

---

## ***Balinese Y-Chromosome Perspective on the Peopling of Indonesia: Genetic Contributions from Pre-Neolithic Hunter-Gatherers, Austronesian Farmers, and Indian Traders***

TATIANA M. KARAFET,<sup>1</sup> J. S. LANSING,<sup>2,3</sup> ALAN J. REDD,<sup>1</sup> JOSEPH C. WATKINS,<sup>4</sup> S. P. K. SURATA,<sup>5</sup> W. A. ARTHAWIGUNA,<sup>6</sup> LAURA MAYER,<sup>1</sup> MICHAEL BAMSHAD,<sup>7</sup> LYNN B. JORDE,<sup>7</sup> AND MICHAEL F. HAMMER<sup>1,2</sup>

**Abstract** The island of Bali lies near the center of the southern chain of islands in the Indonesian archipelago, which served as a stepping-stone for early migrations of hunter-gatherers to Melanesia and Australia and for more recent migrations of Austronesian farmers from mainland Southeast Asia to the Pacific. Bali is the only Indonesian island with a population that currently practices the Hindu religion and preserves various other Indian cultural, linguistic, and artistic traditions (Lansing 1983). Here, we examine genetic variation on the Y chromosomes of 551 Balinese men to investigate the relative contributions of Austronesian farmers and pre-Neolithic hunter-gatherers to the contemporary Balinese paternal gene pool and to test the hypothesis of recent paternal gene flow from the Indian subcontinent. Seventy-one Y-chromosome binary polymorphisms (single nucleotide polymorphisms, SNPs) and 10 Y-chromosome-linked short tandem repeats (STRs) were genotyped on a sample of 1,989 Y chromosomes from 20 populations representing Indonesia (including Bali), southern China, Southeast Asia, South Asia, the Near East, and Oceania. SNP genotyping revealed 22 Balinese lineages, 3 of which (O-M95, O-M119, and O-M122) account for nearly 83.7% of Balinese Y chromosomes. Phylogeographic analyses suggest that all three major Y-chromosome haplogroups migrated to Bali with the arrival of Austronesian speakers; however, STR diversity patterns associated with these haplogroups are complex and may be explained by multiple waves of Austronesian expansion to Indonesia by different routes. Approximately 2.2% of contemporary Balinese Y chromosomes (i.e., K-M9\*, K-M230, and

<sup>1</sup>Division of Biotechnology, Biosciences West, University of Arizona, Tucson, AZ 85721.

<sup>2</sup>Anthropology Department, University of Arizona, Tucson, AZ 85721.

<sup>3</sup>Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87521.

<sup>4</sup>Mathematics Department, University of Arizona, Tucson, AZ 85721.

<sup>5</sup>Maharaswati University, Tabanan, Bali, Indonesia.

<sup>6</sup>Balai Penelitian dan Pengkajian Teknologi Pertanian, Denpasar, Bali, Indonesia.

<sup>7</sup>Department of Human Genetics, University of Utah, Salt Lake City, UT.

*Human Biology*, February 2005, v. 77, no. 1, pp. 93–113

Copyright © 2005 Wayne State University Press, Detroit, Michigan 48201-1309

KEY WORDS: BALI, INDONESIA, Y CHROMOSOME, AUSTRONESIAN EXPANSION, INDIAN TRADERS, PRE-NEOLITHIC HUNTER-GATHERERS, VIETNAMESE, MALAYSIANS, PHILIPPINOS, TAIWANESE ABORIGINALS, SOUTHERN CHINESE, INDIANS, SRI LANKANS, SYRIANS, SAUDI ARABIANS, MELANESIA, MICRONESIA, POLYNESIA, PAPUA NEW GUINEA, OCEANIA.

M lineages) may represent the pre-Neolithic component of the Indonesian paternal gene pool. In contrast, eight other haplogroups (H, J, L, and R) making up approximately 12% of the Balinese paternal gene pool appear to have migrated to Bali from India. These results indicate that the Austronesian expansion had a profound effect on the composition of the Balinese paternal gene pool and that cultural transmission from India to Bali was accompanied by substantial levels of gene flow.

Bali is one of the stepping-stones in the land bridge that once connected the islands of Indonesia to the Asian mainland. Archeological and fossil evidence suggests that the earliest inhabitants of Australia and Papua New Guinea crossed this bridge 40,000 to 60,000 years ago (Thorne et al. 1999; Bowler et al. 2003). Nineteenth-century Dutch archeologists envisaged continuous human occupation in Bali stretching back into the Pleistocene (Lansing 1995). More recent archeological evidence indicates that Austronesian-speaking peoples settled on the coast of Indonesia (including Bali), Malaysia, southern Vietnam, and the Philippines before colonizing most of the inhabitable islands of the Pacific (Bellwood 1997). This has led to the consensus view among archeologists that Austronesian-speaking peoples migrated to Indonesia between 4,500 and 3,000 years ago from southern China and Taiwan and displaced an aboriginal population of hunter-gatherers. This view implies a southern Chinese or Taiwanese origin of Balinese genes. Rapid eastward migrations resulted in the spread of both the Austronesian language family and associated culture to coastal Melanesia and throughout Polynesia between 3,000 B.C. and A.D. 400 (Bellwood 1997).

An alternative view posits an indigenous origin of Austronesian languages in Melanesia or Southeast Asia (Dyen 1962; Oppenheimer 1998) with much less population replacement by Neolithic farmers. Under this model substantial genetic contributions from pre-Neolithic hunter-gatherers may be expected in contemporary Balinese. Genetic data from both mitochondrial DNA and the Y chromosome have recently contributed to this controversy. By mainly focusing on Polynesian populations, some researchers favor the rapid migration or “express train” (Diamond 1988) model (Sykes et al. 1995; Redd et al. 1995; Melton et al. 1998; Su, Jin et al. 2000), and others favor a major contribution to the Polynesian gene pool from eastern Indonesia and Melanesia (M. Richards et al. 1998; Capelli et al. 2001; Kayser et al. 2001; Hurles et al. 2002). Until now, there has been no analysis focusing on the genetic composition of the island of Bali.

A second set of questions pertains to the more recent history of Indonesia. Between the 3rd and 13th centuries A.D. dozens of Indic kingdoms appeared across Southeast Asia, from the plains of Cambodia and central Java to remote corners of Borneo and highland Burma (Lansing 1983). Bali became as deeply “Indianized” as any Southeast Asian society—worshipping Hindu and Buddhist deities, celebrating the great Hindu and Buddhist myths, and measuring social behavior against the standards of a caste system. Classical Indic civilization survived on Bali until the 20th century, long after the destruction of the other Indic

states. But the question of how Indian culture came to Bali has never been fully answered.

Two competing hypotheses concern the degree of demic versus cultural diffusion. Although there is no evidence that Bali was initially populated by Indians, archeological excavations provide evidence of Indian trade contacts going back about 2,000 years (Ardika and Bellwood 1991). Majumdar (1963) postulated wholesale colonization by Indian exiles, whereas van Leur (1955) argued that Indianization was wholly initiated by Southeast Asians who summoned Brahmins to visit their courts, creating merely a “thin and flaking glaze” of Indic language and customs. Although the spread of Islam was much more recent than that of Indian religions, it was also more successful: Islam became the dominant religion of Malaysia and Indonesia. Therefore, apart from Indianization, the possible influence of Persians and Arab Muslims has to be taken into account.

In the past three years the publication of a robust Y-chromosome haplogroup tree defined by more than 250 binary polymorphisms (Y Chromosome Consortium 2002; Jobling and Tyler-Smith 2003) has provided an opportunity to understand paternal population origins, relationships, and dispersals with more phylogenetic and geographic resolution than was previously possible. The addition of microsatellite data facilitates the estimation of haplogroup ages and population divergence times, both of which can be used to infer the chronology of recent human dispersal events. Indeed, the combination of these two kinds of data provides a valuable source of information for more precise identification of ancestral relationships, patterns of gene flow, and the effects of various demographic processes. Here, we present the first large-scale survey of Balinese paternal diversity using a battery of 71 binary polymorphisms (mainly single nucleotide polymorphisms, or SNPs) and 10 microsatellites on a sample of 551 Balinese Y chromosomes. Through comparisons of variation at the same markers in a large Asian sample we assess the relative contributions of Austronesian farmers, pre-Neolithic hunter-gatherers, and Indian traders to the contemporary Balinese paternal gene pool. The results of these analyses reveal a complex history of genetic contributions from different source populations.

## **Subjects and Methods**

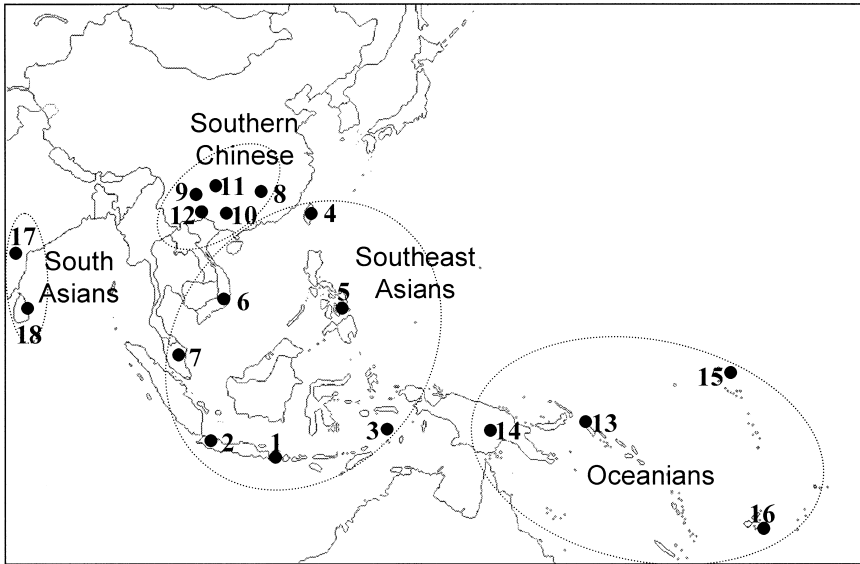
**Samples.** We analyzed 71 SNPs and 10 short tandem repeats (STRs) in a sample of 1,989 Y chromosomes from 20 populations. Our Indonesian samples included 551 Balinese Y chromosomes and 76 Y chromosomes from western ( $n = 21$ ) and eastern ( $n = 55$ ) Indonesia. Also included in this survey were 1,187 Y chromosomes from 13 populations representing other Southeast Asians (Vietnamese, Malaysians, Philipinos, and Taiwanese aboriginals), Southern Chinese (Han, Miao, She, Tujians, and Yao), South Asians (southern Indians and Sri Lankans), and Near Easterners (Syrians and Saudi Arabians) and 175 Y chromosomes from Melanesia, Micronesia, Polynesia, and Papua New Guinea (collectively referred to as Oceania) (Table 1; Figure 1). Many of the non-Balinese

**Table 1.** Frequencies of Major Y-Chromosome Lineages in Bali and 19 Additional Population Samples

Population	N	C <sup>a</sup> (RPS4Y) (YAP)	DE <sup>a</sup> (YAP)	F (P14)	G <sup>a</sup> (M201)	H <sup>a</sup> (M69)	I (P19)	J <sup>a</sup> (p12f)	K <sup>a</sup> (M9)	L (M20)	M <sup>a</sup> (M5)	N <sup>a</sup> (LLY22g)	O (M122*)	O (MI34)	O (LINE)	O <sup>a</sup> (M119)	O <sup>a</sup> (P31)* <sup>b</sup>	O (M95)	P (P27*)	Q (P36)	R <sup>a</sup> (M207)	R <sup>a</sup> (M173)
Southeast Asians	825	<b>35</b>	<b>3</b>	<b>1</b>	<b>2</b>	<b>19</b>	<b>19</b>	<b>14</b>	<b>52</b>	<b>13</b>	<b>12</b>	<b>2</b>	<b>68</b>	<b>19</b>	<b>24</b>	<b>165</b>	<b>5</b>	<b>359</b>	<b>7</b>	<b>9</b>	<b>16</b>	<b>16</b>
Balinese	551	10				19		12	8	13	4		33	1	4	100	1	323	2	8	13	
East Indonesians	55		18						18		7		5									
West Indonesians	21	1		1					1				3	2	4	4		3			1	1
Taiwanese																						
aboriginals	48	1											3			43		1				
Philippines	48	1							23				12	2	3	5		2				
Vietnamese	70	3	2					2				2	7	11	10	4	4	19		5		1
Malaysians	32	1	1		2				2	1			5	3	3	2	11					1
Southern Chinese	<b>384</b>	<b>37</b>	<b>8</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>2</b>	<b>2</b>	<b>1</b>	<b>17</b>	<b>32</b>	<b>70</b>	<b>120</b>	<b>26</b>	<b>22</b>	<b>45</b>	<b>1</b>	<b>1</b>	<b>2</b>	<b>2</b>
Han	166	10	1		1			1	2		2	15	14	47	31	16	8	19		1		
Miao	58	3	5										12	6	22	4	6					
She	51													3	29	1	18					
Tujians	49	12	1									2	6	11	9	4	2					2
Yao	60	12	1											3	29	1	12	2				
Oceanians	<b>175</b>	<b>55</b>					<b>1</b>	<b>1</b>	<b>50</b>		<b>35</b>		<b>19</b>	<b>3</b>	<b>1</b>	<b>7</b>	<b>1</b>		<b>1</b>			<b>1</b>
Melanesians	53	10						1	17		19		3		2							1
Papua New																						
Guineans	46	5							25	13			2		1	1						
Micronesians	16	2							6	1			3		2	1			1			
Polynesians	60	38					1		2	2			11	3	1	2						
South Asians	<b>496</b>	<b>13</b>	<b>49</b>	<b>11</b>	<b>133</b>			<b>53</b>		<b>63</b>				<b>1</b>	<b>1</b>	<b>1</b>	<b>39</b>	<b>3</b>	<b>4</b>	<b>46</b>	<b>80</b>	<b>80</b>
Indians	405	10	40	6	110			35		49				1	1		38	3	1	43	68	68
Sri Lankans	91	3	9	5	23			18		14								1				
Near Easterners	<b>109</b>		<b>11</b>	<b>1</b>	<b>4</b>	<b>1</b>	<b>1</b>	<b>57</b>	<b>5</b>	<b>7</b>									<b>2</b>	<b>3</b>	<b>21</b>	<b>21</b>
Saudi Arabians	22		2		1			8		5										1		5
Syrians	87		9	1	3			49	5	2										1		16

a. Downstream markers typed but not shown. See "Subjects and Methods" section and Figure 1 for a complete list of markers typed.

b. O-P31\* and O-SRY<sub>465</sub>.



**Figure 1.** Approximate geographic positions of 18 populations sampled in this study (see Table 1 and text for names and sizes of population samples). The populations are grouped into four major geographic areas (dotted circles). Near Eastern populations (19 and 20) are not shown.

samples analyzed here have been described in previous studies (Bamshad et al. 2001; Hammer et al. 2001; Karafet et al. 2001; Redd, Roberts-Thomson et al. 2002). The Chinese Han sample was composed of individuals from the Guangdong and Shaanxi provinces and of individuals from different parts of Taiwan (Karafet et al. 2001).

Buccal swabs were collected in 2001–2002 from 551 Balinese volunteers, who gave informed consent. All sampling protocols were procedures approved by the University of Arizona Human Subjects Committee and Balai Pengkajian Teknologi Pertanian (Bali). Buccal cell DNA was isolated according to the method of B. Richards et al. (1993).

**Genetic Markers.** The polymorphic sites in our survey included a set of 63 previously published binary Y-chromosome markers (Karafet et al. 2002) and the following eight polymorphisms: M69, M70, M110, M111, M214, and M230 (Underhill et al. 2000; Kayser et al. 2003), Apt (Pandya et al. 1998), and P34. This is the first report describing PCR conditions for P34, a polymorphism recently discovered in a panel of 92 globally distributed Y chromosomes (Hammer et al. 2003). The Apt and M230 markers were genotyped as reported by Pandya et al. (1998) and Kayser et al. (2003), respectively. We designed allele-specific

PCR reactions to type M69, M70, M110, M111, and M214 (conditions available on request from T. Karafet).

A G → A mutation (referred to as P34) at position 956 of the DYS190 locus was genotyped by allele-specific PCR. The following primers were used to amplify the 399-base pair (bp) control band and the 170-bp band that was allele specific for mutant chromosomes: P34U (5'-CCTGGAAAAGTCAAATCATCG-3'), P34L (5'-CGTGGCATCTTGTCATGTCT-3'), and P34A (5'-CTGTGTCTT-TGTCTGTGTGTA-3'). The cycling conditions were 94°C for 3 min, followed by 35 cycles at 94°C, 64°C, 72°C for 30 s, with a final extension step at 72°C for 2 min. Reactions were run in a final volume of 15 µL containing 10 ng of genomic DNA, 0.2 mM each dNTP, 1 mM each primer, 0.046 mM of TaqStart Antibody (Clontech), 0.0016 mM of Taq DNA polymerase (Eppendorf), 1.5 mM MgCl<sub>2</sub>, 75 mM KCl, and 10 mM Tris-HCl (pH 8.3). The P36 marker was typed as previously reported by Karafet et al. (2002), with the exception that substitution was incorrectly described as G → A instead of G → T.

For the microsatellite analysis 10 STRs (DYS19, DYS388, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS426, and DYS439) were typed in two multiplex PCR reactions. Primer sequences were published by Kayser et al. (1997) and Redd, Roberts-Thomson et al. (2002), and PCR conditions were given by Redd, Agellon et al. (2002). PCR products were electrophoresed on a 3100 Genetic Analyzer (Applied Biosystems) using a 36-cm array and filter set D. The data were analyzed with Genescan, version 3.7 (Applied Biosystems), and Genotyper, version 1.1 (Applied Biosystems). For all statistical analyses DYS389I was subtracted from DYS389II because the DYS389II PCR product also contains DYS389I. SNP and STR frequency data are available on request from T. Karafet.

**Terminology.** We follow the conventions recommended by the Y Chromosome Consortium (2002) for naming Y-chromosome lineages. Capital letters A–R identify the 18 major Y-chromosome clades or haplogroups. We use the shorthand mutation-based naming system, which retains the major haplogroup information (i.e., 19 capital letters) followed by the name of the terminal mutation that defines a given haplogroup. Lineages not defined on the basis of a derived character state represent interior nodes of the tree (paragroups) and are distinguished from terminal haplogroups defined by the derived state at a particular marker by an asterisk. The term haplogroup rather than paralog (Y Chromosome Consortium 2002) is often used to refer to these internal nodes. When no farther downstream markers in the latest version of the Y-chromosome tree (Jobling and Tyler-Smith 2003) were typed for this study, we considered the most derived typed marker to represent a haplogroup. As suggested by de Knijff (2000), distinct Y chromosomes identified by STRs are designated as haplotypes.

**Statistical Analyses.** Y-chromosome STR diversity, variance in repeat number, and three population pairwise genetic distances—absolute size difference

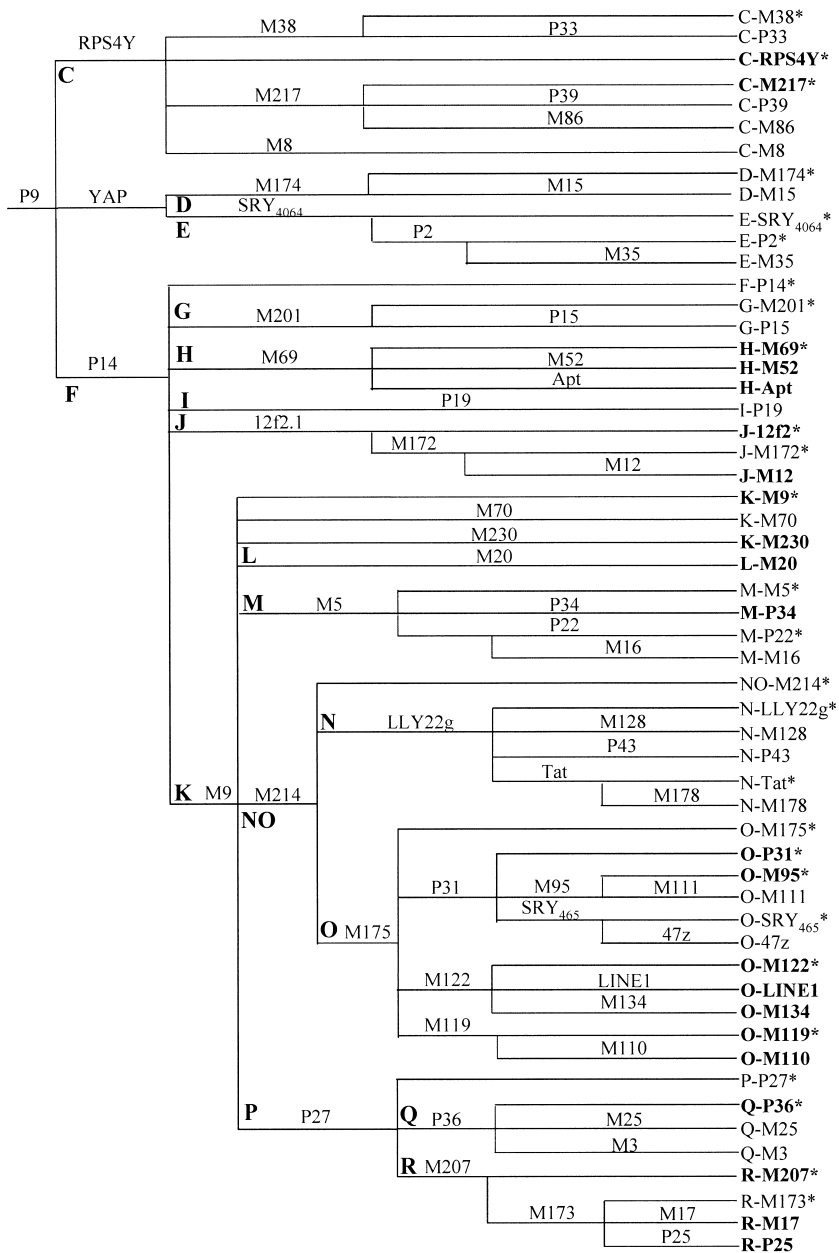
( $D_{ad}$ ), Goldstein's (1995)  $(\delta\mu)^2$  ( $D_{dm}$ ), and distance based on shared alleles ( $D_{ps}$ )—were calculated using the software package Microsat 1.5d (Minch et al. 1997). Haplotypic distance matrices were used to reconstruct neighbor-joining trees using the software package PHYLIP (Felsenstein 1995). Median-joining networks (Bandelt et al. 1999) were constructed using the Network 2.0c program. For network calculations STRs were weighted according to their repeat number variances such that higher weights were assigned to the least variable loci. The reduced median output was used as input for the median-joining network. This procedure reduces the chances of obtaining large reticulations within the network (Hurles et al. 2002). Two different coefficients of admixture,  $m_C$  (Chakraborty et al. 1992) and  $m_R$ , a least-squares estimator (Roberts and Hiorns 1965), were estimated by means of Admixture 1.0 (Bertorelle and Excoffier 1998).

In the Bali population we estimated the time to the most recent common ancestors of particular lineages by means of the YMRCA program (Stumpf and Goldstein 2001). For all STRs we used the mutation rate estimate  $2.8 \times 10^{-3}$  proposed by Kayser, Roewer et al. (2000) and its 95% confidence interval ( $1.72 \times 10^{-3}$  to  $4.27 \times 10^{-3}$ ) with a generation time of 25 years.

To avoid complications arising from the analysis of paragroups, we combined all descendant lineages of a given marker when comparing STR diversity levels among populations. Occasionally, when the geographic distribution of descendant haplogroups differed markedly from that of the ancestral paralog, we also calculated STR diversity on those lineages associated with the paralog only. This facilitated comparisons of ancestral variation among populations with the assumption that there were no missing SNPs marking the included lineages. In the case of median-joining networks, there is often a problem in visualizing the results when many STRs and samples are included in the analysis. Therefore in some analyses we present the results of networks for the paralog only.

## Results and Discussion

Table 1 reports the frequencies of 23 of the 44 lineages found in our survey that were defined on the basis of 71 binary markers typed in 1,989 Y chromosomes from 20 populations, including 551 males from different parts of the island of Bali. Figure 2 shows the evolutionary relationships among 55 haplogroups resulting from typing 57 of the 71 markers used in this survey. Three lineages within haplogroup O (O-M95, O-M122, and O-M119) account for 83.7% of the Y chromosomes in our Balinese sample. Surprisingly, most of the remaining Y chromosomes (accounting for 11.8% of the total) are members of clades H, J, L, and R. These haplogroups are absent or present at low frequencies in southern China, Southeast Asia, and Oceania (Table 1). In our survey, these haplogroups are limited almost entirely to South Asia and/or the Near East. In the following sections we present the results of phylogeographic analyses of these Balinese Y-chromosome haplogroups and discuss implications of these results for hypotheses on the peopling of Bali and the Indonesian region.



**Figure 2.** Evolutionary tree for the 55 Y-chromosome lineages defined by 57 of the 71 markers used in this survey. The names of the markers are shown along the branches of the tree, and lineage names are shown on the right-hand side. Lineage names with an asterisk refer to internal nodes of the tree, or paragroups (see “Subjects and Methods” section). Haplogroups found in Bali are shown in boldface.



**Austronesian Roots of the Balinese Population.** Two Y-chromosome haplogroups, O-M122 and O-M119, are widely distributed in Asia and are likely of Asian origin. Several researchers have suggested that the distribution of these haplogroups provides evidence for an Austronesian expansion from eastern Asia into island Southeast Asia, Melanesia, and Polynesia (Su, Jin et al. 2000; Kayser, Brauer et al. 2000; Kayser et al. 2001, 2003; Capelli et al. 2001; Hurles et al. 2002). Although there is general agreement that haplogroups O-M122 and O-M119 trace the spread of agricultural technology, different source populations for these lineages have been proposed by different investigators. Su and colleagues (Su et al. 1999; Su, Jin et al. 2000) postulated that southern Chinese populations were the source of two independent migrations, one toward Taiwan and the other toward Polynesia through insular Southeast Asia. In contrast, Capelli et al. (2001) and Hurles et al. (2002) found little support for an Austronesian expansion originating in southeast China and Taiwan. They suggested that most of present-day Austronesian speakers trace their paternal origins to Pleistocene hunter-gatherers and that the dispersal of the Austronesian languages was mainly a cultural process. Kayser and colleagues (Kayser, Brauer et al. 2000; Kayser et al. 2001, 2003) support an intermediate scenario—that Polynesian ancestors did indeed originate in mainland Asia and Taiwan. However, they argue that the colonization of the Pacific was preceded by extensive admixture with indigenous Melanesians.

In our sample, 6.9% of Balinese men have Y chromosomes that are members of haplogroup O-M122, which contains the subclades O-M134 and O-LINE (Figure 2). The data presented here and elsewhere indicate that O-M122 is a major Southeast Asian and Oceanian lineage that is absent or nearly absent in the Near East and South Asia (Table 1) as well as in Central and North Asia (Karafet et al. 2001). Here, we note that the distribution of O-M122\* chromosomes (i.e., paragroup O-M122\*) is also more widespread than chromosomes with the derived state at the M134 and LINE-1 mutations (i.e., haplogroups O-M134 and O-LINE-1). Haplogroups O-LINE1 and O-M134 are prevalent in China (Su et al. 1999; Su, Jin et al. 2000; Su, Xiao et al. 2000; Santos et al. 2000; Karafet et al. 2001), making up almost 50% of all sampled chromosomes. These two haplogroups probably originated in China and subsequently migrated into surrounding regions (Su, Xiao et al. 2000; Santos et al. 2000). Paragroup O-M122\* is found in East and Southeast Asia and in Oceania (Kayser, Brauer et al. 2000; Kayser et al. 2001, 2003; Santos et al. 2000; Capelli et al. 2001). The highest average STR heterozygosity on the O-M122 haplogroup (and O-M122\* paragroup) background is found in Southeast Asia, followed by southern China, with lower diversity in Oceania (Table 2).

In our Balinese sample, 38 O-M122 individuals have 17 different STR haplotypes. Within O-M122\*, Y-chromosome STR sharing is observed mainly within populations; however, the Balinese have one haplotype that is widely distributed among Austronesian speakers from other parts of Indonesia, the Philippines, Taiwan, and parts of Oceania. There are too many O-M122 haplotypes to

**Table 2.** STR Diversity Associated with Y-Chromosome Haplogroup/Paragroup Lineages

Population	O-M122			O-M122*			O-M119			O-M119*			O-M95			O-M95*			
	N	Het.	Var.	N	Het.	Var.	N	Het.	Var.	N	Het.	Var.	N	Het.	Var.	N	Het.	Var.	
Southeast Asians <sup>a</sup>	<b>73</b>	<b>0.510</b>	<b>0.555</b>	<b>35</b>	<b>0.445</b>	<b>0.445</b>	<b>65</b>	<b>0.327</b>	<b>0.301</b>	<b>0.318</b>	<b>49</b>	<b>0.330</b>	<b>0.318</b>	<b>36</b>	<b>0.392</b>	<b>0.383</b>	<b>19</b>	<b>0.357</b>	<b>0.337</b>
Balinese	36	0.392	0.492	33	0.336	0.470	100	0.376	0.508	98	0.376	0.508	322	0.316	0.219	322	0.316	0.219	0.219
East Indonesians	5			5			7			5			0			0			
West Indonesians	9			3			4			3			3			3			
Taiwanese aboriginals	3			3			43			34			0.258			1			
Philippines	17	0.421	0.467	12	0.335	0.340	5	0.323	0.273	34	0.301	0.258	2	0.316	0.272	5	0.366	0.336	
Vietnamese	28	0.472	0.565	7			4			4			2	0.402	0.400	10	0.367	0.353	
Malaysians	11	0.518	0.583	5			2			2			11	0.324	0.340	10	0.276	0.331	
East Asians	<b>225</b>	<b>0.420</b>	<b>0.499</b>	<b>31</b>	<b>0.374</b>	<b>0.490</b>	<b>26</b>	<b>0.342</b>	<b>0.305</b>	<b>26</b>	<b>0.342</b>	<b>0.305</b>	<b>45</b>	<b>0.367</b>	<b>0.353</b>	<b>13</b>	<b>0.317</b>	<b>0.408</b>	
Chinese Han	95	0.474	0.586	14	0.554	0.370	16	0.350	0.299	16	0.350	0.299	19	0.324	0.340	10	0.276	0.331	
Chinese Miao	40	0.386	0.498	12	0.158	0.270	4			4			6			2			
Chinese She	32	0.166	0.156	0			1			1			18	0.160	0.085	0			
Chinese Tujians	26	0.411	0.432	5			4			4			0			0			
Chinese Yao	32	0.213	0.290	0			1			1			2			1			
Oceania	<b>23</b>	<b>0.243</b>	<b>0.981</b>	<b>19</b>	<b>0.231</b>	<b>0.854</b>	<b>7</b>			<b>6</b>			<b>0</b>			<b>0</b>			
Melanesians	3			3			2			1			0			0			
Papua New Guineans	2			2			1			1			0			0			
Micronesians	3			3			2			2			0			0			
Polynesians	15	0.336	0.907	11	0.190	0.700	2			2			0			0			
South Asians	<b>2</b>			<b>0</b>			<b>0</b>			<b>0</b>			<b>39</b>	<b>0.332</b>	<b>0.229</b>	<b>39</b>	<b>0.332</b>	<b>0.229</b>	
Indians	2			0			0			0			38			38			
Sri Lankans	0			0			0			0			1			1			
Near Easterners	<b>0</b>			<b>0</b>			<b>0</b>			<b>0</b>			<b>0</b>			<b>0</b>			
Saudi Arabians	0			0			0			0			0			0			
Syrians	0			0			0			0			0			0			

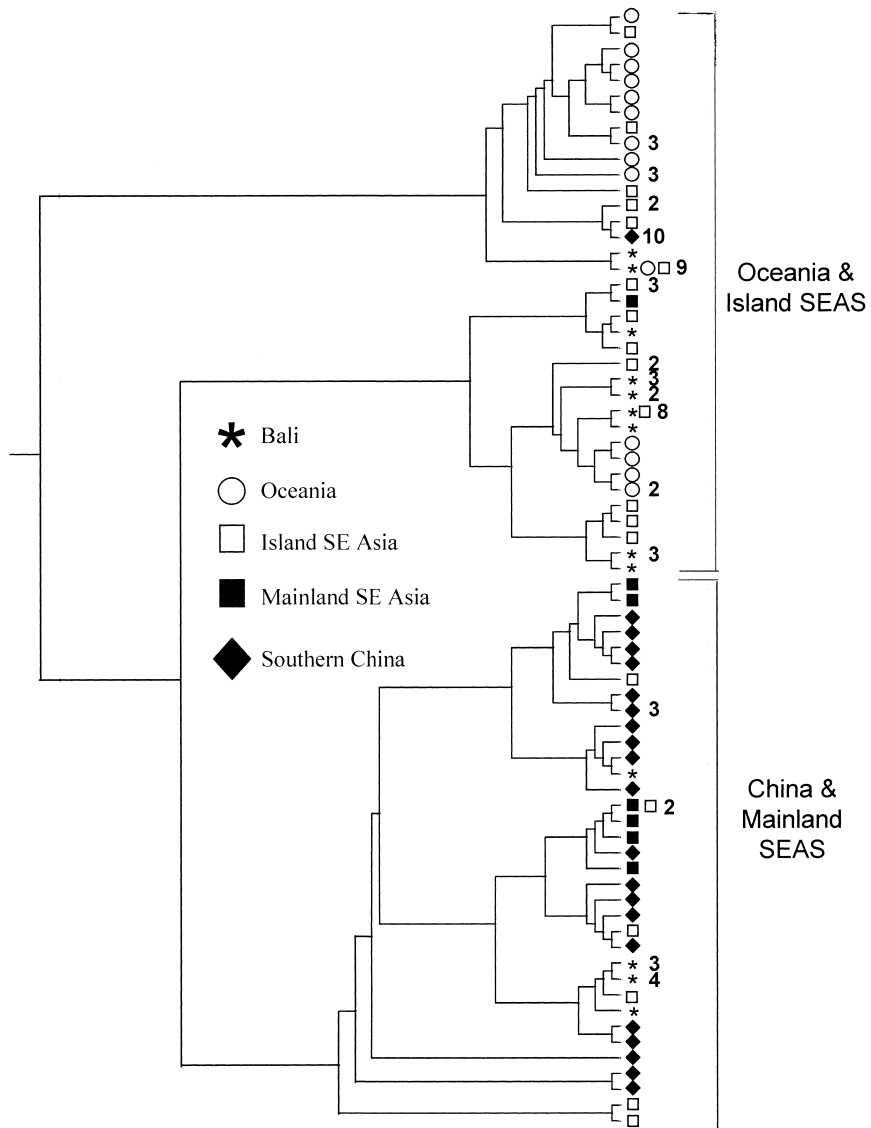
a. Not including Balinese samples.

construct a readable reduced median-joining network, so phylogenetic relationships among O-M122\* STR haplotypes are reconstructed in a neighbor-joining tree based on absolute allele size distances (Figure 3). Three distinct clusters were observed: Two clusters contain “Austronesian” Y chromosomes mainly from island Southeast Asia (e.g., Malaysia, the Philippines, and aboriginal Taiwan) and Oceania, whereas the third cluster is distinguished by a preponderance of haplotypes from mainland China and Vietnam. Most Balinese O-M122 Y chromosomes (23 out of 32) and haplotypes (9 out of 13) are associated with the first and second clusters. All three genetic distance methods reveal the closest relationship between O-M122 haplotypes from Bali and the Philippines (data not shown). The most plausible explanation for the combined data is that O-M122 lineages migrated to Indonesia and Bali from Southeast Asia through the Philippines. Alternatively, our data could be viewed as a spread of Y chromosomes out of Southeast Asia to Oceania and the Philippines by way of Indonesia.

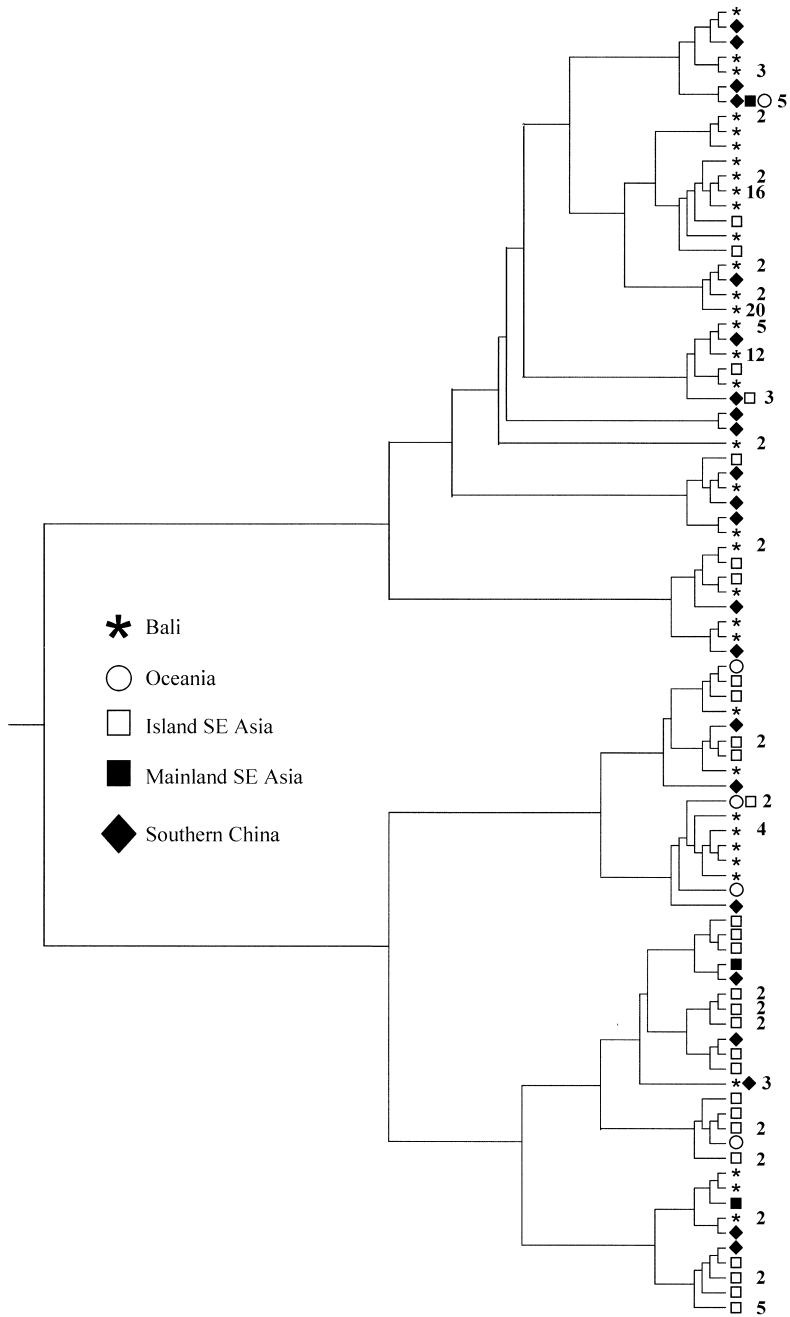
Haplogroup O-M119, which has a similar geographic distribution as O-M122 (Table 1), is another haplogroup that may have been brought to Bali by migrating Austronesian speakers. This haplogroup is observed in 100 Balinese individuals (18.1%) with 38 different STR haplotypes. The M119 marker has drawn particular interest because chromosomes with the derived state are most frequent among Taiwanese aboriginal groups (Su et al. 1999; Su, Jin et al. 2000; Kayser, Brauer et al. 2000; Kayser et al. 2001, 2003; Santos et al. 2000; Capelli et al. 2001; Karafet et al. 2001). Caution is advised in interpreting this to indicate a Taiwanese origin for this mutation because genetic drift resulting from extreme isolation and/or a founder effect may have increased the frequency of this haplogroup in aboriginal Taiwanese. Evidence in support of this hypothesis comes from the slightly reduced STR diversity associated with O-M119 chromosomes in aboriginal Taiwanese compared with other populations (Table 2), despite their high frequency (70.8%) in this population (Table 1). For example, Bali exhibits the second highest frequency of O-M119 chromosomes (17.8%) in our survey; however, the average variance in STR allele size on Balinese O-M119 chromosomes is almost twice as high as that on aboriginal Taiwanese O-M119 chromosomes (Table 2).

In contrast to the O-M122\* haplogroup, a neighbor-joining tree for O-M119\* STR haplotypes (Figure 4) shows little if any geographic structure. O-M119\* chromosomes from Bali, China, and Indonesia are distributed throughout the tree, although Taiwanese aboriginal chromosomes exhibit some clustering. The absence of geographic structure may suggest a more ancient dispersal, several migrations from different source populations, and/or continued gene flow from Southeast Asia after an initial expansion of O-M119\* lineages into this region. O-M119\* chromosomes might also represent a heterogeneous group of not-yet-identified haplogroups. Additional markers within this lineage may allow further resolution of the O-M119\* network.

Interestingly, the O-M110 haplogroup (a descendant of O-M119; see Figure 2) is not found in China. In our survey its geographic distribution is restricted



**Figure 3.** Phylogenetic relationships among O-M122\* STR haplotypes. The neighbor-joining tree was constructed using the PHYLIP program on the basis of absolute allele size distances calculated with the computer program Microsat. The geographic origins of each individual with respect to the four major groupings in Figure 1 are indicated. SEAS, Southeast Asia.



**Figure 4.** Phylogenetic relationships among O-M119\* STR haplotypes. The geographic origins of each individual with respect to the four major groupings in Figure 1 are indicated.

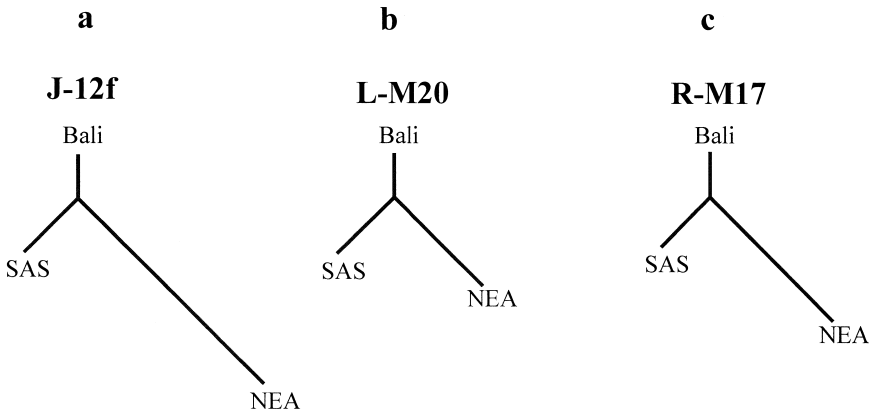
to Austronesian speakers of Taiwan, the Philippines, Indonesia, Bali, and Melanesia. Su and colleagues (Su et al. 1999; Su, Jin et al. 2000) also found this haplogroup in Southeast Asia and Micronesia. An STR-based age for chromosomes carrying the M110 mutation is approximately 3,420 years (95% CI = 2,245–5,570). This pattern raises the possibility that O-M110 chromosomes could have been brought to Bali along with O-M122\* chromosomes by way of the Philippines.

**Gene Flow from the Indian Subcontinent.** Eight different haplogroups within the H, J, L, and R clades are observed in Bali (see Figure 2). Altogether these four major lineages account for 12% of Balinese Y chromosomes. Interestingly, the H-M69\*, H-M52, H-Apt, J-p12f\*, J-M12, L-M20, R-M207\*, and R-M17 lineages are virtually absent in southern China, Southeast Asia, and Oceania and are almost entirely restricted to the Indian subcontinent and the Near East (Table 1). Two of these lineages, H-Apt and H-M69\*, have been previously reported to be limited to Indian populations (Pandya et al. 1998; Kivisild et al. 2003), whereas H-M52 and R-M207\* have also been found in Central Asia, albeit at low frequencies (Semino et al. 2000; Karafet et al. 2001; Wells et al. 2001).

Although the combined results of these surveys strongly support the hypothesis of an Indian origin for these four lineages, it is more difficult to infer the source(s) for the Balinese J-p12f\*, J-M12, L-M20, and R-M17 lineages. In our survey these chromosomes are found in both Indian and Near Eastern populations at relatively high frequencies. The most common L-M20 microsatellite haplotype in Bali is also the most frequent haplotype among Indian L-M20 chromosomes, suggesting that these Balinese chromosomes may be of recent Indian origin. Interestingly, the Balinese do not share identical J-p12f\*, and R-M17 STR haplotypes with either Indians or Near Easterners.

To assess whether Balinese J-p12f\* and R-M17 chromosomes are more likely to have originated in India or in the Near East, we compared modal haplotypes for our samples from these regions. The Balinese modal J-p12f\* haplotype exhibits a difference of two steps from the Indian modal haplotype and seven differences from the Near Eastern modal haplotype. With respect to R-M17, the Balinese modal haplotype differs by one step from the Indian modal haplotype, whereas two steps differentiate the Balinese and Near Eastern modal haplotypes.

We also calculated genetic distances and constructed phylogenetic trees based on our Y-chromosome STR data. All three trees give broadly consistent results: smaller genetic distances and closer genetic relationships among Balinese and Indians than among Balinese and populations from the Near East (Figure 5). It is also possible that some Balinese R-M17 chromosomes are of recent European origin, because this haplogroup is found at low frequencies in Western Europe (Hammer et al. 2001).



**Figure 5.** Neighbor-joining networks depicting relationships among haplogroups (a) J, (b) L, and (c) R-M17 in Balinese, South Asian (SAS), and Near Eastern (NEA) populations based on absolute allele size distances estimated from 10 STR loci.

**What Is the Origin of the Most Common Balinese Y-Chromosome Haplogroup?**

Most Balinese males (58.6%) carry haplogroup O-M95. This haplogroup shows a geographic distribution that is different from haplogroups O-M122 and O-M119, and so its presence in Bali may not be explained solely on the basis of Austronesian expansions. O-M95 is found at higher frequencies in Southeast Asia (13.1%) than in southern China (11.7%) and so far has been found in only two individuals from Oceania (see Table 1) (Su, Jin et al. 2000; Capelli et al. 2001; Hurles et al. 2002; Kayser et al. 2003). The presence of O-M95 was reported in Indian caste and tribal populations at low frequencies (Wells et al. 2001); however, Ramana et al. (2001) recently found this haplogroup at high frequencies in three tribal populations from South India (17–48%). In the present survey we also find a relatively high frequency (22.4%) of this haplogroup among tribal populations from southern India and an average frequency of 7.9% in our combined South Asian sample. There is a narrow range of average STR heterozygosity values associated with the O-M95 haplogroup, with the highest value ( $0.392 \pm 0.024$ ) in Southeast Asia and slightly lower values in southern China and South Asia ( $0.367 \pm 0.015$  and  $0.332 \pm 0.021$ , respectively) (see Table 2). Thus our data provide weak support for Kayser et al.'s (2003) hypothesis that Southeast Asia was the original homeland of the O-M95 haplogroup, because both the highest frequency of O-M95 and the greatest STR diversity are found in Southeast Asia.

Given that the O-M95 haplogroup is present in both Southeast Asia and the Indian subcontinent and considering that our data provide evidence of gene flow from India to Bali, we can envisage three scenarios for how this haplogroup

migrated to Bali. The most plausible explanation is that O-M95 was brought to Bali by Austronesian speakers from Southeast Asia or the Philippines along with O-M122 and/or O-M119. This scenario fits well with the estimated age of the M95 mutation of 8,800 years (Kayser et al. 2003), the geographic distribution of O-M95, and the inferred route of the Austronesian expansion. However, the virtual absence of O-M95 chromosomes in Oceania may reflect a later entry of haplogroup O-M95 to Indonesia and Bali from Southeast Asia (i.e., after the initial colonization of the Pacific by Austronesian farmers). Alternatively, Indians might have carried O-M95 chromosomes together with H, J, L, and R lineages. This model seems less likely because of the high frequency of O-M95 on Bali, the relatively high frequency of O-M95 chromosomes throughout Southeast Asia, and their patchy distribution in India (i.e., concentrated in few tribal populations). Finally, O-M95 chromosomes may have arrived in Bali through separate migrations from Southeast Asia and India.

We addressed these hypotheses by examining patterns of STR diversity associated with our sample of O-M95 chromosomes. A reduced median-joining network depicts geographic clustering of the southern Chinese haplotypes, with an absence of similar structure for O-M95 haplotypes from Bali, India, Malaysia, Vietnam, and other parts of Indonesia (data not shown). To explore the genetic relationships among O-M95 chromosomes from Bali, Southeast Asia, and India, we calculated genetic distances based on STR data and analyzed modal haplotypes among different populations. Modal haplotypes for the Balinese and for the combined Southeast Asian sample are identical, whereas there are two mutational step differences from the Indian modal haplotype. Genetic distance estimations reveal that the divergence between Balinese and Indian O-M95 haplotypes is at least twice as great as the divergence between Balinese and Southeast Asian chromosomes. Surprisingly, the Balinese share 10 O-M95 haplotypes with Indians, but only 4 haplotypes are shared between the Balinese and Southeast Asians. These conflicting results may indicate that most O-M95 chromosomes sampled from Bali derive from a Southeast Asian ancestor and that a few others descend from recent migrants arriving in Bali from the Indian subcontinent (i.e., a two-wave model).

#### **Traces of an Aboriginal Population of Balinese Hunter-Gatherers?**

Haplogroup M was previously recognized as a predominant haplogroup in Melanesia and New Guinea (Hammer et al. 2001; Kayser et al. 2001, 2003), with some presence in eastern Indonesia, Polynesia, and Malaysia (Su, Jin et al. 2000; Capelli et al. 2001; Karafet et al. 2001; Hurler et al. 2002). In the present survey, haplogroup M chromosomes are also mainly restricted to eastern Indonesia and Oceania, with a low frequency in Southeast Asia and Bali (Table 1). An STR-based estimate of the coalescence time of M-M4 chromosomes is approximately 8,200 years (Kayser et al. 2003), whereas an approach based on SNPs yields an age of the M4 mutation of approximately  $12,700 \pm 7,200$  years (Hammer and Zegura 2002).



Based on its higher haplotype diversity in Melanesia, Kayser et al. (2003) proposed a Melanesian rather than an eastern Indonesian origin of the M4 mutation before the Austronesian expansion. Haplogroup M-P34, defined by a newly discovered polymorphism reported here, is also mainly an Oceanian haplogroup (Table 1). It turns out that all the M chromosomes in Bali (0.7%) and eastern Indonesia (12.7%) are haplogroup M-P34. Papua New Guineans exhibit the highest STR diversity and contain haplotypes that link M-P34 and M-M4 SNP haplogroups in median-joining networks (data not shown). Thus our data support the hypothesis of a Papua New Guinean origin of the P34 mutation. An STR-based estimate of the coalescence time for chromosomes carrying the P34 mutation is approximately 5,350 years (95% CI = 3,500–8,720). The presence of haplogroup M-P34 in Bali and Indonesia may reflect a contribution from early Pleistocene settlers before the Austronesian expansion, and/or relatively recent gene flow from Papua New Guinea.

Three other lineages that may represent pre-Neolithic Balinese Y chromosomes are haplogroups C, K-M230, and K-M9\*. Haplogroup K-M230 is the major Y-chromosome lineage in Melanesia and New Guinea, and it also occurs in other parts of Indonesia (Kayser et al. 2003) (see Table 1). Kayser et al. (2003) proposed a Melanesian rather than an eastern Indonesian origin of this haplogroup and suggested that the K-M230 mutation arose before the Austronesian expansion about 8,200 years ago.

Chromosomes carrying the M9-G mutation are widely distributed across Asia; however, chromosomes carrying the ancestral state at seven of the eight downstream SNPs that mark immediate descendant lineages from M9 (K-M9\*) on the most recent Y-chromosome tree (Jobling and Tyler-Smith 2003) are found at relatively high frequencies only in the Philippines, Indonesia, Melanesia, Papua New Guinea, and Micronesia. Capelli et al. (2001) hypothesized that all Y chromosomes carrying the M9-G marker initially expanded out of Melanesia. If this is the case, the small proportion of K-M9\* chromosomes (1.1%) in Bali may be a signature of pre-Neolithic settlements. In contrast to M-P34 and K-M230, the distribution of K-M9\* chromosomes is not restricted to Oceania. They are also present in Malaysia and the Near East, albeit at low frequencies.

To test the possibility that Balinese K-M9\* chromosomes were carried by people from the Near East and/or Southeast Asia, we constructed a Y-chromosome STR network for this paralog (data not shown). Three haplotypes observed in Bali form a tight cluster that is closely affiliated with chromosomes from Indonesia, Melanesia, Micronesia, and Papua New Guinea. The lowest genetic distances are found between the Balinese and Papua New Guineans. Similar to haplogroups M-P34 and K-M230, K-M9\* chromosomes in Bali and Indonesia most likely represent a pre-Neolithic contribution before the Austronesian expansion.

Contrary to the situation for the K-M9\*, K-M230, and M-P34 haplogroups, three STR haplotypes associated with C\* chromosomes in Bali form a tight cluster that is connected to neighboring haplotypes from western Indonesia and East

Asia. Compared with other haplogroups analyzed, C\* has a low diversity in Bali. The occurrence of C\* chromosomes in Bali most probably reflects recent gene flow to Bali.

The observed frequencies of K-M9\*, K-M230, and M-P34 on Bali suggest that 2.2% of the pre-Neolithic gene pool survived the invasion of Austronesian Y chromosomes. Interestingly, a similar extent of replacement is evident in Java (3.8%), and a slightly higher proportion of pre-Neolithic Y chromosomes is observed in Borneo (15.0%) (Kayser et al. 2003). A similar analysis of our sample of eastern Indonesians indicates major pre-Neolithic (78.2%) and minor Austronesian (21.8%) components of their paternal gene pool (Table 1).

## Summary

This study was intended to compare the composition of the Balinese paternal gene pool with those of its Asian neighbors to infer the relative contributions of pre-Neolithic hunter-gatherer and Austronesian farmer Y chromosomes and to test the hypothesis of recent gene flow from India. At least two paternal lineages (O-M122 and O-M119) seem to trace the Austronesian expansion to Bali. Our estimate of the coalescence times of these haplogroups on Bali—4,400 years (95% CI = 2,900–7,200) and 5,630 years (95% CI = 3,700–9,170), respectively—are older than archeological dates. This is not surprising, because both haplogroups arose outside Bali, and early migrants to Bali most likely carried diverse Y chromosomes within these haplogroups. Time estimates of these haplogroups have overlapping confidence intervals, suggesting that both O-M122 and O-M119 chromosomes may represent a Y-chromosome signal of the earliest Austronesian expansion. However, our data based on median-joining networks, neighbor-joining trees, and genetic distances suggest that several expansion waves of Austronesian groups may have reached Bali by different routes. Although haplogroup O-M95 was probably brought by one of the Austronesian migration waves, the data suggest that a subset of these chromosomes may have entered Bali more recently from the Indian subcontinent. The influence of the Austronesian expansion on the preexisting hunter-gatherer gene pool appears to be substantial. The spread of these three haplogroups to Bali has nearly erased the signature of the ancestral Y-chromosome pool.

We also assessed the possible influence of three more recent historical events on the composition of the Balinese gene pool: a period of Indianization characterized by the spread of the Hindu religion and culture to Bali, the expansion of Islam to Indonesia, and the persistence of the Dutch empire on Bali for almost 100 years. Although we do not find evidence for a significant West Asian or Dutch contribution to the Balinese paternal gene pool, solid evidence for an Indian contribution comes from the presence of multiple Indian-specific haplogroups on Bali. An estimate of the Balinese frequency of these Indian haplogroups (about 12%) provides a direct measure of the Indian contribution, whereas

admixture methods suggest that the total contribution of Indian Y chromosomes to the Balinese gene pool might be as high as 19% (data not shown). Although we cannot define how long Indian haplogroups have persisted on Bali, it is worth noting that three of these lineages (J-12f\*, R-M17, and R-M207\*) reveal relatively high average STR diversity (i.e.,  $h = 0.308, 0.279, \text{ and } 0.281$ , respectively), comparable with that seen in the major Balinese haplogroup, O-M95.

Coalescence time estimates of the Indian haplogroups on Bali vary from 2,600 to 3,100 years. As expected, these dates are older than the archeological evidence for the earliest Balinese Indicized kingdoms (late first millennium A.D.); however, they are consistent with archeological evidence for the earliest trade contacts with India (Ardika and Bellwood 1991). Supporting evidence for 2,000-year-old contact with India comes from an analysis of ancient mtDNA (Lansing et al. 2004). Although there is no evidence of Indian Y chromosomes in our eastern Indonesian sample, it is possible that a small number of Indian Y chromosomes are present in other Indonesian populations. However, it is not possible to confirm this until larger sample sizes (Table 1) and/or appropriate markers are typed (Kayser et al. 2003).

The discovery that about 12% of Balinese haplogroups are of relatively recent Indian origin and that these haplogroups are rare or absent in neighboring Indonesian islands should prompt a thorough re-examination of the archeological evidence for contact between India and Bali. The genetic evidence we have presented suggests that the magnitude of such trade and other cultural contacts between India and Bali was much greater than has hitherto been imagined (Lansing 1983).

*Acknowledgments* This research was supported by the National Science Foundation's Biocomplexity Program (through grant BCS 0083524 awarded to J. S. Lansing and M. F. Hammer) and by the National Science Foundation (through grants SBR-9514733, SBR-9512178, and SBR-9818215 awarded to L. B. Jorde). Research in Indonesia was carried out under the auspices of the Balai Pengkajian Teknologi Pertanian, Denpasar, Bali, with the kind support and assistance of its director, Dr. Suprpto. We wish to express our gratitude to Wayan Ardika of Udayana University; to the staff of the Balai Purbakala, Bali, and its director, Ayu Kusumawati; and to the Lembaga Ilmu Pengetahuan Indonesia. We thank Svetlana Reznikova, Christine Ponder, Rupesh Amin, Amit Indap, Jennifer Gulick, and Veronica Contreras for excellent technical assistance.

*Received 19 January 2004; revision received 29 June 2004.*

## **Literature Cited**

- Ardika, I. W., and P. Bellwood. 1991. Sembiran: The beginnings of Indian contact with Bali. *Antiquity* 65:221–232.
- Bamshad, M., T. Kivisild, W. S. Watkins et al. 2001. Genetic evidence on the origins of Indian caste populations. *Genome Res.* 11:994–1,004.

- Bandelt, H. J., P. Forster, and A. Rohlf. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* 16:37–48.
- Bellwood, P. 1997. *Prehistory of the Indo-Malaysian Archipelago*. Honolulu: University of Hawaii Press.
- Bertorelle, G., and L. Excoffier. 1998. Inferring admixture proportions from molecular data. *Mol. Biol. Evol.* 15:1,298–1,311.
- Bowler, J. M., H. Johnston, J. M. Olley et al. 2003. New ages for human occupation and climatic change at Lake Mungo, Australia. *Nature* 421:837–840.
- Capelli, C., J. F. Wilson, M. Richards et al. 2001. A predominantly indigenous paternal heritage for the Austronesian-speaking peoples of insular Southeast Asia and Oceania. *Am. J. Hum. Genet.* 68:432–443.
- Chakraborty, R., M. I. Kamboh, M. Nwankwo et al. 1992. Caucasian genes in American blacks: New data. *Am. J. Hum. Genet.* 50:145–155.
- de Knijff, P. 2000. Messages through bottlenecks: On the combined use of slow and fast evolving polymorphic markers on the human Y chromosome. *Am. J. Hum. Genet.* 67:1,055–1,061.
- Diamond, J. 1988. Express train to Polynesia. *Nature* 336:307–308.
- Dyen, I. 1962. A lexicostatistical classification of the Austronesian languages. *Language* 38:38–46.
- Felsenstein, J. 1995. *PHYLIP: Phylogeny Inference Package*. Seattle: Department of Genetics, University of Washington.
- Goldstein, D. B., A. Ruiz Linares, L. L. Cavalli-Sforza et al. 1995. An evaluation of genetic distances for use with microsatellite loci. *Genetics* 139:463–471.
- Hammer, M. F., F. Blackmer, D. Garrigan et al. 2003. Human population structure and its effects on sampling Y-chromosome sequence variation. *Genetics* 164:1,495–1,509.
- Hammer, M. F., T. M. Karafet, A. J. Redd et al. 2001. Hierarchical patterns of global human Y-chromosome diversity. *Mol. Biol. Evol.* 18:1,189–1,203.
- Hammer, M. F., A. J. Redd, E. T. Wood et al. 2000. Jewish and Middle Eastern non-Jewish populations share a common pool of Y-chromosome biallelic haplotypes. *Proc. Natl. Acad. Sci. USA* 97:6,769–6,774.
- Hammer, M. F., and S. L. Zegura. 2002. The human Y-chromosome haplogroup tree: Nomenclature and phylogeography of its major divisions. *Annu. Rev. Anthropol.* 31:303–321.
- Hurles, M. E., J. Nicholson, E. Bosch et al. 2002. Y chromosomal evidence for the origins of Oceanic-speaking peoples. *Genetics* 160:289–303.
- Jobling, M. A., and C. Tyler-Smith. 2003. The human Y chromosome: An evolutionary marker comes of age. *Nat. Rev. Genet.* 4:598–612.
- Karafet, T. M., L. P. Osipova, M. A. Gubina et al. 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.
- Karafet, T., L. Xu, R. Du et al. 2001. Paternal population history of East Asia: Sources, patterns, and microevolutionary processes. *Am. J. Hum. Genet.* 69:615–628.
- Kayser, M., S. Brauer, G. Weiss et al. 2000. Melanesian origin of Polynesian Y chromosomes. *Curr. Biol.* 10:1,237–1,246.
- Kayser, M., S. Brauer, G. Weiss et al. 2001. Independent histories of human Y chromosomes from Melanesia and Australia. *Am. J. Hum. Genet.* 68:173–190.
- Kayser, M., S. Brauer, G. Weiss et al. 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., P. de Knijff, P. Deltjes et al. 1997. Applications of microsatellite-based Y chromosome haplotyping. *Electrophoresis* 18:1,602–1,607.
- Kayser, M., L. Roewer, M. Hedman et al. 2000. Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs. *Am. J. Hum. Genet.* 66:1,580–1,588.
- Kivisild, T., S. Rootsi, M. Metspalu et al. 2003. The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. *Am. J. Hum. Genet.* 72:313–332.

- Lansing, J. S. 1983. The "Indianization" of Bali. *J. South Southeast Asian Stud.* 14:409–421.
- Lansing, J. S. 1995. *The Balinese*. Orlando: Harcourt Brace.
- Lansing, J. S., A. J. Redd, T. M. Karafet et al. 2004. An Indian trader in Bali? *Antiquity* 78:287–293.
- Majumdar, R. C. 1963. *Ancient Indian Colonization in Southeast Asia*. Calcutta: Baroda.
- Melton, T., S. Clifford, J. Martinson et al. 1998. Genetic evidence for the proto-Austronesian homeland in Asia: mtDNA and nuclear DNA variation in Taiwanese aboriginal tribes. *Am. J. Hum. Genet.* 63:1,807–1,823.
- Minch, E., A. Ruiz-Linares, D. B. Goldstein et al. 1997. *Microsat 1.5: A Computer Program for Calculating Various Statistics on Microsatellite Allele Data*. Palo Alto, CA: Department of Genetics, Stanford University.
- Oppenheimer, S. 1998. *Eden in East: The Drowned Continent of Southeast Asia*. London: Weidenfeld & Nicholson.
- Pandya, A., T. E. King, F. R. Santos et al. 1998. A polymorphic human Y-chromosomal G to A transition found in India. *Ind. J. Hum. Genet.* 4:52–61.
- Ramana, G. V., B. Su, L. Jin et al. 2001. Y-chromosome SNP haplotypes suggest evidence of gene flow among caste, tribe, and the migrant Siddi populations of Andhra Pradesh, South India. *Eur. J. Hum. Genet.* 9:695–700.
- Redd, A. J., A. B. Agellon, V. A. Kearney et al. 2002. Forensic value of 14 novel STRs on the human Y chromosome. *Forensic Sci. Int.* 130:97–111.
- Redd, A. J., J. Roberts-Thomson, T. Karafet et al. 2002. Gene flow from the Indian subcontinent to Australia: Evidence from the Y chromosome. *Curr. Biol.* 12:673–677.
- Redd, A. J., N. Takezaki, S. T. Sherry et al. 1995. Evolutionary history of the COII/tRNALys intergenic 9 base pair deletion in human mitochondrial DNAs from the Pacific. *Mol. Biol. Evol.* 12:604–615.
- Richards, B., J. Skoletsky, A. P. Shuber et al. 1993. Multiplex PCR amplification from the CFTR gene using DNA prepared from buccal brushes/swabs. *Hum. Mol. Genet.* 2:159–163.
- Richards, M., S. Oppenheimer, and B. Sykes. 1998. mtDNA suggests Polynesian origins in Eastern Indonesia. *Am. J. Hum. Genet.* 63:1,234–1,236.
- Roberts, D. F., and R. W. Hiorns. 1965. Methods of analysis of the genetic composition of a hybrid population. *Hum. Biol.* 37:38–43.
- Santos, F. R., A. Pandya, M. Kayser et al. 2000. A polymorphic L1 retroposon insertion in the centromere of the human Y chromosome. *Hum. Mol. Genet.* 9:421–430.
- Semino, O., G. Passarino, P. J. Oefner et al. 2000. The genetic legacy of Paleolithic *Homo sapiens sapiens* in extant Europeans: A Y-chromosome perspective. *Science* 290:1,155–1,159.
- Stumpf, M. P., and D. B. Goldstein. 2001. Genealogical and evolutionary inference with the human Y chromosome. *Science* 291:1,738–1,742.
- Su, B., L. Jin, P. Underhill et al. 2000. Polynesian origins: Insights from the Y chromosome. *Proc. Natl. Acad. Sci. USA* 97:8,225–8,228.
- Su, B., C. Xiao, R. Deka et al. 2000. Y chromosome haplotypes reveal prehistorical migrations to the Himalayas. *Hum. Genet.* 107:582–590.
- Su, B., J. Xiao, P. Underhill et al. 1999. Y-chromosome evidence for a northward migration of modern humans into Eastern Asia during the last ice age. *Am. J. Hum. Genet.* 65:1,718–1,724.
- Sykes, B., A. Leiboff, J. Low-Beer et al. 1995. The origins of the Polynesians: An interpretation from mitochondrial lineage analysis. *Am. J. Hum. Genet.* 57:1,463–1,475.
- Thorne, A., R. Grun, G. Mortimer et al. 1999. Australia's oldest human remains: Age of the Lake Mungo 3 skeleton. *J. Hum. Evol.* 36:591–612.
- Underhill, P. A., P. Shen, A. A. Lin et al. 2000. Y-chromosome sequence variation and the history of human populations. *Nat. Genet.* 26:358–361.
- van Leur, J. C. 1955. *Indonesian Trade and Society: Essays in Asian Social and Economic History*. Nijhoff, Netherlands: Hague Banding.
- Wells, R. S., N. Yuldasheva, R. Ruzibakiev et al. 2001. The Eurasian heartland: A continental perspective on Y-chromosome diversity. *Proc. Natl. Acad. Sci. USA* 98:10,244–10,249.
- Y-Chromosome Consortium. 2002. A nomenclature system for the tree of Y-chromosomal binary haplogroups. *Genome Res.* 12:339–348.

