Hum Genet (2003) 113 : 24–33
DOI 10.1007/s00439-003-0945-1

ORIGINAL INVESTIGATION

Atsushi Tajima · Cheih-Shan Sun · I-Hung Pan
Takafumi Ishida · Naruya Saitou · Satoshi Horai

Mitochondrial DNA polymorphisms in nine aboriginal groups of Taiwan: implications for the population history of aboriginal Taiwanese

Received: 22 November 2002 / Accepted: 28 February 2003 / Published online: 10 April 2003

Abstract Mitochondrial DNA (mtDNA) polymorphisms in the D-loop region and the intergenic COII/tRNA^{Lys} 9-bp deletion were examined in 180 individuals from all nine aboriginal Taiwanese groups: Atayal, Saisiat, Bunun, Tsou, Rukai, Paiwan, Ami, Puyuma, and Yami. A comparison of 563-bp sequences showed that there were 61 different sequence types, of which 42 types were specific to respective aboriginal groups. D-loop sequence variation and phylogenetic analysis enabled the 180 aboriginal lineages to be classified into eight monophyletic clusters (designated C1–C8). Phylogeographic analysis revealed that two (C2 and C4) of the eight clusters were new characteristic clusters of aboriginal Taiwanese and accounted for 8.3% and 13.9% of the aboriginal lineages, respectively. From the estimated coalescent times for the two unique clusters, the mtDNA lineages leading to such clusters were inferred to have been introduced into Taiwan approximately 11,000–26,000 years ago, suggesting ancient immigrations of the two mtDNA lineages. Genetic distances, based on net nucleotide diversities between populations, revealed three distinct clusters that were comprised of northern mountain (Atayal and Saisiat), southern mountain (Rukai and Paiwan), and middle mountain/east coast (Bunun, Tsou, Ami, Puyuma, and Yami) groups, respectively. Furthermore, phylogenetic analysis of 16 human populations (including six other Asian populations and one African population) confirmed that the three clusters for aboriginal Taiwanese had remained largely intact. Each of the clusters (north, south, and middle-east coast) was characterized by a high frequency of a particular lineage (C4, C2, and 9-bp deletion, respectively). This may result from random genetic drift among the aboriginal groups after a single introduction of all the mtDNA lineages into Taiwan, but another plausible explanation is that at least three genetically distinct ancestral populations have contributed to the maternal gene pool of aboriginal Taiwanese.

Introduction

Taiwan is a group of islands situated in the South China Sea and has a population of more than 22 million according to the census data in 2000. The population of Taiwan is largely comprised of early immigrants from mainland China from the 17th century and recent immigrants after World War II, whereas the aboriginal population accounts for around 1.5% of the total population (Jin et al. 1999; Lin et al. 2000). On the basis of linguistic, cultural, and geographical features of such aboriginal people, they can be divided into nine ethnic groups: the Atayal, Saisiat, Bunun, Tsou, Rukai, Paiwan, Ami, Puyuma, and Yami. As shown in Fig. 1, six groups inhabit the central mountain area of the main island of Taiwan: the Atayal and Saisiat in the north, the Bunun and Tsou in the middle, and the Rukai and Paiwan in the south. The Ami and Puyuma inhabit the plains of the eastern coast surrounded by high mountains and the Pacific Ocean, whereas the Yami live on Lanyu Island, located off the eastern coast of the main island. Although there are several hypotheses regarding the origin of these aboriginal people, none has

Electronic database information: the accession numbers for the data in this article are as follows:
DDBJ/EMBL/GenBank databases,
accession nos. AB093682–AB093861

A. Tajima · S. Horai (✉)
Department of Biosystems Science,
Graduate University for Advanced Studies (Sokendai),
240-0193 Hayama, Kanagawa, Japan
Tel.: +81-46-8581575, Fax: +81-46-8581544,
e-mail: horai@soken.ac.jp

C.-S. Sun
Taidon Hospital, Taidon, Taiwan

I-H. Pan
National Taiwan University, Taipei, Taiwan

T. Ishida
Department of Biological Sciences, Graduate School of Science,
University of Tokyo, Tokyo, Japan

N. Saitou
Division of Population Genetics, National Institute of Genetics,
Mishima, Japan

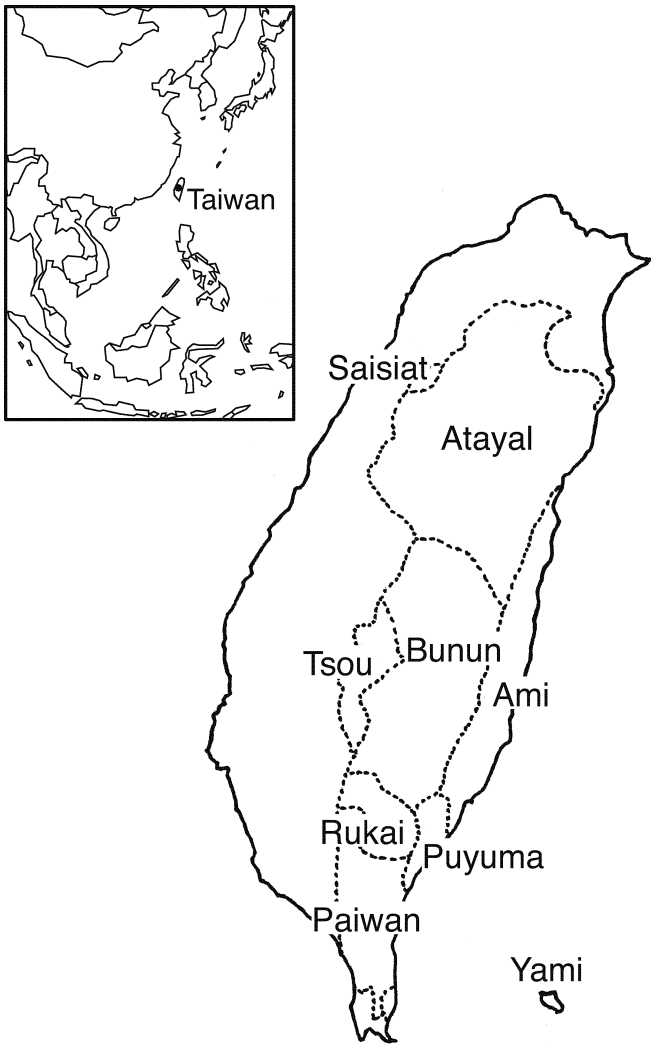


Fig. 1 Geographic distribution of the nine aboriginal groups in Taiwan. The approximate boundaries for these groups are indicated by *dotted lines*. *Insert* Map showing an expanded view of the surrounding regions of Taiwan in east and southeast Asia

gained general acceptance (Melton et al. 1998; Jin et al. 1999).

One of the common characteristics of nine aboriginal Taiwanese groups is that they are included in the Austronesian language group, although their languages can be further classified into four subgroups: Atayalic, Paiwanic, Tsouic, and Malayo-Polynesian (Table 1; Ruhlen 1987). From a linguistic viewpoint regarding the worldwide diversification of the Austronesian language group, Taiwan can be treated as the homeland of the Austronesian languages (Blust 1984–1985; Diamond 1988; Bellwood 1991; Gray and Jordan 2000). In other words, the Austronesian colonization in the Pacific and other regions including Madagascar is thought to have originated in Taiwan (or southern mainland China) approximately 6,000 years ago, although opposition to this has been expressed in recent studies (Oppenheimer 1998; Richards et al. 1998; Su et al. 2000; Oppenheimer and Richards 2001; Lum 2001). Thus, an understanding of the temporal and geographic origins of aboriginal Taiwanese may shed light on questions involved in the Neolithic expansion of Austronesian-speaking populations.

Studies of the genetic variation in human populations allow us to reconstruct their evolutionary history. In particular, mitochondrial DNA (mtDNA) has been widely used to investigate the genetic relationships among closely related populations because of its unique genetic properties, including its high mutation rate and maternal inheritance. Previous studies of mtDNA polymorphisms in aboriginal Taiwanese population have contributed, to some degree, to clarifying their genetic background (Melton et al. 1995, 1998; Sykes et al. 1995). However, such genetic studies have sampled only four (the Atayal, Bunun, Paiwan, and Ami) of the nine aboriginal Taiwanese groups, and therefore comprehensive analyses of all the nine aboriginal groups are necessary to depict the entire scenario regarding the formation of the aboriginal Taiwanese population.

In the present study, we have examined mtDNA polymorphisms in the major noncoding (D-loop) region and intergenic COII/tRNA^{Lys} 9-bp deletion (Cann and Wilson

Table 1 Summary statistics for mtDNA polymorphisms within aboriginal Taiwanese groups

Population	Linguistic affiliation[a]	No. of individuals	k[b]	h[c] (mean±SE)	S[d]	π (%)[e] (mean±SE)
Atayal	Atayalic	20	9	0.816±0.073	23	1.16±0.27
Saisiat	Paiwanic	20	9	0.884±0.049	26	1.35±0.28
Bunun	Paiwanic	20	11	0.921±0.035	25	1.15±0.25
Tsou	Tsouic	20	10	0.905±0.037	26	1.21±0.26
Rukai	Tsouic	20	14	0.916±0.055	29	1.20±0.25
Paiwan	Paiwanic	20	8	0.842±0.060	20	1.17±0.29
Ami	Paiwanic	20	13	0.937±0.035	27	1.10±0.25
Puyuma	Paiwanic	20	11	0.916±0.041	30	1.51±0.30
Yami	Malayo-Polynesian	20	10	0.895±0.043	22	1.36±0.29
total		180	61	0.973±0.004	67	1.36±0.25

[a]According to Ruhlen (1987), in which all are members of Austronesian group
[b]Number of mtDNA sequence types
[c]Haplotype diversity
[d]Number of polymorphic sites
[e]Nucleotide diversity (%)

1983) in order to investigate the genetic background and relationships among all nine aboriginal Taiwanese groups. These analyses have revealed that the maternal gene pool of the aboriginal Taiwanese consisted of eight mtDNA lineage clusters, two of which are rare in other Asian populations. Phylogeographic analysis of such two unique lineages has provided new insights into the population history of the aboriginal Taiwanese.

Materials and methods

Subjects and DNA samples

We examined a total of 180 unrelated individuals from nine aboriginal Taiwanese groups (20 each from the respective groups). Genetic studies related to this research have previously established the presence of serum protein polymorphisms (Umetsu et al. 1994, 1995; Yuasa et al. 2001), red cell enzyme polymorphisms (Jin et al. 1995, 1999), and the prevalence of human T-lymphotropic retrovirus (Ishida et al. 1993) in all of the nine aboriginal groups. All the individuals gave their informed consent prior to their inclusion in this study. Total DNA was isolated from the buffy coat in blood by treatment with sodium dodecyl sulfate and proteinase K and subsequent phenol/chloroform extraction.

Amplification and direct sequencing of the D-loop region of mtDNA

A fragment of the D-loop region of mtDNA was amplified with primer pair A and E (see below) by the polymerase chain reaction (PCR), as described elsewhere (Fucharoen et al. 2001; Qian et al. 2001). The amplified PCR products were purified on MicroSpin S-400 HR columns (Amersham Biosciences, Piscataway, N.J., USA) and were directly sequenced with primers B–E. Nucleotide sequences of the primers were as follows: A, 15897-5′-GTATAAACTAATACACCAGTCTTGT-3′-15921; B, 15985-5′-AGCACCCAAAGCTAAGATTC-3′-16004; C, 16204-5′-AGCAAGTACAGCAATCAACC-3′-16223; D, 16451-5′-GCGAGGAGAGTAGCACTCTT-3′-16432; E, 100-5′-CAGCGTCTCGCAATGCTATCGCGTG-3′-76. The notation of Anderson et al. (1981) was used for numbering of bases. Sequencing reactions were performed with the Dye Terminator Cycle Sequencing FS Kit (Applied Biosystems, Foster City, Calif., USA) according to the manufacturer's instructions. Nucleotide sequences of both strands were determined with an automated DNA sequencer (ABI model 377).

Typing of the COII/tRNA^{Lys} intergenic deletion

The mtDNA 9-bp deletion in the COII/tRNA^{Lys} intergenic region was examined by the PCR method as described by Horai et al. (1996). The amplified fragments (100 bp or 91 bp in length) encompassing the intergenic region were separated by electrophoresis on 4% agarose gels, followed by staining with ethidium bromide for typing.

Data analyses

A total of 563 nucleotide sites (positions 16048–16569 followed by positions 1–41 (numbering according to Anderson et al. 1981) were compared to estimate the number of nucleotide substitutions per site between individual sequences under the two-parameter model of nucleotide substitutions (Kimura 1980). On the basis of the estimated number of nucleotide substitutions, a phylogenetic tree for the mtDNA sequences was constructed by using the neighbor-joining (NJ) method (Saitou and Nei 1987). The reliability of the NJ tree was analyzed by the bootstrap method (Felsenstein 1985) with 1,000 replications.

Nucleotide diversity within and between populations (d_X and d_{XY}, respectively) and net nucleotide diversity between populations (d_A; Nei and Li 1979) were computed with the SENDBS program (provided by N. Takezaki), by using Kimura's two-parameter correction. A phylogenetic tree for human populations was constructed from the matrix of pairwise d_A distances by the NJ method. At the intrapopulational level, haplotype diversity (h: equivalent to heterozygosity for diploid genes; Nei 1987) was calculated with ARLEQUIN software ver 2.001 (Schneider et al. 2001). Statistical comparison was performed by using the χ^2-test; $P<0.05$ was considered significant.

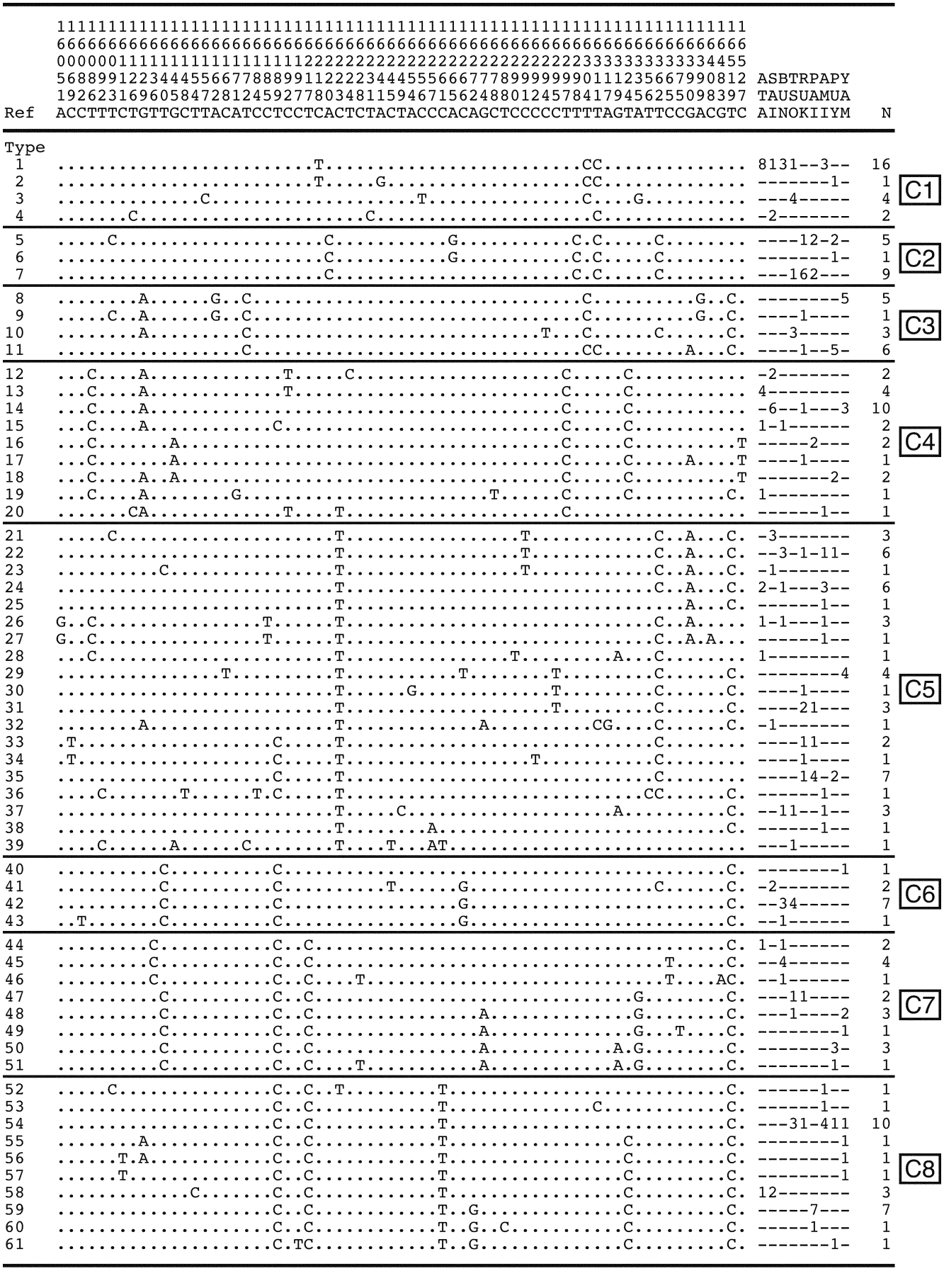
Results and discussion

Sequence diversities in the D-loop region of mtDNA

Nucleotide sequences of 563-bp fragments of the D-loop region were determined for 180 individuals from nine aboriginal Taiwanese groups. There were 61 distinct types of sequence defined by 67 polymorphic sites, as summarized in Fig. 2. A total of 42 types was unique to their respective groups, whereas 19 were shared by 105 individuals (58.3% of the aboriginal population) among two or more groups. Fourteen of the 42 unique types were shared by two or more individuals within each group (a total of 47 individuals; 26.1% of the aboriginal population), whereas each of the remaining 28 types was found in only one individual.

As shown in Table 1, haplotype diversities (h) at the sequence type level varied in the various aboriginal Taiwanese groups, ranging from 0.816 for the Atayal to 0.937 for the Ami. The total diversity for aboriginal Taiwanese was 0.973, which was lower than corresponding estimates for most of the other east and southeast Asian populations (e.g., Yao et al. 2002b). The relatively low haplotype diversities both for each aboriginal group and for the entire population are consistent with recent findings regarding polymorphisms of HLA genes (Lin et al. 2000; Chu et al. 2001) and several serum protein loci (Yuasa et al. 2001). At the nucleotide level, however, nucleotide diversities (π) for the nine aboriginal groups were in the range of 1.10% (for the Ami) to 1.51% (for the Puyuma; Table 1). The overall diversity for the entire aboriginal population (1.36%) was the same magnitude as that for east Asian populations (1.34%; Horai et al. 1996), indicating a high level of DNA variation in the contemporary aboriginal Taiwanese population.

Fig. 2 Nucleotide sequence differences in the D-loop region of mtDNA for 180 aboriginal Taiwanese. The positions of 67 polymorphic sites observed are numbered according to the published reference sequence for human mtDNA (Anderson et al. 1981), as indicated by an *uppercase* letter (A, T, G, or C) *below* the *vertical* five-digit *numerals*. For other sequences (*types 1–61*), only the differences from the reference sequence are indicated, whereas *dots* show identity with the reference. The number of individuals for each sequence type from nine aboriginal Taiwanese groups is represented *right*, where abbreviated population names are as follows: *ATA* Atayal, *SAI* Saisiat, *BUN* Bunun, *TSO* Tsou, *RUK* Rukai, *PAI* Paiwan, *AMI* Ami, *PUY* Puyuma, *YAM* Yami. The total number (*N*) for each type is shown in the column of the extreme right. *Boxed letters right* Phylogenetic clusters characterized in Fig. 3 ▶


Ref
Type
C1
C2
C3
C4
C5
C6
C7
C8

Two unique lineages of mtDNA in aboriginal Taiwanese

Figure 3 shows a phylogenetic tree for the 180 sequences of the D-loop region from nine aboriginal Taiwanese groups. This unrooted NJ tree revealed that the 180 mtDNA lineages could be classified into eight monophyletic clusters (named C1–C8), although the lineages from different aboriginal groups were intermingled in each cluster. All of the eight clusters except C1 possessed either specific combinations of three or four shared polymorphic sites (C2, C3, C6, C7, and C8) or single unique polymorphisms (C4 and C5; Fig. 2). In particular, an A-to-C transversion at 16220 and a T-to-C transition at 16298 were unique polymorphisms among 15 individuals in C2, whereas a T-to-C transition at 16297 was specific to C4. Thus, most of these phylogenetic clusters seem to retain the ancestral polymorphic state before genetic diversification within and among clusters, thereby suggesting several distinct maternal lineages in the aboriginal Taiwanese population.

To ensure the robustness of the phylogenetic clusters further, we inferred the mtDNA haplogroup status of the clusters from their respective D-loop sequence motifs. According to mtDNA haplogroup nomenclature based on complete mtDNA sequences from Asian populations (Kivisild et al. 2002), clusters C1, C3, C6, C7, and C8 would correspond to haplogroups F*, F1a, B5a, B4*(+B4b), and B4a, respectively, whereas C5 contained several other haplogroups such as haplogroup D belonging to Eurasian superhaplogroup M. Thus, these haplogroup assignments reveal that most of the clusters appear to be monophyletic at the haplogroup level, although the present cluster identification in the NJ tree is somewhat arbitrary. Moreover, because two haplogroups (B and F) belong to Eurasien super-haplogroup N (Kivisild et al. 2002; Yao et al. 2002a),

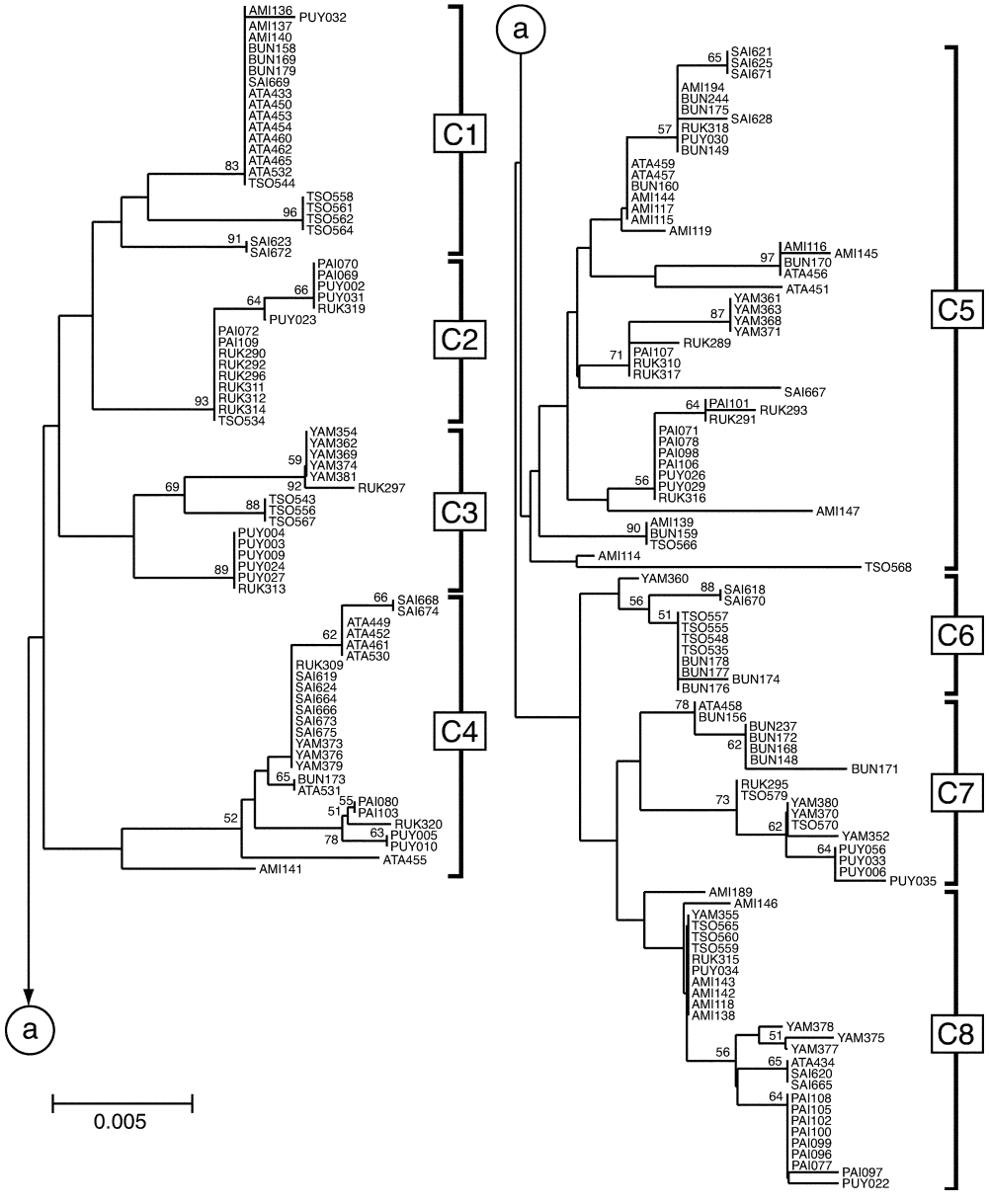


Fig. 3 Phylogenetic tree for 180 mtDNA sequences from nine aboriginal Taiwanese groups. Bootstrap values (more than 50%) are attached to the internal branches. The *scale* for the distance is shown *bottom left*. The eight monophyletic clusters in the tree are indicated by *brackets* with *cluster numbers* (C1–C8). The individual sequences at the tips of branches are represented by the abbreviations of population names (as in Fig. 2) followed by the three-digit *numerals* for the sample numbers

Table 2 Frequency distribution of mtDNA phylogenetic clusters in nine aboriginal Taiwanese groups

Population	Frequency (%)								
	C1	C2	C3	C4	C5	C6	C7	C8	9-bp deletion (C6+C7+C8)
Atayal	40	–	–	30	20	–	5	5	10
Saisiat	15	–	–	40	25	10	–	10	20
Bunun	15	–	–	5	30	20	30	–	50
Tsou	25	5	15	–	10	20	10	15	45
Rukai	–	35	10	10	35	–	5	5	10
Paiwan	–	20	–	10	30	–	–	40	40
Ami	15	–	–	5	50	–	–	30	30
Puyuma	5	15	25	10	15	–	20	10	30
Yami	–	–	25	15	20	5	15	20	40
average	12.8	8.3	8.3	13.9	26.1	6.1	9.5	15.0	30.6
π (%)[a]±SE	0.393±0.136	0.177±0.123	0.610±0.230	0.404±0.126	0.791±0.126	0.182±0.087	0.603±0.211	0.329±0.136	

[a]Nucleotide diversity (%) within cluster. Standard error (SE) of each diversity is calculated by the bootstrap method (500 replications)

the two super-haplogroups contribute to the maternal gene pool of modern aboriginal Taiwanese. In general, haplogroups B, F, and M constitute the majority of mtDNAs in southeast Asian populations (Schurr and Wallace 2002), and hence the present findings indicate that aboriginal Taiwanese have some degree of genetic affinity with southeast Asian populations in terms of mtDNA haplogroup distribution.

On the other hand, we were not able to determine mtDNA haplogroups for clusters C2 and C4, which accounted for 8.3% and 13.9% of the aboriginal mtDNA lineages, respectively (Table 2), because they had no characteristic sets of D-loop mutations required for assignments. On the basis of the previous findings of mtDNA polymorphisms in aboriginal Taiwanese (Melton et al. 1998), two of 28 aboriginal sequences apparently belong to clusters C2 (sample PAI16: A16220C-T16298C-T16311C-T16362C) and C4 (ATA13: T16086C-G16129A-T16297C-T16324C). These two clusters remain monophyletic in an NJ tree rooted with two African orthologus sequences (accession nos. D38112 and AF381988) from mtDNA haplogroup L1a (Maca-Meyer et al. 2001; data not shown). Thus, the inability to assign mtDNA haplogroups for the two clusters implies the presence of unique haplogroups or new branches of previously defined haplogroups in aboriginal Taiwanese groups. The currently available sequence data in the DDBJ/EMBL/GenBank databases reveal no identical sequence types belonging to C2 (types 5–7) and C4 (types 12–20) in other human populations, whereas the related lineages to each cluster have been found at very low frequencies. For example, each of two haplotypes, viz., haplotype 119 (C16168T-A16220C-A16265G-T16298C-T16362C; C2) and 81 (T16086C-T16297C-T16324C; C4) in Sykes et al. (1995) has been found in one Philippine. Thus, the two phylogenetic clusters are not specific to aboriginal Taiwanese but are rare in other Asian populations.

Cluster C2 was observed in four (the Tsou, Rukai, Paiwan, and Puyuma) of the nine aboriginal Taiwanese groups, whereas all of the nine groups except Tsou exhibited C4 at various frequencies (Table 2). Nucleotide diversities for the coalescence of the two clusters (C2 and C4) were estimated as 0.177±0.123% and 0.404±0.126%, respectively (Table 2). Based on complete mtDNA sequences from three human individuals (Horai et al. 1995), we calculated an average substitution rate for the 563-bp mtDNA segment examined in this study as 7.8×10^{-8}/site per year. Assuming this rate under a panmictic population model with constant effective population size, we estimated coalescent times for the two clusters as 11,000±7,900 years for C2 and 26,000±8,100 years for C4. These estimates for the clusters would be linked to their arrival times in Taiwan if there were little or no diversity within the respective founding lineages. As described above, the mtDNA sequences related to the two clusters have been infrequently observed in neighboring Asian populations. Therefore, it is plausible that the genetic diversification within clusters C2 and C4 occurred mainly in Taiwan, but not prior to settlement of individuals carrying the clusters in Taiwan. This interpretation suggests an ancient genetic isolation of the mtDNA lineages belonging to the two clusters in the aboriginal Taiwanese. The implications for this genetic isolation will be discussed later in Genetic relationships among aboriginal Taiwanese groups.

Distribution of COII/tRNA^{Lys} intergenic 9-bp deletion

A deletion polymorphism in a 9-bp tandem repeat of the mtDNA COII/tRNA^{Lys} intergenic region is one of the characteristics not only of Asians (Horai and Matsunaga 1986; Stoneking and Willson 1989; Ballinger at al. 1992; Harihara et al. 1992), but also of the Asian descendents, such as Pacific populations (Hertzberg et al. 1989; Hagelberg and Clegg 1993; Lum et al. 1994) and Native Americans (Schurr et al. 1990; Ward et al. 1991; Torroni et al. 1992; Horai et al. 1993). In particular, the geographic distribution of this deletion in the Pacific region has been associated with past range expansions of Austronesian-language speakers (Melton et al. 1995, 1998; Redd et al. 1995; Sykes et al. 1995).

Consistent with the preliminary findings in six aboriginal Taiwanese groups (Horai 1991), the aboriginal groups exhibited the mtDNA 9-bp deletion at various frequencies, ranging from 10% for the Atayal and Rukai to 50% for the Bunun (Table 2). These differences among the aboriginal groups were statistically significant at the 5% level (P=0.039; χ^2-test), indicating an uneven distribution of this deletion among aboriginal Taiwanese. The average frequency in aboriginal Taiwanese was 30.6%, which was slightly lower than those from previous studies in four (the Atayal, Bunun, Paiwan, and Ami) of the entire aboriginal group (41.5%, Melton et al. 1995; 36%, Sykes et al. 1995). A total of 55 aboriginal Taiwanese with the 9-bp deletion appeared in the three clusters C6, C7, and C8 on the basis of D-loop sequences (Fig. 3). The nucleotide sequences (Fig. 2) revealed that all the individuals with the deletion showed a specific combination of two polymorphisms (T16189C and T16519C). Thus, this deletion event appears to have occurred once on the background of mtDNA carrying such a characteristic set of mutations.

A unique combination of three transitions (T16217C, A16247G, and C16261T) with the 9-bp deletion has been referred to as the "Polynesian motif" because of its high frequencies in Polynesian populations (Melton et al. 1995; Redd et al. 1995). With reference to the evolutionary history of Polynesian populations and/or Austronesian-language speakers, much attention has been turned to where and when this motif emerged in Asia-Pacific region (e.g., Richards et al. 1998). In this survey, we have found mtDNA sequence types carrying only one (T16217C in cluster C7) or two (T16217C and C16261T in C8) of the three mutations in aboriginal Taiwanese population with the 9-bp deletion. However, there is no A16247G mutation among the aboriginal Taiwanese (Fig. 2). The lack of the Polynesian motif in the aboriginal Taiwanese groups agrees with previous findings (Melton et al. 1995, 1998; Sykes et al. 1995), indicating that the motif itself is unlikely to have arisen in Taiwan.

Genetic relationships among aboriginal Taiwanese groups

To investigate genetic relationships among the nine aboriginal Taiwanese groups, we estimated net (d_A) nucleotide diversities between each pair of the aboriginal groups. Table 3 shows the d_A distances and interpopulational (d_{XY}) and intrapopulational (d_X or d_Y) diversities among the nine aboriginal groups. The d_A distances ranged from 0.038% to 0.226% (between the Atayal and Paiwan). On the basis of the geographic distribution of the aboriginal groups in Taiwan (Table 3), the smallest values (0.038%) were found in two combinations between the middle mountain groups and east coast groups (the Bunun-Ami and the Tsou-Puyuma), suggesting close genetic relationships between these aboriginal groups.

Based on the d_A distances, we constructed a phylogenetic tree for nine aboriginal Taiwanese groups by the NJ method (Fig. 4). The tree showed that there were four clades of two aboriginal groups with bootstrap probabilities over 50%: the Atayal-Saisiat (92%), the Tsou-Yami (69%), the Bunun-Ami (55%), and the Rukai-Paiwan (54%). Furthermore, the two clades of the Bunun-Ami and Tsou-Yami formed a larger cluster, together with the Puyuma. The remaining two clades (the Atayal-Saisiat and the Rukai-Paiwan) correspond to northern and southern mountain groups, respectively. It is worth noting that high frequencies of the unique lineage cluster C2 are found in the southern (Rukai and Paiwan) groups, and those of another cluster (C4) in the northern (Atayal and Saisiat) groups (Table 2). On the other hand, the large cluster of five aboriginal groups suggests close genetic affinities between middle mountain groups and east coast groups including the Yami. In spite of the heterogeneous distribution of the mtDNA 9-bp deletion among nine aboriginal Taiwanese groups, higher frequencies of the deletion (over 30%) are common to all of the five aboriginal groups (Bunun, Tsou, Ami, Puyuma, and Yami) in the

Table 3 Estimates of interpopulational (d_{XY}), intrapopulational (d_X or d_Y), and net (d_A) nucleotide diversities among nine aboriginal Taiwanese groups

Population	Central mountain groups[a]						East coast groups		
	North		Middle		South		On coast		Off coast
	Atayal	Saisiat	Bunun	Tsou	Rukai	Paiwan	Ami	Puyuma	Yami
Atayal	**1.161**[b]	1.297	1.325	1.377	1.331	1.391	1.282	1.445	1.464
Saisiat	0.042	**1.348**	1.394	1.473	1.401	1.421	1.365	1.544	1.479
Bunun	0.170	0.145	**1.149**	1.252	1.318	1.319	1.165	1.388	1.381
Tsou	0.190	0.192	0.070	**1.214**	1.368	1.378	1.271	1.399	1.365
Rukai	0.152	0.129	0.144	0.162	**1.197**	1.290	1.249	1.406	1.493
Paiwan	0.226	0.163	0.160	0.187	0.107	**1.169**	1.266	1.451	1.462
Ami	0.150	0.139	0.038	0.112	0.098	0.130	**1.104**	1.372	1.384
Puyuma	0.110	0.116	0.060	0.038	0.053	0.113	0.066	**1.507**	1.542
Yami	0.203	0.124	0.125	0.076	0.213	0.197	0.151	0.107	**1.362**

[a]The classification is based on the places inhabited by the aboriginal Taiwanese groups(Chu et al. 2001)

[b]All values are multiplied by 100. The figures *on* the *diagonal* (*bold*) refer to d_X (or d_Y), those *above* the *diagonal* give d_{XY}, and those *below* the *diagonal* represent the values of $d_A=d_{XY}-(d_X+d_Y)/2$

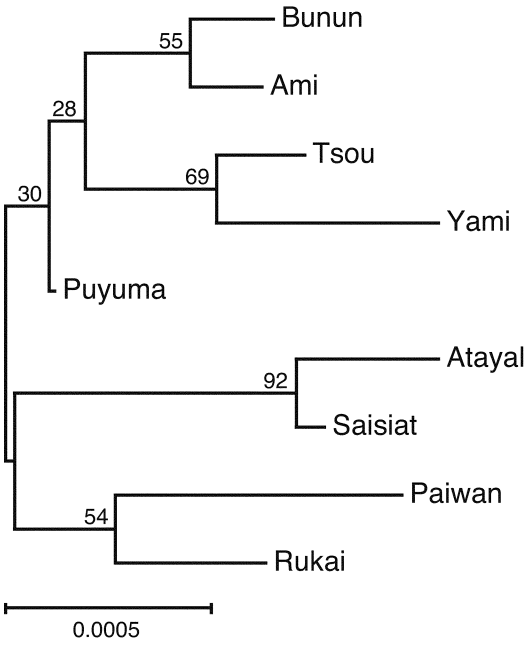


Fig. 4 Neighbor-joining (NJ) tree, showing the relationships of the nine aboriginal Taiwanese groups on the basis of d_A distances with Kimura's two-parameter correction. The numbers shown for the internal branches are the bootstrap probabilities. The *scale* for the genetic distance is shown *below*

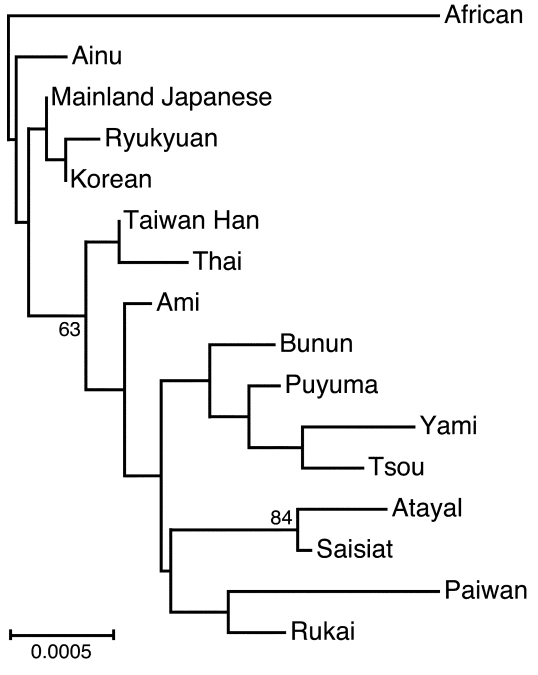


Fig. 5 Neighbor-joining (NJ) tree for the nine aboriginal Taiwanese and other Asian and African populations, on the basis of d_A distances estimated from the nucleotide sequences of 482-bp fragments of the D-loop region (positions 16129–16569 followed by positions 1–41 in the reference sequence of Anderson et al. 1981). The data sources for the seven other human populations are as follows: five east Asians (Horai et al. 1996), Thais Fucharoen et al. 2001), and Africans (Horai and Hayasaka 1990; Vigilant et al. 1991). Bootstrap probabilities (over 50%) are attached to the internal branches. The *scale* for the genetic distance is shown *below*

middle mountain and east coast regions including the Lanyu Island (Table 2). Thus, the present observations imply that the aboriginal groups inhabiting the middle mountain-east coast regions including the Lanyu Island share some extent of a common maternal ancestry. Thus, the phylogenetic relationships among the nine aboriginal groups are likely to fit with their geographic distribution (north, south, and middle-east coat), although the bootstrap support is not so strong.

To infer phylogenetic relationships among the nine aboriginal Taiwanese groups and other human populations, we constructed an NJ tree for a total of 16 populations (Fig. 5), combined with the published data for five east Asian (Horai et al. 1996), Thai (Fucharoen et al. 2001) and African (Horai and Hayasaka 1990; Vigilant et al. 1991) populations. The tree indicated that the three population clusters (north, south, and middle-east coat) for aboriginal Taiwanese remained largely intact, except that the middle-east coat cluster lacked the Ami. The nine aboriginal groups clustered further with Taiwan Han and Thais (bootstrap value 63%), but there appeared to be two separate sub-clusters, which consisted of the aboriginal Taiwanese and Taiwan Han-Thai, respectively. The genetic difference between aboriginal Taiwanese and Taiwan Han at the population level is consistent with those in the phylogenetic analyses undertaken with HLA loci (Lin et al. 2000; Chu et al. 2001), although they have been neighbors on Taiwan for the past four centuries.

The three population clusters (north, south, and middle-east coat) in aboriginal Taiwanese also appear to correspond to the geographic distribution of three mtDNA lineages: C4, C2, and the 9-bp deletion (C6–C8), respectively. This implies that at least three genetically distinct human populations were involved in the prehistoric settlement of Taiwan, although the three lineages (C2, C4, and 9-bp deletion) account only for 52.7% of the aboriginal gene pool studied. Another explanation for the population relationships is that a single introduction of all three mtDNA lineages into Taiwan was coupled with subsequent isolation and random genetic drift among the aboriginal groups. One way of evaluating these scenarios is to determine ancestral source populations of these three mtDNA lineages, especially the geographic origins of the unique lineage clusters C2 and C4. However, the infrequent distribution of the two unique lineages outside Taiwan prevents us from verifying these in such a way. The divergence times for the clusters C2 (11,000±7,900 years) and C4 (26,000± 8,100 years) suggest that mtDNA lineages leading to the two clusters were separately brought into Taiwan. Thus, these postulated human migrations may favor, even if only partially, the former interpretation. There is no discrepancy in dating between this suggestion and the earliest human remains in Taiwan (Lien 1981); modern humans probably settled here 20,000–30,000 years before present (BP). Paleogeographic data (Voris 2000) have suggested that Taiwan was connected to the Asian continent by the exposed continental shelf when sea levels were at 120–75 m below present-day levels, during the last glacial period. Subsequently, at approximately 11,000 years BP

(50 m below the present levels), there were no land connections between Taiwan and the continent, and evidently, Taiwan has been isolated since then. The deglaciation accompanied by sea-level rise at the end of the Pleistocene would result in the characteristic geographic distribution of the two mtDNA lineages (C2 and C4) to Taiwan. At any rate, we should be able to characterize the unique mtDNA lineages more minutely when obtaining complete mtDNA sequences from such lineages and extensive geographic sampling from neighboring human populations.

Linguistic and cultural evidence for Austronesian-language speakers (Blust 1984–1985; Diamond 1988; Bellwood 1991) has been mostly used to interpret the maternal genetic background of aboriginal Taiwanese to date (Melton et al. 1995, 1998; Sykes et al. 1995). In other words, the population history of aboriginal Taiwanese is supposed to date back to 6,000–4,000 BC (Melton et al. 1998), although there is no genetic evidence that excludes the possibility of the earlier arrival of modern humans in Taiwan. The present phylogeographic analysis of the two unique lineage clusters allows us to infer that modern humans carrying the lineages migrated into Taiwan around 11,000–26,000 years ago, suggesting a longer history of the aboriginal people in Taiwan. It should be noted that comparative analyses of the geographic distribution of other distinct mtDNA lineages (clusters C1, C3, C5, C6, C7, and C8: haplogroups B, F, and M) in east and southeast Asia are necessary to address the questions as to when and how the respective lineages were brought into Taiwan. These analyses should also be useful for clarifying the prehistoric dispersal of the Austronesian-speaking populations. This is because the analyses would help us to test two opposing views of aboriginal Taiwanese as a source of the Austronesian expansion (Melton et al. 1995, 1998; Sykes et al. 1995), and an isolated population of Austronesian-language speakers/southeast Asians (Oppenheimer 1998; Su et al. 2000). To attain a deeper understanding of the peopling of Taiwan, the maternal genetic view has also to be integrated with wider knowledge from other genetic studies, such as Y-chromosome haplotype analysis, and many different disciplines.

Acknowledgements We thank Ms. Naoko Anaguchi for technical assistance. This study was supported in part by Grants-in-Aid from the Ministry of Education, Culture, Sports, Science, and Technology of Japan to S. Horai.

References

Anderson S, Bankier AT, Barrell BG, Bruijn MH de, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJ, Staden R, Young IG (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 KH, Wallace DC (1992) Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient Mongoloid migrations. Genetics 130:139–152

Bellwood P (1991) The Austronesian dispersal and the origin of languages. Sci Am 265:70–75

Blust R (1984–1985) The Austronesian homeland: a linguistic perspective. Asian Perspect 26:45–67

Cann RL, Wilson AC (1983) Length mutations in human mitochondrial DNA. Genetics 104:699–711

Chu CC, Lin M, Nakajima F, Lee HL, Chang SL, Juji T, Tokunaga K (2001) Diversity of HLA among Taiwan's indigenous tribes and the Ivatans in the Philippines. Tissue Antigens 58:9–18

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

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

Fucharoen G, Fucharoen S, Horai S (2001) Mitochondrial DNA polymorphisms in Thailand. J Hum Genet 46:115–125

Gray RD, Jordan FM (2000) Language trees support the express-train sequence of Austronesian expansion. Nature 405:1052–1055

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

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

Hertzberg M, Mickleson KN, 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 (1991) A genetic trail of human mitochondrial DNA. In: Mukohata Y (ed) New era of bioenergetics. Academic Press, Tokyo, pp 273–299

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

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, Hayasaka K, Kondo R, Tsugane K, Takahata N (1995) Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. Proc Natl Acad Sci USA 92:532–536

Horai S, Murayama K, Hayasaka K, Matsubayashi S, Hattori Y, Fucharoen G, Harihara S, Park KS, Omoto K, Pan IH (1996) mtDNA polymorphism in east Asian Populations, with special reference to the peopling of Japan. Am J Hum Genet 59:579–590

Ishida T, Pan IH, Horai S, Saitou N, Sun CS (1993) Seroepidemiological survey of human T-lymphotropic retrovirus among indigenous populations in Taiwan. Int J Epidemiol 22:927–930

Jin F, Saitou N, Ishida T, Sun CS, Pan IH, Omoto K, Horai S (1995) Genetic polymorphism of ADA in Taiwan aboriginal populations: new ADA variants detected by isoelectric focusing method. Anthropol Sci 103:227–234

Jin F, Saitou N, Ishida T, Sun CS, Pan IH, Omoto K, Horai S (1999) Population genetic studies on nine aboriginal ethnic groups of Taiwan. I. Red cell enzyme systems. Anthropol Sci 107:229–246

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

Kivisild T, Tolk HV, Parik J, Wang Y, Papiha SS, Bandelt HJ, Villems R (2002) The emerging limbs and twigs of the east Asian mtDNA tree. Mol Biol Evol 19:1737–1751

Lien CM (1981) On the occurrence of fossil *Homo sapiens* in Taiwan. Bull Dept Archaeol Anthropol 42:53–74

Lin M, Chu CC, Lee HL, Chang SL, Ohashi J, Tokunaga K, Akaza T, Juji T (2000) Heterogeneity of Taiwan's indigenous population: possible relation to prehistoric Mongoloid dispersals. Tissue Antigens 55:1–9

Lum JK (2001) The colonization of Remote Oceania and the drowning of Sundaland. In: Jin L, Seielstad M, Xiao C (eds) Genetic, linguistic and archaeological perspectives on human diversity in southeast Asia. World Scientific, Singapore, pp 147–170

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

Maca-Meyer N, González AM, Larruga JM, Flores C, Cabrera VM (2001) Major genomic mitochondrial lineages delineate early human expansions. BMC Genet 2:13

Melton T, Peterson R, Redd AJ, Saha N, Sofro AS, 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, Clifford S, Martinson J, Batzer M, Stoneking M (1998) Genetic evidence for the proto-Austronesian homeland in Asia: mtDNA and nuclear DNA variation in Taiwanese aboriginal tribes. Am J Hum Genet 63:1807–1823

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

Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci USA 76:5269–5273

Oppenheimer S (1998) Eden in the East: the drowned continent of southeast Asia. Phoenix, London

Oppenheimer S, Richards M (2001) Fast trains, slow boats, and the ancestry of the Polynesian islanders. Sci Prog 84:157–181

Qian YP, Chu ZT, Dai Q, Wei CD, Chu JY, Tajima A, Horai S (2001) Mitochondrial DNA polymorphisms in Yunnan nationalities in China. J Hum Genet 46:211–220

Redd AJ, Takezaki N, Sherry ST, McGarvey ST, Sofro AS, 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

Richards M, Oppenheimer S, Sykes B (1998) mtDNA suggests Polynesian origins in Eastern Indonesia. Am J Hum Genet 63: 1234–1236

Ruhlen M (1987) A guide to the world's languages, vol 1. Classification. Stanford University Press, Stanford

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

Schneider S, Roessli D, Excoffier L (2001) Arlequin ver. 2.001: a software for population genetic analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland

Schurr TG, Wallace DC (2002) Mitochondrial DNA diversity in southeast Asian populations. Hum Biol 74:431–452

Schurr TG, Ballinger SW, Gan YY, Hodge JA, Merriwether DA, Lawrence DN, Knowler WC, Weiss KM, Wallace DC (1990) Amerindian mitochondrial DNAs have rare Asian mutations at high frequencies, suggesting they derived from four primary maternal lineages. Am J Hum Genet 46:613–623

Stoneking M, Wilson AC (1989) Mitochondrial DNA. In: Hill A, Serjeantson S (eds) The colonization of the Pacific: a genetic trail. Clarendon, Oxford, pp 215–245

Su B, Jin L, Underhill P, Martinson J, Saha N, McGarvey ST, Shriver MD, Chu J, Oefner P, Chakraborty R, Deka R (2000) Polynesian origins: insights from the Y chromosome. Proc Natl Acad Sci USA 97:8225–8228

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

Torroni A, Schurr TG, Yang CC, Szathmary EJ, Williams RC, Schanfield MS, Troup GA, Knowler WC, Lawrence DN, Weiss KM, Wallace DC (1992) Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations. Genetics 130:153–162

Umetsu K, Yuasa I, Suzuki T, Sun CS, Pan IH, Ishida T, Saitou N, Horai S (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 IH, Ishida T, Saitou N, Horai S (1995) Orosomucoid phenotyping with monoclonal antibodies: polymorphic occurrence of *ORM1*Q0* in aboriginal Taiwanese populations. Hum Hered 45:181–185

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

Voris HK (2000) Maps of Pleistocene sea levels in southeast Asia: shorelines, river systems and time durations. J Biogeography 27:1153–1167

Ward RH, Frazier BL, Dew-Jager K, Pääbo S (1991) Extensive mitochondrial diversity within a single Amerindian tribe. Proc Natl Acad Sci USA 88:8720–8724

Yao YG, Kong QP, Bandelt HJ, Kivisild T, Zhang YP (2002a) Phylogeographic differentiation of mitochondrial DNA in Han Chinese. Am J Hum Genet 70:635–651

Yao YG, Nie L, Harpending H, Fu YX, Yuan ZG, Zhang YP (2002b) Genetic relationship of Chinese ethnic populations revealed by mtDNA sequence diversity. Am J Phys Anthropol 118:63–76

Yuasa I, Umetsu K, Ago K, Sun CS, Pan IH, Ishida T, Saitou N, Horai S (2001) Population genetic studies on nine aboriginal ethnic groups of Taiwan. II. Serum protein systems. Anthropol Sci 109:257–273