Am. J. Hum. Genet. 63:1807–1823, 1998

Genetic Evidence for the Proto-Austronesian Homeland in Asia: mtDNA and Nuclear DNA Variation in Taiwanese Aboriginal Tribes

Terry Melton,[1,*] Stephanie Clifford,[1] Jeremy Martinson,[2] Mark Batzer,[3] and Mark Stoneking[1]

[1]Department of Anthropology, Pennsylvania State University, University Park; [2]Institute of Molecular Medicine, University of Oxford, Oxford; and [3]Department of Pathology, Louisiana State University Medical Center, New Orleans

Summary

Previous studies of mtDNA variation in indigenous Taiwanese populations have suggested that they held an ancestral position in the spread of mtDNAs throughout Southeast Asia and Oceania (Melton et al. 1995; Sykes et al. 1995), but the question of an absolute proto-Austronesian homeland remains. To search for Asian roots for indigenous Taiwanese populations, 28 mtDNAs representative of variation in four tribal groups (Ami, Atayal, Bunun, and Paiwan) were sequenced and were compared with each other and with mtDNAs from 25 other populations from Asia and Oceania. In addition, eight polymorphic *Alu* insertion loci were analyzed, to determine if the pattern of mtDNA variation is concordant with nuclear DNA variation. Tribal groups shared considerable mtDNA sequence identity ($P > .90$), where gene flow is believed to have been low, arguing for a common source or sources for the tribes. mtDNAs with a 9-bp deletion have considerable mainland-Asian diversity and have spread to Southeast Asia and Oceania through a Taiwanese bottleneck. Only four Taiwanese mtDNA haplotypes without the 9-bp deletion were shared with any other populations, but these shared types were widely dispersed geographically throughout mainland Asia. Phylogenetic and principal-component analyses of *Alu* loci were concordant with conclusions from the mtDNA analyses; overall, the results suggest that the Taiwanese have temporally deep roots, probably in central or south China, and have been isolated from other Asian populations in recent history.

Received July 24, 1997; accepted for publication September 30, 1998; electronically published December 1, 1998.

Address for correspondence and reprints: Dr. Terry Melton, Mitotyping Technologies, 1981 Pine Hall Drive, State College, PA 16801. E-mail: twm107@mitotyping.com

*Present affiliation: Mitotyping Technologies, State College, PA

0002-9297/98/6306-0028$02.00

Introduction

Taiwan has been occupied by Austronesian tribal groups whose continuous habitation predates third century A.D. historical records by an unknown length of time. These populations have been in conflict with waves of mainland Han Chinese invaders during much of the last two millennia and have been isolated from outsiders at least since routine movement of Han Chinese onto the island began in the eighteenth century (Kao 1958). The eight extant tribes regionally distributed throughout the rugged internal mountains and along the eastern coast (fig. 1) have individually distinct and sometimes mutually unintelligible languages and different material cultures and social organizations (Chai 1967). These facts have stimulated two primary lines of inquiry for anthropologists. First, what are the most likely temporal and geographic origins of Taiwanese aboriginal groups, which are especially significant because the Taiwanese have played a significant role in the prehistory of the Austronesian expansion throughout Oceania (Bellwood 1978, 1985, 1995; Melton et al. 1995; Sykes et al. 1995)? Second, with respect to intertribal differentiation, have these tribes shared a common origin and differentiated through geographic isolation reinforced by extreme topography, or could their differences reflect separate origins from diverse populations that arrived in successive waves to inhabit Taiwan?

There are several theories about Taiwanese origins, including that they are (1) descendants of Ryukyu Islanders who found their way to Taiwan >2,000 years ago, (2) Malayans who followed warm ocean currents from the south to the island, or (3) pioneers from the Chinese mainland who shared an ancestral root with the Miao hill tribes or Kweichow aborigines, ancestors of the Yuëh people (summarized in Goddard 1963). One theory proposes that proto-Malayans colonized the southern part of the island and in expanding north met with preexisting populations from Japan who had taken up residence, represented by today's Ainu (Goddard 1966). These early hypotheses were based primarily on phenotypic and cultural similarities with the aforemen-

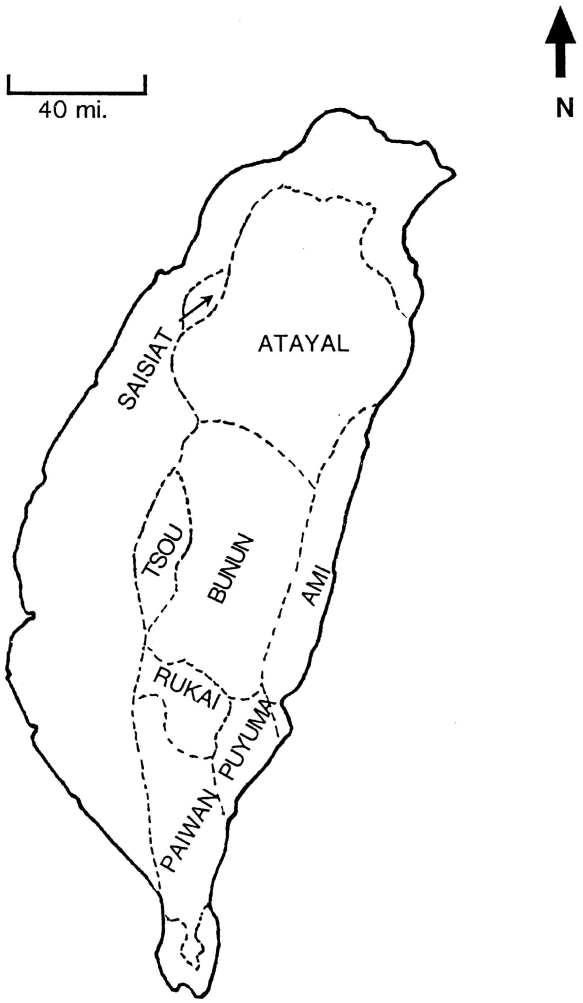

Figure 1 Map of Taiwan, showing the approximate geographic boundaries for the eight main island aboriginal tribes.

tioned populations, rather than on archaeological or linguistic studies.

A complex and uncertain archaeological record indicates that Taiwan has been occupied by humans since at least 6000–3000 B.C. (Chang 1969). Overall, excavated sites provide links primarily with the southern Chinese mainland and suggest that there have been several distinct introductions of material culture to the island. Linguistic reconstruction places the earliest speakers of an Austronesian language in Taiwan well before 2000 B.C.; from there, proto–Malayan-Polynesian languages spread throughout island Southeast Asia and Oceania (Bellwood 1978, 1991; Blust 1988). However, there is little evidence to suggest an ultimate geographic origin for a pre-Austronesian language group, because today no Austronesian languages are spoken in south China. Controversies about the clustering of extant Taiwanese languages also exist. For example, although these languages generally are divided into the Atayalic, Tsouic, and Paiwanic subgroups, there is uncertainty about whether Paiwanic and Tsouic languages were introduced later (Ferrell 1969; Bellwood 1978) and whether the language spoken by the Ami tribe should represent a separate branch from which extra-Formosan proto-Malayan languages arose (Ross 1994).

Studies of classical genetic markers point to overall genetic affinities with south China and Southeast Asia (Kutsuna and Matsuyama 1939; Chou 1959; Huang 1964; Huang and Sheen 1966; Ikemoto et al. 1966; Nakajima et al. 1967, 1971; Nakajima and Ohkura 1971). A study of the Toroko, a branch of the Atayal tribe, found affinities primarily with the Philippines and Thailand and to a lesser extent with south China and Vietnam (Chen et al. 1985). Significant heterogeneity for serum complement proteins was observed among tribes, supporting a scenario of several migratory events from multiple sources in mainland China and Southeast Asia (Umetsu et al. 1994). High frequencies of some orosomucoid alleles that were more common in Chinese than in Southeast Asians were observed (Umetsu et al. 1995). More significantly, with respect to relationships among Taiwanese aborigines, the ORM1*Q0 allele was present at a low frequency in the eight extant mainland tribes, although it is rare worldwide. This suggests either that the tribes have a common origin or that there has been gene flow among the tribes. Tang et al. (1995) measured the frequency of glucose-6-phosphate dehydrogenase deficiency in three tribes and found that the most common disease allele in the Ami tribe had been observed in one individual from China.

We previously used mtDNA to outline a scenario for the spread of proto-Polynesians from Taiwan throughout island Southeast Asia (Melton et al. 1995). We used as markers a 9-bp deletion, which has been observed at moderate to high frequencies in Asia and is nearly fixed in Polynesia (Wrischnik et al. 1987; Hertzberg et al. 1989; Ballinger et al. 1992; Harihara et al. 1992; Hagelberg and Clegg 1993; Horai et al. 1993; Passarino et al. 1993; Lum et al. 1994; Melton et al. 1995; Redd et al. 1995), and three polymorphisms at nucleotide positions 16217, 16247, and 16261 in the mtDNA control region (termed the "Polynesian motif," because these substitutions are found at high frequencies in Oceanic populations with the Asian 9-bp deletion), as well as an additional 15 polymorphic sites in the control region, determined via hybridization with sequence-specific oligonucleotide (SSO) probes (Stoneking et al. 1991; Melton et al. 1995; Melton and Stoneking 1996).

To clarify the relationships among Taiwanese tribes and other Asian populations, we (1) analyzed SSO profile data for Asian populations, using analyses of molecular variance (AMOVAs; Excoffier et al. 1992), (2) selected a subset of Taiwanese tribal mtDNAs for se-

quence analysis of both hypervariable segments of the control region, to compare intertribal variation at the sequence level, (3) compared variation at the first hypervariable segment with previously published Asian mtDNA sequence data, and (4) used allele-frequency data from eight nuclear *Alu* insertion polymorphisms from 14 populations to explore relationships among the Taiwanese and other Asian and Oceanic populations.

Material and Methods

Analysis of SSO Profile Data

Asian population samples, including 82 aboriginal Taiwanese from four tribal groups, and their SSO typing have been described elsewhere (Melton et al. 1995; Melton and Stoneking 1996); however, in this study, mtDNAs with and those without the 9-bp deletion were analyzed separately, since subtle associations with mainland Asia may be difficult to detect, owing to the robustness of 9-bp–deletion affinities within island Southeast Asia. Variants at the IE probe position (Melton et al. 1995), which detect the Polynesian motif, were omitted from this analysis, since they are highly correlated with the 9-bp deletion. Genetic diversity (*h;* Tajima 1989) was estimated for SSO types with and for those without the 9-bp deletion. AMOVAs were performed on data sets of SSO profiles with and on those without the 9-bp deletion, to generate genetic distances among pairs of populations and to test their statistical significance (Melton and Stoneking 1996).

Samples

Twenty-eight aboriginal Taiwanese samples were selected for sequence analysis of hypervariable segments 1 and 2 of the mtDNA control region (table 1). To aid in the selection of samples, for sequence analysis, that represent the range of mtDNA SSO-type variation, a neighbor-joining tree (Saitou and Nei 1987) was constructed from a matrix of the number of inferred substitutions between each pair of the 36 SSO types found in the sample of 82 individuals, and samples were selected from this tree, including a few samples sharing the same SSO type. From published sequences from another 23 aboriginal Taiwanese from these populations (Sykes et al. 1995) we determined that there are, to date, no known major sequence motifs present in these populations other than those sequenced for this study. Therefore, the SSO-typing approach appears to have been very successful in capturing most of the sequence variation in this sample. Of the 28 samples selected for sequencing, 11 (39.3%) had the 9-bp deletion, a proportion similar to that in the entire sample of 82 individuals (41.5%; Melton et al. 1995).

Table 1

Deletion Status of the 28 Samples Sequenced

	No. of Samples		
Population	With Deletion[a]	Without Deletion[b]	Total[c]
Ami	3	4	7
Atayal	2	5	7
Bunun	3	4	7
Paiwan	3	4	7
Total	11	17	28

[a] $h = .982$; mean sequence divergence = .93%.
[b] $h = .993$; mean sequence divergence = 1.06%.
[c] $h = .997$; mean sequence divergence = 1.17%.

Control-Region Amplification and Sequencing

mtDNA was amplified by use of PCR, as described elsewhere (Melton et al. 1995), with the exception that primers L15996 and H16401 were used for region 1 and primers L29 and H408 were used for region 2 (Vigilant et al. 1989). Amplifications for each region were performed in duplicate, with one of the two primers biotinylated at the 5′ end in each reaction. Purification of DNA template and dideoxy sequencing were performed as described elsewhere (Redd et al. 1995).

Data Analysis

After removal from the sequences of sites with insertions and deletions, *h* (Tajima 1989) and the mean sequence divergence (determined by use of the proportion of nucleotide differences between each pair of sequences [MEGA program; Kumar et al. 1993]) were estimated for samples with and for those without the 9-bp deletion. The 28 Taiwanese sequences were aligned with a !Kung African sequence (Vigilant et al. 1989) as an outgroup, and a phylogenetic tree was created by use of the neighbor-joining method, with a matrix of the proportion of nucleotide differences between each pair of sequences (MEGA program; Kumar et al. 1993). The reliability of the structure of the tree's internal branches was tested by use of 500 bootstrap replications (Felsenstein 1985). AMOVAs incorporating pairwise sequence genetic distances (based on the proportion of nucleotide differences) or SSO genetic distances (Melton and Stoneking 1996) were used to determine if the four subsets of tribal mtDNAs were significantly different with respect to their mtDNA sequences or SSO types. The variance among and within populations quantified the amount of heterogeneity among the tribal groups. Permutation testing (1,000 replications) was used to test the significance of the pairwise population genetic distances and variance components (Melton and Stoneking 1996).

Pairwise difference distributions among sequences from individuals with and from those without the 9-bp deletion ($n = 10$ and $n = 17$, respectively) were used to

estimate expansion and divergence times for the Taiwanese, on the basis of the model of Rogers and Harpending (1992). We used a substitution rate of $u = 1.142 \pm 0.333 \times 10^{-7}$ substitutions per site per year per lineage (Stoneking et al. 1992) to estimate tau (τ) units of mutational time, for which $\tau = 2tu$ (u indicates total substitution rate over all sites, and t indicates time, in generations). Computations were performed with the IWaVe program (Sherry 1994). Tajima D tests were performed by use of the mean number of pairwise nucleotide differences and the number of segregating sites, to assess the likelihood of population expansions (Tajima 1989).

Comparison with Other Asian Populations

Previously published mtDNA sequences from Asian populations ($n = 764$) were edited to a standard length (nucleotide positions 16092–16362) and were compared with the first hypervariable region from the Taiwanese sequences. Data were included for populations from north Asia (Altai, Chukchi, Siberian Eskimo, Evenk, Nivkh, and Udegey [Shields et al. 1993; Torroni et al. 1993*b*]), northeast coastal Asia (Japan, Korea, Ainu, and Ryukyu [Horai and Hayasaka 1990; Torroni et al. 1993*a*; Redd et al. 1995; Horai et al. 1996]), central and southern mainland Asia (Han/China, Han/Tibet, Han/Taiwan, Mongolia, and aboriginal/Taiwan [Vigilant et al. 1989; Horai and Hayasaka 1990; Redd et al. 1995; Sykes et al. 1995; Kolman et al. 1996]), island Southeast Asia (the Philippines, Indonesia, and Borneo [Vigilant et al. 1989; Horai and Hayasaka 1990; Redd et al. 1995; Sykes et al. 1995]), and Oceania (Papua New Guinea, the Cook Islands, Tahiti, Aoteoroa, Marquesas, Vanuatu, Australes, the Marshall Islands, Tonga, and Samoa [Horai and Hayasaka 1990; Stoneking et al. 1992; Redd et al. 1995; Sykes et al. 1995]). Only data from individuals without the 9-bp deletion, from the study by Sykes et al. (1995), were used, since those with the deletion were sequenced only from nucleotide positions 16189–16390. All unique sequences ($n = 470$) were used in a neighbor-joining tree (data not shown). This tree was too large for bootstrap analysis; so, those sequences that occurred in the Taiwanese, in three or more Asian individuals without the 9-bp deletion ($n = 51$), or in two or more Asian individuals with the 9-bp deletion ($n = 18$) were used in a neighbor-joining tree with a !Kung sequence as an outgroup (Vigilant et al. 1989). No major differences were noted between the overall structure of this tree and that of the larger tree of 470 sequences. Interior-branch reliability for the tree was tested by means of 500 bootstrap replications.

Alu *Insertion Polymorphisms*

An average of 624 individuals from Taiwan, China, the Philippines, Malaysia, Java, highland Papua New Guinea, coastal Papua New Guinea, east Indonesia (Moluccas and Nusa Tenggaras), Australia, and Samoa were typed for eight *Alu* nuclear insertion polymorphisms, as reported elsewhere (Stoneking et al. 1997). For this project, the Taiwanese were further subdivided into the Ami, Atayal, Bunun, and Paiwan, and new analyses were performed. A principal-component (PC) analysis based on gene-frequency data (presence or absence of the alleles) was used to infer population affinities, by use of the POPSTR program (provided by H. Harpending). Correlations between interpopulation distances for the *Alu* loci—calculated both as Nei standard genetic distances (Nei 1972), by the DISPAN program (Ota 1993), and as Wright's F-statistics (F_{ST}) distances, by the POPSTR program—and interpopulation F_{ST} analog (Φ_{ST}) values for the mtDNA SSO types (AMOVA; Excoffier et al. 1992) were measured by means of the Mantel test (R package; Legendre and Vaudor 1991). An unrooted neighbor-joining population tree was constructed from the matrix of Nei standard genetic distances, and interior-branch reliability was tested by means of 100 bootstrap replications (PHYLIP program; Felsenstein 1993).

Results

SSO Type and Sequence Diversity

Diversity estimates for Taiwanese SSO types with and for those without the 9-bp deletion were $.925 \pm .015$ ($n = 34$) and $.945 \pm .011$ ($n = 48$), respectively, which are not significantly different. AMOVA results on SSO profile data for eight Asian populations subdivided by 9-bp–deletion status are shown in table 2 (*top matrix*). Results for Taiwanese without the 9-bp deletion were not significantly different from those for Filipinos only (table 2, *top right matrix*) ($P = .024$; Bonferroni test $P = .007$, based on seven tests [Weir 1990]), but this was not a robust association, as evidenced by the low P value. Results for Taiwanese with the 9-bp deletion were not significantly different from those for any of the seven other populations, with P values within a range of .028 (Borneo) to .554 (the Philippines) (table 2, *top left matrix*). Overall, the Taiwanese appeared to be most closely linked with island Southeast Asia, with regard to the 9-bp deletion, but more subtle associations were not apparent when SSO profiles were compared.

Whereas the samples selected for sequence analysis had 17 different SSO profiles, there were 27 unique control-region sequences among these 28 samples (average sequence length was 673 nucleotides), with one sequence appearing twice (samples ATA5 and ATA6). Diversity estimates (h) and mean sequence divergences are given in table 1. These estimates should be considered cautiously, because the samples were not chosen randomly; rather, they were chosen to represent the range of variation present in the total sample of 82 individuals. Var-

Table 2

Results of Significance Tests of Genetic Distances between Pairs of Populations

	P VALUE							
POPULATION	Borneo	China	Philippines	Java	Orang Asli	East Indonesia	Taiwan	Malaysia
Borneo	...	.004	.000	.026	.000	.022	.000	.012
China	.174	...	.006	.033	.000	.014	.001	.252
Philippines	.004	.091	...	.001	.000	.002	.024	.041
Java	.591	.318	.136	...	.000	.564	.002	.111
Or Asli	.000	.003	.055	.000	...	.000	.000	.000
East Indonesia	.002	.001	.141	.006	.010	...	.003	.043
Taiwan	.028	.125	.554	.172	.063	.178	...	.007
Malaysia	.362	.504	.241	.525	.005	.072	.232	...
	Ami	Atayal	Bunun	Paiwan				
Ami	...	.799	.503	.782				
Atayal	.795	...	.818	.685				
Bunun	.352	.376	...	.497				
Paiwan	.810	.814	.649	...				

NOTE.—The top matrix gives values for SSO types without the 9-bp deletion (*right*) and for SSO types with the 9-bp deletion (*left*). The bottom matrix gives values for control-region sequences for the Taiwanese (*right*) and for SSO types from these same sequences (*left*).

iable sites, with respect to the Cambridge reference sequence (CRS; Anderson et al. 1981), are shown in table 3. Control-region motifs discriminating the samples with the 9-bp deletion were quite different from those of samples without the deletion; nucleotide substitutions associated with the 9-bp deletion in other Asians (Melton et al. 1995; Redd et al. 1995; Sykes et al. 1995) appeared in all but one sample (AMI2). The 9-bp–deletion status of this sample was reconfirmed in the laboratory (the deletion in this sample was also found in another laboratory [B. Sykes, personal communication]). In this sequence, the lack of substitutions commonly associated with the deletion in Oceania may indicate that an independent 9-bp deletion event has occurred. The 9-bp deletion has been observed, by other researchers, on the background of different control-region substitutions, giving strong evidence for multiple occurrences of the deletion, and the polymorphisms in this sample do not match those reported by others (Vigilant 1990; Ballinger et al. 1992; Chen et al. 1995; Redd et al. 1995; Soodyall et al. 1996). An independent deletion event might be confirmed by sequencing the 9-bp region of the questioned sample, to determine if the deletion region has the commonly observed sequence of the Asian 9-bp deletion. Because of uncertainty about the origin of the 9-bp deletion in this sample, it was not included in further analyses.

A neighbor-joining tree (fig. 2) shows that there is no sequence-specific tribal or subpopulation identity: sequences from different tribes are completely interleaved in the phylogeny. AMOVA results (table 2, *bottom matrix*) indicate that, both at the sequence level (table 2, *bottom right matrix*) and for this subset of SSO types (table 2, *bottom left matrix*), there were no significant differences among the Ami, Atayal, Bunum, and Paiwan (P always >.35). This analysis lumped together samples with and those without the deletion, to increase the overall sample sizes, but the proportion of each type was similar in each tribal group. In the AMOVA based on the sequences, the variance among tribes was −3.7%, whereas the variance within tribes was 103.7% (P = .90, 1,000 permutations). In the AMOVA based on the SSO types, the variance among tribes was −44.5%, whereas the variance within tribes was 144.5% (P = .74, 1,000 permutations). Negative variances in both tests indicate that some sequences and especially some SSO types were, on average, more similar among tribes than within tribes. Most likely the extreme SSO-type similarity among tribes can be attributed to the control-region variation associated with presence or absence of the 9-bp deletion.

Mismatch Analyses

Mismatch and intermatch distributions are shown in figure 3. The peaks for the distribution with and for that without the 9-bp deletion were at ~4.1 and ~6.4 substitutions, giving expansion times of ~23,500 years and ~36,500 years, respectively, for these mtDNA types (95% confidence intervals were approximately 6,600–40,400 years and 22,500–50,500 years, respectively). These times place the coalescent for the present-day mtDNA variation in Taiwan prior to habitation of the island, on the basis of archaeological data. Although the low, wide profiles of these waves seem to indicate that the Taiwanese have not undergone any recent severe bottlenecks in population size, it is important to give this analysis a conservative appraisal, given the small

Table 3

Variable Sites for 28 Taiwanese Sequences, with Respect to the CRS

			Nucleotide Position																					
			16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16
			30	60	60	80	90	29	36	40	48	45	72	85	89	17	22	20	22	23	48	66	67	72
Sample[a]	HV1 HT[b]	9-bp Deletion[c]	07	15	19	66	29	99	63	04	84	44	22	55	99	78	80	02	33	48	81	62	16	02
Reference	…		A	A	C	T	T	G	T	T	C	T	T	C	T	T	C	A	C	T	C	C	C	A
AMI20	36	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.
ATA5	36	2	?	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.
ATA6	36	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.
BUN4	36	2	.	.	T	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.
AMI5	28	2	.	G	.	C	.	.	.	.	.	.	.	T	.	.	.	.	T	.	.	.	.	.
BUN10	28	2	G	G	.	C	.	.	.	.	.	.	.	T	.	.	.	.	T	.	.	.	.	.
ATA1	28	2	.	.	.	C	.	.	.	.	.	.	.	T	.	.	.	.	T	.	.	.	.	.
AMI1	1	2	.	.	.	.	C	.	.	.	T	.	.	T	C	.	.	.	T	.	.	.	.	.
PAI5	32	2	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	T	.	.	.	.	.
BUN12	43	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.
PAI8	39	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.
BUN11	24	2	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	T	.	T	.	.	.
PAI12	51	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.
AMI6	51	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.
ATA4	51	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.
PAI16	37	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.
ATA13	21	2	.	.	.	C	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AMI2	…[d]	1	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.
AMI13	66	1	.	.	.	.	.	.	.	.	.	.	.	.	C	C	.	.	.	.	.	T	.	.
AMI17	69	1	.	.	.	.	.	.	.	.	.	.	.	.	C	C	.	.	.	.	.	T	.	.
PAI9	66	1	.	.	.	.	.	.	.	.	.	.	.	.	C	C	.	.	.	.	.	T	.	.
ATA8	63	1	.	.	.	.	.	.	.	.	.	C	.	.	C	C	.	.	.	.	.	T	.	.
PAI20	67	1	.	.	.	.	.	.	.	.	.	.	.	.	C	C	.	.	.	.	.	T	.	G
ATA11	56	1	.	.	.	.	.	.	C	.	.	.	.	.	C	C	.	.	.	.	.	.	.	.
BUN14	56	1	.	.	.	.	.	.	C	.	.	.	.	.	C	C	.	.	.	.	.	.	.	.
BUN6	61	1	.	.	.	.	.	.	.	C	.	.	.	.	C	.	.	.	.	.	.	.	G	.
BUN13	61	1	.	.	.	.	.	.	.	C	.	.	.	.	C	.	.	.	.	.	.	.	G	.
PAI1	62	1	.	.	.	.	.	.	.	C	.	.	.	.	C	.	.	.	.	.	.	.	G	.

[a] AMI = Ami, ATA = Atayal, BUN = Bunun, and PAI = Paiwan.

[b] Hypervariable region 1 haplotype number in the neighbor-joining tree in figure 4.

[c] 1 = deletion present, and 2 = deletion absent.

[d] Not used in the neighbor-joining tree in figure 4.

number of sequences. Although Tajima *D* test values were not statistically significant (−1.454, −1.489, and −1.134, for samples with the 9-bp deletion, those without the 9-bp deletion, and the total sample, respectively), the negative values are consistent with the possibility that population expansion has occurred (Tajima 1989). The intermatch distribution, which estimates the time of divergence of these two populations, essentially illustrates the history of the 9-bp deletion. The divergence time between mtDNAs with and those without the deletion is placed at ~55,000 years ago, which is similar to a previous estimate of ~58,000 years ago for the coalescent of the 9-bp deletion (Redd et al. 1995).

Phylogenetic Analyses

Figure 4 shows how Taiwanese mtDNAs fit into a phylogeny with other Asians (hypervariable region 1 only; Taiwanese lineage numbers are also shown in table 3; other Taiwanese lineages are from Sykes et al. 1995). Sixty-nine mtDNA lineages that occurred in the Taiwanese, in two or more Asians with the deletion, or in three or more Asians without the deletion were included in this analysis, but a larger tree of 470 Asian sequences (mtDNAs with and those without the 9-bp deletion; data not shown) was not different with respect to the conclusions drawn from this tree. Although bootstrap values are not high for the major clusters, there is geographical coherence to the structure of the tree. Notably, the Taiwanese are scattered throughout the tree, with lineages in every major cluster except the last, which is almost exclusively Oceanic. Although the Taiwanese are scattered, however, they actually share with other populations only 7 of their 21 lineages. The shared lineages often are widely spread over more than one population,

Nucleotide Position																												
16291	16294	16295	16297	16298	16304	16311	16324	16362	16365	16390	73	93	146	150	152	153	185	195	198	199	204	207	210	246	248	263	292
C	C	C	T	T	T	T	T	T	C	G	A	A	T	C	T	A	G	T	C	T	T	G	A	T	A	A	T
.	.	.	.	.	C	C	.	.	.	.	G	.	C	.	.	.	.	.	T	.	.	.	.	.	del	G	.
.	.	.	.	.	C	C	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	del	G	.
.	.	.	.	.	C	C	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	del	G	.
.	.	.	.	.	C	C	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	del	G	.
.	.	.	.	.	.	.	.	C	.	A	G	.	.	.	.	.	.	C	.	.	.	.	.	C	.	G	.
.	.	.	.	.	.	.	.	C	.	A	G	.	.	.	.	.	.	C	.	.	.	.	.	.	.	G	.
.	.	.	.	.	.	.	.	C	.	A	G	.	.	.	.	.	.	C	.	.	.	.	.	.	.	G	.
.	.	.	.	.	.	.	.	C	.	.	G	.	.	T	C	.	A	.	.	.	.	.	.	.	.	G	.
.	.	.	.	.	.	.	.	C	.	.	G	.	.	T	.	.	.	.	.	.	.	.	.	.	.	G	.
T	.	.	.	.	.	.	.	C	.	A	G	.	.	.	C	G	.	.	.	.	.	.	.	.	.	G	.
T	T	.	.	.	.	.	.	C	.	A	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.
T	.	.	.	.	.	.	.	.	.	A	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.
.	.	.	.	.	.	.	.	C	.	.	G	.	C	.	.	.	.	.	.	C	.	.	.	.	.	G	.
.	.	.	.	.	.	.	.	C	.	A	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	C
.	.	.	.	.	.	.	.	C	.	A	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.
.	.	.	.	C	.	C	.	C	.	.	G	.	.	T	.	.	.	.	.	.	.	.	.	.	del	G	.
.	.	.	C	.	.	.	C	.	.	.	G	.	.	.	.	.	.	.	.	C	.	.	.	.	.	G	.
.	.	T	.	.	.	.	.	C	.	.	G	.	C	.	C	.	.	.	.	C	.	.	.	.	.	G	.
.	.	.	.	.	.	.	.	.	.	.	G	.	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.
.	.	.	.	.	.	C	.	.	.	.	G	.	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.
.	.	.	.	.	.	.	.	.	.	.	G	.	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.
.	.	.	.	.	.	.	C	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.
.	.	.	.	.	.	.	C	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.
.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	A	.	.	.	G	.
.	.	.	.	.	.	.	.	.	T	.	G	.	.	.	.	.	.	.	.	.	.	A	.	.	.	G	.
.	.	.	.	.	.	.	.	.	.	.	G	G	.	.	C	.	.	.	.	.	.	.	G	.	.	G	.
.	.	.	.	.	.	.	.	.	.	.	G	G	.	.	.	.	.	.	.	.	.	.	G	.	.	G	.
.	.	.	.	.	.	.	.	C	.	.	G	G	.	.	.	.	.	.	.	.	C	.	G	.	.	G	.

and there is no preponderance of sharing with any particular geographic region. Lineages with the 9-bp deletion were observed in individuals from north and central Asia, as well as in individuals from Oceania, and included lineages related to the Polynesian motif (lineage 64). With respect to lineages without the 9-bp deletion, lineage 51 is shared with central and northeast Asian populations, lineage 43 is shared mostly with Southeast Asians, lineage 32 is shared with northeast Asians, and lineage 19 is shared with a Mongolian. Lineage 51 is common ($n = 24$) and is deep in the top cluster, arguing for its ancestral position in Asia.

Alu *Insertion Polymorphisms*

Table 4 shows the average heterozygosities and allele frequencies of the eight *Alu* insertion polymorphisms. All eight *Alu* loci were polymorphic in the Taiwanese, and the average heterozygosities were not significantly different from those of other Asian populations. Since these polymorphisms arise because of the insertion of an *Alu* element into a new chromosomal location, the lack of the *Alu* element is known to be the ancestral state; hence, a hypothetical ancestral population can be included in which the frequency of the *Alu* element is 0 at each of the eight loci (Batzer et al. 1994). A PC analysis (fig. 5) showed a cluster of Taiwanese populations that has been invaded only by the Filipino sample. The first, second, and third PCs accounted for 67.6%, 14.2%, and 8.2%, respectively, of the total variance.

Table 5 shows distance matrices for the overall SSO types (including mtDNAs with and those without the 9-bp deletion; Φ_{ST}, *right matrix*) and *Alu* insertion frequencies (Nei standard genetic distances, *left matrix;* matrix of F_{ST} distances not shown). Mantel correlation tests showed high correlations between the AMOVA Φ_{ST} distances based on mtDNA SSO types and the *Alu* distances, for the 12 populations that were typed for both markers (*Alu* F_{ST} distances: $r = .505$, $P = .008$; *Alu* Nei distances: $r = .517$, $P = .014$). Therefore, mtDNA SSO types and *Alu* insertion polymorphisms place these Asian

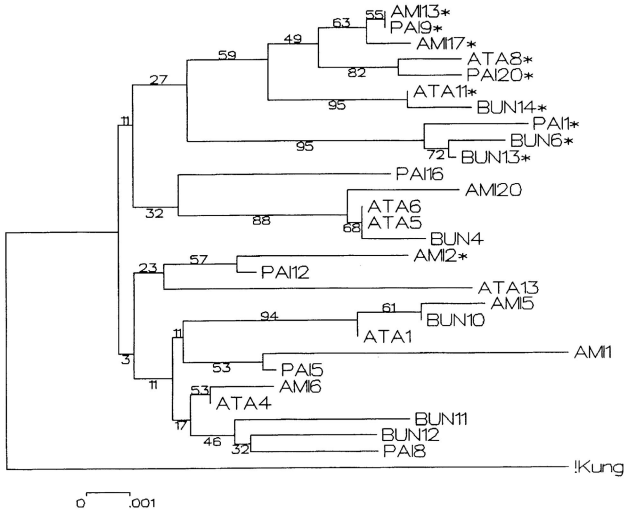

Figure 2 Neighbor-joining tree of 28 Taiwanese control-region sequences, based on the proportion of nucleotide differences as genetic distance. Samples with the 9-bp deletion are indicated by an asterisk (*). Bootstrap values based on 500 replications are shown for each internal branch. AMI = Ami, ATA = Atayal, BUN = Bunun, and PAI = Paiwan.

populations in very similar relationships with each other. However, although the four Taiwanese subpopulations were not found to differ significantly with respect to mtDNA variation, χ^2 contingency tests on the *Alu*-frequency data showed these populations to be significantly different at four of the eight *Alu* loci ($P < .05$). When each subpopulation was removed in turn from each of these four significant χ^2 tests, the Ami were implicated in causing the observed differences for TPA, A25, and B65, whereas the Paiwan and the Atayal were responsible for significant differences for PV92 and TPA, respectively. Therefore, of the four subpopulations, the Ami is the most different with respect to *Alu* insertions.

Figure 6 shows an unrooted neighbor-joining tree based on Nei standard genetic distances for the *Alu* insertion frequencies. As in the PC analysis, the four Taiwanese tribes were found to be the closest neighbors, with their cluster invaded only by Filipinos. Bootstrap values provide reasonable support for these associations. In this tree, the Ami are very close to the Filipinos, and, since there were significant χ^2 differences between the Ami and the other three Taiwanese subpopulations, some gene flow appears to have occurred among the Ami and the Filipinos. Indeed, some gene flow among the Taiwanese and the Filipinos was presumed to have occurred, on the basis of overall mtDNA SSO-type similarity, but the association may be the most robust for the Ami, as shown by means of the *Alu* analysis. The Paiwan also appeared to be similar to Filipinos with respect to overall SSO types (table 5) and to be close in the PC analysis (fig. 6). Interestingly, the Ami and Paiwan are closest geographically to the Philippines.

Discussion

Chai's (1967) comparison of Taiwanese tribal subpopulations demonstrated significant phenotypic, cultural, and linguistic differences among them and revealed that biological relationships among the tribes closely approximated geographic ones. For example, the Ami, living in an eastern coastal corridor, and the Bunun, isolated in the mountains, were the most different from the other groups, in a discriminant function analysis. However, biological relationships did not always support linguistic and cultural classifications, which suggests that some admixture had occurred in regions where tribal borders were in contact over long ranges. Although the degree of migration among tribes was not measured directly, these analyses suggested that gene flow had occurred often among the Atayal, Saisiat, and Tsou and for the Rukai and Paiwan but had occurred very little for either the Ami or Bunun; therefore, there is no strong support for significant admixture having occurred among the four subpopulations that we studied. There is evidence that the Atayal once inhabited the western coastal plain and over time moved into the higher elevations. The Paiwan, living at the southernmost end of the island, in the Central Mountains, once may have also lived on the west coast and then moved inland to a higher elevation. These population movements most likely resulted from conflict with waves of invaders from the Chinese mainland beginning by at least A.D. 230 (the earliest recorded invasion). A Chinese geography written in the third century A.D. reported that the Taiwanese had separate social or breeding groups, indicating that some, if not all, prehistoric tribal identities were mostly in place nearly 1,700 years ago (Chai 1967).

The mtDNA sequence variation that we observed does not support a hypothesis that Taiwanese subpopulations had separate deep origins owing to different, clearly defined source populations. Although they cluster with respect to the 9-bp deletion, sequences from all four tribes were otherwise thoroughly interleaved in a phylogenetic tree, and the AMOVAs for the sequences and the SSO types from these sequences showed no heterogeneity among tribes. The *Alu* PC map showed that the four tribes are neighbors, and their grouping was invaded only by Filipinos. However, in the *Alu* analysis, the tribes appeared to be more distant from each other than are some other Southeast Asian populations, such as Moluccans, Malaysians, and Chinese; indeed, χ^2 tests showed that the subpopulations are not homogeneous at four of the eight *Alu* loci. This finding, plus the ex-

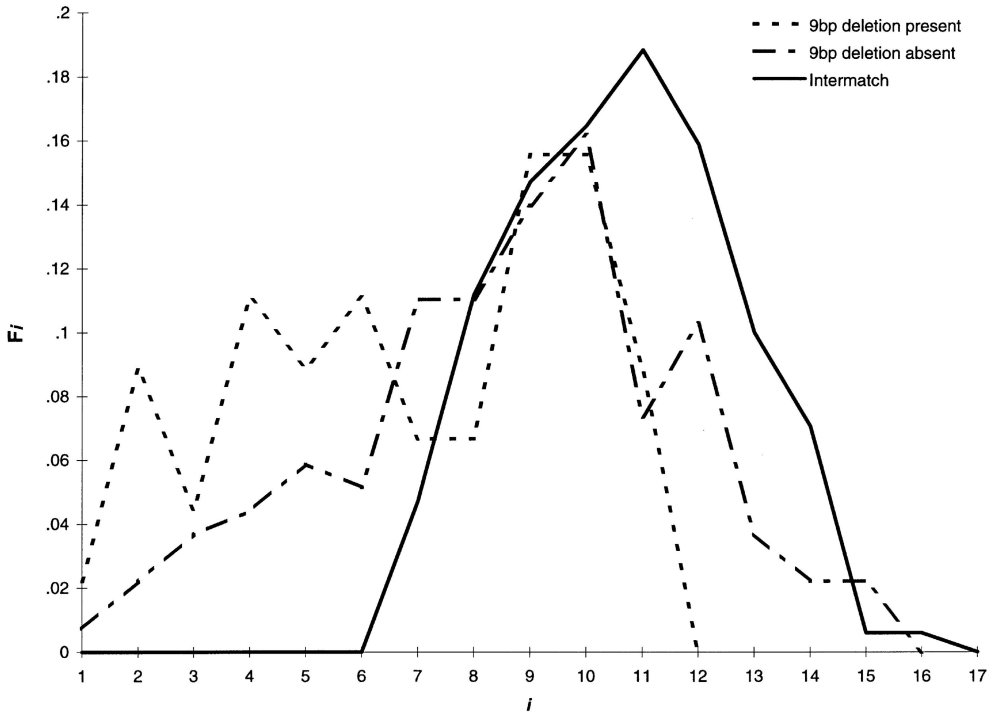

Figure 3 Mismatch and intermatch distributions created by use of the frequency (F_i) of the number of nucleotide differences (i) in pairwise sequence comparisons, for Taiwanese with (n = 10) and those without (n = 17) the 9-bp deletion.

istence of significant intertribal cultural and linguistic diversity, leads us to speculate that gene flow between at least these four tribes has been relatively low over at least the last 2,000 years since they were driven inland, a time period that would not allow for the accumulation of much intertribal mtDNA differentiation but that has allowed for accumulation of some *Alu* insertion–frequency differences.

Although there could have been a single genetically diverse common founding population for the Taiwanese, there more likely have been several introductions of mtDNAs onto the island, which is supported by the archaeological record. Taiwanese Corded Ware, excavated at Fengpitou (southwest Taiwan), Tapenkeng (north Taiwan), and other sites (central Taiwan) (Chang 1969, 1974) and predating 2500 B.C. by an unknown margin of time (unlikely to be earlier than 5000–4000 B.C.), generally relates to Corded Ware of south China and north Indochina that is dated to a much earlier time (7000 B.C.; Bellwood 1978). Lungshanoid pottery (2400–500 B.C.), found at sites generally restricted to Taiwan's west coast, also is derived from China, as determined on the basis of earlier mainland dates for this period. Contemporaneous or later cultures for which there are fewer data, the Yuan-shan culture (north coast, 2000 B.C. to 0) and the Tai-yuan culture (east coast), appear to have links with the Corded Ware culture. In particular, the Tai-yuan culture includes groups of standing stones that are Austronesian in character (Bellwood 1978). However, although archaeological studies to date have suggested that multiple exposures to mainland culture occurred in Taiwan in prehistoric times, as evidenced by at least two major industries, the Corded Ware and Lungshanoid cultures, dates consistent with the proto-Polynesian expansion (beginning at ~4000 B.C.) link the Taiwanese most firmly to the Corded Ware culture, which culturally was the most Austronesian and is found all over the island, whereas the Lungshanoid culture began later (2500 B.C.) and is located predominantly on the west coast. It is possible that the present-day genetic diversity reflects introduction from a number of waves of settlement over several thousand years, but it is tempting to regard the early Corded Ware culture, which dates to ~4000 B.C., as the most influential on Austronesian expansion and language and, therefore, as the predominant genetic and cultural contributor to the present-day Austronesian population.

Placing the Taiwanese in a pan-Asian temporal and geographic context is somewhat difficult. The mtDNA data appear to support at least two ideas. First, the Taiwanese appear to have been mostly isolated from mainland Asians for some unknown period of time, as evidenced by their relative lack of sharing of contemporary control-region sequences and, conversely, by their high

Figure 4 *Left,* Neighbor-joining tree and bootstrap values for 69 mtDNA hypervariable region 1 lineages in Asians. *Right,* Table showing the geographic clustering of the lineages. The lineages are placed in the table to correspond to their placement in the tree. The number of times each lineage was observed is indicated for the following populations: SE = Siberian Eskimo, CK = Chukchi, EV = Evenk, NI = Nivkh, UD = Udegey, AL = Altai, JA = Japan, AI = Ainu, RY = Ryukyu, KO = Korea, MO = Mongolia, HN = Han (Chinese, Taiwanese), TA = aboriginal Taiwanese, PH = Philippines, IN = Indonesia, BO = Borneo, PN = Papua New Guinea, VA = Vanuatu, MI = Marshall Islands, AU = Australes, TO = Tonga, SA = Samoa, CI = Cook Islands, AO = Aoteoroa, TH = Tahiti, and MA = Marquesas.

		N. Asia						N.E. Asia				C. Asia			S.E.Asia			Oceania								
Lineage		SE	CK	EV	NI	UD	AL	JA	AI	RY	KO	MO	HN	TA	PH	IN	BO	PN	VA	MI	AU	TO	SA	CI	AO	TH	MA
61														**2**													
62														**1**													
60													1			1											
53											1	1															
59												2				1											
58													3			2											
57																2											
55													2														
56												2		**2**		1											
69														**1**		1											
54																2											
64																1		12					17				
52																		2									
68	9 base pair deletion lineages											2	1														
66											1		1	**2**		5		2					2				
63														**1**													
67														**1**													
65																							2				
26																			3								
6								3																			
7								7																			
23																	3										
4													3														
14													2		1		1										
25								1																			2
36														**4**													
16								2	1		1																
33												3															
30								1	1	1																	
31											3																
32								3		1	1			**2**													
27								1				1	1														
47				1		1						1	2														
38								1		5	1	1			1												
40														**1**													
1														**1**													
28														**3**													
34									8		1																
35										2	3				1												
24														**1**													
39														**2**													
43				1										**1**	1	1	2										
45											3						4			1							
46												1			6		1										
5		2	2																								
42			1						1			1	1													1	
8											1		2														
29								2			2																
49				1			1	1		1																	
3					1							4															
37														**1**													
48								1	3			3															
41											1	3															
44											3																
51				1				5	1		8	4	2	**3**													
22														**1**													
50									3																		
13																			7								
2									4																		
18									2								1										
19												2		**1**													
20														**1**													
15								2			2																
17												1	5														
21														**1**													
12																		3									
9																								3			
10																					3		1	11	1	2	
11																		2				1	1				

Table 4

Average Heterozygosities and Allele Frequencies of the Eight *Alu* Insertion Polymorphisms Included in the PC Analysis and Neighbor-Joining Tree

Population	Heterozygosity ± SE	Allele Frequency (No. of Individuals Typed)							
		ACE	TPA25	PV92	APO	MABD1	A25	B65	FXIIIB
Ami	.27 ± .07	.59 (22)	.78 (20)	.89 (23)	.91 (23)	.43 (21)	.00 (21)	.74 (21)	.98 (22)
Atayal	.34 ± .07	.58 (19)	.43 (14)	.92 (19)	.94 (18)	.47 (19)	.24 (19)	.42 (18)	.97 (18)
Bunun	.30 ± .07	.40 (20)	.58 (20)	.98 (20)	.98 (20)	.32 (19)	.25 (20)	.30 (20)	.95 (19)
Paiwan	.36 ± .05	.53 (20)	.69 (21)	.60 (21)	.93 (20)	.33 (21)	.14 (21)	.29 (21)	.88 (21)
Moluccas	.35 ± .05	.67 (49)	.56 (49)	.69 (49)	.76 (49)	.19 (47)	.00 (49)	.26 (48)	.78 (48)
Nusa Tengarras	.37 ± .05	.64 (91)	.38 (91)	.50 (91)	.78 (91)	.19 (88)	.05 (94)	.40 (86)	.81 (90)
Coastal PNG	.35 ± .05	.66 (47)	.16 (47)	.36 (47)	.66 (47)	.17 (49)	.02 (49)	.27 (47)	.30 (49)
Highland PNG	.29 ± .06	.74 (69)	.16 (69)	.24 (69)	.68 (69)	.01 (69)	.04 (69)	.18 (68)	.30 (66)
Australia	.29 ± .05	.91 (99)	.13 (99)	.15 (99)	.87 (99)	.04 (42)	.35 (43)	.39 (33)	.65 (40)
Philippines	.37 ± .06	.53 (52)	.63 (51)	.80 (30)	.98 (51)	.36 (50)	.14 (51)	.57 (50)	.72 (38)
China	.35 ± .04	.67 (50)	.35 (49)	.86 (50)	.82 (50)	.17 (49)	.10 (49)	.35 (48)	.71 (45)
Malaysia	.38 ± .05	.64 (48)	.50 (44)	.72 (47)	.76 (48)	.27 (43)	.02 (46)	.42 (44)	.73 (45)
Java	.33 ± .05	.86 (32)	.39 (32)	.84 (32)	.78 (32)	.42 (32)	.06 (31)	.58 (32)	.92 (31)
Samoa	.33 ± .07	.89 (46)	.48 (46)	.53 (47)	.71 (47)	.19 (48)	.01 (48)	.45 (47)	.91 (47)

Note.—Populations are described by Stoneking et al. (1997). PNG = Papua New Guinea. SE = standard error.

number of unique lineages, but this separation time cannot be dated, because the few clusters of uniquely Taiwanese sequences are so small. The Taiwanese also lack the most common Asian mtDNA SSO type (234), which has been found in every other east Asian population sampled. Furthermore, an AMOVA of 11 Asian populations revealed that the Taiwanese account for more Asian substructure than does any other single population (Melton and Stoneking 1996). Second, the Taiwanese have a deep position with respect to Asian mtDNA control-region variation, as evidenced by the appearance of their sequences throughout neighbor-joining phylogenies of the most commonly shared types. This may indicate that they are derived from an early diverse pool of types that spread, from a centralized location, throughout Asia; certainly, the mismatch expansion times indicate that this substantial diversity has temporally deep roots. In fact, the four shared lineages without the deletion were observed to be from central Asia, northeast coastal Asia, island Southeast Asia, and Mongolia. The presence of related types in Taiwan, Japan, Korea, Ryukyu, and the Ainu is quite intriguing, and, although nothing suggests tight associations among any of these populations, we cannot rule out more recent introduction directly onto Taiwan, via migration, of mtDNAs from these northern coastal regions (however, the flow of mtDNAs could have been in the other direction).

Overall, for the markers we have described, the Taiwanese most resemble populations from the Philippines, but this probably is the result of migration from Taiwan—and perhaps especially from the Ami—south to the Philippines, as suggested by an earlier analysis of mtDNA diversity (Melton et al. 1995). However, although a connection between Taiwanese with the deletion and Southeast Asia is robust, mainland Asian connections for the deletion are present as well and have been found in additional mtDNAs observed primarily in north Asia and Taiwan. mtDNA substitutions associated with the Polynesian motif (e.g., 16217 and 16261) were also observed in China and Mongolia (as well as in Indonesia and the Philippines), providing another mainland link. Curiously, although the 9-bp deletion and related control-region motifs are easily detectable in Southeast Asia, there is little evidence of Taiwanese mtDNAs without the deletion in this region. Only one haplotype of Taiwanese without the 9-bp deletion was shared with another island Southeast Asian population, although there were a number of similar lineages. Unfortunately, at the present time, we lack mtDNA sequence data from Vietnam, Thailand, and other areas of mainland Southeast Asia, and, since theories about Taiwanese origins emphasize both south China and northern Indochina as potential sources, these would be valuable populations to sample.

Languages remotely ancestral to Austronesian languages may have been spoken in south China before 5000 B.C. but may have been obliterated by the expansion of Chinese languages (Bellwood 1978). Benedict (1942) hypothesized an ancient link between Austronesian languages and the Thai-Kadai language group, which may have arisen in Neolithic rice-cultivating communities in China south of the Yangtze River, including northern Thailand and Indochina, around 5000–4000 B.C. This theory has been controversial, but there are Chinese records of southern mainlanders (the Yuëh) who may have spoken Austronesian languages, although evidence for this is weak (Bellwood 1978). Therefore, associations between Austronesian languages, the Neolithic archaeology of the Corded Ware culture, and the diverse deep pool of mtDNAs in these tribes provide a

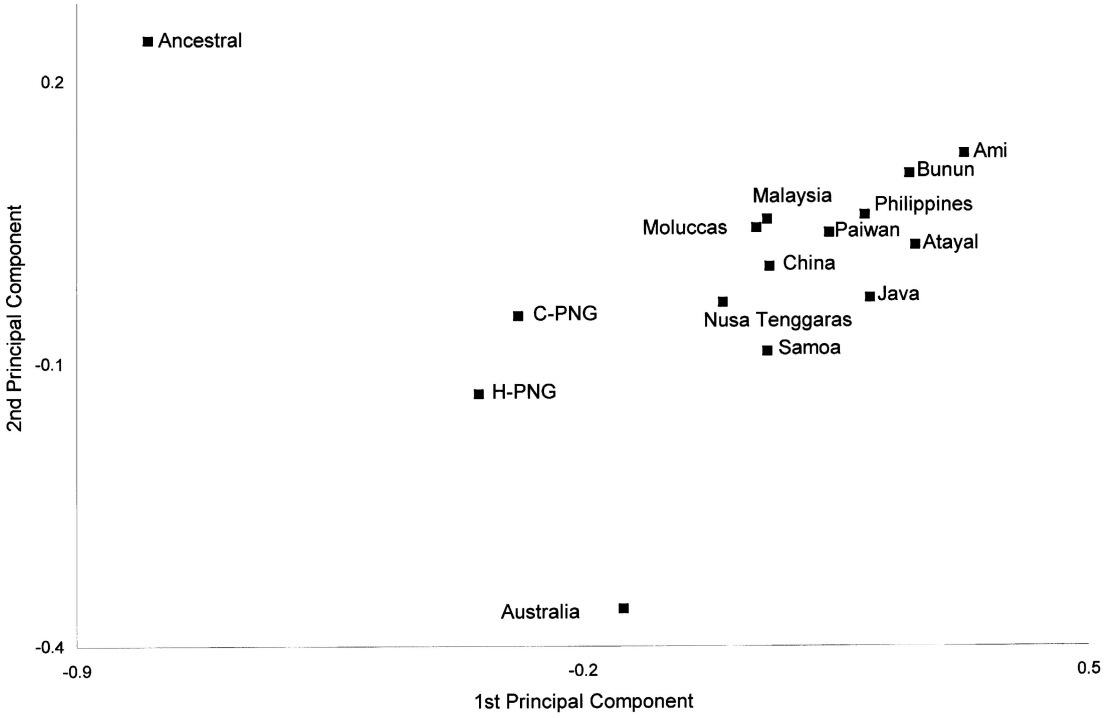

Figure 5 PC map based on first and second scaled vectors derived from the frequency of eight *Alu* insertion polymorphisms. The trace values for the first, second, and third PCs were 67.6%, 14.2%, and 8.2%, respectively.

tantalizing hint that present-day aboriginal Taiwanese are relict descendants of south Chinese or Indochinese Neolithic populations.

Previously, we showed a directional component for gene flow from Taiwan south through the Philippines and east Indonesia (Melton et al. 1995). Here, we have tried to shed some light on what might be the ultimate source for Austronesian populations in Asia. Although additional genetic markers, populations, and archaeological data would be useful, a scenario for the early colonization of Taiwan could be sketched as follows. During 6000–4000 B.C. Neolithic proto-Austronesian speakers, spreading from early centers of rice cultivation in central and south China, expanded to coastal China and across the Formosa Straits (and to some extent into the northern regions of China), taking with them considerable mtDNA diversity, including the 9-bp deletion. Arriving in Taiwan, they spread over the island and left evidence of the Corded Ware culture. Several or many waves of migration occurred, contributing to the genetic diversity that we observed. Later, introduction onto the west coast of the Lungshanoid culture perhaps introduced more mtDNA variation, but during this time the Austronesian-associated Corded Ware culture was already moving through the Pacific, carrying a limited selection of mtDNAs with and those without the 9-bp deletion. Meanwhile, proto-Austronesian languages were being extinguished in mainland Asia. In more recent times, perhaps the last 3,000 years, invasion by modern populations drove the linguistically isolated Taiwanese into remote areas of the island, where they experienced tribal linguistic and cultural differentiation, but the time depth for this did not allow for tribal differentiation at the mtDNA level. While some gene flow may have maintained overall mtDNA and nuclear DNA homogeneity, genetic drift may have operated to provide the mixed picture observed for the eight *Alu* loci typed in this study. If this scenario is correct, all Taiwanese Austronesian languages most likely sprang from a common deep root and were not introduced by separate population settlements. Although this scenario probably oversimplifies complex human prehistory, the genetic variation described here supports (1) common ancient origins for modern aboriginal Taiwanese populations; (2) general long-term isolation of the Taiwanese from other Asian populations, with the possible exception of coastal northeast Asians; (3) derivation of Taiwanese mtDNAs from a diverse genetic pool with roots in mainland central or southern Asia; and (4) tribal separations in historic times, suggested by the *Alu* data and supported by their extreme cultural diversity. Overall, these conclusions are temporally and geographically consistent with genetic, linguistic, and archaeological studies performed to date.

Table 5

Genetic Distances

Population	Ancestral	Ami	Atayal	Bunun	Paiwan	Moluccas	Nusa Tengrarras	Coastal PNG	Highland PNG	Australia	Philippines	China	Malaysia	Java	Samoa
Ami	.937	...	.0292	.0718	−.0219	.0294	.0477	.3282	...	...	.0063	.0682	.0376	.0696	.3556
Atayal	.7628	.0395	...	−.0053	.0919	.071	.0312	.4752	...	...	.0286	.1084	.087	.0696	.5128
Bunun	.7226	.0541	.0048	...	.1139	.0799	.0244	.5634	...	...	.042	.0788	.0723	.0465	.6021
Paiwan	.5694	.0498	.0299	.0232	...	.0574	.0429	.4001	...	...	.0038	.084	.0561	.0794	.4266
Moluccas	.4563	.0745	.0498	.051	.0144	...	.0172	.4302	...	...	.0334	.0123	−.0044	.0093	.4508
Nusa Tengarras	.3993	.0941	.0601	.0804	.0285	.0136	...	.3554	...	...	.0222	.0257	.0166	.0006	.375
Coastal PNG	.1784	.3011	.2182	.2476	.1586	.097	.0682	...	...	...	.3147	.4789	.4441	.4335	.0001
Highland PNG	.1799	.3735	.2785	.3011	.1882	.1198	.0875	.0048	...	...	...	...	...	...	...
Australia	.3986	.3126	.2122	.2656	.1619	.1371	.0731	.0733	.0575	...	...	...	...	...	...
Philippines	.6645	.0193	.0214	.0263	.0198	.0402	.0518	.174	.2278	.202	...	.0511	.0339	.0402	.3254
China	.4862	.0859	.0313	.0422	.0447	.0138	.0247	.0932	.1253	.1312	.0362	...	−.0018	.0078	.4965
Malaysia	.4655	.0499	.0336	.047	.0194	.0014	.0105	.0908	.127	.137	.0193	.0069	...	.0044	.4599
Java	.7352	.0413	.0219	.072	.0633	.0445	.0451	.1739	.225	.169	.0428	.0317	.0232	...	.4504
Samoa	.5383	.0841	.0753	.1117	.0467	.0204	.013	.1123	.1274	.0901	.0718	.0416	.0207	.0279	...

NOTE.—The top right of the matrix gives the interpopulation AMOVA Φ_{ST} genetic distances for mtDNA SSO types. The bottom left of the matrix gives the interpopulation Nei standard genetic distances for *Alu* allele frequencies. The populations from highland Papua New Guinea and Australia were not SSO typed. PNG = Papua New Guinea.

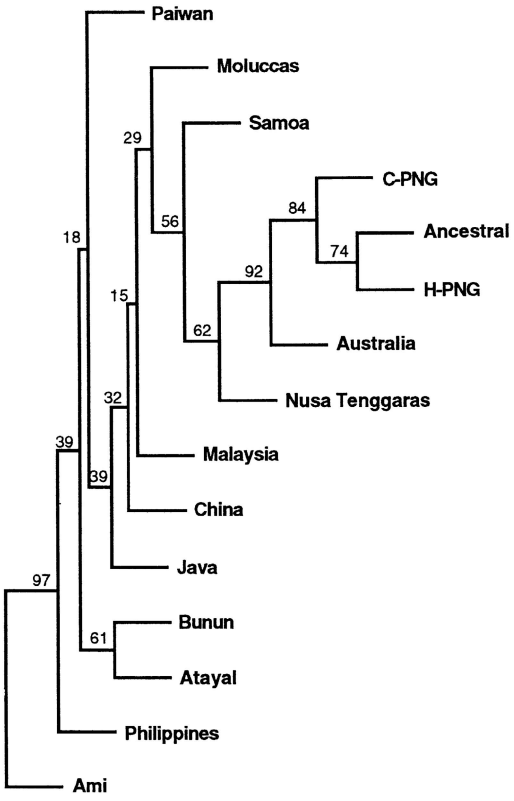

Figure 6 Neighbor-joining tree based on matrix of Nei standard genetic distances, for the eight *Alu* insertion polymorphisms.

Acknowledgments

We thank Alan Redd, Martin Richards, Vincent Macaulay, and Bryan Sykes for their helpful discussions and Martin Richards and Vincent Macaulay for their assistance with Asian sequence data and with the software used. We also appreciate the software and advice provided by Henry Harpending and the computation of the mismatch-distribution confidence intervals by Steve Sherry. T.M. is grateful to Bryan Sykes and the Institute of Molecular Medicine in Oxford, for the facilities provided to complete this project. The project was supported by an National Science Foundation grant (to M.S.).

References

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

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

Batzer MA, Stoneking M, Alegria-Hartman M, Bazan H, Kass DH, Shaikh TH, Novick GE, et al (1994) African origin of human-specific polymorphic Alu insertions. Proc Natl Acad Sci USA 91:12288–12292

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

——— (1985) Prehistory of the Indo-Malayan archipelago. Academic Press, London

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

——— (1995) Austronesian prehistory in Southeast Asia: homeland, expansion, and transformation. In: Bellwood P, Fox JJ, Tryon D (eds) The Austronesians: historical and comparative perspectives. Department of Anthropology, Comparative Austronesian Project, Research School of Pacific and Asian Studies, Australian National University, Canberra, pp 96–111

Benedict P (1942) Thai, Kadai and Indonesian: a new alignment in south-eastern Asia. Am Anthropologist 44:576–601

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

Chai CK (1967) Taiwan aborigines: a genetic study of tribal variations. Harvard University Press, Cambridge

Chang KC (1969) Fengpitou, Tapenkeng, and the prehistory of Taiwan. No. 73 in: Publications in anthropology. Yale University, New Haven

——— (1974) Man in the Choshui and Tatu River Valleys in central Taiwan: preliminary report of an interdisciplinary project, 1972–1973 season. Asian Perspect 17:36–55

Chen KH, Cann H, Chen TC, Van West B, Cavalli-Sforza L (1985) Genetic markers of an aboriginal Taiwanese population. Am J Phys Anthropol 66:327–337

Chen Y-S, Torroni A, Excoffier L, Santa-Benerecetti AS, Wallace DC (1995) Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups. Am J Hum Genet 57:133–149

Chou PL (1959) A study on the blood groups of Paiwan tribe in Mu-Tan district, Pin-Tung prefecture, Formosa. Zinxuigaku Kenkyu 6:127–132

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

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

——— (1993) PHYLIP: Phylogeny Inference Package, version 3.5p. University of Washington, Seattle

Ferrell R (1969) Taiwanese aboriginal groups: problems in cultural and linguistic classification. Monograph 17, Institute of Ethnology, Academia Sinica, Taipei

Goddard WG (1963) The makers of Taiwan. China Publishing, Taipei

——— (1966) Formosa. Macmillan, London

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

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

Hertzberg MS, Mickleson KNP, Serjeantson SW, Prior JF, Trent RJ (1989) An Asian-specific 9-bp deletion of mitochondrial

DNA is frequently found in Polynesians. Am J Hum Genet 44:504–510

Horai S, Hayasaka K (1990) Intraspecific nucleotide sequence differences in the major noncoding region of human mitochondrial DNA. Am J Hum Genet 46:828–842

Horai S, Kondo R, Nakagawa-Hattori Y, Hayashi S, Sonoda S, Tajima K (1993) Peopling of the Americas, founded by four major lineages of mitochondrial DNA. Mol Biol Evol 10:23–47

Horai S, Murayama K, Hayasaka K, Matsubayashi S, Hattori Y, Fucharoen G, Harihara S, et al (1996) mtDNA polymorphism in East Asian populations, with special reference to the peopling of Japan. Am J Hum Genet 59:579–590

Huang M-C (1964) Studies on the distribution of Rh blood types among various racial tribes in Formosa. Jpn J Leg Med 18:135–142

Huang M-C, Sheen N-L (1966) Studies on the distribution of blood types among various racial tribes in Formosa. Paper presented at the 11th Pacific Scientific Congress, Tokyo

Ikemoto S, Ming C-T, Haruyama N, Furumata T (1966) Blood group frequencies in the Ami tribe (Formosa). Proc Jpn Acad 1942:173–177

Kao TY (1958) Taiwan historical events. Cheng Chung Publishing, Taipei

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

Kumar S, Tamura K, Nei M (1993) Manual for MEGA: Molecular Evolutionary Genetic Analysis software. Institute of Molecular Evolutionary Genetics, Pennsylvania State University, University Park

Kutsuna M, Matsuyama M (1939) On the blood groups of Formosa aborigines. J Taiwan Med Assoc 38:1153–1178

Legendre P, Vaudor A (1991) The R package: multidimensional analysis, spatial analysis. Department of Biological Sciences, University of Montreal, Montreal

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

Melton T, Peterson R, Redd AJ, Saha N, Sofro ASM, Martinson J, Stoneking M (1995) Polynesian genetic affinities with Southeast Asian populations as identified by mtDNA analysis. Am J Hum Genet 57:403–414

Melton T, Stoneking M (1996) Extent of heterogeneity in mitochondrial DNA of ethnic Asian populations. J Forensic Sci 41:591–602

Nakajima H, Ohkura K (1971) The distribution of several serological and biochemical traits in East Asia. III. The distributions of gamma globulin (Gm[1], Gm[2], Gm[5] and Inv[1]) and Gc groups in Taiwan and Ryukyu. Hum Hered 21:362–370

Nakajima H, Ohkura K, Huang M-C, Saito R, Seto T (1971) The distribution of several serological and biochemical traits in East Asia. IV. The distribution of the blood groups in Taiwanese mountain aborigines. Jpn J Hum Genet 16:57–68

Nakajima H, Ohkura K, Sheen Y-Z, Chow Z-S, Lee S-P, Orita Y, Masuda Y (1967) The distributions of ABO, MN, Q, Lewis, and Rh blood groups in Taiwan. Jpn J Hum Genet 11:244–251

Nei M (1972) Genetic distances between populations. Am Nat 106:283–292

Ota T (1993) Manual for DISPAN: genetic distance and phylogenetic analysis. Institute of Molecular Evolutionary Genetics, Pennsylvania State University, University Park

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

Redd 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

Rogers A, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. Mol Biol Evol 9:552–569

Ross MD (1994) Some current issues in Austronesian linguistics. In: Tryon DT (ed) Comparative Austronesian dictionary. Mouton de Gruyter, Berlin

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

Sherry S (1994) Documentation for IWaVe: a sequence comparison module. Department of Anthropology, Pennsylvania State University, University Park

Shields GF, Schmeichen AM, Frazier BL, Redd A, Voevoda MI, Reed JK, Ward RH (1993) mtDNA sequences suggest a recent evolutionary divergence for Beringian and northern North American populations. Am J Hum Genet 53:549–562

Soodyall H, Vigilant L, Hill AV, Stoneking M, Jenkins T (1996) mtDNA control-region sequence variation suggests multiple independent origins of an "Asian-specific" 9-bp deletion in sub-Saharan Africans. Am J Hum Genet 58:595–608

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

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

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

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

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

Tang TK, Huang W-Y, Chang Tang C-J, Hsu M, Cheng T-A, Chen K-H (1995) Molecular basis of glucose-6-phosphate

dehydrogenase (G6PD) deficiency in three Taiwan aboriginal tribes. Hum Genet 95:630–632

Torroni A, Schurr TG, Cabell MF, Brown MD, Neel JV, Larsen M, Smith DG, et al (1993*a*) Asian affinities and continental radiation of the four founding Native American mtDNAs. Am J Hum Genet 53:563–590

Torroni A, Sukernik RI, Schurr TG, Starikovskaya YB, Cabell MF, Crawford MH, Comuzzie AG, et al (1993*b*) mtDNA variation of aboriginal Siberians reveals distinct genetic affinities with Native Americans. Am J Hum Genet 53: 591–608

Umetsu K, Yuasa I, Suzuki T, Sun C-S, Pan I-H, Ishida T, Saitou N, et al (1994) Polymorphisms of complement component I and C1R subcomponent of C1 in nine aboriginal Taiwanese populations. Hum Biol 66:339–348

Umetsu K, Yuasa I, Harada A, Suzuki T, Pan I-H, Ishida T, Saitou N, et al (1995) Orosomucoid phenotyping with monoclonal antibodies: polymorphic occurrence of ORM*Q0 in aboriginal Taiwanese populations. Hum Hered 45:181–185

Vigilant L (1990) Control region sequences from African populations and the evolution of human mitochondrial DNA. PhD thesis, Department of Biochemistry and Molecular Biology, University of California, Berkeley

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

Weir BS (1990) Genetic data analysis. Sinauer, Sunderland, MA

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