

Rachel A. Chow · Jose L. Caeiro · Shu-Juo Chen  
Ralph L. Garcia-Bertrand · Rene J. Herrera

## Genetic characterization of four Austronesian-speaking populations

Received: 7 June 2005 / Accepted: 2 August 2005 / Published online: 6 October 2005  
© The Japan Society of Human Genetics and Springer-Verlag 2005

**Abstract** Ascertaining the genetic relationships between Austronesian populations is paramount to understanding their dispersal throughout the islands of the Pacific and Indian Oceans. The start of the Austronesian expansion has been dated to approximately 6,000 years ago, and from linguistic and archeological evidence, the origin of this dispersal may have been the island of Formosa. Consequently, the Taiwanese aboriginal populations and their phylogenetic relationship to the Austronesian-speaking groups from Madagascar at the occidental fringes of the expansion are of great interest. In this study, allelic frequencies from six polymorphic point mutation loci were assessed in the Austronesian-speaking populations of Madagascar, the Atayal aborigines of Formosa, and the general populations of Bali and Java. These allelic frequencies were compared and analyzed with the corresponding values from eight other worldwide populations from geographically targeted regions. The group from Madagascar is genetically distinct from their east-African neighbor from Zimbabwe. Our data also indicates that the Ami and the Atayal aborigines in the island of Taiwan, which occupy adjacent territories, differ sharply genetically. Genetic differences were also found between the populations of Bali and Java, belying their geographical proximity. Our

results indicate that the east-African population from Madagascar phylogenetically segregates intermediate between mainland east-African and east-Asian groups, corroborating linguistic data indicating the Austronesian influence on this population.

**Keywords** Madagascar · Austronesian · HLA-DQ $\alpha$ 1 · Point mutations · Ami · Atayal

### Introduction

The colonization of the islands of the Pacific and Indian Oceans that started approximately 6,000 years ago represents one of the major human dispersals (Cann 2001). There are many theories surrounding the peopling of these islands and no single one enjoys full approval (Jin et al. 1999). Both archeological and linguistic data indicate that Austronesian-speaking populations were responsible for this expansion (Diamond 1988). Currently, there are 1,200 Austronesian languages and 200 million native speakers spanning approximately two-thirds the circumference of the globe (Bellwood 1991; Diamond 2000). Austronesian languages are found as far west as Madagascar, off the coast of Mozambique in east Africa, to the reaches of Easter Island in the remote east Pacific, and include the languages of the islands of Southeast Asia, such as those of the aborigines of Formosa (Kirch 2000). The discontinuous pattern of insular habitation led to social and linguistic isolation which resulted in unique and autonomous cultural evolution in distant islands. In turn, the art, linguistics, physical anthropology, and genetics of the present day native populations can be used to illuminate the Austronesian diaspora, a past that is complicated to reconstruct since no preserved Austronesian writings are dated earlier than 670 AD near the end of the language family's expansion (Diamond 2000). From the compiled archeological evidence, the Express Train to Polynesia theory was formulated to explain the Austronesian expansion.

R.A. Chow · R.J. Herrera (✉)  
Department of Biological Sciences, Florida International  
University, University Park Campus, Miami, FL 33199, USA  
E-mail: herrerar@fiu.edu  
Tel.: +1-305-3481258  
Fax: +1-305-3481259

J.L. Caeiro  
Seccion de Antropoloxia, Facultade de Bioloxia,  
Universidade de Santiago de Compostela, Galicia, Spain

S.-J. Chen  
Department of Anthropological Sciences, Stanford University,  
Stanford, CA 94305, USA

R.L. Garcia-Bertrand  
Department of Biological Sciences, Colorado College,  
Colorado Springs, CO 80903, USA

According to this theory, in this major human spread, people migrated from southern China to Formosa, the Philippines, Indonesia, and finally to the coast of New Guinea and Polynesia (Diamond 1988). Yet, there is still some debate over whether the actual homeland is in Taiwan, southern China, or elsewhere, including eastern Indonesia (Oppenheimer 1998; Richards et al. 1998; Su et al. 2000; Lum 2001; Oppenheimer and Richards 2001). Depending on the actual source population, the island of Formosa may represent a gateway to the Austronesian expansion.

Taiwan, also known as Formosa from colonial times, has been home to Austronesian tribal groups since before 4,000 BC (Ruhlen 1994). Furthermore, Taiwan is the location of the greatest Austronesian linguistic variety with nine of the ten subgroups present, and thus, according to linguistic theory, is believed to be near the origin of the language family (Diamond 2000). The oldest archeological sites are in Taiwan and are known to belong to the Ta-p'en-k'eng culture, the earliest ceramic phase (Kirch 1997, 2000). This ceramic cultural tradition then diffused from Formosa through the Philippines and into the equatorial islands of Southeast Asia. The phylogeny of Austronesian languages derived from previous linguistic studies parallels the pattern of island settlement supported by this archeological data with Taiwan as the projected homeland (Gray and Jordan 2000).

There are nine tribal groups that currently constitute the aboriginal population of Formosa: Ami, Atayal, Bunun, Paiwan, Puyama, Rukai, Saisiya, Tsou, and Yami (Jin et al. 1999). They comprise approximately 1.5% of Taiwan's total population which currently numbers 21 million (Chu 1997; Jin et al. 1999). The remainder of the population consists mainly of the Han Chinese who have been instrumental in the displacement and redistribution of the native groups since ancient times to the rugged inland mountains and the eastern coastline (Kao 1958; Chai 1967). The Atayal is Taiwan's second largest tribe after the Ami. Currently, there are about 90,000 Atayal tribal members residing in a large area in northern Formosa.

There are few genetic studies involving the aborigines of Formosa, providing a limited understanding on the origins of each tribe and the relationship between the tribes and other worldwide populations (Sewerin et al. 2002). Some investigations have examined the blood types and their frequencies among the tribes (Ikemoto et al. 1942; Huang 1964; Nakajima and Ohkura 1971). Tajima and collaborators studied all nine of the aboriginal groups from Formosa and concluded that specific mtDNA lineages were introduced into Taiwan 11,000–16,000 years ago suggesting ancient migrations of two mtDNA lineages (Horai et al. 1995; Tajima et al. 2003). In addition, Horai's group delineated three distinct clades of tribes representing the northern mountain (Atayal and Saisiat), southern mountain (Rukai and Paiwan), and the middle mountain/east coast (Bunun, Tsou, Ami, Puyama, and Yami) geographical zones

(Tajima et al. 2003). Melton and colleagues found evidence supporting a common source for the tribes in south-central China and hypothesized that in 6,000–4,000 BC Neolithic proto-Austronesians migrated from mainland China to Taiwan bringing with them the mtDNA 9 base pair deletion (Melton et al. 1998). Melton et al. (1998) further suggested that since the origin of the "Polynesian motif" can be traced to Taiwan, the island of Formosa may be the origin of the expansion. The genetic influence of these aboriginal groups from Formosa in Melanesia has also been supported by other mtDNA investigations (Merriwether et al. 1999). Other mtDNA studies have suggested that the Indonesian archipelago may be the origin of the expansion some 17,000 years ago (Richards et al. 1998). Along these lines, Y-chromosome data have been interpreted to suggest that the origin for the Austronesian-speaking populations of insular Southeast Asia and Oceania may be the pre-Neolithic groups that first populated the area (Capelli et al. 2001). Some researchers have postulated that the Taiwanese aborigines may represent population isolates outside the mainstream of the Austronesian dispersal (Oppenheimer 1998; Su et al. 2000). ABO blood group data indicate strong genetic affinities among Austronesian groups from Melanesia and Polynesia that are not shared with non-Austronesian populations from the Pacific (Ohashi et al. 2004). It is evident that these studies provide conflicting results and portray a history that remains fragmented and vague.

More recently, Sewerin and colleagues focused on the genetic polymorphisms in six-point mutation loci of the Ami tribe and provided evidence for the genetic uniqueness of the group (Sewerin et al. 2002). The results of studies such as Sewerin's have generated a series of questions involving the intertribal associations and the relationships between the aborigines and other Austronesian-speaking populations. In the present study, using the same six genetic markers, we provide information on the genetic diversity and phylogenetic affinities of the Austronesian-speaking Madagascar population. In addition, we report on the genetic relationship between the Atayal and the Ami, along with those of the Atayal and Ami to other Austronesian populations from Bali and Java. Genotypic and allelic frequencies from these loci are reported for the first time from the populations of Bali, Java, Madagascar and the Atayal tribe of Taiwan, and compared to data from eight other geographically targeted worldwide reference groups.

---

## Materials and methods

### Populations studied

Sixty-nine unrelated individuals from the east-African island of Madagascar were sampled. This collection from Madagascar represents the general population of the island. In addition, 40 unrelated individuals from the

Atayal tribe of Formosa were collected. The Atayalan samples were procured from several villages within their traditional territory, as illustrated in Fig. 1. Also studied were 24 individuals from Java and 34 from Bali. The population from Bali was collected in sites from throughout the 2,200-square-mile island. In Java, the sampling was performed in locations all over the island. Java and Bali are both islands within the borders of Indonesia whereas Madagascar is located about 250 miles off the southeast coast of Africa and approximately 6,000 miles from the island of Formosa. Individuals from Madagascar are known to speak Malagasy, an Austronesian language. Figure 1 illustrates the geographical position of the populations studied. These four populations were compared with the data from eight other worldwide populations previously reported in the literature. Individuals were identified as Atayal members by tracing back biographical information for at least two generations. Table 1 lists the populations examined and their location along with their reference.

#### Blood collection and DNA purification

The samples were collected as whole blood in EDTA Vacutainer tubes in adherence to the guidelines set forth by Florida International University's Institutional Review Board. Cells were lysed and leukocyte nuclei

were isolated by centrifugation, followed by digestion of nuclei with proteinase K as previously described (Antunez de Mayolo et al. 2002). Total genomic DNA was then isolated by a standard phenol/chloroform extraction and ethanol precipitation (Luis et al. 2003).

#### DNA amplification and genotyping

The extracted DNA was amplified by polymerase chain reaction (PCR) using the AmpliType PM-DQ $\alpha$ 1 PCR Amplification and Typing Kit (Perkin Elmer Corp, Norwalk, CT, USA) using the conditions specified by the manufacturer. PCR was performed using a Perkin-Elmer 480 thermal cycler. Following amplification, samples were screened for successful amplification by electrophoresis in a 1 $\times$  TAE, 2% agarose gel followed by ethidium bromide staining and ultraviolet (UV) visualization. The HLA-DQ $\alpha$ 1, LDLR, GYPA, HBGG, D7S8, and GC loci were then genotyped for each sample. The chromosomal locations of the loci are: HLA-DQA1, 6p21.3; LDLR, 19p13.1-13.3; GYPA, 4q28-31; HBGG, 11p15.5; D7S8, 7q22-31.1 and GC, 4q11-13 (Sewerin et al. 2002). Typing of these samples involves reverse dot blot technology with allele-specific oligonucleotide probes bound to strips that allow the typing of multiple loci at one time. All alleles can be typed from a single PCR reaction. These loci have been extensively

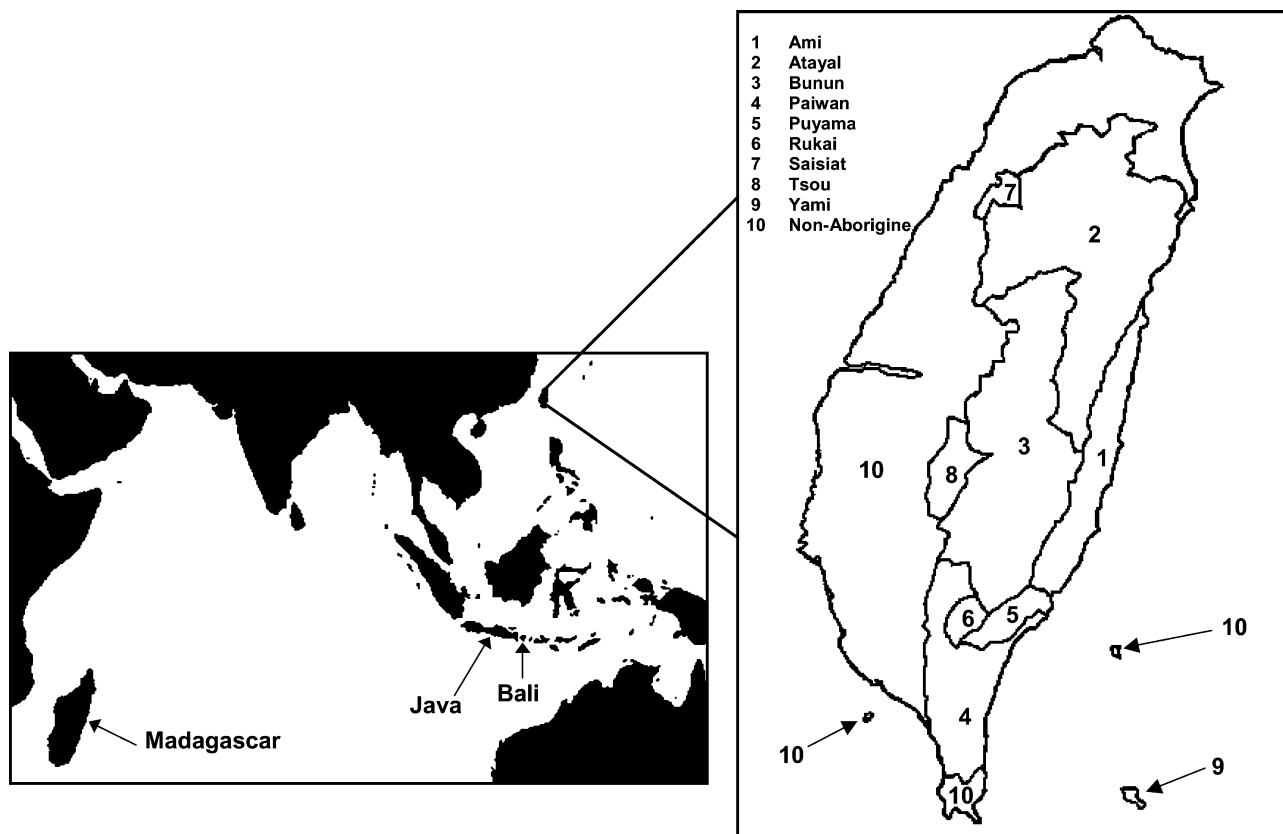


Fig. 1 Geographical map of the populations studied with expanded map of Taiwan and the distribution of its aboriginal groups

**Table 1** Name and origin of populations

Population	Location	References
African American	General population from the United States	Budowle et al. (1995)
Ami	Aborigines from east-central Taiwan	Sewerin et al. (2002)
Alaska Eskimo	Native Americans from North Slope Borough, Alaska	Walkinshaw et al. (1996)
Atayal	Aborigines from northern Taiwan	Present study
Bali	General population from Bali	Present study
Basque	Individuals from the Goierri Valley, Gipuzkoa Province, Spain	Brown et al. (2000)
Chinese	General population from mainland China	Huang and Budowle (1995)
Java	General population from Java	Present study
Madagascar	General population from Madagascar	Present study
Navajo	Nadene-speaking tribe from southwest United States	Scholl (1996)
United States Caucasian	General population from United States	Budowle et al. (1995)
Zimbabwe	Black African population from Zimbabwe	Wolfarth et al. (2000)

investigated in population genetics studies and were found to satisfy Hardy–Weinberg equilibrium expectations (Sewerin et al. 2002).

### Phylogenetic and statistical analyses

For all loci, genotypic and allelic frequencies were calculated using the gene-counting method (Li 1976). To ascertain the phylogenetic relationships between populations, Maximum Likelihood (ML) analysis based on the allelic frequency distributions of the loci were generated using the PHYLIP 3.52c program (Felsenstein 1993). Bootstrap consensus phylogenies (1,000 replications) were generated by the SEQBOOT and CONTML options programs of PHYLIP. The CONTML and CONSENSE programs determined the best-fit tree.

A Principal Component (PC) analysis was performed using the numerical taxonomy and multivariate analysis system (NTSYS) PC program to summarize genetic relationships among the populations (Rohlf 2002). Centroid analysis was conducted to examine the relative gene flow experienced by populations and/or effective population size (Harpending and Ward 1982). The centroid model assumes an island model of population structure and expects a linear relationship between heterozygosity and genetic distance from the centroid. The centroid is defined as the overall mean allelic frequency of the populations. The theory is particularly useful in detecting and analyzing outliers. If a population is receiving gene flow from elsewhere at a higher-than-average rate, then the heterozygosity would be higher than expected (Sewerin et al. 2002). Those populations would plot above a linear regression line. Conversely, if the population has remained genetically isolated, the heterozygosity would be lower than expected due to less-than-average gene flow and would segregate below the linear regression line (Sewerin et al. 2002). Alternatively, populations plotting above or below the regression line may be indicative of higher- or lower-than-average, respectively, effective population size. The BIOSYS II program was used to generate expected heterozygosities and to detect any deviations from Hardy–Weinberg equilibrium expectations using the chi-square deviation

test. Power of discrimination (PD) values were also determined for all six loci. G-tests were performed with 2×2 contingency tables to ascertain statistical significance and determine whether the populations are homogeneous with each other (Lewontin and Felsenstein 1965).

## Results

### Genetic parameters of Atayal, Bali, Java and Madagascar populations

Table 2 displays the allelic and genotypic frequencies for the LDLR, GYPA, HBGG, D7S8, and GC loci for all four populations. Allelic and genotypic frequencies for the HLA-DQ $\alpha$ 1 locus are shown in Table 3. The expected and observed heterozygosities for all six loci in the four populations are presented in Table 4. The observed heterozygosities range from 10% (HBGG) to 83% (HLA-DQ $\alpha$ 1), both exhibited within the Atayal population. Except for one instance, all loci in the four populations were found in Hardy–Weinberg equilibrium (Table 5). The one exception was GYPA in the population from Java, which following a Bonferroni correction (0.008), was also in Hardy–Weinberg equilibrium. PD values are presented in Table 6. The Atayal population exhibits the largest range in PD, with 0.176 in the HBGG locus and 0.913 in the HLA-DQA1 locus. The six loci G-test indicated that all differences are statistically significant ( $P < 0.0001$ ) for all pair-wise comparisons involving the four populations reported in the present study as well as the eight reference groups.

### Phylogenetic analyses

Figure 2 displays the ML phylogeny and bootstrap values. In the resulting dendrogram, four clusters were observed: (1) Zimbabwe, African American, and Madagascar; (2) Caucasian; (3) Ami and Native American; and (4) Atayal and Han Chinese. The Atayal group segregates with the Han Chinese whereas the Ami tribe clusters with the Native Americans. The Madagascar

**Table 2** Genotypic and allelic frequencies for loci LDLR, GYPA, HBGG, D7S8, and GC

Locus	LDLR	GYPA	HBGG	D7S8	GC
Atayal ( <i>N</i> =40)					
Genotype					
AA	0.000	0.450	0.000	0.400	0.025
AB	0.175	0.375	0.100	0.450	0.375
BB	0.825	0.175	0.900	0.150	0.325
BC	NA	NA	0.000	NA	0.125
CC	NA	NA	0.000	NA	0.075
AC	NA	NA	0.000	NA	0.075
Allele					
A	0.087	0.637	0.050	0.625	0.250
B	0.913	0.363	0.950	0.375	0.575
C	NA	NA	0.000	NA	0.175
Java ( <i>N</i> =24)					
Genotype					
AA	0.083	0.208	0.042	0.208	0.000
AB	0.458	0.708	0.333	0.500	0.208
BB	0.459	0.084	0.625	0.292	0.125
BC	NA	NA	0.000	NA	0.167
CC	NA	NA	0.000	NA	0.208
AC	NA	NA	0.000	NA	0.292
Allele					
A	0.313	0.563	0.208	0.458	0.250
B	0.688	0.438	0.792	0.542	0.313
C	NA	NA	0.000	NA	0.438
Bali ( <i>N</i> =34)					
Genotype					
AA	0.147	0.441	0.029	0.235	0.000
AB	0.353	0.471	0.324	0.559	0.000
BB	0.500	0.088	0.647	0.206	0.412
BC	NA	NA	0.000	NA	0.412
CC	NA	NA	0.000	NA	0.147
AC	NA	NA	0.000	NA	0.029
Allele					
A	0.309	0.676	0.191	0.515	0.015
B	0.691	0.324	0.809	0.485	0.618
C	NA	NA	0.000	NA	0.376
Madagascar ( <i>N</i> =69)					
Genotype					
AA	0.015	0.420	0.159	0.377	0.000
AB	0.420	0.348	0.159	0.478	0.261
BB	0.565	0.232	0.188	0.145	0.536
BC	NA	NA	0.145	NA	0.145
CC	NA	NA	0.204	NA	0.029
AC	NA	NA	0.145	NA	0.029
Allele					
A	0.225	0.594	0.333	0.616	0.145
B	0.775	0.406	0.326	0.384	0.739
C	NA	NA	0.341	NA	0.116

NA not applicable because allele or genotypes do not exist at specific loci, *N* number of individuals studied

population segregates intermediate between the east-African and east-Asian groups. All bootstrap values, except for one, were at or above 50%.

Figure 3 displays the PC analysis. PC 1 (the *x*-axis) and PC 2 (the *y*-axis) represents 37% and 27% of the total variability, respectively. The Caucasian populations segregate in the right upper region of the map. PC

**Table 3** HLA-DQ $\alpha$ 1 genotypic and allelic frequencies

Genotype	Frequency <sup>a</sup>	Allele	Frequency <sup>a</sup>
Atayal ( <i>N</i> =40)			
1.1, 1.1	0.075	1.1	0.312
1.1, 1.2	0.075	1.2	0.112
1.1, 1.3	0.050	1.3	0.163
1.1, 3	0.150	3	0.175
1.1, 4.1	0.100	4.1	0.163
1.1, 4.2/4.3	0.075	4.2/4.3	0.075
1.2, 1.3	0.050		
1.2, 3	0.050		
1.2, 4.1	0.050		
1.3, 1.3	0.025		
1.3, 3	0.025		
1.3, 4.1	0.100		
1.3, 4.2/4.3	0.025		
3, 3	0.050		
3, 4.1	0.050		
3, 4.2/4.3	0.025		
4.1, 4.2/4.3	0.025		
Bali ( <i>N</i> =34)			
1.1, 1.1	0.147	1.1	0.397
1.1, 1.2	0.207	1.2	0.250
1.1, 2	0.029	1.3	0.015
1.1, 4.1	0.059	2	0.029
1.1, 4.2/4.3	0.207	3	0.015
1.2, 1.2	0.029	4.1	0.044
1.2, 1.3	0.029	4.2/4.3	0.250
1.2, 3	0.029		
1.2, 4.1	0.029		
1.2, 4.2/4.3	0.147		
2, 4.2/4.3	0.029		
4.2/4.3, 4.2/4.3	0.059		
Java ( <i>N</i> =24)			
1.1, 1.2	0.042	1.1	0.167
1.1, 1.3	0.083	1.2	0.104
1.1, 3	0.042	1.3	0.083
1.1, 4.2/4.3	0.166	2	0.042
1.2, 3	0.042	3	0.083
1.2, 4.2/4.3	0.124	4.1	0.021
1.3, 4.2/4.3	0.083	4.2/4.3	0.500
2, 3	0.042		
2, 4.2/4.3	0.042		
3, 4.2/4.3	0.042		
4.1, 4.2/4.3	0.042		
4.2/4.3, 4.2/4.3	0.250		
Madagascar ( <i>N</i> =69)			
1.1, 1.2	0.072	1.1	0.152
1.1, 3	0.044	1.2	0.181
1.1, 4.1	0.101	1.3	0.065
1.1, 4.2/4.3	0.087	2	0.043
1.2, 1.2	0.029	3	0.102
1.2, 1.3	0.044	4.1	0.196
1.2, 2	0.014	4.2/4.3	0.261
1.2, 3	0.072		
1.2, 4.2/4.3	0.101		
1.3, 2	0.014		
1.3, 4.1	0.044		
1.3, 4.2/4.3	0.029		
2, 2	0.014		
2, 4.2/4.3	0.029		
3, 4.1	0.044		
3, 4.2/4.3	0.044		
4.1, 4.1	0.059		
4.1, 4.2/4.3	0.087		
4.2/4.3, 4.2/4.3	0.072		

<sup>a</sup>If genotype or allele is not listed, then frequency is zero  
*N* number of individuals studied

**Table 4** Expected and observed heterozygosities for all six loci and all four populations

	LDLR	GYP A	HBGG	D7S8	GC	DQA1
<b>Java</b>						
Expected	0.439	0.503	0.337	0.507	0.662	0.689
Observed	0.458	0.708	0.333	0.500	0.667	0.708
<b>Atayal</b>						
Expected	0.162	0.468	0.096	0.475	0.584	0.788
Observed	0.175	0.375	0.100	0.450	0.575	0.825
<b>Bali</b>						
Expected	0.444	0.444	0.314	0.507	0.490	0.702
Observed	0.353	0.471	0.324	0.559	0.441	0.765
<b>Madagascar</b>						
Expected	0.351	0.494	0.671	0.477	0.422	0.724
Observed	0.420	0.362	0.551	0.478	0.435	0.739

**Table 5** Hardy–Weinberg equilibrium tests for all six loci

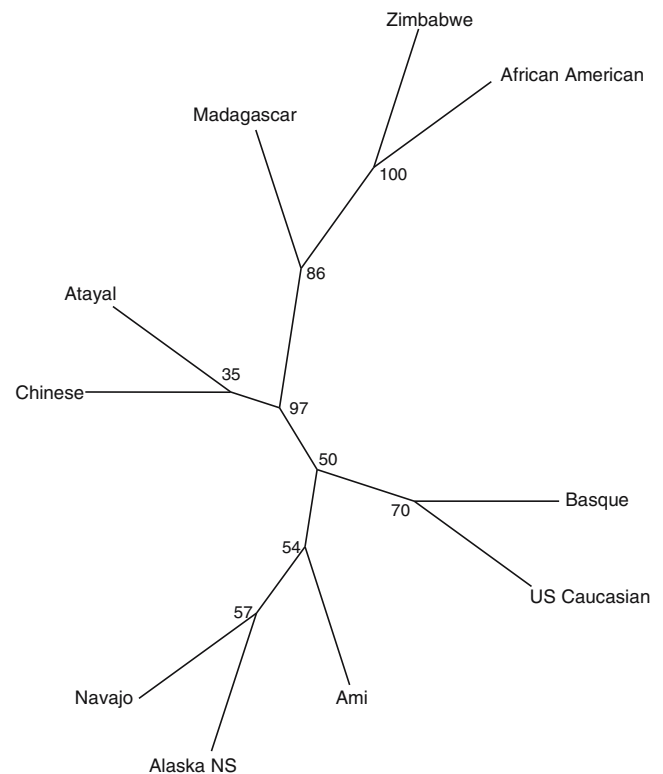
	LDLR	GYP A	HBGG	D7S8	GC	DQA1
<b>Chi-square</b>						
Java	0.823	0.040	0.957	0.944	0.298	0.763
Atayal	0.577	0.202	0.774	0.739	0.109	0.853
Bali	0.222	0.724	0.853	0.545	0.568	0.933
Madagascar	0.096	0.111	0.176	0.976	0.279	0.107

**Table 6** Power of discrimination values for all six loci

Population	LDLR	GYP A	HBGG	D7S8	GC	DQA1
Java	0.581	0.619	0.496	0.623	0.802	0.877
Atayal	0.280	0.604	0.176	0.608	0.755	0.913
Bali	0.581	0.588	0.475	0.625	0.618	0.868
Madagascar	0.515	0.616	0.363	0.610	0.625	0.946

1 separates the African American, Zimbabwe, and Madagascar groups from the rest of the populations. The Native American groups cluster together in the extreme right on the plot and include Na-Dene and Eskaleut populations. The Han Chinese, Bali, and Atayal plot loosely together at the lower center of the map with the populations from Java at the periphery. The Ami represent an outlier segregating by itself in the lower right-hand quadrant. Within Fig. 3, the populations of Bali and Java segregate apart from each other with Bali plotting closer to the Atayal and Ami. Java, on the other hand, plots at the fringes of the Caucasian group. Interestingly, the Madagascar population plots midway between the African groups and the loose cluster of Atayal, Bali, and Han Chinese.

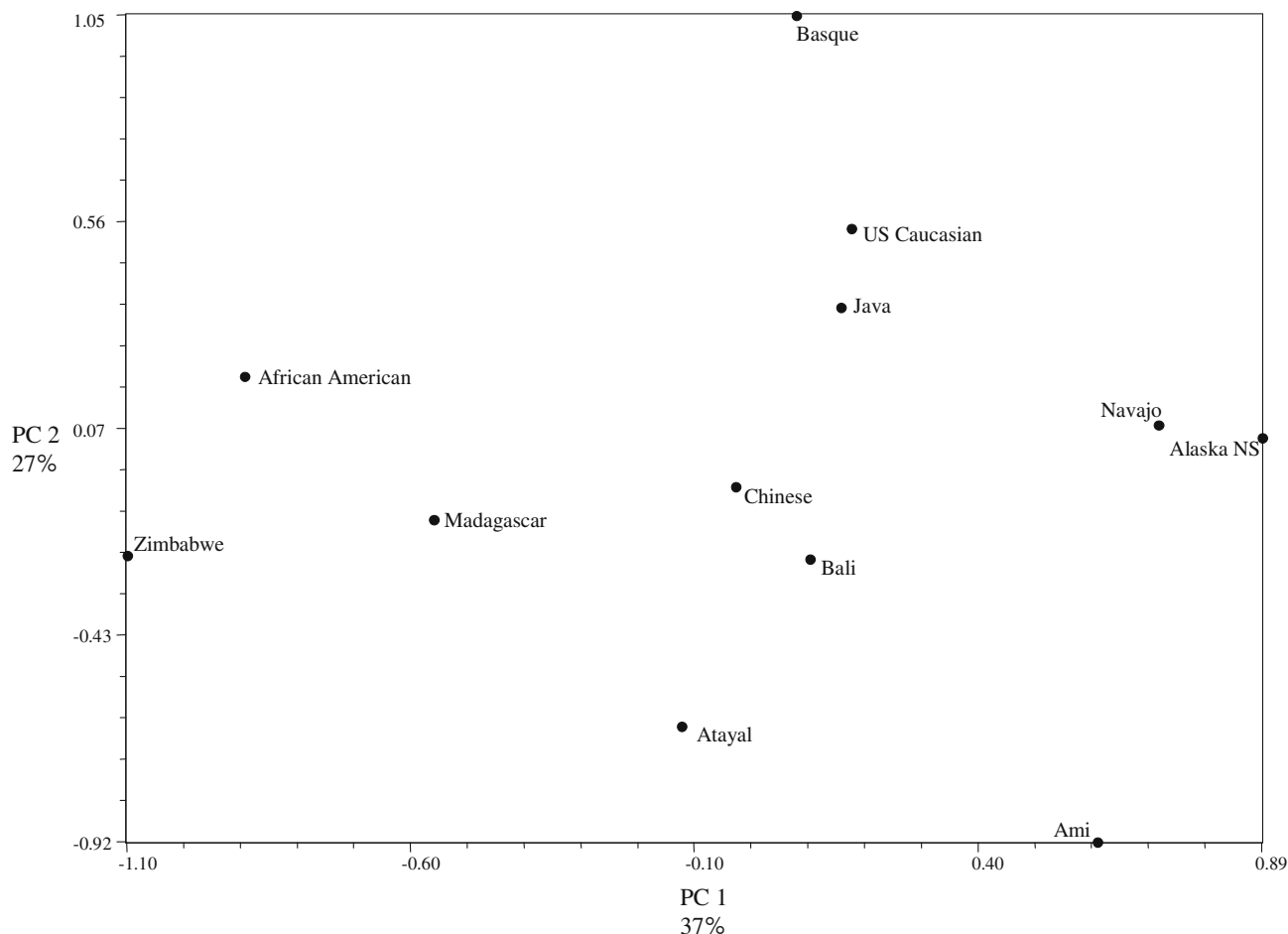
Figure 4 depicts the centroid analysis of the Atayal, Bali, Java, and Madagascar groups along with the eight other reference populations. The Caucasian and African populations plot above the linear regression whereas the Native American populations are located beneath the regression line. As outliers, the Ami and Atayal are the most distant groups under the regression line. The Han Chinese, Java, and Zimbabwe map nearly on the regression line while Madagascar plots above.

**Fig. 2** Maximum-likelihood (ML) tree illustrating human phylogenetic relationships. Tree was generated using allelic frequency data of ten populations, using PHYLIP 3.52c. NS North Slope Borough population

## Discussion

In this study, six polymorphic loci containing point mutations were examined in four geographically targeted populations from the Austronesian language family. Because the island of Formosa has been hypothesized as the potential Austronesian homeland, studies of the country's aborigines are of paramount importance (Bellwood 1991). It is likely that the diversification of the Austronesian language family occurred in Taiwan (Diamond 2000) and possibly only one of the groups is responsible for the successive colonization of other islands during the diaspora. This idea emphasizes the importance of ascertaining the relationships among tribes as well as between the tribes and other Austronesian groups and worldwide populations (Tajima et al. 2003).

Figure 2 depicts a ML tree with four clusters. One clade contains populations from Africa (Zimbabwe and Madagascar) and of African descent (African American). In another cluster, the Native Americans segregate with the Ami while the Atayal group cluster with the mainland Han Chinese in a third clade. Caucasians are found together in a fourth cluster. It is not surprising that the Zimbabwe, African American, and Madagascar populations are found together within a cluster. The island of Madagascar lies just off the east coast of Africa

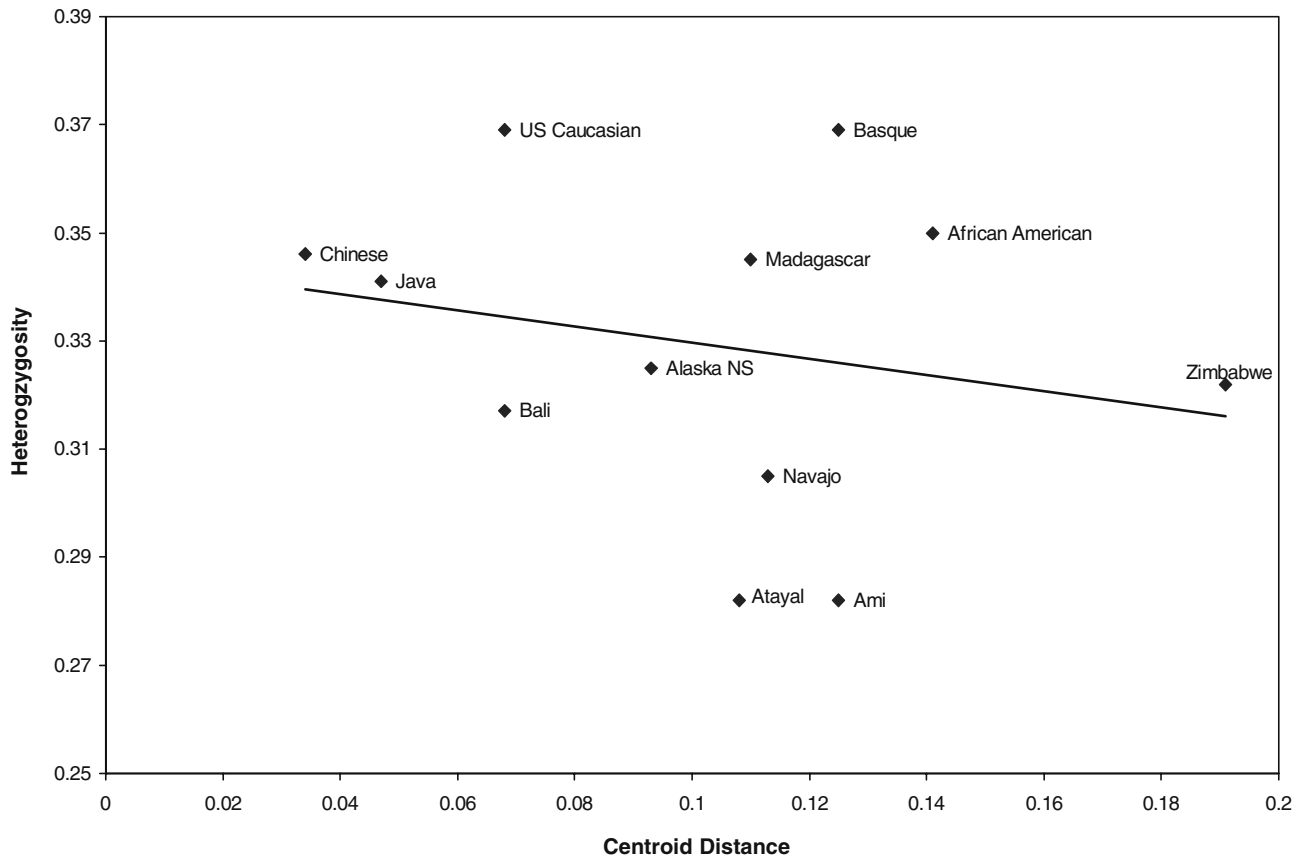


**Fig. 3** Principal component (PC) map. Worldwide populations from diverse geographical areas were used in the analysis. PC1 and PC2 represent the first and second PC values, respectively. PC1 and PC2 account for 37% and 27% of the variability, respectively

in close proximity (about 500 miles) from Zimbabwe. Yet, the relationship exhibited by the groups from Africa and the East Asian/Native American populations underscore two important issues. First is the separation of the Madagascar group away from the Zimbabwe and the African Americans in the African clade and its proximity to the Atayal and Han Chinese. It is significant that the Zimbabwe, an East African population, is genetically closer to African Americans, an admixed group with a major West African genetic contribution than to the sample from Madagascar just off the coast of east Africa. It is possible that the Madagascar population's genetic affinity to the Atayal and the continental East Asians (Hans) may be the result of the Austronesian migration into Madagascar approximately 3,200 years ago (Ruhlen 1994). The Ami aborigines, on the other hand, segregate more distant from the Madagascar group in a cluster with two Native American groups. These results point to the Atayal and not the Ami aborigines as a stronger candidate for the Formosan source population responsible for the westward Austronesian dispersal. The Navajo and the Alaskan Eskimos, which group close to each other, are recent

arrivals (approximately within the last 2,000 years) to the New World representing the distinct language groups Na-Dene and Eskaleut, respectively. The segregation of these African and East Asian/Native American groups in the ML dendrogram mirror the genetic affinities reflected in the PC plot discussed below.

The segregation of the Atayal tribe with the Han Chinese supports the connection between the Austronesian language family and Southeast Asia. Previous studies, using 13 classical loci, indicate that the Toroko, a branch of the Atayal, have higher affinities to the general populations from the Philippines and Thailand than to the groups from southern China and Vietnam (Chen et al. 1985). It is interesting to note that the two Taiwanese aboriginal groups segregate into different clades. In phylogenetic studies using mtDNA haplogroups, the Ami and the Atayal also segregate into distinct clades (Tajima et al. 2003). The Ami and the Atayal have been living for thousands of years in close proximity in adjacent territories (see Fig. 1) on the island of Formosa. There are two possible explanations for this. One is that their differences may reflect separate origins from diverse populations that arrived in succes-



**Fig. 4** Centroid gene flow analysis plot. The heterozygosity of each population (y-axis) is plotted versus the centroid (overall mean allelic frequency of population) (x-axis). Plot includes eight worldwide reference populations

sive waves to inhabit Formosa followed by isolation, and the second is that these variations may simply be due to subsequent geographical partitioning and genetic differentiation of tribes from a common origin. The known strong cultural and linguistic differences could have generated barriers capable of preventing gene flow between the two groups preserving their uniqueness.

The PC plot, a two-dimensional illustration of allelic variability between the populations, is depicted in Fig. 3. The first and second PCs account for 64% of the total variation. Within the plot, four groups are evident: (1) Caucasians; (2) a scattered set including the Atayal, Bali and Han Chinese; (3) Native Americans; and (4) Africans and populations of African descent. As with the ML analysis, the population from Madagascar clusters away from the African Americans and the Zimbabwe group, and toward the Atayal, Bali and Han Chinese. Also as observed in the ML study, the intermediate position of the Madagascar group between the African groups and the East Asian populations may reflect the Austronesian contribution to the island population. The large genetic distance between the Ami and the Atayal belies the fact that these tribal groups are close neighbors and corroborate the ML data. Again, independent source populations in mainland Asia and/or extreme genetic drift and/or genetic

isolation may be at least partially responsible for their phonetic differences. It is surprising that the Bali and Java groups plot distantly from each other in spite of their geographic proximity. It is obvious that these populations are genetically unique and distinct even though they represent adjacent islands only approximately 10 miles from each other with populations that belong to the same Austronesian language family. Compared with Bali, the island of Java is approximately 23 times larger and more culturally diverse. Subsequent to the Austronesian diaspora, Hindus, Buddhists, and Muslims have invaded Java. Today, Bali is predominately Hindu. Greater genetic flow from different groups and/or effective population size in Java may be an explanation for the observed differences between these two islands. The position of the Java population above and the Bali group below the linear regression in the centroid analysis (see next paragraph) corroborate this scenario. In addition, these populations may have undergone extreme drift and/or genetic isolation subsequent to migration into the two islands.

In the centroid analysis (Fig. 4), populations that plot above the regression line are expected to possess larger effective size and/or experience more gene flow than those below the regression line. The African American group plots above the regression line which supports the



fact that this population is highly admixed. Madagascar may also have experienced high levels of gene flow from diverse sources. It is likely that the Austronesian groups that made it to Madagascar admixed with the populations originally from mainland Africa. The position of the Zimbabwe group slightly above the linear regression most likely is the result of the greater diversity and heterozygosity of sub-Saharan African populations (Cavalli-Sforza et al. 1994). Within the plot, all Caucasian populations fall above the regression line. The Han Chinese and Java lie above and nearly on the line. The Native Americans plot below the regression line with the Alaskan population closest to the line. The Bali population is positioned below the linear regression line as well. The location of the Java group above the linear regression and the Balinese below is in agreement with their relative effective population size and corroborate the contention that Java has experienced greater gene flow from different source populations. The Ami and Atayal are outliers located furthest below the line proximal to each other. This data, together with the ML and PC results, indicates that while these two groups are genetically different, they both experienced little gene flow and have remained separated. Population size may have also had an influence as both populations are small and limited in diversity and heterozygosity. Data from several serum protein loci corroborate the low genetic diversity among Taiwanese aborigines (Yuasa et al. 2001). This genetic data supports the idea that the aboriginal tribes of Taiwan have remained isolated from the Han Chinese population and from each other. The fact that these two tribes have different material cultures and social organizations substantiates this conclusion (Chai 1967).

As demonstrated by the ML and PC analyses, it is significant that the Atayal and the Ami aboriginal groups are genetically unique. These results are possibly an indication that the Ami and Atayal may have had different ancestral source populations originating in mainland Asia and subsequent cultural isolation. The centroid analysis argues for a small effective population size along with a limited amount of gene flow for both the Ami and the Atayal. Genetic drift due to founder effect and/or isolation may also be responsible for the differences between the two aboriginal groups. Also, based on the dendrogram and the PC analysis, the Madagascar population exhibits an intermediate genetic relationship between the African/African descent groups and the Atayal/Han Chinese cluster which may be due to an Austronesian influence on the island some 3,200 years ago. It is significant that Madagascar segregates in the direction of the East Asians in the PC plot and away from Zimbabwe which is geographically much closer by one order of magnitude. Madagascar displayed genetic affinity with the Atayal while maintaining a larger genetic distance from the Ami. This data supports previous archeological and linguistic data and supports a westward expansion of Austronesians originating from or near Formosa reaching the island of Madagascar.

**Acknowledgements** We would like to acknowledge Maria Cristina Terreros, Erica M. Shepard, and Diane J. Rowold for their assistance in the preparation of this article.

## References

- Antunez de Mayolo G, Antunez de Mayolo A, Antunez de Mayolo P, Papiha SS, Hammer M, Yunis JJ, Yunis EJ, Damodaran C, Martinez de Pancorbo M, Caeiro JL, Puzyrev VP, Herrera RJ (2002) Phylogenetics of worldwide human populations as determined by polymorphic Alu insertions. *Electrophoresis* 23:3346–3356
- Bellwood P (1991) The Austronesian dispersal and the origin of the languages. *Sci Am* 265:88–93
- Brown RJ, Rowold D, Tahir M, Barna C, Duncan G, Herrera RJ (2000) Distribution of the HLA-DQA1 and polymarker alleles in the Basque population of Spain. *Forensic Sci Int* 108:145–151
- Budowle B, Lindsey JA, Decon JA, Koons BW, Giusti AM, Comay CT (1995) Variation and population studies of the loci LDLR, GYPA, HBGG, D7S8, and Gc (PM loci) and HLA-DQ $\alpha$  using a multiplex amplification and typing procedure. *J Forensic Sci* 40:45–54
- Cann RL (2001) Genetic clues to dispersal in human populations: retracing the past from the present. *Science* 291:1742–1748
- Capelli C, Wilson JF, Richards M, Stumpf MPH, Gratrix F, Oppenheimer S, Underhill P, Pascali VL, Ko TM, Goldstein DB (2001) A predominantly indigenous paternal heritage for the Austronesian-speaking peoples of insular Southeast Asia and Oceania. *Am J Hum Genet* 68:432–443
- Cavalli-Sforza LL, Menozzi P, Piazza A (1994) The history and geography of human genes. Princeton University Press, Princeton
- Chai CK (1967) Taiwan aborigines: a genetic study of tribal variations. Harvard University Press, Cambridge
- 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
- Chu HL (1997) An introduction to the indigenous culture of Taiwan. Charity Printing Industrial Co., Taipei
- Diamond JM (1988) Express train to Polynesia. *Nature* 336:307–308
- Diamond JM (2000) Taiwan's gift to the world. *Nature* 403:709–710
- Felsenstein J (1993) PHYLIP (Phylogeny Inference Package). Seattle, WA
- Gray RD, Jordan FM (2000) Language trees support the express train sequence of Austronesian expansion. *Nature* 405:1052–1055
- Harpending HC, Ward RH (1982) Chemical systematics and human population. In: Nitecki M (ed) Biological aspects of evolutionary biology. University of Chicago Press, Chicago
- 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
- Huang MC (1964) Studies in the distribution of Rh blood types among various racial tribes in Formosa. *Nippon Koigaku Zasshi [Jpn J Leg Med]* 18:135–142
- Huang NE, Budowle B (1995) Chinese population data on the PCR-based loci HLA-DQ alpha, low density lipoprotein receptor, glycophorin A, hemoglobin  $\alpha$ ,  $\gamma$ , D7S8 and group specific component. *Hum Hered* 45:34–40
- Ikemoto S, Ming CT, Haruyama N, Furumata T (1942) Blood group frequencies in the Ami tribe (Formosa). *Proc Jpn Acad*:173–177
- Jin F, Saitou N, Ishida T, Sun CS, Pan IH, Omoto K, Horai S (1999) Population studies on nine aboriginal ethnic groups of Taiwan. I. Red cell enzyme systems. *Anthropol Sci* 107:229–246

- Kao TY (1958) Taiwan historical events. Cheng Chung Publishing, Taipei
- Kirch PV (1997) The Lapita peoples: ancestors of the oceanic world. Blackwell, Massachusetts
- Kirch PV (2000) On the road of the winds: an archeological history of the Pacific Islands before European contact. University of California Press, California
- Lewontin RC, Felsenstein J (1965) The robustness of homogeneity test in  $2 \times N$  tables. *Biometrics* 21:19–33
- Li CC (1976) First course in population genetics. Boxwood Press, Pacific Grove, CA
- Luis JR, Terreros MC, Martinez L, Rojas D, Herrera RJ (2003) Two problematic human polymorphic Alu insertions. *Electrophoresis* 24:2290–2294
- 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
- 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
- Merriwether DA, Friedlaender JS, Mediavilla J, Mgone C, Gentz F, Ferrell RE (1999) Mitochondrial DNA variation is an indicator of Austronesian influence in Island Melanesia. *Am J Phys Anthropol* 110:243–270
- Nakajima H, Ohkura K (1971) The distribution of several serological and biological traits in East Asia, vol 3, the distribution of gamma globulin (Gm [1], Gm [2], Gm [5], and Ivn [1]) and Gc groups in Taiwan and Ryukyu. *Hum Hered* 21:362–370
- Ohashi J, Naka I, Ohtsuka R, Inaoka T, Ataka Y, Nakazama M, Tokunaga K, Matsumura Y (2004) Molecular polymorphism of ABO blood group gene in Austronesian and non-Austronesian populations in Oceania. *Tissue Antigens* 63:355–361
- 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
- Richards M, Oppenheimer S, Skyes B (1998) MtDNA suggests Polynesian origins in eastern Indonesia. *Am J Hum Genet* 63:1234–1236
- Rohlf FJ (2002) NTSys pc (Numerical Taxonomy System). Exeter Publishing, Setauket
- Ruhlen M (1994) The origin of language: tracing the origin of the mother tongue. Wiley, New York, pp 177–180
- Sewerin B, Cuza FJ, Szmulewicz MN, Rowald DJ, Bertrand-Garcia RL, Herrera RJ (2002) On the genetic uniqueness of the Ami aborigines of Formosa. *Am J Phys Anthropol* 119:240–248
- Su B, Jin L, Underhill P, Martinson J, Saha N, McGarvey ST, Shriver MD, Chu J, Oefner P, Chakraborty R, Deika R (2000) Polynesian origins: insights from the Y chromosome. *Proc Natl Acad Sci USA* 97:8225–8228
- Tajima A, Sun CS, Pan IH, Saitou N, Horai S (2003) Mitochondrial DNA polymorphisms in nine aboriginal groups of Taiwan: implications for the population history of aboriginal Taiwanese. *Hum Genet* 113:24–33
- Walkinshaw M, Strickland L, Hamilton H, Denning K, Gayley T (1996) DNA profiling in two Alaskan native populations using HLA-DQA1, PM, and D1S80 loci. *J Forensic Sci* 41:478–474
- Wolfarth R, Nhari LT, Budowle B, Kanoyangwa SB, Masuka E (2000) Polymarker, HLA-DQA1 and D1S80 allele data in a Zimbabwean Black sample population. *Int J Legal Med* 113:300–301
- 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 synthesis. *Anthropol Sci* 109:257–273

Copyright of Journal of Human Genetics is the property of Springer Science & Business Media B.V.. The copyright in an individual article may be maintained by the author in certain cases. Content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.