Philippine Mitochondrial DNA Diversity: A Populated Viaduct between Taiwan and Indonesia?

Kristina A. Tabbada,[1,2] Jean Trejaut,[3] Jun-Hun Loo,[3] Yao-Ming Chen,[4] Marie Lin,[3] Marta Mirazón-Lahr,[2] Toomas Kivisild,[2] and Maria Corazon A. De Ungria*[,1]

[1]DNA Analysis Laboratory, Natural Sciences Research Institute, University of the Philippines Diliman, Quezon City, Philippines

[2]Leverhulme Centre for Human Evolutionary Studies, University of Cambridge, Cambridge, United Kingdom

[3]Transfusion Medicine Research Laboratory, Mackay Memorial Hospital, Taipei, Taiwan

[4]Department of Health, Tainan Hospital, Tainan City, Taiwan

***Corresponding author:** E-mail: mariadeungria@gmail.com.

Associate editor: Lisa Matisoo-Smith

Abstract

Relatively little is known about the genetic diversity of the Philippine population, and this is an important gap in our understanding of Southeast Asian and Oceanic prehistory. Here we describe mitochondrial DNA (mtDNA) variation in 423 Philippine samples and analyze them in the context of the genetic diversity of other Southeast Asian populations. The majority of Philippine mtDNA types are shared with Taiwanese aboriginal groups and belong to haplogroups of postglacial and pre-Neolithic origin that have previously been identified in East Asian and Island Southeast Asian populations. Analysis of hypervariable segment I sequence variation within individual mtDNA haplogroups indicates a general decrease in the diversity of the most frequent types (B4a1a, E1a1a, and M7c3c) from the Taiwanese aborigines to the Philippines and Sulawesi, although calculated standard error measures overlap for these populations. This finding, together with the geographical distribution of ancestral and derived haplotypes of the B4a1a subclade including the Polynesian Motif, is consistent with southward dispersal of these lineages "Out of Taiwan" via the Philippines to Near Oceania and Polynesia. In addition to the mtDNA components shared with Taiwanese aborigines, complete sequence analyses revealed a minority of lineages in the Philippines that share their origins—possibly dating back to the Paleolithic—with haplogroups from Indonesia and New Guinea. Other rare lineages in the Philippines have no closely related types yet identified elsewhere.

Key words: mitochondrial DNA, phylogeography, Philippines, Southeast Asia, Austronesian.

Introduction

The people of Island Southeast Asia (ISEA) and Oceania are remarkable for the contrasting patterns of diversity seen in their languages, cultures, and physical appearance. Their origins have been the subject of scientific inquiry and discussion for over two centuries, and the study of the region's prehistory has played a key role in the broader debates over the first dispersals of humans out of Africa as well as the influence of agriculture and the role of demographic expansions in recent human evolution.

It is generally accepted that anatomically modern humans entered ISEA and reached as far as New Guinea and Australia approximately 50,000 years before the present (YBP) (Détroit et al. 2004; O'Connell and Allen 2004; Barker et al. 2007). However, there is little consensus over subsequent events in the region's prehistory or the contribution of these early occupants to the present diversity observed in ISEA and Oceania.

Various models have been proposed to explain population histories in ISEA and Oceania. These models differ from each other in terms of the postulated center of dispersal, the direction of movement, and the timeframe of events. The "Out of Taiwan" group of models traces the origin of a significant part of the gene pool of Austronesian-speaking populations in Oceania to a dispersal event from mainland Southeast Asia over Taiwan. One variant of the Out of Taiwan model, popularly called the "Express Train to Polynesia" hypothesis, suggests an expansion of agriculturalist populations out of southern China into Taiwan approximately 6,000 BP, followed by movement south through the Philippines and Indonesia (ca. 4,000 BP), and then eastwards into the Pacific (Diamond 1988; Bellwood 1989). Under this model, population dispersal occurs in a north to south direction through the Philippines within a very short time, presumably with only minor admixture with existing Philippine populations along the way. Archaeological and linguistic data have been interpreted to support this theory of Southeast Asian and Oceanic prehistory (Hung et al. 2007; Gray et al. 2009). Other authors have proposed modified models; one variant is the "Slow Boat to Polynesia," in which population movement follows the same path as the "Express Train," but at a slower pace, possibly predating the Neolithic (Trejaut et al. 2005) and with more admixture along the way (Kayser et al. 2000; Mona et al. 2009). All Out of Taiwan models assume the existence of a common genetic component of continental Southeast Asian origin among the Austronesian-speaking populations of Near Oceania and Polynesia. However, in the case of Y chromosome

lineages, it has been shown that the Southeast Asian genetic contribution has become sequentially diluted during the process of dispersal (Kayser et al. 2000, 2006).

An alternative set of hypotheses identifies the primary center of Southeast Asian population dispersal to be within the region of the southern Philippines and Sulawesi (Meacham 1988; Solheim 1988; Oppenheimer and Richards 2001). Under this hypothesis, the path of movement of dispersing populations is northwards with respect to the Philippines and Taiwan. Thus, Polynesian populations are not expected to carry a significant proportion of genetic components originating recently from Taiwan or continental Southeast Asia. One variant of this model proposes that the motivating force behind this movement of populations was the rise of sea levels at the end of the last glacial period (ca. 10,000 YBP). As a consequence, the period of *in situ* development of ISEA populations was longer than the time proposed in the Express Train/Out of Taiwan model (Richards et al. 1998; Oppenheimer and Richards 2001).

Although mitochondrial DNA (mtDNA) and Y chromosome analyses have indicated differing degrees of genetic contributions by mainland Southeast Asians to the Austronesian-speaking populations of remote Oceania (Hurles et al. 2003), recent studies covering hundreds of autosomal length variants (STRs and indels) have shown that Austronesian-speaking populations of Polynesia and Micronesia share a substantial portion of their gene pool with Taiwanese Aboriginal or East Asian populations, with no more than a 20% (or smaller) genetic contribution from Near Oceania (Friedlaender et al. 2008; Kayser et al. 2008). The number of nuclear sites sampled in these studies makes the inferences drawn from the data extremely robust. Therefore, the hypothesis that all Austronesian speakers share a substantial East Asian genetic component is well supported by these new lines of evidence. However, the autosomal variant frequency patterns do not inform us about the timing of the Austronesian dispersal or about local population histories in detail. It is therefore interesting to reexamine the mtDNA phylogenetic signals in ISEA in the light of these estimates of average genomic regional contribution.

Patterns of mtDNA variation have been extensively studied in some Southeast Asian and Oceanic populations. Ancient maternal lineages, possibly dating from the first incursion of *Homo sapiens* into the region, are retained at high frequencies among isolated groups in Peninsular Malaysia and persist at low frequencies in the general population of Southeast Asians (Macaulay et al. 2005; Hill et al. 2006, 2007). Mona et al. (2009) proposed that haplogroups P and Q—which were previously thought to have originated in Near Oceania and Australia (Forster et al. 2001; Friedlaender et al. 2005, 2007)—may instead have arisen in Eastern Indonesia during Late Pleistocene.

Many studies on maternal lineages in the region have focused on the "Polynesian Motif," a set of mutations at four nucleotide positions (np) within the mtDNA hypervariable segment I (HVS-I), which—together with base changes in the coding region recently identified by Trejaut et al. (2005)—characterize a unique subclade of haplogroup B4a1a (Pierson et al. 2006). The Polynesian Motif and related lineages have been linked to population expansion in Southeast Asia and the Pacific (Melton et al. 1995; Redd et al. 1995; Sykes et al. 1995). However, no detailed phylogenetic study on B4a1a or other mtDNA lineages has been performed in the Philippines, which lies on the most plausible migration route from Taiwan to Near Oceania. Here, we report mtDNA variation in the Philippine population and describe the distribution and diversity of Filipino maternal (mtDNA) lineages at the complete sequence level in the context of surrounding Southeast Asian populations.

Materials and Methods

Population Sampling

Two different sample sets were used in this study. The first consisted of FTA-bound DNA sourced from blood samples of 100 individuals from urban populations representing the Philippines's three major island groups (fig. 1)—the National Capital Region in Luzon ($N = 47$), Cebu City in the Visayas ($N = 26$), and Zamboanga City located in Mindanao ($N = 27$). These samples were selected randomly from the existing genetic bank of the University of the Philippines, Natural Sciences Research Institute, DNA Analysis Laboratory (DAL), for inclusion in the study. The second DNA sample set was extracted from the blood of 323 volunteers of Philippine descent residing in Taiwan and 74 Hakka individuals. Samples were collected at the Transfusion Medicine Research Laboratory of the Mackay Memorial Hospital (MMH) in Taipei and extracted according to previously described procedures (Trejaut et al. 2005). Philippine participants were assigned into geographical categories (Luzon, Visayas, and Mindanao) based on their birthplace, as determined in an interview. DNA samples of individuals whose origins could not be adequately determined were collated into a "Geographically Undefined" group.

Sequence Analysis

Four hundred twenty-three samples of Philippine origin were sequenced for HVS-I [the minimum sequence length captured np 16,032—16,365] using Big Dye Terminator chemistry. Additional sequencing was performed on 169 samples covering the HVS-II (np 73–340 minimum). Whole mtDNA genome sequence was determined for a selection of 31 Philippine samples in Taipei following methods described by Torroni et al. (2001).

Haplogroup affiliation of samples was confirmed by additional analyses (see supplementary table S1, Supplementary Material online). All samples from the DAL were tested by at least one PCR–RFLP reaction to confirm haplogroup identity; this was done at the Leverhulme Centre for Human Evolutionary Studies, Cambridge, based on techniques first described by Torroni et al. (1992, 1993). Samples analyzed at the MMH were sequenced at np 9,820–10,890; PCR-RFLP typing was performed on selected samples (109 reactions). Sequence data at np 7,990–8,990

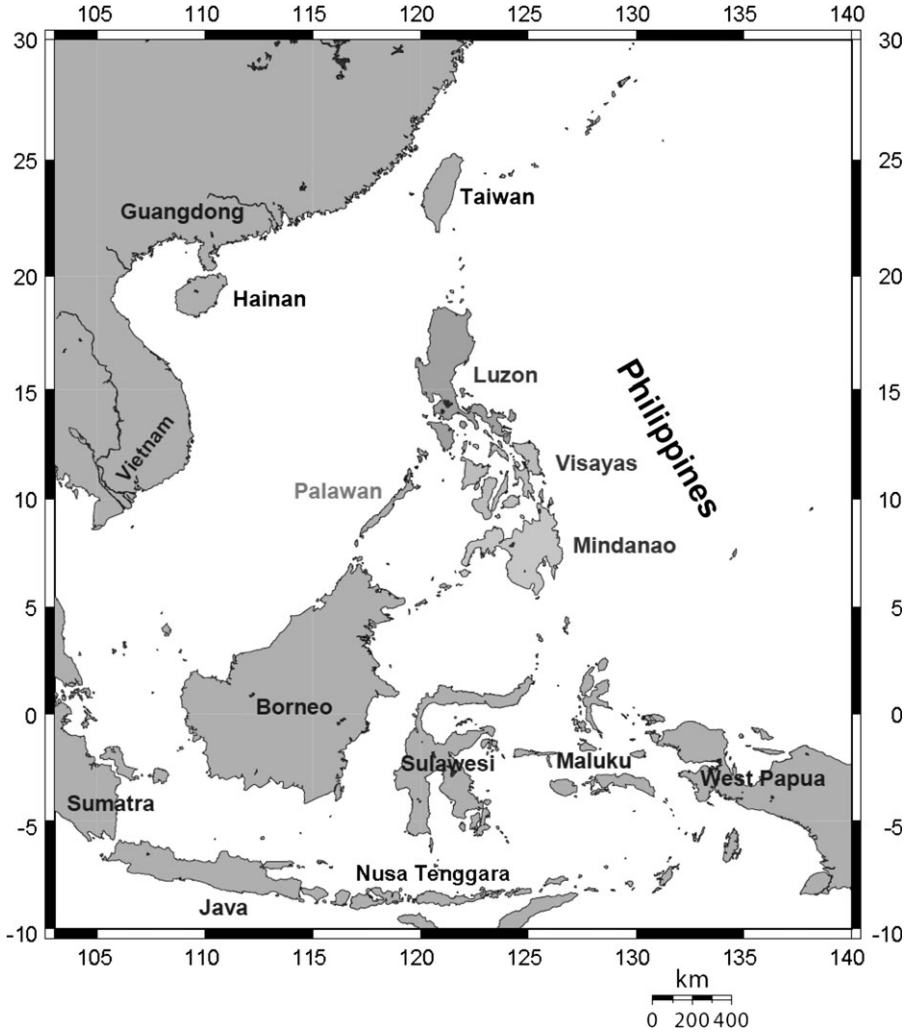

Fig. 1. Map of sampling sites in the Philippines, Taiwan and surrounding populations in Southeast Asia. Map created using The Generic Mapping Tools (Wessel and Smith 1995).

($N = 97$) and np 14,000–15,000 ($N = 82$) was also collected for the MMH samples.

Data Analysis

Mitochondrial DNA haplogroups are referred to following existing mtDNA haplogroup nomenclature (van Oven and Kayser 2009), *mtDNA tree Build 6* (28 September 2009) available at http://www.phylotree.org/.

Mitochondrial DNA HVS-I sequences ($N = 423$) were compared with data from nine Austronesian-speaking Taiwanese aboriginal groups ($N = 640$) (Trejaut et al. 2005) and Indonesian samples from Sulawesi ($N = 218$) (Hill et al. 2007). Sequence length was standardized to a common stretch of 275 bp, spanning np 16,090—16,365. Sequences were aligned using DNA Alignment 1.1.3.0 (Fluxus Technology, Ltd).

Haplogroup frequencies for East Asian and Southeast Asian groups were obtained for comparison from 102 Han Chinese from Liaoning (Yao et al. 2002; Wen et al. 2004) and 99 from Guangdong (Kivisild et al. 2002; Yao et al. 2002), Indonesian samples from Borneo ($N = 157$), Java ($N = 99$), Sumatra ($N = 180$), Maluku ($N = 66$), Austronesian groups in Nusa Tenggara ($N = 152$) and a non-Austronesian group from Nusa Tenggara ($N = 38$)(Cox 2003; Hill et al. 2007; Mona et al. 2009). Other groups added for comparison were Taiwanese Hakka population ($N = 74$) (Trejaut et al. forthcoming), the Hlai of Hainan ($N = 27$) (Li et al. 2007), Kinh of Vietnam ($N = 41$) (Li et al. 2007) and 52 urban Thais (Allard et al. 2004). Haplogroup frequency data from West Papuan ($N = 59$) and Polynesian ($N = 9$) populations was taken from Friedlaender et al. (2005).

An exact test of population differentiation for Philippine groups was performed using Arlequin 2.000 (Goudet et al. 1996; Schneider et al. 2000), as were the calculations of pairwise distances between populations (F_{ST}) and Mantel tests between F_{ST} matrices generated using all haplogroups and selected individual haplogroups to investigate the contribution of different lineages to genetic distances between populations. *P* values were calculated using 100,000 steps of a Markov chain and 3,000 dememorization steps were performed.

Table 1. MtDNA Haplogroup Frequencies (%) in Three Philippine Subpopulations.

	Luzon	Visayas	Mindanao	Geographically Undefined	Total
B4a	0.56	0.00	1.43	0.00	0.47
B4a1a	11.30	9.82	10.00	12.50	10.87
B4a1a1	0.00	0.00	1.43	1.56	0.47
B4b1	7.34	7.14	7.14	9.38	7.57
B4c1b	5.65	6.25	0.00	6.25	4.96
B5a	1.13	0.00	0.00	1.56	0.71
B5b	10.17	4.46	4.29	10.94	7.80
B7	0.56	0.00	0.00	0.00	0.24
D*	0.56	0.00	0.00	0.00	0.24
D5b	0.56	0.00	0.00	1.56	0.47
D6	1.13	0.89	2.86	1.56	1.42
E1	0.00	0.89	0.00	0.00	0.24
E1a1a	7.91	16.07	14.29	7.81	11.11
E1a2	0.00	0.00	0.00	1.56	0.24
E1b	2.82	0.00	1.43	0.00	1.42
E2*	1.13	4.46	5.71	3.13	3.07
E2b	0.56	0.00	0.00	0.00	0.24
F1a1a	0.56	0.00	0.00	0.00	0.24
F1a3	3.39	2.68	1.43	4.69	3.07
F1a4	3.39	6.25	2.86	4.69	4.26
F3b	2.82	0.89	2.86	0.00	1.89
F4b	1.13	0.00	0.00	0.00	0.47
M*	2.26	1.79	4.29	1.56	2.36
M21a	0.00	0.89	0.00	1.56	0.47
M21c1a	0.56	0.00	0.00	0.00	0.24
M17c1a	2.82	0.00	0.00	0.00	1.18
M71	0.00	0.89	1.43	0.00	0.47
M72	0.56	0.00	0.00	0.00	0.24
M73a	0.00	1.79	0.00	0.00	0.47
M7a	0.56	0.00	0.00	0.00	0.24
M7b	0.00	0.00	1.43	0.00	0.24
M7b1	1.69	1.79	0.00	0.00	1.18
M7b3	2.82	5.36	2.86	1.56	3.31
M7c	1.13	1.79	1.43	0.00	1.18
M7c3c	12.99	7.14	12.86	12.50	11.35
M8a	0.56	0.00	0.00	0.00	0.24
N*	0.00	0.89	0.00	0.00	0.24
N22	0.00	0.00	1.43	0.00	0.24
P8	1.13	0.89	0.00	1.56	0.95
P10	0.00	0.89	1.43	0.00	0.47
R24	0.56	8.04	2.86	1.56	3.07
R9c	4.52	5.36	5.71	4.69	4.96
R9e	0.00	0.00	1.43	1.56	0.47
Y2	4.52	2.68	7.14	6.25	4.73
Z4	0.56	0.00	0.00	0.00	0.24
Total	100.00	100.00	100.00	100.00	100.00
Sample size, *N*	177	112	70	64	423

Significance level was set at $\alpha = 0.05$. Multidimensional scaling (MDS) analysis was performed on distances between all populations using SPSS 13.0 (SPSS Inc., 1999–2004).

Reduced median networks (Bandelt et al. 1995) for each haplogroup were drawn for Taiwanese aboriginal, Philippine and Indonesian sequences using Network 4.2.0.1 (Fluxus Technology, Ltd). Ancestral and derived haplotypes were designated; the diversity of each haplogroup was calculated using the ρ statistic (Morral et al. 1994; Forster et al. 1996) with standard error σ (Saillard et al. 2000). Values of ρ in mutational time were converted to years using the estimated rate of one mutation per 20,180 years in the stretch between np 16,090—16,365 (Forster et al. 1996).

Results

Classification of Philippine mtDNA Sequences

The combined approach of control region sequencing and coding region SNP typing (supplementary table S2, Supplementary Material online) showed that 94% of the Philippine samples fall into mtDNA haplogroups previously identified in East Asian and Southeast Asian populations. These include subclades of haplogroups B4, B5, D, E, F, R9, M7, and Y, as well as the recently identified M17 (formerly M45) (Kivisild et al. 2002; Trejaut et al. 2005; Kong et al. 2006; Hill et al. 2007; van Oven and Kayser 2009). The most common types were B4a1a, E1a1a, and M7c3c (table 1). The remaining 6% of the samples were previously uncharacterized types, some of which are grouped into newly defined clades.

Comparison of Philippine Groups with Southeast Asian and East Asian Populations

Comparison of the mtDNA haplogroup frequency distributions in the three major island groups of the Philippines showed similar haplogroup profiles. Southeast Asian populations cluster closely together in an MDS plot including groups from Near Oceania and Polynesia (fig. 2*A*). However, an MDS plot of Asians and Southeast Asians reveals genetic differentiation between these groups. The Philippines, Taiwanese Aborigines, and Sulawesi cannot be clearly separated in the first dimension of the MDS plot; these groups can only be distinguished in the second dimension (fig. 2*B*). However, the MDS plot does enable us to differentiate between the latter populations and other Island Southeast Asians (Maluku, Java, Borneo, Sumatra, and Austronesian-speaking groups in the Nusa Tenggaras). Han Chinese populations from Taiwan (Hakka) and southern China group closely together, whereas mainland Southeast Asian populations from Vietnam and Thailand are interspersed with other East Asian groups. Genetic distances between these populations are most closely correlated with the distributions of haplogroups B5b, M7b3, and M7c3c ($r \geq 0.25$); less so with other frequent haplogroups such as E1a1a and B4a1a.

Estimation of Coalescent Times in the Taiwanese Aboriginal, Philippine, and Sulawesi populations

Analysis of the distribution and diversity of the most common haplogroups shared between populations showed that there is a general trend of decreasing HVS-I diversity from Taiwan to the Philippines and then Sulawesi (table 2). The majority of the mtDNA haplogroups (M7b3, E1a1a, B4a1a, B4b1, M7c3c, and B4c1b) found in both Taiwan and the Philippines have greater values of ρ and earlier coalescent dates among the aboriginal Taiwanese. Haplogroups that are found at substantial frequencies in both the Philippines and Sulawesi—such as M7c3c and F1a4—have a more recent coalescent date or very limited diversity in the Indonesian population (table 2). In the Philippines, mtDNA haplogroups

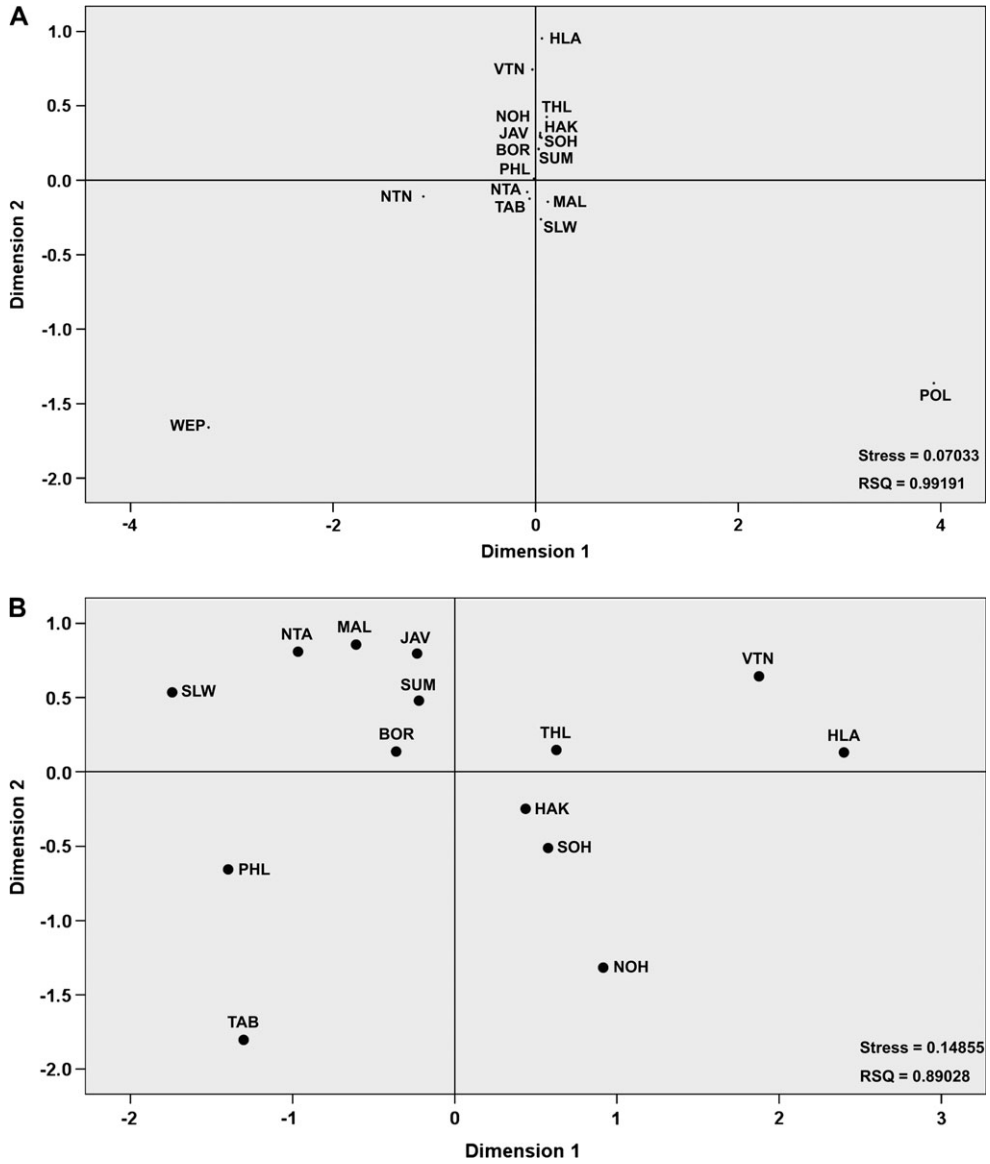

Fig. 2. Plot of first two dimensions produced by MDS analysis of mtDNA haplogroup frequencies in selected East, Southeast Asian, Near Oceanic, and Polynesian populations, including Borneo (BOR), Hakka (HAK), Hlai (HLA), Java (JAV), Maluku (MAL), Nusa Tenggara—Austronesian (NTA), Northern Han (NOH), Philippines (PHL), Southern Han (SOH), Sulawesi (SLW), Sumatra (SUM), Taiwanese Aborigines (TAB), Thailand (THL) and Vietnam (VTN). (*A*) Including Nusa Tenggara—Non-Austronesian (NTN), Polynesian (POL), and West Papuan (WEP); Borneo and Sumatra overlap. (*B*) East Asian and Southeast Asian groups only.

with estimated coalescent dates before the Holocene (>10,000 YBP) make up 30% of the samples, whereas 38% of lineages coalesce to the early Holocene (10,000—5,000 YBP), and 13% within the last 5,000 years. The remaining 19% of the haplogroups for which coalescent times were not determined each occur at low frequency.

Identification of Common mtDNA Haplogroups in the Philippines, Taiwanese Aborigines, and ISEA

Haplogroup B4a1a occurs at similar frequencies in Taiwan, the Philippines, and Sulawesi (fig. 3 and supplementary table S3, Supplementary Material online) and shows earlier coalescent dates among Taiwanese aboriginals (9,500 YBP) than in the Philippine population (7,900 YBP). Its diversity in Sulawesi remains to be determined, as the HVS-I data do not allow for its differentiation from other B4a subclades (Hill et al. 2007). Although haplogroup B4a1a1, the immediate precursor to the Polynesian Motif (Pierson et al. 2006), was not identified among 640 Taiwanese aboriginals (Trejaut et al. 2005), it is notable that in the Philippine population both the ancestral and the derived allele at np 14,022 can be observed (table 1 and fig. 4*B*).

Aside from B4a1a, several haplogroups are shared between the Philippines, Taiwan, and ISEA (fig. 3). Of these, M7c3c is observed at the highest frequency throughout ISEA and its distribution shows a comparatively high degree of correlation ($r = 0.32$) to the genetic distances between populations. Although the M7c3 clade shows several derived variants in China, subclade M7c3c is rare in the

Table 2. Estimated Ages of Major mtDNA Haplogroups in the Philippines, Taiwan, and Sulawesi Based on ρ.

	Taiwanese Aborigines (Trejaut et al. 2005)		Philippines		Sulawesi (Hill et al. 2007)	
	ρ ± σ	ky ± SE	ρ ± σ	ky ± SE	ρ ± σ	ky ± SE
R24	NO	NO	0.15 ± 0.11	3.1 ± 2.2	NO	NO
M7b3	0.51 ± 0.29	10.2 ± 5.8	0.21 ± 0.12	4.3 ± 2.5	NC	NC
B5b1	NC	NC	0.23 ± 0.12	4.7 ± 2.4	0.50 ± 0.37	10.1 ± 7.5
Y2	NC	NC	0.25 ± 0.11	5.0 ± 2.3	NC	NC
F1a3	NC	NC	0.31 ± 0.19	6.2 ± 3.8	NC	NC
E1a1a	0.59 ± 0.26	11.9 ± 5.3	0.36 ± 0.22	7.3 ± 4.5	0.24 ± 0.10	4.9 ± 2.1
B4a1a	0.47 ± 0.23	9.5 ± 4.6	0.39 ± 0.12	7.9 ± 2.4	NC	NC
B4b1	0.53 ± 0.49	10.7 ± 9.9	0.44 ± 0.16	8.8 ± 3.2	NC	NC
F1a4	NC	NC	0.50 ± 0.30	10.1 ± 6.0	NC	NC
M7c3c	0.70 ± 0.39	14.0 ± 7.9	0.56 ± 0.18	11.4 ± 3.6	0.22 ± 0.14	4.4 ± 2.9
R9c	NC	NC	0.57 ± 0.20	11.5 ± 4.1	NC	NC
E1b1	NO	NO	0.67 ± 0.41	13.4 ± 8.2	0.33 ± 0.25	6.7 ± 5.0
B4c1b	1.46 ± 0.99	29.5 ± 20.1	0.81 ± 0.72	16.3 ± 14.5	1.12 ± 0.89	22.7 ± 18.0
M7b1	1.00 ± 0.66	20.2 ± 16.6	1.20 ± 0.57	24.2 ± 11.4	NC	NC
F3b2	0.80 ± 0.52	16.1 ± 10.6	1.75 ± 0.79	35.3 ± 16.0	NO	NO

NOTE.—NO, haplogroup not observed; NC, ρ not calculated due to insufficient variation and sample size.

mainland. M7c3c is found at higher frequencies in ISEA and has therefore been identified as a potentially informative marker of Southeast Asian prehistory (Trejaut et al. 2005; Hill et al. 2007). An examination of the M7c3 subclade distribution shows that the Taiwanese aboriginals have both M7c3a and M7c3c types, whereas in the Philippines, Sulawesi, and other Southeast Asian populations, only M7c3c is observed. M7c3c makes up an increasing percentage of the population from the Taiwanese aboriginals to the Philippines (where it is the second most frequent haplogroup) and Sulawesi. Haplogroup M7c3c has an older estimated age in Taiwan (ca. 14,000 years) than in the Philippines (11,400 YBP) and Sulawesi (4,400 YBP). However, the value of ρ in the aboriginal Taiwanese population may be inflated by the presence of a variant that is confined to one subpopulation (np 16,168–16,266 in the Yami).

Haplogroup F1a4 has a lower frequency than M7c3c, but a similar pattern of distribution. This haplogroup is rare in Taiwanese aboriginals (<1%) but appears at greater frequencies in the Philippines and Sulawesi (4–5%). Despite its slightly higher frequency in Sulawesi, the F1a4 lineage has a lower diversity in the latter population compared with the Philippines.

Other haplogroups shared by Philippine, Taiwanese aboriginal, and other ISEA populations include Y2, D5b, and M7b3. Y2 has a slightly higher frequency in the Philippines compared with surrounding groups. In contrast, haplogroup D5 (Kong et al. 2006) is found throughout Southeast Asia but is rare in the Philippines (table 1). One D5 subclade was previously detected in the Taiwanese Ami tribe (Trejaut et al. 2005), in Borneo, and Sulawesi, and labeled as a new clade D5d with an approximate 4 ky

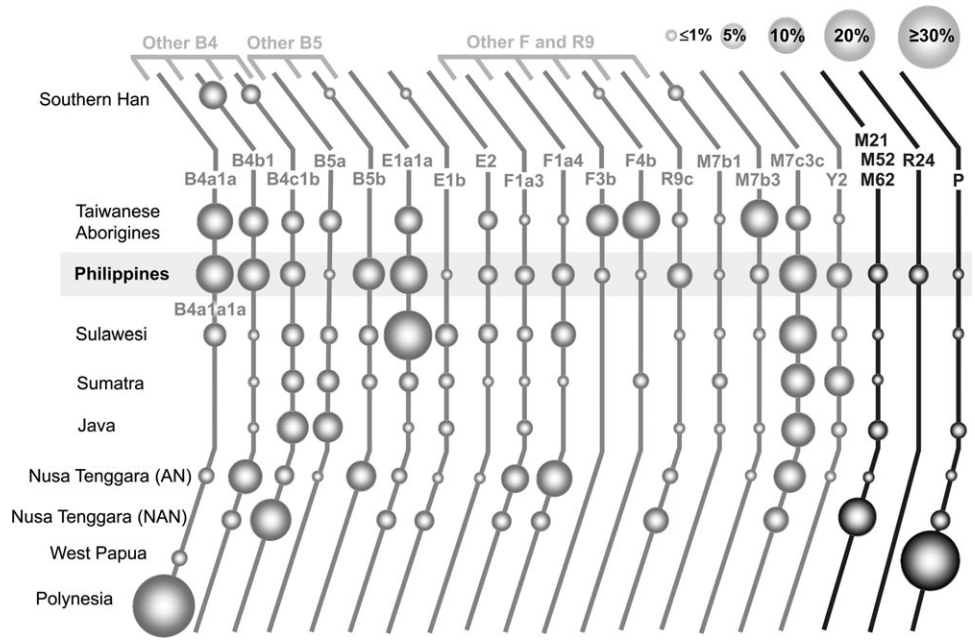

FIG. 3. Frequencies of major mtDNA haplogroups in East Asian, Southeast Asian, Papuan, and Polynesian groups.

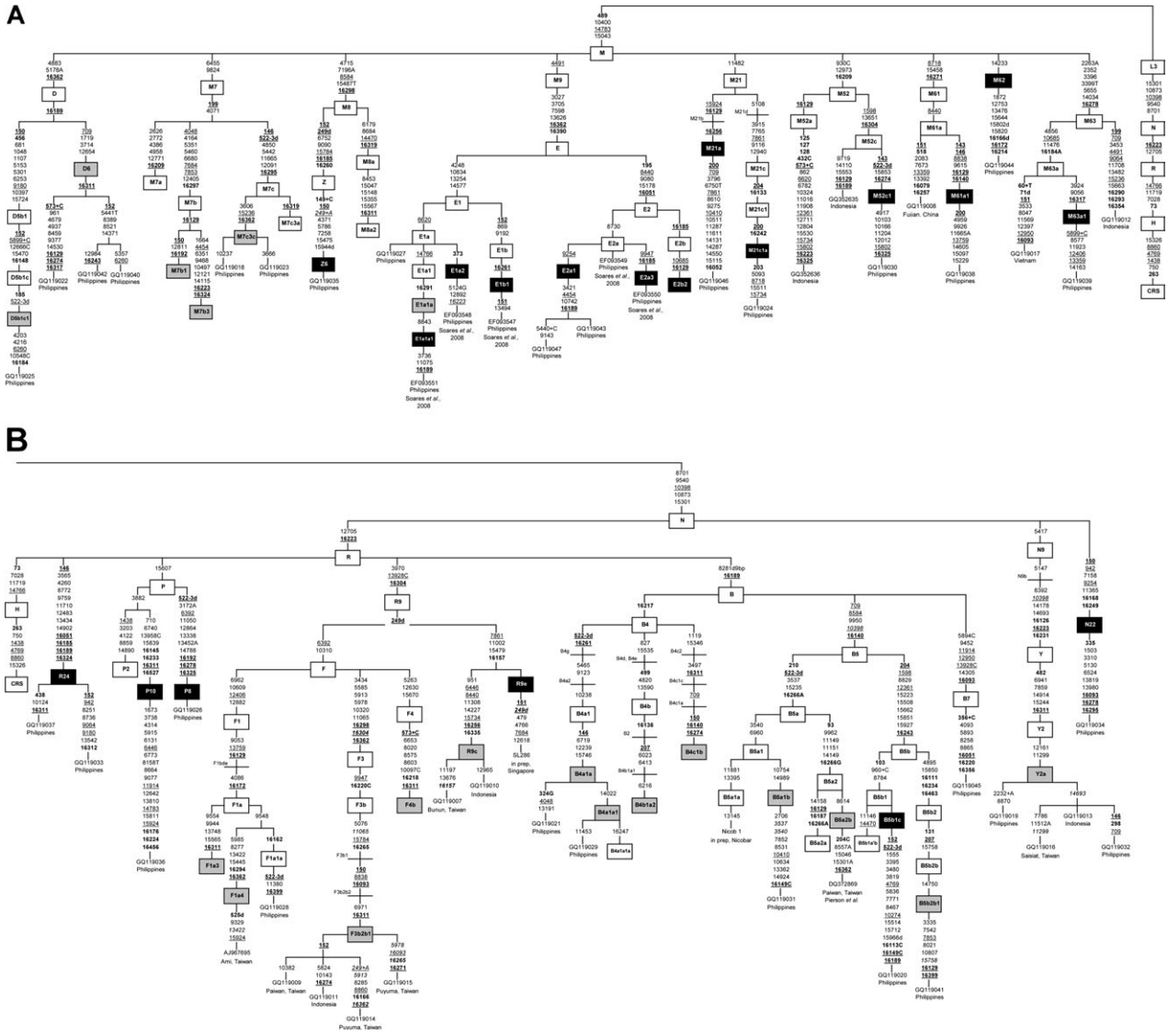

Fig. 4. Complete mtDNA sequences from the Philippines as well as other East and Southeast Asian populations. Haplogroup clades that are confined to the Philippines and Indonesia are in black. Haplogroups also found in Taiwan are in grey. Base changes in the control region are printed in bold, back mutations are in italics, and recurrent mutations are underlined. (A) Macrohaplogroup M. (B) Macrohaplogroup N.

coalescent date in ISEA (Hill et al. 2007). Using complete sequence information, we determined that the lineage named D5d is, in fact, a subclade of haplogroup D5b1, which we propose to label D5b1c (fig. 4*A*) according to the complete sequence-based nomenclature of Kong et al. (2006). Although haplogroup M7b is widely distributed in mainland Asia, its daughter clade M7b3 is virtually absent in China, with only a variant showing the motif 16,086–16,297–16,324 (presumably with a back mutation at np 16,129) observed once in northwestern China (Yao et al. 2004). A single sequence with the motif 16,129–16,297–16,324 has been found among 102 samples from Liaoning in the northeast (Wen et al. 2004). However, M7b3 is found at a relatively high frequency (11%) in the Taiwanese aboriginal population. The Taiwanese aboriginal population includes M7b3 haplotypes with and without a mutation at np 16,086. M7b3 is found at a low frequency and diversity in the Philippines and Sulawesi, and in these populations, as well as in Java and the Nusa Tenggaras (the only other Island Southeast Asian populations in which M7b3 has been found), the sequences belonging to this haplogroup all exhibit a base change at np 16,086 (Hill et al. 2007). The distribution of M7b3 is correlated ($r = 0.26$) with genetic distances between Asian and Southeast Asian groups. There is a decrease in diversity (as measured by ρ) within haplogroup M7b3 from Taiwan to the Philippines and Sulawesi (table 2). Other haplogroups that are common in Taiwan but are found at decreasing frequency along the proposed Out of Taiwan route are F3b and F4b (fig. 3).

Identification of Philippine mtDNA Haplogroups Shared with ISEA and Mainland Southeast Asia

Lineages that comprise a significant proportion ($\geq$5%) of the Philippine population and are generally shared with both Island and/or Mainland Southeast Asians include B4b1, B4c1b, B5b, E, and R9c. Among these mtDNA

lineages, haplogroup E is unusual in that it is virtually absent in mainland Asia (Soares et al. 2008). The E1a1a lineage, which is the most common subclade of E, shows a decrease in diversity and coalescent time from Taiwan (11,900 YBP) to the Philippines (7,300 YBP) to Sulawesi (4,900 YBP). In Sulawesi, a lower diversity within haplogroup E1a1a is observed despite this lineage making up a larger proportion (17.7%) of the population. In the Philippines, E1a lineages that do not belong to E1a1a are extremely rare: only one example is present in the current data set. Other E subclades, such as E1b, E2a, and E2b are found at low frequencies in the Philippines.

Identification of Novel Haplogroups in the Philippine Population

Novel mtDNA haplogroups observed in the Philippines are described in figure 4. Philippine samples with a HVS-I motif including a transition at np 16,209 belong to a newly identified subclade of Southeast Asian haplogroup M17 (Hill et al. 2006; van Oven and Kayser 2009). This subclade, M17c, was observed in five Philippine individuals as well as in Indonesian samples. In the Philippines, M17c1 has so far been found only in Luzon; however, samples with the M17c1 haplogroup HVS-I motif have also been reported in Nusa Tenggara (Mona et al. 2009). Variants of haplogroup M21 (M21a and M21c) are also observed in the Philippines at low frequencies (table 1). Although samples belonging to M21 have been identified in various populations of ISEA, the specific subclade motifs reported here appear to be confined to the Philippines. However, given their low frequency, substantially larger sample sizes would be needed to determine the phylogeographic signature of these lineages.

Three new M-haplogroups were detected at extremely low frequencies in the Philippine population and defined at complete sequence resolution. M71 is defined by a transition at np 16,271 as well as by coding region mutations. The two samples observed in the Philippines also exhibit a distinct set of control region mutations and therefore fall into subclade M71a1 (fig. 4A). Related (putative M71) samples have been observed in populations as disparate as Fujian and Java; however, aside from one sample exhibiting the M71a1 HVS-I motif reported in Bali by Hill et al. (2007), this subclade has not been detected outside the Philippines. Although haplogroup M72 is defined only by a single coding region transition at np 14,233, Philippine samples belonging to M72 also have a characteristic HVS-I motif, which has been observed in Mataram, Flores, and Pantar (Hill et al. 2007; Mona et al. 2009). Another haplogroup, M73a, is observed in the Philippines; the transversion at np 16,184 that characterizes this lineage has also been found in Vietnam (fig. 4A).

A novel mitochondrial lineage belonging to macrohaplogroup R has been observed in the Philippine population at a frequency of 3%. This haplogroup, called R24, appears to be unrelated to other haplogroups within the R macroclade (fig. 4B). In the Philippines, R24 is observed at its highest frequency in the Visayas, whereas matching HVS-I motifs have not yet been reported in neighboring Southeast Asian populations.

Haplogroup P is found at highest frequencies in New Guinea, Near Oceania, and Australia, whereas infrequent occurrences of P1 lineages, which carry a distinctive HVS-I transition at np 16,357, have been reported as far west as Sumba, Sulawesi, and the Malaysian Peninsula (Hill et al. 2007). Unexpectedly, we detected two previously uncharacterized subclades of haplogroup P among our Philippine samples and assigned them to new haplogroups P8 and P10 (fig. 4B). Control region sequences of two samples with the characteristic motif of P10 have been reported in previously published Philippine data sets (Tajima et al. 2004; Hill et al. 2007). However, as with P8, this lineage has not been detected outside the Philippines.

Discussion

Analysis of Philippine mtDNA lineages yields insights on shared prehistory with other populations in Southeast Asia. Clustering of the Philippine population with other Asian or Southeast Asian populations in the MDS plot (fig. 2) is supported by a high degree of haplogroup sharing with Taiwanese aboriginal groups and populations of Indonesia (Sulawesi and Austronesian groups in Nusa Tenggara); almost 80% of the mtDNA lineages in each population are also found in the other groups in the cluster (fig. 3, supplementary table S3, Supplementary Material online). Other populations in ISEA (Borneo, Java, Sumatra, and Maluku), which form a second cluster in the MDS plot, have a smaller percentage (57–70%) of mtDNA lineages in common with the Philippines (Cox 2003; Hill et al. 2007). The combined frequency of haplogroups found in both the Philippines and Mainland Asia is substantially lower (50% or less). Given the substantial haplogroup sharing across Island Southeast Asian populations, the Express Train variation of the Out of Taiwan hypothesis, implying fast coastal penetration of dispersing populations into the Philippines, is unlikely to explain the present day mtDNA variation.

The majority of mtDNA haplogroups observed in the Philippines, Taiwan, and Sulawesi show a general decrease in HVS-I diversity from Taiwan to the Philippines and Sulawesi, although the calculated standard errors show substantial overlap. This suggests a progressive loss of variation resulting from population dispersal from a source to sink areas. This pattern is consistent with the direction of movement proposed in the Out of Taiwan hypothesis. It should be noted that the estimated coalescent dates for many haplogroups (including those linked with the Out of Taiwan dispersal) are considerably older in Taiwan and the Philippines than would be expected under the Express Train model. However, several demographic phenomena in the population under study–for example, substructuring and drift among the Taiwanese aborigines–may affect the estimated coalescence dates (Cox 2008). Furthermore, the precise rate of mutation remains controversial given the uncertainties in calibration (Ho and

Larson 2006; Soares et al. 2009 and references therein). Therefore, greater importance should be placed on the relative values of ρ rather than point estimates of the dates.

The Polynesian Motif and its precursors have received considerable attention in studies of mtDNA types in Asian and Pacific populations, and the development of the motif is thought to be connected to the dispersal of Austronesian-speaking populations from Taiwan (Melton et al. 1995; Redd et al. 1995; Sykes et al. 1995). Haplogroup B4a1a is highly diverse in Taiwan, but the subclade (B4a1a1) characterized by a mutation at np 14,022 is absent there. The identification of haplogroup B4a1a1 in the Philippines may indicate a stage of development of the Polynesian Motif along the north to south pathway proposed in the general Out of Taiwan model for the Austronesian population expansion. This apparently completes a series of genetic links from Taiwan (where the B4a1a motif may have originated), through the Philippines (where the np 14,022 mutation might have evolved) and finally to Indonesia (where the full Polynesian Motif first occurs). However, the observation of a B4a1a1 sample in the Philippine population is not necessarily incompatible with models that argue for an extended development period for the Polynesian Motif in ISEA, if the proposed area of development of the motif is expanded to include the Philippines. Another alternative explanation is that the B4a1a1 lineages might have been brought to the Philippines by a back migration from Indonesia.

Whereas genetic studies have focused on lineages (like the Polynesian motif and M7c3c) that increase in frequency along the proposed path of population movement Out of Taiwan, those lineages (such as F3b, F4, and M7b3) that decrease in frequency along this route may also be informative. Hill et al. (2007) hypothesized that haplogroup M7b3 was a marker for early Holocene population expansion in Southeast Asia. However, a detailed examination of the distribution of M7b3 types may support alternative explanations. Whereas more than one M7b3 clade is present in the Taiwanese aborigines, only a single variant of this haplogroup (with a base change at np 16,086) predominates in the Philippines and Sulawesi. This may indicate that in the latter two populations, M7b3 occurs as a result of migration of individuals carrying a subset of the diversity found in Taiwan. However, the low frequency of this haplogroup in Sulawesi makes it necessary to exercise caution when interpreting these findings. An alternative interpretation is that the increased diversity of haplogroup M7b3 in the Taiwanese aboriginal population occurs due to the high frequency of samples without a mutation at np 16,086 in a single group (the Saisiat), possibly as a result of drift.

Patterns of diversity at several of the most frequent Island Southeast Asian mtDNA haplogroups appear to show a north to south movement of populations, consistent with the Out of Taiwan theory. However, Out of Taiwan models that postulate expansion as a single defined event seem to be inconsistent with the variety of shared maternal lineages in Taiwan, the Philippines, and Sulawesi. Furthermore, the area that shows extensive sharing of lineages is limited; beyond Maluku and Nusa Tenggara, only a handful of the diverse Southeast Asian haplogroups made their way into the Pacific, perhaps indicating that a population bottleneck occurred along this route (Green 1991; Kayser et al. 2000; Paz 2002). Alternatively, an increased amount of admixture with indigenous populations may have taken place in this region (Kayser et al. 2000; Mona et al. 2009).

Haplogroup E has been studied as a potential marker of early Holocene population expansion stemming from within ISEA (Hill et al. 2007; Soares et al. 2008). This hypothesized role is supported by the presence in the Philippine population of subclades E1b, E2a, and E2b at low frequencies, whereas the last two lineages are absent from the Taiwanese aboriginal population (Soares et al. 2008). The greater diversity of haplogroup E in ISEA compared with Taiwan is consistent with expansion of lineages from the south (Hill et al. 2007; Soares et al. 2008). However, as Soares et al. (2008) note, the distribution and diversity of the most common E subclade, E1a1a, is not as readily explained under this hypothesis. E1a1a has a lower diversity in the Philippine population and in Sulawesi than among the Taiwanese aborigines, despite making up a larger proportion of these populations. Whereas the increased value of ρ in Taiwan may be attributed, in part, to subdivision within the aboriginal population, it is perhaps significant that lineages within E1a1, but exclusive of E1a1a, are found in Taiwan (Soares et al. 2008) and yet are virtually absent from the Philippines. It would therefore appear that the deeper branches of E1a1 are found within the Taiwanese aboriginal population, rather than in the Philippines, reversing the pattern of genetic variation expected if E1a1 expanded from ISEA. E1a1a is found in east New Guinea and Near Oceania and, along with E1b, is the only E lineage identified in appreciable numbers outside of Southeast Asia (Friedlaender et al. 2007); this haplogroup may therefore have been carried out of ISEA during the expansion of populations into the Pacific. Thus, although haplogroup E may be a marker of postglacial expansion, clades within this haplogroup possibly reflect the impact of later population events.

Whereas as much as 14% of the mtDNA diversity in some Southeast Asian populations is made up of region-specific haplogroups (Hill et al. 2007), such lineages make up a smaller proportion of the Philippine population (table 1). Whole mtDNA genome sequencing of Philippine samples belonging to such uncommon lineages has enabled us to identify new subclades of haplogroup P in the Philippines; P subclades were previously thought to be found predominantly in Near Oceania and Australia and to have originated in that area. Although one of these Philippine subclades, P8, does not appear to be related to any known P subclade, P10 shares a transition at np 3,882 with haplogroup P2, a haplogroup found in New Guinea and Near Oceania. This may be evidence of an ancient association between the two haplogroups (P2 and Philippine haplogroup P10). The estimated founder age of P2 is 47,100 YBP (Friedlaender et al. 2007), which places the origin of the haplogroup within the timeframe of the first

colonization of Asia and the Pacific by modern humans. Friedlaender et al. (2005) suggested that the presence of haplogroup P in Southeast Asia may be the result of migration (or back migration) of women from New Guinea. However, a detailed characterization of Southeast Asian haplotypes belonging to P indicates that they are distinct from those found in New Guinea. Furthermore, the Philippine samples belonging to haplogroup P have HVS-I motifs that are different from those found in Indonesia (Hill et al. 2007; Mona et al. 2009). These lines of evidence suggest that the P haplotypes in Southeast Asia and those found in New Guinea and the South Pacific have ancient, rather than recent, common origins. Furthermore, these findings appear to support the suggestion of Mona et al. (2009) that branches of haplogroup P (and Q) may have originated in ISEA before expanding to Near Oceania and Australia.

Other indigenous Southeast Asian haplogroups also show patterns of long-term *in situ* evolution. Whole mtDNA analysis of lineages belonging to M21 and newly characterized haplogroup M73 suggests that although these haplogroups have a wide distribution in ISEA, there is evidence of substructuring within the lineages, which may result from local evolution of derived types. Similarly, M71, R24, and subclades of P appear to be highly localized and limited to (or present at very low frequencies outside) the Philippines. The restricted distribution of these lineages is in contrast to that of haplogroups such as E1a1a and M7c3c, which are dispersed over a broad swath of ISEA. This may indicate that the newly identified types are not derived from migrations out of mainland China or Taiwan, or from recent demic expansion within Southeast Asia. Rather, their distribution is more consistent with prolonged periods of population diversification *in situ*. The analysis of the diversity of such lineages is expected to provide insights on ancient population events in Southeast Asia, possibly dating back to the first dispersal of modern humans to the region. If these lineages represent the contribution of the first anatomically modern humans to enter ISEA during the Pleistocene, it would be interesting to investigate whether they have survived in isolated Philippine populations (such as the "Negritos"), as they have in the Orang Asli of Malaysia (Macaulay et al. 2005; Hill et al. 2006).

Supplementary Material

Supplementary tables S1, S2, and S3 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/). The control region sequence data for this paper appear in GenBank under accession numbers: FJ971490–FJ971589 (sequenced at DAL) and GQ119048–GQ119339 (sequenced at MMH). Whole mtDNA sequences have accession numbers GQ1190007-GQ119047, GQ352635, and GQ352636.

Acknowledgments

K.A.T. and M.C.A.D.U. acknowledge financial assistance from the Natural Sciences Research Institute and the Office of the Vice Chancellor for Research and Development of the University of the Philippines (Project No.: 060613PNSE), as well as Dr Peter Forster for the primers used at DAL and Dr Peter M. Schneider for training in mtDNA sequencing. K.A.T. thanks the British Foreign and Commonwealth Office for a Chevening Scholarship. Gayvelline Calacal, Henry Perdigon, and Jazelyn Salvador assisted in the collection and extraction of samples from the DAL. K.A.T. and M.C.A.D.U. acknowledge Dr Michael Purugganan's input in the preliminary conceptualization of the work done in the Philippines, as well as advice on statistical analysis from Dr Shuhua Xu and Frederick Delfin. J.T., J.H.L., and M.L. acknowledge financial support from the National Science Council of Taiwan (grant 95-2627-H-195-001). The authors thank Dr Melanie Pierson and an anonymous reviewer for insightful and constructive comments and suggestions. The authors are grateful to all Filipinos who donated DNA samples for this study. Finally, the authors thank Dr Mannis van Oven for advice on haplogroup nomenclature.

References

Allard MW, Wilson MR, Monson KL, Budowle B. 2004. Control region sequences for East Asian individuals in the Scientific Working Group on DNA Analysis Methods forensic mtDNA data set. *Legal Med.* 6:11–24.

Bandelt HJ, Forster P, Sykes BC, Richards MB. 1995. Mitochondrial portraits of human populations using median networks. *Genetics.* 141:743–753.

Barker G, Barton H, Bird M, et al. (27 co-authors). 2007. The 'human revolution' in lowland tropical Southeast Asia: the antiquity and behaviour of anatomically modern humans at Niah Cave (Sarawak, Borneo). *J Hum Evol.* 52:243–261.

Bellwood P. 1989. The colonization of the Pacific: some current hypotheses. In: Hill AVS, Serjeantson SW, editors. The colonization of the Pacific: a genetic trail. Clarendon Press, Oxford (UK). p. 1–59.

Cox MP. 2003. Genetic patterning at Austronesian contact zones. [PhD dissertation]. University of Otago, Dunedin, New Zealand.

Cox MP. 2008. Accuracy of molecular dating with the rho statistic: deviations from coalescent expectations under a range of demographic models. *Hum Biol.* 80(4):335–357.

Détroit F, Dizon E, Falguères C, Hameau S, Ronquillo W, Sémah F. 2004. Upper Pleistocene *Homo sapiens* from the Tabon cave (Palawan, The Philippines): description and dating of new discoveries. *Comptes Rendus Pale.* 3:705–711.

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

Forster P, Harding R, Torroni A, Bandelt HJ. 1996. Origin and evolution of Native American mtDNA variation: a reappraisal. *Am J Hum Genet.* 59:935–945.

Forster P, Torroni A, Renfrew C, Roehl A. 2001. Phylogenetic star contraction applied to Asian and Papuan mtDNA evolution. *Mol Biol Evol.* 18:1864–1881.

Friedlaender J, Schurr T, Gentz F, et al. (13 co-authors). 2005. Expanding Southwest Pacific mitochondrial haplogroups P and Q. *Mol Biol Evol.* 22(6):1506–1517.

Friedlaender JS, Friedlaender FR, Hodgson JA, Stoltz M, Koki G, Horvat G, Zhadanov S, Schurr TG, Merriwether DA. 2007. Melanesian mtDNA complexity. *PLoS ONE.* 2(2):e248.

Friedlaender JS, Friedlaender FR, Reed FA, et al. (12 co-authors). 2008. The genetic structure of Pacific islanders. *PLoS Genet.* 4(1):173–190.

Goudet J, Raymond M, de Meeüs T, Rousset F. 1996. Testing differentiation in diploid populations. *Genetics.* 144:1933–1940.

Gray RD, Drummond AJ, Greenhill SJ. 2009. Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science.* 323:479–483.

Green RC. 1991. Near and remote Oceania: disestablishing 'Melanesia' in culture history. In: Pawley A, editor. Man and a half: essays in honour of Ralph Bulmer. Auckland Polynesia Society, Auckland, New Zealand. p. 491–502.

Hill C, Soares P, Mormina M, et al. (12 co-authors). 2006. Phylogeography and ethnogenesis of aboriginal Southeast Asians. *Mol Biol Evol*. 23(12):2480–2491.

Hill C, Soares P, Mormina M, et al. (11 co-authors). 2007. A mitochondrial stratigraphy for Island Southeast Asia. *Am J Hum Genet*. 80:29–43.

Ho SYW, Larson G. 2006. Molecular clocks: when times are a-changin.'. *Trends Genet*. 22(2):79–83.

Hung HC, Iizuka Y, Bellwood P, Nguyen KD, Bellina B, Silapanth P, Dizon E, Santiago R, Datani I, Manton JH. 2007. Ancient jades map 3,000 years of prehistoric exchange in Southeast Asia. *Proc Natl Acad Sci USA*. 104(50):19745–19750.

Hurles ME, Matisoo-Smith E, Gray RD, Penny D. 2003. Untangling Oceanic settlement: the edge of the knowable, the edge of the knowable. *Trends in Ecol Evol*. 18:531–540.

Kayser M, Brauer S, Cordaux R, et al. (15 co-authors). 2006. Melanesian and Asian origins of Polynesians: mtDNA and Y chromosome gradients across the Pacific. *Mol Biol Evol*. 23:2234–2244.

Kayser M, Brauer S, Weiss G, Underhill PA, Roewer L, Schievenhoevel W, Stoneking M. 2000. Melanesian origin of Polynesian Y-chromosomes. *Curr. Biol*. 10:1237–1246.

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.

Kivisild T, Tolk HV, Parik J, Wang Y, Papiha S, Bandelt HJ, Villems R. 2002. The emerging limbs and twigs of the East Asian mtDNA tree. *Mol Biol Evol*. 19(10):1737–1751.

Kong QP, Bandelt HJ, Sun C, et al. (12 co-authors). 2006. Updating the East Asian phylogeny: a prerequisite for identification of pathogenic mutations. *Hum Mol Genet*. 15(13): 2076–2086.

Li H, Cai X, Winograd-Cort ER, et al. (12 co-authors). 2007. Mitochondrial DNA diversity and population differentiation in southern East Asia. *Am J Phys Anthropol*. 134:481–488.

Macaulay V, Hill C, Achilli A, et al. (21 co-authors). 2005. Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science*. 308:1034–1036.

Meacham W. 1988. On the improbability of Austronesian origins in Southeast China. *Asian Perspect*. (1984–1985). 26:89–106.

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

Mona S, Grunz KE, Brauer S, et al. (11 co-authors). 2009. Genetic admixture history of eastern Indonesia as revealed by Y-chromosome and mitochondrial DNA analysis. *Mol Biol Evol*. 26(8):1865–1877.

Morral N, Bertranpetit J, Estivill X, et al. (31 co-authors). 1994. The origin of the major cystic fibrosis mutation (△F508) in European populations. *Nat Genet*. 7:169–175.

O'Connell JF, Allen J. 2004. Dating the colonization of Sahul (Pleistocene Australia-New Guinea): a review of recent research. *J Arch Sci*. 31:835–853.

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

Paz V. 2002. Island Southeast Asia: spread or friction zone?. In: Bellwood P, Renfrew C, editors. Examining the farming/language dispersal hypothesis. McDonald Institute Monographs, Cambridge (UK). p. 275–285.

Pierson M, Martinez-Arias R, Holland BR, Gemmell NJ, Hurles ME, Penny D. 2006. Deciphering past human population movements in Oceania: provably optimal trees of 127 mtDNA genomes. *Mol Biol Evol*. 23(10):1966–1975.

Redd AJ, Takezaki N, Sherry ST, McGarvey ST, Sofro ASM, Stoneking M. 1995. Evolutionary history of the COII/tRNALys intergenic 9 base pair deletion in human mitochondrial DNAs from the Pacific. *Mol Biol Evol*. 12:604–615.

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

Saillard J, Forster P, Lynnerup N, Bandelt HJ, Nørby S. 2000. mtDNA variation among Greenland Eskimoes: the edge of the Beringian expansion. *Am J Hum Genet*. 67:718–726.

Schneider S, Roessli D, Excoffier L. 2000. A Software for Population Genetics Data Analysis. Arlequin: Ver 2.000. Genetics and Biometry Laboratory, University of Geneva, Geneva (Switzerland).

Soares P, Ermini L, Thomson N, Mormina M, Rito T, Röhl A, Salas A, Oppenheimer S, Macaulay V, Richards MB. 2009. Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am J Hum Genet*. 84:740–759.

Soares P, Trejaut JA, Loo JH, et al. (14 co-authors). 2008. Climate change and post-glacial human dispersals in Southeast Asia. *Mol Biol Evol*. 25(6):1209–1218.

Solheim WG II. 1988. The Nusantao hypothesis: the origin and spread of Austronesian speakers. *Asian Perspect*. (1984–1985). 26:77–89.

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

Tajima A, Sun CS, Pan IH, Ishida T, Saitou N, Horai S. 2004. Mitochondrial DNA polymorphisms in nine aboriginal groups of Taiwan: implications for the population history of aboriginal Taiwanese. *Hum Genet*. 113:24–33.

Torroni A, Rengo C, Guida V, et al. (12 co-authors). 2001. Do the four clades of the mtDNA haplogroup L2 evolve at different rates? *Am J Hum Genet*. 69:1348–1356.

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

Torroni A, Schurr TG, Yang CC, et al. (11 co-authors). 1992. Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations. *Genetics*. 130:153–162.

Trejaut JA, Kivisild T, Loo JH, Lee CL, He CL, Hsu CJ, Li ZY, Lin M. 2005. Traces of archaic mitochondrial lineages persist in Austronesian-speaking Formosan populations. *PLoS Biol*. 3:1–11.

Trejaut JA, Yen J-C, Loo J-H, He C-L, Lin M. Forthcoming. Genetic polymorphism of the plain tribes of Taiwan.

van Oven M, Kayser M. 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat*. 30(2):E386–E394 Available from: http://www.phylotree.org, mtDNA tree Build 6 (28 September 2009).

Wen B, Li H, Lu D, et al. (18 co-authors). 2004. Genetic evidence supports demic diffusion of Han culture. *Nature*. 431:301–305.

Wessel P, Smith WHF. 1995. New version of the Generic Mapping Tools released. *EOS Trans. Amer. Geophys. U*. 76(33):329.

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

Yao YG, Kong QP, Wang CY, Zhu CL, Zhang YP. 2004. Different matrilineal contributions to genetic structure of ethnic groups in the Silk Road region in China. *Mol Biol Evol*. 21(12): 2265–2280.