

# Major East–West Division Underlies Y Chromosome Stratification across Indonesia

Tatiana M. Karafet,<sup>1</sup> Brian Hallmark,<sup>1,2</sup> Murray P. Cox,<sup>1</sup> Herawati Sudoyo,<sup>3</sup> Sean Downey,<sup>2</sup> J. Stephen Lansing,<sup>2,4</sup> and Michael F. Hammer<sup>\*,1,2</sup>

<sup>1</sup>ARL Division of Biotechnology, University of Arizona

<sup>2</sup>School of Anthropology, University of Arizona

<sup>3</sup>Eijkman Institute for Molecular Biology, Jakarta, Indonesia

<sup>4</sup>Santa Fe Institute, Santa Fe, New Mexico

\*Corresponding author: E-mail: mfh@u.arizona.edu.

Associate editor: Lisa Matisoo-Smith

## Abstract

The early history of island Southeast Asia is often characterized as the story of two major population dispersals: the initial Paleolithic colonization of Sahul ~45 ka ago and the much later Neolithic expansion of Austronesian-speaking farmers ~4 ka ago. Here, in the largest survey of Indonesian Y chromosomes to date, we present evidence for multiple genetic strata that likely arose through a series of distinct migratory processes. We genotype an extensive battery of Y chromosome markers, including 85 single-nucleotide polymorphisms/indels and 12 short tandem repeats, in a sample of 1,917 men from 32 communities located across Indonesia. We find that the paternal gene pool is sharply subdivided between western and eastern locations, with a boundary running between the islands of Bali and Flores. Analysis of molecular variance reveals one of the highest levels of between-group variance yet reported for human Y chromosome data (e.g.,  $\Phi_{ST} = 0.47$ ). Eastern Y chromosome haplogroups are closely related to Melanesian lineages (i.e., within the C, M, and S subclades) and likely reflect the initial wave of colonization of the region, whereas the majority of western Y chromosomes (i.e., O-M119\*, O-P203, and O-M95\*) are related to haplogroups that may have entered Indonesia during the Paleolithic from mainland Asia. In addition, two novel markers (P201 and P203) provide significantly enhanced phylogenetic resolution of two key haplogroups (O-M122 and O-M119) that are often associated with the Austronesian expansion. This more refined picture leads us to put forward a four-phase colonization model in which Paleolithic migrations of hunter-gatherers shape the primary structure of current Indonesian Y chromosome diversity, and Neolithic incursions make only a minor impact on the paternal gene pool, despite the large cultural impact of the Austronesian expansion.

**Key words:** Y chromosome, Paleolithic, colonization, Austronesian, Wallace's line.

## Introduction

Indonesia, the world's largest archipelago, is a chain of more than 17,000 islands that stretches between the continents of Asia and Australia, dividing the Pacific and Indian oceans. The ~240 million inhabitants are extremely diverse, speaking more than 750 languages and representing >300 different ethnic groups. Within this extreme diversity, several large-scale patterns give clues to the early settlement history of the region. Early explorers noticed morphological differences from east to west that were dramatic enough to lead Alfred Russell Wallace to designate a human phenotypic boundary demarcating the transition between Asian and Melanesian features. Relative to his more well-known biogeographic boundary, this line lies slightly east, running between the islands of Sumbawa and Flores (Wallace 1869). The languages of the region follow a similar pattern, with the majority belonging to the extensive Austronesian language family but with more distantly related Papuan languages occurring in the Far Eastern provinces, especially in areas where Melanesian features predominate (Wallace 1869; Howells 1973; Pietrusewsky 1994; Cox 2008).

To explain these patterns, the prehistory of this region has often been framed as the story of two major range expansions: the initial Paleolithic colonization of Sahul ~45 ka ago (Kirch 2000; Roberts et al. 2001; Leavesley and Chappell 2004; O'Connell and Allen 2004; Barker et al. 2007) and the much later Neolithic expansion of Austronesian-speaking farmers (4–6 ka ago) out of mainland Asia or Taiwan into Indonesia and the Pacific (Kirch 1997; Diamond 2000; Gray and Jordan 2000; Su et al. 2000; Capelli et al. 2001; Hurler et al. 2002; Kayser et al. 2003, 2006; Karafet et al. 2005, 2008; Li et al. 2008). In this scenario, the Austronesian expansion shaped the primary genetic, linguistic, and cultural diversity of the region, and the distribution of Papuan languages and phenotypic features are remnants of the initial colonization, which survived for thousands of years amid significant climatic changes and cultural shifts.

Although appealing in its simplicity, this two-phase model is inconsistent with information from numerous sources that point to a far more complex history for the region. Recent work on human mitochondrial DNA (mtDNA) suggests that the majority of the region's maternal gene pool has a pre-Austronesian origin and that the

distribution of mtDNA haplogroups is better explained by climatic and sea-level changes following the Last Glacial Maximum rather than the expansion of farmers out of mainland Asia and/or Taiwan (Hill et al. 2007; Soares et al. 2008). Ethnobotanical and linguistic evidence suggests a significant pre-Austronesian westward dispersal of bananas and their cultivators from New Guinea into eastern Indonesia and possibly even further west (Denham and Donohue 2009). Work on pig mtDNA points to multiple distinct migrations not only eastward out of Southeast Asia but also within Wallacea itself (Bellwood and White 2005; Larson et al. 2005; Lum et al. 2006). These data suggest that the pigs of Melanesia and Oceania trace their maternal origin to Southeast Asia rather than to Taiwan, which has been proposed as the place of origin of the Austronesian language family based on linguistic diversity (Blust 1995; Bellwood 1997).

Even within linguistics, the source of the Austronesian languages remains controversial, and dispersals both into and out of Taiwan are still debated (Meacham 1984; Gray and Jordan 2000; Diamond and Bellwood 2003). In addition, recent small-scale studies (Lansing et al. 2007, 2008) have added to our understanding of the contact dynamics that occurred as Austronesian speakers moved into places where preexisting Papuan populations already lived. More recently, Indonesia, especially in the west, has experienced ever-increasing spheres of Eurasian influence, which have all led to major cultural changes, including trade with India and the establishment of Hindu kingdoms (from ~2.5 ka ago), the arrival of Arab traders and Islam from the Near East (~1 ka ago), and European contact within the past 500 years.

Although Indonesian populations have been surveyed in many previous genetic studies (Kayser et al. 2000, 2001, 2003, 2006; Hurles et al. 2002; Redd et al. 2002b; Karafet et al. 2005; Li et al. 2008; Mona et al. 2009), there are still many open questions regarding the settlement history of the archipelago. One challenge is the sheer size and complexity of the Indonesian region, which makes comprehensive sampling difficult. To better characterize the paternal genetic landscape and to further disentangle the complex demographic history of the region, we present the largest survey of Indonesian Y chromosome diversity to date. In 2003, we established a close collaboration with the Eijkman Institute for Molecular Biology in Jakarta and have since sampled 1,917 men from 32 locations across Indonesia. Here, we report the results of genotyping these samples with an extensive battery of Y chromosome markers, both single-nucleotide polymorphisms (SNPs,  $n = 85$ ) and short tandem repeats (Y-STRs,  $n = 12$ ). Included in our set of SNPs are three markers (P201, P203, and JST002611) that have not been surveyed in this geographic region thus far and provide increased phylogenetic resolution of two of the major lineages (haplogroups O-M119 and O-M122) associated with the Austronesian expansion. In addition, our database contains the largest number of samples yet reported ( $n = 873$ ) from the eastern Indonesian province of Nusa Tenggara Timur, which is

notable as a contact zone with high linguistic and cultural diversity.

## Subjects and Methods

### Samples

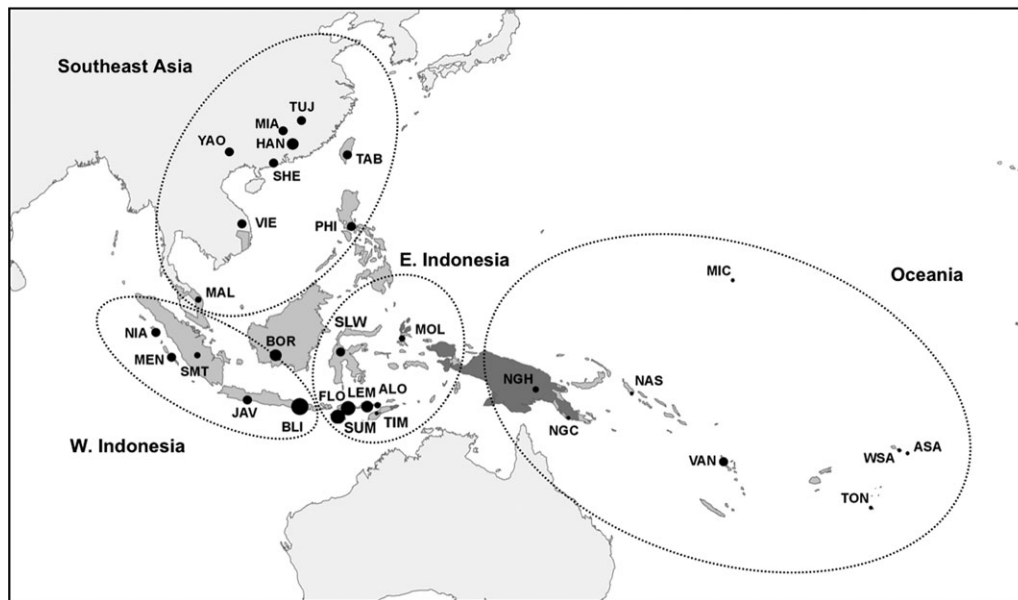
Our Indonesian sample comprised 1,917 males from 32 communities on 13 islands, including Sumatra ( $n = 38$ ), Nias ( $n = 60$ ), Mentawai ( $n = 74$ ), Java ( $n = 61$ ), Borneo ( $n = 86$ ), Bali ( $n = 641$ ), Sulawesi ( $n = 54$ ), Flores ( $n = 394$ ), Lembata ( $n = 92$ ), Sumba ( $n = 350$ ), Alor ( $n = 28$ ), Timor ( $n = 9$ ), and a composite group from the Maluku Islands ( $n = 30$ ). For many of these islands, the samples were collected at multiple traditional villages but have been pooled together in this study for a broad-scale analysis (supplementary table S1, Supplementary Material online). Buccal swabs were collected from volunteers by HS and/or JSL from 2003 to 2007 with informed consent by the donors. All sampling procedures were approved by the University of Arizona Human Subjects Committee, Balai Pengkajian Teknologi Pertanian (Bali), and the Government of Indonesia. For comparative analysis, we included an additional 763 previously reported samples from our database, representing 19 populations across Southeast Asia and Oceania (Hammer et al. 2001; Karafet et al. 2001, 2005; Redd et al. 2002b). Figure 1 shows a map of the sampling locations, and additional information can be found in supplementary tables S1 and S2 (Supplementary Material online).

### Genetic Markers

Polymorphic sites from the nonrecombining portion of the human Y chromosome (NRY) included a set of 82 binary markers published previously (Karafet et al. 2005) together with a set of three new polymorphisms, JST002611, P201, and P203 (Karafet et al. 2008), that have not been typed at this scale before. The phylogenetic position of these markers and the haplogroups they define in our samples are shown in figure 2. A hierarchical genotyping strategy was used in which major haplogroups were predicted based on the array of Y-STR alleles contained on each Y chromosome (Schlecht et al. 2008) and then confirmed by genotyping of a smaller set of SNPs. Once the correct major haplogroup was identified, additional genotyping was restricted to the appropriate downstream mutations along the haplogroup tree (fig. 2). We follow the nomenclature recommended by the Y Chromosome Consortium (YCC 2002; Karafet et al. 2008), focusing mainly on the mutation-based naming system. Potentially paraphyletic paralogous groups are distinguished from haplogroups by the asterisk symbol. We also analyzed 12 Y-STRs (DYS19, DYS385a, DYS385b, DYS388, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS426, and DYS439) using methods described by Redd et al. (2002a).

### Statistical Analysis

The software package Arlequin 3.0 (Excoffier Schneider et al. 2005) was used to calculate several population parameters, including  $\Phi_{ST}$  distances, Nei's (1987) heterozygosity, and the mean number of pairwise differences among



**Fig. 1.** Map showing the approximate geographic positions of the population samples included in this survey. The size of each circle is proportional to the size of the sample. The populations are grouped (within dotted lines) into four major geographic areas (three-letter codes refer to populations sampled in this survey): Southeast Asia (SEA, including both mainland and island populations: HAN, Han Chinese; TUJ, Tujia; MIA, Miao; YAO, Yao; SHE, She; VIE, Vietnamese; MAL, Malaysians; TAB, Taiwanese aboriginals; and PHI, Filipinos), western Indonesia (SMT, Sumatra; NIA, Nias; MEN, Mentawai; JAV, Java; BLI, Bali; and BOR, Borneo), eastern Indonesia (FLO, Flores; SUM, Sumba; LEM, Lembata; ALO, Alor; TIM, Timor; SLW, Sulawesi; and MOL, Moluccas), and Oceania (NGC, Coastal Papua New Guinea; NGH, Highland Papua New Guinea; VAN, Vanuatu; NAS, Nasioi; MIC, Micronesia; ASA, American Samoa; WSA, western Samoa; TON, Tonga; and RNU, Rapa Nui/Easter Island and TAH, Tahiti not shown). The distribution of Austronesian and Papuan languages is shown in light and dark shades of gray, respectively.

haplogroups. Multidimensional scaling (MDS) (Kruskal 1964) was performed on the Slatkin's linearized  $\Phi_{ST}$  distances using the software package NTSYS (Rohlf 1998). Median-joining (MJ) networks of haplotypes within certain haplogroups (Bandelt et al. 1999) were constructed using the Network 4.1c program. STRs were weighted according to their repeat number as described elsewhere (Karafet et al. 2005). The program ArcMap 9.3 was used to produce geographical maps.

## Results

### Distribution of Y Chromosome Lineages

A total of 55 haplogroups (or paragroups) were found in our database of populations from Indonesia, Southeast Asia, and Oceania (supplementary table S1, Supplementary Material online). Figure 2 presents a maximum parsimony tree showing the evolutionary relationships of the clades present in our database. The majority of samples (~94.7%) belong to either haplogroup C (~22.5%) or three major subclades of haplogroup K (~72.2%): haplogroups M, O, and S, as well as K\*. Figure 2 also shows frequencies of the most common haplogroups in four broad geographic regions: mainland/island Southeast Asia (we refer to this region, which includes both mainland and some island populations, as SEA), western Indonesia (WIN), eastern Indonesia (EIN), and Oceania (OCE).

### Haplogroup C

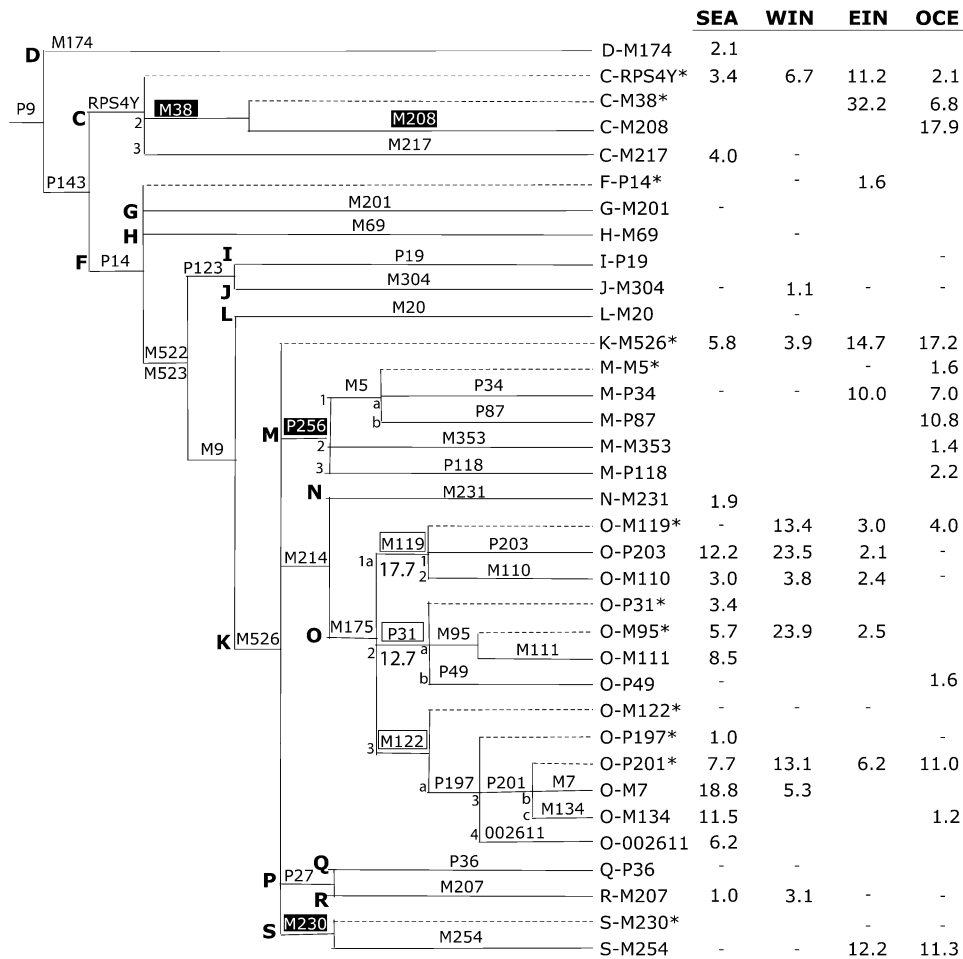
The geographic distribution of C lineages (i.e., downstream from the RPS4Y mutation) is shown in supplementary

figure S1 (Supplementary Material online). The high frequency and widespread distribution of these lineages are indicative of their long history in the region. Our results generally echo previous findings (Kayser et al. 2003, 2006, 2008; Cox and Mirazon Lahr 2006; Scheinfeldt et al. 2006; Cox et al. 2007; Mona et al. 2007) with paralog C-RPS4Y\* chromosomes found from Southeast Asia to eastern Indonesia, C-M38\* found predominantly in eastern Indonesia, and C-M208 being limited to Melanesia and the Pacific islands (i.e., remote Oceania).

On a finer geographic scale, the C-RPS4Y\* paralog has a patchy distribution throughout SEA and Indonesia and is absent or at very low frequency further east in Melanesia and Polynesia. In SEA, the highest frequency is among the Yao in China (20%). It is also present in eastern Indonesia, achieving relatively high frequencies in Flores (29.2%) and Lembata (22.8%). The C-M38\* paralog is absent in western Indonesia but is the most frequent lineage in eastern Indonesia (33.5%), varying from 57% in Sumba to 11% in Sulawesi. It is also found at lower frequency in Melanesia and Oceania. Haplogroup C-M208, a subgroup of C-M38, is not found in eastern Indonesia and its geographic distribution is restricted to Melanesia and Polynesia.

### Haplogroup K and Its Derived Lineages

The vast majority of samples belong to haplogroups derived from the M9 mutation. Within this group, we find a large number of samples belonging to haplogroups M (6.4%), O (55.0%), and S (6.0%). The K-M9\* paralog has a wide geographic distribution and is found on almost all the sampled



**Fig. 2.** Maximum parsimony tree of 35 Y chromosome haplogroups together with their frequencies in four geographic groups. Sample sizes for each region are mainland Southeast Asia (SEA), 581; western Indonesia (WIN), 960; eastern Indonesia (EIN), 957; and Oceania (OCE), 182. Major clades (i.e., C–S) are labeled with uppercase letters to the left of each clade. Mutation names are given along the branches. Dotted lines indicate internal nodes not defined by downstream markers (i.e., paragroups). The names of 35 haplogroups observed in our samples are shown to the right of the branches using the mutation-based nomenclature of the YCC (2002). Lineage-based nomenclature information is provided next to internal branches for key O and M haplogroups. Haplogroup frequencies (shown to the far right) are weighted to normalize the contribution of each population. (Hyphen indicates <1%) Clades with assumed Asian mainland and Papuan origins are shown in white and black boxes, respectively.

islands. It reaches relatively high frequency in the Philippines (45.8%) and is common in Melanesia, including Papua New Guinea.

The new polymorphism P256 joins haplogroups M-M5, K-M353, and K-P118 into superhaplogroup M (Karafet et al. 2008). In our survey, paragroup M-P256\* is found at low incidence on Flores (2.5%) and New Guinea (6.3%). Many M haplogroups (M-P87, M-P22, M-M16, M-M353, and M-SRY9138) are observed primarily in Melanesia. The M-P34 haplogroup has a wider geographic distribution and is found at moderate frequency in eastern Indonesia (11.6%) and New Guinea (25%). M-P34 chromosomes are also observed in Bali and Malaysia, albeit at very low frequency. The associated Y-STR variance within the M lineage is not significantly different in eastern Indonesia and Oceania (0.587 vs. 0.706,  $P = 0.213$ ; supplementary table S3, Supplementary Material online).

Haplogroup S (S-M230) is another haplogroup with a primarily Melanesian distribution. The majority of S-M230 chromosomes are marked by a downstream poly-

morphism at M254, the derived form of which is present at 13.3% in eastern Indonesia and 12.1% in Melanesia. Only nine M230 chromosomes were found to have the ancestral state at the M254 site (i.e., eight in eastern Indonesia and one in Melanesia).

Haplogroup O, also derived from K, represents one of the major lineages in Indonesia. A total of 14 distinct O haplogroups/paragroups are found in our survey. Haplogroups defined by the M175 mutation fall into three main clades: O-M119, O-P31, and O-M122. Additional markers were typed in each clade, giving increased resolution to the geographic distribution of these haplogroups. O-M119 chromosomes were typed for two additional downstream markers: M110 and P203. P203 is a new marker that adds significant spatial resolution to this haplogroup (supplementary fig. S2, Supplementary Material online). Although the O-M119\* paragroup is virtually absent in our mainland/Southeast Asian sample (the exceptions are a single Han Chinese and a Malaysian), it is found at notable frequencies in western and eastern

Indonesia (11% and 3.2%, respectively). O-P203 now replaces what was previously O-M119\* throughout much of SEA and Indonesia. It is especially notable that the majority of Taiwanese aboriginals in our survey are O-P203\* (i.e., no individuals belong to O-M119\* when P203 is typed). Subhaplogroup O-M110 is present in Indonesia, as well as among Taiwanese aboriginals and Filipinos.

The geographic distribution of O-P31 and downstream markers is shown in [supplementary figure S3](#) (Supplementary Material online). In addition to the defining marker of this clade (P31), four additional markers were typed: M95, M111, SRY465, and 47z. O-M95 is of greatest importance in Indonesia and is present in both WIN and SEA, reaching its highest frequency on Bali (57.3%). It is nearly absent in Oceania and eastern Indonesia, with the exception of Sulawesi. Although it is at highest frequency in western Indonesia, there is greater Y-STR diversity associated with the O-M95 lineage in SEA than in Indonesia. Other O-P31 haplogroups are limited to SEA and are not found in our sample of Indonesians.

M122 chromosomes were further analyzed for the biallelic markers P197 (phylogenetically equivalent to M324), P201=JST021354, M7, M134, and JST002611. [Supplementary figure S4](#) (Supplementary Material online) shows their geographic distribution. The derived O-M134, O-M7, and O-JST002611 subhaplogroups are absent or found at very low frequencies in Indonesia but are prevalent in different ethnic groups in China and SEA. Paragroup O-P197\* is found at low frequency in Vietnam, the Philippines, and China (Han). A total of 17 O-M122\* chromosomes were found, but only 2 of them were observed outside of Indonesia. Only paragroup O-P201\* has a wide geographic distribution and is found in all geographic regions surveyed.

To explore the potential root of O-P201 chromosomes in Indonesia, we calculated genetic distances ( $R_{ST}$ ) based on O-P201\* STR haplotypes ([supplementary table S4](#), Supplementary Material online). The divergence between Filipinos/Taiwanese aboriginals and Indonesians was insignificant and 10-fold less than that between Southeast Asians and Indonesians (0.027,  $P = 0.11$  vs. 0.349,  $P = 0.00$ ). Genetic distances between Oceania and Philippines/Taiwanese aboriginals were even lower (0.007,  $P = 0.35$ ).

#### Minor Lineages

The remaining ~5.3% of samples fall into a wide range of haplogroups. Other than haplogroups N and Q, which are found at highest frequencies in SEA (3.6% and 1.0%, respectively), most of the minor lineages (i.e., within haplogroups H, J, L, and R) reach their highest frequency in western Indonesia. This reflects their widespread distribution in this region, with lineages within the H, J, L, and P clades being found at low-to-moderate frequencies (i.e., ranging from ~3% to ~15%) on all western Indonesian islands, except Mentawai and Nias.

#### Population Structure

To examine population relationships, an MDS analysis was performed using haplogroup frequencies ([fig. 3](#)). All popu-

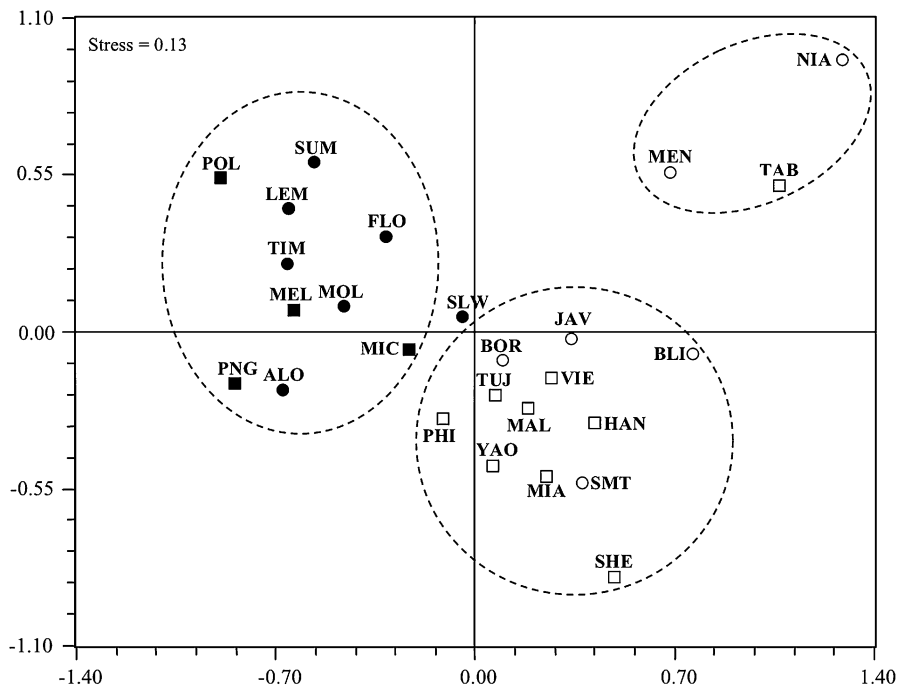
lations from eastern Indonesia cluster together on the left side of the plot with populations from Oceania scattered within this group. Western Indonesian groups show a closer affinity to populations from SEA rather than to populations from eastern Indonesia, with the exception of Sulawesi. Populations from Nias and Mentawai cluster together with Taiwanese aboriginals and form a cluster on the upper right side of the plot. This clustering may result from the high frequency of O-P203 chromosomes in all three populations.

In western Indonesia, 27 NRY haplogroups are present, whereas only 17 haplogroups are observed in eastern Indonesia ([fig. 2](#), [supplementary table S2](#), Supplementary Material online). The  $\Phi_{ST}$  value based on Y haplogroup frequencies using all 32 Indonesian populations was 0.40 ([table 1](#)). When the 32 populations were divided into 13 islands, the  $\Phi_{ST}$  parameter became slightly higher (0.41), with high variation among populations within groups ( $\Phi_{SC} = 0.28$ ;  $\Phi_{CT} = 0.19$ ). When the 32 Indonesian populations were subdivided into two groups along Wallace's biogeographical line, the corresponding  $\Phi_{ST}$  value reached its highest value of 0.47, with a significantly higher between-group variation ( $\Phi_{CT} = 0.28$ ). Because the sample from Sulawesi had an intermediate position between EIN and WIN in the MDS plot, we also performed analyses of molecular variance with Sulawesi as a part of western Indonesia, but the  $\Phi_{ST}$  value remained 0.47 (data not shown).

## Discussion

Here, we present a unified summary of paternally inherited diversity across Indonesia. Our typing of additional markers paired with new sampling, especially in eastern Indonesia, suggests a more refined interpretation of the distribution of several lineages and provides a more detailed picture of Y chromosome variation in Indonesia than has been previously presented. Perhaps the most striking result is the dramatic difference in haplogroup frequencies between western and eastern Indonesia. Levels of between-group variation, especially when populations are subdivided into western and eastern Indonesian groups, are some of the highest observed thus far for Y chromosome data ([table 1](#)). Although a lack of data in central Indonesia (e.g., Lombok and Sumbawa) and insufficient sampling of Sulawesi make the exact location of this boundary uncertain, it is clear that the transition occurs somewhere between Bali and Flores in the vicinity of both Wallace's biogeographical and phenotypic lines (Wallace 1869; Cox 2008). Limited published data (Li et al. 2008) from Lombok and Sumbawa indicate that this boundary may be at the eastern end, between Sumbawa and Flores. However, the low number of samples and limited number of SNPs that were typed make that result tentative.

[Figure 2](#) sheds light on the haplogroups that differ most significantly between these regions. Notably, haplogroups C-M38\*, M-P34, and S-M254 account for more than half of Y chromosomes in the east, and yet they are nearly absent in the west. On the other hand, the O lineages O-P203, O-M95\* and O-M119\* account for more than 60% of



**Fig. 3.** MDS plot of 26 populations based on  $\Phi_{ST}$  genetic distances. Open white circles—western Indonesia; full black circles—Oceania; open white squares—eastern Indonesia; and full black squares—Southeast Asian populations. For three-letter codes, see figure 1. Populations from Papua New Guinea, Melanesia, and Polynesia are indicated with letter codes PNG, MEL, and POL, respectively.

Y chromosomes in the west, and these lineages are found at less than 10% east of Wallace’s biogeographical line. The hypotheses of a Melanesian origin of haplogroups C-M38\*, M-P34, and S-M254 and an Asian origin of haplogroup O are fairly well accepted in the literature (see below); however, this scenario in itself does not explain why we find such an abrupt transition in eastern Indonesia. In the next sections, we address this question by performing a comparative analysis of the distribution of Y chromosome haplogroups found in Southeast Asia, Indonesia, and Oceania and then formulate a model that incorporates multiple episodes of colonization across the region.

### Y Chromosomes of the Initial Settlement of Indonesia

The current consensus is that the mutations defining haplogroup C originated somewhere in Asia after the initial exodus of anatomically modern humans from sub-Saharan Africa (Hammer et al. 1998; Underhill et al. 2000; Ke et al. 2001; YCC 2002; Karafet et al. 2008). The wide geographic

distribution of paragroup C-RPS4Y\* and its associated high Y-STR variance is consistent with an early entry into Indonesia. The finding that the derived lineages, paragroup C-M38\* and haplogroup C-M208, are found only east of Wallace’s biogeographical line parallels results of previous studies concluding that these mutations arose in eastern Indonesia or Melanesia (Kayser et al. 2003, 2006, 2008; Scheinfeldt et al. 2006; Cox et al. 2007; Mona et al. 2007).

The fine-scale sampling in this survey shows that C-M38\* attains its highest frequency (and STR diversity) in eastern Indonesia rather than in Melanesia (supplementary table S3, Supplementary Material online), suggesting its origin may be further west than previously hypothesized. Although this general west-to-east pattern of ancestral and derived C lineages may reflect an ancient west-to-east settlement pattern, it is important to note that both C-RPS4Y\* and C-M38\* are paragroups, and as such, there are likely to be undiscovered downstream SNPs associated with these lineages. So, if we can accept the inference that M208 is an old Melanesian marker (Cox et al. 2007), then

**Table 1.** Analysis of Molecular Variance.

	N	Number of Populations	Number of Groups	Within Populations		Among Populations within Groups		Among Groups	
				$\Phi_{ST}$	Variance (%)	$\Phi_{SC}$	Variance (%)	$\Phi_{CT}$	Variance (%)
Indonesian populations	1917	32	1	0.40	59.8				
Indonesian islands	1917	32	13	0.41	59.0	0.28	22.5	0.19	18.6
Western vs. eastern populations	1917	32	2	0.47	52.8	0.27	19.6	0.28	27.6
Global (Hammer et al. 2001)	2858	50	1	0.36	64.0				
Siberia (Cruciani et al. 2002)	1432	28	1	0.35	65.0				
Africa (Karafet et al. 2002)	608	22	1	0.34	65.8				
Africa (Wood et al. 2005)	1122	40	1	0.32	68.2				

we should also consider all underived Indonesian C\* lineages as the genetic legacy of the initial Paleolithic settlers of region. Because the presumed initial migration route took the early settlers through what is now western Indonesia, the lower frequency of both C\* and its derived subclades in the west (fig. 2) could be the result of either a continual eastward migration of the initial settlers (i.e., settlements were not permanently established in western Indonesia) or later waves of (partial) replacement.

Like haplogroup C, K-M9 also represents an early founding lineage of the region. Today, paragroup K-M9\* is widely distributed from Southeast Asia to Oceania and is associated with high levels of STR diversity (supplementary table S3, Supplementary Material online). Two major subclades of K (M and S) are thought to have arisen in Melanesia sometime after its initial colonization (Kayser et al. 2001, 2003, 2006). In addition, derived haplogroups within M and S (M-P34 and S-M254) have been proposed to be markers of the expansion of speakers of Trans–New Guinea phylum languages from the interior of mainland New Guinea to surrounding Melanesian and eastern Indonesian islands 6–10 ka ago (Kayser et al. 2006; Mona et al. 2007). Interestingly, we find M-P34 and S-M254 at moderate frequencies (i.e., 10% and ~12%, respectively) in our eastern Indonesian sample, and Y-STR diversity associated with M-P34 in eastern Indonesia outpaces that in New Guinea (supplementary table S3, Supplementary Material online).

Given that C-M38, M-P256, M-P34, and S-M230 are common both in eastern Indonesia and in Melanesia, it is difficult to differentiate between these two regions as the homeland of these lineages. One possible explanation for the observed haplogroup distribution is bidirectional dispersal between eastern Indonesia and New Guinea across the Banda and Arafura seas. Indeed, the Papuan languages spoken in eastern Indonesia may be relatively recent translocations from western New Guinea (Pawley 2002). However, a simple model of recurrent gene flow does not explain why some haplogroups in figure 2 are specific to New Guinea/Oceania or the observation that none of the 71 different M-P34 12-locus STR haplotypes associated with M-P34 is shared between New Guinea and eastern Indonesia (data not shown). As in the case of C-RPS4Y\*, it is important to remember that paragroup K-M9\* chromosomes are widespread across Indonesia and reflect the ancient substrate upon which “Papuan” mutations occurred (i.e., P256 and M230). The moderate frequency of K-M9\*, especially in eastern Indonesia (7%) and Oceania (11%), suggests that additional undiscovered markers that are informative for tracing other migratory trends may be present on these chromosomes.

### Paleolithic versus Austronesian Paternal Contribution to Indonesia

In contrast with the C, M, and S haplogroups discussed above, haplogroup O and its major subclades (O-M122, O-M119, and O-M95) are thought to have originated on the Pleistocene Asian mainland (Su et al. 1999, 2000; Capelli

et al. 2001; Kayser et al. 2001, 2003, 2006; Shi et al. 2005; Li et al. 2008). These haplogroups, especially O-M119 and O-M122, have drawn considerable attention as possible markers of the Austronesian expansion and the spread of agricultural technology from China/Taiwan into island Southeast Asia, Melanesia, and Polynesia (Kayser et al. 2000, 2001, 2003, 2006, 2008; Su et al. 2000; Capelli et al. 2001; Hurles et al. 2002). Accordingly, we suggest that lineages within haplogroup O represent a later contribution to the genetic strata of Indonesia. A careful examination of the distributions of particular O lineages does, indeed, support the hypothesis of an Austronesian connection; however, the presence of the most common O haplogroups in Indonesia may be better explained by earlier Paleolithic contributions.

In western Indonesia, O-M175-derived lineages account for more than 80% of Y chromosomes. Paragroup O-M95\* is notable because it is common in western Indonesia (e.g., achieving frequencies of >50% on Java and Bali) but is virtually absent east of Flores. This paragroup is widespread in south and southeast Asia (Karafet et al. 2005), and it is unclear when it initially entered western Indonesia. The ancient time to the most recent common ancestor of this lineage and Asian distribution supports the hypothesis of a pre-Austronesian incursion of O-M95\* into western Indonesia from Southeast Asia (Kumar et al. 2007). It has also been suggested that O-M95\* chromosomes appeared in Indonesia after the initial colonization of the Pacific by Austronesian farmers and that the high frequency of O-M95\* in Bali and Java may reflect an even more recent influx of males from the Indian subcontinent (e.g., possibly concomitant with the spread of Hinduism and the establishment of Indian kingdoms in the first millennium) (Karafet et al. 2005).

O-M122, a haplogroup with a wide Southeast Asian distribution, has been proposed as a marker of the spread of Austronesian-speaking populations (Kayser et al. 2000, 2006; Capelli et al. 2001; Karafet et al. 2005; Shi et al. 2005; Scheinfeldt et al. 2006). The typing of additional SNPs within the O-M122 lineage results in a much more fine-grained picture of the distribution of this clade in mainland and island Southeast Asia. High frequencies of the derived O-M134, O-M7, and O-002611 subclades are observed in different ethnic groups in China and mainland Southeast Asia. In contrast, these haplogroups are absent or only marginally present among Indonesian, Taiwanese aboriginal, and Pacific populations. The low frequency of O-M7 in western Indonesia (i.e., Bali, Java, and Borneo) and O-M134 in Polynesia most likely reflects recent connections with mainland China (see next section).

More notable are the results of typing the novel marker P201, which has the effect of converting almost all chromosomes outside of mainland Asia that were previously identified as O-M122\* to O-P201\*. Given this widespread pattern, it may be that O-M122 chromosomes previously found at high frequency on many Pacific Islands (Kayser et al. 2006) are actually O-P201\* (in our small sample of 64 Polynesians typed here, 11 of 15 O-M122 carriers are

also O-P201\*). Unlike the lineages discussed so far, the frequencies of O-P201\* chromosomes are fairly constant (~7%) right across the major regions sampled here (fig. 2). Importantly, despite its relatively low frequency on Taiwan (~6%), genetic distances based on STR variation associated with P201 chromosomes reveal a much closer relationship among Taiwanese aboriginals/Filipinos, Indonesians, and Oceanians than between any of these groups and mainland Southeast Asians (supplementary table S4, online). Therefore, we hypothesize that this new marker traces the large population expansion associated with the spread of Austronesian languages and culture.

With regard to the hypothesis of a Taiwanese origin of the Austronesian people (Bellwood 2007), the M119 mutation has drawn particular interest because these chromosomes are dominant among Taiwanese aboriginal groups (Su et al. 1999, 2000; Kayser et al. 2000, 2001, 2003, 2006, 2008; Capelli et al. 2001; Hurler et al. 2002; Karafet et al. 2005; Li et al. 2008). Previously, we found little geographic structure for O-M119\* STR haplotypes and proposed that the absence of such structure might reflect an ancient dispersal, migrations from different source populations, and/or sustained gene flow from Southeast Asia (Karafet et al. 2005). We also suggested that O-M119\* chromosomes might represent a heterogeneous group of not-yet-identified haplogroups. Here, we have substantiated the second claim and find that the additional mutation P203 allows significantly better geographic resolution of O-M119 chromosomes than previously possible. Many (but not all) chromosomes that were previously identified as O-M119\* are now O-P203, including those from the Taiwanese aboriginals sampled here (i.e., all 34 M119 chromosomes are marked with the derived P203 mutation). The current results reveal that while the “ancestral” O-M119\* lineage is virtually absent in mainland Southeast Asia, the derived O-P203 subclade is frequent there, as well as in western Indonesia (fig. 2).

The high frequency of O-P203 in western Indonesia (~24%) and much lower frequency in eastern Indonesia (~2%) and Oceania (<1%) (fig. 2) are not easily explained by a model that links its spread solely with the Austronesian expansion. A highly reticulated MJ network based on Y-STR diversity is uninformative on the question of a Taiwanese affinity of Indonesian O-P203 chromosomes (data not shown). Genetic distances based on O-P203 Y-STR haplotypes are equally similar between Taiwanese aboriginals and Indonesians (RST = 0.151) and Taiwanese aboriginals and mainland Southeast Asian populations (RST = 0.156). However, sharing of 12-locus Y-STR haplotypes associated with P203 chromosomes was found between Taiwanese aboriginals and Indonesians (i.e., from Nias, Mentawai, Java, and Bali), whereas no such sharing was found between the Taiwanese aboriginal and mainland Southeast Asian P203 chromosomes in our sample. This suggests that some portion of Indonesian O-P203 chromosomes may have migrated from Taiwan.

A stronger case can be made for a Taiwanese affinity of Indonesian O-M110 chromosomes (Karafet et al. 2005;

Kayser et al. 2008). An MJ network of 12-locus Y-STR haplotypes associated with O-M110 chromosomes has a central node composed of 2 Taiwanese aboriginal and 14 western Indonesian chromosomes (fig. 4). However, although present in ~19% of Taiwanese aboriginals sampled here, O-M110 chromosomes are found at low frequencies elsewhere (with the exception of Nias), with only 1 of 182 samples in Oceania belonging to this haplogroup (fig. 2, supplementary table S2, Supplementary Material online). In fact, a larger number of O-M119\* chromosomes was found in Oceania (4 of 182). Thus, while it is possible that haplogroups O-M110 and O-P203 mark an expansion out of Taiwan, it is also possible that at least part of their distribution, especially in western Indonesia, reflects a pre-Austronesian dispersal of these haplogroups in island Southeast Asia. In this way, these lineages are reminiscent of mtDNA haplogroup E whose distribution in Taiwan and island Southeast Asia has been suggested to predate the Austronesian expansion (Hill et al. 2007; Soares et al. 2008).

### Y Chromosomes Entering Indonesia in Historic Times

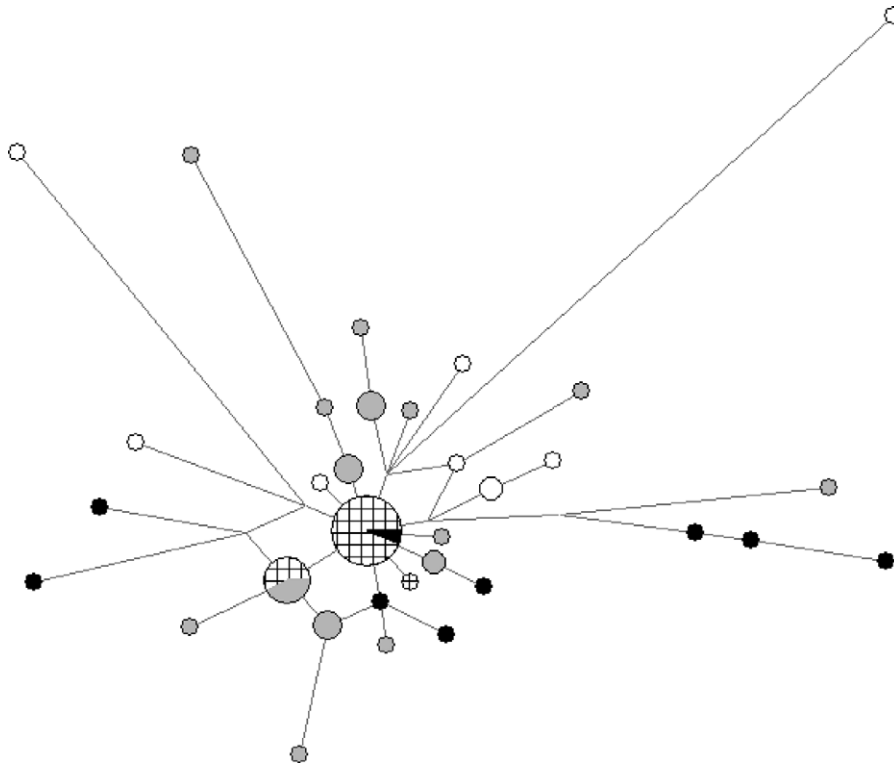
The remaining Indonesian chromosomes represent a range of haplogroups and account for only ~6% of the total sample. Some of these lineages have clear affinities with distant geographic regions and may mark incursions into Indonesia in recent times. For example, Indian contact <2.5 ka ago may have introduced lineages within haplogroups H (H-M69, H-Apt, and H-M52), R (R-M124), and Q (Q-M346) to Indonesia (which accounts for ~2% of Indonesian Y chromosomes; supplementary table S2, Supplementary Material online). Interestingly, H-M69, H-Apt, and Q-M346 are observed only in Bali, which has the highest frequency of Indian Y chromosomes (Kivisild et al. 2003; Karafet et al. 2005; Gutala et al. 2006; Sengupta et al. 2006), whereas Y chromosomes with Indian affinity (e.g., H-M52 and R-M124) are also found at lower frequencies in Java, Borneo, and Sumatra.

Although it is more difficult to pinpoint the source of J-M304, L-M20, and R-M17 lineages because they are present at relatively high frequencies in both Indian and Near Eastern populations (Karafet et al. 2005), it is plausible that some of these lineages entered Indonesia recently with the spread of Islam. Overall, Indian and Arab influences are restricted to western Indonesia, particularly the adjacent islands of Java and Bali, where Indian and Arab cultural influences are self-evident (Lansing 1983). Chinese influences occur here too (e.g., O-M7 is present at <1% in Bali and ~11% in Java); however, they are more important in Borneo (e.g., O-M7 is found at ~20%), a major Chinese outpost dating from the Han dynasty (Ricklefs 1993; Taylor 2003).

### A Four-Stage Colonization Model for the Region

To integrate our phylogeographic inferences for the diverse set of haplogroups and populations surveyed here, we now formulate a four-stage colonization model that attempts to





**FIG. 4.** Median-joining network for haplogroup O-M110 based on variation at 12 Y-STRs. Haplotypes are represented by circles with the area proportional to the number of individuals carrying that haplotype. Branch lengths are proportional to the number of one-repeat mutations separating two haplotypes. Color coding of haplotypes: black (Taiwanese aboriginals), cross-hatched (eastern Indonesians), gray (western Indonesians), and white (other).

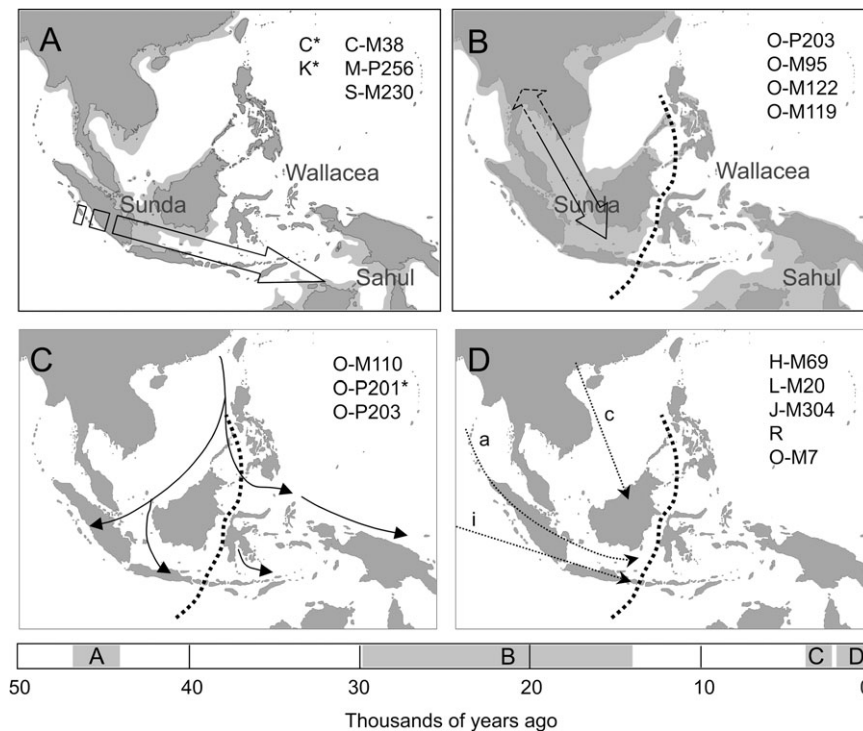
account for the current pattern of Y chromosome variation in Indonesia (fig. 5). In the first stage, a Late Pleistocene arrival of the first anatomically modern settlers introduces basal C and K lineages to the entire region (which eventually give rise to haplogroups C-M38, M-P256, and S-M230 in eastern Indonesia/Melanesia). At the time of this expansion (~45–50 ka ago), sea levels were much lower and the shape of the coastline was very different from that of today (fig. 5A). For example, Sumatra, Java, Borneo, and other small groups of islands formed a direct extension of the Asian mainland—the Sunda continental shelf or Sundaland—and the deep-water channel separating Bali and Lombok marked the end of this land mass. Further travel eastward to Wallacea and Sahul required crossing water.

Climate change during, and at the end of, the last glacial period (33–16 ka ago) may have had an important effect on human diversity in the region, including an overall population decline during this period (Bird et al. 2005; Pope and Terrell 2008). After 19 ka ago, the sea level began to rise again, with Southeast Asia reaching its present coastline by around 8 ka ago (Mulvaney and Kamminga 1999). These climatic changes may have spurred a second round of expansion of hunter-gatherers into Sundaland from further north on the mainland (Soares et al. 2008). Indeed, the spread of the Southeast Asian Hoabinhian culture into Sumatra may be one tangible marker of these movements (Bellwood 2007).

We posit that dispersals of hunter-gatherers radiating over an extended period of time (e.g., 8–35 ka ago) intro-

duced several major subclades of haplogroup O to Indonesia (e.g., O-M119, O-M95, O-P203, and O-M122) (fig. 5B). However, the current data do not inform us about the age of the sharp Y chromosome boundary between western and eastern Indonesia. Cox et al. (2010) genotyped a small set of ancestry informative markers on the autosomes and X chromosome and found a similar transition from Asian to Melanesian ancestry over a narrow geographic region in eastern Indonesia. Although clines that extend over long distances and that originate from well-differentiated source populations may be remarkably stable (Wijsman and Cavalli-Sforza 1984; Cavalli-Sforza et al. 1994; Fix 1999), it is not clear how divergent human groups practicing similar subsistence strategies and living in close geographic proximity could have remained (semi-) isolated since the Late Pleistocene or why they may have done so. An alternate explanation is that the boundary is fleeting and formed recently through the mixing of groups that shared ancient common ancestry. Future studies of many more genome-wide markers should help to determine whether this ancestry cline reflects contact between differentiated hunter-gatherer groups in the Late Pleistocene (i.e., during this second stage of colonization in fig. 5B) or more recent mixing resulting from the spread of farming populations.

The third stage of colonization corresponds to the Austronesian expansions. This maritime dispersal of rice agriculturists from southern China/Taiwan, beginning between 5.5 and 4.0 ka ago, resulted in the expansion of Austronesian languages throughout the region (Bellwood 2007). We



**Fig. 5.** A multistage colonization model for Indonesia. (A) Initial wave of colonization 40–50 ka ago, (B) Paleolithic contribution from mainland Asia, (C) Austronesian expansion, and (D) migration in historic times (see text for details). Haplogroups and paragroups listed in each panel are postulated to have arrived during this stage. Arrows do not denote precisely defined geographic routes. The dotted arrowhead in panel (B) indicates possible bidirectional gene flow. The dotted line in panels (B), (C), and (D) represents Wallace’s biogeographic boundary. Arrows in panel (C) adapted from Gray and Jordan (2000). Small letters associated with arrows in panel (D) refer to migrations from India (i), Arabia (a), and China (c). Shaded areas with capital letters (A–D) under panels refer to approximate time frame for each stage of colonization. Coastlines in panels (A) and (B) are drawn with sea levels 50 and 120 m below current levels (C and D) ([http://www.fieldmuseum.org/research\\_collections/zoology/zoo\\_sites/seamaps/mapindex1.htm](http://www.fieldmuseum.org/research_collections/zoology/zoo_sites/seamaps/mapindex1.htm)).

posit that it also led to the migration of haplogroups O-P201 and possibly O-M110 and O-P203 to both sides of Wallace’s line as it penetrated the Indonesian region from the north by sea (fig. 5C). The final phase of settlement involves several incursions, especially into western Indonesia, during historic times (fig. 5D). The first of these is the spread of Hinduism and the establishment of Indian kingdoms, which took place between the 3rd and 13th centuries (Peter 1982; Karafet et al. 2005). This resulted in the introduction of multiple haplogroups that derive from south Asia and today are found at low frequency in Bali, Java, Borneo, and Sumatra. The second is the spread of Islam (Tibbetts 1979), ultimately from Arabia, which may have introduced paternal lineages within haplogroups J, L, and R. Finally, haplogroup O-M7 is a marker of recent Chinese influence.

Although the initial colonizers entered a landscape previously unoccupied by anatomically modern humans, subsequent expansions occurred into territory already inhabited by genetically differentiated groups. However, there is no evidence that any of these expansions resulted in a complete replacement of the Y chromosomes of previous inhabitants. This is evidenced by the survival of older genetic strata in both western and eastern Indonesia. Likewise, haplogroup O lineages associated with the Austronesians are only found at <20% in western Indonesia and less

frequently in eastern Indonesia and Melanesia, despite the fact that Austronesian languages predominate throughout most of the region. Finally, the migration processes that led to the most dramatic cultural and social change, the “Indianization” and “Islamization” of Indonesia, resulted in the smallest amount of genetic change (i.e., they only account for a small percentage of Indonesian Y chromosomes). Interestingly, the earliest settlers left the most durable pattern in the Y chromosome data: a sharp transition from Asian O haplogroups to Melanesian haplogroups (C, M, and S) over a small area in eastern Indonesia. From a Y chromosome perspective, more recent incursions of culture occurred without much of a corresponding genetic effect.

### Supplementary Material

Supplementary tables S1–S4 and figures S1–S4 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

### Acknowledgments

Indonesian samples were obtained by JSL and HS and by Golfiani Malik, Wuryantari Setiadi, Loa Helena Suryadi, and Meryanne Tumonggor of the Eijkman Institute for Molecular Biology, Jakarta, Indonesia, with the assistance

of Indonesian Public Health clinic staff, following protocols for the protection of human subjects established by both the Eijkman Institute and the University of Arizona institutional review boards. Permission to conduct research in Indonesia was granted by the Indonesian Institute of Sciences. This research was supported by the National Science Foundation.

## References

- Bandelt HJ, Forster P, Rohl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol.* 16:37–48.
- Barker G, Barton H, Bird M, et al. (27 co-authors). 2007. 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.
- Bellwood P. 1997. Prehistory of the Indo-Malaysian archipelago. Honolulu (HI): University of Hawaii Press.
- Bellwood P. 2007. Prehistory of the Indo-Malaysian archipelago. Canberra (Australia): Australian National University.
- Bellwood P, White P. 2005. Domesticated pigs in eastern Indonesia. *Science* 309: (author reply).
- Bird MI, Taylor D, Hunt C. 2005. Environments of insular Southeast Asia during the Last Glacial Period: a savanna corridor in Sundaland? *Quat Sci Rev.* 24:2228–2242.
- Blust R. 1995. The prehistory of the Austronesian-speaking peoples—a view from language. *J World Prehist.* 9:453–510.
- Capelli C, Wilson JF, Richards M, Stumpf MP, Gratrix F, Oppenheimer S, Underhill P, Pascali VL, Ko TM, Goldstein DB. 2001. A predominantly indigenous paternal heritage for the Austronesian-speaking peoples of insular Southeast Asia and Oceania. *Am J Hum Genet.* 68:432–443.
- Cavalli-Sforza LL, Menozzi P, Piazza A. 1994. The history and geography of human genes. Princeton (NJ): Princeton University Press.
- Cox MP. 2008. The genetic environment of Melanesia: clines, clusters and contact. In: Koven VT, editor. Population genetics research progress. New York: Nova Science Publishers. p. 45–83.
- Cox MP, Karafet TM, Lansing JS, Sudoyo H, Hammer MF. 2010. Autosomal and X-linked SNPs reveal a steep Asian-Melanesian ancestry cline in eastern Indonesia and a sex-bias in admixture rates. *Proc R Soc B.* Advance Access published March 27, 2010, doi:10.1098/rspb.2009.2041.
- Cox MP, Mirazon Lahr M. 2006. 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.
- Cox MP, Redd AJ, Karafet TM, Ponder CA, Lansing JS, Sudoyo H, Hammer MF. 2007. A Polynesian motif on the Y chromosome: population structure in remote Oceania. *Hum Biol.* 79:525–535.
- Cruciani F, Santolamazza P, Shen P, et al. (16 co-authors). 2002. A back migration from Asia to sub-Saharan Africa is supported by high-resolution analysis of human Y-chromosome haplotypes. *Am J Hum Genet.* 70:1197–1214.
- Denham T, Donohue M. 2009. Pre-Austronesian dispersal of banana cultivars West from New Guinea: linguistic relics from Eastern Indonesia. *Archaeol Oceania.* 44:18–28.
- Diamond J, Bellwood P. 2003. Farmers and their languages: the first expansions. *Science* 300:597–603.
- Diamond JM. 2000. Taiwan’s gift to the world. *Nature* 403:709–710.
- Excoffier L, Laval G, Schneider S. 2005. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinform Online.* 1:47–50.
- Fix AG. 1999. Migration and colonization in human microevolution. Cambridge: Cambridge University Press.
- Gray RD, Jordan FM. 2000. Language trees support the express-train sequence of Austronesian expansion. *Nature* 405:1052–1055.
- Gutala R, Carvalho-Silva DR, Jin L, Yngvadottir B, Avadhanula V, Nanne K, Singh L, Chakraborty R, Tyler-Smith C. 2006. A shared Y-chromosomal heritage between Muslims and Hindus in India. *Hum Genet.* 120:543–551.
- Hammer MF, Karafet T, Rasanayagam A, Wood ET, Altheide TK, Jenkins T, Griffiths RC, Templeton AR, Zegura SL. 1998. Out of Africa and back again: nested cladistic analysis of human Y chromosome variation. *Mol Biol Evol.* 15:427–441.
- Hammer MF, Karafet TM, Redd AJ, Jarjanazi H, Santachiara-Benerecetti S, Soodyall H, Zegura SL. 2001. Hierarchical patterns of global human Y-chromosome diversity. *Mol Biol Evol.* 18:1189–1203.
- Hill C, Soares P, Mormina M, et al. (11 co-authors). 2007. A mitochondrial stratigraphy for island southeast Asia. *Am J Hum Genet.* 80:29–43.
- Howells WW. 1973. The Pacific Islanders. New York: Charles Scribner’s Sons.
- Hurles ME, Nicholson J, Bosch E, Renfrew C, Sykes BC, Jobling MA. 2002. Y chromosomal evidence for the origins of oceanic-speaking peoples. *Genetics* 160:289–303.
- Karafet T, Xu L, Du R, Wang W, Feng S, Wells RS, Redd AJ, Zegura SL, Hammer MF. 2001. Paternal population history of East Asia: sources, patterns, and microevolutionary processes. *Am J Hum Genet.* 69:615–628.
- Karafet TM, Lansing JS, Redd AJ, et al. (11 co-authors). 2005. 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.
- Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, Hammer MF. 2008. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res.* 18:830–838.
- Karafet TM, Osipova LP, Gubina MA, Posukh OL, Zegura SL, Hammer MF. 2002. High levels of Y-chromosome differentiation among native Siberian populations and the genetic signature of a boreal hunter-gatherer way of life. *Hum Biol.* 74:761–789.
- Kayser M, Brauer S, Cordaux R, et al. (15 co-authors). 2006. Melanesian and Asian origins of Polynesians: mtDNA and Y chromosome gradients across the Pacific. *Mol Biol Evol.* 23:2234–2244.
- Kayser M, Brauer S, Weiss G, Schiefenovel W, Underhill P, Shen P, Oefner P, Tommaseo-Ponzetta M, Stoneking M. 2003. Reduced Y-chromosome, but not mitochondrial DNA, diversity in human populations from West New Guinea. *Am J Hum Genet.* 72: 281–302.
- Kayser M, Brauer S, Weiss G, Schiefenovel W, Underhill PA, Stoneking M. 2001. Independent histories of human Y chromosomes from Melanesia and Australia. *Am J Hum Genet.* 68:173–190.
- Kayser M, Brauer S, Weiss G, Underhill PA, Roewer L, Schiefenovel W, Stoneking M. 2000. Melanesian origin of Polynesian Y chromosomes. *Curr Biol.* 10:1237–1246.
- Kayser M, Choi Y, van Oven M, Mona S, Brauer S, Trent RJ, Suarika D, Schiefenovel W, Stoneking M. 2008. The impact of the Austronesian expansion: evidence from mtDNA and Y chromosome diversity in the Admiralty Islands of Melanesia. *Mol Biol Evol.* 25:1362–1374.
- Ke Y, Su B, Song X, et al. (23 co-authors). 2001. African origin of modern humans in East Asia: a tale of 12,000 Y chromosomes. *Science* 292:1151–1153.
- Kirch PV. 1997. The Lapita peoples: ancestors of the Oceanic world. Oxford: Blackwell.
- Kirch PV. 2000. On the road of the winds: an archaeological history of the Pacific islands before European contact. Berkeley (CA): University of California Press.

- Kivisild T, Rootsi S, Metspalu M, et al. (18 co-authors). 2003. The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. *Am J Hum Genet.* 72:313–332.
- Kruskal JB. 1964. Multidimensional scaling by optimizing goodness of fit to a nometric hypothesis. *Psychometrika* 29:1–27.
- Kumar V, Reddy AN, Babu JP, Rao TN, Langstieh BT, Thangaraj K, Reddy AG, Singh L, Reddy BM. 2007. Y-chromosome evidence suggests a common paternal heritage of Austro-Asiatic populations. *BMC Evol Biol.* 7:47.
- Lansing JS. 1983. Perfect order: recognizing complexity in Bali. Princeton (NJ): Princeton University Press.
- Lansing JS, Cox MP, Downey SS, et al. (11 co-authors). 2007. Coevolution of languages and genes on the island of Sumba, eastern Indonesia. *Proc Natl Acad Sci U S A.* 104:16022–16026.
- Lansing JS, Watkins JC, Hallmark B, Cox MP, Karafet TM, Sudoyo H, Hammer MF. 2008. Male dominance rarely skews the frequency distribution of Y chromosome haplotypes in human populations. *Proc Natl Acad Sci U S A.* 105:11645–11650.
- Larson G, Dobney K, Albarella U, et al. (13 co-authors). 2005. Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science* 307:1618–1621.
- Leavesley MG, Chappell J. 2004. Buang Merabak: additional early radiocarbon evidence of the colonization of the Bismarck Archipelago, Papua New Guinea. *Antiquity* 78:No. 301.
- Li H, Wen B, Chen SJ, et al. (19 co-authors). 2008. Paternal genetic affinity between Western Austronesians and Daic populations. *BMC Evol Biol.* 8:146.
- Lum JK, McIntyre JK, Greger DL, Huffman KW, Vilar MG. 2006. Recent Southeast Asian domestication and Lapita dispersal of sacred male pseudohermaphroditic “tuskers” and hairless pigs of Vanuatu. *Proc Natl Acad Sci U S A.* 103:17190–17195.
- Meacham W. 1984. On the improbability of Austronesian origins in South China. *Asian Perspect.* 26:89–106.
- Mona S, Grunz KE, Brauer S, et al. (11 co-authors). 2009. Genetic admixture history of Eastern Indonesia as revealed by Y-chromosome and mitochondrial DNA analysis. *Mol Biol Evol.* 26:1865–1877.
- Mona S, Tommaseo-Ponzetta M, Brauer S, Sudoyo H, Marzuki S, Kayser M. 2007. Patterns of Y-chromosome diversity intersect with the Trans-New Guinea hypothesis. *Mol Biol Evol.* 24:2546–2555.
- Mulvaney DJ, Kamminga J. 1999. Prehistory of Australia. Washington: Smithsonian Institution Press.
- Nei M. 1987. Molecular evolutionary genetics. New York: Columbia University Press.
- O’Connell JF, Allen J. 2004. Dating the colonization of Sahul (Pleistocene Australia-New Guinea): a review of recent research. *J Archaeol Sci.* 31:835–853.
- Pawley A. 2002. The Austronesian dispersal: languages, technologies and people. In: Bellwood P, Renfrew C, editors. Examining the farming/language dispersal hypothesis. Cambridge: McDonald Institute for Archaeological Research. p. 251–273.
- Peter L. 1982. The next Empire. *Futures* 14:47–61.
- Pietruszewsky M. 1994. Pacific-Asian relationships: a physical anthropological perspective. *Ocean Linguist.* 33:407–429.
- Pope KO, Terrell JE. 2008. Environmental setting of human migrations in the circum-Pacific region. *J Biogeogr.* 35:1–21.
- Redd AJ, Agellon AB, Kearney VA, Contreras VA, Karafet T, Park H, de Knijff P, Butler JM, Hammer MF. 2002a. Forensic value of 14 novel STRs on the human Y chromosome. *Forensic Sci Int.* 130:97–111.
- Redd AJ, Roberts-Thomson J, Karafet T, Bamshad M, Jorde LB, Naidu JM, Walsh B, Hammer MF. 2002b. Gene flow from the Indian subcontinent to Australia: evidence from the Y chromosome. *Curr Biol.* 12:673–677.
- Ricklefs M. 1993. A history of modern Indonesia since c. 1300. London: MacMillan.
- Roberts RG, Flannery TF, Ayliffe LK, Yoshida H, Olley JM, Prideaux GJ, Laslett GM, Baynes A, Smith MA, Jones R. 2001. New ages for the last Australian megafauna: continent-wide extinction about 46,000 years ago. *Science* 1888–1892.
- Rohlf FJ. 1998. NTSYS-pc: Numerical taxonomy and multivariate analysis system. Setauket (NY): Exeter Software.
- Scheinfeldt L, Friedlaender F, Friedlaender J, Latham K, Koki G, Karafet T, Hammer M, Lorenz J. 2006. Unexpected NRY chromosome variation in Northern Island Melanesia. *Mol Biol Evol.* 23:1628–1641.
- Schlecht J, Kaplan ME, Barnard K, Karafet T, Hammer MF, Merchant NC. 2008. Machine-learning approaches for classifying haplogroup from Y chromosome STR data. *PLoS Comput Biol.* 4:e1000093.
- Sengupta S, Zhivotovsky LA, King R, et al. (15 co-authors). 2006. Polarity and temporality of high-resolution y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of Central Asian pastoralists. *Am J Hum Genet.* 78:202–221.
- Shi H, Dong YL, Wen B, Xiao CJ, Underhill PA, Shen PD, Chakraborty R, Jin L, Su B. 2005. Y-chromosome evidence of southern origin of the East Asian-specific haplogroup O3-M122. *Am J Hum Genet.* 77:408–419.
- Soares P, Trejaut JA, Loo JH, et al. (14 co-authors). 2008. Climate change and postglacial human dispersals in Southeast Asia. *Mol Biol Evol.* 25:1209–1218.
- Su B, Jin L, Underhill P, et al. (11 co-authors). 2000. Polynesian origins: insights from the Y chromosome. *Proc Natl Acad Sci U S A.* 97:8225–8228.
- Su B, Xiao J, Underhill P, et al. (21 co-authors). 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.
- Taylor JG. 2003. Indonesia. New Haven (CT): Yale University Press.
- Tibbetts G. 1979. A study of the Arabic texts containing material on South-East Asia. Leiden (The Netherlands): E. J. Brill.
- Underhill PA, Shen P, Lin AA, et al. (21 co-authors). 2000. Y chromosome sequence variation and the history of human populations. *Nat Genet.* 26:358–361.
- Wallace AR. 1869. The Malay archipelago: the land of the orang-utan, and the bird of paradise. London: Macmillan and Co.
- Wijsman EM, Cavalli-Sforza LL. 1984. Migration and genetic population-structure with special reference to humans. *Ann Rev Ecol Syst.* 15:279–301.
- Wood ET, Stover DA, Ehret C, et al. (11 co-authors). 2005. Contrasting patterns of Y chromosome and mtDNA variation in Africa: evidence for sex-biased demographic processes. *Eur J Hum Genet.* 13:867–876.
- Y Chromosome Consortium (YCC). 2002. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res.* 12:339–348.