; Redd et al. 1995; Sykes et al. 1995; Lum and Cann 1998) and with comparative linguistics (see fig. 2). The exception to this pattern is the western Micronesian population of the Marianas, which appears near Australia, Japan, highland Papua New Guinea, and Vanuatu.

The three Melanesian populations examined appear in two distinct places on the mtDNA PC map. Fiji clusters with Micronesian and Polynesian populations, whereas Vanuatu and Papua New Guinea appear to be distinct from other Pacific Islanders. The latter grouping is contrary to linguistic classification, which includes both Vanuatu and Fiji with Micronesians and Polynesians within the Oceanic branch of the Austronesian language family (see fig. 2). Note that Vanuatu is the only Oceanic Austronesian-speaking population that does not appear in the Remote Oceanic cluster described above. These results suggest high maternal gene flow into Vanuatu from Near Oceania, across linguistic boundaries, consistent with previous studies within Melanesia (Serjeantson et al. 1983; Welsch et al. 1992).

The first two eigenvectors extracted from the STR genetic-distance matrix account for 54% of the genetic variance and are presented in figure 4. The lower amount of variance summarized in two dimensions indicates a more complex pattern relative to the mtDNA data. The first dimension of this PC map (*X*-axis) accounts for 31% of the genetic variance and separates all Asian populations from Near and Remote Oceanic populations (except the Marianas). In general, Remote Oceanic Pacific Island populations appear intermediate between populations from Asia and highland Papua New Guinea. Interestingly, the population of Aboriginal Australians is found within the Pacific Island cluster. Also, many of the Pacific Island populations with high STR diversities appear close to Asia and Near Oceania on the STR PC map. Note that the western Micronesian populations from Palau and the Marianas are closest to Asian populations, whereas Fiji, Vanuatu, Pohnpei, Samoa, and Yap are closest to Papua New Guinea, in general agreement with their geographic locations (fig. 1). Thus, in contrast to results from mtDNA data, STR data reveal

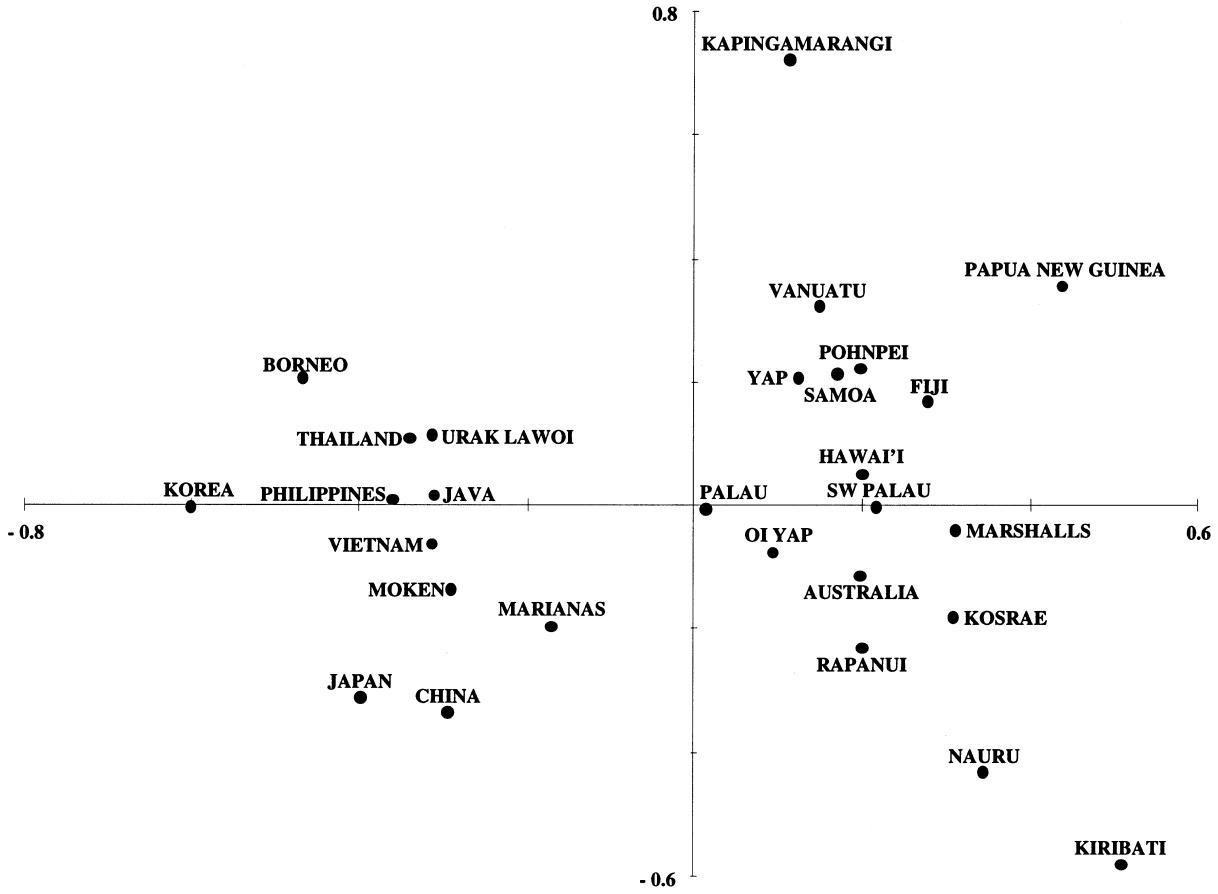

Figure 4 PC map based on STR genetic distances. The first (*X*-axis) and second (*Y*-axis) eigenvectors summarize 31% and 23% of the genetic variance, respectively. OI Yap = Outer Island Yap; and SW Palau = Southwest Palau.

that Micronesians and Polynesians have had substantial contributions to their nuclear genome from both Asian and Near Oceanic populations, consistent with previous nuclear genetic (Serjeantson 1985; O'Shaughnessy et al. 1990; Martinson 1996; Roberts-Thomson 1996) and craniometric (Pietrusewsky 1990) analyses.

G_{ST} Estimates

Table 2 lists the mtDNA- and STR-based G_{ST} estimates for each of the four potential spheres of interaction. In three of the four data sets, G_{ST} estimates from mtDNA sequences vary between 0.36 and 0.39, indicating that 36%–39% of the genetic variation is attributable to between-population differences. These values fall within the range of previous G_{ST} estimates based on mtDNA restriction–site data (0.31–0.41) (Stoneking et al. 1990; Merriwether et al. 1991) but are nearly twice that based on hypervariable II control-region sequences from worldwide populations (0.20) (Jorde et al. 1995).

G_{ST} estimates from STR loci vary between 0.05 and 0.08. Like the mtDNA G_{ST} estimates, these values are about twice that estimated from worldwide populations (0.03) (Jorde et al. 1995). In three of the four comparisons, the STR G_{ST} estimates are about fivefold smaller than the corresponding mtDNA G_{ST} estimates. These STR G_{ST} estimates do not take allele size into account and thus may actually overestimate the amount of population structure.

The lowest G_{ST} estimates from both mtDNAs (0.13) and STRs (0.05) are found in the Oceanic Austronesian Micronesian subset. These results indicate high gene flow throughout Micronesia, consistent with ethnographic, archaeological, and linguistic data. Micronesia consists of extensive arcs of low atolls interspersed with a few high islands. The inhabitants of the atoll systems have developed and maintained complex navigational techniques to reduce environmental vulnerability (Ridgell et al. 1994). Many of the atoll societies participated in extensive trade networks with each other and with the high islands throughout prehistory, as evidenced by pottery exchange (Intoh 1997) and by the current distribution of languages (Marck 1986). It appears that

Table 2

G_{ST} Estimates among Sets of Populations

G_{ST} Estimate	Polynesians and Austronesian Melanesians	Oceanic Austronesian Micronesians	All Austronesians	All Populations
mtDNA	.385	.130	.387	.363
STR	.063	.053	.071	.076

both linguistic divergence and biparental genetic affinities within central and eastern Micronesia reflect exchange and male gene flow throughout prehistory (Lum, in press).

Mantel Correlations

Table 3 displays the results from the pairwise and partial correlation analyses for each of the four potential spheres of interaction. A low but significant correlation ($r = .25$, $P < .01$) is observed between mtDNA and STR genetic distances when all 28 populations are compared but not in the other comparisons.

For each pair of distance matrices compared, each of the other distance matrices is controlled, to identify potentially complex patterns of covariance. Such a pattern is observed in one of the sets of comparisons. Geographic and linguistic distances among all Austronesian-speaking populations are significantly correlated ($r = .28$, $P < .05$). When mtDNA distances are controlled, however, the correlation nearly disappears and is no longer significant ($r = .09$). In contrast, note that correlations between mtDNA and both geography and language remain significant ($P < .05$) when the other matrices are controlled. Thus, partial correlations reveal that the correlation between geography and language among this set of populations results from correlations between mtDNA and both geography and language.

The highest correlation observed is between mtDNA and linguistic distances ($r = .84$, $P < .01$). In contrast, the STR genetic and linguistic distances are not significantly correlated in any of the comparisons. These results suggest that patterns of mtDNA affinities may retain the signature of initial settlement, whereas STR genetic patterns may reflect subsequent male-biased gene flow.

Geographic distances are not significantly correlated with both STR and mtDNA genetic distances in any case. Geographic and STR genetic distances are correlated ($r > .6$) in both the Polynesian and Melanesian (marginally significant, $P < .06$) and Oceanic Austronesian Micronesian ($P < .01$) comparisons but not in the comparisons of larger sets of populations. These results indicate that, within the Pacific Island sets of populations, gene flow may be male biased and geographically restricted (Slatkin 1993). In contrast, geographic distances are significantly correlated with mtDNA genetic distances only in comparisons including Asian populations ($r > .4$, $P < .05$). This significant but moderate correlation suggests that the correspondence reflects the genetic clustering of Pacific Islanders relative to other populations (see fig. 3) but not among themselves. Moreover, the discordance between mtDNA genetic and geographic distances within Pacific Island sets of populations suggests that mitochondrial gene flow has been low and that the expansion into Remote Oceania has been recent.

G_{ST} and Mantel Correlations

Here we relate G_{ST} estimates to nongenetic characteristics. Significant correlations ($r > .5$, $P < .05$) between maternal genetic and linguistic distances are observed in two of the three comparisons involving exclusively Austronesian-speaking populations. In these two cases, mtDNA G_{ST} estimates are relatively high (>0.38), indicating low within-group variance and substantial genetic isolation. In the Oceanic Austronesian Micronesian comparison, the mtDNA G_{ST} estimate is 0.13, indicating relatively high gene flow, and the correlation with linguistic distances is negative ($r = -.28$). These results indicate that genetic and linguistic distances may coevolve and that their correspondence may be preserved under conditions of genetic isolation.

A significant correlation ($r = .65$, $P < .01$) between biparental genetic and geographic distances is observed in the Oceanic Austronesian Micronesian subset. This correlation is coincident with a low STR G_{ST} estimate (0.05). In the comparisons including Asian populations, the G_{ST} estimates are >0.07, and the correlation between STR genetic and geographic distances is nonsignificant and low ($r < .20$). The G_{ST} estimate for the Polynesian and Melanesian data set is intermediate (0.06), and the correlation between STR genetic and geographic distances is marginally significant ($r = .64$, $P < .06$). Thus, a trend is suggested; lower STR G_{ST} estimates are coincident with higher correlations between STR and geographic distances. These results suggest that isolation by distance increases as nuclear gene flow increases, as predicted by simulations (Slatkin 1993).

Table 3

Mantel Correlations among Genetic, Geographic, and Linguistic Distances

Pairwise and Partial Correlations[a]	Polynesians and Austronesian Melanesians	Oceanic Austronesian Micronesians	All Austronesians	All Populations
mtDNA × STR	.167	−.092	.216	.252**
mtDNA × STR (geography)	.137	−.102	.152	.196*
mtDNA × STR (language)	.316	−.062	.131	
mtDNA × geography	.097	−.022	.419**	.430***
mtDNA × geography (STR)	−.014	.049	.393*	.404***
mtDNA × geography (language)	.591	−.021	.331*	
STR × geography	.643	.647**	.193	.182
STR × geography (mtDNA)	.639	.648**	.116	.085
STR × geography (language)	.668	.651**	.144	
mtDNA × language	.748*	−.280	.518***	
mtDNA × language (STR)	.769*	−.273	.495***	
mtDNA × language (geography)	.843**	−.280	.458***	
STR × language	−.056	.116	.205	
STR × language (mtDNA)	−.277	.094	.112	
STR × language (geography)	.245	.146	.160	
Geography × language	−.360	.007	.284*	
Geography × language (mtDNA)	−.654	.001	.087	
Geography × language (STR)	−.423	−.090	.255*	

[a] Distance in parentheses is controlled.

* $P < .05$.

** $P < .01$.

*** $P < .001$.

Discussion

Caveats

Since both sets of genetic data were determined from the same individuals, there is an a priori expectation of correspondence. The two genetic systems examined, however, have different mutation rates, different effective population sizes, and different inheritance patterns, any of which can potentially lead to discordance. Phylogenetic studies indicate a mutation rate of 3×10^{-6} substitutions per site per generation for the mtDNA control region (Sherry et al. 1994). Recent estimates based on extensive pedigrees indicate that the mutation rate may be as high as 5×10^{-5} substitutions per site per generation (Howell et al. 1996; Parsons et al. 1997). These two estimates most likely refer to mitochondrial polymorphism fixation and mutation rates, respectively. STR mutation rates have been estimated at 10^{-4} to 10^{-3} (Edwards 1992; Weber and Wong 1993). Thus, it appears that the STR mutation rate is as much as 1,000-fold higher than the mtDNA mutation rate. The potential difference in mutation rates is probably not very important for our purposes, however, since the settlement time of Remote Oceania is <4,000 years BP, and most genetic differences between these island populations are likely due to differential sampling of pre-existing variation rather than to the generation of novel variation in situ.

The fourfold-smaller effective population size of the mitochondrial genome relative to the nuclear genome enhances the effect of genetic drift (Birky et al. 1983). The effect of genetic drift on the mtDNA genome would be even more extreme if colonizations of Pacific Islands were achieved by maternally related kin groups. Colonization by maternally related clans, common throughout Micronesia, may preserve most nuclear variation but little mtDNA variation (Wade et al. 1994). Thus, the nuclear genome is expected to be less affected by and to rebound faster from a bottleneck than is the mitochondrial genome.

Extensive sex-biased gene flow may also be a source of discordance. Since mtDNA is maternally inherited, mtDNA data reflect female gene flow. Nuclear STR loci, in contrast, are biparentally inherited and reflect both female and male gene flow. We expect that dispersal events that resulted in successful colonization differed in sex ratio from those involved in trading, warfare, and perhaps exploration. Thus, extensive sex-biased, predominantly male gene flow could create discordance by altering genetic relationships based on nuclear STR loci but not on mtDNA.

Genetic and linguistic distances are expected to coevolve under conditions of isolation, with both patterns and amounts of variation reflecting the same history of population fissions and fusions (Smouse and Long 1992). Thus, a discordance between genetic and linguistic distances may result from gene flow across linguistic boundaries. Significant correlations between ge-

netic and geographic distances indicate isolation by distance (Wright 1943). Such an accumulation of genetic differences under geographically restricted dispersal is expected to be detectable only when a population is near or at equilibrium under current patterns of gene flow (Slatkin 1993). In contrast, a lack of congruence between genetic and geographic distances suggests a recent colonization event.

The mtDNA and STR genetic distances depict different patterns of relationships among Pacific Island and Asian populations (see figs. 3 and 4). The major difference between these patterns is the substantial affinities between Remote Oceanic Islanders and Near Oceanic populations from highland Papua New Guinea and Australia, observed in the STR data but not in the mtDNA data. We examined genetic diversity, gene flow, and correlations among genetic, linguistic, and geographic distances within four sets of populations representing potential spheres of interaction, to evaluate whether the differences observed were due to differences in effective population sizes or to sex-biased gene flow. In particular, we wondered whether the two genetic patterns reflect settlement or postcolonization gene flow. Several results are instructive on this point. mtDNA genetic diversity within populations is reduced, especially within geographically marginal populations. This is consistent with a recent bottleneck and a paucity of female gene flow. In cases where mtDNA G_{ST} estimates are high (>0.38), we observe significant correlations between mitochondrial and linguistic distances ($r > .5$, $P < .05$). These results suggest that a paucity of female gene flow may have both preserved settlement patterns in mtDNAs and maintained correlations between mitochondrial and linguistic distances. As the STR G_{ST} estimates decrease, the correlations between nuclear genetic and geographic distances increase. In addition, Remote Oceanic Islanders geographically close to Near Oceania and island Southeast Asia have relatively high levels of STR diversity, consistent with higher migration rates. These results indicate geographically restricted male-biased gene flow at or near equilibrium, suggesting that the current nuclear genetic pattern may reflect postcolonization gene flow rather than initial settlement. Our inferences of male-biased gene flow, based on the differences between maternally and biparentally inherited genetic markers, provide testable predictions of Y-chromosome affinities. We are currently collecting Y-chromosome data that we expect will confirm the results of our present study.

The Express Train and the Entangled Bank

There is general agreement about the dispersal of Pacific Islanders throughout Remote Oceania, but their origin west of Melanesia remains controversial. Based largely on linguistic evidence and interpretations of archaeology, the prevailing view depicts island Southeast Asia as the origin of Remote Oceanic Islanders (Diamond 1988). The extreme version of this view describes this colonization as an express train that leaves southern China and transports its occupants through Near Oceania and into Remote Oceania without stopping to pick up any additional passengers. A competing view, championed most enthusiastically by Terrell (1988), sees the colonization of Remote Oceania as an expansion of Near Oceanic populations through long-term interactions with other western Pacific populations. These complex, long-term interactions, described metaphorically as an entangled bank, gave rise to the cultural innovations characteristic of Remote Oceania (Terrell 1988; Terrell et al. 1997). This view is supported by studies in Melanesia that document a general lack of congruence between genetic and linguistic patterns throughout the region (Serjeantson et al. 1983; Welsch et al. 1992). Our present analyses indicate that these views may not be incompatible.

By examining different genetic systems from the same individuals, we have generated patterns consistent with both views. As described above, our mtDNA data are correlated with linguistic data and suggest island Southeast Asia as the origin of Remote Oceanic Islanders. These data are consistent with the express train model. Our STR data, in contrast, are not correlated with linguistic data and highlight affinities between Near Oceanic and Remote Oceanic populations. We have argued that the differences between these patterns result from postcolonization male-biased gene flow. Genetic interactions between populations after initial colonization may have been mediated by a predominantly male segment of voyaging societies, engaged in the control of resources. This bias served to preserve pre-existing linguistic differences, lines of status, and hierarchical divisions among matrilineal kinship groups. Thus, we see female settlement as an express train and male gene flow as an entangled bank.

Acknowledgments

We thank all the people who made this study possible by donating DNA samples. We also express our appreciation to those who facilitated the collection of samples: P. Acevedo, D. Addison, M. Andrew, S. Auerbach, R. K. Blaisdell, C. Ching, J. Clegg, G. Dever, A. di Piazza, J. Filefney, S. George, Y. Gibbons, J. Gilmatam, J. Haleyalmang, G. Heathcote, N. Hinshiranan, L. Humphreys, T. Hunt, Y. Hsia, S. Izutsu, A. Kugfas, M. Mesngon, J. Mooteb, D. Otobed, N. Patrick, A. Polloi, E. Petrick, R. Puri, F. Rehuher, J. Scott, S. Serjeantson, D. Sol, J. Winter, and J. Wozniak. We also thank R. Blust and K. Rehg, for their insights on Austronesian linguistics, and M. Bamshad, G. Heathcote, and two anonymous reviewers for comments on earlier versions of this manuscript. This work was supported by Wenner-Gren Foundation for Anthropological Re-

search grant Gr. 5640, by National Science Foundation (NSF)/ Sloan grant BIR-9510748, and by NSF grants SBR-9514733 and SBR-9512178.

Electronic-Database Information

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

GenBank, http://www.ncbi.nlm.nih.gov/Web/Genbank (for 711 sequences, accession numbers AF066083–AF066182, AF066184–AF066383, and AF066385–AF066795; the remaining 34 sequences were reported by Lum et al. [1994])
Genome Database, http://www.gdb.org (for STR markers used)

References

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

Bender BW (1971) Micronesian languages. In: Sebeok TA (ed) Current trends in linguistics. Mouton, The Hague, pp 426–465

Birky CW, Muruyama T, Fuerst P (1983) An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts, and some results. Genetics 103: 513–527

Boom R, Sol CJA, Salimans MMM, Jansen CL, Wertheim-Van Dillen PME, Van Der Noordaa J (1990) Rapid and simple method for purification of nucleic acids. J Clin Microbiol 28:495–503

Darwin C (1859) [1967] On the origin of species by means of natural selection. Atheneum, New York

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

Edwards A, Hammond HA, Jin L, Caskey CT, Chakraborty R (1992) Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. Genomics 12:241–253

Felsenstein J (1993) PHYLIP (Phylogeny Inference Package) version 3.5c. Department of Genetics, University of Washington, Seattle (distributed by the author)

Finney B (1994) Voyage of rediscovery: a cultural odyssey through Polynesia. University of California Press, Berkeley

Ganczakowski M, Bowden DK, Maitland K, Williams TN, O'Shaughnessy D, Viji J, Lucassen A, et al (1995) Thalassaemia in Vanuatu, south-west Pacific: frequency and haematological phenotypes of young children. Br J Haematol 89:485–495

Green RC (1991) Near and remote Oceania: disestablishing "Melanesia" in cultural history. In: Pawley A (ed) Man and a half: essays in Pacific anthropology and ethnobotony in honour of Ralph Bulmer. The Polynesian Society, Auckland, pp 491–502

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

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

Higuchi R (1989) Simple and rapid preparation of samples for PCR. In: Erlich HA (ed) PCR technology. Stockton Press, New York, pp 31–43

Howell N, Kubacka I, Mackey DA (1996) How rapidly does the human mitochondrial genome evolve? Am J Hum Genet 59:501–509

Intoh M (1997) Human dispersal into Micronesia. Anthropol Sci 105:15–28

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

Jackson FH (1983) The internal and external relationships of the Trukic languages of Micronesia. PhD thesis, University of Hawaii at Manoa, Honolulu

——— (1986) On determining the external relationships of the Micronesian languages. In: Geraghty PL, Wurm SA (eds) FOCAL II: papers from the Fourth International Conference on Austronesian Linguistics. Pacific Linguistics C-94, Australian National University, Canberra, pp 201–238

Jorde LB, Bamshad MJ, Watkins WS, Zenger R, Fraley AE, Krakowiak PA, Carpenter KD, et al (1995) Origins and affinities of modern humans: a comparison of mitochondrial and nuclear genetic data. Am J Hum Genet 57:523–538

Jorde LB, Rogers AR, Bamshad M, Watkins WS, Krakowiak P, Sung S, Kere J, et al (1997) Microsatellite diversity and the demographic history of modern humans. Proc Natl Acad Sci USA 94:3100–3103

Kimura M (1980) A simple model for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120

Lalouel JM (1973) Topology of population structure. In: Morton NE (ed) Genetic structure of populations. University Press of Hawaii, Honolulu, pp 139–152

Long JC (1996) Matrix Correlation Analysis, version 1.0. Section on Population Genetics and Linkage, Laboratory of Neurogenetics, National Institute of Alcohol and Alcohol Abuse, NIH, Rockville, Maryland (distributed by the author)

Lum JK. Central and eastern Micronesia: genetics, the overnight voyage, and linguistic divergence. Man and Culture in Oceania (in press)

Lum JK, Cann RL (1998) mtDNA and language support a common origin of Micronesians and Polynesians in Island Southeast Asia. Am J Phys Anthropol 105:109–119

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

Mantel N (1967) The detection of disease clustering and a generalized regression approach. Cancer Res 27:209–220

Marck JC (1986) Micronesian dialects and the overnight voyage. J Polynesian Soc 95:253–258

Martinson JJ (1996) Molecular perspectives on the colonization of the Pacific. In: Boyce AJ, Macie-Taylor CGN (eds) Molecular biology and human diversity. Cambridge University Press, London, pp 171–195

Merriwether DA, Clark AG, Ballinger SW, Schurr TG, Soodyall H, Jenkins T, Sherry ST, et al (1991) The structure of human mitochondrial DNA variation. J Mol Evol 33: 543–555

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

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

Parsons TJ, Munciec DS, Sullivan K, Woodyatt N, Alliston-Greiner R, Wilson MR, Berry DL, et al (1997) A high observed substitution rate in the human mitochondrial DNA control region. Nat Genet 15:363–368

Pawley A, Green R (1973) Dating the dispersal of the Oceanic languages. Oceanic Linguistics 12:1–67

Pawley A, Ross M (1993) Austronesian historical linguistic and cultural history. Annu Rev Anthropol 22:425–459

Pietrusewsky M (1990) Craniofacial variation in Micronesia and the Pacific: a multivariate study. Micronesica Suppl 2: 373–402

Redd AJ, Takezaki N, Sherry ST, McGarvey ST, Sofro ASM, Stoneking M (1995) Evolutionary history of the COII/ $tRNA^{Lys}$ intergenic 9 base pair deletion in human mitochondrial DNAs from the Pacific. Mol Biol Evol 12:604–615

Rehg K (1995) The significance of linguistic interaction spheres in reconstructing Micronesian prehistory. Oceanic Linguistics 34:305–326

Ridgell R, Ikea M, Uruo I (1994) The persistence of Central Carolinian navigation. ISLA: J Micronesian Stud 2:181–206

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 Guinea Highlanders is supported by α-globin haplotype analysis. Am J Hum Genet 58:1017–1024

Ross M (1996) Is Yapese Oceanic? In: Nothofer B (ed) Reconstruction, classification, description: festschrift in honor of Isidore Dyen. Verlag Meyer, Hamburg, pp 121–166

Serjeantson SW (1985) Migration and admixture in the Pacific: insights provided by human leucocyte antigens. In: Kirk R, Szathmary E (eds) Out of Asia: peopling the Americas and the Pacific. The Journal of Pacific History Inc, Canberra, pp 133–145

Serjeantson SW, Kirk RL, Booth PB (1983) Linguistic and genetic differentiation in New Guinea. J Hum Evol 12:77–92

Sherry ST, Rogers AR, Harpending H, Soodyall H, Jenkins T, Stoneking M (1994) Mismatch distributions of mtDNA reveal recent human population expansions. Hum Biol 66: 761–775

Shriver MD, Jin L, Boerwinkle E, Deka R, Ferrell RE, Chakraborty R (1995) A novel measure of genetic distance for highly polymorphic tandem repeat loci. Mol Biol Evol 12: 914–920

Slatkin M (1993) Isolation by distance in equilibrium and nonequilibrium populations. Evolution 47:264–279

Smouse PE, Long JC (1992) Matrix correlation analysis in anthropology and genetics. Yearbook Phys Anthropol 35: 187–213

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

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 (1990) Relationship between migration and DNA polymorphism in a local population. Genetics 126:231–234

Terrell J (1988) History as a family tree, history as an entangled bank: constructing images and interpretations of prehistory in the South Pacific. Antiquity 62:642–657

Terrell JE, Hunt TL, Gosden C (1997) The dimensions of social life in the Pacific. Curr Anthropol 38:155–195

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

Wade MJ, McKnight ML, Shaffer HB (1994) The effect of kin-structured colonization on nuclear and cytoplasmic genetic diversity. Evolution 48:1114–1120

Weber JL, Wong C (1993) Mutation of human short tandem repeats. Hum Mol Genet 2:1123–1128

Welsch RL, J Terrell, Nadolski JA (1992) Language and culture on the north coast of New Guinea. Am Anthropol 94: 568–600

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

Wright S (1943) Isolation by distance. Genetics 28:114–138

——— (1965) The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution 19:395–420

Zhang L, Cui X, Schmitt K, Hubert R, Navidi W, Arnheim N (1992) Whole genome amplification from a single cell: implications for genetic analysis. Proc Natl Acad Sci USA 89: 5847–5851