

Ascertaining the Role of Taiwan as a Source for the Austronesian Expansion

Sheyla Mirabal,^{1,2} Alicia M. Cadenas,² Ralph Garcia-Bertrand,³ and Rene J. Herrera^{2*}

¹*Department of Molecular and Human Genetics, College of Medicine, Florida International University, Miami, FL 33199*

²*Department of Molecular and Cell Biology, The Rockefeller University, New York, NY 10065*

³*Department of Biological Sciences, Colorado College, Colorado Springs, CO*

KEY WORDS YSTRs; Austronesian expansion; Taiwanese aborigines

ABSTRACT Taiwanese aborigines have been deemed the ancestors of Austronesian speakers which are currently distributed throughout two-thirds of the globe. As such, understanding their genetic distribution and diversity as well as their relationship to mainland Asian groups is important to consolidating the numerous models that have been proposed to explain the dispersal of Austronesian speaking peoples into Oceania. To better understand the role played by the aboriginal Taiwanese in this diaspora, we have analyzed a total of 451 individuals belonging to nine of the tribes currently residing in Taiwan, namely the Ami, Atayal, Bunun, Paiwan, Puyuma, Rukai, Saisiyat, Tsou, and the Yami from Orchid Island off the coast of Taiwan across 15 autosomal short tandem repeat loci. In addition, we have compared the genetic profiles of these tribes to populations from mainland China as well as to collections at key points

throughout the Austronesian domain. While our results suggest that Daic populations from Southern China are the likely forefathers of the Taiwanese aborigines, populations within Taiwan show a greater genetic impact on groups at the extremes of the current domain than populations from Indonesia, Mainland, or Southeast Asia lending support to the “Out of Taiwan” hypothesis. We have also observed that specific Taiwanese aboriginal groups (Paiwan, Puyuma, and Saisiyat), and not all tribal populations, have highly influenced genetic distributions of Austronesian populations in the Pacific and Madagascar suggesting either an asymmetric migration out of Taiwan or the loss of certain genetic signatures in some of the Taiwanese tribes due to endogamy, isolation, and/or drift. *Am J Phys Anthropol* 150:551–564, 2013. © 2013 Wiley Periodicals, Inc.

Language of the Austronesian family are spoken in a wide geographical range bound by Madagascar to the west, Easter Island to the East, New Zealand to the south and Taiwan to the north (Gray and Jordan, 2000). Though several theories have been proposed to explain the origin, routes of dispersal, timing of migratory pauses and pulses, as well as the current genetic relationships of Austronesian groups, no single conjecture has been able to bridge all of the data (linguistic, molecular and archaeological) available. Since nine out of the ten currently spoken Austronesian linguistic subgroups can only be found in Taiwan (Blust, 1999), the island lies at the root of these interrogatives.

Taiwan's current population is largely composed of Han Chinese (of the Min and Hakka subgroups) who migrated to the island from southeast China in the last half millennium (Chiung, 2001). It is the Taiwanese aborigines, however, who comprise about 1.5% of the populace and speak Austronesian languages, which are considered indigenous to the country (Trejaut et al., 2005). Though these tribal groups are believed to be descendants of mainland Asians, the timing of their arrival in Formosa is still debatable and it is unclear whether they represent a single or several Paleolithic or Neolithic waves of migration (Tsang, 2002; Sagart, 2004; Sagart, 2005).

Archaeological evidence suggests that humans were present in Taiwan about 15,000 years before present (YBP) (Tsang, 2002), but most remains are more recent dating to Neolithic times and are largely associated with the arrival of pastoral agriculturalists from mainland

China around 8,000 YBP (Ruhlen, 1994; Rolett and Tsang, 2001; Diamond and Bellwood, 2003). Several studies indicate that southern Chinese Daic populations are the likely ancestors of Formosan tribal groups (Li et al., 2008; Wu et al., 2009) and it has been claimed that the Daic domain within South China, and not Taiwan, is the homeland of all Austronesian groups (Li et al., 2008). The Daic are considered to be the original inhabitants of China's southeast coast and their origins can be traced to 20,000 YBP (Li et al., 2007). Today, Daic people are second only to the Han in numbers within China and have a strong presence in Thailand, Laos, Vietnam, Myanmar and India (Grimes, 2002).

It is generally accepted that the Austronesian diaspora began around 6,000 years ago (Zhang, 1987; Diamond, 1988; Bellwood et al., 1995). Of the several theories proposed to explain the series of events that ensued, three have received widespread genetic, linguistic and

Additional Supporting Information may be found in the online version of this article.

*Correspondence to: R. J. Herrera, College of Medicine, Florida International University, University Park, OE 304, Miami, FL 33199, USA. E-mail: ReneJustoHerrera@gmail.com

Received 10 July 2012; accepted 14 December 2012

DOI: 10.1002/ajpa.22226
Published online 26 February 2013 in Wiley Online Library (wileyonlinelibrary.com).

archeological support; namely the “express train,” “entangled bank,” and “slow boat” hypotheses. The “express train” model (also coined as “out of Taiwan”) postulates that Austronesians originated in Taiwan and traversed rapidly through Micronesia and Melanesia, leading to minimal admixture with the pre-existing substrata of the area, before settling throughout Oceania (Melton et al., 1995; Bellwood, 1997; Lum, 1998; Green, 1999; Hagelberg et al., 1999; Diamond, 2000; Gray and Jordan, 2000; Trejaut et al., 2005; Gray et al., 2009; Tofanelli et al., 2009; Greenhill et al., 2010). On the other hand, the “entangled bank” hypothesis proposes that a migration from Mainland Asia to the Pacific during the Mid-Holocene, along with continued and extensive integration with natives throughout the routes traversed, is responsible for the current genetic characteristics of Austronesian groups (Terrel et al., 1997; Capelli et al., 2001; Oppenheimer and Richards, 2001; Hurler et al., 2002). The “slow boat” model, which combines some of the main elements of both aforementioned ideas, posits that though Austronesian migrants are of Asian descent (most likely Taiwanese), they traveled slowly through Island Southeast Asia assimilating some of the pre-existing genetic substrata and eventually reaching Near and Far Oceania (Kayser et al., 2000, 2003, 2006, 2008).

The distributions of mtDNA haplogroup B4a1a1 (the so dubbed “Polynesian motif” and its ancestral lineage B4a1a provided the first direct genetic evidence tying Polynesians to Taiwanese aborigines (Trejaut et al., 2005). A phylogeny based on these mtDNA subtypes directly mirrors linguistic relationships (Bellwood, 1990; Donohue and Denham, 2010) and follows a route originating in Taiwan, where it is most diverse, traveling south toward the Philippines and Indonesia (Melton et al., 1995; Hill et al., 2007; Tabbada et al., 2010; Razafindrazaka et al., 2010). However, subsequent studies have unearthed previously undiscovered genetic affinities between mtDNAs in the Bismarck Archipelago of Indonesia and Oceanian populations that may predate the Austronesian expansion by as much as 3,000 years obscuring previous conclusions (Soares et al., 2011).

Y-chromosomal studies have also proven controversial and though some early analyses alluded to ties between Asian populations and Polynesian groups (Su et al., 2000), most reports indicate close genetic ties between Melanesia and Polynesia and no clear connections to Taiwan (Lum, 1998; Kayser et al., 2000, 2003, 2006, 2008). Recently, however, Karafet et al. (2010) demonstrated that a subgroup of haplogroup O3, specifically O3a2, is widely distributed throughout Island Southeast Asians, Indonesians and Polynesians but absent from mainland Asian populations. Subsequently, Mirabal et al. (2012) reported that a close genetic relationship exists between the Ami (one of the Taiwanese aboriginal tribes) and Polynesian populations by finding that all O3a2 (P201) chromosomes in the Ami and Polynesians also belong to the rare downstream undifferentiated sub-haplogroup O3a2c* (P164), that had previously only been detected at low levels in some mainland East Asian populations (Yan et al., 2011). These findings established a direct genetic link between Taiwanese aborigines and Polynesian groups previously undetected due to the lack of resolution of O3 derived Y-chromosomes afforded by previous studies (Su et al., 2000; Kayser et al., 2000, 2003, 2006, 2008).

In an attempt to consolidate the wealth of data obtained from uniparental loci, other authors have turned to auto-

somal markers. By using autosomal short tandem repeat (STR) loci, Kayser et al. (2008) and Regueiro et al. (2008) both proposed independently that East Asians and Melanesians are the source of 79% (Kayser) / 76% (Regueiro) and 21% (Kayser) / 24% (Regueiro) of the current Polynesian gene pool, respectively. By using genome wide scans including both microsatellites and insertions/deletions, Fried Laender et al. (2008) resolved close genetic ties among Austronesian speakers and concluded that signals shared by populations of Austronesian ancestry were only present at low frequencies in Austronesian speaking Melanesian groups leading to the suggestion that movement through Melanesia was rapid. Subsequently, high density autosomal single nucleotide polymorphisms (SNPs) have also facilitated admixture time estimates between Asian and Papuan populations in Eastern Indonesia which are highly congruent with linguistic time estimates corresponding to the time of the Austronesian expansion (Xu et al., 2012). Other autosomal SNPs investigations have found only minute contributions from Austronesian groups throughout Southeast Asia and instead suggest that mainland Asians have shaped these regions (Jinam et al., 2012).

In the current study, we have analyzed a total of 451 individuals from Taiwan (95 Ami, 45 Atayal, 47 Bunun, 40 Paiwan, 40 Puyuma, 33 Rukai, 39 Saisiyat, 42 Tsou and 70 Yami) across a set of 15 autosomal STR loci and compared them to previously published collections from mainland Asia, Island Southeast Asia, Indonesia, Australia, Polynesia and Madagascar (an Austronesian-speaking island approximately 200 km off the south east coast of Africa) in order to assess their current genetic relationships.

Autosomal STRs are selectively neutral, hypervariable markers particularly helpful when elucidating genetic relationships between closely related, recently separated or admixed populations given their large number of alleles, high heterozygosity, abundance and widespread distribution throughout the genome (Jorde et al., 1997; Rowold and Herrera, 2003; Perez-Miranda et al., 2005; Shepard et al., 2005; Shepard and Herrera, 2006; Ibarra-Rivera et al., 2008; Regueiro et al., 2008). They vary solely on the basis of mutation and drift, lending themselves to varied applications including detailed forensic analyses (Leibelt et al., 2003) and population genetics investigations since they provide the high resolution needed to investigate closely related or recently separated populations (Rowold and Herrera, 2003). In the present study, we aim to shed light onto the populations from mainland Asia that contributed to the genetic makeup of Taiwanese aborigines, to understand the differentiation among these Formosan tribal groups and their role in the dispersal of Austronesian speakers. Moreover, we hope to establish whether influences exerted throughout the Austronesian domain are homogenous among the Taiwanese tribes, or whether certain tribes have played more significant roles.

MATERIALS AND METHODS

Populations, sample collection, and DNA isolation

Peripheral blood samples or buccal swabs were collected from a total of 451 unrelated individuals native to nine of the aboriginal tribes present in Taiwan (95 Ami, 45 Atayal, 47 Bunun, 40 Paiwan, 40 Puyuma, 33 Rukai, 39 Saisiyat, 42 Tsou and 70 Yami). Genealogical

information was collected for a minimum of two generations to ascertain descent. Blood samples were extracted using the standard phenol-chloroform method (Novick et al., 1995; Antunez de Mayolo et al. 2002) while buccal swabs were processed utilizing the Gentra Buccal Cell Kit (Puregene, Gentra Systems, Minneapolis, MN) according to the manufacturers' specifications. Samples were stored as stock solutions in 10 mM Tris-EDTA at -80°C . All samples were procured from donors voluntarily while closely adhering to the ethical guidelines stipulated by Tsu Chi University, Hualien, Taiwan, Colorado College, Colorado Springs Colorado, USA and Florida International University, Miami, Florida, USA Institutional Review Boards. The project was revised and approved by all the institutions' ethical committees.

Reference populations

Thirty-five reference populations were employed for comparison across the 15 STR loci under study. The geographical locations, abbreviations used to define them throughout the article, number of individuals, references, as well as average heterozygosities are all provided in Table 1. Although not included in most statistical analyses herein, the Taiwanese samples analyzed by Lee et al. (2002) were compared to ours across 13 STR loci to assess collection-based differences (Fig. 2b).

DNA amplification and STR genotyping

Samples were typed utilizing the commercially available AmpF/STR Identifier kit which allows for analysis of 15 autosomal STR loci (D8S1179, D21S11, D7S820, CSF1PO, D3S1358, TH01, D13S317, D16S539, D2S1338, D19S433, vWA, TPOX, D18S51, D5S818, and FGA) (Applied Biosystems, 2001). Amplifications were performed in an Eppendorf Mastercycler gradient (Eppendorf, Hamburg, Germany) according to protocols established by the manufacturer (Applied Biosystems, 2001). Amplicons were separated via capillary electrophoresis in an ABI Prism 3130xl Genetic Analyzer following denaturation with formamide and addition of the ABI GeneScan 500 LIZ internal size standard. Genotyping was conducted by comparison to the allelic ladder and internal size standards with the GeneMapper software v3.2.

Statistical analyses

Allelic frequencies were calculated utilizing the GenePop v3.4 program (Raymond and Rousset, 1995). Departures from Hardy-Weinberg equilibrium were assessed by calculating observed and expected heterozygosities (H_o and H_e , respectively) and associated P -values with the aid of the Arlequin software package v3.5.1.2 (Levene, 1949; Guo and Thompson, 1992; Schneider et al., 2000). Significance was ascertained before and after application of the Bonferroni correction ($\alpha=0.05/15=0.0033$ for 15 loci). The PowerStats v1.2 Software (Jones, 1972; Brenner and Morris, 1989; Tereba, 1999) was employed to calculate parameters of population genetics interest including Matching Probability (MP), Power of Discrimination (PD), Polymorphic Information Content (PIC), Power of Exclusion (PE) and Typical Paternity Index (TPI). Overall heterozygote deficiencies as well as inbreeding coefficients (F_{IS}) on a locus by locus basis were calculated utilizing GenePop (Raymond and Rousset, 1995).

The DISPAN program (Ota, 1993) was used to tabulate the inter, intra and total population genetic

variance components (Gst, Hs and Ht, respectively). The populations were partitioned into groups as follows: (1) African (Angola, Equatorial Guinea, Hutu, Kenya, Madagascar, Mozambique, South Africa and Tutsi), (2) Chinese (Chao Shan, Henan, Huis, Jinan, Maonans, Miaos, Mulaos, Shaanxi and Yis), (3) Indonesian Archipelago (Bali, East Timor, Java and Surabaya), (4) Pacific populations (Australian Aborigines, New Zealand Asian, New Zealand Eastern Polynesia, New Zealand Western Polynesia, Samoa and Tonga), (5) Taiwanese aborigines (Ami, Atayal, Bunun, Paiwan, Puyuma, Rukai, Saisiyat, Tsou and Yami), and (6) All populations which represents a composite of all the aforementioned groups.

Multidimensional Scaling (MDS) plots were constructed with the aid of the ALSCAL tool within IBM's SPSS v20 statistical software package (SPSS for windows); two dimensional spatial modeling was performed based on chi square distances. Three MDS graphs were generated; the first was used to assess the general differences among Taiwanese aborigines and to compare between the collections genotyped in this study and those previously reported by Lee et al. (2002) given the discordant frequencies reported for Taiwanese aborigines across several molecular markers (Lin and Broadberry, 1998; Lin et al., 2000, 2005; Su et al., 2000; Capelli et al., 2001; Chu et al., 2001; Lee et al., 2002; Trejaut et al., 2005). The second and third MDS plots explored the relationships of Taiwanese aboriginal populations in the context of both their genetic connection to other East Asian populations and their role in the Austronesian expansion at the level of both 15 and 13 autosomal STR loci, respectively. The statistical significance of the MDS analyses was evaluated according to Sturrock and Rocha (2000). Population substructuring was explored with the aid of the *Structure* software v2.3.3 (Pritchard et al., 2000) available from <http://pritch.bsd.uchicago.edu/structure/html>. Clustering was conducted based on the 15 and 13 STR loci sets utilizing the admixture model that does not take into account population information in order to reduce bias at $k=2-20$. The k value exhibiting the highest degree of structure was calculated according to Evanno et al. (2005).

To further ascertain these relationships, admixture estimates were undertaken using the nonlinear regression tool within the program SPSS v20 (Long et al., 1991; Perez-Miranda et al., 2006) that uses correlations between current populations to ascertain ancestry. These analyses assume that the numbers of individuals sampled are sufficient to circumvent sampling bias and that the markers typed are selectively neutral and experience independent assortment. The genetic contributions of putative parent populations (donor or source populations) to hybrid populations (considered to be the product of genetic mixture between parental populations) were assessed while understanding that the results attained may reflect shared ancestry between the proposed parent and hybrid rather than direct genetic contributions. To investigate the populations within the mainland that have most impacted the gene pools of Taiwan's aboriginal tribes, we employed each of the Chinese populations as a parent and every Taiwanese group as a hybrid. In addition, to explore Taiwan's role as part of the Austronesian expansion, we used the Austronesian Pacific populations from New Zealand, Tonga and Samoa as well as Madagascar as hybrid groups and estimated their admixture profiles using geographical groups consisting of Chinese populations, Taiwanese aborigines,

TABLE 1. Populations analyzed

| Region | Population | Abbreviation | Average Het | Linguistic affiliation | Sample size | References | |
|-------------------------------|-------------------------|-------------------|-------------------|------------------------|--|------------------------------|----------------------|
| Africa | Angola | ANG | 0.80526 | Niger-Congo | 110 | Beleza et al., 2004 | |
| | Equatorial Guinea | EGU | 0.80311 | Niger-Congo | 134 | Alves et al., 2005 | |
| | Hutu (Rwanda) | HUT | 0.80343 | Niger-Congo | 95 | Shepard and Herrera, 2006 | |
| | Kenya | KEN | 0.80876 | Niger-Congo | 65 | Shepard and Herrera, 2006 | |
| | Madagascar | MAD | 0.81237 | Malayo-Polynesian | 67 | Regueiro et al., 2008 | |
| | Mozambique | MOZ | 0.79641 | Niger-Congo | 144 | Alves et al., 2004 | |
| | South Africa | SAF | 0.80021 | Niger-Congo | 98 | Kido et al., 2007 | |
| Australia | Tutsi (Rwanda) | TUT | 0.79941 | Niger-Congo | 126 | Regueiro et al., 2004 | |
| | Australian Aborigines | AUS | 0.77682 | Australian | 1172 | Eckoff et al., 2007 | |
| China | Chao Shan | CCS | 0.77773 | Sinitic | 144 | Hu et al., 2005 | |
| | Henan | HEN | 0.78132 | Sinitic | 400 | Xu et al., 2009 | |
| | Huis | HUI | 0.77914 | Varied* | 144 | Liu et al., 2006a | |
| | Jinan | JIN | 0.78201 | Sinitic | 420 | Tang et al., 2009 | |
| | Maonans | MAO | 0.78208 | Daic | 108 | Liu et al., 2006b | |
| | Miaos | MIA | 0.77850 | Hmong-Mien | 141 | Liu et al., 2006b | |
| | Mulaos | Mul | 0.78944 | Daic | 167 | Liu et al., 2006a | |
| | Shaanxi | CSH | 0.78317 | Sinitic | 446 | Wu et al., 2008 | |
| | Yis | YIS | 0.77582 | Tibeto-Burman | 165 | Liu et al., 2006a | |
| | Indonesian Archipelago | Bali | BAL | 0.78030 | Malayo-Polynesian | 75 | Shepard et al., 2005 |
| | | East Timor | ETI | 0.79613 | Malayo-Polynesian/ Trans-New Guinea | 186 | Souto et al., 2005 |
| | Pacific Islanders | Java | JAV | 0.79686 | Malayo-Polynesian | 60 | Shepard et al., 2005 |
| | | Surabaya | ISU | 0.79080 | Malayo-Polynesian | 105 | Dobashi et al., 2005 |
| | | New Zealand Asian | NZA | 0.79556 | Malayo-Polynesian | 153 | Bright et al., 2010 |
| New Zealand Eastern Polynesia | | NZEP | 0.78375 | Malayo-Polynesian | 1970 | Bright et al., 2010 | |
| New Zealand Western Polynesia | | NZWP | 0.78648 | Malayo-Polynesian | 644 | Bright et al., 2010 | |
| Southeast Asia | Samoa | SAM | 0.78407 | Malayo-Polynesian | 95 | Shepard et al., 2005 | |
| | Tonga | TON | 0.79155 | Malayo-Polynesian | 51 | Regueiro et al., 2008 | |
| | Kuala Lumpur (Malaysia) | MKL | 0.79863 | Malayo-Polynesian | 110 | Maruyama et al., 2008 | |
| | Malay (Malaysia) | MMA | 0.79757 | Malayo-Polynesian | 210 | Seah et al., 2003 | |
| | Iban (Malaysia) | MIB | 0.78042 | Malayo-Polynesian | 195 | Suadi et al., 2007 | |
| | Biduyah (Malaysia) | MBI | 0.77691 | Malayo-Polynesian | 195 | Suadi et al., 2007 | |
| | Melenau (Malaysia) | MME | 0.77196 | Malayo-Polynesian | 128 | Suadi et al., 2007 | |
| | Philippines | PHI | 0.78873 | Malayo-Polynesian | 106 | De Ungria et al., 2005 | |
| Taiwan | Thailand | THA | 0.79301 | Daic | 210 | Rerkamnuaychoke et al., 2006 | |
| | Ami | AMI | 0.77668 | East Formosan | 95 | Present study | |
| | Atayal | ATA | 0.73762 | Atayalic | 45 | Present study | |
| | Bunun | BUN | 0.76859 | Bunun | 47 | Present study | |
| | Paiwan | PAI | 0.79165 | Paiwanic | 39 | Present study | |
| | Puyuma | PUY | 0.78273 | Puyuma | 40 | Present study | |
| | Rukai | TUK | 0.77313 | Rukai | 33 | Present study | |
| | Saisiyat | SAI | 0.75627 | Northwest Formosan | 40 | Present study | |
| | Taiwan Han | TAIH | 0.78380 | Sinitic | 597 | Wang et al., 2003 | |
| | Tsou | TSO | 0.74884 | Tsouic | 42 | Present study | |
| Yami | YAM | 0.70817 | Malayo-Polynesian | 70 | Present study | | |

Indonesian Archipelago, and Southeast Asian groups as parental groups. Subsequently, we disassembled the aforementioned groups into individual populations to better understand the possible input of specific individual collections. We used the Australian aborigines, which are geographically proximal to the Pacific Islanders but have experienced distinctly divergent demographic histories, as a basis for comparison.

Familial relationships among individuals were estimated utilizing the ML-Relate program (Kalinowski et al., 2006) on a tribe by tribe basis. This program assesses the degrees of haplotype sharing per pair of individuals and determines whether any familial relationships may exist within a given pair and the extent of these relationships is

designated as full sibling (FS), half sibling (HS), or parent-offspring (PO). Samples deemed to be FS and PO were excluded from the MDS plot presented in Supporting Information Figure 4 in order to assess whether any differences would arise in population relationships when individuals that may be related are excluded from the analysis. Although removing HS would be ideal, it should be noted that every individual within each population was deemed "related" to at least one other individual within the group. These findings, along with endogamy practices and high inbreeding coefficients, may be indicative of ancestral ties between individuals rather than recent kinship. In addition, in our study, biographical information was procured to prevent collection from relatives.

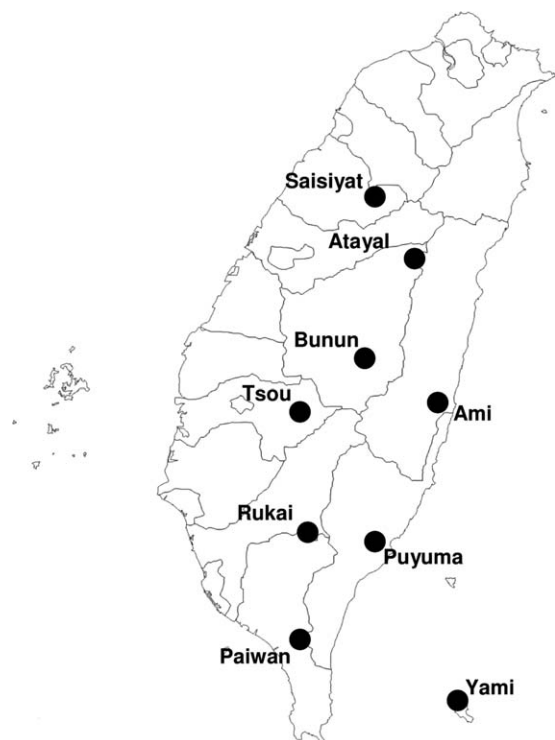


Fig. 1. Map delineating the locations within Taiwan of populations included in this study.

RESULTS

Intra-population diversity

Allelic frequency distributions, observed and expected heterozygosities along with associated P -values, as well as parameters of population genetics interest are presented in Supporting Information Tables 1–9. Statistically significant departures from Hardy–Weinberg equilibrium (HWE) expectations can be observed in several loci within every population (see Supporting Information Tables 1–9). However, the only statistically significant departures after applying the Bonferroni correction for multiple observations ($\alpha=0.05/15=0.003$) were D21S11 in the Atayal population (P -value=0.0018) and D19S433 in the Yami collection (P -value=0.0000).

Once analyzed as a whole (using all loci as a complete dataset), heterozygote deficiencies were more prominent and statistically significant in the Bunun, Puyuma, Rukai, Saisiyat and Yami (Supporting Information Table 10). Combined Matching Probabilities (CMP) are two orders of magnitude lower in the Atayal and Tsou groups (Supporting Information Table 10) than in the other populations while Combined Power of Discrimination (CPD) and Combined Power of Exclusion (CPE) are the lowest in the Tsou and Bunun collections, respectively (Supporting Information Table 10), when either 13 or 15 autosomal STR loci are employed. It should be noted that the robustness of the values for all the population genetics important parameters examined were affected by the inclusion of the two additional loci genotyped in this study; specifically, CMP values are two orders of magnitude higher when employing 13 versus 15 STR loci (Supporting Information Table 10). Inbreeding coefficients on a locus-by-locus basis are presented in Supporting Information Table 11. The

Bunun collection contains the highest number of loci (10) with P -values indicating statistically insignificant diversity correlations suggesting that it has experienced the most inbreeding/endogamy. On the other hand, the Tsou population only contains two loci that exhibit statistically insignificant P -values which is indicative of less inter-familial coupling. We also employed the ML-Relate program to investigate the degree of relatedness among individuals genotyped in our study (Supporting Information Table 12). Although some potential relationships are indicated, these aboriginal populations have likely experienced continued inbreeding, founder and/or bottleneck events in the past, and therefore, it is expected that there may be extensive allele sharing between some individuals. Number of alleles in common should not necessarily be taken as indicative of different degrees of familial relationships since they instead may be representative of distant common ancestry rather than a shared bloodline. When grouped based on geographic divides, intra-population variance is lowest among Taiwanese aborigines ($H_S=0.743899$) and highest in African populations ($H_S=0.795324$) (Table 2) again attesting to the decreased diversity in these aboriginal groups.

Inter-population diversity and population relationships

Inter-population diversity values (G_{ST}) are lowest among the Chinese ($G_{ST}=0.009179$) and Southeast Asian ($G_{ST}=0.009993$) populations (Table 2). The Taiwanese tribes display the highest variance values ($G_{ST}=0.040195$) of any of the groups examined, followed by the Pacific collections ($G_{ST}=0.026919$). The marked differences in inter-population diversity observed in the Taiwanese aborigines are also readily observed in *Structure* projections (Fig. 3) and in both the global and Taiwan specific MDS plots (Fig. 2a,b).

In the global MDS (Stress=0.12633), the Taiwanese aborigines plot to the right of the chart with the Ami, Paiwan and Puyuma populations partitioning close to a grouping composed of South Chinese, Indonesian and Southeast Asian collections (Fig. 2a). The rest of the tribes are found scattered to the right of this cluster and distant from each other in both axes. With the exemption of the Yis, all of the southern Chinese populations are found in a tight conglomerate at the center of the Chinese/Southeast Asian/Taiwanese/Indonesian grouping. The Northern Chinese are found to the left of this ensemble except for the Jihnan collection which segregates with the Yis while the Malaysian populations, the Philippine group, and the Thailand collection are found intertwined and surrounding the southern and northern Chinese aggregates. The Pacific populations are located to the lower left quadrant following an almost geographical partition. The New Zealand Asians (NZA) collection on the other hand plots near the Southeast Asian populations following ethnic rather than geographic relationships. The Australian aborigines are located at a similar location within the first dimension of the graph; however, they are found completely isolated from any other collection across the second dimension of the plot. The African populations form a tightly packed agglutination at the extreme left of the projection with the Malagasy plotting in between this aggregate and the Asian cluster. There are some differences in positioning observed within the 13 loci Global MDS plot (Stress=0.13267),

TABLE 2. Inter (G_{st}), Intra (H_s), and Total (H_t) Population Variance

| Locus | Africa | | | China | | | Indonesian Archipelago | | | Pacific populations | | | Southeast Asia | | | Taiwanese aborigines | | | All populations | | |
|---------|----------|----------|----------|----------|----------|----------|------------------------|----------|----------|---------------------|----------|----------|----------------|----------|----------|----------------------|----------|----------|-----------------|----------|----------|
| | G_{st} | H_s | H_t | G_{st} | H_s | H_t | G_{st} | H_s | H_t | G_{st} | H_s | H_t | G_{st} | H_s | H_t | G_{st} | H_s | H_t | G_{st} | H_s | H_t |
| D8S1179 | 0.011441 | 0.778393 | 0.787402 | 0.008241 | 0.846774 | 0.853811 | 0.007614 | 0.845357 | 0.851842 | 0.015338 | 0.819269 | 0.832030 | 0.010157 | 0.845634 | 0.854311 | 0.070845 | 0.769315 | 0.827973 | 0.035708 | 0.814310 | 0.844464 |
| D21S11 | 0.008926 | 0.853657 | 0.861345 | 0.004005 | 0.816930 | 0.820216 | 0.008777 | 0.849134 | 0.856652 | 0.017495 | 0.819664 | 0.834259 | 0.009099 | 0.838645 | 0.846345 | 0.037838 | 0.814309 | 0.846333 | 0.029584 | 0.829880 | 0.865179 |
| D7S820 | 0.001035 | 0.759868 | 0.767648 | 0.008192 | 0.751428 | 0.757635 | 0.003065 | 0.771376 | 0.773747 | 0.023009 | 0.795827 | 0.814570 | 0.004970 | 0.760555 | 0.764354 | 0.400620 | 0.741298 | 0.772684 | 0.032394 | 0.760670 | 0.786136 |
| CSF1PO | 0.031482 | 0.772736 | 0.787854 | 0.027997 | 0.731118 | 0.732828 | 0.005778 | 0.772828 | 0.771704 | 0.011209 | 0.711074 | 0.719772 | 0.011223 | 0.706941 | 0.714965 | 0.027576 | 0.712276 | 0.732475 | 0.050785 | 0.727388 | 0.766305 |
| D3S1358 | 0.009851 | 0.726665 | 0.733894 | 0.009677 | 0.718748 | 0.725771 | 0.010625 | 0.707418 | 0.715015 | 0.006112 | 0.721023 | 0.725457 | 0.009360 | 0.734915 | 0.742309 | 0.023178 | 0.694551 | 0.711031 | 0.014308 | 0.717385 | 0.727798 |
| TH01 | 0.008052 | 0.733299 | 0.739251 | 0.008625 | 0.667912 | 0.673723 | 0.023866 | 0.764891 | 0.783583 | 0.076644 | 0.743742 | 0.805477 | 0.016681 | 0.748371 | 0.761066 | 0.068909 | 0.677258 | 0.727381 | 0.062892 | 0.714071 | 0.761994 |
| D13S317 | 0.011208 | 0.728255 | 0.736510 | 0.009611 | 0.789574 | 0.797236 | 0.005852 | 0.783208 | 0.787819 | 0.027634 | 0.780455 | 0.802635 | 0.013393 | 0.784972 | 0.795628 | 0.044930 | 0.688956 | 0.721367 | 0.057370 | 0.755458 | 0.801437 |
| D15S139 | 0.009348 | 0.734605 | 0.742009 | 0.021980 | 0.778645 | 0.786144 | 0.005237 | 0.780046 | 0.784153 | 0.016466 | 0.782389 | 0.795487 | 0.020617 | 0.777718 | 0.794090 | 0.028799 | 0.770164 | 0.793002 | 0.025986 | 0.778625 | 0.798398 |
| D2S1338 | 0.009069 | 0.884462 | 0.892557 | 0.007362 | 0.856669 | 0.863022 | 0.010642 | 0.843022 | 0.852080 | 0.016009 | 0.860264 | 0.874260 | 0.011895 | 0.852154 | 0.862412 | 0.032771 | 0.816744 | 0.844416 | 0.027342 | 0.852374 | 0.876334 |
| D19S433 | 0.010840 | 0.827658 | 0.836728 | 0.003315 | 0.810767 | 0.813464 | 0.007100 | 0.816281 | 0.822118 | 0.033621 | 0.801717 | 0.829609 | 0.006271 | 0.807996 | 0.813095 | 0.042871 | 0.770510 | 0.805022 | 0.029562 | 0.804310 | 0.828811 |
| vWA | 0.008079 | 0.814126 | 0.820757 | 0.007281 | 0.789113 | 0.794900 | 0.014347 | 0.802726 | 0.814411 | 0.021796 | 0.785993 | 0.803506 | 0.010067 | 0.790185 | 0.798221 | 0.022487 | 0.789826 | 0.807995 | 0.034147 | 0.794907 | 0.823009 |
| TPOX | 0.018326 | 0.764880 | 0.779158 | 0.004463 | 0.604731 | 0.607442 | 0.035480 | 0.621513 | 0.644375 | 0.084569 | 0.651579 | 0.711773 | 0.005101 | 0.605996 | 0.609048 | 0.039790 | 0.597916 | 0.622693 | 0.054058 | 0.640204 | 0.676790 |
| D18S51 | 0.007758 | 0.865423 | 0.872189 | 0.004716 | 0.859470 | 0.863543 | 0.010332 | 0.829943 | 0.838607 | 0.025463 | 0.841024 | 0.862999 | 0.009704 | 0.840805 | 0.849044 | 0.032125 | 0.788350 | 0.814517 | 0.030903 | 0.837835 | 0.864552 |
| D5S818 | 0.009096 | 0.758130 | 0.765089 | 0.006724 | 0.768634 | 0.773837 | 0.010130 | 0.735808 | 0.743338 | 0.015670 | 0.772990 | 0.785296 | 0.004388 | 0.777419 | 0.780845 | 0.041028 | 0.711467 | 0.741906 | 0.037136 | 0.754406 | 0.783502 |
| FGA | 0.007607 | 0.877704 | 0.884431 | 0.006394 | 0.865037 | 0.870603 | 0.009999 | 0.862847 | 0.871562 | 0.017363 | 0.829773 | 0.844435 | 0.005737 | 0.851139 | 0.856050 | 0.048356 | 0.815540 | 0.856981 | 0.024876 | 0.849994 | 0.871678 |
| All | 0.011350 | 0.795324 | 0.804455 | 0.009179 | 0.777037 | 0.784235 | 0.010953 | 0.782148 | 0.790810 | 0.026919 | 0.781161 | 0.802771 | 0.009993 | 0.781563 | 0.789452 | 0.040195 | 0.743899 | 0.775052 | 0.036095 | 0.775454 | 0.804493 |

however, general assemblages remain the same in both projections (Supporting Information Fig. 1).

The Taiwanese specific MDS (Stress=0.13951) projection echoes the major differences observed in both the global plot and G_{st} values (Fig. 2b and Table 2). There are considerable differences between the aboriginal Taiwanese collections genotyped in this study and those in Lee et al. (2002). Most of the collections identified as belonging to the same tribe map in the general vicinities of each other, however, the Tsou, Atayal and Saisiyat counterpart collections segregate more distant from each other. The Puyuma and Paiwan populations' genotype in the present study as well as in Lee et al. (2002) are found closely associated with the Han collections from Taiwan suggesting that these populations may have experienced admixture with the Min Nan and Hakka speaking majorities in the island. Stress values for every plot were compared to the cutoff values calculated by Sturrock and Rocha (2000) and iTwas deemed that the comparisons are sound given the number of objects evaluated.

The *Structure* plot based on 15 loci illustrates considerable sub-structuring (Fig. 3) among Taiwanese aborigines with a clear signature shared by the Ami, Paiwan, Puyuma and Rukai groups observed at $k=5$ and 6 in yellow. At $k=6$ (light blue) some clustering between the Bunun and Tsou is also apparent while the Atayal, Saisiyat and Yami emerge as independent entities within this plot (Fig. 3). Parallelisms among Taiwanese aborigines, in relation to Madagascar and Tonga, are most evident at $k=2$, with Paiwan, Puyuma, Rukai and Saisiyat conferring the greatest similarities. The greatest differences observed in the *Structure* graph performed at 13 loci from the aforementioned results involve the demarcation between the Atayal and Bunun collections. Whereas these differences are apparent at 15 loci at every k explored, this divergence is not noted until $k=7$ in the 13 loci projection (Supporting Information Fig. 3). No further subdivisions were observed beyond $k=7$ (data not shown) at the 15 and 13 loci levels. Upon calculating delta k , it was assessed that the highest level of structure for the 15 loci data set exists at $k=4$ (Supporting Information Fig. 2).

In the admixture proportion analyses, the major contributors from mainland China to the Taiwanese aborigines are the Mulaos, contributing as much as 100% of the Bunun gene pool and the Maonans, contributing as much as 43.9% to the Paiwan collection (Table 3). In contrast, the Taiwan Han population receives 49.5% of its genetic diversity from the Shaanxi Han group. Of the tribal groups, the only population that does not derive most of its autosomal components from the southern Daic populations (Mulaos and Maonans) is the Rukai tribe, receiving 48.2% of its genetic input from the Huis instead.

The Taiwanese aborigines are the major contributors to the New Zealand Western Polynesians [WPNZ (43.3%)], Samoans (39.9%) and Tongans (45.1%) (Table 4). The Malagasy population is most impacted by Indonesia (36.8%) while the New Zealand Eastern Polynesians (NZEP) receives 66.3% of their autosomal gene pool from South East Asia. The NZA ethnic group obtains 70.5% of their genetics from the China group (Table 4) while the Australian aborigines derive their entire autosomal component from Indonesian populations. All of the hybrid populations included, except for the Australian aborigines, display aboriginal Taiwanese

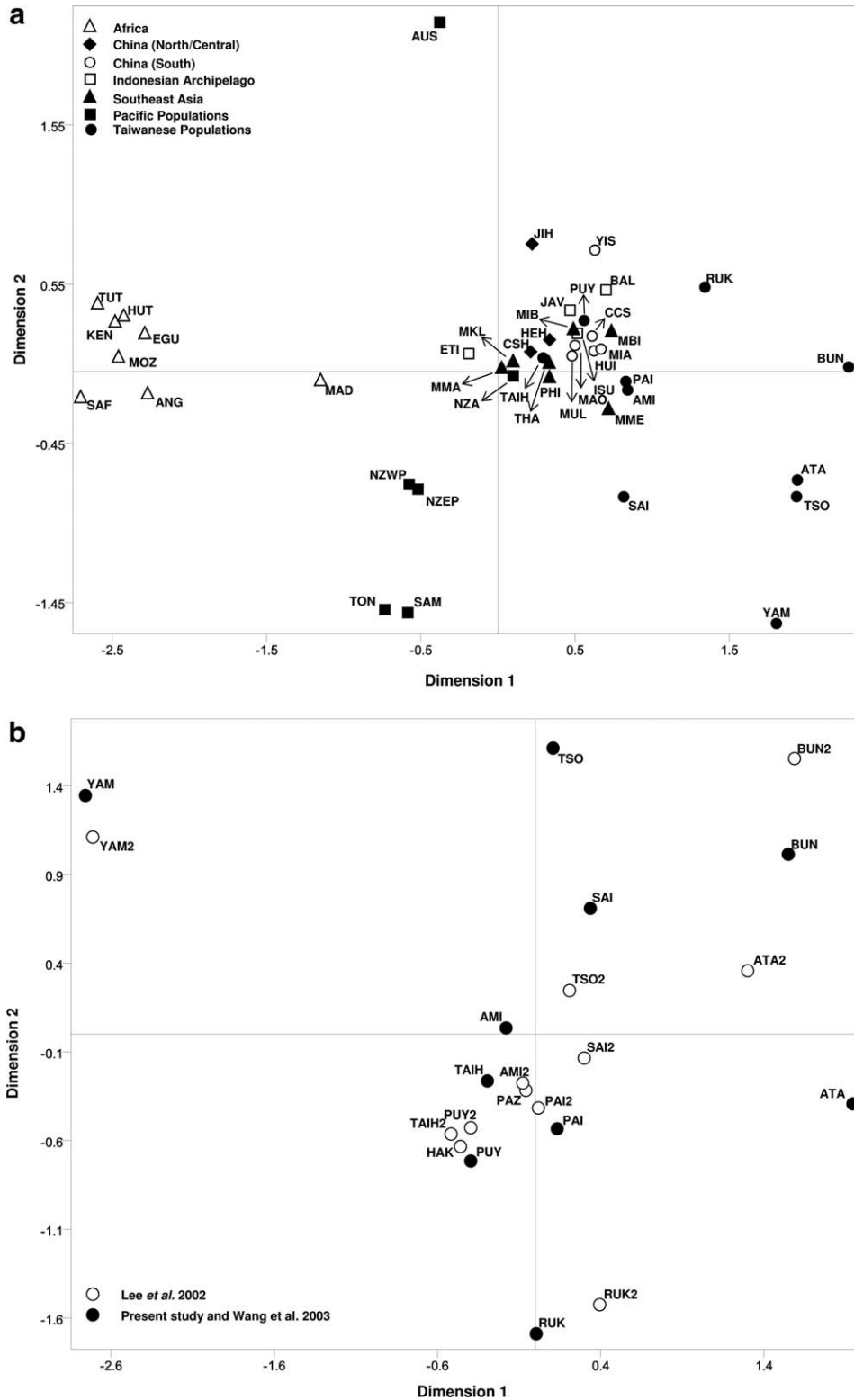


Fig. 2. (a) Global MDS plot based on 15 autosomal STRs (Stress=0.12633 and RSQ=0.95154). Please refer to Table 1 for population abbreviations. (b) Multidimensional Scaling (MDS) plot of Taiwanese populations at 13 autosomal STR loci (Stress=0.13951 RSQ=0.94741). AMI (Ami), ATA (Atayal), BUN (Bunun), HAK (Hakka), PAI (Paiwan), PAZ (Pazze), PUY (Puyuma), RUK (Rukai), SAI (Saisiyat), TAIH (Taiwan Han), TSO (Tsou), YAM (Yami or Tao).

influences (Table 4). Of the Southeast Asian populations, the Malaysian Malays and Kuala Lumpur collections appear to be the most genetically influential for all of

the populations, donating over 60% of each hybrid's Southeast Asian derived autosomal component except for the NZA which are affected by Thailand instead

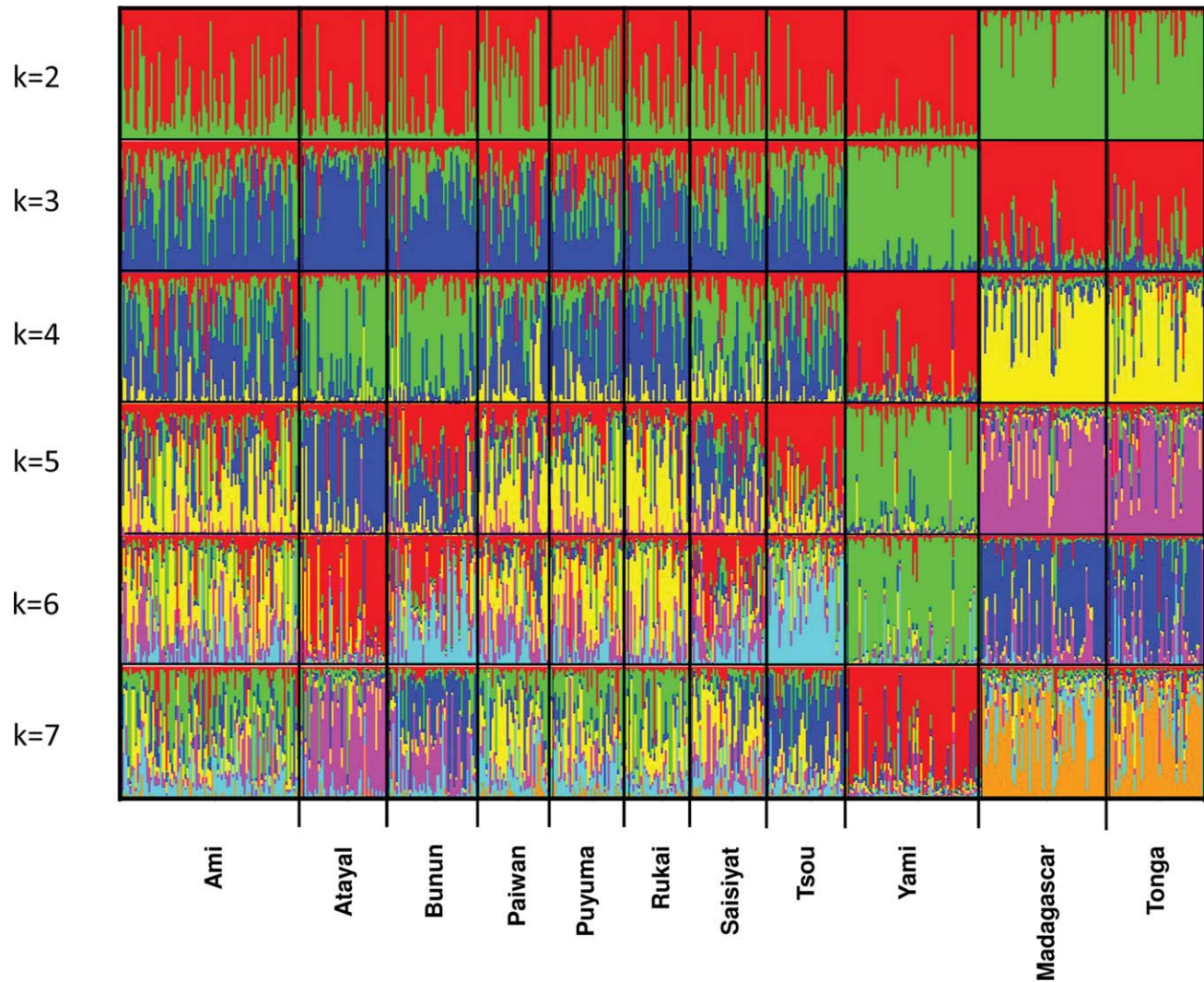


Fig. 3. *Structure* projections of Taiwanese aborigines, Madagascar and Tonga utilizing 15 autosomal STRs.

(Table 4). East Timor (the only Melanesian collection in this study) is the greatest individual donor within the Indonesian grouping with values ranging from 88.6% in the Australian aborigines to 30.2% in the NZA (Table 4).

Pair-wise familial relationships between individuals within tribes are presented in Supporting Information Table 12. It is apparent that considerable allele sharing exists within all the Taiwanese tribes examined in this study, given that each individual within every population was determined by the software to be a HS to at least one other individual within the tribes investigated. In order to ascertain whether familial ties lie at the root of the population relationships observed in our study, individuals that appear to be FS or PO were removed from the analyses employed to generate the MDS plot presented in Supporting Information Figure 4. It is apparent that although some minor changes have occurred (for example, the Puyuma population typed in this article no longer maps together with the Taiwan Han, Hakka and Puyuma populations from Lee et al., 2002, see Fig. 2b), the overall relationships depicted in this plot have not changed dramatically suggesting that the trends observed throughout this study are not

merely due to recent familial relationships but likely to ancestral relatedness.

DISCUSSION

Genetic relationships within Taiwan and with mainland China

Population relationships in the global MDS plot (Fig. 2a) display a clear agglutination of Southern Chinese groups independent of ethno-linguistic affiliations (see Table 1 for linguistic affiliations). Three Taiwanese populations (Ami, Paiwan and Puyuma) are found in close proximity to the aforementioned cluster; however, the rest of the Taiwanese aborigines are found scattered to the right of the chart with no clear partitioning among them. The relationship between southern mainland populations and aboriginal groups is further supported by admixture proportions indicating strong genetic relationships between the Tai-Kadai (Daic) speaking groups (Mulaos and Maonans) and all Formosan collections (Table 3). It is interesting to note that the other (non-Daic) southern mainland groups display little influence in comparison, and only the Huis make

TABLE 3. Autosomal Contributions of Chinese Populations to Taiwanese Aborigines

| Parental population | Hybrid Populations | | | | | | | | | | Taiwan Han |
|---------------------|--------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|------------|
| | Ami | Atayal | Bunun | Paiwan | Puyuma | Rukai | Saisiyat | Tsou | Yami | | |
| Chao Shan | 0.000±0.095 | 0.000±0.053 | 0.000±0.008 | 0.000±0.074 | 0.000±0.075 | 0.000±0.084 | 0.000±0.054 | 0.000±0.015 | 0.000±0.028 | 0.197±0.073 | |
| Henan | 0.142±0.154 | 0.000±0.058 | 0.000±0.016 | 0.000±0.036 | 0.000±0.154 | 0.000±0.131 | 0.000±0.039 | 0.000±0.083 | 0.329±0.273 | 0.023±0.068 | |
| Huis | 0.000±0.002 | 0.153±0.222 | 0.000±0.092 | 0.000±0.147 | 0.217±0.187 | 0.482±0.287 | 0.000±0.087 | 0.000±0.040 | 0.000±0.086 | 0.081±0.060 | |
| Jinan | 0.000±0.007 | 0.000±0.022 | 0.000±0.047 | 0.073±0.081 | 0.000±0.019 | 0.000±0.039 | 0.056±0.092 | 0.000±0.028 | 0.275±0.188 | 0.010±0.022 | |
| Maos | 0.426±0.170 | 0.000±0.083 | 0.000±0.079 | 0.439±0.241 | 0.000±0.122 | 0.347±0.222 | 0.000±0.101 | 0.047±0.183 | 0.000±0.085 | 0.040±0.046 | |
| Miaos | 0.000±0.077 | 0.000±0.114 | 0.000±0.015 | 0.000±0.056 | 0.000±0.087 | 0.000±0.033 | 0.000±0.082 | 0.043±0.220 | 0.000±0.171 | 0.054±0.047 | |
| Mulaos | 0.215±0.141 | 0.847±0.246 | 1.000±0.140 | 0.370±0.216 | 0.418±0.213 | 0.025±0.153 | 0.925±0.174 | 0.814±0.265 | 0.380±0.255 | 0.099±0.057 | |
| Shaanxi | 0.217±0.185 | 0.000±0.021 | 0.000±0.049 | 0.073±0.130 | 0.365±0.233 | 0.031±0.138 | 0.000±0.043 | 0.000±0.022 | 0.000±0.051 | 0.495±0.087 | |
| Yis | 0.000±0.038 | 0.000±0.044 | 0.000±0.014 | 0.045±0.068 | 0.000±0.064 | 0.116±0.113 | 0.019±0.081 | 0.097±0.122 | 0.016±0.086 | 0.000±0.009 | |

a major contribution (48.2% in the Rukai) attesting to the plausibility of a Daic origin for Taiwanese aborigines (Li et al., 2008). This presence from the Huis in the Rukai and to a lesser extent in the Atayal (15.3%) and Puyuma (21.7%) may emanate from the highly admixed nature of the Huis which represents an amalgamation of descendants from Silk Road travelers that distinguish themselves from other Chinese ethnic groups by their practice of Islam. On the other hand, the Taiwanese Han collection is most influenced by the Shaanxi (a northern Han population) and the Chao Shan (a southern Han group) and receives only 1–10% of its genetics from each of the remaining populations included except for the Yis who are Tibeto-Burman speakers (Table 3). This is expected, considering that the Min Nan and Hakka languages (Han Chinese linguistic groups) are both largely represented within Taiwan (Chiung, 2001).

Despite some overall similarities in the populations that exert genetic effects on the Formosan tribes (Table 3), there exists high variability between the degrees of admixture with Han groups and in the contributions of any given southern Chinese collection to each of the aboriginal populations (Table 3). These differences are echoed in all MDS plots (Fig. 2a,b, Supporting Information Fig. 1) where Taiwanese aborigines exhibit no clear partitioning among them and appear as separate from each other as they are from Island Southeast Asian, Indonesians and mainland Asian populations (Fig. 2a). Furthermore, sampling appears to make a definitive impact on the results from studies utilizing Taiwanese aborigines [For example Kayser et al., 2003, 2006, 2008 see very few O-M122 derived Y-chromosomes whereas Su et al. (2000) and Mirabal et al. (2012) observe this haplogroup at considerable frequencies in the Ami (37%)], leading to the widespread differences between groups identified as belonging to the same tribes but published elsewhere (Lee et al., 2002) and our results (Fig. 2b). These differences may be the result of reduced gene flow among sub-tribal communities which has previously been observed in the Yami of Orchid Island (Loo et al., 2011). It should also be noted that when individuals that may be related to each other are removed from the general collections, the observed differences in the MDS plot are not substantial (see Supporting Information Fig. 4 and Fig. 2b) suggesting that the relationships observed throughout this study are not due to recent contact but may emanate from ancestral ties.

Genetic distances among Taiwanese aborigines ($G_{st}=0.040195$) are such that inter-population variance indices are higher than those found among Pacific populations ($G_{st}=0.026919$) which are geographically isolated and contain major biogeographical barriers between them in the form of vast open oceanic distances. Corroborating these statements are the low heterozygosity values (Table 1), severe heterozygote deficiencies (Supporting Information Table 10), high number of loci exhibiting positive inbreeding coefficient correlations (Supporting Information Table 11) and the lowest intra-population variance of all the groups of populations analyzed ($H_s=0.743899$) exhibited by the Taiwanese tribes (Table 2). Altogether, these data suggest that while Taiwanese tribes are genetically different from each other, they are internally highly homogenous, likely due to continued endogamy and genetic drift. These data parallel linguistic divergences between tribes given that though some words are shared across the various Formosan

TABLE 4. Admixture Proportions of Austronesian Populations

| Parental groups | Hybrid populations | | | | | | | Australian Aborigines | |
|----------------------|--------------------|------------------|------------------|-------------|-------------|-------------|-------------|-----------------------|--|
| | New Zealand (WP) | New Zealand (EP) | New Zealand (As) | Madagascar | Samoa | Tonga | | | |
| Chinese | 0.000±0.164 | 0.135±0.218 | 0.705±0.062 | 0.000±0.111 | 0.000±0.133 | 0.000±0.100 | 0.000±0.066 | | |
| Indonesia | 0.285±0.180 | 0.000±0.127 | 0.076±0.054 | 0.368±0.228 | 0.221±0.203 | 0.392±0.223 | 1.000±0.083 | | |
| Southeast Asia | 0.282±0.260 | 0.663±0.280 | 0.105±0.083 | 0.353±0.271 | 0.380±0.247 | 0.158±0.235 | 0.000±0.000 | | |
| Taiwanese Aborigines | 0.433±0.155 | 0.202±0.175 | 0.114±0.041 | 0.279±0.177 | 0.399±0.176 | 0.451±0.183 | 0.000±0.047 | | |
| Parental populations | | | | | | | | | |
| Chinese | | | | | | | | | |
| Henan | - | 0.000±0.130 | 0.000±0.076 | - | - | - | - | | |
| Jinan | - | 0.000±0.075 | 0.063±0.046 | - | - | - | - | | |
| Chao Shan | - | 0.000±0.032 | 0.015±0.048 | - | - | - | - | | |
| Shaanxi | - | 0.359±0.237 | 0.463±0.094 | - | - | - | - | | |
| Huis | - | 0.000±0.181 | 0.135±0.103 | - | - | - | - | | |
| Yis | - | 0.124±0.116 | 0.037±0.025 | - | - | - | - | | |
| Miaos | - | 0.000±0.030 | 0.000±0.021 | - | - | - | - | | |
| Mulaos | - | 0.510±0.231 | 0.257±0.064 | - | - | - | - | | |
| Maos | - | 0.007±0.122 | 0.030±0.051 | - | - | - | - | | |
| Indonesia | | | | | | | | | |
| Surabaya | 0.314±0.157 | - | 0.487±0.100 | 0.382±0.145 | 0.330±0.169 | 0.272±0.164 | 0.000±0.032 | | |
| Bali | 0.082±0.118 | - | 0.211±0.070 | 0.000±0.016 | 0.106±0.115 | 0.000±0.077 | 0.065±0.123 | | |
| Java | 0.000±0.033 | - | 0.000±0.057 | 0.000±0.026 | 0.000±0.035 | 0.000±0.011 | 0.049±0.098 | | |
| East Timor | 0.603±0.146 | - | 0.302±0.090 | 0.618±0.142 | 0.564±0.147 | 0.728±0.179 | 0.886±0.151 | | |
| Southeast Asia | | | | | | | | | |
| Kuala Lumpur | 0.488±0.244 | 0.085±0.170 | 0.062±0.077 | 0.357±0.246 | 0.188±0.213 | 0.306±0.248 | - | | |
| Malay | 0.302±0.215 | 0.643±0.253 | 0.089±0.087 | 0.528±0.256 | 0.572±0.263 | 0.397±0.265 | - | | |
| Iban | 0.000±0.004 | 0.000±0.052 | 0.000±0.000 | 0.000±0.054 | 0.000±0.002 | 0.000±0.052 | - | | |
| Biduyar | 0.000±0.040 | 0.000±0.086 | 0.000±0.004 | 0.000±0.003 | 0.000±0.047 | 0.000±0.032 | - | | |
| Menelau | 0.047±0.081 | 0.000±0.053 | 0.010±0.032 | 0.030±0.069 | 0.209±0.123 | 0.133±0.129 | - | | |
| Philippines | 0.163±0.160 | 0.235±0.151 | 0.177±0.061 | 0.084±0.123 | 0.031±0.116 | 0.164±0.181 | - | | |
| Thailand | 0.000±0.126 | 0.038±0.149 | 0.662±0.097 | 0.000±0.056 | 0.000±0.111 | 0.000±0.060 | - | | |
| Taiwanese Aborigines | | | | | | | | | |
| Ami | 0.033±0.119 | 0.270±0.177 | 0.341±0.088 | 0.187±0.132 | 0.000±0.109 | 0.000±0.112 | - | | |
| Atayal | 0.027±0.059 | 0.003±0.058 | 0.025±0.041 | 0.069±0.078 | 0.000±0.048 | 0.041±0.070 | - | | |
| Bunun | 0.000±0.043 | 0.000±0.048 | 0.009±0.037 | 0.000±0.041 | 0.000±0.036 | 0.000±0.051 | - | | |
| Rukai | 0.002±0.079 | 0.000±0.053 | 0.000±0.034 | 0.000±0.021 | 0.000±0.083 | 0.000±0.045 | - | | |
| Paiwan | 0.328±0.110 | 0.291±0.128 | 0.171±0.080 | 0.382±0.123 | 0.395±0.143 | 0.293±0.121 | - | | |
| Puyuma | 0.190±0.110 | 0.230±0.131 | 0.290±0.091 | 0.177±0.129 | 0.129±0.106 | 0.199±0.140 | - | | |
| Saisiyat | 0.271±0.098 | 0.129±0.104 | 0.100±0.063 | 0.185±0.105 | 0.359±0.111 | 0.238±0.115 | - | | |
| Tsou | 0.056±0.074 | 0.066±0.076 | 0.000±0.035 | 0.000±0.053 | 0.026±0.075 | 0.146±0.114 | - | | |
| Yami | 0.092±0.070 | 0.010±0.055 | 0.065±0.046 | 0.000±0.015 | 0.091±0.075 | 0.083±0.073 | - | | |

Hyphens are indicative of populations without input at the group level were not analyzed at the individual level for that given set of populations.

languages, they are mostly intelligible from each other (Blust, 1999).

The *Structure* analyses provide another clue into the current genetic relationships within Taiwan, illustrating signatures shared by the Ami, Paiwan, Puyuma and Rukai (yellow) and between the Bunun and Tsou (blue), while the Atayal, Saisiyat and Yami emerge as highly differentiated from each other and any of the other groups ($k=6$, Fig. 3). No further subpopulation structure was observed by increasing the number of sources (k values).

The Yami collection appears the most distant from all other Taiwanese aboriginal populations and displays the lowest average heterozygosity (0.70817) both within this grouping and of all the populations analyzed in this study (Fig. 2a,b, Table 1 Supporting Information Fig. 1). This population is located in small Orchid Island, a 45 km² volcanic island about 100 km from Taiwan's southeastern coast, which is located in what has been regarded as a via-point out of Taiwan into the Philippines (Tsang, 2005). The natives currently speak a Malayo-Polynesian language belonging to the Batanic sub-branch which is closely related to languages spoken in the islands found between the Philippines and Taiwan (Loo et al., 2011). In the current study, no significant genetic similarities were observed between the Yami and Filipino groups, echoing findings by Loo et al. (2011) based on Y-chromosomal studies suggesting that these linguistic similarities are the result of acculturation emanating from continued, well documented, trade between Orchid Islanders and the Batanes Archipelago of the Philippines from ~2,500 and 300 YBP (Loo et al., 2011).

Beyond Taiwan: The Austronesian diaspora in the Pacific and Indian Oceans

It is of interest that only two of the Pacific populations, NZA and NZEP, both from New Zealand, are impacted by mainland Chinese groups in the admixture proportions (Table 4). The fact that the NZA population is of Asian descent is the likely explanation for its rejection. In contrast, all the Austronesian Pacific populations studied (as well as Madagascar), exhibit widespread influences from Taiwanese aborigines, these influences are even greater, in most instances, than the impact from populations more geographically proximal (for example Indonesia). Notably, the three Taiwanese aboriginal groups that seem largely responsible for the aforementioned effects (Paiwan, Puyuma and Saisiyat) have been largely understudied. The Ami, which not only is the largest tribe (population wise) within Taiwan but the most represented in the literature, only affects the New Zealand groups and Madagascar but not Samoa and Tonga (Table 4), indicating that it is essential to include a complete survey of these tribal groups when ascertaining phylogenetic relationships. It is of importance to define that Taiwanese populations have seemingly experienced high levels of endogamy, isolation, and/or genetic drift allowing for the extensive molecular differentiation among groups. It is plausible that the initial genetic signatures that have now become the core of what we understand to be Austronesian specific genetic signals may have been lost in some of the tribes [e.g., subhaplogroups O3 (M122), and further O3a2 (P201) and undifferentiated O3a2c* (P164) are only present at significant levels in the Ami and only minimally in other

Taiwanese aboriginal groups]. Therefore, autosomal as well as mtDNA and Y chromosomal studies surveying the complete set of aboriginal populations should provide a deeper understanding of the Austronesian diaspora.

CONCLUSION

The Austronesian diaspora represents one of the largest human dispersals in the peopling of the world and though the genetic and linguistic footprints of Austronesian speakers have been widely studied, many questions remain regarding their origins and migratory routes. We have explored the genetic relationships among Taiwanese tribes and found them to be not only highly heterogeneous when compared to each other but highly homogenous on an intra-population basis supporting previous statements suggesting that genetic drift, long periods of isolation and/or endogamy within each tribe have occurred. Additionally, in the hope of shedding light onto the likely source for these Formosan populations, we have compared these tribes in admixture analyses and found support (in congruence with previous reports) to the idea that the southern Chinese Daic domain may be the cradle of Proto-Austronesian migrants to Taiwan. We compared each Taiwanese aboriginal population to groups found at the fringes of the Austronesian range in the Pacific and Indian oceans and found that the Puyuma, Saisiyat and Paiwan are the most genetically influential tribes to modern Austronesian populations and theorize that other Taiwanese tribes may have lost these genetic signatures due to isolation, drift and/or endogamy. Altogether, our data lend support to the slow boat hypothesis of Austronesian dispersal where clear genetic fingerprints from Taiwan can be discerned despite various levels of admixture from the surrounding regions. Furthermore, theories suggesting a Daic origin to all Austronesians are also supported. These results should be tempered, however, due to the limited number of Melanesian reference populations in the current report due to unavailability in the literature.

LITERATURE CITED

- Alves C, Gusmao L, Damasceno A, Soares B, Amorim A. 2004. Contribution for an African autosomic STR database (AmpFISTR Identifiler and Powerplex 16 system) and a report on genotypic variations. *Forensic Sci Int* 139:201–205.
- Alves C, Gusmao L, Lopez-Parra AM, Soledad-Mesa M, Amorim A, Arroyo-Pallo E. 2005. STR allelic frequencies for an African population sample (Equatorial Guinea) using AmpFISTR Identifiler and Powerplex 16 kits. *Forensic Sci Int* 148:239–242.
- 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.
- Applied Biosystems. 2001. AmpFISTR identifiler PCR amplification kit user's manual. Foster City, California: Applied Biosystems.
- Bellwood P. 1990. From Late Pleistocene to Early Holocene in Sundaland. In: Gable C, Sofer O, editors. *The world at 18,000 BP*, Vol. 2. London: Unwin Hyman. p 255–263.
- Bellwood P. 1997. *Prehistory of the Indo-Malaysian Archipelago*, revised edition. Honolulu: University of Hawaii Press.
- Bellwood PA, James JF, Barrell T. 1995. *The Austronesians: historical and comparative perspectives*. Canberra: Department of Anthropology, Research School of Pacific and Asian Studies, Australian National University.

- Beleza S, Alves C, Reis F, Amorim A, Carracedo A, Gusmao L. 2004. 17 STR data (AmpFISTR Identifier and Powerplex 16 system) from Cabinda (Angola). *Forensic Sci Int* 14:193–196.
- Bright JA, Buckelton JS, McGovern CE. 2010. Allele frequencies for the four major sub-populations of New Zealand for the 15 Identifier loci. *Forensic Sci Int Genet* 4:265–266.
- Blust R. 1999. Subgrouping, circularity and extinction: some issues in Austronesian comparative linguistics. *Symp Ser Inst Linguist Acad Sinica* 1:31–94.
- Brenner C, Morris J. 1989. Paternity index calculations in single locus hypervariable DNA probes: validation and other studies. In: *Proceedings for the International Symposium on Human Identification Promega Corporation, Madison*. p 21–53.
- Capelli C, Wilson JF, Richards M, Stumpf MP, 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. *AJHG* 68:432–443.
- Chiung WVT. 2001. Language attitudes toward written Taiwanese. *J Multiling Multicult Dev* 22:502–523.
- Chu CC, Nakajima LM, 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.
- De Ungria MCA, Roby RK, Tabbada KA, Rao Coticone S, Tan MM, Hernandez KN. 2005. Allele frequencies of 19 STR loci in a Philippine population generated using AmpFISTR multiplex and ALF singleplex systems. *Forensic Sci Int* 152:281–284.
- Diamond JM. 1988. Express train to Polynesia. *Nature* 336:307–308.
- Diamond JM. 2000. Taiwan's gift to the world. *Nature* 403:709–710.
- Diamond JM, Bellwood P. 2003. Farmers and their languages: the first expansions. *Science* 300:597–603.
- Dobashi Y, Kido A, Fujitani N, Hara M, Susukida R, Oya M. 2005. STR data for the AmpFISTR Identifier loci in Bangladesh and Indonesian populations. *Legal Med* 7:222–226.
- Donohue M, Denham T. 2010. Farming and language in island Southeast Asia: reframing Austronesian history. *Curr Anthropol* 51:223–256.
- Eckhoff C, Walsh S, Buckleton J. 2007. Population data from sub-populations of the Northern Territory of Australia for 15 autosomal short tandem repeat (STR) loci. *Forensic Sci Int* 171:237–249.
- Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14:2611–2620.
- Friedlaender JS, Friedlaender FR, Reed FA, Kidd KK, Kidd JR, Chambers GK, Lea RA, Loo JH, Koki G, Hodgson JA, Merriwether DA, Weber JL. 2008. The genetic structure of Pacific islanders. *PLoS Genet* 4:e19.
- Gray RD, Drummond AJ, and Greenhill SJ. 2009. Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science* 323:479–483.
- Gray RD, Jordan FM. 2000. Language trees support the express-train sequence of Austronesian expansion. *Nature* 405:1052–1055.
- Green RC. 1999. Integrating historical linguistics with archaeology: insights from research in Remote Oceania. *Indo-Pac Prehist Assoc Bull* 18:3–16.
- Greenhill SJ, Drummond AJ, Gray RD. 2010. How accurate and robust are the phylogenetic estimates of Austronesian language relationships? *PLoS One* 5:e9573.
- Grimes BF. 2002. *Ethnologue: languages of the world*, 14th ed. Dallas: International Academic Bookstore.
- Guo S, Thompson E. 1992. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* 48:361–372.
- Hagelberg E, Goldman N, Lio P, Whelan S, Schiefenhover W, Clegg JV, Bowden DK. 1999. Evidence for mitochondrial DNA recombination in a human population of island Melanesia. *Proc R Soc Lond B* 226:485–492.
- Hill C, Soares P, Mormina M, Macaulay V, Clarke D, Blumbach PB, Vizuete-Forster M, Forster P, Bulbeck D, Oppenheimer S, Richards M. 2007. A mitochondrial stratigraphy for island Southeast Asia. *Am J Hum Genet* 80:29–43.
- Hu SP, Yu XJ, Liu JW, Cai KL. 2005. Analysis of STR polymorphisms in the Chao Shan population in South China. *Forensic Sci Int* 147:93–95.
- Hurles ME, Nicholson J, Bosch E, Renfrew C, Sykes BC, Jobling MA. 2002. Y Chromosomal evidence for the origins of Oceanic-speaking peoples. *Genetics* 160:289–303.
- Ibarra-Rivera L, Mirabal S, Regueiro M, Herrera RJ. 2008. Delineating genetic relationships among the Maya. *Am J Phys Anthropol* 135:329–347.
- Jinam TA, Hong LC, Phipps ME, Stoneking M, Ameen M, Edo J. 2012. Evolutionary history of continental Southeast Asians: “early train” hypothesis based on genetic analysis of mitochondrial and autosomal DNA data. *HUGO Pan-Asian SNP Consortium, Saitou N. Mol Biol Evol* 29:3513–3527.
- Jones DA. 1972. Blood samples: probability of discrimination. *J Forensic Sci Soc* 12:355–359.
- Jorde LB, Rogers AR, Bamshad M, Watkins WS, Krakowiak P, Sung S, Kere J, Harpending HC. 1997. Microsatellite diversity and the demographic history of modern humans. *Proc Natl Acad Sci USA* 94:3100–3103.
- Kalinowski ST, AP Wagner, ML Taper. 2006. ML-Relate: a computer program for maximum likelihood estimation of relatedness and relationship. *Mol Ecol Notes* 6:576–579.
- Karafet T, Hallmark B, Cox M, Sudoyo H, Downey S, Lansing S, Hammer M. 2010. Major East-West division underlies Y chromosome stratification across Indonesia. *Mol Biol Evol* 27:1833–1844.
- Kayser M, Brauer S, Weiss G, Underhill PA, Roewer L, Schiefenhövel W, Stoneking M. 2000. Melanesian origin of Polynesian Y chromosomes. *Curr Biol* 10:1237–1246.
- Kayser M, Brauer S, Weiss G, Schiefenhövel W, Underhill P, Shen P, Oefner P, Tommaseo-Ponzetta M, Stoneking M. 2003. Reduced Y-chromosome, but not mitochondrial DNA, diversity in human populations from West New Guinea. *Am J Hum Genet* 72:281–302.
- Kayser M, Brauer S, Cordaux R, Casto A, Lao O, Zhivotovskiy LA, Moise-Faurie C, Rutledge RB, Schiefenhoevel W, Gil D, Lin AA, Underhill PA, Oefner PJ, Trent RJ, Stoneking M. 2006. Melanesian and Asian origins of Polynesians: mtDNA and Y chromosome gradients across the Pacific. *Mol Biol Evol* 23:2234–2244.
- Kayser M, Lao O, Saar K, Brauer S, Wang X, Nürnberg P, Trent RJ, Stoneking M. 2008. Genome-wide analysis indicates more Asian than Melanesian ancestry of Polynesians. *Am J Hum Genet* 82:194–198.
- Kido A, Dobashi Y, Fujitani N, Hara M, Susukida R, Kimura H, Oya M. 2007. Population data on the AmpFISTR Identifier loci in Africans and Europeans from South Africa. *Forensic Sci Int* 168:232–235.
- Lee JC, Lin M, Tsai L, Hsu C, Hsieh H, Huang N, Shih RT, Wun J, Chang J, Ko Y, Tzeng C, Linacre A. 2002. Population study of polymorphic microsatellite DNA in Taiwan. *Forensic Sci J* 1:31–37.
- Leibelt C, Budowle B, Collins P, Daoudi Y, Moretti T, Nunn G, Reeder D, Roby R. 2003. Identification of a D8S1179 primer binding site mutation and the validation of a primer designed to recover null alleles. *Forensic Sci Int* 133:220–227.
- Levene H. 1949. On a matching problem arising in genetics. *Ann Math Stat* 20: 91–94.
- Li H, Cai X, Winograd-Cort ER, Wen B, Cheng X, Qin Z, Liu W, Liu Y, Pan S, Qian J, Tan C, Jin L. 2007. Mitochondrial DNA diversity and population differentiation in southern East Asia. *Am J Phys Anthropol* 134:481–488.
- Li H, Wen B, Chen S, Su B, Pramoongjago P, Liu Y, Pan S, Qin Z, Liu W, Cheng X, Yang N, Li X, Tran D, Lu D, Hsu M, Dekar M, Marzuki S, Tan C, Jin L. 2008. Paternal genetic affinity between western Austronesians and Daic populations. *BMC Evol Biol* 8:146.
- Lin M, Broadberry RE. 1998. Immunohematology in Taiwan. *Transfus Med Rev* 12:56–72.
- Lin M, Chu CC, Broadberry RE, Yu LC, Loo JH, Trejaut JA. 2005. Genetic diversity of Taiwan's indigenous peoples: possible relationship with insular Southeast Asia. In: Sagart L, Blench R, and A Sanchez-Mazas, editors. *The peopling of*

- East Asia: putting together archaeology, linguistics and genetics. London and New York: RoutledgeCurzon. p 230–247.
- Lin M, Chu CC, Lee HL, Chang SL, Ohashi J, Tokunaga K, Akaza T, Fuji T. 2000. Heterogeneity of Taiwan's indigenous population: possible relation to prehistoric Mongoloid dispersals. *Tissue Antigens* 55:1–9.
- Liu C, Liu C, Wang H. 2006a. STR data for the 15 loci from three minority populations in Guangxi municipality in South China. *Forensic Sci Int* 162:49–52.
- Liu C, Liu C, Wang H. 2006b. Genetic diversity at 15 STR loci in two tribal populations in Southern China. *Forensic Sci Int* 162:28–32.
- Long JC, Williams RC, McAuley JE, Medis R, Partel R, Tregellas WM, South SF, Rea AE, McCormick SB, Iwaniec U. 1991. Genetic variation in Arizona Mexican Americans: estimation and interpretation of admixture proportions. *Am J Phys Anthropol* 84:141–157.
- Loo J, Trejaut J, Yen J, Chen Z, Lee C, Lin M. 2011. Genetic affinities between the Yami tribe people of Orchid Island and the Philippine islanders of the Batanes Archipelago. *BMC Genet* 12:21.
- Lum JK. 1998. Central and Eastern Micronesia: genetics, the overnight voyage, and linguistic divergence. *Man Cult Oceania* 14:69–80.
- Maruyama S, Minaguchi K, Takezaki N, Nambiar P. 2008. Population data on 15 STR loci using AmpF/STR Identifier kit in a Malay population living in and around Kuala Lumpur, Malaysia. *Legal Med* 10:160–162.
- Melton T, Peterson R, Redd AJ, Saha N, Sofro AS, Martinson J, Stoneking M. 1995. Evolutionary history of the CoII/tRNALys intergenic 9 base pair deletion in human mitochondrial DNAs from the Pacific. *Mol Biol Evol* 12:60–615.
- Mirabal S, Herrera KJ, Gayden T, Regueiro M, Underhill PA. 2012. Increased Y-chromosome resolution of haplogroup O suggests genetic ties between the Ami aborigines of Taiwan and the Polynesian Islands of Samoa and Tonga. *Gene* 492:339–348.
- Novick GE, Novick CC, Yunis J, Yunis E, Martinez K, Duncan GG, Troup GM, Deininger PL, Stoneking M, Batzer MA, Herrera RJ. 1995. Polymorphic human specific *Alu* insertions as markers for human identification. *Electrophoresis* 16:1596–1601.
- Oppenheimer SJ, Richards M. 2001. Polynesian origins. Slow boat to Melanesia? *Nature* 410:166–167.
- Ota T. 1993. DISPAN: genetic distance and phylogenetic analysis. University Park, PA: Institute of Molecular Evolutionary Genetics, Pennsylvania State University.
- Perez-Miranda AM, Alfonso-Sanchez MA, Pena JA, Herrera RJ. 2006. Qatari DNA variation at a crossroad of human migrations. *Hum Hered* 61:67–79.
- Perez-Miranda AM, Alfonso-Sanchez MA, Peña JA, Pancorbo MM de, Herrera RJ. 2005. Genetic polymorphisms at 13 STR loci in autochthonous Basques from Alava province (Spain). *Legal Med* 7:58–61.
- Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945–959.
- Razafindrazaka H, Ricaut FX, Cox MP, Mormina M, Dugoujon JM, Randriamarolaza LP, Guitard E, Tonasso L, Ludes B, Crubézy E. 2010. Complete mitochondrial DNA sequences provide new insights into the Polynesian motif and the people of Madagascar. *Eur J Hum Genet* 18:575–581.
- Raymond M, Rousset F. 1995. An exact test for population differentiation. *Evolution* 149:1280–1283.
- Regueiro M, Carril JC, Pontes ML, Pinheiro MF, Luis JR, Caeiro B. 2004. Allele distribution of 15 PCR-based loci in the Rwanda Tutsi population by multiplex amplification and capillary electrophoresis. *Forensic Sci Int* 143:61–63.
- Regueiro M, Mirabal S, Lacau H, Caeiro JL, Garcia-Bertrand R, Herrera RJ. 2008. Austronesian genetic signature in East African Madagascar and Polynesia. *J Hum Genet* 53:106–120.
- Rerkamnuaychoke B, Rinthachai T, Shotivaranon J, Jomsawat U, Siriboonpiputtana T, Chaiatchanarat K, Pasomsub E, Chantratita W. 2006. Thai population data on 15 tetrameric STR loci—D8S1179, D21S11, D7S820, CSF1PO, D3S1358, TH01, D13S317, D16S539, D2S1338, D19S433, vWA, TPOX, D18S51, D5S818 and FGA. *Forensic Sci Int* 158:234–237.
- Rolett VB, Tsang CH. 2001. The Austronesian homeland: new evidence for early voyaging. In: *International Symposium on Austronesian cultures: Issues Relating to Taiwan* (Archaeology). Dec 8–11. Taipei (Taiwan): Academia Sinica. p 1–2.
- Rowold DJ, Herrera RJ. 2003. Inferring recent phylogenies using forensic STR technology. *Forensic Sci Int* 133:260–265.
- Ruhlen M. 1994. The origin of language: tracing the evolution of the mother tongue. New York: Wiley. p170–180.
- Sagart L. 2004. Sino-Tibetan-Austronesian: an updated and improved argument. In: Sagart L, Blench R, Sanchez-Mazas A, editors. *The peopling of East Asia: putting together archaeology, linguistics and genetics*. London: RoutledgeCurzon. p 161–176.
- Sagart L. 2005. The higher phylogeny of Austronesian and the position of Tai-Kadai. *Oceanic Linguist* 43:411–444.
- Schneider S, Roessli D, Excoffier L. 2000. Arlequin v.2000: a software for population genetics data analysis. Geneva: Genetics and Biometry Laboratory, University of Geneva.
- Shepard EM, Chow RA, Suafo E, Addison D, Perez-Miranda PE, Garcia-Bertrand RL, Herrera RJ. 2005. Autosomal STR variation in five Austronesian populations. *Hum Biol* 77:825–851.
- Seah LH, Jeevan NH, Othman MI, Jaya P, Ooi YS, Wang PR, Kee SS. 2003. STR data for the AmpFISTR Identifier loci in three ethnic groups (Malay, Chinese, Indian) of the Malaysian population. *Forensic Sci Int* 138:134–137.
- Shepard EM, Herrera RJ. 2006. Genetic encapsulation among Near Eastern populations. *J Hum Genet* 51:467–476.
- Soares P, Rito T, Trejaut J, Mormina M, Hill C, Tinkler-Hundal E, Braid M, Clarke DJ, Loo JH, Thomson N, Derham T, Donohue M, Macaulay V, Lin M, Oppenheimer S, Richards MB. 2011. Ancient voyaging and Polynesian origins. *Am J Hum Gen* 88:239–247.
- Souto L, Alves C, Gusmao L, Ferreira E, Amorim A, Corte-Real F, Vieira DN. 2005. Population data on 15 autosomal STRs in a sample from East Timor. *Forensic Sci Int* 155:77–80.
- Sturrock K, Rocha J. 2000. A multidimensional scaling stress evaluation table. *Field Methods* 12:49–60.
- Suadi Z, Siew LC, Tie R, Hui WB Adam A, Thiew SH, Boon LK. 2007. STR data for the AmpFISTR Identifier loci from the three main ethnic indigenous population groups (Iban, Bidayuh) and Melanau) in Sarawak, Malaysia. *J Forensic Sci* 52:231–234.
- 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.
- Tabbada KA, Trejaut J, Loo JH, Chem YM, Lin M, Mirazon-Lahr M, Kivisild T, De Ungria MC. 2010. Philippine mitochondrial DNA diversity: a population viaduct between Taiwan and Indonesia? *Mol Biol Evol* 27:21–31.
- Tang J, Zhang J, Jiang F, Yu X. 2009. Genetic analyzing of 15 STR loci in a Han population of Jinan (northern China). *Legal Med* 11:144–146.
- Tereba A. 1999. Tools for analysis of population statistics. *Profiles in DNA* 2:14–16.
- Terrel JE, Hunt TL, Gosden C. 1997. The dimensions of social life in the Pacific. *Curr Anthropol* 38:155–195.
- Tofanelli S, Bertoni S, Castri L, Luiselli D, Donati G, Paoli G. 2009. On the origins and admixture of Malagasy: new evidence from high-resolution analyses of paternal and maternal lineages. *Mol Biol Evol* 26:2109–2124.
- Trejaut JA, Kivisild T, Loo JH, Lee CL, He CL, Hsu CJ, Li ZY, Lin M. 2005. Traces of archaic mitochondrial lineages in Austronesian-speaking Formosan populations. *PLoS Biol* 3:e247.
- Tsang CH. 2002. The prehistory of Taiwan: a brief introduction. In: *Seventeenth Congress of the Indo-Pacific Prehistory Association* Sep 9–15. Taipei (Taiwan): Academia Sinica. p 47–52.
- Tsang CH. 2005. On the origin of the Yami People of Lanyu as viewed from archaeological data. *J Austronesian Stud* 1:135–151.

- Wang C, Chen D, Chen C, Lu S, Sun C. 2003. STR data for the AmpFlSTR SGM Plus and Profiler loci from Taiwan. *Forensic Sci Int* 138:119–122.
- Wu F, Ho C, Pu C, Hu K, Liu DH. 2009. Genetic polymorphisms of 17 Y-chromosomal short tandem repeat loci in Atayal population of Taiwan. *Croat Med J* 50:313–320.
- Wu Y, Zhang X, Zhou Y, Chen Z, Wang X. 2008. Genetic polymorphisms of 15 STR loci in Chinese Han population living in Xi'an city of Shaanxi province. *Forensic Sci Int: Gen* 2:e15–e18.
- Xu S, Pugach I, Stoneking M, Kayser M, Jin L, Pan Asian SNP Consortium. 2012. Genetic dating indicates that the Asian-Papuan admixture through Eastern Indonesia corresponds to the Austronesian expansion. *PNAS* 109:4574–4579.
- Xu L, Hu S, Feng G. 2009. STR polymorphisms of the Henan population and investigation of the Central Plains Han origin of Chaoshanese. *Biochem Genet* 47:569–581.
- Yan S, Wang CC, Li H, Li SL, Jin L, Genographic Consortium. 2011. An updated tree of Y haplogroup O and revised phylogenetic positions of mutations P164 and PK4. *Eur J Hum Genet* 19:1013–1015.
- Zhang GZ. 1987. *Archaeology on the coast of Southeast China and origin of Austronesian*. Southern ethnic archaeology 1. Chengdu: Sichuan University Press.