npg

ORIGINAL ARTICLE

The Indonesian archipelago: an ancient genetic highway linking Asia and the Pacific

Meryanne K Tumonggor[1,2], Tatiana M Karafet[3], Brian Hallmark[3], J Stephen Lansing[1,4], Herawati Sudoyo[2], Michael F Hammer[3] and Murray P Cox[5]

Indonesia, an island nation linking mainland Asia with the Pacific world, hosts a wide range of linguistic, ethnic and genetic diversity. Despite the complexity of this cultural environment, genetic studies in Indonesia remain surprisingly sparse. Here, we report mitochondrial DNA (mtDNA) and associated Y-chromosome diversity for the largest cohort of Indonesians examined to date—2740 individuals from 70 communities spanning 12 islands across the breadth of the Indonesian archipelago. We reconstruct 50 000 years of population movements, from mitochondrial lineages reflecting the very earliest settlers in island southeast Asia, to Neolithic population dispersals. Historic contacts from Chinese, Indians, Arabs and Europeans comprise a noticeable fraction of Y-chromosome variation, but are not reflected in the maternally inherited mtDNA. While this historic immigration favored men, patterns of genetic diversity show that women moved more widely in earlier times. However, measures of population differentiation signal that Indonesian communities are trending away from the matri- or ambilocality of early Austronesian societies toward the more common practice of patrilocal residence today. Such sex-specific dispersal patterns remain even after correcting for the different mutation rates of mtDNA and the Y chromosome. This detailed palimpsest of Indonesian genetic diversity is a direct outcome of the region's complex history of immigration, transitory migrants and populations that have endured *in situ* since the region's first settlement.
Journal of Human Genetics (2013) **58**, 165–173; doi:10.1038/jhg.2012.154; published online 24 January 2013

INTRODUCTION

Indonesia, a maritime nation comprising over 17 000 islands straddling the Pacific and Indian Oceans, links mainland Asia with the Pacific world. Although a single lingua franca is spoken widely across the archipelago today (Bahasa Indonesia), Indonesia hosts over 730 indigenous languages and associated ethnic groups.[1] Most of these languages belong to the geographically dispersed Austronesian language family, but Papuan languages are spoken by some groups in the far east of the archipelago.[2] Relative to its land area, Indonesia is one of the most varied regions on earth in terms of ethnic, linguistic and genetic diversity.

As an island nation, past changes in global climate have had an especially strong influence on Indonesia.[3] Lands in the west, now largely submerged, once formed a vast continental shelf jutting out from Asia (Sundaland). In the east, Australia and New Guinea were linked into a single continent (Sahul). During most of the late Pleistocene, it was possible to walk—with only minor water crossings—from Bangkok to Sydney. However, following the end of the last glacial period, ∼18 kya, ice melting in the arctic fueled a rapid rise in global sea levels. Continental Indonesia swiftly fragmented into the long chain of islands that characterizes the nation today.

The human history of Indonesia played out against the backdrop of this dynamically changing geography. Settled by anatomically modern humans at least 47 kya[4,5] and perhaps much earlier,[6] our species has inhabited Indonesia longer than Europe.[7,8] Archeology provides sporadic, but widespread, evidence of early hunter-gatherer groups throughout the Pleistocene.[9] However, the biggest cultural change occurred within the last 10 kya when the archeological record abruptly chronicles the appearance of agricultural communities together with pottery, plant cultivation and animal domestication. The vast spread of Austronesian languages likely also occurred during this time.[10–12] Whether the Neolithic era was ushered in by population movements from Taiwan[13] or was instead dominated by regional developments[14,15] remains a highly contentious topic of discussion—a question that genetics is increasingly striving to answer.

For such a large and ethnically diverse nation at the pivot point of Asia and the Pacific, studies of Indonesian genetic diversity are surprisingly sparse.[16–24] The most complete study of mitochondrial DNA (mtDNA) posited that Indonesian diversity has largely been shaped by two forces: population movements driven by sea level changes, and by farming populations expanding from the Asian mainland into the islands of southeast Asia.[25,26] Y-chromosome

[1]Department of Anthropology, University of Arizona, Tucson, AZ, USA; [2]Eijkman Institute for Molecular Biology, Jakarta, Indonesia; [3]Arizona Research Laboratories, Division of Biotechnology, University of Arizona, Tucson, AZ, USA; [4]Santa Fe Institute, Santa Fe, NM, USA and [5]Institute of Fundamental Sciences, Massey University, Palmerston North, New Zealand
Correspondence: Professor MF Hammer, Arizona Research Laboratories, Division of Biotechnology, University of Arizona, Tucson, AZ 85721, USA.
E-mail: mfh@email.arizona.edu

studies, reflecting the history of Indonesian men, mirror these themes, emphasizing a complex multifaceted history of the region's islands and communities.[20] Although autosomal markers are rapidly changing our understanding of regional prehistory,[27–29] the geographical resolution and widespread availability of comparative data ensure that haploid markers remain extremely powerful tools for studying the human past. Here, we present the largest sample of Indonesian mtDNA diversity assembled to date—2740 individuals from 70 populations on 12 islands spanning the full range of Indonesia's geographical, ethnic and linguistic diversity. We reconstruct a partial history of Indonesia's women, piecing together 50 ky of population movements that have shaped the diversity of Indonesians living today. Finally, by comparison with Y-chromosome data for the same set of individuals,[20] we contrast the histories of Indonesian men and women, presenting common patterns of shared inheritance with key points of demographic and social difference.

MATERIALS AND METHODS

Samples

Genetic diversity was screened in 2740 consenting, closely unrelated and seemingly healthy individuals drawn from across the Indonesian archipelago. Permission to conduct research in Indonesia was granted by the Indonesian Institute of Sciences. Biological samples (peripheral blood and buccal swabs) were obtained by MKT, JSL, HS, Golfiani Malik, Wuryantari Setiadi and Loa Helena Suryadi of the Eijkman Institute for Molecular Biology, Jakarta, Indonesia, with the assistance of Indonesian Public Health clinic staff, and followed protocols for the protection of human subjects established by both the Eijkman Institute and the University of Arizona Institutional Review Boards. Participant interviews confirmed ethnic, linguistic and geographic classifications for at least two generations into the past.

Seventy populations from 12 island groups were sampled across the Indonesian archipelago, including (from west to east) Sumatra ($n = 42$), Mentawai ($n = 128$), Nias ($n = 59$), Java ($n = 51$), Bali ($n = 487$), Sulawesi ($n = 200$), Sumba ($n = 634$), Flores ($n = 469$), Lembata ($n = 92$), Pantar ($n = 29$), Timor ($n = 526$) and Alor ($n = 23$) (Supplementary Table S1). For comparative purposes, we also explored previously published genetic data from mainland Asian populations: northwestern Chinese (Yili, Xinjiang; $n = 47$), northeastern Chinese (Fencheng, Liaoning; $n = 51$), southwestern Chinese (Kunming, Yunnan; $n = 43$) and southeastern Chinese (Zhanjiang, Guangdong; $n = 30$),[30] Thai ($n = 52$) and Vietnamese ($n = 41$),[31,32] and indigenous Malaysians ($n = 260$).[33] We also compared our Indonesian data with neighboring island southeast Asian populations, including indigenous Taiwanese ($n = 640$),[34] Filipinos ($n = 423$), Papua New Guineans ($n = 231$), Island Melanesians ($n = 1366$) and Micronesians ($n = 47$).[35,36] The geographical locations of these populations are illustrated in Figure 1.

DNA extraction and genetic screening

DNA was extracted from peripheral blood samples using the salting-out procedure of Miller, Dykes and Polesky.[37] DNA from buccal swabs was extracted using standard phenol–chloroform protocols.

The first hypervariable segment (HVS I) of mtDNA was amplified using primers L15926 (5′-TCAAAGCTTACACCAGTCTTGTAAACC-3′) and H639 (5′-GGGTGATGTGAGCCCGTCA-3′). PCR amplicons were sequenced in both forward and reverse directions using primers L15965 (5′-CAAGGACAAAT CAGAGAA-3′) and H11 (5′-GTGGTTAATAGGGTGATAG-3′). Traditional Sanger sequences were aligned and edited with Sequencher v. 5.0 (Gene Codes Corporation, Ann Arbor, MI, USA; http://www.genecodes.com). Polymorphisms were scored relative to the revised Cambridge reference sequence.[38]

Haplogroups were initially predicted from HVS I sequences using known reference genomes.[23,25,36,39–42] Assignments were subsequently confirmed using Taqman and restriction fragment length polymorphism assays (Supplementary Table S2).
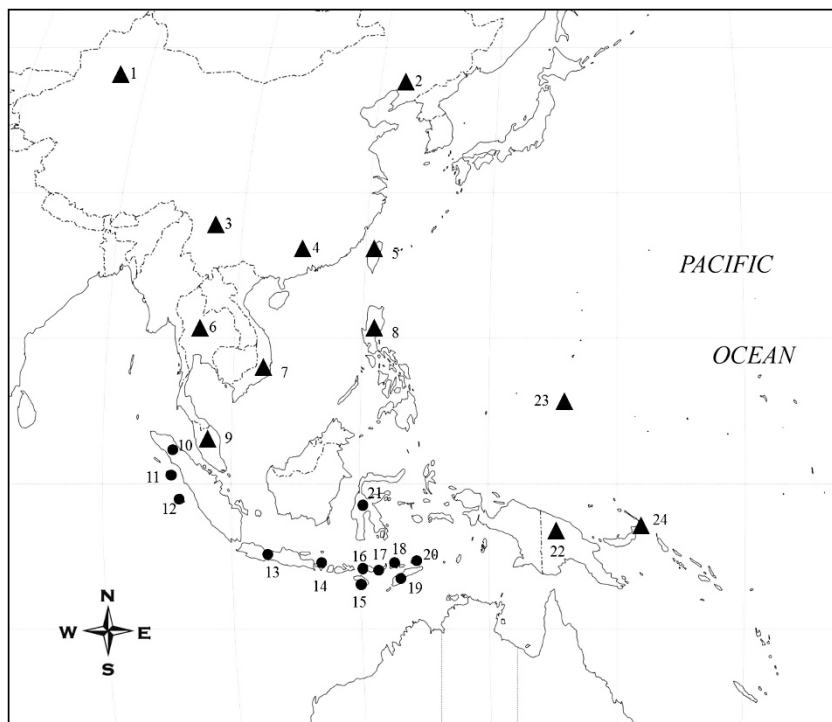
Figure 1 Locations of Studied Populations. (1) Northwest China,[30] (2) northeast China,[30] (3) southwest China,[30] (4) southeast China,[30] (5) Indigenous Taiwanese,[34] (6) Thailand,[31] (7) Vietnam,[32] (8) Philippines,[36] (9) Indigenous Malaysians,[33] (10) Sumatra (present study), (11) Nias (present study), (12) Mentawai (present study), (13) Java (present study), (14) Bali (present study), (15) Sumba (present study), (16) Flores (present study), (17) Lembata (present study), (18) Pantar (present study), (19) Timor (present study), (20) Alor (present study), (21) Sulawesi (present study), (22) Papua New Guinea,[35] (23) Micronesia[35] and (24) Melanesia.[35]

Table 1 Molecular diversity indices and growth summary statistics for Indonesian island groups

Island	N	h	S	Diversity indices						Growth Statistics			
				Haplotype diversity	s.d.	MNPD	s.d.	Nucleotide diversity	s.d.	Tajima's D	P	Fu's Fs	P
Sumatra	42	29	48	0.973	0.013	7.90	3.75	0.015	0.008	−1.03	0.146	*−13.2*	*<0.001*
Mentawai	128	20	38	0.890	0.012	6.77	3.21	0.013	0.007	−0.10	0.504	0.500	0.632
Nias	59	24	44	0.862	0.040	7.02	3.34	0.013	0.007	−0.87	0.196	−4.54	0.104
Java	51	33	52	0.968	0.013	7.68	3.64	0.014	0.008	−1.16	0.084	*−16.1*	*<0.001*
Bali	487	129	99	0.975	0.002	8.16	3.79	0.015	0.008	−1.29	0.071	*−24.1*	*0.002*
Sulawesi	200	97	87	0.976	0.005	7.53	3.53	0.014	0.007	*−1.51*	*0.040*	*−24.5*	*0.001*
Sumba	634	159	108	0.981	0.001	7.89	3.68	0.015	0.008	*−1.40*	*0.038*	*−24.0*	*0.001*
Flores	469	149	109	0.987	0.001	8.22	3.82	0.016	0.008	*−1.44*	*0.038*	*−24.1*	*0.002*
Lembata	92	47	68	0.968	0.008	9.12	4.23	0.017	0.009	−1.04	0.154	*−20.7*	*<0.001*
Timor	526	117	96	0.955	0.005	8.72	4.03	0.016	0.008	−1.08	0.115	*−24.0*	*0.002*
Alor	23	22	45	0.996	0.014	9.12	4.36	0.017	0.009	−0.98	0.176	*−13.4*	*<0.001*
Pantar	29	22	52	0.978	0.015	10.0	4.73	0.019	0.010	−0.89	0.192	*−6.78*	*0.016*

Abbreviations: h, number of haplotypes; MNPD, mean number of pairwise differences; N, number of sequences; P, probability value; S, number of polymorphic sites.
Significant growth summary statistics are bold and italicized.

Statistical analyses

Molecular diversity, population structure estimates and genetic distances between populations were calculated using Arlequin v. 3.11[43] (http://cmpg.unibe.ch/software/arlequin3). The statistical significance of F_{ST} values was evaluated using 10^4 permutations of the data. Pairwise genetic distances between populations were computed as the linearized value, $F_{ST}/(1 - F_{ST})$.[44,45] Differences in haplogroup frequencies between eastern and western Indonesia were determined via parametric bootstrapping using code implemented in R (available from the authors on request).[46]

Median-joining networks were built using Network v. 4.5.1.6 (Fluxus Engineering; http://www.fluxus-engineering.com).[47] Haplogroups were tentatively dated with the ρ statistic method[48] using a rate of one mutation every 19 171 years.[49] Dates are only intended as a rough guide for relative haplogroup ages.[50]

MtDNA and Y-chromosome comparisons

Y-chromosome diversity has also been screened for the same panel of individuals described above.[20] Differences in mtDNA and Y-chromosome diversity between populations were analyzed using an analysis of molecular variance implemented in Arlequin. A measure of interlocus differentiation $\acute{G}_{ST}$,[51] standardized for different mutation rates, was calculated using code implemented in R (available from the authors on request).[46]

MtDNA HVS I sequences have been deposited in GenBank (accession numbers: KC113641–KC115854). Y-chromosome STR data are provided as Supplementary Data Set S1.

RESULTS

We screened 2740 individuals from 70 communities on 12 Indonesian islands. MtDNA HVS I sequences showed high levels of diversity, as measured by the number of polymorphic sites, number of haplotypes, the mean number of pairwise differences and Nei's haplotype diversity (Table 1). The haplotype diversity of communities ranged from 0.862–0.996, which indicates that most individuals within these groups carry unique mtDNA lineages. The most diverse communities are found in eastern parts of the archipelago (Sumba, Flores, Pantar and Alor), where both Asian and Papuan lineages occur side by side. The western barrier islands of Nias and Mentawai are least diverse, even when compared with other Asian populations. For instance, Nias and Mentawai are the only Indonesian populations with diversity as low as that of indigenous Taiwanese groups (0.838–0.924).[34,52]

Summary statistics such as Fu's Fs and Tajima's D can be informative about the roles of selection and demography. For mtDNA control region sequences, which seem little affected by natural selection, observed values are suggestive of low levels of growth across the archipelago, with the exception of the barrier islands, Nias and Mentawai (Table 1). Growth seems strongest in the central cluster of islands (Bali, Sulawesi, Sumba and Flores), but is less pronounced in the extreme west and east of the archipelago.

Individuals were assigned to mtDNA haplogroups using a combination of HVS I sequence motifs and single-nucleotide polymorphisms (SNPs) distributed around the coding region of the mtDNA genome. Fifty-one haplogroups were identified, with all lineages falling into macrohaplogroups M (47.05%) and N (52.95%). The 51 Indonesian haplogroups are plotted on a tree of mtDNA diversity constructed using previously published HVS I sequences and coding-region SNPs[23,25,36,39–42] (Figure 2).

Haplogroup frequencies differ between western and eastern Indonesia (Supplementary Table S3). In the west, haplogroups B5a (12%), B4c1b3 (9%) and Y2 (10.5%) are carried by a third of individuals. These haplogroups, frequent in western Indonesia, are notable by their near absence in eastern Indonesia. In the east, haplogroups F1a4 (8.7%), Q including Q1 and Q2 (7.7%), P (2.8%) and B4a1a1a (2.3%) represent nearly a quarter of individuals. Correspondingly, these haplogroups are rare or absent in western Indonesia, which is expected for lineages with strong Papuan connections (P and Q), but more surprising for lineages like the Polynesian motif (B4a1a1a). The Polynesian motif is found as far west as Bali, albeit in just two individuals (0.4%). However, it was not detected in samples from the western Indonesian islands of Java, Sumatra, Nias and Mentawai, even though this region is thought to have contributed to the settlement of Madagascar where the Polynesian motif is carried by nearly a third of individuals.[53,54] The prevalence of the Polynesian motif in Madagascar, and its absence from the island region where the inhabitants of Madagascar originated, has yet to be satisfactorily explained.

We compared the distribution of Indonesian mtDNA haplogroups with those of surrounding populations (Supplementary Table S4). Most haplogroups are shared. The deep maternal lineages M17a, M73, M47, N21, N22, R21, R22 and R23 have patchy distributions across
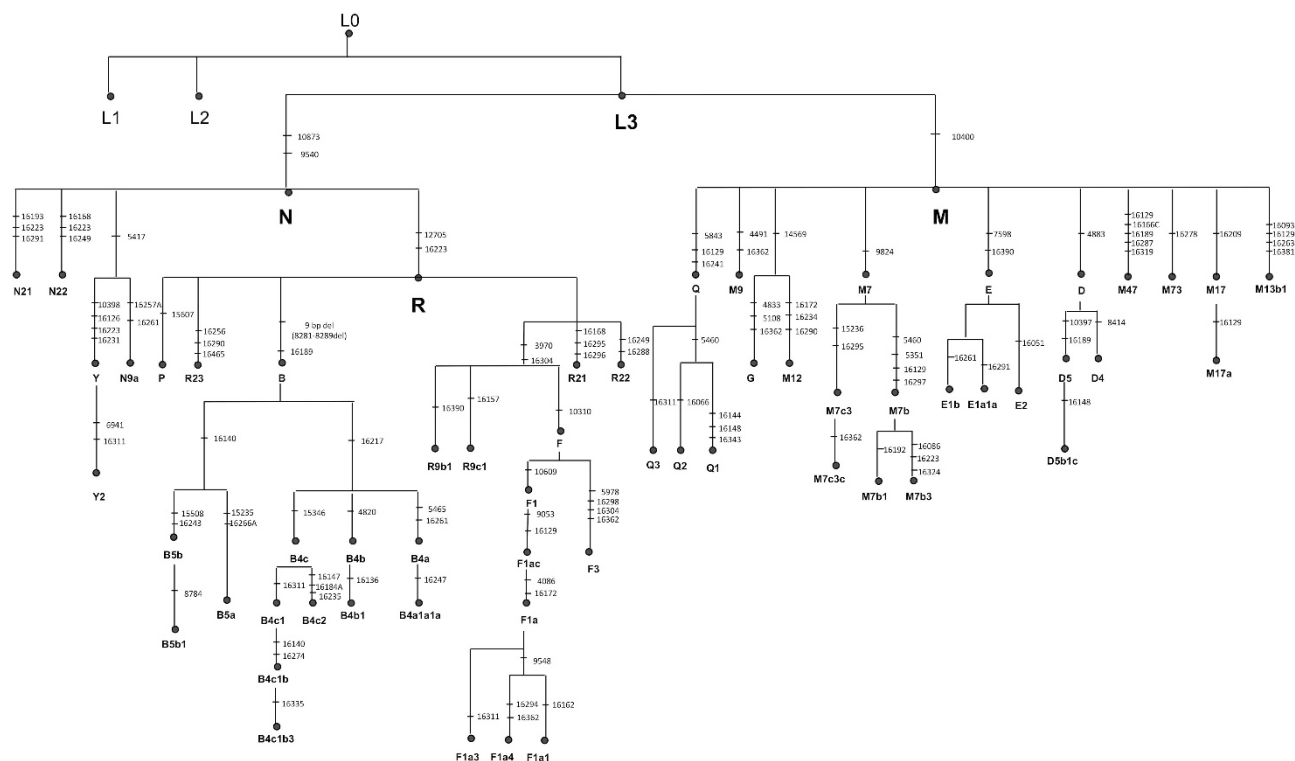
Figure 2 Phylogeny of mitochondrial DNA haplogroups observed in Indonesian populations in the present study.

mainland and island southeast Asia, likely reflecting ancient maternal lineages tracing back to the first settlers in this region.[25,33,36,55] Four of these lineages (M17a, N21, R22 and R23) reach higher frequencies in western compared with eastern Indonesia (parametric bootstrap, all $P \ll 0.01$) (Supplementary Table S3).

Owing to proposed population origins of Austronesian language speakers in Taiwan, Indonesian links to the Philippines and indigenous Taiwanese are of especial interest. These three locations share four haplogroups (E1a1a, M7b3, M7c3c and Y2), which have previously been suggested as candidates for a mid-Holocene dispersal out of Taiwan.[25,36] Figure 3 illustrates that lineages shared with Filipinos and indigenous Taiwanese are generally more common in the east than in the west of Indonesia. The exceptions are Y2 and M7c3c, where high frequencies in Nias and Mentawai may perhaps be caused by genetic drift or strong founder events in these extremely small and geographically isolated populations on the barrier islands of Sumatra.[52]

Median-joining networks were constructed for haplogroups shared between Indonesia, the Philippines and Taiwan (Figure 3). We note that at least one haplotype was shared between Indonesia and Taiwan for each of these four lineages (E1a1a, M7b3, M7c3c and Y2). In all cases, this shared lineage was the ancestral haplotype, and descendent lineages depict a star-like expansion indicative of population growth and/or geographical expansion. Unfortunately, the networks are not informative about the direction of migration: the data would fit a model of rapid expansion from Taiwan to the Philippines and Indonesia, but are equally consistent with population movements in the opposite direction.

To explore population relationships further, multidimensional scaling was performed on all Indonesian HVS I sequences using Slatkin's linearized F_{ST} as the genetic distance between groups (Supplementary Figure S1). Although there is no simple population division, western Indonesian groups cluster away from eastern Indonesian populations. Multidimensional scaling analysis with regional neighbors shows that Indonesians fall together with Asian groups, but away from Oceanian populations (Supplementary Figure S2). Among Asian populations, Indonesians cluster most closely with the Philippines and Vietnam, and more distantly with Taiwan (Supplementary Figure S3).

An analysis of molecular variance illustrates that the Y-chromosome STRs ($F_{ST} = 0.202$) have markedly higher variation among populations than mtDNA HVS I ($F_{ST} = 0.073$) (Table 2), suggesting that women have dispersed more widely in the past than men. This trend is maintained when populations are collapsed to their 12 island groups or an even broader east–west division. The trend holds, albeit more weakly, when distances are standardized for the ~ 400-fold higher mutation rate of the Y chromosome (on the order of 10^{-5} mutation events per STR per year)[56–59] relative to mtDNA (on the order of 10^{-7} mutation events per base pair per year).[49] The $\acute{G}_{ST}$ of Y-chromosome STRs ($\acute{G}_{ST} = 0.972$) is still notably higher than that of mtDNA HVS I ($\acute{G}_{ST} = 0.862$) (Supplementary Table S5). Curiously, the trend even holds on very small geographical scales, such as among communities situated along the highland river systems of Bali.[60]

DISCUSSION

Austronesian languages, spoken from Madagascar in the west to Rapanui/Easter Island in the east, form one of the world's largest and most geographically dispersed language families. Despite intense archeological, linguistic and genetic research, where people speaking these languages came from, how they dispersed and what their ancestral communities looked like remain major open questions. One hypothesis states that Austronesian speakers originated in Taiwan, which is supported by the fact that the most basal languages of the
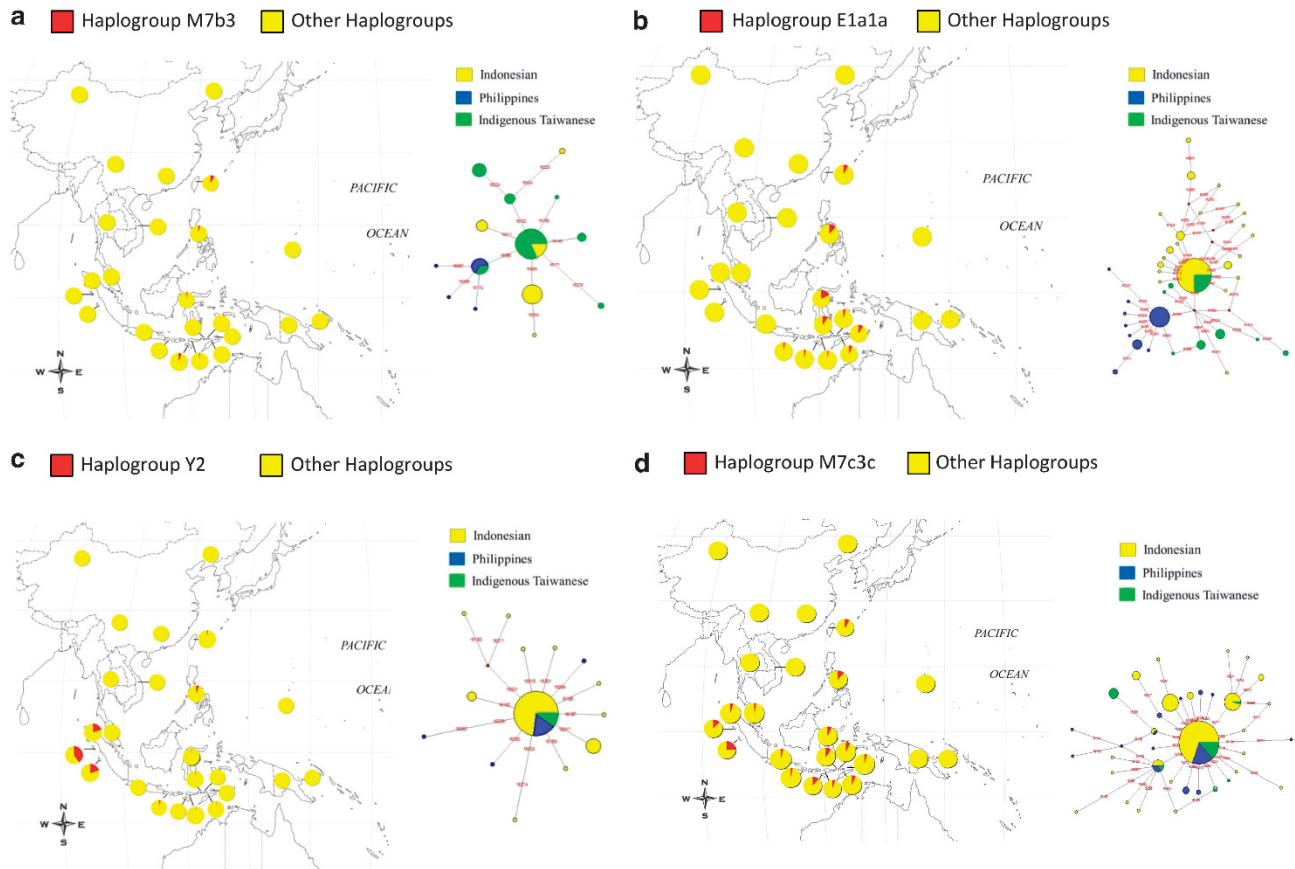
Figure 3 Frequency distributions and median-joining networks of mitochondrial DNA haplogroups (**a**) M7b3, (**b**) E1a1a, (**c**) Y2 and (**d**) M7c3c shared by indigenous Taiwanese (green), Filipino (blue) and Indonesian populations (yellow). Note that haplogroup networks are largely agnostic about the direction of population movements between Taiwan, the Philippines and Indonesia.

Table 2 Analysis of molecular variance (AMOVA) for subsets of Indonesian populations

		Within populations		Among populations within groups		Among groups	
		F_{ST}	Var (%)	F_{SC}	Var (%)	F_{CT}	Var (%)
Y-SNPs							
All Indonesian populations	70 populations	0.361	63.9				
Islands	12 islands	0.384	61.6	0.195	14.9	0.236	23.5
East vs West	2 main groups	0.475	52.5	0.247	17.2	0.304	30.3
Y-STRs							
All Indonesian populations	70 populations	0.202	79.8				
Islands	12 islands	0.210	78.9	0.152	14.2	0.069	6.90
East vs West	2 main groups	0.222	77.8	0.188	18.0	0.042	4.20
mtDNA SNPs							
All Indonesian populations	70 populations	0.092	90.8				
Islands	12 islands	0.096	90.4	0.067	6.45	0.032	3.15
East vs West	2 main groups	0.109	89.1	0.080	7.76	0.031	3.14
mtDNA HVS I							
All Indonesian populations	70 populations	0.073	92.7				
Islands	12 islands	0.077	92.3	0.048	4.61	0.031	3.09
East vs West	2 main groups	0.088	91.2	0.062	6.05	0.028	2.75

Abbreviations: HVS I, first hypervariable segment; mtDNA, mitochondrial DNA; SNP, single-nucleotide polymorphism; Var, variance.
All values are statistically significant ($P \ll 0.0001$).

Austronesian family are spoken there.[61] From Taiwan, nascent farming groups are believed to have expanded 3–4000 years ago through the Philippines, Indonesia and out into the Pacific.[3,13] Alternative hypotheses, such as Austronesian groups originating in and dispersing from Indonesia, remain possible,[14,62] with many genetic lineages in Indonesia showing old and local connections. Models combining these two extremes may ultimately be the best predictors. We envisage some genetic contributions from Taiwan, possibly including speakers of early Austronesian languages, with a substantial biological heritage from waves of ancestral populations arriving in island southeast Asia following its first settlement 50 kya.

We consider how the mitochondrial data aligns with this spectrum of origin models. Some older lineages appear to trace back to the very earliest settlers in southeast Asia. R21, which is found only in Mentawai (0.8%), diverged from the common haplogroup R ancestor ~60 kya.[55,63] Other basal mtDNA lineages (for example, N21 and N22) are shared at extremely low levels by Indonesians and mainland southeast Asian groups.

Candidate mtDNA lineages of a Taiwanese dispersal (E1a1a, M7b3, M7c3c and Y2) have also been proposed.[25,36] These four haplogroups have similar distributions, with basal haplotypes shared between indigenous Taiwanese, Filipinos and Indonesians. However, the direction of dispersal is inconclusive: ancestral-derived haplotype orders are consistent with a rapid expansion from Taiwan to the Philippines and Indonesia, but population dispersals in the opposite direction are equally likely. Although we provide haplogroup dates with some reluctance,[50] we note that relative ages are inconsistent with a simple dispersal from Taiwan to the Philippines, and thence to Indonesia. Instead, they seem a better fit to widespread population movements within island southeast Asia during the Holocene. However, we note that sample sizes differ substantially between these three locations (2740, 423 and 640 for Indonesians, Filipinos and indigenous Taiwanese, respectively), which adds variance—and perhaps bias—to diversity and dating estimates. Considered together with their large confidence intervals, we are reluctant to draw strong conclusions from molecular dates alone.

The Polynesian motif is also generally associated with a Taiwanese dispersal, but actually possesses an unusual geographical distribution. The ancestral form occurs widely throughout mainland and island southeast Asia. However, the Polynesian motif itself is found only at low frequency in the Philippines (0.5%)[36] and eastern Indonesia (2.3%).[16] Although frequencies reach as high as 7.4% on Timor, the lineage is found no further west than Bali (0.4%, or just 2 of 457 individuals). This is consistent with a proposed origin in island Melanesia,[64] but notably conflicts with the high frequency of the Polynesian motif in Madagascar, which was settled ~1200 years ago from western Indonesian sources.[53] We suggest that an inclusive framework that describes the full distribution of this unusual mtDNA lineage is still lacking. Nevertheless, an unambiguous connection with population dispersals from Taiwan during the Neolithic seems increasingly unlikely.

MtDNA evidence does suggest that most Indonesian groups may have increased in size. Negative values of Fu's Fs and Tajima's D point toward population growth and/or geographical expansion (Table 1), and reinforce similar conclusions that might be drawn from the star-like phylogenies of shared island southeast Asian mtDNA lineages (Figure 3). Signals of growth are greatest in the center of the archipelago (Bali, Sulawesi, Sumba and Flores), but weaker towards the eastern and western peripheries. Key exceptions are the barrier islands, Nias and Mentawai, whose mtDNA profiles are statistically more consistent with constant population size. These two groups also show some of the lowest haplotype diversities of any island southeast Asian populations (for instance, 43.5% of the Nias population carry haplogroup Y2). These outlier patterns may be due to genetic drift, as settlements on Nias and Mentawai are small even by indigenous Indonesian standards, or they may be caused by founder events and therefore reflect the unusual genetic profile of the islands' first settlers. The strong retained cultural heritage of these barrier islands may speak to this point. The inhabitants of Nias still practice old traditions, including the construction and maintenance of megaliths, and stone jumping (*hombo batu*), whereby young men show prowess by repeatedly jumping up onto a tall stone. Megalithic traditions associated with Austronesian culture were once common across Indonesia, but are now largely restricted to peripheral communities including those on Nias, Mentawai and Sumba.

In comparisons with neighboring populations, Indonesia's closest genetic connections lie toward mainland and island southeast Asia rather than Oceania (Supplementary Figures S2 and S3). Western Indonesian groups are notably distinct from Papuan groups (Figure 4), largely owing to low levels of haplogroups P and Q.[65] Whether these lineages are an enduring local presence from the original inhabitants of the region, or instead reflect recent westward movements from New Guinea is unclear. Soares *et al.*[64] suggest that the Polynesian motif spread westward from the Bismarck Archipelago, and similar movements have been proposed to explain the distribution of Papuan languages in eastern Indonesia.[2,66] SNPs from across the genome have been used to argue for an indigenous presence of Papuan genotypes rather than back migration from New Guinea, but strong statistical support for this is currently lacking.[67] Unfortunately, the data presented here, while powerful in helping to distinguish shared connections, are less informative about directions of movement.

The individuals screened for mtDNA were also assayed for Y-chromosome diversity.[20] Analysis of Austronesian languages and cultural systems,[10,11,68] as well as autosomal markers,[27,29] suggests that the men and women of island southeast Asia have followed quite different social histories. We compared mtDNA and Y-chromosome diversity to explore this further. Genetic divisions between populations are far weaker for mtDNA HVS I ($F_{ST} = 0.073$) than for Y-chromosome STRs ($F_{ST} = 0.202$), and this effect is even more pronounced at the haplogroup level when Indonesia is separated into its eastern and western parts (mtDNA SNPs $F_{ST} = 0.109$; Y-chromosome SNPs $F_{ST} = 0.475$) (Table 2). This discrepancy suggests that men and women have had different patterns of dispersal, with women moving widely between communities, while men have historically stayed local. One possible social explanation is patrilocality, where men remain in their natal community, but women move to the home village of their husband. Indeed, most of the populations presented in this study are patrilineal today. Interestingly, matrilineal systems are thought to have dominated ancestral Austronesian societies.[11,27,69] These mtDNA/Y-chromosome patterns may have been laid down after that cultural shift, and therefore reflect only the last few thousand years. Alternately, patrilocality could instead be the long-term standard with a transient switch to matrilocality during the Austronesian era. Differences in mutation rates could potentially confound this analysis; Y-STRs have high mutation rates, even when compared with the fast evolving mtDNA. However, these general patterns are still supported even when using standardized measures of genetic variation (Supplementary Table S5).

However, while the genetic evidence indicates that men and women experienced divergent histories, shared characteristics are more pronounced than differences. Karafet *et al.*[20] proposed a four-stage
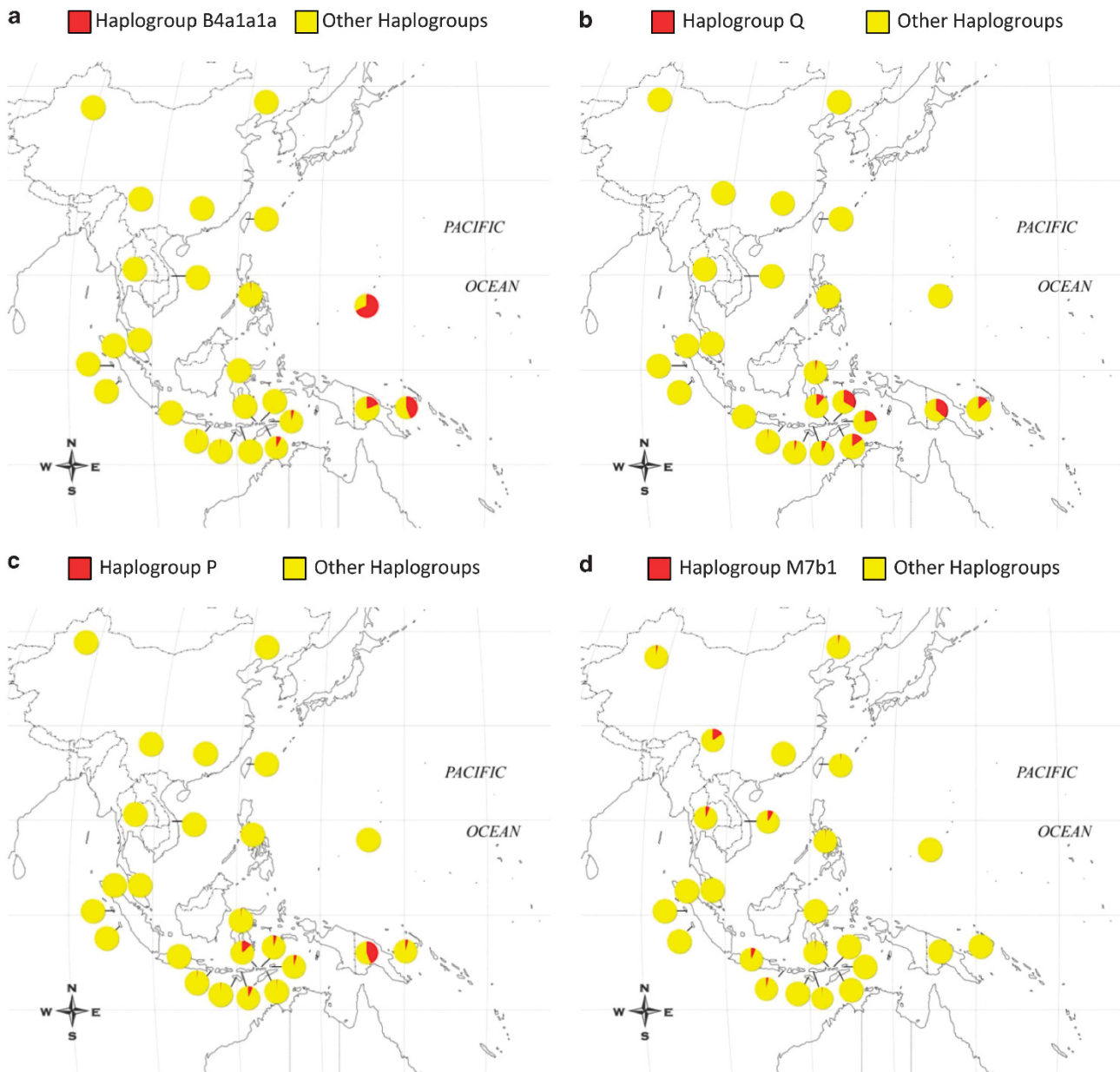
Figure 4 Frequency distribution of mitochondrial DNA haplogroups (**a**) B4a1a1a, (**b**) Q, (**c**) P and (**d**) M7b1.

colonization model built around variation on the Y chromosome. Here, we integrate the mtDNA evidence to provide a broader framework for reconstructing the history of Indonesia.

The first stage of Indonesian prehistory represents the archipelago's initial settlement as part of the African dispersal ~50 kya. The geography of the region was then markedly different from today. Sea levels were much lower, most modern islands had merged into larger landmasses and the westernmost parts of Indonesia were physically contiguous with mainland Asia. This first stage is recorded by deep mtDNA lineages (M17a, M73, M47, N21, N22, R21, R22 and R23), which trace back to the main branching of macrohaplogroups M and N, and have a spotty distribution across both mainland and island southeast Asia today.[25,33,36,55]

The second stage reflects recurrent colonization events from mainland Asia throughout the later Paleolithic. Many haplogroups

(B4a, B4b, B4c, B4c1b3, B5a, B5b, B5b1, D and E) show origin dates of 10–40 kya (Supplementary Table S6)[25,63] and are distributed across a wide range of mainland and island southeast Asian populations. As these lineages vary considerably in diversity (and hence, probable age) and show quite different geographical distributions, it is unlikely that any single demic event brought them to Indonesia. Indeed, many of these mtDNA haplogroups have been identified as key components of populations in peninsular Malaysia.[63] Therefore these lineages likely reflect multiple population movements from mainland Asia, possibly hunter-gatherers who followed the now-submerged river systems that once ran from mainland Asia between the modern islands of Sumatra, Java and Borneo.

The third stage represents Neolithic movements into and around island southeast Asia. Some of these may involve population dispersals from (and perhaps to) Taiwan, while others reflect

movements between Indonesian island groups. Representative haplogroups include M7b3, E1a1a, M7c3c and Y2. Autosomal data strongly support large demic movements of Asian populations into eastern Indonesia from around 4 kya.[67] Putatively associated with the spread of Austronesian languages, the direction of these movements remains unknown. If population movements during the Holocene were largely restricted to eastern Indonesia, this may explain why basal mtDNA lineages (such as M17a, N21, R22 and R23) are much rarer in the east. Although the Neolithic seems to have been a period of high population movement, it is curious that distinctions between eastern and western parts of the archipelago persist to the present. Within these regions, genetic similarity between island groups suggests that population movements were relatively frequent, consistent with archeological and linguistic evidence of a strong maritime culture. Borders such as Wallace's biogeographical line are poor boundaries for communities with significant ocean-going capabilities,[17] and the reason for an east–west distinction so clearly persisting in the genetic data is still imperfectly understood.[17,27,70] The most inclusive explanations invoke changes in social behavior, such as the emergence of matrilocal 'house societies' (sociétés à maison) during the early Austronesian period, and their subsequent disappearance as communities increasingly turned to patrilocality when the expansion period drew to a close.[70]

The fourth stage reflects historic movements into Indonesia, largely involving trade and the associated spread of major religions from India, Arabia and China.[71] Although found at relatively low frequency today, Y-chromosome lineages representing these movements occur across Indonesia,[20] notably in the west, such as the Hindu dominated island of Bali.[21] It is therefore interesting that no mtDNA lineages representing this period of invasion and migration were observed in our substantial data set. Unlike previous stages of Indonesian settlement, trade and religious connections during the historic era must have involved only male travelers, who subsequently took local wives. This difference between mtDNA and the Y chromosome emphasizes the complex, plural nature of most social processes. While most indigenous Indonesian communities practice patrilocality today (and hence exhibit preferential movement of women), long-distance genetic contributions are still effectively driven by men. As is often the case with complex biological systems, seemingly diametrically opposed processes can even act simultaneously in the same community at the same time.

The enormous diversity of language, culture and genetics in Indonesia is a direct outcome of the region's complex history of migration and settlement. Demic and cultural processes are both apparent: movements of people are indicated by related mtDNA lineages, but many aspects of culture—notably the widespread dispersal of Austronesian languages—are not obviously associated with genetics. Indeed, the only mtDNA lineage found across all Indonesian island groups is M7c3c, but this haplogroup, while also present in Taiwan and the Philippines, appears to be absent from other Austronesian-speaking populations in Oceania. Therefore, no single shared mtDNA lineage links all speakers of Austronesian languages, even if only at low frequency. Instead, Austronesian populations are characterized more by their diversity than by any shared genetic inheritance. However, there are limits to the power of uniparentally inherited markers, and with nearly 3000 individuals now screened, we wonder whether further sampling will substantially change the picture of mtDNA diversity portrayed here. Admixture analysis on genome-wide data sets indicates the extent of Asian immigrants in Indonesian populations and assigns the time of admixture to the mid to late Holocene.[67] Therefore, we are hopeful that new studies of autosomal data may help to answer many of the questions that remain outstanding, and we look forward to the clarification that this new wave of genetic evidence promises to bring.

1 Lewis, M. P. *Ethnologue: Languages of the World* (SIL International, Dallas, TX, USA, 2009).
2 Pawley, A. The chequered career of the Trans New Guinea hypothesis: Recent research and its implications. In *Papuan Pasts: Cultural, Linguistic and Biological Histories of Papuan-Speaking Peoples* (eds Pawley, A., Attenborough, R., Golson, J. & Hide, R.) 67–107 (Pacific Linguistics, Canberra, 2005).
3 Bellwood, P. *Prehistory of the Indo-Malaysian Archipelago* (Australian National University, Canberra, 2007).
4 Barker, G. The archaeology of foraging and farming at Niah Cave, Sarawak. *Asian Perspect.* **44,** 90–106 (2005).
5 Barker, G., Barton, H., Bird, M., Daly, P., Datan, I., Dykes, A. *et al.* The 'human revolution' in lowland tropical Southeast Asia: the antiquity and behavior of anatomically modern humans at Niah Cave (Sarawak, Borneo). *J. Hum. Evol.* **52,** 243–261 (2007).
6 Rasmussen, M., Guo, X., Wang, Y., Lohmueller, K. E., Rasmussen, S., Albrechtsen, A. *et al.* An Aboriginal Australian genome reveals separate human dispersals into Asia. *Science* **334,** 94–98 (2011).
7 Mellars, P. A new radiocarbon revolution and the dispersal of modern humans in Eurasia. *Nature* **439,** 931–935 (2006).
8 Mellars, P. Going east: new genetic and archaeological perspectives on the modern human colonization of Eurasia. *Science* **313,** 796–800 (2006).
9 O'Connell, J. F. & Allen, J. Dating the colonization of Sahul (Pleistocene Australia–New Guinea): a review of recent research. *J. Archaeol. Sci.* **31,** 835–853 (2004).
10 Gray, R. D., Drummond, A. J. & Greenhill, S. J. Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science* **323,** 479–483 (2009).
11 Jordan, F. M., Gray, R. D., Greenhill, S. J. & Mace, R. Matrilocal residence is ancestral in Austronesian societies. *Proc. R. Soc. B* **276,** 1957–1964 (2009).
12 Currie, T. E., Greenhill, S. J., Gray, R. D., Hasegawa, T. & Mace, R. Rise and fall of political complexity in island south-east Asia and the Pacific. *Nature* **467,** 801–804 (2010).
13 Bellwood, P. *The First Farmers: The Origins of Agricultural Societies* (Blackwell Publishing, Oxford, 2005).
14 Oppenheimer, S. *Eden in the East: The Drowned Continent of Southeast Asia* (Phoenix, London, 1999).
15 Oppenheimer, S. J. & Richards, M. Polynesian origins. Slow boat to Melanesia? *Nature* **410,** 166–167 (2001).
16 Cox, M. P. Indonesian mitochondrial DNA and its opposition to a Pleistocene era origin of proto-Polynesians in Island Southeast Asia. *Hum. Biol.* **77,** 179–188 (2005).
17 Cox, M. P. The genetic environment of Melanesia: Clines, clusters and contact. In *Population Genetics Research Progress* (ed. Koven, V. T.) 45–83 (Nova Science Publishers, New York, 2008).
18 Cox, M. P. & Lahr, M. M. Y-Chromosome diversity is inversely associated with language affiliation in paired Austronesian- and Papuan-speaking communities from Solomon Islands. *Am. J. Hum. Biol.* **18,** 35–50 (2006).
19 Cox, M. P., Redd, A. J., Karafet, T. M., Ponder, C. A., Lansing, J. S., Sudoyo, H. *et al.* A Polynesian motif on the Y chromosome: population structure in Remote Oceania. *Hum. Biol.* **79,** 525–535 (2007).
20 Karafet, T. M., Hallmark, B., Cox, M. P., Sudoyo, H., Downey, S. S., Lansing, J. S. *et al.* Major east-west division underlies Y chromosome stratification across Indonesia. *Mol. Biol. Evol.* **27,** 1833–1844 (2010).
21 Karafet, T. M., Lansing, J. S., Redd, A. J., Reznikova, S., Watkins, J. C., Surata, S. P. *et al.* Balinese Y-chromosome perspective on the peopling of Indonesia: genetic contributions from pre-Neolithic hunter-gatherers, Austronesian farmers, and Indian traders. *Hum. Biol.* **77,** 93–114 (2005).
22 Kayser, M. The human genetic history of Oceania: near and remote views of dispersal. *Curr. Biol.* **20,** R194–R201 (2010).
23 Kayser, M., Brauer, S., Cordaux, R., Casto, A., Lao, O., Zhivotovsky, L. A. *et al.* Melanesian and Asian origins of Polynesians: mtDNA and Y-chromosome gradients across the Pacific. *Mol. Biol. Evol.* **23,** 2234–2244 (2006).
24 Kayser, M., Brauer, S., Weiss, G., Schiefenhövel, W., Underhill, P., Shen, P. *et al.* Reduced Y-chromosome, but not mitochondrial DNA, diversity in human populations from West New Guinea. *Am. J. Hum. Genet.* **72,** 281–302 (2003).

25 Hill, C., Soares, P., Mormina, M., Macaulay, V., Clarke, D., Blumbach, P. B. et al. A mitochondrial stratigraphy for island southeast Asia. Am. J. Hum. Genet. 80, 29–43 (2007).

26 Soares, P., Trejaut, J. A., Loo, J.-H., Hill, C., Mormina, M., Lee, C.-L. et al. Climate change and postglacial human dispersals in Southeast Asia. Mol. Biol. Evol. 25, 1209–1218 (2008).

27 Cox, M. P., Karafet, T. M., Lansing, J. S., Sudoyo, H. & Hammer, M. F. Autosomal and X-linked single nucleotide polymorphisms reveal a steep Asian-Melanesian ancestry cline in eastern Indonesia and a sex bias in admixture rates. Proc. R. Soc. B 277, 1589–1596 (2010).

28 HUGO Pan-Asian SNP Consortium. Mapping human genetic diversity in Asia. Science 326, 1541–1545 (2009).

29 Wollstein, A., Lao, O., Becker, C., Brauer, S., Trent, R. J., Nürnberg, P. et al. Demographic history of Oceania inferred from genome-wide data. Curr. Biol. 20, 1983–1992 (2010).

30 Yao, Y. G., Kong, Q. P., Bandelt, H. J., Kivisild, T. & Zhang, Y. P. Phylogeographic differentiation of mitochondrial DNA in Han Chinese. Am. J. Hum. Genet. 70, 635–651 (2002).

31 Allard, M. W., Wilson, M. R., Monson, K. L. & Budowle, B. Control region sequences for East Asian individuals in the Scientific Working Group on DNA Analysis Methods forensic mtDNA data set. Leg. Med. 6, 11–24 (2004).

32 Li, H., Cai, X., Winograd-Cort, E. R., Wen, B., Cheng, X., Qin, Z. et al. Mitochondrial DNA diversity and population differentiation in southern East Asia. Am. J. Phys. Anthropol. 134, 481–488 (2007).

33 Hill, C., Soares, P., Mormina, M., Macaulay, V., Meehan, W., Blackburn, J. et al. Phylogeography and ethnogenesis of aboriginal southeast Asians. Mol. Biol. Evol. 23, 2480–2491 (2006).

34 Trejaut, J. A., Kivisild, T., Loo, J. H., Lee, C. L., He, C. L., Hsu, C. J. et al. Traces of archaic mitochondrial lineages persist in Austronesian-speaking Formosan populations. PLoS Biol. 3, e247 (2005).

35 Friedlaender, J. S., Friedlaender, F. R., Hodgson, J. A., Stoltz, M., Koki, G., Horvat, G. et al. Melanesian mtDNA complexity. PLoS ONE 2, e248 (2007).

36 Tabbada, K. A., Trejaut, J., Loo, J.-H., Chen, Y.-M., Lin, M., Mirazón-Lahr, M. et al. Philippine mitochondrial DNA diversity: a populated viaduct between Taiwan and Indonesia? Mol. Biol. Evol. 27, 21–31 (2010).

37 Miller, S. A., Dykes, D. D. & Polesky, H. F. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res. 16, 1215 (1988).

38 Andrews, R. M., Kubacka, I., Chinnery, P. F., Lightowlers, R. N., Turnbull, D. M. & Howell, N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat. Genet. 23, 147 (1999).

39 Tanaka, M., Cabrera, V. M., Gonzalez, A. M., Larruga, J. M., Takeyasu, T., Fuku, N. et al. Mitochondrial genome variation in eastern Asia and the peopling of Japan. Genome Res. 14, 1832–1850 (2004).

40 Kong, Q. P., Yao, Y. G., Sun, C., Bandelt, H. J., Zhu, C. L. & Zhang, Y. P. Phylogeny of east Asian mitochondrial DNA lineages inferred from complete sequences. Am. J. Hum. Genet. 73, 671–676 (2003).

41 Friedlaender, J., Gentz, F., Friedlaender, F., Kaestle, F., Schurr, T., Koki, G. et al. Mitochondrial genetic diversity and its determinants in Island Melanesia. In Papuan Pasts: Cultural, Linguistic and Biological Histories of Papuan-Speaking Peoples (eds Pawley, A., Attenborough, R., Golson, J. & Hide, R.) 693–716 (Pacific Linguistics, Canberra, 2005).

42 van Oven, M. Revision of the mtDNA tree and corresponding haplogroup nomenclature. PNAS 107, E38–E39 (2010).

43 Excoffier, L. & Lischer, H. E. L. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Res. 10, 564–567 (2010).

44 Slatkin, M. Inbreeding coefficients and coalescence times. Genet. Res. Camb. 58, 167–175 (1991).

45 Slatkin, M. A measure of population subdivision based on microsatellite allele frequencies. Genetics 139, 457–462 (1995).

46 R Development Core Team. R: A Language and Environment for Statistical Computing (2012). http://www.r-project.org/.

47 Bandelt, H. J., Forster, P. & Röhl, A. Median-joining networks for inferring intraspecific phylogenies. Mol. Biol. Evol. 16, 37–48 (1999).

48 Forster, P., Harding, R., Torroni, A. & Bandelt, H.-J. Origin and evolution of native American mtDNA variation: a reappraisal. Am. J. Hum. Genet. 59, 935–945 (1996).

49 Soares, P., Ermini, L., Thomson, N., Mormina, M., Rito, T., Röhl, A. et al. Correcting for purifying selection: an improved human mitochondrial molecular clock. Am. J. Hum. Genet. 84, 740–759 (2009).

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

51 Hedrick, P. W. A standardized genetic differentiation measure. Evolution 59, 1633–1638 (2005).

52 van Oven, M., Hämmerle, J. M., van Schoor, M., Kushnick, G., Pennekamp, P., Zega, I. et al. Unexpected island effects at an extreme: reduced Y-chromosome and mitochondrial DNA diversity in Nias. Mol. Biol. Evol. 28, 1349–1361 (2011).

53 Cox, M. P., Nelson, M. G., Tumonggor, M. K., Ricaut, F. X. & Sudoyo, H. A small cohort of Island Southeast Asian women founded Madagascar. Proc. R. Soc. B 279, 2761–2768 (2012).

54 Razafindrazaka, H., Ricaut, F.-X., Cox, M. P., Mormina, M., Dugoujon, J.-M., Randriamarolaza, L. P. et al. Complete mitochondrial DNA sequences provide new insights into the Polynesian motif and the peopling of Madagascar. Eur. J. Hum. Genet. 18, 575–581 (2010).

55 Macaulay, V., Hill, C., Achilli, A., Rengo, C., Clarke, D., Meehan, W. et al. Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. Science 308, 1034–1036 (2005).

56 Dupuy, B. M., Stenersen, M., Egeland, T. & Olaisen, B. Y-chromosomal microsatellite mutation rates: differences in mutation rate between and within loci. Hum. Mutat. 23, 117–124 (2004).

57 Fenner, J. N. Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. Am. J. Phys. Anthropol. 128, 415–423 (2005).

58 Gusmão, L., Sánchez-Diz, P., Calafell, F., Martín, P., Alonso, C. A., Álvarez-Fernández, F. et al. Mutation rates at Y chromosome specific microsatellites. Hum. Mutat. 26, 520–528 (2005).

59 Zhivotovsky, L. A., Underhill, P. A., Cinnioğlu, C., Kayser, M., Morar, B., Kivisild, T. et al. The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. Am. J. Hum. Genet. 74, 50–61 (2004).

60 Lansing, J. S., Karafet, T. M., Schoenfelder, J. & Hammer, M. F. A DNA signature for the expansion of irrigation in Bali? In Past Human Migrations in East Asia: Matching Archaeology, Linguistics and Genetics (eds Sanchez-Mazas, A., Blench, R., Ross, M. D., Peiros, I. & Lin, M.) 377–395 (Routledge, London, 2008).

61 Blust, R. The Austronesian Languages. Research School of Pacific and Asian Studies. (Australian National University, Canberra, Australia, 2009).

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

63 Jinam, T. A., Hong, L.-C., Phipps, M. E., Stoneking, M., Ameen, M., Edo, J. et al. Evolutionary history of continental southeast Asians: 'Early train' hypothesis based on genetic analysis of mitochondrial and autosomal DNA data. Mol. Biol. Evol 29, 3513–3527 (2012).

64 Soares, P., Rito, T., Trejaut, J., Mormina, M., Hill, C., Tinkler-Hundal, E. et al. Ancient voyaging and Polynesian origins. Am. J. Hum. Genet. 88, 239–247 (2011).

65 Forster, P., Torroni, A., Renfrew, C. & Röhl, A. Phylogenetic star contraction applied to Asian and Papuan mtDNA evolution. Mol. Biol. Evol. 18, 1864–1881 (2001).

66 Ross, M. Pronouns as a preliminary diagnostic of grouping Papuan languages. In Papuan Pasts: Cultural, Linguistic and Biological Histories of Papuan-Speaking Peoples (eds Pawley, A., Attenborough, R., Golson, J. & Hide, R.) 15–65 (Pacific Linguistics, Canberra, 2005).

67 Xu, S., Pugach, I., Stoneking, M., Kayser, M., Jin, L. & the HUGO Pan-Asian SNP Consortium. Genetic dating indicates that the Asian-Papuan admixture through Eastern Indonesia corresponds to the Austronesian expansion. Proc. Natl Acad. Sci. USA 109, 4574–4579 (2012).

68 Gray, R. D. & Jordan, F. M. Language trees support the express-train sequence of Austronesian expansion. Nature 405, 1052–1054 (2000).

69 Hage, P. & Marck, J. Matrilineality and the Melanesian origin of Polynesian Y chromosomes. Curr. Anthropol. 44, S121–S127 (2003).

70 Lansing, S. J., Cox, M. P., de Vet, T. A., Downey, S. S., Hallmark, B. & Sudoyo, H. An ongoing Austronesian expansion in Island Southeast Asia. J. Anthropol. Archaeol. 30, 262–272 (2011).

71 Tibbetts, G. A Study of the Arabic Texts Containing Material on South-East Asia (E. J. Brill, Leiden, 1979).