

# Traces of Archaic Mitochondrial Lineages Persist in Austronesian-Speaking Formosan Populations

Jean A. Trejaut<sup>1\*</sup>, Toomas Kivisild<sup>2</sup>, Jun Hun Loo<sup>1</sup>, Chien Liang Lee<sup>1</sup>, Chun Lin He<sup>1</sup>, Chia Jung Hsu<sup>1</sup>, Zheng Yuan Li<sup>1</sup>, Marie Lin<sup>1\*</sup>

**1** Transfusion Medicine Laboratory, Mackay Memorial Hospital, Taipei, Taiwan, **2** Estonian Biocentre, Tartu, Estonia

**Genetic affinities between aboriginal Taiwanese and populations from Oceania and Southeast Asia have previously been explored through analyses of mitochondrial DNA (mtDNA), Y chromosomal DNA, and human leukocyte antigen loci. Recent genetic studies have supported the “slow boat” and “entangled bank” models according to which the Polynesian migration can be seen as an expansion from Melanesia without any major direct genetic thread leading back to its initiation from Taiwan. We assessed mtDNA variation in 640 individuals from nine tribes of the central mountain ranges and east coast regions of Taiwan. In contrast to the Han populations, the tribes showed a low frequency of haplogroups D4 and G, and an absence of haplogroups A, C, Z, M9, and M10. Also, more than 85% of the maternal lineages were nested within haplogroups B4, B5a, F1a, F3b, E, and M7. Although indicating a common origin of the populations of insular Southeast Asia and Oceania, most mtDNA lineages in Taiwanese aboriginal populations are grouped separately from those found in China and the Taiwan general (Han) population, suggesting a prevalence in the Taiwanese aboriginal gene pool of its initial late Pleistocene settlers. Interestingly, from complete mtDNA sequencing information, most B4a lineages were associated with three coding region substitutions, defining a new subclade, B4a1a, that endorses the origin of Polynesian migration from Taiwan. Coalescence times of B4a1a were  $13.2 \pm 3.8$  thousand years (or  $9.3 \pm 2.5$  thousand years in Papuans and Polynesians). Considering the lack of a common specific Y chromosomal element shared by the Taiwanese aboriginals and Polynesians, the mtDNA evidence provided here is also consistent with the suggestion that the proto-Oceanic societies would have been mainly matrilineal.**

Citation: Trejaut JA, Kivisild T, Loo JH, Lee CL, He CL, et al. (2005) Traces of archaic mitochondrial lineages persist in Austronesian-speaking Formosan populations. *PLoS Biol* 3(8): e247.

## Introduction

Present day Taiwan is a home to heterogeneous groups of people. The main part of the Taiwanese population is composed of the Minnan (73.5%) and Hakka (17.5%) who descend from immigrants from the Fujian and Guangdong provinces of Southeast China during the last 400 years. After World War II, migration from different provinces of China brought a present-day share of 7.5% to 13% to Taiwan [1]. Only 1.5% of today's population of Taiwan is represented by Austronesian speakers—Atayalic, East Formosan, Puyuma, Paiwan, Rukai, Tsouic, Bunun, Western plains, Northwest Formosan, and Malayo-Polynesian languages [2]—who are generally considered indigenous to the country. According to the year 2000 census, Yami with 4,050 individuals and Amis with 146,796 individuals represent the smallest and the largest tribal populations of Taiwan respectively.

The archaeological record suggests that a substantial cultural change occurred in Taiwan approximately in the sixth millennium BC [3]. Whether or not the transition to Neolithic technology in Taiwan corresponded to a substantial gene flow from China is unclear. It is also not clear whether there were one or multiple waves of Neolithic migrations to northern and southern Taiwan from southeast China [4–6]. There is archaeological evidence for the occupation of caves in southern Taiwan by humans by at least 15 thousand YBP [4]. Yet, the bulk of archaeological material started to accumulate during the Neolithic period.

Three different views have been put forward to explain the origin of present day Austronesian-speaking tribes in Taiwan, including (a) origin from expanding regions of early rice cultivation in central China [7], (b) origin in insular Southeast Asia or (c) local settlement after the rise of the sea levels [8]. During the massive immigration of Han speakers to the western plains of Taiwan, most tribes took refuge in remote regions such as the central mountain ranges or the east coast of Taiwan. It is believed that this geographical isolation has largely contributed toward maintaining their culture and languages until the present day in contrast to the plain tribes that are characterized by high levels of admixture [9].

Received April 29, 2005; Accepted May 11, 2005; Published July 5, 2005  
DOI: 10.1371/journal.pbio.0030247

Copyright: © 2005 Trejaut et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abbreviations: COII/tRNA<sup>LYS</sup>, mtDNA intergenic region between the COII gene and the Lysine tRNA gene; D-loop, displacement loop; HLA, human leukocyte antigen; HVS-I, hyper-variable segment I; MRCA, most recent common ancestor; mtDNA, mitochondrial DNA; np, nucleotide position; PC, principal components; RFLP, restriction fragment length polymorphism

Academic Editor: David Penny, Massey University, New Zealand

\* To whom correspondence should be addressed. E-mail: jtrejaut@ms1.mmh.org.tw (JAT), marilin@ms2.mmh.org.tw (ML)

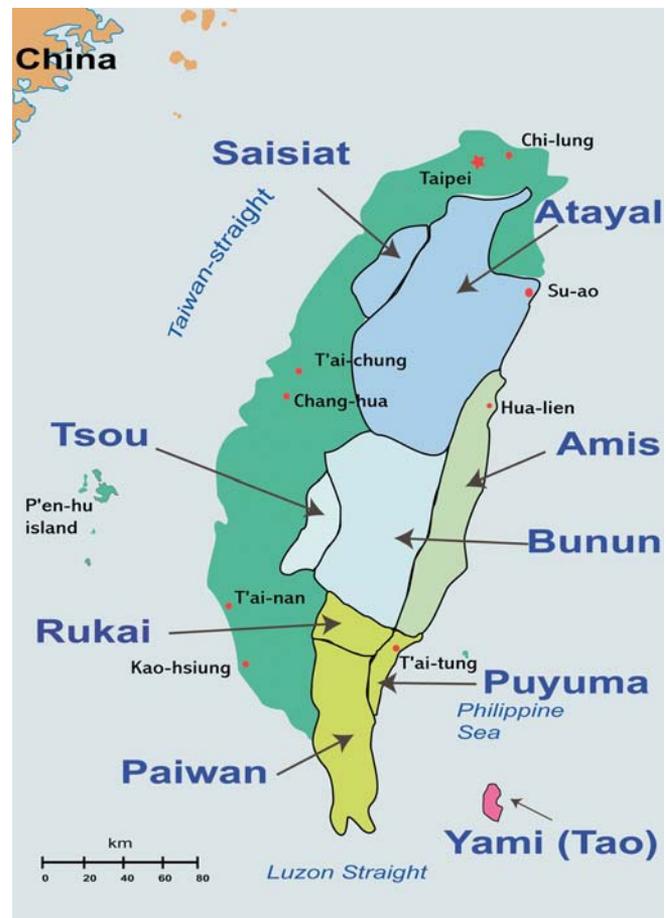
‡ Current address: Transfusion Medicine Laboratory, Mackay Memorial Hospital, Taipei, Taiwan

The role of Taiwanese indigenous populations in the Austronesian settlement of island Southeast Asia and Polynesia has come under intense discussion during the last two decades, in particular among geneticists working with mitochondrial DNA (mtDNA), Y chromosome, and HLA loci [9–21]. Although the landscape of possible migration out of Taiwan and interaction scenarios is quite complex [22], two opposing models, the “express train” [23] and the “entangled bank” models [24], emerging from among other alternatives [8,16,25] have commonly been tested with genetic data. Proponents of the express train model [23] hold that the Polynesian islands were settled in a relatively short period with a migration originating from southern China to Taiwan. According to this model, the Austronesian migration from Taiwan was coupled by the spread of Lapita culture, whose supposed precursor pottery in Taiwan is dated to at least 6000 YBP [26]. The entangled bank model [24], along with the “slow boat” model [8,16,25], on the other hand, proposes that the ancestors of Polynesians did not move rapidly through Southeast Asia and Melanesia. According to these two models, the interaction with Melanesians before colonizing the Pacific was the major factor determining the genetic composition of the Polynesian ancestors. The proponents of the intermediary slow boat model [8,16,25], also popularly introduced in the book *Eden in the East* [27], propose a date and a location in Southeast Asia for the origin of the Polynesian expansion and question the Taiwanese genetic contribution to the ancestors of Polynesians.

Previous mtDNA studies [13,28] on four tribal groups (Amis, Atayal, Bunun, and Paiwan) and on the Taiwanese general population [12,29] have revealed their common ancestral origins with populations of China and/or Southeast Asia. Tajima et al. [9] characterized two hyper-variable segment I (HVS-I) lineage groups as unique to Taiwanese aboriginals, accounting for 22% of their mtDNA variation, and estimated their origin to approximately 11,000–26,000 years ago. Three groupings of aboriginal Taiwanese populations were suggested by this study [9]: southern (Rukai and Paiwan), northern (Atayal and Saisiat), and central-eastern (Amis, Bunun, Tsou, Puyuma, and Yami). However, the data did not conclusively show whether these three distinctive clusters reflected three different migrations to Taiwan or whether they could be explained by the effect of drift and long-term isolation of the island’s populations.

A common HVS-I motif 16189–16217–16261, classified within haplogroup B4a [30], is shared between Taiwanese and Polynesian populations. This motif was considered first as the genetic link supporting Polynesian origins in Taiwan [10,13]. This root haplotype of B4a from which the Polynesian motif derives by a single HVS-I mutation is, however, widely spread in East Asia. Besides Taiwan and Southeast Asia, B4a occurs frequently in southern China and has been observed as far inland as among Mongolians [21,30–35]. Its deep time depth,  $40,400 \pm 9,600$  years in China [34], significantly predates the timeframe that is relevant to the peopling of Polynesia. Therefore, this general genetic link between Taiwan and Polynesia has now been refuted as being supportive of Polynesian origins specifically in Taiwan within a recent timeframe as implicated by the fast train model [8].

An overwhelming majority of mainland East Asian mtDNA lineages are nested within haplogroups A, G, M, N9, and R9 [30,31,34–38]. Although previous mtDNA research on Taiwa-



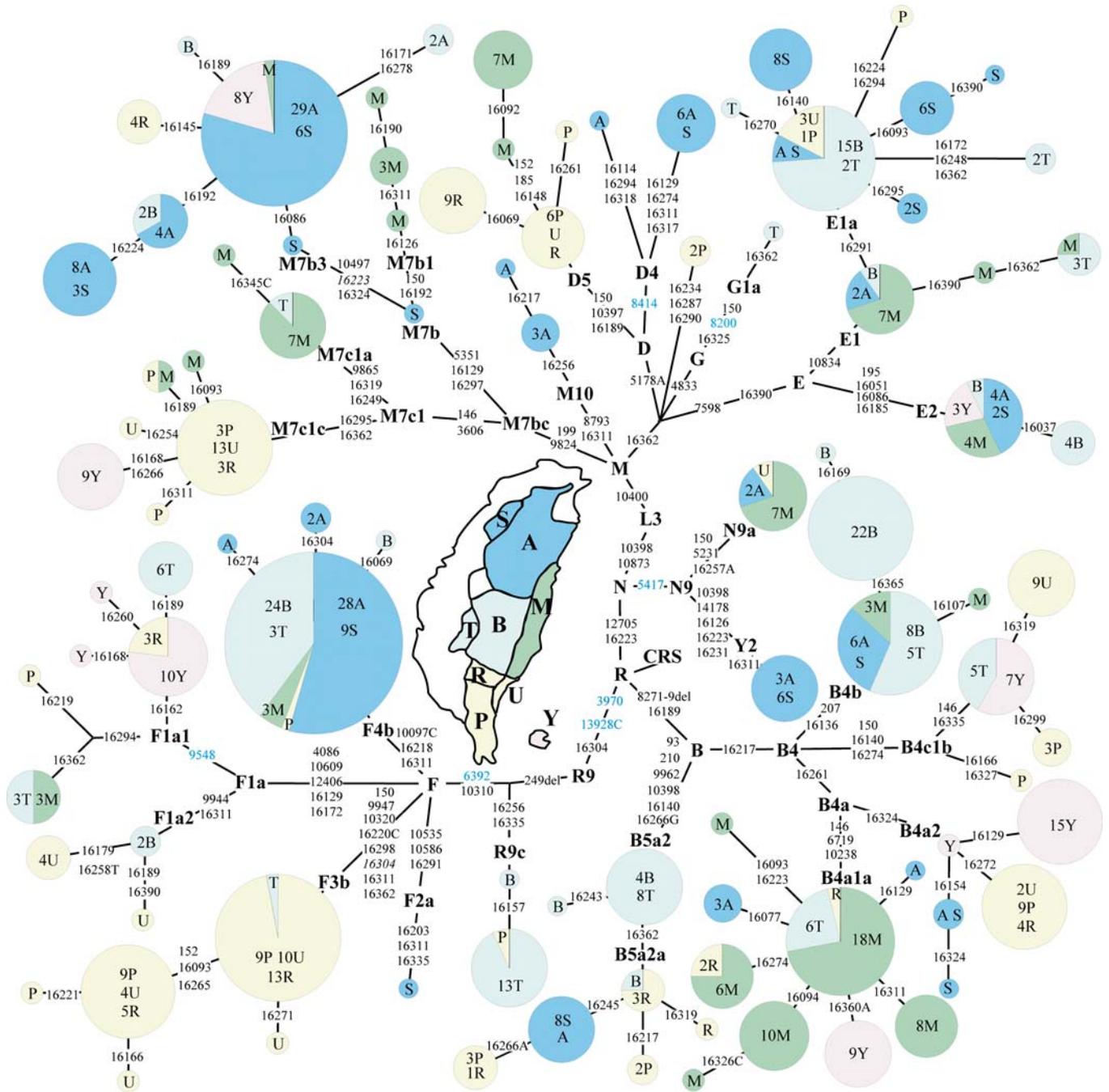
**Figure 1.** Geographic Distribution of Nine Indigenous Tribes of Taiwan  
DOI: 10.1371/journal.pbio.0030247.g001

nese aboriginal populations has focused only on hyper-variable displacement loop (D-loop) data, we attempt here to define the distribution of these haplogroups and their region-specific twigs in nine Taiwan indigenous tribes (Figure 1). Using the coalescent approach, we examine the hypothesis that the distinct mtDNA variation and genetic structure seen in present-day Taiwanese mountain tribes might be the result of a Neolithic migration or of a long-term separation from their common ancestors with the modern Chinese populations of mainland East Asia. Finally, to determine if Taiwanese B4a lineages share any coding region variants with Polynesian B4a lineages [39], we determined the complete mtDNA sequence for eight different B4a lineages from indigenous Taiwanese.

## Results

### Haplogroup Structure and Distribution in Taiwanese Aboriginal Populations

Ninety-six distinct haplotypes were identified by 81 variable sites of the mtDNA control region in 640 Taiwanese aboriginal samples representing all nine mountain tribes (Figure S1). Among these, 39 individuals possessed a unique HVS-I haplotype. From the 96 haplotypes, 30 were shared by at least two different tribes. Haplotype sharing occurred mostly between adjacent tribes (Figures 2 and S1). Phylogenetic analysis using control and coding region information



**Figure 2.** Tree Drawn from a Median-Joining Network of 96 mtDNA Haplotypes Observed in Nine Indigenous Taiwanese Populations  
 The tree is based on sequences of HVS-I (16024–16390) and a coding region segment covering 9,793 to 10,899 bps. Haplogroup defining HVS-II mutations were manually added after the generation of the network. Additional coding region mutations, ascertained through complete mtDNA sequencing of an individual of each subclade of the haplogroup defined by the mutation, were used to generate the network and are shown in blue. All nps are numbered according to reference sequence [71]. Mutations in italic indicate back conversions. Nucleotide change is specified only for transversions. Node areas are proportional to haplotype frequencies of the pooled nine tribes. The population codes are as follows: A, Atayal; B, Bunun; M, Amis; P, Paiwan; R, Rukai; S, Saisiat; U, Puyuma; T, Tsou; and Y, Yami. CRS = Cambridge reference sequence [70].  
 DOI: 10.1371/journal.pbio.0030247.g002

clustered the observed haplotypes into 20 distinct haplogroups and subgroups whose distribution among tribes was compared to available information from other Asian populations (Table 1). Four basic haplogroups—B, E, R9, and M7—accounted for more than 90% of the variation observed in aboriginal Taiwanese. In comparison, the combined frequency of these haplogroups in China averaged less than

40%, with southern Chinese having significantly higher frequency than northern Chinese (Table 1). Among these four basic haplogroups, haplogroup E was nearly absent in continental Asia. On the other hand, haplogroups A, D4, G, and M8–M10 accounted approximately for 41% of the sequences from the mainland, whereas it was rare or absent in Taiwanese Aborigines. Only two unique D4 haplotypes

**Table 1. Haplogroup Frequencies in Taiwan, East Asia, and Oceania**

Haplogroups	Sub-haplogroups	Taiwan Tribes %										Northern Mountain Tribes		Central Mountain Tribes		Southern Mountain Tribes		East Coast Tribes		Southeast Asia		Indo-China		Southern Han Asia		Northern Han Asia	
		Nine		Northern		Central		Southern		East Coast		Southeast		Indo-China		Southern Han Asia		Northern Han Asia									
		Atayal	Saisiat	Tsou	Bunun	Paiwan	Rukai	Puyuma	Annis	Tao	Luzon	Philippines	Indonesia	Moluccas	South Vietnam	Malaysia	Thailand	Taiwanese	South China	North China	Japan	Korea	Ryukyu				
		(n = 109)	(n = 63)	(n = 60)	(n = 89)	(n = 55)	(n = 50)	(n = 52)	(n = 98)	(n = 64)	(n = 37)	(n = 59)	(n = 54)	(n = 62)	(n = 35)	(n = 51)	(n = 33)	(n = 155)	(n = 142)	(n = 189)	(n = 62)	(n = 320)	(n = 48)				
A		6.37	0.92	3.17	16.36	8.00	3.85	25.00	1.69	—	—	—	—	—	1.96	—	—	6.45	1.41	8.99	6.45	8.44	—				
B (others)		6.04	5.50	1.58	8.33	34.83	4.08	—	—	—	—	—	—	—	—	—	—	0.65	1.41	1.06	—	—	—				
B4		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1.29	2.11	1.59	6.45	0.94	—				
B4a (all)		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	10.56	3.70	—	—	—	—				
B4a		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	9.15	3.70	—	—	—	—				
B4a1		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	5.81	1.41	—	—	—	—				
B4a1a (all)		8.74	3.67	10.00	6.00	14.06	44.90	14.06	15.25	9.26	24.19	9.26	9.26	9.26	—	—	—	1.94	1.41	—	—	—	—				
B4a1a		8.74	3.67	10.00	6.00	14.06	44.90	14.06	15.25	9.26	24.19	9.26	9.26	9.26	—	—	—	—	—	—	—	—	—				
B4a1a	(np 16247)	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—				
B4a2		6.37	0.92	3.17	16.36	8.00	3.85	25.00	1.69	—	—	—	—	—	—	—	—	3.87	4.93	3.17	3.23	0.63	2.08				
B4b (all)		6.04	5.50	1.58	8.33	34.83	4.08	—	—	—	—	—	—	—	—	—	—	0.65	4.93	3.17	3.23	0.63	2.08				
B4b1		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—				
B4b <sup>a</sup>		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—				
B4c1		4.87	—	8.33	7.30	17.30	10.90	2.70	6.78	3.70	4.84	2.86	1.96	—	—	—	—	3.52	0.53	—	—	—	—				
B5 (all)		5.86	0.90	12.70	13.33	6.70	9.10	10.00	5.08	7.41	8.84	7.84	12.12	12.12	12.12	12.12	12.12	3.23	2.82	4.23	8.06	2.81	—				
B5a		5.86	0.90	12.70	13.33	6.70	9.10	10.00	2.70	3.39	7.41	3.23	5.71	7.84	7.84	7.84	7.84	3.23	2.82	4.23	8.06	2.81	—				
B5b		—	—	—	—	—	—	—	10.81	1.69	—	—	—	—	—	—	—	—	—	—	—	—	—				
C + Z + M8a		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—				
C		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—				
Z		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—				
M8a		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—				
D5		4.76	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—				
E		11.38	6.40	31.74	13.33	23.59	3.63	5.76	13.26	4.68	8.11	11.11	9.68	—	—	—	—	—	—	—	—	—	—				
F1a (all)		6.06	—	15.00	2.25	1.82	6.00	7.69	3.06	18.75	13.51	5.08	16.67	24.19	11.43	5.88	15.15	5.81	14.08	4.23	1.61	0.94	2.08				
F1a		2.20	—	5.00	2.25	1.82	6.00	7.69	3.06	18.75	13.51	5.08	16.67	24.19	11.43	5.88	15.15	5.81	14.08	4.23	1.61	0.94	2.08				
F1a1		3.86	—	10.00	—	—	6.00	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—				
F1b		0.21	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—				
F1c		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—				
F2		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—				
F4b		11.44	8.97	28.44	14.29	23.80	1.66	5.00	28.09	3.40	2.70	1.69	—	—	—	—	—	—	—	—	—	—	—				
M7 (all)		8.97	39.44	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—				
M7a		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—				
M7b1		0.74	—	1.60	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—				
M7b3		8.23	39.44	22.20	3.40	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—				
M7 <sup>b</sup>		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—				
M7c (all)		8.97	—	1.70	7.30	6.00	28.85	10.24	26.60	21.62	6.78	7.41	—	—	—	—	—	—	—	—	—	—	—				
M7c1c		7.87	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—				
M7c <sup>a</sup>		1.10	—	1.70	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—				
M9		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—				
M10/M4		0.41	3.70	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—				
M <sup>a</sup> and others		1.48	6.40	1.60	1.70	3.60	3.60	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—				
M <sup>b</sup>		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—				
M(D4)		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—				
G		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—				
N9a		1.20	1.80	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—				
R (all)		2.91	1.58	21.66	1.12	1.81	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—				
R <sup>b</sup>		2.91	1.58	21.66	1.12	1.81	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—				
T1		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—				
U2		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—				
Y		1.37	2.80	9.50	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—				
Gene diversity		NA	0.853	0.922	0.906	0.838	0.91	0.881	0.868	0.924	0.864	0.936	0.997	0.984	0.986	0.996	0.995	0.995	0.995	0.995	0.986	0.992	0.990				
± SD		± 0.022	± 0.012	± 0.018	± 0.021	± 0.018	± 0.026	± 0.025	± 0.013	± 0.016	± 0.018	± 0.006	± 0.008	± 0.005	± 0.008	± 0.003	± 0.001	± 0.001	± 0.001	± 0.001	± 0.006	± 0.002	± 0.002				
Nucleotide diversity		NA	0.014	0.018	0.015	0.014	0.016	0.017	0.014	0.017	0.014	0.014	0.018	0.021	0.022	0.018	0.019	0.018	0.019	0.017	0.017	0.015	0.012				
± SD		± 0.007	± 0.009	± 0.008	± 0.008	± 0.008	± 0.008	± 0.009	± 0.008	± 0.008	± 0.008	± 0.008	± 0.009	± 0.011	± 0.010	± 0.009	± 0.011	± 0.009	± 0.009	± 0.009	± 0.009	± 0.008	± 0.007				

Populations references are shown in Figure 3. A dash indicates that the haplogroup is not seen. Frequencies in bold type were used for PC analysis.  
<sup>a</sup>All other subtypes.  
 all, compound frequency; n, number of individuals; NA, not applicable; nd, not determined.  
 DOI: 10.1371/journal.pbio.009247.t001



were observed in the Atayal and Saisiat populations and a single G1a lineage in Tsou (Table 1; Figures 2 and S1).

### Principal Components Analysis

Principal components (PC) analysis using haplogroup frequencies revealed a high level of differentiation between Taiwanese aborigines as compared with other Asian populations (Figure 3). In particular, Taiwanese aboriginal populations appeared closer to island Southeast Asian populations (Luzon, Philippines, Moluccas, and Indonesia) than to populations from mainland East Asia (Fujian, South Vietnam, Malaysia, and Thailand). The three southernmost populations of Taiwan (Puyuma, Paiwan, and Rukai) and, more distantly, Yami from Orchid Island clearly differentiated the southern populations from the northern and central populations of Taiwan from which the Bunun sample emerged as an outlier. Although the Amis population clustered closely with northern and central tribes in the first two dimensions of the PC analysis, their haplogroup structure and relatively high frequency of B4a, D5, and M7c clades at the same time showed an affinity toward southern tribes of Taiwan. The grouping of Taiwanese indigenous populations revealed by our analysis differs from the population tree drawn from pairwise nucleotide differences by Tajima et al. [20] in which Puyuma and Yami, for example, clustered with central and eastern populations.

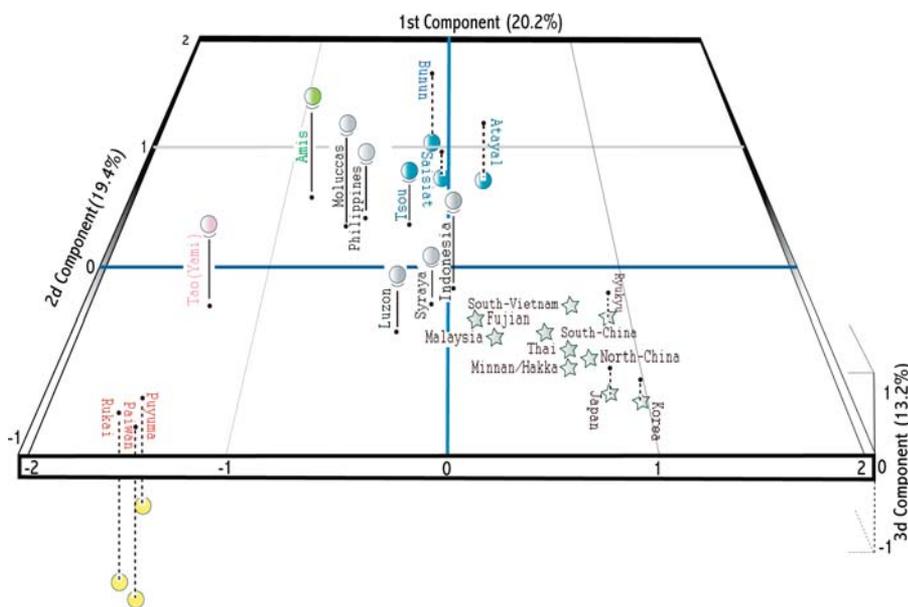
### Tribes of Northern and Central Taiwan

Haplogroups B4b, B5a2, E, F4b, and M7b covered more than 80% of the mtDNA variation observed in Atayal, Saisiat, and Bunun populations (north and central Taiwan; see Figure 2). The frequency of these haplogroups elsewhere in Taiwan was significantly lower (24%;  $p < 0.0001$ ). Haplogroup E was most divergent and frequent in the Saisiat population in the northern mountain ranges of Taiwan.

Two subclades of haplogroup B5a can be defined on the basis of complete sequence information [37,38], available restriction fragment length polymorphism (RFLP) and HVS-I information from Southeast Asia [11,30,33,34,40,41], and our data: The loss of a *Hae*III site at nucleotide position (np) 6957 in association with 16266A allele defines the major subclade B5a1 spread in Southwest China, Thailand, Vietnam, and Nicobar Islands whereas the presence of +11146 *Dde*I site at np 11146 defines a B5a2 clade that is predominantly in association with 16266G allele. Within the B5a2 clade, available D-loop information allows us to postulate the presence of two further branches that are highly region specific in Southeast and East Asia: HVS-I motif 16140–16189–16266G–16362 (B5a2a), which so far has been observed exclusively in Taiwan [9,28], whereas another motif, 16140–16187–16189–16266A/G, characterizes all B5a variants observed so far in Korean, Japanese, and Han lineages from northeast China [33,34,40,42]. In Taiwanese aboriginals, B5a2 lineages were found all over the island whereas their frequency, likely due to drift, was the highest in north-central regions, among the Tsou and Saisiat (see Figure 2).

Over one quarter of Taiwanese from northern and central mountain regions (Atayal, Saisiat, and Bunun) belonged to a novel subclade of haplogroup F4 [38,39,43,44] called F4b because of its distinctive mutation motif 10097C–16218–16311 (see Figure 2). Considering available mtDNA datasets for Southeast Asia, haplogroup F4b has a marginally low frequency (<1%) in China [30,31,34] and was not found among 519 Thai samples [45,46], in which the overall haplogroup F frequency was one of the highest in Asia (22%).

Most northern Taiwanese aboriginals from the M7b clade possessed a combination of four HVS-I substitutions (at nps 16086–16129–16324), which defines the clade M7b3 [47]. HVS-I sequences with a similar motif have been observed



**Figure 3.** Principal Components Analysis

Principal components map obtained from a matrix of haplogroup frequencies in nine Taiwan indigenous tribes (Table 1), northern and southern Chinese [34], Taiwan urban population (Minnan and Hakka) [29], Japan [73], Korea [33], Luzon [12], Moluccas [14], Indonesia, Ryukyu [42] Malaysia, South Vietnam, the Philippines [20], and Thai [45,46]. Austronesian-speaking populations are represented by circles, non-Austronesian-speaking populations by stars.

DOI: 10.1371/journal.pbio.0030247.g003

previously only in one Hui, one Uighur, one Dai, and one Mongolian sequence [32,43,47,48]. Consistent with the study of Tajima et al. [9] in which the HVS-I–defined cluster C4 corresponds to our M7b3, we found no exact matches to the Taiwanese M7b3 sequences in the available literature. Two Taiwanese complete sequences with M7b3 HVS-I motif [39], one matching our haplotype 56, the other deriving from haplotype 60 by a single HVS-I mutation, shared seven coding region mutations (at nps 1664, 4454, 6351, 9468, 10497, 12121, and 14115) that have not been seen in other M7b sequences from China or Japan [37,38,49]. Furthermore, the Taiwanese M7b3 sequences lacked a mutation at np 12811 that is shared by sister clades M7b1 and M7b2 [37]. Interestingly, all Taiwanese aboriginal populations lacked the M7b2 subclade that occurs with remarkable haplotype diversity ( $h = 1$ ) and frequency (8%) in neighboring Ryukyu island populations [42]. Also, subclade M7b1, which is common in southern Chinese [30] and the non-indigenous Taiwanese population (9% frequency; [42]) was found at low frequency only among the Amis. Interestingly, all five Amis M7b1 sequences derived from their ancestral haplotype by a substitution at np 16126 that was not observed among any other 28 M7b1 samples from Asia [30]. This observation makes unlikely the possibility that admixture with Han populations from China would be the source of the M7b1-16126C lineages seen in Amis.

Highly homogeneous haplogroup B4b lineages were concentrated in the northern and central regions of Taiwan, likely involving a founder effect in Bunun. This clade is frequent in southern China where an exact match for the dominant B4b haplotype is also observed among the Taiwanese [30].

Haplotype 30 (see Figure S1), observed in six Saisiat and three Atayal samples within haplogroup Y, included only a match with two Han Chinese from Shanghai [40]. Similarly, exact haplotype matches with Han Chinese from China were found among tribal samples grouped in haplogroups B4b and D4, which were observed only in the northern-central parts of Taiwan, in the Atayal, Saisiat, and Bunun populations. Haplogroups B4b and D4 could represent recent admixture mediated through urban populations. Nonetheless, this appears to be quite unlikely for the Bunun who have historically been one of the most isolated tribes in Taiwan, who practiced head hunting until the early 20th century. The high incidence of two specific B4b haplotypes among them, in contrast, can be well explained by drift operating in a strictly endogamous small community.

Thirteen Tsou sequences, one Paiwan, and one Bunun sample shared a characteristic haplogroup R9c substitution pattern [50] in the control region (see Figure 2). This haplogroup has been observed at low frequency in mainland Southeast Asia [34,50,51].

#### Yami and Tribes of Southern and Southeastern Taiwan

Haplogroups B4a, D5, F3b, M7c, and N9a characterized 72.2% of the mtDNA variation of the populations of south and southeast Taiwan. Two distinct subclades of M7c were observed. The first subclade, M7c1c [30], occurs frequently all over the southeast coast, and is most frequent among Puyuma (29%). The second subclade, M7c1a, is seen in Amis (8%) and one Tsou sample. Haplotype diversity ( $h = 0.59$ ) of M7c1 in Taiwan is likely reduced by genetic drift and founder effects among the Puyuma and Yami tribes. Similarly, without any

downstream variation, only the founder HVS-I haplotype (16223–16257A) of haplogroup N9a was observed in Atayal, Amis, and Puyuma populations. Subclade B4c1b, frequent in southern Taiwan, has been previously detected at low frequency in China, Japan, Malaysia, Vietnam, the Philippines, and Java [13,20,30,38,51–53]. Two distinct subgroups of B4a can be found in Taiwanese aboriginals. The first of them, B4a1a, distinguished by a substitution at np 6719 (see Figures 2, S1), occurs frequently among the Amis and Yami, whereas the second, B4a2, is common in the southern tribes Paiwan, Puyuma, Rukai, and Yami.

Haplogroup F3b, previously labeled as R9a, has been encountered at low frequency in south and west China [30,34,37]. In Taiwanese, a high frequency of F3b was observed specifically among the three southernmost populations and is virtually absent in other tribes. This is consistent with the distribution of the respective HVS-I cluster C2 described in the study of Tajima et al. [9]. Interestingly, the HVS-I motif of all Taiwanese F3b samples was different from the motif commonly observed among the Chinese. Namely, Taiwanese F3b samples were characterized by a transversion at np 16220 (16220C) and lacked the transition at np 16335 seen in R9c. The 16220C variant of haplogroup F3b has been observed previously in two Bai samples from Yunnan Province of China and in one Japanese sample from Honshu [42,43]. In both cases, however, the difference with the closest Taiwanese HVS-I haplotype is two or more substitutions in HVS-I. Haplogroup F1a distribution is concentrated among the Tsou and Yami populations. Interestingly, among Yami, only the F1a1 subclade, defined by the HVS-I motif 16129–16162–16172–16304, was detected, although other haplogroup F subclades occur at considerably high frequency throughout other populations of Taiwan.

Finally, haplogroup D5 is spread at a moderately low frequency throughout East Asia, being most frequent in the south and absent or rare in Central Asia and Siberia [30]. Among Taiwanese aboriginal groups, distribution of D5 lineages is restricted to the three southernmost populations and the Amis. Although samples from Paiwan, Puyuma, and Rukai carry the root HVS-I haplotype of D5 or its one-step mutation descendants, the Amis harbor a derived motif with two substitutions (16148 and 16092) that has not been detected elsewhere in Asia so far.

#### Coalescence Times

A number of mtDNA haplogroups in Taiwan showed limited haplotype and nucleotide diversity. Consequently, according to control sequence information, sequences belonging to haplogroups B4b, D4, E2, F4b, N9a, and Y (and possibly R9c when ignoring the unique Bunun specimen) coalesced in their most recent common ancestors (MRCAs) within the last two thousand years (Table 2). These low coalescence times may be due either to drift or to bottlenecks affecting locally evolving lineages, or they may be due to founder effects following admixture with recent immigrants from outside the island. The presence or absence of matching sequence types elsewhere, as detailed above, can be considered the best possible way to evaluate these two scenarios. The nearly complete absence of haplogroup E in China makes it unlikely that the shallow time depth of E2 lineages in Taiwan could be explained by a migration from China. However, many populations from island Southeast Asia are

**Table 2.** Coalescence Times of Major mtDNA Haplogroups in Taiwanese Aborigines

Haplogroup Nodes	n	$\rho \pm \sigma$	Age $\pm$ SE (Years in Thousands)
E2 <sup>a</sup>	18	0.0 to 0.06 <sup>a</sup>	0 to 1.1 <sup>a</sup>
N9a <sup>a</sup>	10	0.0 to 0.10 <sup>a</sup>	0 to 2.0 <sup>a</sup>
Y <sup>a</sup>	9	0.0 to 0.11 <sup>a</sup>	0 to 2.2a
D4 <sup>a</sup>	7	0.0 to 0.14 <sup>a</sup>	0 to 2.9 <sup>a</sup>
B4b	47	0.04 $\pm$ 0.03	0.8 $\pm$ 0.6
F4b	72	0.04 $\pm$ 0.03	0.9 $\pm$ 0.6
R9c <sup>b</sup>	15	0.07 $\pm$ 0.07 <sup>b</sup>	1.3 $\pm$ 1.3 <sup>b</sup>
F1a1	21	0.38 $\pm$ 0.29	7.7 $\pm$ 5.9
B4a1a	66	0.47 $\pm$ 0.23	12.2 $\pm$ 4.7
B4c1b	24	0.50 $\pm$ 0.4	10.1 $\pm$ 8.0
M7b3	73	0.51 $\pm$ 0.23	10.3 $\pm$ 4.5
E1a	44	0.59 $\pm$ 0.26	11.91 $\pm$ 5.2
D5	26	0.61 $\pm$ 0.38	12.3 $\pm$ 7.7
M7c1c	33	0.70 $\pm$ 0.4	14.1 $\pm$ 8.1
F1a2	7	0.71 $\pm$ 0.59	14.4 $\pm$ 11.9
F3b	54	0.79 $\pm$ 0.22	16.1 $\pm$ 4.5
R9c	15	0.93 $\pm$ 0.87	18.8 $\pm$ 17.6
B4a2	34	1.00 $\pm$ 0.63	20.2 $\pm$ 12.7
M7b1	5	1.00 $\pm$ 0.66	20.2 $\pm$ 13.4
B5a2	33	1.12 $\pm$ 0.57	22.6 $\pm$ 11.7
E1	59	1.25 $\pm$ 0.47	25.2 $\pm$ 9.7

<sup>a</sup>When haplogroups have no descendent-associated lineage, one-step mutation is introduced in the estimation of  $\rho$  and age.

<sup>b</sup>Possible alternative calculation in which np 16157 is treated as a reverse conversion in a unique F9c Bunun lineage. DOI: 10.1371/journal.pbio.0030247.t002

still poorly covered, and the possibility that these lineages can derive from a migration from outside should be kept open.

More than half of Taiwanese mtDNA lineages fall into clades B4a1a, B4c1b, E1a, F1a1, F1a2, F3b, M7c1c, and M7b3 that show, with a broad range of standard errors, average coalescence times between 7.7 and 16.1 thousand years. The dates for clades M7c1c and M7b3 are similar to those of M7b1 and M7b2 previously reported in Southeast Asia [30]. It is likely that these four M7b daughter clades, together with other subclades of haplogroups B4, E, and F3b, began to diversify at the time of the rise of the sea levels after the end of the Younger Dryas cold spell approximately 11,000 years ago in distinct islands close to mainland Southeast Asia.

### Taiwanese Affinities with Austronesian Populations of Oceania

PC analysis reveals that Taiwanese populations in general and the Amis specifically are more closely related to island Southeast Asian populations than to populations from mainland East Asia (see Figure 2). However, only haplogroups B4a1a and M7c1c substantiate the close affinity between Taiwanese populations and those from islands beyond the Philippines. Characteristic HVS-I motifs of haplogroups B5a, F3b, R9c, and E2 were also observed in the Philippines and Sabah from Borneo; haplogroups Y and M7b3 were seen as shared only with Philippines' populations. Finally, F4b and R9c lineages are noncharacteristic of Near and Remote Oceania [11,12,20,39].

The B4a sequences, in general, are frequent along the southeastern coast of Taiwan, but also have a wide and ancient distribution all over East Asia and island Pacific. To provide a more detailed phylogenetic resolution, we deter-

mined the complete mtDNA sequence of eight different HVS-I haplotypes observed in Taiwan that belong to haplogroup B4a. Phylogenetic analysis, including 23 published B4a complete sequences, reveals an interesting substructure of the clade. One major subclade, B4a1, can be defined now by three coding region mutations at nps 5465, 9123, and 10238 (Figure 4) that cover all Taiwanese and Oceanian B4a sequences. The B4a clade also includes samples from Korea, Japan, and China. Using a mutation rate of  $1.26 \pm 0.08 \times 10^{-8}$  in the mtDNA coding region [54], the coalescence time of B4a1 ( $28.9 \pm 7.4$  thousand years) is consistent with its origin, likely somewhere in Southeast Asia before the last glacial maximum. Notably, besides the B4a1 subclade, which occurs at minor frequency in Han Chinese (8/332; [30,34]), at least two other subclades of B4a can be found in southern China. Similarly, among Japanese, haplogroup B4a1 is represented by three different subclades that appear as sister groups to the Taiwanese/Polynesian-specific clade B4a1a (Figure 4). On the other hand, the minor B4a2 subclade among Taiwanese has one closely related lineage among Japanese.

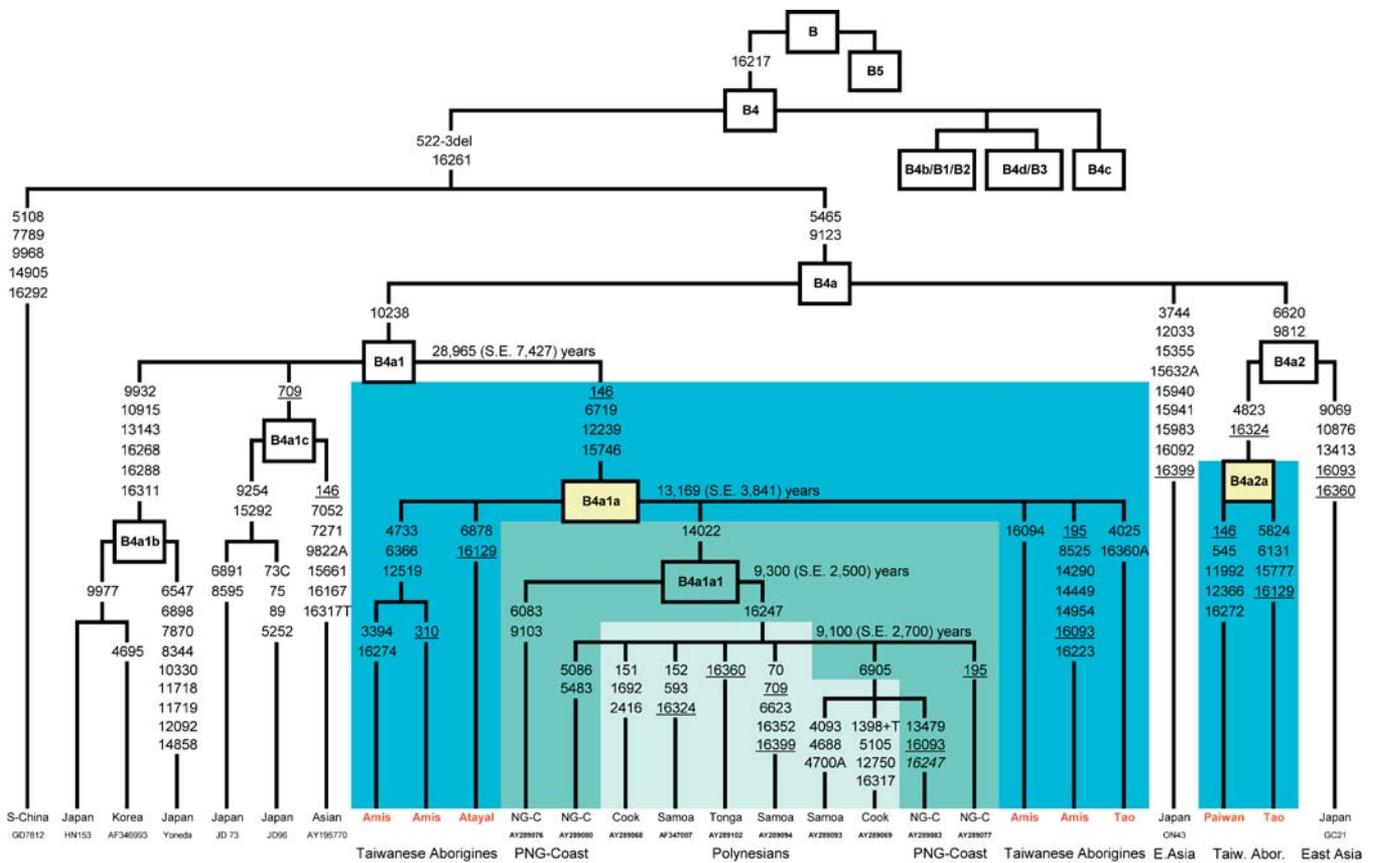
The coalescence time of haplogroup B4a1a, based on coding region mutations, is  $13.2 \pm 3.8$  thousand years, consistent with its HVS-I-based age estimate. To test whether B4a1a could have been imported recently to Taiwan from the mainland, we genotyped the defining markers of this clade in 47 North Vietnamese and 79 Han speakers from Fujian (unpublished data), the closest province of China to Taiwan, and we did not observe any carriers of haplogroup B4a1a. The clustering pattern within B4a1a therefore provides a unique link between Polynesian, Papuan, and Taiwanese lineages, supporting their common origin around the Younger Dryas period somewhere in island Southeast Asia, possibly in Taiwan. The Polynesian and Papuan sequences occupy a derived position in the B4a1a tree, descending from their MRCA defined by a mutation at np 14022 that was not observed in Taiwanese. The derived B4a1a1 clade in Papuans and Polynesians has a coalescence time of  $9.3 \pm 2.5$  thousand years.

## Discussion

### Genetic Composition of mtDNA Lineages in Aboriginal Taiwanese

Analysis of the mtDNA haplogroup composition in aboriginal populations of Taiwan revealed highly distinctive patterns of the spread of the subgroups of common Asian specific haplogroups as compared to populations of China. On the one hand, haplogroups B, E, R9, and M7, which cover most of the lineages in aboriginal Taiwanese, have significantly lower frequencies in China. On the other hand, haplogroups A, C, G, and M8–M10, which are frequent in China, are completely absent in the nine tribes of Taiwan. In contrast to mainland China, the low haplotype diversity seen among small indigenous populations suggests that genetic drift may be responsible for generating the distinct frequency patterns reported in this study (Table 1).

Interestingly, however, haplogroups that make up the majority of lineages in Taiwanese aboriginals were also observed at high frequency (>60%; Table 1) among other island Southeast Asian populations who similarly present significant differences in their combined haplogroup frequencies with mainland East Asian populations. It is highly



**Figure 4.** Phylogenetic Tree Relating Haplogroup B4a1 Complete Sequences  
 Base pair exchange is specified only for transversions. Recurrent mutations are underlined. Coalescence times are shown beside nodes.  
 DOI: 10.1371/journal.pbio.0030247.g004

unlikely that the increment of the compound frequency of haplogroups B, E, R9, and M7 seen among all aboriginal populations of Taiwan could be explained by drift. It seems more plausible that the ancestral population of coastal East Asia and island Southeast Asia was already enriched by the founder lineages of these haplogroups and that drift affected the frequencies of individual haplogroups while their combined frequency throughout this region remained high.

In light of this study, the complete absence among aboriginal populations of several mtDNA haplogroups common on the Chinese mainland would suggest that the Neolithic colonizers (a) did not contribute significantly to the mtDNA pool of the pre-existing Formosan population; or (b) that the Neolithic migrants were a small group already lacking haplogroups A, C, G, D4, and M8-M10 that might have become frequent in southern China more recently.

Although possessing the same general haplogroup composition, the nine Taiwanese aboriginal populations are all significantly different from each other (in genetic distances), with southern populations showing a frequency pattern different from that of the central and northern groups (see Figure 3) and each population revealing its own specific founder haplotypes (see Figure 2). Given the low geographic distances between the sampling locations, this striking diversity between the tribes, seen also in HLA studies [18], may denote prevalence of strong inter-group cultural differences preserved through social isolation and endogamy of the tribes. Even though intermarriages may have been uncom-

mon between the tribes, it is possible that small-scale admixture could have occurred through child adoption among the mountain tribes [55].

### Fitting the mtDNA Heritage of the Taiwanese Aborigines into the Models of Austronesian Expansion

Genetic models explaining the origins of Austronesian migration can be categorized by the three following components: the place of origin, the time scale, and the correspondence with specific archaeological evidence. Studies based on Y chromosomal and mitochondrial markers have traced in Polynesians the presence of lineages that are characteristic of Melaneseans and East Indonesians but which are absent in mainland East Asia and Taiwan [8,16,25,56]. Evidence for interaction or initiation of a human settlement process, directed toward Remote Oceania, somewhere in East Indonesia or Melanesia is also supported by the phylogenetic analyses of Polynesian rats, *Rattus exulans* [57]. Besides the Melanesian-specific component, both human mtDNA and Y chromosomal haplogroups found in Polynesians include those that are common in populations of mainland East Asia and also in Taiwan [22]. This general share of ancestry refers, first of all, to mtDNA haplogroup B4a, which is the most frequent lineage group among Polynesians. However, the specific HVS-I motif associated with Polynesian expansion occurs only in Near and Remote Oceania whereas its immediate ancestral sequence is common throughout East

Asia and has a coalescence time significantly predating the Polynesian migration [21].

Phylogenetic analysis of complete mtDNA sequences (Figure 4) in this study reveals the presence of a motif of three coding region mutations (nps 6719, 12239, and 15746) that define haplogroup B4a1a and are shared among aboriginal Taiwanese, Melanians, and Polynesians. No mainland East Asian population has yet been found to carry lineages derived from these three positions. This suggests that the motif may have evolved in populations living in or near Taiwan at the end of the Late Pleistocene period. Considering the differences between the Late Pleistocene and present-day shorelines of Southeast Asia, the B4a1 lineages may also have evolved in regions now submerged under the sea. As long as mtDNA lineages of the Philippines and Indonesia, in particular, have not been analyzed in similar detail, the question about the precise origin of B4a1a has to be left open. The presence of an additional coding region mutation (np 14022) that defines haplogroup B4a1a1 and the HVS-I transition at np 16247 in Melanians and Polynesians points to a maturation phase of the Austronesian migration in East Indonesia or Melanesia, during which the final Polynesian motif of mutations was established.

Austronesian is the world's most widely distributed language family, yet nine of its ten subfamilies are restricted to Taiwanese aboriginal populations [2]. The phylogeny of B4a1a mitochondrial genomes resembles the linguistic reconstruction of Austronesian languages with five of its primary branches restricted to Taiwan and the sixth branch spread all over Oceania. The  $9.3 \pm 2.6$  thousand-year-old coalescence date we obtained using only coding region information for B4a1a1 diversification in Papuans and Polynesians predates significantly the earliest signs of Lapita culture in the region (around 3,500 BP). Nonetheless, these findings provide the first direct phylogenetic evidence for the common ancestry of Austronesian and indigenous Taiwanese maternal lineages and their maturation phase in East Indonesia or Melanesia. High frequency of a tagging mitochondrial haplotype in contrast to the absence of such in the Y chromosomes of Polynesians and Taiwanese might lend credence to the suggestion that the proto-Austronesian communities were matriarchal [58] and matrilineal (as the Amis tribe still is in Taiwan) whereby the Y chromosome pool of the initial migrants was lost after being repeatedly diluted on the way toward Polynesia.

## Conclusions

Taiwanese aboriginal populations share their maternal ancestry with populations of mainland East Asia through haplogroups B, R9, and M7 as their main genetic components. At the same time the haplogroup structure at a finer phylogenetic resolution suggests relatively long-term isolation from the mainland populations. The coalescence times of B4a1a, F3b, F4b, R9c, and M7c1c lineages point to founder effects in Taiwan ranging from recent (0–2,000 years) to more ancient times (7,000–20,000 years). These results most likely reflect the drift in small endogamous populations of the island that became isolated by the rising sea levels after the last Ice Age.

The time element ( $13.2 \pm 3.8$  thousand years to the MRCA) obtained from the phylogenetic reconstruction of complete B4a1a sequences requires that we adopt a model

according to which the origin of Austronesian migration can be traced back to Taiwan, and allows for the notion that it was followed by interaction periods elsewhere in Indonesia and finally in Melanesia where the complete motif specific to Polynesian B4a1a1 sequences (Polynesian motif) was developed.

## Materials and Methods

In this study the sequence variation of the mtDNA D-loop HVS-I region, nps 16006–16397 were characterized in 640 samples drawn from nine Taiwan indigenous mountain tribes representative of most languages, cultures, and geographical settlements seen on the island before the last four centuries. In addition, coding region (CR) nps 9959–10917 were sequenced and the mtDNA nine-bp deletion in the mtDNA intergenic region between the COII gene and the Lysine tRNA gene (COII/tRNA<sup>LYS</sup>) was screened by sequencing specific polymorphism (SSP) with primer pair L8215/H8297 [34]. Finally, significant RFLPs were used to remove unavoidable errors in haplogroup designation. The following tribes were selected: in the north of Taiwan, the Atayal ( $n = 109$ ) and the Saisiat ( $n = 63$ ); in the central mountain ranges, the Tsou ( $n = 60$ ) and the Bunun ( $n = 89$ ); in the east, the Amis ( $n = 98$ ) and the Yami ( $n = 64$ ); and in the south, the Rukai ( $n = 50$ ), the Paiwan ( $n = 55$ ), and the Puyuma ( $n = 52$ ) (see Figure 1). All indigenous people had both parents belonging to the same tribe, and gave consent to participation in this study.

**DNA extraction.** Blood samples were collected in ACD tubes and treated at the Transfusion Medicine Research Laboratory of the Mackay Memorial Hospital in Taipei. Genomic DNA was extracted from 500  $\mu$ l of buffy coat using the QIAmp DNA kit (QIAml blood kit, Qiagen inc. Chatsworth, California, United States) with minor modifications to the procedure recommended by the manufacturer.

**DNA sequence analysis.** Using primers L15997 and H16401 [59], a segment of 404 nucleotides from the D-loop HVS-I of the mtDNA was obtained. Primer pairs 5, 7, 8, 11–14, 19, and 24 F&R described in Rieder et al. [60] were used for typing informative RFLP sites using the following restriction enzymes: *AluI* (nps 5176 and 13262), *BspEI* (np 4710), *HaeIII* (np 663, 3391, and 8391), *HhaI* (nps 4830, 7598, and 9053), *HinfI* (np 9820), *HpaI* (np 12405), *MboII* (np 12704), and *Tsp509I* (np 5416) [34,61]. Primers 15F [60] and H10917 (5' gAACAgCTAAA-TAggTTg 3') were used to amplify a segment of 959 nucleotides used for sequencing both CR strands (nps 9959 to 10917). Length variation in the COII/tRNA<sup>LYS</sup> (nine-bp deletion) region was assayed by PCR-SSP method, using primer pair L8215/H3274 [34]. PCR was performed in 20  $\mu$ l reactions and carried out for 36 cycles using the following final conditions: 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 0.2 mM deoxyribonucleotide triphosphate each, 2 mM MgCl<sub>2</sub>, 0.25 U recombinant Taq DNA polymerase (Life Technologies, Taiwan), 5% glycerol, cresol red 0.075 mg/ml, 0.2 mM of each primer, and 30–100 ng of DNA sample. Thermal cycling conditions were 1 min. at 95 °C followed by 36 cycles of 95 °C for 10 s, 60 °C for 30 s, and 72 °C for 30 s. Following the thermal cycling, the samples were maintained at 75 °C for 10 min. Prior to sequencing, the amplified products were separated from excess dNTP and primers by first pre-treating with shrimp alkaline phosphatase (Sap) and Exo I enzymes (USB Product number US 70995 pre-sequencing Kit, Pharmacia, Taiwan) following the conditions recommended by the manufacturer (37 °C for 30 min and 85 °C for 15 min). For each individual, sequencing of nps 16006 to 16397 was performed on both strands using the Perkin-Elmer/Applied Biosystems Division (ABI Taiwan) DyeDeoxy Terminator Cycle Sequencing Kit (Foster City, California, United States) according to the recommendations of the manufacturer. Purification on a G50 sephadex column was performed before the final run on an automated DNA sequencer (ABI Model 377). In our inter-population analysis with other studies, the region including nps 16182 to 16183 was excluded because an upstream adenosine homo-polymeric region may cause poor sequencing. Consequently substitutions at nps 16182 and 16183 were not indicated in Figure 2.

**Data analysis.** Indices of molecular variance were calculated in ARLEQUIN package 2.0 [62]. A neighbor-joining tree was constructed with 1,000 bootstrap replicas using MEGA2 package [63]. A PC map was constructed with ADE-4 package [64–66] using HVS-I haplogroups frequencies obtained from the nine indigenous tribes of Taiwan and other regions of Asia shown in table 1. Coalescence times were calculated using the  $\rho$  statistic and HVS-I mutation rate of one transition per 20,180 years in bps 16090–16365 [67,68]. For complete coding region sequences, a mutation rate of  $1.7 \times 10^{-8}$  substitutions

per site per million years was used assuming 5 million years for the split between human and chimpanzee lineages [69]. A more conservative rate calibration  $1.26 \times 10^{-8}$  [54], assuming a 6.5 million-year-old split of human and chimpanzee mtDNA lineages would imply that all the coalescence times, based on complete sequence data, that are discussed in the text would be approximately 1.3 times older.

## Supporting Information

### Figure S1. Ninety-Six Haplotypes Seen in 640 Taiwan Indigenous Individuals

For HVS-I, HVS II, and nps 9793–10899, sequences are numbered according to the revised reference sequence [70,71]. Only transversions are specified and the number in bracket indicates that only this region has been sequenced. Restriction enzymes have sometimes been used instead of sequencing to infer nps 9824, 10238, 10310, 10320, 10397, 10398, and 10400 (HinfI, HphI, BspMI, Tsp509I, BsrI, DdeI, and AluI, respectively). Significant diagnostic restriction enzymes are indicated at the top of each RFLP column with losses or gains of RFLP sites shown as plus or minus “9-bp del” refers to the length variation between nps 8271 and 8280 in the COII/tRNA<sup>LYS</sup> region as assayed by PCR-SSP method [34] with 1 denoting presence of the nine-bp deletion and 2 denoting non-deletion. The colored columns represent lineage distribution among populations; colors are region specific and are conserved in Figure 2. The minimum evolution tree was constructed with Mega2 [63] using Tamura-Nei

pairwise distances [72]. Specimens sequenced from 9982 to 10838 only are annotated with a number sign.

Found at DOI: 10.1371/journal.pbio.0030247.sg001 (2.9 MB TIF).

### Accession Numbers

The GenBank accession numbers (<http://www.ebi.ac.uk/embl/index.html>) for HVS-I and HVS-II data in this article are as follows: HVS-I (AJ967036–AJ967675) and HVS-II (AJ967676–AJ968315). Complete sequence data: accession numbers AJ842744–AJ842751.

## Acknowledgments

We are grateful to all Taiwan indigenous people who participated in this project. We would like to acknowledge Richard Villemes (Tartu University), Victor Mair, Peter Underhill, Koji Lum, Qing-Peng Kong, and Hans-Jürgen Bandelt for corrections and comments on the manuscript. The project was supported by grant No 90–2320-B-195–002 from the National Science Council of Taiwan and grant NHRI-EX93-9218BI from the National Health Research Institute. TK received support from Estonian Science Foundation grant 5574.

**Competing interests.** The authors have declared that no competing interests exist.

**Author contributions.** JAT and ML conceived and designed the study. ML contributed DNA samples. JHL, CJH, CLL, CLH, and ZYL performed the sequence analysis. JAT and TK performed population genetic analyses. JAT, TK, and ML wrote the paper. ■

## References

- Chiang WVT (2001) Language attitudes toward written Taiwanese. *J Multiling Multicult Devel* 22: 502–523.
- Blust R (1999) Subgrouping, circularity and extinction: Some issues in Austronesian comparative linguistics. *Symp Ser Inst Linguist Acad Sinica* 1: 31–94.
- Rolett BV, Tsang CH (2001) The Austronesian homeland: New evidence for early voyaging. In: *International Symposium on Austronesian cultures: Issues Relating to Taiwan* (Archeology). Dec 8–11. Taipei (Taiwan): Academia Sinica. pp 1–2
- 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. pp 47–52
- 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. pp 161–176
- Sagart L (2005) The higher phylogeny of Austronesian and the position of Tai-Kadai. *Oceanic Linguist* 43: 411–444.
- Diamond J, Bellwood P (2003) Farmers and their languages: The first expansions. *Science* 300: 597–603.
- Oppenheimer SJ, Richards M (2001) Polynesian origins. *Slow boat to Melanesia?* *Nature* 410: 166–167.
- Tajima A, Sun CS, Pan IH, Ishida T, Saitou N, et al. (2003) Mitochondrial DNA polymorphisms in nine aboriginal groups of Taiwan: Implications for the population history of aboriginal Taiwanese. *Hum Genet* 113: 24–33.
- Lum JK, Rickards O, Ching C, Cann RL (1994) Polynesian mitochondrial DNAs reveal three deep maternal lineage clusters. *Hum Biol* 66: 567–590.
- Ballinger SW, Schurr TG, Torroni A, Gan YY, Hodge JA, et al. (1992) Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient mongoloid migrations. *Genetics* 130: 139–152.
- Sykes B, Leifoff 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.
- Melton T, Peterson R, Redd AJ, Saha N, Sofro AS, et al. (1995) Polynesian genetic affinities with Southeast Asian populations as identified by mtDNA analysis. *Am J Hum Genet* 57: 403–414.
- Redd AJ, Takezaki N, Sherry ST, McGarvey ST, Sofro AS, et al. (1995) Evolutionary history of the COII/tRNA<sup>LYS</sup> intergenic 9 base pair deletion in human mitochondrial DNAs from the Pacific. *Mol Biol Evol* 12: 604–615.
- Su B, Xiao J, Underhill P, Dekar R, Zhang W, et al. (1999) Y-Chromosome evidence for a northward migration of modern humans into Eastern Asia during the last Ice Age. *Am J Hum Genet* 65: 1718–1724.
- Kayser M, Brauer S, Weiss G, Underhill PA, Roewer L, et al. (2000) Melanesian origin of Polynesian Y chromosomes. *Curr Biol* 10: 1237–1246.
- Chen YF, Chen SJ, Chen CH, Chu CL, Hsu M, et al. (2001) A study of microsatellite loci on Y chromosome of Taiwan aborigines. *Academia Sinica*. 8–11 p.
- Lin M, Chu CC, Lee HL, Chang SL, Ohashi J, et al. (2000) Heterogeneity of Taiwan's indigenous population: possible relation to prehistoric Mongoloid dispersals. *Tissue Antigens* 55: 1–9.
- Chu CC, Lin M, Nakajima F, Lee HL, Chang SL, et al. (2001) Diversity of HLA among Taiwan's indigenous tribes and the Ivatans in the Philippines. *Tissue Antigens* 58: 9–18.
- Tajima A, Hayami M, Tokunaga K, Juji T, Matsuo M, et al. (2004) Genetic origins of the Ainu inferred from combined DNA analyses of maternal and paternal lineages. *J Hum Genet* 49: 187–193.
- Richards M, Oppenheimer S, Sykes B (1998) mtDNA suggests Polynesian origins in Eastern Indonesia. *Am J Hum Genet* 63: 1234–1236.
- Hurles ME, Matisoo-Smith E, D. GR, Penny D (2003) Untangling Oceanic settlement: The edge of the knowable. *Trends Ecol Evol* 18: 531–540.
- Diamond JM (1988) Express train to Polynesia. *Nature* 336: 307–308.
- Terrell JE, Kelly KM, Rainbird P (2001) Foregone conclusions? In search of “Papuans” and “Austronesians.” *Curr Anthropol* 42: 97–124.
- Gibbons A (2001) The Peopling of the Pacific. Archaeologists, linguists, and geneticists struggle to understand the origins of the bold seafarers who settled the remote Pacific Islands. *Science* 291: 1735–1737.
- Bellwood P (1991) The Austronesian dispersal and the origin of languages. *Sci Am* 265: 88–93.
- Oppenheimer S (1999) Eden in the East: The drowned continent of the Southeast Asia. London: Trafalgar Square Publishing. 560 p.
- Melton T, Clifford S, Martinson J, Batzer M, Stoneking M (1998) Genetic evidence for the proto-Austronesian homeland in Asia: mtDNA and nuclear DNA variation in Taiwanese aboriginal tribes. *Am J Hum Genet* 63: 1807–1823.
- Tsai LC, Lin CY, Lee JC, Chang JG, Linacre A, et al. (2001) Sequence polymorphism of mitochondrial D-loop DNA in the Taiwanese Han population. *Forensic Sci Int* 119: 239–247.
- Kivisild T, Tolk HV, Parik J, Wang Y, Papiha SS, et al. (2002) The emerging limbs and twigs of the East Asian mtDNA tree. *Mol Biol Evol* 19: 1737–1751.
- Wen B, Xie X, Gao S, Li H, Shi H, et al. (2004) Analyses of genetic structure of Tibeto-Burman populations reveals sex-biased admixture in southern Tibeto-Burmans. *Am J Hum Genet* 74: 856–865.
- Kolman C, Sambuughin N, Bermingham E (1996) Mitochondrial DNA analysis of Mongolian populations and implications for the origin of New World founders. *Genetics* 142: 1321–1334.
- Lee S, Shin C, Kim K, Lee Y, Lee J (1997) Sequence variation of mitochondrial DNA control region in Koreans. *Forensic Sci Int* 87: 99–116.
- 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, Zhang YP (2002) Phylogeographic analysis of mtDNA variation in four ethnic populations from Yunnan Province: New data and a reappraisal. *J Hum Genet* 47: 311–318.
- Torroni A, Miller JA, Moore LG, Zamudio S, Zhuang J, et al. (1994) Mitochondrial DNA analysis in Tibet: Implications for the origin of the Tibetan population and its adaptation to high altitude. *Am J Phys Anthropol* 93: 189–199.
- Kong QP, Yao YG, Sun C, Bandelt HJ, Zhu CL, et al. (2003) Phylogeny of East Asian mitochondrial DNA lineages inferred from complete sequences. *Am J Hum Genet* 73: 671–676.
- Tanaka M, Cabrera VM, Gonzalez AM, Larruga JM, Takeyasu T, et al. (2004) Mitochondrial genome variation in eastern Asia and the peopling of Japan. *Genome Res* 14: 1832–1850.

39. Ingman M, Gyllensten U (2003) Mitochondrial genome variation and evolutionary history of Australian and New Guinean aborigines. *Genome Res* 13: 1600–1606.
40. Nishimaki Y, Sato K, Fang L, Ma M, Hasekura H, et al. (1999) Sequence polymorphism in the mtDNA HV1 region in Japanese and Chinese. *Legal Med* 1: 238–249.
41. Thangaraj K, Sridhar V, Kivisild T, Reddy AG, Chaubey G, et al. (2005) Different population histories of the Mundari- and Mon-Khmer-speaking Austro-Asiatic tribes inferred from the mtDNA 9-bp deletion/insertion polymorphism in Indian populations. *Hum Genet* 116: 507–517.
42. Horai S, Murayama K, Hayasaka K, Matsubayashi S, Hattori Y, et al. (1996) mtDNA polymorphism in East Asian Populations, with special reference to the peopling of Japan. *Am J Hum Genet* 59: 579–590.
43. Yao YG, Nie L, Harpending H, Fu Y, Yuan ZG, et al. (2002) Genetic relationship of Chinese ethnic populations revealed by mtDNA sequence diversity. *Am J Phys Anthropol* 118: 63–76.
44. Kong QP, Yao YG, Sun C, Zhu CL, Zhong L, et al. (2004) Phylogeographic analysis of mitochondrial DNA haplogroup F2 in China reveals T12338C in the initiation codon of the ND5 gene not to be pathogenic. *J Hum Genet* 49: 414–423.
45. Oota H, Settheetham-Ishida W, Tiwawech D, Ishida T, Stoneking M (2001) Human mtDNA and Y-chromosome variation is correlated with matrilineal versus patrilineal residence. *Nat Genet* 29: 20–21.
46. Fucharoen G, Fucharoen S, Horai S (2001) Mitochondrial DNA polymorphisms in Thailand. *J Hum Genet* 46: 115–125.
47. 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: 2265–2280.
48. Comas D, Calafell F, Mateu E, Perez-Lezaun A, Bosch E, et al. (1998) Trading genes along the Silk Road: mtDNA sequences and the origin of Central Asian populations. *Am J Hum Genet* 63: 1824–1838.
49. Ozawa T (1995) Mechanism of somatic mitochondrial DNA mutations associated with age and diseases. *Biochim Biophys Acta* 1271: 177–189.
50. Wen B, Li H, Gao S, Mao X, Gao Y, et al. (2005) Genetic structure of Hmong-Mien speaking populations in East Asia as revealed by mtDNA lineages. *Mol Biol Evol* 22: 725–734.
51. Oota H, Kitano T, Jin F, Yuasa I, Wang L, et al. (2002) Extreme mtDNA homogeneity in continental Asian populations. *Am J Phys Anthropol* 118: 146–153.
52. Qian YP, Chu ZT, Dai Q, Wei CD, Chu JY, et al. (2001) Mitochondrial DNA polymorphisms in Yunnan nationalities in China. *J Hum Genet* 46: 211–220.
53. Lum JK, Cann RL (2000) mtDNA lineage analyses: Origins and migrations of Micronesians and Polynesians. *Am J Phys Anthropol* 113: 151–168.
54. Mishmar D, Ruiz-Pesini E, Golik P, Macaulay V, Clark AG, et al. (2003) Natural selection shaped regional mtDNA variation in humans. *Proc Natl Acad Sci USA* 100: 171–176.
55. Keliher M (2003) Out of China. Yu Yonghe's tales of Formosa A history of 17th century Taiwan. Taipei (Taiwan): SMC Publishing. 242 p.
56. Hurles ME, Nicholson J, Bosch E, Renfrew C, Sykes BC, et al. (2002) Y-chromosomal evidence for the origins of oceanic-speaking peoples. *Genetics* 160: 289–303.
57. Matisoo-Smith E, Robins JH (2004) Origins and dispersals of Pacific peoples: Evidence from mtDNA phylogenies of the Pacific rat. *Proc Natl Acad Sci U S A* 101: 9167–9172.
58. Hage P, Marck J (2003) Matrilineality and the Melanesian origin of Polynesian Y chromosomes. *Curr Anthropol* 44: S121–S126.
59. Wilson MR, Polansky D, Butler J, Dizinno JA, Replogle J, et al. (1995) Extraction, PCR amplification and sequencing of mitochondrial DNA from human hair shafts. *Biotechniques* 18: 662–669.
60. Rieder MJ, Taylor SL, Tobe VO, Nickerson DA (1998) Automating the identification of DNA variations using quality-based fluorescence resequencing: Analysis of the human mitochondrial genome. *Nucleic Acids Res* 26: 967–973.
61. Forster P, Torroni A, Renfrew C, Röhl A (2001) Phylogenetic star contraction applied to Asian and Papuan mtDNA evolution. *Mol Biol Evol* 18: 1864–1881.
62. Schneider S, Roessli D, Excoffier L (2000) Arlequin version 2.000: A software for population genetics data analysis. Geneva: University of Geneva, Genetics and Biochemistry laboratory.
63. Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: Molecular evolutionary genetics analysis software. *Bioinformatics* 17: 1244–1245.
64. Thioulouse J, Chessel D, Doledec S, Olivier JM (1997) ADE-4: A multivariate analysis and graphical display software. *Statistic and Computing* 7: 75–83.
65. Benzécri JP (1969) Statistical analysis as a tool to make patterns emerge from data. In: Watanabe S, editor. *Methodologies of pattern recognition*. New York: Academic Press. pp. 35–60
66. Greenacre MJ, Hastie T (1987). The geometric interpretation of correspondence analysis. *J Amer Statist Assoc* 82: 437–447
67. 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.
68. Saillard J, Forster P, Lynnerup N, Bandelt H-J, Nørby S (2000) mtDNA variation among Greenland Eskimos: The edge of the Beringian expansion. *Am J Hum Genet* 67: 718–726.
69. Ingman M, Kaessmann H, Pääbo S, Gyllensten U (2000) Mitochondrial genome variation and the origin of modern humans. *Nature* 408: 708–713.
70. Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, et al. (1981) Sequence and organization of the human mitochondrial genome. *Nature* 290: 457–465.
71. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, et al. (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 23: 147.
72. Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol* 10: 512–526.
73. Seo Y, Stradmann-Bellinghausen B, Rittner C, Takahama K, Schneider PM (1998) Sequence polymorphism of mitochondrial DNA control region in Japanese. *Forensic Sci Int* 97: 155–164.

## Correction: Highly Nonrandom Features of Synaptic Connectivity in Local Cortical Circuits

Sen Song, Per Jesper Sjöström, Markus Reigl, Sacha Nelson, Dmitri B. Chklovskii

DOI: 10.1371/journal.pbio.0030068

In *PLoS Biology*, volume 3, issue 3.

In the caption for Figure 5A, the formula for the lognormal fit was given as follows:

$$p[w] = 0.426 \exp[-(\ln[w] + 0.702)^2 / (2 \times 0.9355^2)] / w.$$

The second squaring exponent should have been inside the parentheses, so it reads as follows:

$$p[w] = 0.426 \exp[-(\ln[w] + 0.702)^2 / (2 \times 0.9355^2)] / w.$$

This correction note may be found online at DOI: 10.1371/journal.pbio.0030350.

Published October 11, 2005

Citation: (2005) Correction: Highly nonrandom features of synaptic connectivity in local cortical circuits.

PLoS Biol 3(10): e350.

## Correction: Directed Migration of Positively Selected Thymocytes Visualized in Real Time

Colleen M. Witt, Subhadip Raychaudhuri, Brian Schaefer, Arup K. Chakraborty, Ellen A. Robey

DOI: 10.1371/journal.pbio.0030160

In *PLoS Biology*, volume 3, issue 6.

The following statement in the last paragraph of the Results contains a factual error: "A total of 34% of cortical thymocytes expressing the P14 TCR had motility rates greater than 13  $\mu\text{m}/\text{min}$  compared to approximately 2% for wild-type thymocytes expressing diverse TCRs." The sentence should read as follows: "A total of 20% of cortical thymocytes expressing the P14 TCR had motility rates greater than 13  $\mu\text{m}/\text{min}$  compared to approximately 2% for wild-type thymocytes expressing diverse TCRs."

The authors apologize for any confusion this error may have caused.

This correction note may be found online at DOI: 10.1371/journal.pbio.0030373.

Published October 11, 2005

Citation: (2005) Correction: Directed migration of positively selected thymocytes visualized in real time. *PLoS Biol* 3(10): e373.

## Correction: Traces of Archaic Mitochondrial Lineages Persist in Austronesian-Speaking Formosan Populations

Jean A. Trejaut, Toomas Kivisild, Jun Hun Loo, Chien Liang Lee, Chun Lin He, Chia Jung Hsu, Zheng Yuan Lee, Marie Lin

DOI: 10.1371/journal.pbio.0030247

In *PLoS Biology*, volume 3, issue 8.

The seventh author's name was originally misspelled as Zheng Yuan Li; it should be Zheng Yuan Lee.

The first sentence of the Materials and Methods should read as follows:

"In this study the sequence variation of the mtDNA D-loop HVS-I and HVS-II regions, nps 16006 to 16397 and nps 53 to 404, respectively, was analyzed in 640 samples drawn from nine Taiwan indigenous mountain tribes representative of most languages, cultures, and geographical settlements seen on the island before the last four centuries."

The first two sentences in the section "Data Sequence Analysis" in the Materials and Methods should read as follows:

"Using primer pairs L15997–H16401 and L048–H408 [59], segments of 404 bps and 360 bps from the D-loop HVS-I and HVS-II, respectively, were obtained. Primer pairs 5, 7, 8, 11–14, 19, and 24 F&R described in Rieder et al. [60]...."

This correction note may be found online at DOI: 10.1371/journal.pbio.0030376.

Published October 11, 2005

Citation: (2005) Correction: Traces of archaic mitochondrial lineages persist in Austronesian-speaking Formosan populations. *PLoS Biol* 3(10): e376.