Genetic Admixture History of Eastern Indonesia as Revealed by Y-Chromosome and Mitochondrial DNA Analysis

S. Mona, K. Grunz, S. Brauer, Brigitte Pakendorf, L. Castri, H. Sudoyo, S. Marzuki, H. Barnes, J. Schmidtke, M. Stoneking, et al.

Genetic Admixture History of Eastern Indonesia as Revealed by Y-Chromosome and Mitochondrial DNA Analysis

Stefano Mona,†[1] Katharina E. Grunz,‡[1,2] Silke Brauer,*§ Brigitte Pakendorf,‡‖[3] Loredana Castrì,¶ Herawati Sudoyo,# Sangkot Marzuki,# Robert H. Barnes,** Jörg Schmidtke,†† Mark Stoneking,‡ and Manfred Kayser**

*Department of Forensic Molecular Biology, Erasmus University Medical Center Rotterdam, Rotterdam, The Netherlands; †Department of Biology and Evolution, University of Ferrara, Ferrara, Italy; ‡Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany; §Netherlands Forensic Institute, The Hague, The Netherlands; ‖Department of Linguistics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany; ¶Department of Evolutionary Biology, University of Bologna, Bologna, Italy; #Eijkman Institute for Molecular Biology, Jakarta, Indonesia; **Institute of Social and Cultural Anthropology, University of Oxford, Oxford, United Kingdom; and ††Institute of Human Genetics, Hannover Medical School, Hannover, Germany

Eastern Indonesia possesses more linguistic diversity than any other region in Southeast Asia, with both Austronesian (AN) languages that are of East Asian origin, as well as non-Austronesian (NAN) languages of likely Melanesian origin. Here, we investigated the genetic history of human populations from seven eastern Indonesian islands, including AN and NAN speakers, as well as the relationship between languages and genes, by means of nonrecombining Y-chromosomal (NRY) and mitochondrial DNA (mtDNA) analysis. We found that the eastern Indonesian gene pool consists of East Asian as well as Melanesian components, as might be expected based on linguistic evidence, but also harbors putative indigenous eastern Indonesian signatures that perhaps reflect the initial occupation of the Wallacea by aboriginal hunter-gatherers already in Palaeolithic times. Furthermore, both NRY and mtDNA data showed a complete lack of correlation between linguistic and genetic relationships, most likely reflecting genetic admixture and/or language shift. In addition, we noted a small fraction of the NRY and mtDNA data shared between eastern Indonesians and Australian Aborigines likely reflecting an ancient link between Asia and Australia. Our data thus provide insights into the complex genetic ancestry history of eastern Indonesian islanders characterized by several admixture episodes and demonstrate a clear example of the lack of the often-assumed correlation between the genes and languages of human populations.

Introduction

Island Southeast Asia links two continental regions and thus has served as a crossroad for migrations between them; eastern Indonesia (EI) is an important yet in general understudied part of this area. Of special interest is the fact that EI has the highest linguistic diversity of any Southeast Asian region, with two completely unrelated groups of languages, Austronesian (AN) and non-Austronesian (NAN) also called Papuan. AN-speaking groups are found all over Southeast Asia including EI, eastward in parts of coastal and most of island Melanesia and throughout Micronesia and Polynesia, as well as westward on Madagascar (Bellwood et al. 1995; Adelaar and Himmelmann 2005). All AN languages trace back to a common ancestral language (Proto-AN) and are thought to have spread by an expansion that started about 5,500–6,000 years ago in Taiwan with an assumed ultimate origin in East Asia (Kirch 1997; Blust 1999; Diamond and Bellwood 2003; Bellwood 2004; Gray et al. 2009). NAN languages dominate the New Guinea mainland and are also found on a few islands north, northeast, and east of mainland New Guinea, as well as eastward in parts of the Solomon Islands and westward in parts of EI (Wurm and Hattori 1981; Foley 1986; Ross 2005; Specht 2005). The extremely large amount of linguistic diversity among the perhaps 800 NAN languages suggests that if there is a common origin for these languages, it was an ancient one, presumably in New Guinea (Foley 1986, 2000). Indeed, although AN languages are coherently classified into a single family by the standard comparative method (i.e., they descend from a single recent common ancestor) (Adelaar and Himmelmann 2005), NAN languages belong to several families that do not have clear phylogenetic relationships (Foley 2000). Furthermore, AN languages share many homologous words (cognates), which have proved useful in reconstructing their phylogenetic history (Gray and Jordan 2000; Gray et al. 2009); conversely, NAN languages have few plausible cognates but nonetheless display some structural similarities, which distinguish them from the AN family (Dunn et al. 2005).

The AN languages of the Nusa Tenggara Timur and eastern Timor area are generally assumed to be related (Klamer 2002) belonging to the Central-Malayo-Polynesian (CMP) subfamily (Adelaar and Himmelmann 2005). However, the existence of this subfamily has been questioned (Blust 1993; Donohue and Grimes 2008), and it is also unknown where exactly in Southeast Asia Proto-Malayo-Polynesian was spoken, probably in the Philippines–Sulawesi area. It is further believed that AN languages exist in EI since about 4,000 ybp (Pawley 1999; Spriggs 2003), and AN speakers—on their dispersal from Taiwan—replaced or admixed with NAN-speaking regional aboriginal hunter-gatherers living in Southeast Asia (including EI) already since Palaeolithic times (Diamond and Bellwood 2003). Like the AN languages of the Nusa Tenggara Timur and eastern Timor area, the NAN languages of this region have also been shown to be related with each other (Klamer et al. 2008). The major linguistic hypothesis suggests that NAN speakers in the region of Nusa Tenggara Timur and

[1] These authors contributed equally to this work.

[2] Present address: Max-Delbrück Center for Molecular Medicine, Berlin, Germany.

[3] Present address: Max Planck Research Group on Comparative Population Linguistics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany.

Key words: eastern Indonesia, admixture, Y-Chromosome, mtDNA, Austronesian, Papuan.

E-mail: m.kayser@erasmusmc.nl.

Mol. Biol. Evol. 26(8):1865–1877. 2009
doi:10.1093/molbev/msp097
Advance Access publication May 4, 2009

eastern Timor trace back to the expansion of the Trans-New Guinea Phylum (TNG) that started about 10,000 ybp in the highlands of eastern (now Papua) New Guinea (Pawley 1998, 2005; Foley 2000; Ross 2005). However, it appears to be rather uncertain whether regional NAN–TNG languages were established already before AN speakers arrived in EI or whether they arrived only at a later date (Pawley 2005). An alternative view suggests that NAN speakers in Nusa Tenggara Timur and eastern Timor are remnants of ancient continuous populations already living in the region from the Pleistocene, rather than assuming that New Guinea highlanders migrated back over water to small EI islands (Klamer et al. 2008).

The currently known archaeological sites for modern humans in EI date back to about 30–37,000 ybp in eastern Timor (O'Connor et al. 2002) and northern Maluku (Bellwood 1996). However, older dates of 43,000 ybp from southern Thailand (Anderson 1997), as well as evidence that modern humans were in New Guinea at least 40,000 ybp (Groube et al. 1986) and in Australia at least 50,000 ybp (Roberts et al. 1990), raise the expectation that the initial colonization of EI by modern humans occurred more than 50,000 ybp (Spriggs 2000). In addition, archaeological sites with Neolithic dates of about 4,500–3,800 ybp are known from Flores and of about 4,400–3,400 ybp from eastern Timor, which have been associated with the Neolithic spread of farmers from Taiwan via Southeast Asia, Island Melanesia into Polynesia (Spriggs 2003). This Neolithic spread is usually associated with the spread of AN speakers (Bellwood 1997; Diamond and Bellwood 2003; Spriggs 2003). It has been suggested that the AN-speaking Neolithic farmers arrived in EI via an eastern route via Sulawesi (and perhaps the northern Maluku) rather than a western route via Java (Pawley 1999; Spriggs 2003).

Notably, some authors have argued for a mostly Island Southeast Asia origin of AN groups in Palaeolithic times based on mitochondrial DNA (mtDNA) dating (Oppenheimer and Richards 2001; Hill et al. 2007). However, this model is not in agreement with the Taiwanese origin of AN languages (Blust 1999) and the topology of the phylogenetic AN-language tree (Gray and Jordan 2000; Gray et al. 2009), nor with genetic evidence from the human-specific bacterial parasite *Helicobacter pylori*, which also points to a Taiwanese origin and Out-of-Taiwan spread of AN speakers (Moodley et al. 2009). In addition, this model disagrees with human genetic data such as the Taiwanese origin of the major Asian nonrecombining Y-chromosomal (NRY) haplogroup in Island Melanesia (O-M110) (Kayser, Choi, et al. 2008), the Taiwanese origin of the B4a1a mtDNA haplogroup shared among Taiwanese Aborigines, Melanesians, and Polynesians (Trejaut et al. 2005), and the genetic affinities of Polynesians and Micronesians with Taiwanese Aborigines as revealed from 890 autosomal DNA markers (Friedlaender et al. 2008).

Overall, the linguistic and archaeological data suggest at least two or three distinct migration waves influenced EI: an initial colonization at least 50,000 ybp; the expansion of AN speakers about 4,000 ybp and possibly an immigration of NAN speakers either before or around the same time of the AN arrival. To what extent is this complex history of EI, as suggested by linguistic and archaeological data, reflected in the gene pool of contemporary islanders? Previously, we found preliminary genetic evidence for both East Asian and Melanesian paternal and maternal lineages in EI (Kayser et al. 2001, 2006). However, these results were based on just a few individuals from a few islands of the Maluku and Nusa Tenggara regions. Paternal lineages of East Asian as well as Melanesian ancestry were also observed in two other studies concerning EI, but only including samples from two of the many regional islands (Sumba and Flores) (Lansing et al. 2007, 2008). In the present study, we investigated the human genetic history of the region of Nusa Tenggara Timur and eastern Timor in more detail by analyzing NRY and mtDNA diversity in 344 and 330 individuals, respectively, from seven islands (Flores, Adonara, Solor, Lembata, Pantar, Alor, and from eastern Timor), including both AN- and NAN-speaking groups. We also examined whether the genetic and linguistic relationships of EI groups were correlated, and additionally investigated the extent to which genetic contributions from the various migrations that have influenced this region can be detected in the gene pool of contemporary EI islanders.

Material and Methods

Samples and Genotyping

Samples were collected from several villages each from six EI islands (Adonara, Alor, Lembata, Flores, Solor, and Pantar) belonging to the Indonesian province Nusa Tenggara Timur, as well as from the Democratic Republic of Timor L'este (referred to here as eastern Timor; fig. 1). Samples were collected as cheek swabs and with individual informed consent by R.B., J.S., and H.S. with the collaboration and assistance of the Eijkman Institute for Molecular Biology, Jakarta, and with the advice and active assistance of local health officials. Some of the sampled villages spoke CMP languages belonging to the AN family, whereas others spoke NAN languages that are mostly considered to belong to the TNG (Pawley 2005; Ross 2005; Klamer et al. 2008). Samples were sorted according to the paternal grandfather's birthplace/language for the NRY analysis and the maternal grandmother's birthplace/language for the mtDNA analysis. Grouping samples by island and language resulted in 10 groups: an AN group for each of the seven islands representing different AN languages, and an additional NAN group for each of Alor, Pantar, and eastern Timor representing different NAN languages (tables 1 and 2).

DNA was extracted from the cheek swabs via a salting-out protocol (Kayser et al. 2003). We analyzed 35 binary NRY markers and seven NRY short tandem repeat polymorphisms (Y-STRs) in 344 EI individuals, using previously described methods (Kayser et al. 2003, 2006). The DYS390 Y-STR was sequenced in individuals belonging to NRY haplogroup C-RPS4Y* to search for the 390.1 deletion polymorphism as described previously (Kayser et al. 2001). The hypervariable region 1 (HV1) of mtDNA was amplified in 330 EI individuals using primers L16001 and H16410 (Handt et al. 1996; Cordaux et al. 2003) and sequenced using Big Dye (Applied Biosystems, Foster City, CA) chemistry on an ABI 377 or ABI 3700 DNA Sequencer (Applied Biosystems). Both DNA strands were sequenced separately; for those samples with the

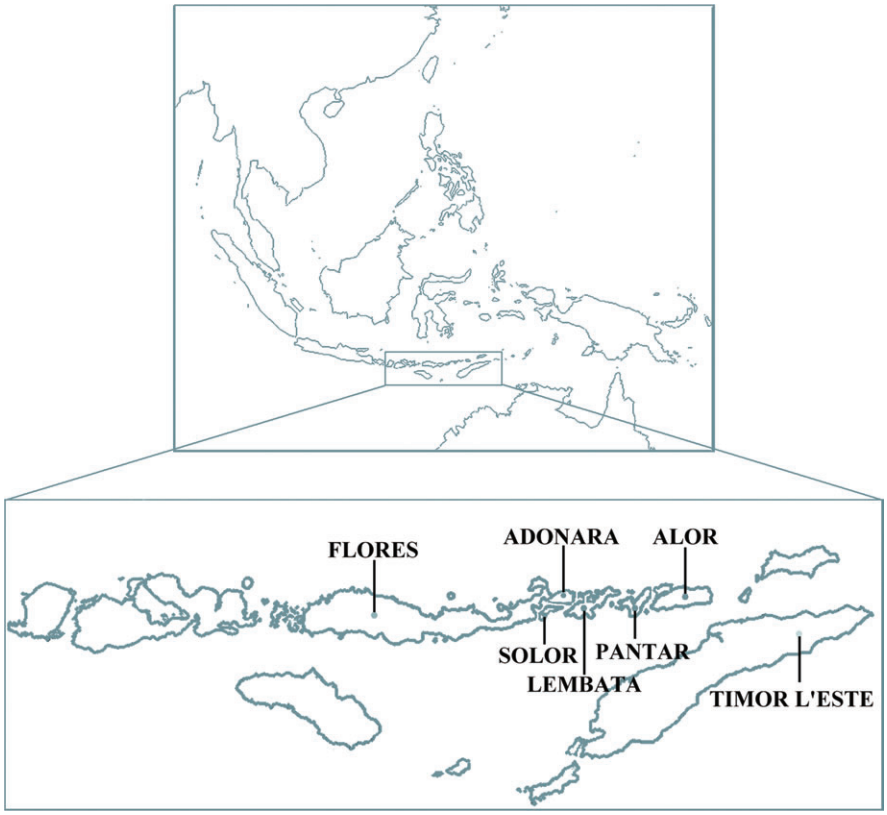

FIG. 1.—Map of Southeast Asia with the sampled islands of the Nusa Tenggara Timur region and eastern Timor (Timor L'este) of EI.

"C-stretch" in the region 16,184–16,193, both strands were sequenced twice. Sequences were analyzed using the SeqManII software from the Lasergene software package (DNASTAR Inc., Madison, WI). The mtDNA 9-bp deletion was typed as described elsewhere (Redd et al. 1995). NRY binary markers were used to characterize NRY haplogroups (supplementary fig. S1, Supplementary Material online) considering the most recent Y-Chromosome Consortium update on nomenclature (Karafet et al. 2008), whereas mtDNA haplogroups were inferred using sequence data and the 9-bp deletion data using the commonly applied nomenclature (supplementary fig. S2, Supplementary Material online). MtDNA sequences characterized by 16217C, 16261T, 16247G, and the 9-bp deletion have traditionally been classified as carrying the "Polynesian Motif." Because this motif comprises only one of several subgroups of haplogroup B4a1a1 that are differentiated by mtDNA polymorphisms outside HV1 (Trejaut et al. 2005;

Table 1
NRY Haplogroups Observed in Eastern Indonesian Populations with Their Previously Assumed Geographic Origin

NRY Haplogroup	Origin[a]	Adonara (AN)	Alor (NAN)	Alor (AN)	E-Timor (AN)	E-Timor (NAN)	Flores (AN)	Lembata (AN)	Pantar (NAN)	Pantar (AN)	Solor (AN)	All AN	All NAN	Overall
C-M38*	M	73	9	3	12	2	28	6	8	2	16	140	19	159
C-M208	M	4	0	0	0	0	0	0	0	0	1	5	0	5
K-M9*	M[b]/U	2	8	0	7	0	8	3	0	1	7	28	8	36
M-M4*	M	2	0	0	1	0	2	0	0	0	0	5	0	5
M-P34	M	0	3	0	2	0	0	0	3	2	0	4	6	10
S-M230*	M	0	0	0	0	0	0	2	0	0	0	2	0	2
S-M254*	M	2	1	1	3	2	9	6	5	3	4	28	8	36
C-RPS4Y*	A	10	1	0	3	0	17	13	0	0	5	48	1	49
NO-M214*	A	0	0	0	0	0	1	0	0	0	0	1	0	1
O-M175*	A	1	0	0	0	0	1	0	1	0	3	5	1	6
O-M119	A	1	0	0	6	1	0	0	1	0	0	7	2	9
O-M122*	A	0	0	0	0	0	5	1	9	2	7	15	9	24
O-M134	A	0	0	0	0	0	0	0	1	0	0	0	1	1
F-M89*	U	1	0	0	0	0	0	0	0	0	0	1	0	1
Sample size		96	22	4	34	5	71	31	28	10	43	289	55	344

[a] As previously assumed M: Melanesian origin, A: Asian origin, and U: unknown.
[b] Previous assumption made for Polynesian Y-chromosome, which may not hold for eastern Indonesian ones (see text).

Table 2
MtDNA Haplogroups Observed in Eastern Indonesian Populations with Their Previously Assumed Geographic Origin

MtDNA Haplogroup	Origin[a]	Adonara (AN)	Alor (NAN)	Alor (AN)	E-Timor (AN)	E-Timor (NAN)	Flores (AN)	Lembata (AN)	Pantar (NAN)	Pantar (AN)	Solor (AN)	All AN	All NAN	Overall
B	A	0	0	0	2	0	0	1	0	0	0	3	0	3
B4a	A	2	1	0	1	1	3	3	0	0	1	10	2	12
B4b1	A	5	2	0	0	0	9	3	1	0	4	21	3	24
B4c1b	A	4	0	0	0	0	1	1	3	2	0	8	3	11
B4c2	A	0	0	0	0	0	1	0	0	0	0	1	0	1
B5a	A	0	0	0	0	0	1	1	0	0	0	2	0	2
B5b	A	4	1	0	0	1	6	0	0	0	1	11	2	13
D	A	6	6	1	0	0	6	1	1	0	1	15	7	22
D5d1	A	0	0	0	0	0	3	1	0	0	1	5	0	5
F1a	A	17	2	0	2	0	10	6	2	0	4	39	4	43
F1a1	A	2	0	0	0	0	1	1	0	0	0	4	0	4
M7b	A	0	0	0	0	0	0	0	1	0	0	0	1	1
M7b3	A	0	0	0	0	0	1	0	0	0	0	1	0	1
M7c1a	A	0	0	0	0	0	0	0	0	0	7	7	0	7
M7c1c	A	5	1	0	0	0	7	0	2	0	2	14	3	17
R9c	A	2	1	0	2	1	1	2	2	0	1	8	4	12
U2	A	1	1	0	0	0	0	0	0	0	0	1	1	2
Y2	A	1	0	0	0	0	0	0	0	0	0	1	0	1
Pre-Z	A	0	0	0	0	0	6	0	0	0	0	6	0	6
P1	M	0	1	0	2	0	0	2	0	1	0	5	1	6
P1e	M	0	0	0	0	0	1	0	0	0	0	1	0	1
P4a	M	0	0	0	0	0	0	0	0	0	1	1	0	1
Q1	M[b]	8	2	2	9	0	6	4	7	6	3	38	9	47
Q2	M[b]	4	0	0	0	0	0	2	0	0	1	7	0	7
R14	M	0	0	0	1	0	0	1	0	0	0	2	0	2
E1a	E	0	1	0	8	2	3	2	0	1	0	14	3	17
E1b	E	1	1	0	1	0	1	0	1	0	0	3	2	5
E2	E	1	0	0	1	0	0	0	0	0	5	7	0	7
Polynesian motif	E	0	0	0	2	0	3	0	0	0	0	5	0	5
M42	AU	1	0	0	0	0	0	0	0	0	0	1	0	1
N12	AU	0	0	0	0	0	2	0	0	0	0	2	0	2
Rest	U	13	4	0	2	0	3	3	8	0	9	30	12	42
Sample size		77	24	3	33	5	75	34	28	10	41	273	57	330

[a] As previously assumed M: Melanesian origin, A: East Asian origin, E: eastern Indonesian origin but ultimately traceable to East Asia by precursor motif, AU: Australian origin, and U: unknown.

[b] Evidence presented here suggests eastern Indonesian origin (see text).

Pierson et al. 2006) that were not analyzed in the present study, we classified these sequences as Polynesian Motif rather than as haplogroup B4a1a1. All EI mtDNA HV1 sequences are available from GenBank under the following accession numbers: FJ838794–FJ839123. Additional NRY and mtDNA data from East and Southeast Asia, Melanesia, Polynesia, and Australia (Tommaseo-Ponzetta et al. 2002; Kayser et al. 2003, 2006; Mona et al. 2007) were included for comparative analyses and are provided in Supplementary tables S1 and S2, Supplementary Material online, respectively.

Statistical Analyses

Diversity indices, F_{st}/R_{st} values, and analysis of molecular variance (AMOVA) were computed using the software ARLEQUIN 3.0 (Excoffier et al. 2005). Genetic distances between populations were displayed using the multidimensional scaling algorithm (MDS), implemented in the software STATISTICA. Median-joining networks (Bandelt et al. 1999) among Y-STR haplotypes as well as mtDNA HVR1 sequences were built using the software NETWORK (http://www.fluxus-technology.com/sharenet.htm); a weighting scheme was applied to the Y-STR loci, based on estimated mutation rates, as described elsewhere (Mona et al. 2007), whereas default settings were used for the mtDNA analyses. NETWORK was also used to compute the time to the most recent common ancestor (TMRCA) of mtDNA haplogroups by means of the rho statistic and its standard deviation (Forster et al. 1996; Saillard et al. 2000).

A Bayesian-based coalescent approach (Wilson and Balding 1998; Wilson et al. 2003), implemented in the software BATWING, was used for demographic inference of NRY haplogroups using Y-STRs and the NRY haplogroup tree topology. The coalescent prior model used for the topology and branch lengths of the gene genealogy was an initial constant population size followed by a demographic expansion (Wilson et al. 2003). The likelihood of the gene genealogy was computed under the stepwise mutation model (Ohta and Kimura 1973). The posterior probability of the gene genealogy, population genetic parameters, and NRY haplogroup dating were approximated through the Metropolis–Hastings algorithm (Metropolis et al. 1953; Hastings 1970). Priors for the Y-STR mutation rates and the coalescent model were applied as described previously (Kayser et al. 2006).

To determine the coalescence time of each haplogroup, the gene genealogy was constrained using the

known NRY phylogeny (supplementary fig. S1, Supplementary Material online). The final analysis was based on two runs of 100 million Markov chain Monte Carlo generations each with a 10% burn-in period. TRACER (Rambaut and Drummond 2004) (available at http://tree.bio.ed.ac.uk/software/tracer/) was used to check for the convergence of the two runs and to compute the effective sample size (always >200) and the 95% high posterior density of all the parameters, combining the two runs.

Results and Discussion

Geographic Origins and Migration History of Eastern Indonesians

Based on the analysis of 35 binary NRY markers, we identified 14 NRY haplogroups in the 10 EI groups from the region of Nusa Tenggara Timur and eastern Timor (table 1, supplementary fig. S1, Supplementary Material online). A Melanesian origin was inferred previously for seven NRY haplogroups (C-M38*, C-M208, M-M4*, M-P34, S-M230*, S-M254*, and K-M9*), whereas for six NRY haplogroups (O-M122*, O-M134, O-M119, O-M175*, NO-M214*, and C-RPS4Y*), an East Asian origin was assumed before (Kayser et al. 2000, 2001, 2003, 2006; Capelli et al. 2001). One man from Adonara was haplogroup F-M89*, for which the origin cannot be traced given the markers analyzed. Haplogroup K-M9* is potentially of polyphyletic origin as it represents one of the oldest paternal lineages in Asia/Oceania (supplementary fig. S1, Supplementary Material online). Although we previously assigned a Melanesian origin for K-M9* in an analysis of Polynesian Y-chromosomes (Kayser et al. 2006), this assumption may not hold for EI given that K-M9* was also observed in Southeast Asia outside EI, albeit in low frequency (supplementary table S1, Supplementary Material online). Therefore, K-M9* was not included in the analyses concerning the ancestry of EI lineages.

Based on the mtDNA HV1 sequences and the 9-bp deletion data, we identified 31 mtDNA haplogroups (table 2, supplementary fig. S2, Supplementary Material online): 19 mtDNA haplogroups (B, B4a, B4b1, B4c1b, B4c2, B5a, B5b, D, D5d1, F1a, F1a1, M7b, M7b3, M7c1c, M7c1a, R9c, U2, Y2, and pre-Z) have a previously inferred East Asian origin (Redd et al. 1995; Kivisild et al. 2002; Yao et al. 2002; Trejaut et al. 2005); six mtDNA haplogroups (Q1, Q2, P1, P1e, P4a, and R14) have a previously inferred Melanesian origin (Forster et al. 2001; Friedlaender et al. 2005, 2007; Hudjashov et al. 2007); and four mtDNA haplogroups (E1a, E1b, E2, and the so-called Polynesian Motif) were previously assumed to be of indigenous EI origin (Lum et al. 1994; Redd et al. 1995; Lum and Cann 1998; Richards et al. 1998; Trejaut et al. 2005; Hill et al. 2007) but can be ultimately traced back to East Asia/Taiwan by their precursor haplogroups (supplementary fig. S2, Supplementary Material online). Two mtDNA haplogroups observed in very low frequency (M42 in one Adonara male and N12 in two Flores males) were previously only found in Australian Aborigines (Pierson et al. 2006; Hudjashov et al. 2007). Finally, 42 individuals (12.7%) from our EI data set could not be assigned to any known mtDNA haplogroups, reflecting inherent limitations in mtDNA haplogroup assignments based solely on HV1 sequence and 9-bp deletion data.

Overall, in EI, the proportion of Melanesian paternal and maternal lineages was 63.1% and 20%, with 60% and 20.5% in NAN speakers, and 63.7% and 17.5% in AN speakers, respectively. The frequency of Melanesian NRY haplogroups ranged from 45.2% in AN speakers from Lembata to 100% in AN speakers from Alor (albeit only four individuals) or 84.4% in AN speakers from Adonara, whereas the frequency of Melanesian mtDNA haplogroups ranged from 0% in NAN speakers from eastern Timor (albeit only five individuals in total) to 70% in AN speakers from Pantar. The overall proportion of Asian paternal and maternal lineages was 26.2% and 67%, respectively, with 26.3% and 61.4% in AN speakers and 25.5% and 68.1% in NAN speakers, respectively. Asian components for NRY haplogroups ranged from 0% in AN speakers from Alor (albeit only four individuals in total) or 4.5% in NAN speakers from Alor to 45.2% in AN speakers from Lembata; and for mtDNA haplogroups from 30% in AN speakers from Pantar (albeit only three individuals) to 100% in NAN speakers from Timor (albeit only five individuals) or 84% in AN speakers from Flores (table 3).

The above results rely on the correct geographic classification of mtDNA/NRY haplogroup origins. Usually, geographic origins of haplogroups are inferred from phylogeography and associated diversity and/or TMRCA estimations. However, this approach requires extensive sampling throughout various regions in order to obtain a detailed view of haplogroup distributions. EI has certainly been underrepresented in human population genetic studies carried out so far, and the paucity of available data could have confounded the correct identification of the origin of some haplogroups. We therefore estimated TMRCA and associated diversity for all NRY and mtDNA haplogroups with reasonable sample size for EI and compared the outcomes with our previously published data from East Asia and Melanesia (supplementary tables S3 and S4, Supplementary Material online).

Notably, mtDNA haplogroup P1, for which a Melanesian origin was inferred previously (Friedlaender et al. 2005, 2007), has an older mean TMRCA in EI (53,814 ybp) relative to Melanesia (46,160 ybp) (supplementary table S3, Supplementary Material online), although the dating intervals largely overlap between regions (supplementary table S3, Supplementary Material online) and the EI estimate is only based on six samples. The P1 network analysis (fig. 2*A*) suggests a Melanesian origin in agreement with previous reports (Friedlaender et al. 2005, 2007). Although it is possible to argue that P1 mtDNAs in EI are the result of a recent migration from New Guinea, this is not fully supported by the haplotype sharing analysis, as only one of the three EI haplotypes is shared with Melanesia (fig. 2*A*, supplementary table S5, Supplementary Material online). However, more EI data are needed to fully establish whether mtDNA haplogroup P1 is of Melanesian or of eastern Indonesian origin.

Furthermore, we identified at least two haplogroups (Q1 and Q2) that were previously suggested to be of Melanesian origin (Friedlaender et al. 2007), whereas

Table 3
NRY and mtDNA Haplogroups Observed in Eastern Indonesian Populations Combined according to Their Assumed Region of Origin

Geographic Origin	Adonara (AN)	Alor (AN)	E-Timor (AN)	Flores (AN)	Lembata (AN)	Solor (AN)	Pantar (AN)	Alor (NAN)	E-Timor (NAN)	Pantar (NAN)	All AN	All NAN	Overall
Asian NRY-DNA (%)[a]	12 (12.5)	0	9 (26.5)	24 (33.8)	14 (45.2)	15 (34.9)	2 (20)	1 (4.5)	1 (20)	12 (42.9)	76 (26.3)	14 (25.5)	90 (26.2)
Melanesian NRY-DNA (%)[b]	81 (84.4)	4 (100)	18 (52.9)	39 (54.9)	14 (45.2)	21 (48.8)	7 (70)	13 (59.1)	4 (80)	16 (57.1)	184 (63.7)	33 (60)	217 (63.1)
Other NRY-DNA (%)[c]	3 (3.1)	0	7 (20.6)	8 (11.3)	3 (9.7)	7 (16.3)	1 (10)	8 (36.4)	0	0	29 (10)	8 (14.5)	37 (10.8)
Asian mtDNA (%)[d]	51 (66.2)	1 (33.3)	19 (57.6)	63 (84)	22 (64.7)	27 (65.9)	3 (30)	17 (70.8)	5 (100)	13 (46.4)	35 (61.4)	186 (68.1)	221 (67)
Melanesian mtDNA (%)[e]	12 (15.6)	2 (66.7)	12 (36.4)	9 (12)	9 (26.5)	5 (12.2)	7 (70)	3 (12.5)	0	7 (25)	10 (17.5)	56 (20.5)	66 (20)
Australian mtDNA (%)[f]	1 (1.3)	0	0	0	0	0	0	0	0	0	0	1 (0.4)	1 (0.3)
Other mtDNA (%)[g]	13 (16.9)	0	2 (6.1)	3 (4)	3 (8.8)	9 (22)	0	4 (16.7)	0	8 (28.6)	12 (21.1)	30 (11)	42 (12.7)
MtDNA hg Q1 and Q2 not considered Melanesian													
Asian mtDNA (%)	51 (66.2)	1 (33.3)	19 (57.6)	63 (84)	22 (64.7)	27 (65.9)	3 (30)	17 (70.8)	5 (100)	13 (46.4)	186 (68.1)	35 (61.4)	221 (67)
Melanesian mtDNA (%)	0	0	3 (9.1)	1 (1.3)	3 (8.8)	1 (2.4)	1 (10)	1 (4.2)	0	0	9 (3.3)	1 (1.8)	10 (3)
Australian mtDNA (%)	1 (1.3)	0	0	2 (2.7)	0	0	0	0	0	0	3 (1.1)	0	3 (0.9)
Other mtDNA (%)	25 (32.5)	2 (66.7)	11 (33.3)	9 (12)	9 (26.5)	13 (31.7)	6 (60)	6 (25)	0	15 (53.6)	75 (27.5)	21 (36.8)	96 (29.1)
MtDNA hg E1a, E1b, E2, Polynesian Motif, Q1 and Q2 considered eastern Indonesian													
Asian mtDNA (%)	49 (63.6)	1 (33.3)	7 (21.2)	56 (74.7)	20 (58.8)	22 (53.7)	2 (20)	15 (62.5)	3 (60)	12 (42.9)	157 (57.5)	30 (52.6)	187 (56.7)
Melanesian mtDNA (%)	0	0	3 (9.1)	1 (1.3)	3 (8.8)	1 (2.4)	1 (10)	1 (4.2)	0	0	9 (3.3)	1 (1.8)	10 (3)
Eastern Indonesian mtDNA (%)	14 (18.2)	2 (66.7)	21 (63.6)	13 (17.3)	8 (23.5)	9 (22)	7 (70)	4 (16.7)	2 (40)	8 (28.6)	74 (27.1)	14 (24.6)	88 (26.7)
Australian mtDNA (%)	1 (1.3)	0	0	2 (2.7)	0	0	0	0	0	0	3 (1.1)	0	3 (0.9)
Other mtDNA (%)	13 (16.9)	0	2 (6.1)	3 (4)	3 (8.8)	9 (22)	0	4 (16.7)	0	8 (28.6)	30 (11)	12 (21.1)	42 (12.7)

[a] O-M119, O-M122*, O-M134, O-M175*, NO-M214*, and C-RPS4Y*.
[b] C-M208, C-M38*, S-M230*, S-M254*, M-M4*, and M-P34.
[c] F-M89* and K-M9*.
[d] B, B4a, B4b1, B4c1b, B4c2, B5a, B5b, D, D5d1, E1a, E1b, E2, F1a, F1a1, M7b, M7b3, M7c1a, M7c1c, Polynesian Motif, R9c, U2, Y2, pre-Z.
[e] P1, P1e, P4a, Q1, Q2, and R14.
[f] M42 and N12.
[g] U.

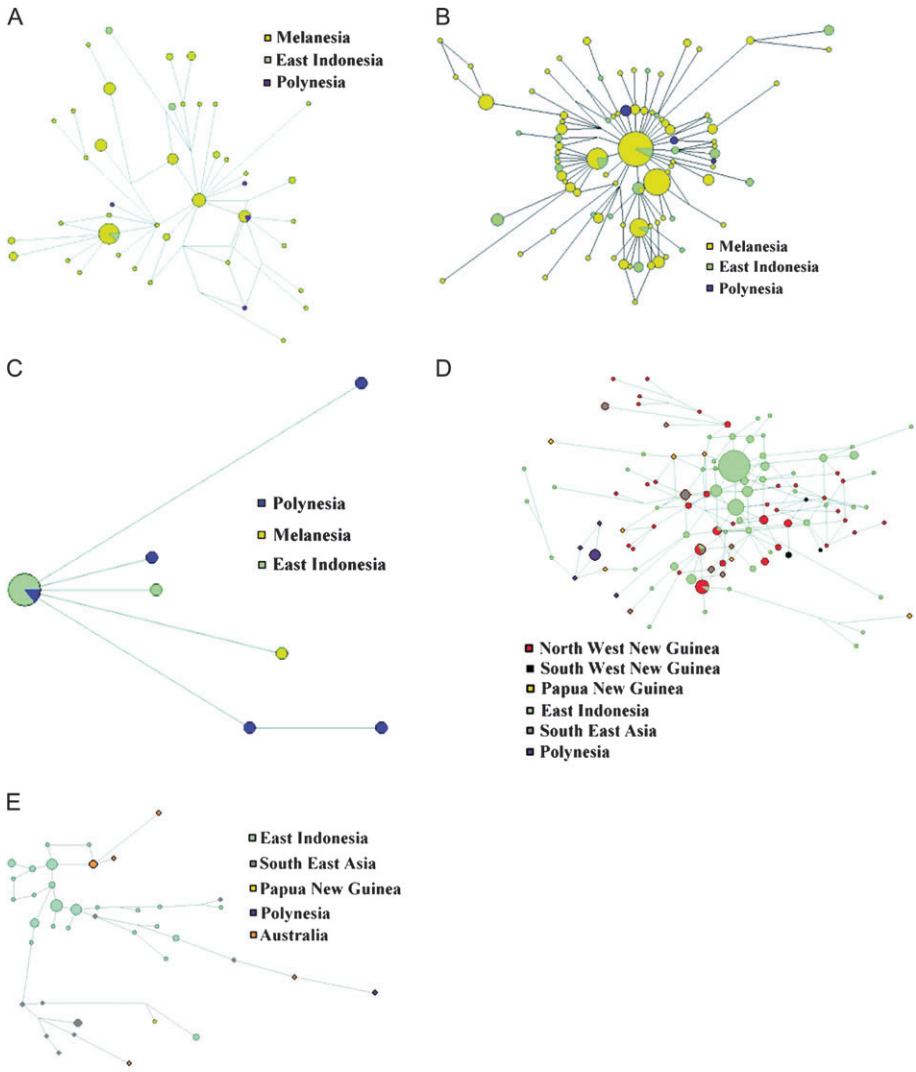

FIG. 2.—Median-joining network of mtDNA and NRY haplogroups (*A*) mtDNA P1, (*B*) mtDNA Q1, (*C*) mtDNA Q2, (*D*) NRY-DNA C-M38*, and (*E*) NRY-DNA C-RPS4Y* based on associated HV1 sequences and associated Y-STR haplotypes, respectively. Nodes represent mutation steps; circles represent haplotypes with the size being proportional to the number of individuals carrying the haplotype. Color code of haplotypes is according to geographic regions as indicated.

our new data suggest they might be of EI origin instead. Most convincing evidence is presented for mtDNA haplogroup Q1, which is of moderate frequency in EI (14.2%) and has a mean TMRCA considerably older in EI (48,089 ybp) relative to Melanesia (34,578 ybp), although dating intervals largely overlap between both regions (supplementary table S3, Supplementary Material online). Moreover, the mtDNA sequence diversity within haplogroup Q1 was very high in EI (0.967 ± 0.01) making it somewhat unlikely that this haplogroup exists in EI as a result of a recent migration from Melanesia; this is further confirmed by the low number of shared haplotypes between the two regions (supplementary able S5, Supplementary Material online). A network analysis was not clear in inferring the region of origin (fig. 2*B*). The TMRCA estimates suggest that Q1 may be a signature of the initial aboriginal hunter-gatherer population in the Pleistocene that occupied both EI and New Guinea.

Similar conclusions may be reached for mtDNA haplogroup Q2, which was observed in somewhat higher frequency in EI than previously reported from Melanesia (supplementary table S2, Supplementary Material online). Although sample size is still relatively small in EI ($N = 7$, but even smaller in Melanesia), a network analysis may suggest a possible EI origin of this mtDNA haplogroup (fig. 2*C*). However, Q2 is divided into two branches, one previously associated with Melanesia and the other with Australia (Hudjashov et al. 2007), which cannot be distinguished on the basis of the HV1 sequences alone. Haplogroup Q2 in EI could therefore be of EI ancestry (like haplogroup Q1), or possibly of Australian ancestry. Sequencing whole mtDNA genomes should shed more light on the origin(s) of haplogroup Q2 in EI.

In sum, at least mtDNA haplogroup Q1, and possibly Q2, could reflect an ancient shared ancestry between EI and New Guinea, that is, the Pleistocene aboriginal

hunter-gather population. Reestimating the Melanesian mtDNA proportions in EI by excluding Q1 and Q2 from the Melanesian component results in an overall Melanesian contribution to EI of only 3%, with 3.3% in AN speakers and 1.8% in NAN speakers (table 3). At the same time, if Q1 and Q2 are considered as being of EI origin, together with E1a, E1b, E2, and the Polynesian Motif as suggested previously (Redd et al. 1995; Richards et al. 1998; Trejaut et al. 2005; Hill et al. 2007), the level of indigenous EI mtDNAs rises to 26.7% (27.1% in AN speakers and 24.6% in AN speakers; table 3). No other previous conclusions on regional geographic origins of NRY and mtDNA haplogroups were altered by the EI data presented here (supplementary tables S3 and S4, Supplementary Material online).

Did the Melanesian contribution to EI happen before or after the arrival of AN speakers in the region? The analysis of haplotype sharing was not informative enough to answer this question as only very few haplotypes were shared between both geographic regions (supplementary tables S5 and S6, Supplementary Material online). Also, the coalescent estimates in EI of NRY haplogroups with a clear East Asian (O-M122* and O-M119) or Melanesian (S-M254*, M-P34) origin are very similar, with their posterior distribution overlapping significantly (supplementary table S4, Supplementary Material online). Moreover, because the TMRCA of a haplogroup depends on the diversity of the founding/migrant population, it is difficult to establish which migration came first (ANs vs. NANs) particularly when the migrations may have occurred at similar times. Furthermore, as described above, there is disagreement concerning the relative timing of the arrival of AN and NAN speakers, based on linguistic data.

The presence of both Melanesian and Asian NRY and mtDNA haplogroups in EI populations is reflected in their intermediate position in the MDS analyses (fig. 3). For NRY data, EI populations appear clustered with AN-speaking Island Melanesians (who also carry Asian and Melanesian Y-chromosomes) due to their higher Melanesian components except for NAN Pantar clustering more with other Southeast Asians due to the high frequency (43%) of Asian NRY haplogroups. The real outlier is Adonara grouping with Australians due to the high frequency of C-RPS4Y*. However, as expected from mtDNA haplogroup frequencies, EI populations cluster with Southeast Asian and mainland East Asian populations in the MDS plot based on mtDNA F_{st} distances except for AN Pantar clustering with NAN groups from New Guinea due to the high frequency (70%) of Melanesian mtDNA haplogroups.

Although less pronounced in EI compared with other regions, we observed a higher eastern Asian component with mtDNA than with NRY-DNA, and conversely a higher Melanesian component with NRY-DNA than with mtDNA in EI, similar to what has been described for Island Melanesia (Kayser, Choi, et al. 2008) and Polynesia (Kayser et al. 2006). As described elsewhere, a history of sex-biased admixture between ANs and NANs might explain this result, confirming previous surveys in Near and Remote Oceania on the modality of the AN migration (Hage and Marck 2003; Kayser et al. 2006; Kayser, Choi, et al. 2008; Kayser, Lao, and Stoneking 2008).

Genetic versus Linguistic Relationships in Eastern Indonesia

The proportions of Asian vers3us Melanesian NRY or mtDNA haplogroups do not differ significantly between AN and NAN groups of EI (table 3). Moreover, an AMOVA provided no support for a population grouping according to linguistic (AN vs. NAN) classification for NRY haplogroups and mtDNA HV1 sequences and only very small (1.4%) and statistically not significant evidence based on NRY haplotypes (table 4). This lack of a correlation between the genetic and linguistic relationships among EI populations is also reflected in the MDS analyses. Indeed, in the scatter plots based on mtDNA and NRY haplogroups, AN and NAN groups from EI usually cluster together (fig. 3). The only exceptions are the AN group from Pantar in the mtDNA analysis (which clustered unexpectedly with NAN groups from New Guinea), and the AN group from Adonara in the NRY analysis (which clustered with Australian). Our data thus demonstrate clear evidence for genetic admixture between AN and NAN groups, which is in line with linguistic evidence showing that some NAN features occur in AN languages and some AN features occur in NAN languages of Nusa Tenggara Timur, eastern Timor, and the Moluccas (Klamer 2002, Klamer et al. 2008). That NAN groups in EI once were more widespread than today is indicated by the previous existence of the NAN language Tambora on Sumbawa island, that—along with almost all its speakers—was lost in a volcanic eruption in 1815 (Donohue 2007). Notably, Sumba lies more than 500 km west of the current NAN languages in EI and is currently populated only by AN speakers. A wider spread of NAN speakers across EI in previous times may explain, by means of genetic admixture, the presence of Melanesian NRY and mtDNA haplogroups on EI islands currently population only by AN speakers, as we observed on Flores, Adonara, Solor, and Lembata, in addition to genetic admixture events between language groups across EI islands.

Several studies have reported a correlation between languages and genes on a continental (Barbujani and Sokal 1990) as well as global perspective (Cavalli-Sforza et al. 1988) or within single islands (Lansing et al. 2007; Hunley et al. 2008). However, regional studies either found no such relationship (Hunley et al. 2008) or observed a gene–language correlation only for maternally inherited mtDNA but not with paternally inherited NRY-DNA due to sex-biased genetic admixture and the inheritance of language via the mothers (Nasidze et al. 2006; Kayser, Choi, et al. 2008). Our finding of a lack of correlation between mtDNA–NRY relationships and AN/NAN language affiliation across EI islands appears to contradict the recent study by Lansing et al. (2007), which came to the opposite conclusion about language–gene relationships in one EI island (Sumba). However, their analysis was at a different geographic scale, as they focused on AN groups only from Sumba and examined the correlation between NRY haplogroups associated with the AN expansion and the retention of proto-AN cognates in the languages of various Sumbanese groups. In contrast, we sampled across several islands and compared the genetic and linguistic relationships of AN and NAN groups, which is a very different geographic and linguistic scale than that of Lansing et al. (2007). A lack of correlation

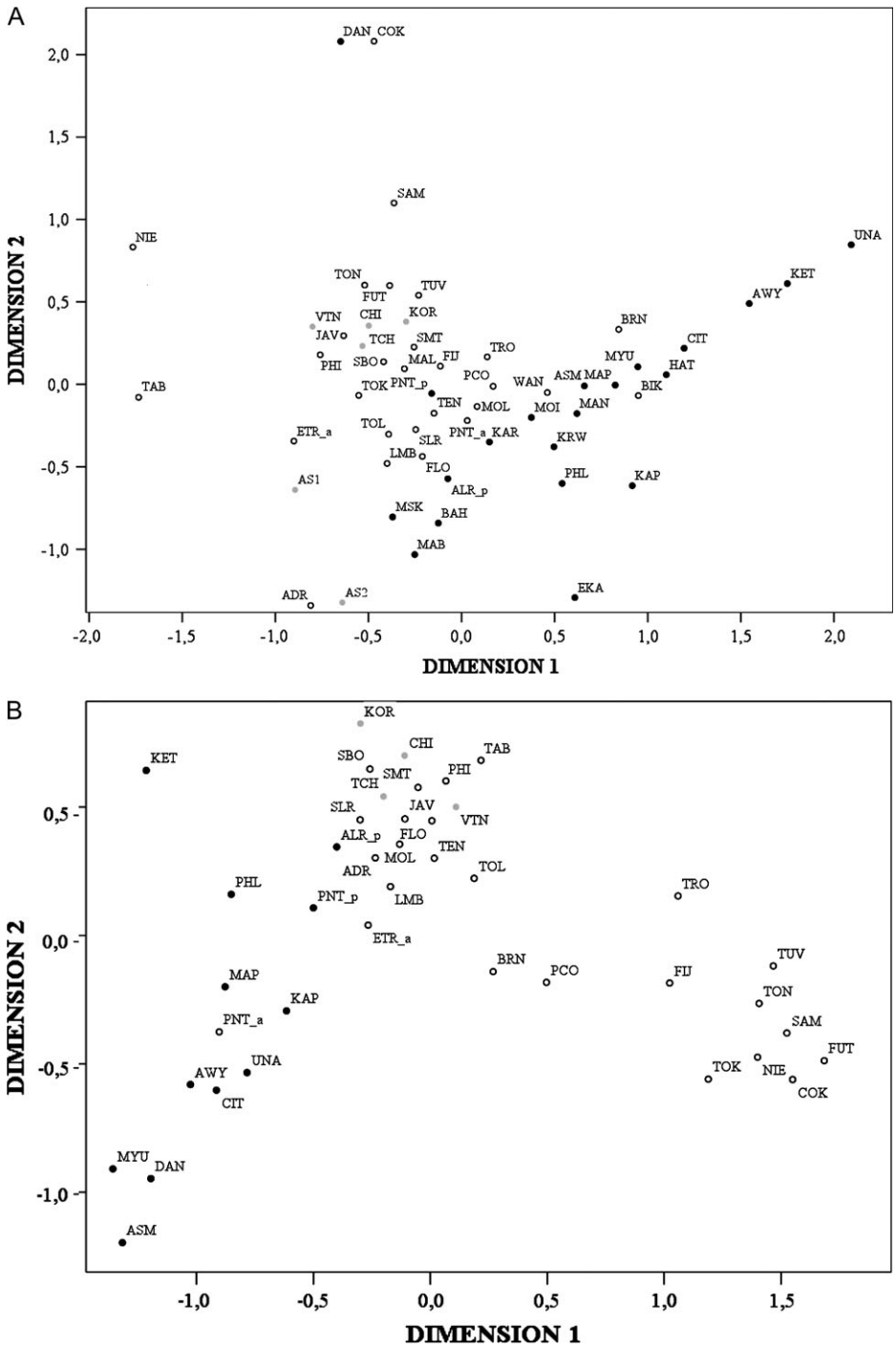

FIG. 3.—Two-dimensional MDS plot based on a F_{st} distance matrix computed from NRY haplogroup frequencies (*a*) and mtDNA sequences (*b*). Stress values: 0.187 and 0.076, respectively. Open white circles: AN-speaking groups; full black circles: NAN (Papuan)-speaking groups; full gray circles: all other populations (irrespective of languages). Groups with more than five samples were not considered. Abbreviations are as following: ADR: Adonara (eastern Indonesia, EI), ALR_p: NAN Alor (EI), ETR_a: AN eastern Timor (EI), FLO: Flores (EI), LMB: Lembata (EI), PNT_p: NAN Pantar (EI), PNT_a: AN Pantar (EI), SLR: Solor (EI). Additional data are from Kayser et al. (2001, 2003, 2006) and additionally from Mona et al. (2007) for NRY data, and from Tommaseo-Ponzetta et al. (2002) for mtDNA data: MOL: Moluccas (pooled from Hiri and Ternate, EI), TEN: Nusa Tenggara (pooled Alor, Flores, Roti, Timur; EI), KAR: Karon (Northwest New Guinea, NWNG), BAH: Baham (NWNG), BIK: Biak (NWNG), EKA: Ekari (NWNG), HAT: Hatam (NWNG), MAB: Maibrat (NWNG), MAN: Mantion (NWNG), MOI: Moi (NWNG), MSK: Moskona (NWNG), WAN: Wandamen (NWNG), IRA: Irarutu (NWNG), ONI: Onin (NWNG), THE: Tehit (NWNG), ASM: Asmat (Southwest New Guinea, SWNG), AWY: Awyu (SWNG), CIT: Citak (SWNG), DAN: Dani/Lani (SWNG), KET: Ketengban (SWNG), KRW: Korowai (SWNG), MAP: Mappi (SWNG), MY: Muyu (SWNG), UNA: Una (SWNG), YAL: Yali (SWNG), KMB: Kombai (SWNG), TRO: Trobriand (Papua New Guinea, PNG), BRN: Bereina (PNG), KAP: Kapuna (PNG), TOL: Tolai New Britain (PNG), PCO: PNG coast, PHL: PNG highlands, FIJ: Fiji, AS1: Australian Aborigines Arnhem Land, AS2: Australian Aborigines Great Sandy Desert, CHI: Han Chinese (East Asia, EA), TAB: Taiwan Aborigines (EA), TCH: Taiwan Han Chinese (EA), VTN: Vietnam (EA), KOR: Korea (EA), MAL: Malay (Southeast Asia, SEA), PHI: Philippines (SEA), SBO: South Borneo (SEA), SMT: Sumatra (SEA), JAV: Java (SEA), COK: Cook Islands (Polynesia, POL), FUT: Futuna (POL), NIE: Niue (POL), TOK: Tokelau (POL), TON: Tonga (POL), TUV: Tuvalu (POL), WES: West Samoa (POL).

Table 4
Analysis of Molecular Variance Performed on NRY and mtDNA Data Dividing the Eastern Indonesian Populations into the Two Linguistic Groups (AN and NAN)

	NRY Haplotypes			NRY Haplogroups			mtDNA Haplotypes		
	Var Components	% of Variation	Fixation Indices	Var Components	% of Variation	Fixation Indices	Var Components	% of Variation	Fixation Indices
AG	0.081	1.4	0.014	−0.001	−0.2	−0.002	-0.037	−1.1	−0.011
APWG	0.620*	10.7	0.108*	0.042*	11.1	0.111*	0.135*	4.0	0.039*
WP	5.113*	88.0	0.120*	0.336*	89.1	0.109*	3.301*	97.1	0.029*

AG = among groups; APWG = among populations within groups; WP = within populations. *indicates a P value < 0.0001 (obtained after 1,000 permutations). See table 1 for AN and NAN groups.

at one level does not preclude a correlation at another level (or vice versa). Similar differences in geographic scales were also observed for northern Island Melanesia, where on the regional scale no correlation between languages and genes were observed, but within one island (New Britain) linguistic and genetic data appeared highly correlated (Hunley et al. 2008). Our results therefore emphasize the importance of carrying out such comparisons at different geographic scales, in order to fully investigate questions about the correlations between genes and languages.

Male-Biased Population Bottleneck in Parts of EI

The NRY data from the Adonara population of EI revealed a strong signal of a population bottleneck with a severely reduced haplogroup ($h = 0.411 \pm 0.061$; mean across eastern Indonesia: $h = 0.786$) as well as haplotype ($h = 0.871 \pm 0.029$; mean: $h = 0.952$) diversity. Additional evidence was provided by the high frequency of NRY haplogroup C-M38* (76%) and the associated Y-STR network analysis (fig. 2*D*). This genetic signature of a population bottleneck was not revealed in the mtDNA analysis, where diversity estimates are in the range of other regional groups (haplotype diversity $h = 0.898 \pm 0.017$; mean: $h = 0.859$). Consequently, these data together allow us to conclude a male-biased bottleneck in the population history of Adonara. Notably, high frequencies of C-M38* (71%) were previously observed in males from Bama on Flores (Lansing et al. 2008) and in some local groups (up to 94%) from Sumba (Lansing et al. 2007). Although no associated Y-STR haplotype diversity, network analysis and mtDNA data were presented by Lansing et al. (2007, 2008), their findings of extremely high frequencies of C-M38* Y-chromosomes on Flores and Sumba may suggest that the male-biased bottleneck involving C-M38*, as clearly revealed here for Adonara, has a somewhat wider geographic distribution in the Nusa Tenggara region, which deserves further investigation.

Paternal and Maternal Lineages Connecting EI with Australia

We found NRY haplogroup C-RPS4Y* in surprisingly high frequency in EI, accounting for 14.2% of all EI Y-chromosomes (table 1), which is the highest frequency of any region in our current and previous data set (supplementary table S1, Supplementary Material online). The RPS4YT mutation characterizing the major haplogroup C represents one of the two oldest branches of the NRY tree in Asia/Oceania (in addition to M9) (supplementary fig. S1, Supplementary Material online) (Underhill et al. 2001). It most likely represents the oldest NRY haplogroup of Asian origin in EI. Haplogroup C-RPS4Y* (excluding the presence of DYS390.1del, and the derived state of M38, M217, and M210) has not been reported in East Asia or Polynesia, is almost completely absent from Melanesia (reported so far from one male from coastal PNG and one Fijian), and is very rare in Southeast Asia (outside EI) (table S1, Supplementary Material online) (Kayser et al. 2001, 2006). However, this NRY lineage was previously found at an appreciable frequency (~10%) in northern Australian Aborigines (Kayser et al. 2001, 2006) and also, albeit less frequently (~1.3%), in central Australian Aborigines (Redd et al. 2002). Extremely high frequencies of C-RPS4Y*(xM38) (92%) were also previously found in males from Cibol on Flores (Lansing et al. 2008), although M217 was not typed and no data on DYS390.1del are reported in this study, their presence is unlikely in an EI population, as demonstrated here.

A possible link between EI and Australia is supported by both the MDS plot (fig. 3*A*) and the network analysis for C-RPS4Y* (fig. 2*E*), which indicates a close relationship between EI and Australian Y-STR haplotypes. The Y-STR DYS390.1 deletion, occurring in complete association with the derived state of RPS4Y (Kayser et al. 2001), was here confirmed to be absent from EI; hence, this deletion remains specific to Australia, where it most likely originated (Kayser et al. 2001). The Y-STR diversity and TMRCA associated with C-RPS4Y* in EI (supplementary table S4, Supplementary Material online), suggests that any link between EI and Australia reflects the pre-Neolithic gene pool of the area, rather than a signature of more recent migration from/to Australia.

C-RPS4Y* was also previously found in India, albeit in low frequency, and was used to support an ancient genetic link between India and Australia (Redd et al. 2002). Our new data on C-RPS4Y* in EI may thus serve as indicator for an ancient migration from Asia via India and EI into Australia. Further support for this scenario comes from two mtDNA haplogroups, M42 and N12, found in EI (albeit at low frequency, table 2), which were previously observed only in Australia (Pierson et al. 2006; Hudjashov et al. 2007): complete mtDNA sequencing of these individuals may shed more light on this question. These NRY and mtDNA haplogroups linking EI and Australia are completely absent (or nearly so) from New Guinea, which

was connected with Australia forming the Sahul continent until about 7–8,000 ybp. Assuming a common origin of Australian and New Guinean Aborigines (Hudjashov et al. 2007) these haplogroups must have been lost by drift in New Guinea; alternatively, the settlement of Sahul from Sunda occurred in several independent waves to New Guinea and to Australia. This latter scenario is supported by genetic differences between New Guinea and Australia, as previously observed with NRY (Kayser et al. 2001) and mtDNA (Redd and Stoneking 1999) data.

Conclusions

Our mtDNA and NRY data suggest a complex genetic history in eastern Indonesia, with components that reflect contributions from AN-speaking migrants from East Asia and NAN-speaking migrants from Melanesia. In addition, we find components that appear to reflect an ancient indigenous eastern Indonesian gene pool that is partly shared with Australia. Thus, genetically, EI is a melting pot. Linguistic data, too, have indicated extensive contact between AN-speaking and NAN-speaking groups in this region (Klamer 2002, 2008). We cannot exclude that eastern Indonesian groups who currently speak AN languages spoke NAN languages in the past, or vice versa, and that some of the Melanesian genetic signals we identified in AN speakers, or East Asian signals we detected in NAN speakers of eastern Indonesia, reflect such complete language shifts. In any case, Asian and Melanesian genetic components appear in all eastern Indonesian groups regardless of their current language, indicating a long history of genetic admixture in the Nusa Tenggara Timur and eastern Timor region of EI.

Overall, although a correlation between genes and languages is often assumed and indeed has been found within single islands (Lansing et al. 2007; Hunley et al. 2008), and at continental (Barbujani and Sokal 1990) as well as global levels (Cavalli-Sforza et al. 1988), exceptions have been reported (Nasidze et al. 2006), including AN and NAN groups from island Melanesia (Hunley et al. 2008). We demonstrated that the region of EI represents a clear example where human migration and admixture history, perhaps together with language shift, have diminished such correlations over time.

Supplementary Material

Supplementary tables S1–S6 and supplementary figures S1–S2 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

Note Added in Proof

Of the nine EI individuals with NRY haplogroup O-M119 additional genotyping of the Y-SNP M110 revealed that from the six eastern Timor ANs five belong to O-M110 (one sample could not be typed due to limited DNA) as well as the single individual from eastern Timor NAN, whereas the single Adonaran and the single Pantar NAN carry O-M119*(xM110). Evidence for a Taiwanese origin of O-M110, which is the most frequent Asian NRY haplogroup in Island Melanesia, is described elsewhere (Kayser, Choi, et al. 2008).

Acknowledgments

We are grateful to all volunteers for having provided samples for this study. Sample collection was carried out under the auspices of the Indonesian Institute of Sciences; we are grateful for this assistance. We are also especially grateful for the collaboration of the Eijkman Institute for Molecular Biology, Jakarta, Indonesia, and for the assistance of several members of the health service in eastern Indonesia and, in Timor L'este, for the assistance of Rev. Dr. Filomeno Jakob Abel, S.J. Additionally, we thank D. Müller and D. Kappei for technical assistance in DNA extraction and genotyping, as well as B. Höffner and A. Weihmann for assistance in DNA sequencing. We are also grateful to G. Reesink, G. Holton, M. Ross, and M. Donohue for advice on linguistic data interpretation and G. Reesink in addition for valuable comments on the manuscript. This study was supported by grants from the British Academy, the Committee for South East Asian Studies of the British Academy, and the Oppenheimer Fund through R.H.B., the Max Planck Society through M.S., and the Erasmus University Medical Center through M.K.

Literature Cited

Adelaar KA, Himmelmann NP. 2005. The Austronesian languages of Asia and Madagascar. London: Routledge/Curzon. London.

Anderson DD. 1997. Cave archaeology in Southeast Asia. Geoarchaeology. 12:607–638.

Bandelt HJ, Forster P, Rohl A. 1999. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol. 16:37–48.

Barbujani G, Sokal RR. 1990. Zones of sharp genetic change in Europe are also linguistic boundaries. Proc Natl Acad Sci USA. 87:1816–1819.

Bellwood P. 1996. The northern Moluccas as a crossroad between Indonesia and the Pacific. International Conference on Linguistics and Culture Relations in East Indonesia, New Guinea and Australia. Gadjah Mada University, Yogyakarta.

Bellwood P. 1997. Prehistory of the Indo-Malayan archipelago. Honululu (HI): University of Hawai'i Press.(revised ed.)

Bellwood P. 2004. The origins and dispersals of agricultural communities in Southeast Asia. In: Glover I, Bellwood P, editors. Southeast Asia: from prehistory to history. London: RoutledgeCurzon. p. 21–40.

Bellwood P, Fox J, Tryon D. 1995. The Austronesians: historical and comparative perspectives. Canberra (Australia): Department of Anthropology (Australian National University).

Blust R. 1993. Central and Central-Eastern Malayo-Polynesian. Ocean Linguist. 32:241–293.

Blust R. 1999. Subgrouping, circularity and extinction: some issues in Austronesian comparative linguistics. Symp Ser Inst Linguist Acad Sinica. 1:31–94.

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, Piazza A, Menozzi P, Mountain J. 1988. Reconstruction of human evolution: bringing together genetic, archaeological, and linguistic data. Proc Natl Acad Sci USA. 85:6002–6006.
Cordaux R, Saha N, Bentley GR, Aunger R, Sirajuddin SM, Stoneking M. 2003. Mitochondrial DNA analysis reveals diverse histories of tribal populations from India. Eur J Hum Genet. 11:253–264.
Diamond J, Bellwood P. 2003. Farmers and their languages: the first expansions. Science. 300:597–603.
Donohue M. 2007. The Papuan language of Tambora. Ocean Linguist. 46:520–537.
Donohue M, Grimes CE. 2008. Yet more on the position of the languages of eastern Indonesia and East Timor. Ocean Linguist. 47:115–159.
Dunn M, Terrill A, Reesink G, Foley RA, Levinson SC. 2005. Structural phylogenetics and the reconstruction of ancient language history. Science. 309:2072–2075.
Excoffier L, Laval G, Schneider S. 2005. Arlequin ver. 3.0: an integrated software package for population genetics data analysis. Evol Bioinform Online. 1:47–50.
Foley WA. 1986. The Papuan languages of New Guinea. Cambridge: Cambridge University Press.
Foley WA. 2000. The language of New Guinea. Annu Rev Anthopol. 29:357–404.
Forster P, Harding R, Torroni A, Bandelt HJ. 1996. Origin and evolution of Native American mtDNA variation: a reappraisal. Am J Hum Genet. 59:935–945.
Forster P, Torroni A, Renfrew C, Rohl A. 2001. Phylogenetic star contraction applied to Asian and Papuan mtDNA evolution. Mol Biol Evol. 18:1864–1881.
Friedlaender JS, Friedlaender FR, Hodgson JA, Stoltz M, Koki G, Horvat G, Zhadanov S, Schurr TG, Merriwether DA. 2007. Melanesian mtDNA Complexity. PLoS ONE. 2:e248.
Friedlaender JS, Friedlaender FR, Reed FA, et al. (11 co-authors). 2008. The genetic structure of Pacific Islanders. PLoS Genet. 4:e19.
Friedlaender J, Schurr T, Gentz F, et al. (12 co-authors). 2005. Expanding Southwest Pacific mitochondrial haplogroups P and Q. Mol Biol Evol. 22:1506–1517.
Gray RD, Drummond AJ, Greenhill SJ. 2009. Language phylogenies reveal expansion pulses and pauses in Pacific settlement. Science. 323:479–483.
Gray RD, Jordan FM. 2000. Language trees support the express-train sequence of Austronesian expansion. Nature. 405:1052–1055.
Groube L, Chappell J, Muke J, Price D. 1986. A 40,000 year-old human occupation site at Huon Peninsula, Papua New Guinea. Nature. 324:453–455.
Hage P, Marck J. 2003. Matrilineality and the Melanesian origin of Polynesian Y chromosomes. Curr Anthropol. 44:121–127.
Handt O, Krings M, Ward RH, Paabo S. 1996. The retrieval of ancient human DNA sequences. Am J Hum Genet. 59:368–376.
Hastings WK. 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika. 57:97–109.
Hill C, Soares P, Mormina M, et al. (10 co-authors). 2007. A mitochondrial stratigraphy for island southeast Asia. Am J Hum Genet. 80:29–43.
Hudjashov G, Kivisild T, Underhill PA, et al. (10 co-authors). 2007. Revealing the prehistoric settlement of Australia by Y chromosome and mtDNA analysis. Proc Natl Acad Sci USA. 104:8726–8730.
Hunley K, Dunn M, Lindstrom E, Reesink G, Terrill A, Healy ME, Koki G, Friedlaender FR, Friedlaender JS. 2008. Genetic and linguistic coevolution in Northern Island Melanesia. PLoS Genet. 4:e1000239.
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.
Kayser M, Brauer S, Cordaux R, Casto A, et al. (14 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, Schiefenhovel 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, Schiefenhovel 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, Schiefenhovel 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, Suarkia D, Schiefenhovel 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.
Kayser M, Lao O, Stoneking M. 2008. Reply to Hedrick "East Asian and Melanesian Ancestry in Polynesians. Am J Hum Genet. 83:140–142.
Kirch PV. 1997. The Lapita peoples: ancestors of the oceanic world. Oxford: Blackwell.
Kivisild T, Tolk HV, Parik J, Wang Y, Papiha SS, Bandelt HJ, Villems R. 2002. The emerging limbs and twigs of the East Asian mtDNA tree. Mol Biol Evol. 19:1737–1751.
Klamer M. 2002. Typical features of Austronesian languages in Central/Eastern Indonesia. Ocean Linguist. 41:363–383.
Klamer M, Reesink G, Van Staden M. 2008. East Nusantara as a linguistic area. In: Muysken P, editor. From linguistic areas to areal linguistics. Amsterdam (the Netherlands): John Benjamins. p. 95–149.
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 USA. 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 USA. 105:11645–11650.
Lum JK, Cann RL. 1998. mtDNA and language support a common origin of Micronesians and Polynesians in Island Southeast Asia. Am J Phys Anthropol. 105:109–119.
Lum JK, Rickards O, Ching C, Cann RL. 1994. Polynesian mitochondrial DNAs reveal three deep maternal lineage clusters. Hum Biol. 66:567–590.
Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. 1953. Equations of state calculations by fast computing machine. J Chem Phys. 21:1087–1091.
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.
Moodley Y, Linz B, Yamaoka Y, et al. (15 co-authors). 2009. The peopling of the Pacific from a bacterial perspective. Science. 323:527–530.
Nasidze I, Quinque D, Rahmani M, Alemohamad SA, Stoneking M. 2006. Concomitant replacement of language

and mtDNA in South Caspian populations of Iran. Curr Biol. 16:668–673.
O'Connor S, Spriggs M, Veth P. 2002. Excavations at Lene Hara Cave establishes occupation in East Timor at least 30,000 to 35,000 years ago. Antiquity. 76:45–50.
Ohta T, Kimura M. 1973. A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. Genet Res. 22:201–204.
Oppenheimer SJ, Richards M. 2001. Polynesian origins. Slow boat to Melanesia? Nature. 410:166–167.
Pawley A. 1998. The Trans New Guinea phylum hypothesis: a reassessment. In: Miedema J, Odé C, Dam RAC, editors. Perspectives on the bird's head of Irian Jaya, Indonesia. Amsterdam (the Netherlands): Editions Rodopi. p. 655–689.
Pawley A. 1999. Chaising rainbows: implications of the rapid dispersal of Austronesian languages for subgrouping and reconstruction. In: Zeitoun E, Li PJ-K, editors. Selected papers from the Eighth International Conference on Austronesian Linguistics. Taipei (Taiwan): Academia Sinica. p. 95–138.
Pawley A. 2005. The chequered career of the Trans New Guina hypothesis: recent research and its implications. In: Pawley A, Attenborough R, Golson J, Hide R, editors. Papuan pasts: cultural, linguistic and biological histories of Papuan-speaking peoples. Canberra (Australia): Pacific Linguistics, Research School of Pacific and Asian Studies, The Australian National University. p. 67–107.
Pierson MJ, Martinez-Arias R, Holland BR, Gemmell NJ, Hurles ME, Penny D. 2006. Deciphering past human population movements in Oceania: provably optimal trees of 127 mtDNA genomes. Mol Biol Evol. 23:1966–1975.
Rambaut A, Drummond AJ. 2004. TRACER. Version 1.3. Available from: http://beast.bio.ed.ac.uk/Tracer.
Redd AJ, Roberts-Thomson J, Karafet T, Bamshad M, Jorde LB, Naidu JM, Walsh B, Hammer MF. 2002. Gene flow from the Indian subcontinent to Australia: evidence from the Y chromosome. Curr Biol. 12:673–677.
Redd AJ, Stoneking M. 1999. Peopling of Sahul: mtDNA variation in aboriginal Australian and Papua New Guinean populations. Am J Hum Genet. 65:808–828.
Redd AJ, Takezaki N, Sherry ST, McGarvey ST, Sofro AS, Stoneking M. 1995. Evolutionary history of the COII/tRNALys intergenic 9 base pair deletion in human mitochondrial DNAs from the Pacific. Mol Biol Evol. 12:604–615.
Richards M, Oppenheimer S, Sykes B. 1998. mtDNA suggests Polynesian origins in Eastern Indonesia. Am J Hum Genet. 63:1234–1236.
Roberts RG, Jones R, Smith MA. 1990. Thermoluminescence dating of a 50,000-year-old human occupation site in northern Australia. Nature. 345:153–156.
Ross M. 2005. Pronouns as a preliminary diagnostic of grouping Papuan languages. In: Pawley A, Attenborough R, Golson J, Hide R, editors. Papuan pasts: cultural, linguistic and biological histories of Papuan-speaking peoples. Canberra (Australia): Pacific Linguistics. p. 15–65.
Saillard J, Forster P, Lynnerup N, Bandelt HJ, Norby S. 2000. mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. Am J Hum Genet. 67:718–726.
Specht J. 2005. Revisiting the Bismarcks: some alternative views. In: Pawley A, Attenborough R, Golson J, Hide R, editors. Papuan pasts: cultural, linguistic and biological histories of Papuan-speaking peoples. Canberra (Australia): Pacific Linguistics. p. 235–288.
Spriggs M. 2000. Out of Asia? The spread of Southeast Asian Pleistocene and Neolithic maritime cultures in Island Southeast Asia and the western Pacific. In: O'Connor S, Veth P, editors. East of Wallace's line: studies of past and present maritime cultures of the Indo-Pacific region. Rotterdam (the Netherlands): A.A. Balkema. p. 51–75.
Spriggs M. 2003. Chronology of the Neolithic transition in Southeast Asia and the Western Pacific: a view from 2003. Rev Archaeol. 24:57–80.
Tommaseo-Ponzetta M, Attimonelli M, De Robertis M, Tanzariello F, Saccone C. 2002. Mitochondrial DNA variability of West New Guinea populations. Am J Phys Anthropol. 117:49–67.
Trejaut JA, Kivisild T, Loo JH, Lee CL, He CL, Hsu CJ, Li ZY, Lin M. 2005. Traces of archaic mitochondrial lineages persist in Austronesian-speaking Formosan populations. PLoS Biol. 3:e247.
Underhill PA, Passarino G, Lin AA, Shen P, Mirazon Lahr M, Foley RA, Oefner PJ, Cavalli-Sforza LL. 2001. The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. Ann Hum Genet. 65:43–62.
Wilson IJ, Balding DJ. 1998. Genealogical inference from microsatellite data. Genetics. 150:499–510.
Wilson IJ, Weale ME, Balding DJ. 2003. Inferences from DNA data: population histories, evolutionary processes and forensic match probabilities. J R Stat Soc Ser A (Stat Soc). 166:155–188.
Wurm SA, Hattori S. 1981. Language Atlas of the Pacific Area. Canberra (Australia): Australian Academy of the Humanities in collaboration with the Japan Academy.
Yao YG, Kong QP, Bandelt HJ, Kivisild T, Zhang YP. 2002. Phylogeographic differentiation of mitochondrial DNA in Han Chinese. Am J Hum Genet. 70:635–651.

Naruya Saitou, Associate Editor

Accepted April 28, 2009