MOLECULAR BIOLOGY AND EVOLUTION

On the Origins and Admixture of Malagasy: New Evidence from High-Resolution Analyses of Paternal and Maternal Lineages

Sergio Tofanelli,[1] *Stefania Bertoncini,*[1] *Loredana Castrì,† Donata Luiselli,† Francesc Calafell,‡ Giuseppe Donati,§ and Giorgio Paoli**

*Dipartimento di Biologia, Unità di Antropologia, Università di Pisa, Pisa, Italy; †Dipartimento di Biologia Evoluzionistica Sperimentale, Università di Bologna, Bologna, Italy; ‡Department de Ciències Experimentals i de la Salut (CEXS), Unitat de Biologia Evolutiva, Universitat Pompeu Fabra, Barcelona, Spain; and §Department of Anthropology and Geography, Oxford Brookes University, Oxford, United Kingdom

The Malagasy have been shown to be a genetically admixed population combining parental lineages with African and South East Asian ancestry. In the present paper, we fit the Malagasy admixture history in a highly resolved phylogeographic framework by typing a large set of mitochondrial DNA and Y DNA markers in unrelated individuals from inland (Merina) and coastal (Antandroy, Antanosy, and Antaisaka) ethnic groups. This allowed performance of a multilevel analysis in which the diversity among main ethnic divisions, lineage ancestries, and modes of inheritance could be concurrently evaluated. Admixture was confirmed to result from the encounter of African and Southeast Asian people with minor recent male contributions from Europe. However, new scenarios are depicted about Malagasy admixture history. The distribution of ancestral components was ethnic and sex biased, with the Asian ancestry appearing more conserved in the female than in the male gene pool and in inland than in coastal groups. A statistic based on haplotype sharing (D_{HS}), showing low sampling error and time linearity over the last 200 generations, was introduced here for the first time and helped to integrate our results with linguistic and archeological data. The focus about the origin of Malagasy lineages was enlarged in space and pushed back in time. Homelands could not be pinpointed but appeared to comprise two vast areas containing different populations from sub-Saharan Africa and South East Asia. The pattern of diffusion of uniparental lineages was compatible with at least two events: a primary admixture of proto-Malay people with Bantu speakers bearing a western-like pool of haplotypes, followed by a secondary flow of Southeastern Bantu speakers unpaired for gender (mainly male driven) and geography (mainly coastal).

Introduction

Most interpretations of archeological and linguistic data support the hypothesis that the island of Madagascar, located in the Indian Ocean, was permanently settled by human groups not earlier than the sixth century AD (Dahl 1951, 1991; Dewar and Wright 1993; Adelaar 1995a). However, drastic biotic changes (i.e., the turning of the vegetation coverage from rain forest to savannah, faunal extinctions, and a sudden increase of charcoal remains) have been inferred since 2,300 yBP (Burney 1987; MacPhee and Burney 1991; Gasse and Van Campo 1998; Burney et al. 2003; Perez et al. 2003) and attributed to human activities.

The present population, known by the general term "Malagasy," is considered an admixed population as it shows a combination of morphological and cultural traits typical of Bantu and Austronesian speakers. Such a combination is present at different degrees in the main subgroups into which Malagasy ethnic diversity is generally classified: Highlanders (HLs) and *Côtiers* (CTs) (Blench 2006, 2007). HLs (the Merina, Betsileo, Sihanaka, and Bezanozano groups) are settled in the central plateaus and are considered to be the most "Asian" group based on light skin, straight black hair, and a rice-based economy. CTs (the Sakalava, Mahafaly, Antanosy, Antandroy, and Antaisaka groups, among others) are coastal dwellers described as being more "African" in physical appearance (darker complexion, curly hair, and prognathic jaws) and in some features of the material culture. A common link across groups is the Malagasy language, which is spoken throughout the island. It belongs to the West Malayo-Polynesian (WMP) branch of the Austronesian family and ~90% of its basic vocabulary has been found to be shared with Maanyan, a language from the region of the Barito River in southeastern Borneo (Dahl 1951; Adelaar 1995b). The remaining 10% of the lexicon presents Bantu, Malay, South Sulawesian, and Javanese borrowings and a small number of Sanskrit loanwords. Some linguists (Adelaar 1995a; Blench 2007) have suggested that people from Southeast Barito would have been brought there as subordinates (slaves, ship crew, and workers) by Malays, at the time of the maximum expansion of the Srivijaya empire (sixth to seventh century AD), when they dominated Indonesia and controlled trade networks across the Indian Ocean. The contribution of Bantu language to Malagasy is mainly from the Sabaki vocabulary, currently spoken North of the Zambesi river (Blench 2007).

It is not known whether the Malagasy founder population came into the island already admixed or admixture was an *in situ* process. Nonetheless, the combined evidence from archeology and linguistics seems to support the theory of Deschamps (1960) that the East African coast may have been visited by Austronesian mariners from an early period, before they definitely settled in Madagascar (Blench 1996; Adelaar 2009). Evidence of recent genetic introgression from pirates, traders, slaves, captives, and colonists of different origins (African, Indian, Arabian, Portuguese, French, British, and Dutch) is historically documented for both Malagasy groups (Kent 1962, 1970; Dewar and Wright 1993).

In the most extensive study published so far on Malagasy evolutionary genetics, Hurles et al. (2005) detected a balanced contribution of lineages with African and Southeast Asian ancestry in both the Y chromosome and the

[1] Joint authorship: Sergio Tofanelli and Stefania Bertoncini contributed equally to the paper.

Key words: Malagasy, mtDNA, Y chromosome, admixture.

E-mail: stofanelli@biologia.unipi.it.

Mol. Biol. Evol. 26(9):2109–2124. 2009
doi:10.1093/molbev/msp120
Advance Access publication June 17, 2009

http://mbe.oxfordjournals.org/

mitochondrial genome. Pairwise F_{ST} distances calculated upon Y haplogroup frequencies suggested southern Borneo as the most likely place of origin of Asian founders, providing genetic support to the linguistic evidence. Moreover, diversity values suggested a smaller migration from Africa than from Asia and traces of a recent introgression were found on Y chromosomes. However, genotyped samples were small (mitochondrial DNA, mtDNA $N = 37$, Y DNA $N = 35$) and limited to HLs. Larger and more ethnically comprehensive samples are needed to obtain a much more reliable picture of Madagascan genetic structure and history.

In order to address the aforementioned issue, we increased Hurles et al.'s phylogenetic and geographic resolution by typing a large set of binary and multistate markers at Y chromosome (14 unique event polymorphisms [UEPs], 17 short tandem repeats [STRs]) and mitochondrial genome (19 Single Nucleotide Polymorphisms, SNPs; HVS-I sequences) in a total of 133 unrelated individuals from one HL (Merina) and three CT (Antandroy, Antanosy, and Antaisaka) Malagasy groups. Y-STRs were typed here for the first time. We also introduced a new time-linear statistical approach (D_{HS}-based simulations) to reconstruct admixture dynamics, performed extensive computer simulations under different evolutionary models, enriched haplotype reference databases, and reassigned Hurles' HVS-I sequences into L, M, and N subhaplogroups. This allowed us to frame more precisely the time and place of origin of the different genetic components and to pool novel and already published data as to keep separate comparisons between African and Asian components, between HLs and CTs, and between maternally and paternally inherited markers.

Materials and Methods

Subjects

A map with the sampling location and the distribution of Malagasy ethnic groups is given in figure 1. Samples were taken in private clinics around Taoloñaro (Fort Dauphin) from unrelated blood donors who gave their informed consent to project aims and data treatment. Ethnic affiliation was established by self-assignment. The individuals sampled ($N = 133$) were from the Highland Merina ($N = 9$) and the CT Antandroy ($N = 59$), Antanosy ($N = 54$), and Antaisaka ($N = 11$) groups. Merina, by far the largest ethnic group of the island, have preserved Austronesian-like traits by discouraging intermarriages with African-looking peoples across a three-level caste system. Antandroy, known as "those of thorns," live in the far southern dry forests and are seminomadic groups of cattle breeders (African zebus) with uncertain origin. Antanosy, or "people from the island," descend from a group settled on the southern coasts from a little island off Taoloñaro. Phenotypic traits, language, and other cultural features are fairly heterogeneous. Antaisaka claim a direct descent from a founding father of Sakalawa origin, a group from the west coast with strong African features (ethnic data were taken from Schraeder 1995).

When analyses between HL and CT subgroups were performed, mtDNA data from highland populations published in or recalculated from the paper of Hurles et al. (2005) were combined with our original Merina data ($N = 9$) into the HL group for a total highland sample size of 46. This was not possible for Y-STR data due to the absence of Y microsatellites in previous studies.

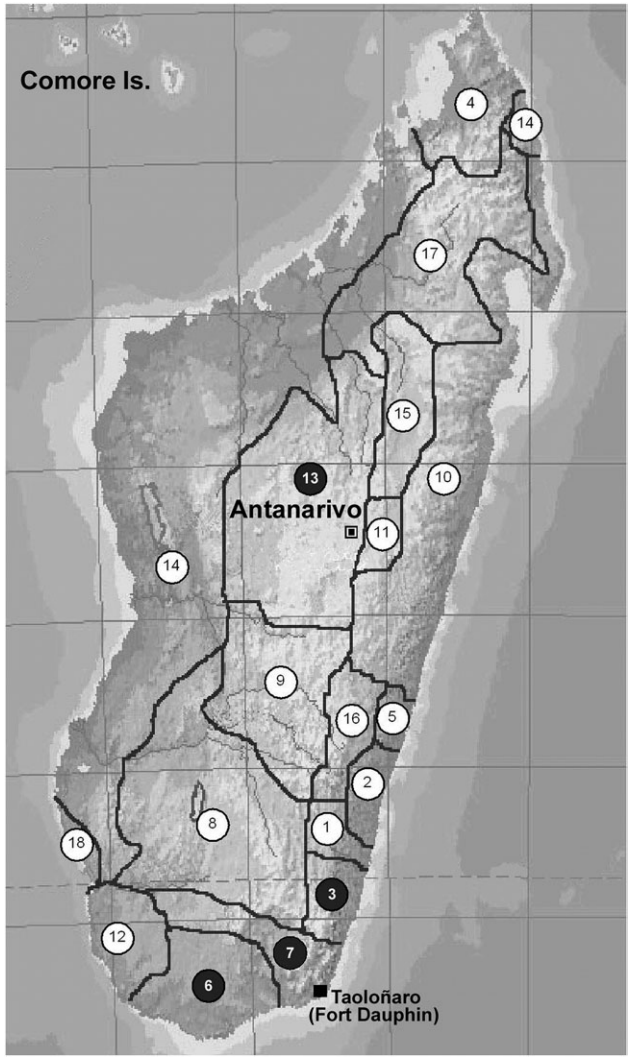

FIG. 1.—Geographic distribution of the 18 Malagasy ethnic groups: 1 (Antaifasy), 2 (Antaimoro), **3 (Antaisaka)**, 4 (Antankarana), 5 (Antambahoaka), **6 (Antandroy)**, **7 (Antanosy)**, 8 (Bara), 9 (Betsileo), 10 (Betsimisaraka), 11 (Bezanozano), 12 (Mahafaly), **13 (Merina)**, 14 (Sakalava), 15 (Sihanaka), 16 (Tanala), 17 (Tsimihety), and 18 (Vezo).

DNA Analyses

Genomic DNA was extracted from dried bloodspots with the DNATM IQ System kit (Promega Corporation). MtDNA hypervariable region I (HVS-I) was amplified using primers L15996 and H16401 (Vigilant et al. 1989) and the polymerase chain reaction (PCR) products purified with Exo-SAP. Sequencing reactions were performed for each strand (using primers L15996 and H16401) with the ABI PRISM BigDye Terminator v1.1 Cycle Sequencing kit (Applied Biosystems) according to supplier's recommendations. All sequences have been deposited in GenBank (accession numbers EU336804–EU336936). Mutations at nps

16182 and 16183 were ignored in interpopulation analyses because they either represent fast-mutating, often heteroplasmic length polymorphisms, or they are sequencing artefacts caused by the long poly-cytosine stretch found in this position. Moreover, they may not always be included in previous reports.

Nineteen SNPs from the mtDNA-coding region, defining 17 haplogroups, were typed by restriction fragment length polymorphism analysis following a hierarchical approach: L1/L2 (+3592 *Hpa*I), L0a (−4310 *Alu*I), L2 (+16389 *Hin*fI), L2a (+13803 *Hae*III), L2b (+4157 *Alu*I), L2c (−13957 *Hae*III), L3 (−3592 *Hpa*I, +10394 *Dde*I, −10871 *Mnl*I), L3b (+10084 *Taq*I), L3e (+2349 *Mbo*I); M (+10397 *Alu*I, +10394 *Dde*I), E (−7598 *Hha*I), D (−5176 *Alu*I), G (+4831 *Hha*I), N (+10871 *Mnl*I, −10397 *Dde*I, −10394 *Dde*I), B (COII/tRNAlys 9-bp deletion), R9 (+12406 *Hin*cII), and F3b (+10319 *Tsp*509I).

A two-step protocol was used to assign each mtDNA molecule to haplogroups: first, the combination of HVS-I sequences and the literature (Kivisild et al. 2002, 2004; Metspalu et al. 2004; Salas et al. 2002, 2004; Beleza et al. 2005; Hurles et al. 2005; Trejaut et al. 2005; Hill et al. 2007) was taken into account to classify mtDNAs into haplogroups and subhaplogroups; then, the 19 SNPs were used to refine the classification. The mitochondrial nomenclature was according to Salas et al. (2002, 2004), Kivisild et al. (2004), Trejaut et al. (2005), and Van Oven and Kayser (2009). Unbiased comparisons with Hurles' data were obtained by reassigning HVS-I sequences to L, M, and N sub-haplogroups according to the results of the assignment method described above. In particular (supplementary table S1, Supplementary Material online): haplogroup mutation motifs L2a1b, L3b1, L3e1a, M(xM7), M7c1c, B4a1a1, E1a, R9, F3b could be clearly identified; two haplotypes within the L* clade (mutations 16223, 16265T and 16209, 16223, 16311) remained unassigned into sub-haplogroups.

A subset of 110 DNAs was amplified at 17 Y-STR loci (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS385a, DYS385b, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, GATA C4, and GATA H4) with the "AmpFlSTRY-filer" kit (Applied Biosystems). The length of PCR-amplified fragments was evaluated with the ABI PRISM 310 sequencer (Applera) using allelic ladders with known sequence and the alleles were assigned with the Genotyper 3.7 software. Alleles at the Y DYS389II locus were counted as differences between DYS389II and DYS389I alleles. The haplotype duplicated at DYS385 (MAD30) was considered as having a single 12–14 allele at the DYS385b locus.

A prior assignment of STR haplotypes into candidate Y haplogroups (named according to Karafet et al. 2008) was performed following different strategies. They included: haplotype matching at the Y-STR Haplotype Reference Database (YHRD) (release 27: 15,956 17-locus haplotypes; 70,286 9-locus haplotypes, http://www.yhrd.org), at the Y Chromosome Consortium database (76 12-locus haplotypes, http://www.rootsweb.com) and at a manually edited archive of published and unpublished data (51,795 9-locus haplotypes); motif matching (the presence of the sub-Saharian specific GATA C4*17 allele); distance-based (Cavalli-Sforza and Edwards 1967) and Bayesian (Rannala and Mountain 1997; Athey 2006) methods of assignment. Bayesian algorithms were applied using the GeneClass 2.0 software package (Institut National de la Recherche Agronomique, http://www.inra.fr/; Piry et al. 2004) and the interface of the Haplogroup Predictor web page (http://home.comcast.net/~hapest5/index.html). Subsequently, only 14 UEPs were typed in singleplex reactions to check the assignment to candidate haplogroups and all predictions were correct. SNP analyses were performed with the SNaPshot ddNTP Primer Extension Kit (Applied Biosystems), using the primers reported in supplementary table S2, Supplementary Material online. Amplification products were subsequently purified using Exo-SAP. The indel YAP was typed according to Hammer (1994) and visualized on 2% low-melting agarose gels.

Calculation of Admixture Proportions

Paleontological (Bowler et al. 2003; Morwood et al. 2004; Mellars 2006) and mitochondrial genetic evidence (Metspalu et al. 2004) would suggest that one of the initial colonizations of Eurasia followed a "southern coastal route" and started around 60–90 kyBP. It could then be argued that the Malagasy parental gene pools have been shaped across at least 60,000 years of reproductive isolation. This fact, coupled with the appropriate level of resolution and phylogeographic informativeness of the chosen set of DNA markers, has driven to the mutual exclusivity of the lineages with African and Indonesian ancestry observed in both the Y and the mitochondrial genomes. Thus, the proportion of the two geographic/linguistic components in HL and CT gene pools was assessed by lineage counting, and the amount of admixture was derived from the relative proportion among lineages. Mitochondrial haplogroups of the L* group and Y haplogroups E1b1a, B2*, E2b were considered of African/Bantu ancestry. Mitochondrial haplogroups M7c1c, M(xM7), F3b, R9, B4a1a1, B4a, E1a and Y haplogroups O1a and O2a were considered of Indonesian/Austronesian ancestry. New statistical inferences (see the Results section) induced us to exclude the Indonesian origin of R1a1 and J2 chromosomes and place them into a heterogeneous clade with alleged Eurasian ancestry together with L*, E1b1b1a, and R1b1 chromosomes.

Place of Origin Analyses

The place of origin was assigned either for mtDNA or YDNA lineages on the basis of multistate data. The use of fast-evolving markers prevented heterogeneity of published SNP data sets from resulting in unbalanced or low-resolved comparisons.

A new statistic, D_{HS}, was introduced to assess ancestry that exploit two similar indexes of genetic similarity following the $\hat{S}$ estimator proposed by Nei and Li (1979) for restriction maps:

1) $D_{HS}=(1-S_n)(1-S_h)$, where
2) $S_n=\{n_x+n_y\}/\{N_x+N_y\}$,with n_x and n_y the absolute frequencies of the chromosomes carrying the haplotypes shared by populations X (Malagasy groups) and Y (reference samples), and N_x, N_y their sample sizes;

Table 1
Performance of D_{HS} and Weir and Cockerman's F_{ST} Statistics in Forward Computer Simulation (200 Iterations) under an Extended Wright–Fisher Model with Varying Priors of Haplotype Diversity (H_0) and Number of Migrants (N_em)

		Multistate Markers		Binary Markers	
		D_{HS}	F_{ST}	D_{HS}	F_{ST}
$H_0 = 0.000$	b	3.666×10^{-3}	0.162×10^{-3}	2.277×10^{-3}	0.185×10^{-3}
	R^2	0.992	0.995	0.989	0.997
	CV	18.7	40.2	16.9	34.4
$H_0 = 0.815$	b	4.671×10^{-3}	0.222×10^{-3}	2.721×10^{-3}	0.371×10^{-3}
	R^2	0.983	0.982	0.996	0.997
	CV	10.4	60.3	15.1	83.2
$H_0 = 0.997$	b	2.884×10^{-3}	0.234×10^{-3}	4.539×10^{-3}	0.352×10^{-3}
	R^2	0.613	0.978	0.986	0.997
	CV	5.0	34.7	15.3	43.9
		Multistate markers		Binary markers	
		D_{HS}	F_{ST}	D_{HS}	F_{ST}
$N_em = 10$	b	3.376×10^{-3}	1.191×10^{-3}	1.202×10^{-3}	0.095×10^{-3}
	R^2	0.562	0.422	0.534	0.443
	CV	11.1	38.4	13.2	45.1
$N_em = 100$	b	7.889×10^{-3}	1.818×10^{-3}	3.575×10^{-3}	1.433×10^{-3}
	R^2	0.875	0.928	0.832	0.909
	CV	14.5	57.6	21.8	53.3
$N_em = 1{,}000$	b	5.944×10^{-3}	0.491×10^{-3}	2.601×10^{-3}	0.518×10^{-3}
	R^2	0.998	0.993	0.992	0.970
	CV	13.6	69.9	17.6	76.3

Two populations were assumed evolving in reproductive isolation and constant size for 200 generations after divergence from a source population at generation t_0. Simulated multistate markers were 9-locus STR haplotypes evolving according to a strict stepwise mutation model (SMM) ($\mu = 1.85 \times 10^{-3}$ mut/locus/gen, Gusmão et al. 2005). Simulated binary markers were 360 D-loop sites evolving under an infinite allele model (IAM) ($\mu = 9{,}5 \times 10^{-6}$ mut/locus/gen, Howell et al. 2003). b = slope of the regression line, R^2 = Pearson's regression coefficient, CV = mean coefficient of variation calculated as $\sqrt{(\sum \mathrm{CV}^2/k)} * 100$ for $k = 200$ generations.

3) $S_h = 2h_{xy}/\{H_x + H_y\}$, with h_{xy} the number of different haplotypes shared by the two populations and H_x, H_y the total number of different haplotypes, respectively, in population X (Malagasy groups) and Y (reference samples).

The above distance varies from 0 (all haplotypes shared by the two populations) to 1 (no shared haplotypes). Its efficiency has been evaluated against the F_{ST} distance (Weir and Cockerham 1984; Michalakis and Excoffier 1996) by means of forward computer simulations under different evolutionary scenarios (table 1) with the Markov chain Monte Carlo method implemented in the program ASHEs (http://ashes.codeplex.com). The extent of the sampling error (measured as Coefficient of Variation or CV), and the linear relationship with time (expressed as both b, the slope of the regression line and R^2, Pearson's coefficient of regression) have been used as performance criteria. Each computer simulation (200 iterations) modeled the increase rate over 200 generations of averaged D_{HS} and F_{ST} values between diverging populations each evolving under reproductive isolation and constant size (Wright–Fisher model). The impact on the two statistics of either the heterogeneity (H) or the effective size (N_e) in the populations was evaluated by simulating different values of H_0 (the haplotype diversity of the ancestral pool of chromosomes whose N_e was set to 5,000) after an initial even split (table 1), and a different number of migrants (N_em = effective size*migration rate) from a parental group with $H_0 = 0.815$ and $N_e = 5{,}000$ (fig. 2). Whatever the type of character considered (360 HVS-I sites or 9-locus Y-STRs), D_{HS} performed much better than F_{ST}, being more linear with time (from 2 to 23 times higher b values) and having much lower variance (from 2 to 9 times lower CV values). D_{HS} curves tended to saturation after the first 20–40 generations only in the case of a high initial level of diversity ($H_0 > 0.997$) at multistate haplotypes or of a marked founder effect (N_em around 10).

D_{HS} was calculated for Malagasy HVS-I and YSTR 9-locus haplotypes of both Asian and African ancestries searching, respectively, against 8,007 and 6,455 entries with known sub-Saharan African and Southeast Asian ancestries (reference in tables 4–7).

Estimation of the Time Since the Admixture Event (TSAE)

Under the assumption that the shorter the mean coalescence time of pairs of haplotypes, the higher the number of exact matchings, D_{HS} statistics gives a tool to estimate the TSAE. Thus, we simulated under different evolutionary models the variation of the D_{HS} statistics concurrently with the variation of haplotype diversity in a parental (H_x) and a migrant (H_y) population diverged at time t_0 from a common pool of haplotypes. To take into account the effects of demographic dynamics occurred since the admixture, realistic prior parameters were set: Haplotype diversity of the source pool (H_0) was calculated as the arithmetic mean between observed H_x and H_y levels; one-sixth of the averaged present-day growth rates in Madagascar, East Africa, and SE Asia (respectively, 0.030, 0.019, 0.015; CIA World Factbook 2009, http://www.cia.gov/library/publications/the-world-factbook/) as increment rate (w). For each model, a generation interval was accepted

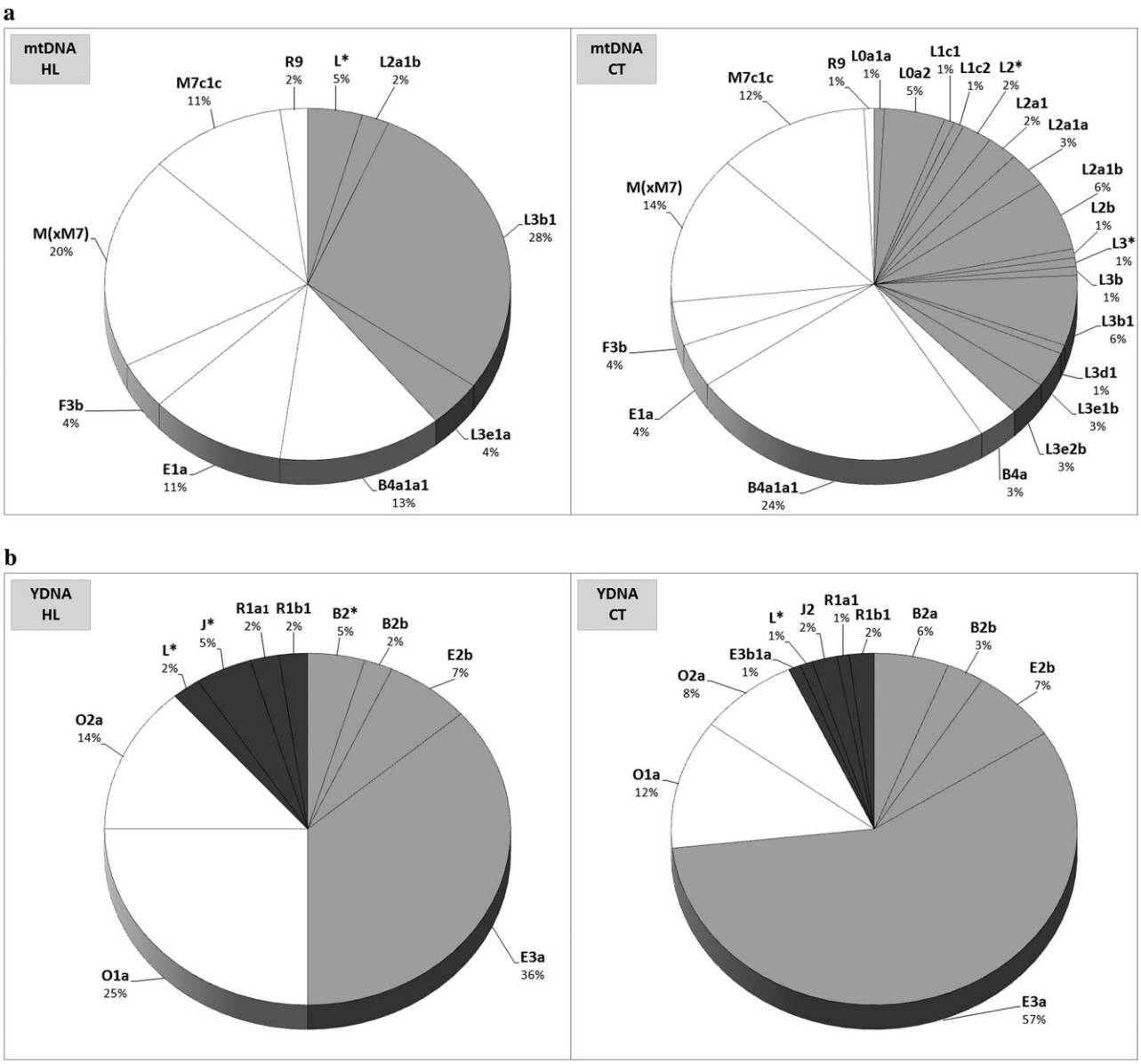

FIG. 2.—MtDNA (*a*) and Y (*b*) haplogroup frequencies in HLs and CTs. In white, Indonesian-derived haplogroups; in light gray, African-derived haplogroups.

as the most likely TSAE when all the observed values of D_{HS}, H_x, and H_y fell within the tolerance interval (95% confidence interval, CI) of the simulated distribution. It should be taken into account that the lack of the "correct" source populations could bias these estimates toward higher D_{HS} values and consequent earlier dates.

Network Analysis

A median-joining network connecting the 17-locus haplotypes of all Malagasy E1b1a chromosomes was constructed with the NETWORK 4.2.0.1 software (Fluxus Technology, http://www.fluxus-engineering.com), weighting each STR locus according to Bosch et al. (2006): The weight of the *i*th STR was calculated as $10*V_m/V_i$, where V_m is the mean variance of all STRs and V_i is the variance of the *i*th STR. We marked those nodes containing allele series that equally (W/E) or preferentially matched (see supplementary table S3, Supplementary Material online) with mainland Africans from western-central (W) or southeastern (E) regions.

Other Statistical Analyses

Indexes of population genetic structure (Analysis of Molecular Variance or AMOVA), pairwise F_{ST} distances, Nei's diversity index (*H*), and the mean number of pairwise differences (MPD) were computed using the ARLEQUIN package ver 3.1 (http://cmpgunibech/software/arlequin3, Excoffier et al. 2005). Differences between distributions of *H* values were evaluated by a *t*-test according to Nei (1987). The weighted intralineage mean pairwise difference (WIMP), which measures the mean within-haplogroup

Table 2
Mitochondrial and Y Chromosome Diversity

	Population	Reference	*N*	Group	MPD	*H*		WIMP
						SNPs	HVS-I	
mtDNA	Antandroy	This research	59		IN 6.54 ± 3.13	IN 0.77 ± 0.03	IN 0.81 ± 0.03	IN 0.873
	Antanosy	This research	54	CT	AF 7.83 ± 3.71	AF 0.91 ± 0.02	AF 0.96 ± 0.01	AF 4.710
	Antaisaka	This research	11					
	Merina	This research	9	HL	IN 5.36 ± 2.67	IN 0.81 ± 0.04	IN 0.86 ± 0.04	IN 0.710
	Bezanozano	Hurles et al. (2005)	37		AF 2.83 ± 1.56	AF 0.48 ± 0.13	AF 0.69 ± 0.11	AF 1.765
	Betsileo							
	Merina							
	Sihanaka							

	Population	Reference	*N*	Group	MPD	*H*		WIMP
						UEPs	STR	
Y chromosome	Antandroy	This research	46	CT	IN 9.01 ± 4.33	IN 0.51 ± 0.06	IN 0.99 ± 0.02	IN 6.094
	Antanosy	This research	47		AF 8.51 ± 3.98	AF 0.39 ± 0.07	AF 0.99 ± 0.00	AF 6.116
	Antaisaka	This research	8					
	Merina	This research	9	HL	IN 6.10 ± 3.49		IN 0.90 ± 0.16	IN 6.100
					AF 8.17 ± 4.81	IN 0.49 ± 0.08	AF 1.00 ± 0.18	AF 8.166
	Bezanozano	Hurles et al. (2005)	35			AF 0.46 ± 0.12		
	Betsileo							
	Merina							
	Sihanaka							

HL—Highland groups, CT—coastal groups, *H*—Nei's diversity, WIMP—Weighted mean Intralineage Mean Pairwise difference. IN, Indonesian lineages, AF, African lineages.

diversity, was calculated as described (Hurles et al. 2002). The statistical significance of contingency tables was tested by the χ^2 Fisher's exact test using the STATISTICA 6.0 software package (StatSoft Inc.).

Results

Genetic Variability in Malagasy Subgroups

Goodness of Population Subgrouping

AMOVA showed that the between-group component of HVS-I variance was highest when Antandroy and Antanosy were pooled in the same group ($F_{CT} = 0.039$, $P = 0.3236$). When HLs were moved to the same group with Antandroy or Antanosy, F_{CT} values decreased and became negative (respectively, with $P = 1.0000$ and $P = 0.6686$), whereas the within-group component or F_{SC} scaled up by six to eight times, from 0.006 ($P = 0.2346$) to 0.038 ($P = 0.2766$) and 0.051 ($P = 0.2033$). This shows that the main source of genetic differentiation in the island separates HLs from CTs and justifies pooling the different subpopulations within that main divide.

Analysis of Maternal Lineages

The pool of mtDNA sequences found in Malagasy samples was a clear admixture of typical Bantu and Austronesian lineages (fig. 2*a*, supplementary table S1, Supplementary Material online). The averaged proportion of the two linguistic–geographic components (listed henceforth as Indonesian:African) were similar in HL (63%:37%) and CT (62%:38%) subgroups. The admixture ratio in coastal groups varied from 67%:33% in Antanosy to 54%:46% in Antaisaka. The two ethnic subgroups shared most of Southeast Asian-specific haplogroups (M[xM7], M7c1c, E1a, F3b, R9) and the "Polynesian motif' B4a1a1 (Soodyall et al. 1995). The African haplogroup inventory was more heterogeneous in CTs, where 11 more lineages than in HLs were found, despite the deviation from a hypothesis of equal diversity not being statistically significant (two-tailed Fisher exact test, $P = 0.78$) and the possible role that nonrandom mating might have played in the loss of some HL lineages. Nonetheless, when comparing mitochondrial lineages between CT and HL (table 2), *H* values for binary variability were significantly more diverse among African ($t = 3.27$, $P \sim 0.001$) than among Indonesian ($t = 0.80$, $P > 0.40$) derived lineages. A higher heterogeneity of the African CT component (*H* values) held true also at fast-mutating markers (HVS-I haplotypes, $t = 2.44$, $P < 0.01$) as well as at the unbiased estimate of intralineage diversity (WIMP values).

Analysis of Paternal Lineages

From a paternal point of view (fig. 2*b*, supplementary table S1, Supplementary Material online), a prevalence of African lineages was observed both in HLs and CTs, but with different proportions (HL 39:50%; CT 20:74%) and with extreme values in Antandroy (14% Indonesian, 86% African). Potential recent contributions from Eurasia (haplogroups R1b1, R1a1, J*, and J2; Francalacci and Sanna 2008), the Indian subcontinent (haplogroups R1a1, J*, J2, and L*; Sengupta et al. 2006), or the Horn of Africa–Arabia (haplogroups J*, J2, and E1b1b1a; Luis et al. 2004) sum to about 11% in HL and 4% in CT, but frequencies differ among coastal ethnic groups (0% in Antandroy and Antaisaka, 9.3% in Antanosy).

Regardless of the unequal apportionment, Asian and African components were virtually defined by the same haplogroups in the two ethnic subgroups and diversity

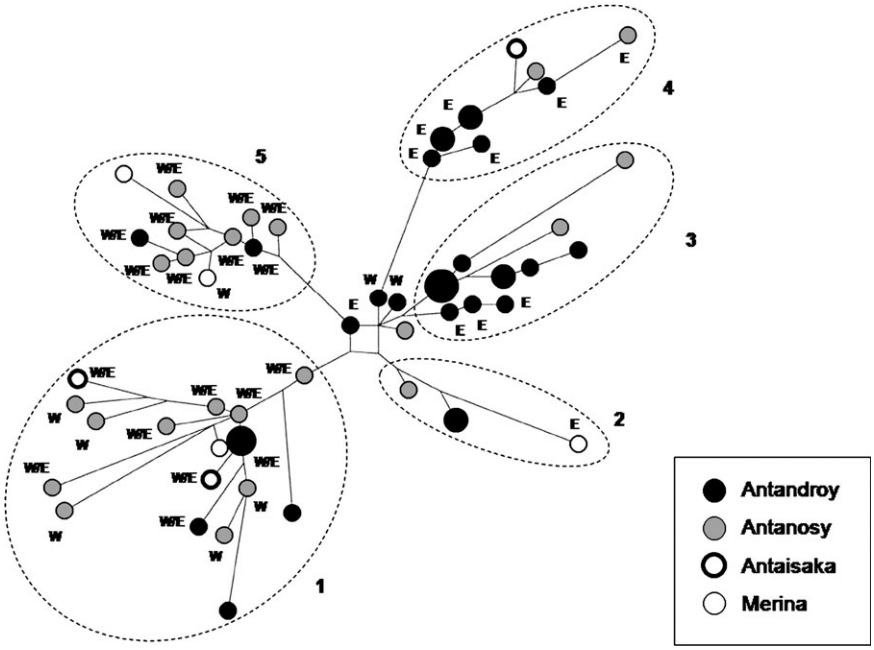

FIG. 3.—Median-joining network of 17-locus haplotypes ("Y-filer" set) belonging to E1b1a chromosomes. Circles represent haplotypes with areas proportional to the number of individuals they contain. Capital letters indicate haplotypes with affinities with western-central (W) and eastern (E) Africans or both (W/E).

indexes were similar, so that oscillations in UEP diversity could be said to be confidently explained by sampling bias (table 2). However, a less homogeneous pattern emerged after a more thoroughly descriptive analysis.

A basic stratification of the African-derived male pool is demonstrated by the architecture of the network (fig. 3) linking haplotypes from E1b1a, the most frequent haplogroup in all population samples (44% Merina, 69% Antandroy, 50% Antanosy, and 37.5% Antaisaka). Haplotypes from the different Malagasy groups appeared unevenly distributed in the five main subclusters. Subclusters 3 and 4, where Antandroy chromosomes concentrated, grouped East-like haplotypes (E), whereas subclusters 1 and 5, where most of Antanosy and Merina chromosomes fell, hosted Western-like haplotypes (W). The association of Antandroy Y chromosomes with E-like haplotypes and Antanosy Y chromosomes with W-like haplotypes is statistically supported (two-tailed Fisher exact test, $P = 0.002$). A direct descent of all the CTs from southern or southeastern African males is pointed out by haplotype sharing results (supplementary table S3, Supplementary Material online) for B2a and E2b haplotypes (16% of CT matchings).

As Y chromosomes varied between coastal and inland populations in terms of relative admixture proportion and lineage ancestry, Y data do confirm the between-group heterogeneity of African-derived lineages observed in the mitochondrial genome. In brief, a sex- and ethnic-biased contribution to the two geographic–linguistic components could be observed.

Individual Ancestry

The mutual exclusivity of Malagasy lineages provides the rare opportunity of calculating the frequency of individuals with homogeneous and heterogeneous ancestry at Y and mt genomes (table 3). The deviations from expected values under a random mating model might be considered the analogue, at complementary haploid markers, of the deviations from the Hardy–Weinberg equilibrium at diploid loci.

Whatever the criterion of subdivision (ethnic group, population subgroup, and ancestry pair), expected and observed distributions closely overlapped, even though the deviations were higher in the ethnic group with stronger social limitations to random mating (Merina). This implies that mating choices, whether due to natural preferences or imposed by social rules, are independent of the ancestry of the genes encoded on the Y chromosome or on the mtDNA.

Origin of Admixture Components

In order to reconstruct the geographic origin of the admixed lineages, shared haplotypes and pairwise D_{HS} distances were always analyzed by taking haplogroups with African and Indonesian ancestry separately (see tables 4–7).

Origin of Maternal Lineages

Of a total of 170 Malagasy HVS-I sequences, 117 (68.8%), belonging to 24 lineages and at least 18 different haplogroups, had an exact counterpart in the database (supplementary table S3, Supplementary Material online) with homogeneous matching rates between subgroups or ancestries (two-tailed Fisher exact tests, all with $P > 0.89$).

Regarding African-type sequences, the links with populations at the roots of the Bantu dispersal (western and central Africans) were closer for HL than for CT subgroups.

Table 3
Individual Ancestry

		Homogeneous Ancestry						Heterogeneous Ancestry												
		AF–AF		IN–IN		Subtotal		AF–IN		IN–AF		EU–AFR		EU–IN		Subtotal				
	N	Exp	Obs	Exp	Obs	Exp	Obs	Exp	Obs	Exp	Obs	Exp	Obs	Exp	Obs	Exp	Obs	χ^2	df	*P*
Antandroy	45	18.20	17	3.20	2	**21.40**	**19**	20.80	22	2.80	4					**23.60**	**26**	1.11	3	0.774
Antanosy	45	9.02	10	6.89	7	**15.91**	**17**	19.98	19	3.11	3	1.87	1	4.13	5	**29.09**	**28**	0.75	5	0.980
Antaisaka	8	3.13	3	1.13	1	**4.25**	**4**	1.88	2	1.88	2					**3.75**	**4**	0.04	3	0.998
Merina	9	1.78	3	2.78	4	**4.56**	**7**	2.22	1	2.22	1					**4.44**	**2**	2.71	3	0.438
Total	107	32.13	33	13.99	14	**46.12**	**47**	44.88	44	10.01	10	1.87	1	4.13	5	**60.88**	**60**	0.63	5	0.987

AF = African ancestry, IN = Indonesian ancestry, EU = Eurasian ancestry. First place terms in ancestry pairs (i.e., AF in AF–IN) refers to Y ancestry.

Among shared African HVS-I, the sequences missing in Eastern Bantu samples (the L1c2 motif and L3b1 motifs 16093–16223–16278–16362 and 16223–16278–16362, supplementary table S3, Supplementary Material online) sum to the 72% of HL and to the 25% of CT sequences.

MtDNA gene pool in CTs (all groups) should have been more heavily influenced than the HLs' pool by contributions from Eastern Bantu-speaking women. In fact, haplogroups L0a1a, L0a2, L2a1a, L3e1b, and the 16192T derived subcluster of L2a1b, which are preferentially observed in ethnic groups settled along the interlacustrine area, the Zambezi river, and from Mozambique (Pereira et al. 2001; Salas et al. 2002; Knight et al. 2003; Castrì et al. 2009) were observed only in Antandroy, Antaisaka, or Antanosy (supplementary table S3, Supplementary Material online). D_{HS} values (table 4) heavily support the above findings as reference population samples from West and South-East Africa scored differently in CT and HL rankings.

Malagasy Indonesian-type sequences matched closely with Insular Southeast Asian haplotypes (supplementary table 5 and S3, Supplementary Material online). In particular,

Table 4
Population Pairwise Comparisons: African Mitochondrial Haplotypes

Region (Population)	*N*	Area	Language	D_{HS}	*H*	Reference
Malagasy (HL)	18		Austronesian WMP	—	0.686 ± 0.112	This research; Hurles et al. (2005)
West Africa (Fulbe)	57	WAF	NC	0.750	0.972 ± 0.011	Watson et al. (1997)
Mozambique	414	SEAF	NC-Bantu	0.852	0.972 ± 0.003	Pereira et al. (2001); Salas et al. (2002)
Malagasy (CT)	47		Austronesian WMP	—	0.961 ± 0.012	This research; Hurles et al. (2005)
Mozambique	414	SEAF	NC-Bantu	0.610	0.972 ± 0.003	Pereira et al. (2001), Salas et al. (2002)
Tanzania (Sukuma)	21	EAF	NC-Bantu	0.798	1.000 ± 0.015	Knight et al. (2003)
Malagasy (CT + HL)	65		Austronesian WMP	—	0.930 ± 0.112	This research; Hurles et al. (2005)
Mozambique	414	SEAF	NC-Bantu	0.562	0.972 ± 0.003	Pereira et al. (2001), Salas et al. (2002)
Angola	154	SWAF	NC-Bantu	0.730	0.993 ± 0.002	Plaza et al. (2004), Beleza et al. (2005)
West Africa (Fulbe)	57	WAF	NC	0.738	0.972 ± 0.011	Watson et al. (1997)
Kenya (Kikuyu)	25	EAF	NC-Bantu	0.801	0.993 ± 0.013	Watson et al. (1996)
Sao Tomè	153	WAF	NC-Bantu	0.811	0.982 ± 0.004	Mateu et al. (1997), Trovoada et al. (2004)
Kenya (Nairobi)	100	EAF	NC-Bantu	0.831	0.995 ± 0.002	Brandstätter et al. (2004)
Tanzania (Sukuma)	21	EAF	NC-Bantu	0.832	1.000 ± 0.015	Knight et al. (2003)
Cameroon	550	WCAF	NC-Bantu	0.837	0.994 ± 0.001	Coia et al. (2005), Destro-Bisol et al. (2004), Cerný et al. (2004)
Equatorial Guinea	56	WCAF	NC-Bantu	0.867	0.940 ± 0.013	Mateu et al. (1997), Pinto et al. (1996)
Sierra Leone	277	WAF	NC-Bantu	0.900	0.991 ± 0.001	Jackson et al. (2005)
Guinea Bissau	372	WAF	NC-Bantu	0.902	0.986 ± 0.002	Rosa et al. (2006)
Senegal	238	WAF	NC-Bantu	0.903	0.987 ± 0.002	Graven et al. (1995), Rando et al. (1998)
Niger-Nigeria	103	WAF	NC-Bantu	0.915	0.994 ± 0.003	Watson et al. (1996, 1997)
Kenya (Turkana)	37	EAF	Nilo-saharan	0.926	0.994 ± 0.009	Watson et al. (1996)
Ethiopia	385	EAF	Afro-Asiatic	0.928	0.994 ± 0.001	Kivisild et al. (2004)
Sudan	75	EAF	Afro-Asiatic	0.950	0.993 ± 0.004	Krings et al. (1999)
Tanzania (Datoga)	18	EAF	Nilo-saharan	0.955	0.987 ± 0.023	Knight et al. (2003)
Sudan (Nubian)	79	EAF	Nilo-saharan	0.967	0.974 ± 0.009	Krings et al. (1999)
Tanzania (Iraqw)	12	EAF	Afro-Asiatic	0.972	0.924 ± 0.058	Knight et al. (2003)
Cabo Verde	292	WAF	NC-Bantu	0.985	0.975 ± 0.004	Brehm et al. (2002)
Somalia	15	EAF	Afro-Asiatic	1.000	1.000 ± 0.024	Watson et al. (1996)
Mauritania	30	WAF	Afro-Asiatic	1.000	0.975 ± 0.017	Rando et al. (1998)
South East Africa	414			0.562	0.972 ± 0.004	
South West Africa	181			0.730	0.982 ± 0.005	
West Central Africa	606			0.839	0.993 ± 0.001	
East Africa	767			0.911	0.997 ± 0.000	
West Africa	1522			0.943	0.993 ± 0.001	

EAF—East Africa, WAF—West Africa, WCAF—West Central Africa, SWAF—South West Africa, SEAF—South East Africa, NC-Bantu—Niger-Congo Bantu.

Table 5
Population Pairwise Comparisons: Asian Mitochondrial Haplotypes

Region (Population)	*N*	Area	Language	D_{HS}	*H*	Reference
Malagasy (CT + HL)	105		Austronesian WMP		0.839 ± 0.033	This research
Indonesia (Ambon)	43	ISEA	Austronesian WMP	0.532	0.971 ± 0.012	Hill et al. (2007)
Sulawesi (Ujung Padang)	46	ISEA	Austronesian WMP	0.544	0.977 ± 0.011	Hill et al. (2007)
Sulawesi (Toraja)	64	ISEA	Austronesian WMP	0.560	0.949 ± 0.013	Hill et al. (2007)
Lombok (Mataran)	44	ISEA	Austronesian WMP	0.566	0.984 ± 0.010	Hill et al. (2007)
Borneo (Banjarmasin)	89	ISEA	Austronesian WMP	0.585	0.989 ± 0.004	Hill et al. (2007)
Sulawesi (Manado)	89	ISEA	Austronesian WMP	0.633	0.959 ± 0.012	Hill et al. (2007)
Philippines	61	ISEA	Austronesian WMP	0.713	0.949 ± 0.012	Hill et al. (2007)
Java (Tengger)	36	ISEA	Austronesian WMP	0.725	0.932 ± 0.025	Hill et al. (2007)
Borneo (Kota Kinabalu)	67	ISEA	Austronesian WMP	0.759	0.982 ± 0.008	Hill et al. (2007)
Bali Flores Java	30	ISEA	Austronesian WMP	0.762	0.982 ± 0.016	Hill et al. (2007)
Taiwan (Puyuma)	52	TW	Austronesian F	0.777	0.868 ± 0.025	Trejaut et al. (2005)
Taiwan (Rukai)	50	TW	Austronesian F	0.777	0.881 ± 0.026	Trejaut et al. (2005)
Sumatra (Pekanbaru)	56	ISEA	Austronesian WMP	0.777	0.975 ± 0.010	Hill et al. (2006)
Sumatra (Palembang)	28	ISEA	Austronesian WMP	0.798	0.958 ± 0.030	Hill et al. (2006)
Sulawesi (Palu)	38	ISEA	Austronesian WMP	0.812	0.969 ± 0.018	Hill et al. (2007)
Sumatra (Bangka)	34	ISEA	Austronesian WMP	0.843	0.975 ± 0.014	Hill et al. (2006)
Taiwan (Paiwan)	55	TW	Austronesian F	0.844	0.910 ± 0.018	Trejaut et al. (2005)
Sumba (Waing)	50	ISEA	Austronesian WMP	0.847	0.976 ± 0.011	Hill et al. (2007)
Malaysia (Aborigens)	96	ISEA	Austronesian WMP	0.851	0.912 ± 0.016	Hill et al. (2006)
Indonesia (Alor)	45	ISEA	Austronesian WMP	0.854	0.975 ± 0.013	Hill et al. (2007)
Taiwan (Amis)	98	TW	Austronesian F	0.864	0.923 ± 0.012	Trejaut et al. (2005)
Bali (Denpasar)	65	ISEA	Austronesian WMP	0.874	0.990 ± 0.005	Hill et al. (2007)
Singapore (Malaysians)	205	ISEA	Austronesian WMP	0.907	0.993 ± 0.002	Wong et al. (2007)
Taiwan (Tsou)	60	TW	Austronesian F	0.910	0.906 ± 0.018	Trejaut et al. (2005)
Sumatra (Padang)	24	ISEA	Austronesian WMP	0.926	0.978 ± 0.019	Hill et al. (2006)
Taiwan (Yami)	64	TW	Austronesian F	1.000	0.864 ± 0.016	Trejaut et al. (2005)
Taiwan (Bunun)	89	TW	Austronesian F	1.000	0.833 ± 0.021	Trejaut et al. (2005)
Taiwan (Saisiat)	63	TW	Austronesian F	1.000	0.919 ± 0.012	Trejaut et al. (2005)
Malaysia (Semang)	112	ISEA	Austronesian WMP	1.000	0.813 ± 0.023	Hill et al. (2006)
Taiwan (Atayal)	109	TW	Austronesian F	1.000	0.853 ± 0.022	Trejaut et al. (2005)
Malaysia (Senoi)	52	ISEA	Austronesian WMP	1.000	0.852 ± 0.031	Hill et al. (2006)
Insular Southeast Asia	1,374			0.859	0.991 ± 0.001	
Taiwan	640			0.893	0.969 ± 0.003	
Continental Southeast Asia	508			0.986	0.997 ± 0.001	
South Asia	1,950			1.000	0.994 ± 0.000	

SA—South Asia, CSEA—Continental Southeast Asia, TW—Taiwan, ISEA—Insular Southeast Asia, F—Formosan.

samples from the Molucca islands (Ambon) and Sunda Islands (Sulawesi, Lombok, and Borneo) scored the lowest D_{HS} distances. The fact that Borneans from the Barito River region (Banjarmasin) were more distant (fifth ranking place) from Malagasy samples than other Southeast Asian populations makes the correspondence between vocabulary and genetic data less obvious than previously reported (Hurles et al. 2005). An increment of the Malagasy–Borneans genetic distance was due to B4a1a1 haplotypes (16189–16217–16247–16261 motif), which are common in the Malagasy Indonesian component (34.3%, this research) and in Ambon (14.0%, Hill et al. 2007) but only sporadically found in Borneo (1.3% Hill et al. 2007). Computer simulations under an extended Wright–Fisher model and stringent priors (growth rate = 0.02, $\mu = 9.5 \times 10^{-6}$) exclude (max upper 95%CI = 15.3%) that lineage sorting or founder effects could have driven B4a1a1 frequency from ~2.2% (that of a putative source population showing the present frequency at Banjarmasin) to 34.3% (present frequency in Malagasy) within the last 2,500 years (125 generations), whatever the effective size of Austronesian founders ($50 < N < 5{,}000$). Hence, ancestors different from the present Banjar people should be invoked to explain the observed scenario. Alternatively, it might be the case that B4a1a1 had a higher frequency in Maanyan properly speaking groups (currently living North of Banjarmasin), which have not been genotyped so far (see also Adelaar 2006).

Taking mitochondrial data on the whole, whereas migration caused a significant loss of diversity in Indonesian-derived ($t = 4.80$ for CT and $t = 3.48$ for HL, $P < 0.001$) and highland African-derived ($t = 2.74$, $P < 0.05$) gene pools, in the CT African component ($t = 0.940$, $P > 0.30$) did not, further supporting a more heterogeneous flow from Africa to the coastal groups than to elsewhere.

Origin of Paternal Lineages

Sixty (55%) Malagasy 9-locus YSTR haplotypes and 597 (30%) one-step neighbors matched with selected database entries (supplementary table S3, Supplementary Material online). As for the origin of African Y-lineages, haplotype sharing and D_{HS} distances mirrored mitochondrial results (table 6 and supplementary table S3, Supplementary Material online): Genetic distances demonstrated a fair affinity with western and central African samples and, again, the lowest values were with South East African haplotypes. Nearly 33% of Mozambican chromosomes could be estimated to be identical to Malagasy chromosomes by descent and the ratio between the relative

Table 6
Population Pairwise Comparisons: African Y-STR Haplotypes

Region (Population)	*N*	Area	Language	D_{HS}	*H*	References
Malagasy (CT + HL)	79		Austronesian WMP	—	0.984 ± 0.005	This research
Mozambique	112	SEAF	NC-Bantu	0.553	0.988 ± 0.004	Alves et al. (2003)
Angola	75	SWAF	NC-Bantu	0.708	0.994 ± 0.004	YHRD
Bubi	133	WAF	NC-Bantu	0.848	0.986 ± 0.004	Barrot et al. (2007)
CAR	165	WCAF	NC-Bantu	0.854	0.991 ± 0.002	Lecerf et al. (2007)
Fang	116	WAF	NC-Bantu	0.887	0.988 ± 0.004	Barrot et al. (2007)
South Africa (Xhosa)	99	SAF	NC-Bantu	0.896	0.974 ± 0.009	Leat et al. (2004)
Equatorial Guinea	100	WCAF	NC-Bantu	0.915	0.994 ± 0.001	Arroyo-Pardo et al. (2005)
Guinea Bissau	161	WAF	NC-Bantu	0.917	0.998 ± 0.001	Rosa et al. (2006)
South Africa	73	SAF	NC-Bantu	0.921	0.968 ± 0.013	Leat et al. (2004)
Cameroon	54	WCAF	NC-Bantu	0.937	0.872 ± 0.043	YHRD
Somalia	201	NEAF	Afro-Asiatic	0.986	0.956 ± 0.007	Hallenberg et al. (2005)
West Africa	79	WAF	NC-Bantu	0.978	0.992 ± 0.004	YHRD
South East Africa	112	SEAF		0.553	0.988 ± 0.004	
South West Africa	75	SWAF		0.708	0.994 ± 0.004	
West Central Africa	568	WCAF		0.876	0.995 ± 0.001	
South Africa	172	SAF		0.893	0.971 ± 0.007	
West Africa	240	WAF		0.936	0.998 ± 0.001	
North East Africa	201	NEAF		0.986	0.956 ± 0.007	

NEAF—Near East Africa, WAF— West Africa, WCAF— West Central Africa, SWAF—South West Africa, SEAF—South East Africa, NC-Bantu—Niger-Congo Bantu.

frequencies of exact and neighboring haplotypes was three times higher in Mozambicans (11.8) than in western (3.1) or central (3.2) population samples.

The analyses of Asian-derived Y chromosomes were consistent with mitochondrial outcomes as well. Southeast Asian populations showed the lowest distances and the highest proportion of haplotype matchings (table 7 and supplementary table S3, Supplementary Material online): about three-fourths of Southeast Asian haplotypes exactly matching Malagasy lineages belonged to Malay people (from Sarawak and mainland Malaysia). However, identical haplotypes were never over 6%, and both, the size of Malagasy Y chromosomes with Asian ancestry and the geographic coverage of reference samples, are inadequate to give a comprehensive picture. Similarly as for mtDNA data, admixture led to a more appreciable decrease of haplotype diversity in the Indonesian ($t = 1.86$, $0.05 < P < 0.10$) than in the African component ($t = 0.63$, $P > 0.50$), in contrast with the hypothesis of a smaller migration from Africa than from Asia (Hurles et al. 2005).

Lineages not directly linked with the former admixture (i.e., of presumed West Eurasian origin) could be recognized only in Antanosy. They correspond to R1a1, R1b1, J2, E1b1b1a, and L* haplogroups. The geographic assignment of exact matching (9-locus $N = 166$) and neighboring ($N = 1811$) haplotypes in the YHRD (release 27) and in our database suggested for J2, R1a1, and R1b1 chromosomes a clear European-Near Eastern origin, for the E1b1b1a chromosome a Somali origin, and an uncertain origin (no matchings) for the L* chromosome.

TSAE

Computer simulations provide a most likely estimate of the timeline for the arrival of each genetic component to Madagascar (table 8). The male African component is compatible with a large size range of founders (200–2,000) and with time windows in the 75- to 800-yBP range. The closeness among observed *H* values in Malagasy (0.983) and Mozambican (0.988) population samples makes simulated values for migrant and parental populations covarying over a large generation interval. Evolutionary scenarios with prior $N_em > 2{,}000$ were unreliable (data not shown).

The estimated TSAE for the African female genetic counterpart was deeper and differed in CTs and HLs (1,820–3,820 yBP with 1,000–2,000 N_em in CTs; >3,500 yBP with 200 N_em in HLs). The former range largely overlaps with the tolerance intervals estimated for the Asian founders whether they calculated taking a single putative parental population (Ambon, 2,340–3,080 yBP, 500–1,000 N_em) or the population pool showing the first five lowest DHS values (from Ambon to Banjarmasin, 1,000–3,080 yBP, 200–1,000 N_em). Simulations for the male Asian component were not performed because of the inadequateness of both analyzed and reference samples.

Discussion

Amount of Admixture

The uniqueness of Malagasy genome in the landscape of human genetic variation is due to a recent balanced mix of gene pools that have been shaped by at least 60,000 years of independent evolution. It offered us the rare opportunity of using in combination mitochondrial and Y markers to assess every parental lineage to its homeland following a mutual exclusive criterion. It also helped estimate how within-lineage variability and lineage ancestries are apportioned into the different ethnic groups. The relevance of the genotyped sample in terms of size and ethnic coverage, as well as the large discriminating power of the chosen markers, allowed us to carry out a three-level analysis in which ethnicity (HL and CT subgroups), lineage ancestry (African, Indonesian), and inheritance (paternal and maternal) could be concurrently considered. Lastly, a novel simulation approach was applied to best place in time and space

Table 7
Population Pairwise Comparisons: Asian Y-STR Haplotypes

Region (Population)	*N*	Area	Language	D_{HS}	*H*	References
Malagasy (CT + HL)	25		Austronesian WMP	—	0.940 ± 0.031	This research
Malaysia (Sarawak–Melanau)	102	ISEA	Austronesian WMP	0.929	0.970 ± 0.006	Chang et al. (2009)
Malaysia (Malay origin)	333	ISEA	Austronesian WMP	0.951	0.999 ± 0.000	Chang et al. (2007)
Malaysia (Sarawak–Iban)	101	ISEA	Austronesian WMP	0.965	0.990 ± 0.003	Chang et al. (2003)
Philippines	76	ISEA	Austronesian WMP	0.969	0.998 ± 0.003	Kwak et al. (2005)
Timor East	138	ISEA	Austronesian CMP	0.980	0.994 ± 0.002	Souto et al. (2006)
Taiwan	200	TW	Austronesian F	0.986	0.998 ± 0.001	Huang et al. (2008)
China (Han)	187	CSEA	Sino-Tibetan Chinese	0.986	1.000 ± 0.001	Yang et al. (2006)
Hong Kong	481	CSEA	Sino-Tibetan Chinese	0.992	0.998 ± 0.000	Yeung et al. (2006)
Malaysia (Sarawak–Bidayuh)	113	ISEA	Austronesian WMP	1.000	0.980 ± 0.005	Chang et al. (2003)
Malaysia (Kensiu)	18	ISEA	Austronesian WMP	1.000	0.843 ± 0.056	Bekaert et al. (2006)
Malaysia (Malay)	36	ISEA	Austronesian WMP	1.000	0.998 ± 0.007	Bekaert et al. (2006)
Malaysia (Jahai)	15	ISEA	Austronesian WMP	1.000	0.943 ± 0.040	Bekaert et al. (2006)
Singapore (Malay origin)	186	ISEA	Austronesian WMP	1.000	0.998 ± 0.001	Yong et al. (2006)
East Java	90	ISEA	Austronesian WMP	1.000	0.998 ± 0.002	Kido et al. (2005)
Indonesia	32	ISEA	Austronesian WMP	1.000	1.000 ± 0.008	Kwak et al. (2005)
Singapore	212	ISEA	Austronesian WMP	1.000	0.996 ± 0.001	Tang et al. (2006)
China (Minnan Han)	109	CSEA	Sino-Tibetan Chinese	1.000	0.992 ± 0.003	Hu (2006)
China (Tibetan minority)	119	CSEA	Sino-Tibetan Burnese	1.000	0.996 ± 0.002	Zhu, Deng, et al. (2006), Zhu, Liu et al. (2006)
China Han (Ningxia)	101	CSEA	Sino-Tibetan Chinese	1.000	0.999 ± 0.002	Zhu, Deng, et al. (2006), Zhu, Liu et al. (2006)
China (Uigur)	107	CSEA	Sino-Tibetan Chinese	1.000	0.999 ± 0.001	Zhu, Shen, et al. (2005), Zhu, Wang, et al. (2005)
China (Yi)	100	CSEA	Sino-Tibetan Chinese	1.000	0.990 ± 0.004	Zhu, Shen et al. (2005); Zhu, Wang et al. (2005)
China (Beijing Han)	49	CSEA	Sino-Tibetan Chinese	1.000	0.999 ± 0.004	Kwak et al. (2005)
China (Yunnan)	29	CSEA	Sino-Tibetan Chinese	1.000	0.998 ± 0.001	Kwak et al. (2005)
China Han (Northeast Liaoning)	141	CSEA	Sino-Tibetan Chinese	1.000	0.999 ± 0.001	Wang and Sawaguchi (2006)
China (Tibetan)	107	CSEA	Sino-Tibetan Burnese	1.000	0.998 ± 0.002	Li et al. (2007)
China (Manchurians)	32	CSEA	Sino-Tibetan Chinese	1.000	0.960 ± 0.023	Kwak et al. (2005)
Thailand	41	CSEA	Tai-Kadal	1.000	1.000 ± 0.005	Kwak et al. (2005)
Vietnam	43	CSEA	Austro-Asiatic	1.000	0.998 ± 0.006	Kwak et al. (2005)
India (Chotanagpur Plateau)	115	SA	Indo-Iranian	1.000	0.998 ± 0.002	Banerjee et al. (2005)
India (Jat Sikhs)	108	SA	Indo-Iranian	1.000	0.977 ± 0.007	Henke et al. (2001)
Bangladesh	72	SA	Indo-Iranian	1.000	0.998 ± 0.003	Dobashi et al. (2005)
India (Jats of Haryana)	84	SA	Indo-Iranian	1.000	0.942 ± 0.016	Nagy et al. (2007)
India (Bengal)	57	SA	Indo-Iranian	1.000	0.996 ± 0.004	Singh et al. (2006)
Insular Southeast Asia	1,452	ISEA		0.979	0.999 ± 0.000	
Taiwan	200	TW		0.986	0.998 ± 0.001	
Continental Southeast Asia	2,127	CSEA		0.996	0.999 ± 0.000	
South Asia	516	SA		1.000	0.993 ± 0.001	

SA—South Asia, CSEA—Continental Southeast Asia, TW—Taiwan, ISEA—Insular Southeast Asia, F—Formosan.

the origins of the migrations. It opened new scenarios on the admixture history of Malagasy ethnic groups with respect to previous analyses.

Our results confirmed that admixture in Malagasy was due to the encounter of people surfing the extreme edges of two of the broadest historical waves of language expansion: the Austronesian and Bantu expansions. In fact, all Madagascan living groups show a mixture of uniparental lineages typical of present African and South East Asian populations with only a minor contribution of Y lineages with different origins. Two observations suggest that the Y lineages with "another origin" entered the island in recent times: 1) they are particularly frequent in the Tanosy area (Fort Dauphin), and around Antananarivo, where commercial networks and slave trade had a focus; 2) they matched with haplotypes typical of present Indo-European (Europeans) and Arabic-speaking (Somali) people.

The proportion of the main ancestral genetic components varied between highland and coastal groups both qualitatively and quantitatively, depending on the sex. As a general rule, the Indonesian ancestry was more conserved in the female than in the male gene pool, in HLs than in CTs. In synthesis, genes rather than language best fit the diversity of the anthropological heritage evident among Malagasy groups.

Origin of Founding Lineages

The deep rooting of ancestral lineages, the high discriminating power of HVS-I and, above all, of YSTR haplotypes, coupled with the availability of large reference databases, allowed us to identify a likely place of origin for each lineage. However, the search of a pinpointed geographic ancestry could be inconclusive even for forthcoming genomewide surveys. Either gene flows in the two areas of origin after the Malagasy migration or the occurrence of complex underlying demographies in the making of Malagasy gene pool would make the assessment of univocal

Table 8
Most likely TSAEs

	Observed Values	Expected TSAE (95% CI in Generations)				Tolerance Interval
Y Africa CT	(Mozambique)	*N_em = 200* *H_0 = 0.9855* *w = 1.003*	*N_em = 500* *H_0 = 0.9855* *w = 1.003*	*N_em = 1000* *H_0 = 0.9855* *w = 1.003*	*N_em = 2000* *H_0 = 0.9855* *w = 1.003*	**200–2000**
D_{HS}	0.563	3–4	8–11	13–20	19–32	
H_x	0.988	3–37	4–49	3–45	3–81	
H_y	0.983	1–6	1–16	1–200	3–93	**3–32**
TSAE		**3–4**	**8–11**	**13–20**	**19–32**	**(75–800 years)**
mt Africa HL	(Fulbe)	*N_em = 200* *H_0 = 0.829* *w = 1.003*	*N_em = 500* *H_0 = 0.829* *w = 1.003*	*N_em = 1,000* *H_0 = 0.829* *w = 1.003*	*N_em = 2,000* *H_0 = 0.829* *w = 1.003*	**200**
D_{HS}	0.750	63–200	127–200	190–200	—	
H_x	0.972	175–200	—	191–200	195–200	
H_y	0.686	13–200	76–200	—	—	**175–200**
TSAE		**175–200**	—	—	—	**(3,500–>4,000)**
mt Africa CT	(Mozambique)	*N_em = 200* *H_0 = 0.9665* *w = 1.003*	*N_em = 500* *H_0 = 0.9665* *w = 1.003*	*N_em = 1,000* *H_0 = 0.9665* *w = 1.003*	*N_em = 2,000* *H_0 = 0.9665* *w = 1.003*	**1,000–2,000**
D_{HS}	0.610	15–42	39–96	97–153	96–196	
H_x	0.972	12–200	13–200	11–200	14–200	
H_y	0.961	1–11	2–31, 108–200	2–96, 106–200	5–200	**96–196**
TSAE		—	—	**106–153**	**96–196**	**(1,820–3,820 years)**
mt Asia HL + CT	(Ambon)	*N_em = 200* *H_0 = 0.905* *w = 1.003*	*N_em = 500* *H_0 = 0.905* *w = 1.003*	*N_em = 1,000* *H_0 = 0.905* *w = 1.003*	*N_em = 2,000* *H_0 = 0.905* *w = 1.003*	**500–1,000**
D_{HS}	0.532	24–83	52–135	85–190	114–200	
H_x	0.971	121–200	117–200	132–200	101–200	
H_y	0.839	7–200	20–200	114–154	—	**117–154**
TSAE		—	**117–135**	**132–154**	—	**(2,340–3,080 years)**
mt Asia HL + CT	(Ambon-Banjarmasin)	*N_em = 200* *H_0 = 0.905* *w = 1.003*	*N_em = 500* *H_0 = 0.905* *w = 1.003*	*N_em = 1000* *H_0 = 0.905* *w = 1.003*	*N_em = 2000* *H_0 =0.905* *w = 1.003*	**200–1000**
D_{HS}	0.532–0.585	24–98	52–153	85–218	114–200	
H_x	0.949–0.989	50–200	54–200	44–350	43–200	
H_y	0.839	7–200	20–200	114–154	—	**50–154**
TSAE		**50–98**	**54–153**	**114–154**	—	**(1,000–3,080 years)**

Simulated multistate markers were 9-locus STR haplotypes evolving according to a strict SMM (μ = 0.00185 mut/locus/gen). Simulated binary markers were 360 D-loop sites evolving under a IAM (μ = 0.0000095 mut/locus/gen). For each model a generation interval was accepted as most likely TSAE when all the observed values of D_{HS}, H_x, and H_y fell within the tolerance interval (95%CI) of the simulated distribution. Prior parameters are in italics.

relationships misleading. On the basis of our results, we could confidently frame a macro-area most likely homeland of the two main components.

Population samples from a region embracing Sunda Islands, Molucca islands, and Malaysia showed the closest genetic affinities with Malagasy "Indonesian"-type lineages. The homogeneous distribution among ethnic groups at binary and multistate markers and the loss of variability with respect to putative founding populations suggest a migration that took place in a few waves. Estimates based on the properties of the D_{HS} statistic never exceeded 1,000 effectives, as regards the size of founders, and the 1,000- to 3,000-yBP time range for the TSAE.

Several considerations make our results consistent with a migration occurred during the second pulse in the spread of Austronesian-speakers started around 3,800 yBP out of Taiwan toward Philippines, Northern Sulawesi, and West Borneo (Belwood 1995; Spriggs 2003; Gray et al. 2009). First, a high proportion of Malagasy lineages, the 47% and 28%, respectively of mitochondrial and Y haplotypes, was observed in Indonesians but not in aboriginal Taiwanese (a total of 10 populations) nor in continental Southeast Asians (a total of 16 populations). Second, Malagasy belongs to the WMP languages, the first clade splitting from the Formosan, the deepest branch of the Austronesian family tree spoken only in Taiwan (Ethnologue 1996). Third, linguistic evidence (Adelaar 2009) points that Malagasy retains more conservative morpho-syntactical features (Philippine-type structure) than Maanyan, which has been under West Indonesian influence (Malay-type structure) derived from more recent contacts with Malays. It supports a major migration occurred earlier than the time of the Malay political and cultural dominance in Indonesia (sixth to seventh century AD).

The second phase of Austronesian expansion would have been pulsed by the acquisition of the ship-building technology (outrigger canoes) needed to expand westward and eastward along thousands of miles in the Indian and Pacific Oceans (Blust 1999; Pawley 2002). It would also be the time when novel genetic lineages would have been acquired. Hence, the fact that the Malagasy basic vocabulary and Maanyan are closely related does not exclude that Madagascar was settled by proto-Malays leaved in an early phase (3–2 thousand years ago) from a different and perhaps

wider area of West Indonesia, rather than specifically from the Southeast Barito. This does not exclude that some genetic lineages, in particular those paternally inherited, as well as some loanwords from Indonesian and Sanskrit languages (Adelaar 2009), entered into the island through secondary contacts mediated by the Malay traders in the Christian Era.

As far as the Malagasy "African" lineages, haplotype sharing and the analysis of the diversity between subgroups pointed to more complex dynamics: A recent layer of southeastern-like Y haplotypes would have superimposed onto an early background formed by haplotypes typical of populations currently living near the roots of Bantu dispersal, with a larger impact on coastal groups.

It is hard to say whether the most recent layer originated by mass migration or by mating with slaves or captives of East African origin. The second hypothesis is much less likely because slave trade in Madagascar was mainly an outward process (Campbell 1981). Similarly, an ultimate answer cannot be given to the question of how western African sequences entered into the Malagasy pool. The genetic evidence in HLs of a more conserved western-like mtDNA profile with 72% of matching lineages absent in present-day SE Bantu speakers suggests a link between Malagasy and western–central Bantu speakers not mediated by indirect gene flow with Mozambicans. It is also in agreement with a previous genetic report based on β-globin haplotypes (Hewitt et al. 1996) and with the origin of the Bantu linguistic borrowings observed in Malagasy (Dahl 1988), both suggesting a close affinity with upper East Africa north of the Zambesi river. The impact of the earliest Bantu contact (Comorian, in Dahl's definition of 1988) on Malagasy phonology would have been substantial (i.e., the development of vocal endings) and occurred before the settlement of Madagascar (Adelaar 2009). An indirect genetic evidence at support of this hypothesis is the fact that Congolese resulted as the most likely parental populations of the African HLA–DRB1 haplotypes in the Comoro islands (300 km NE off Madagascan shores, Gibert et al. 2006).

Evidence of an early interaction between SE-Asia and sub-Saharan Africa is accumulating: the occurrence of banana cultivation (Asian *Musa* spp. phytolits) in Southern Cameroon and Uganda before 500 BCE (Mbida et al. 2000; Lejju et al. 2006); the archeo-zoological evidence for an early (second millennium BCE–first millennium AD) introduction of *Bos indicus* into East Africa from Asian routes (Magnavita 2006); the excavation of chicken bones from a Neolithic limestone cave site at Zanzibar (Chami 2001). Accordingly, increasing support exists for long distance contacts between Austronesians and Bantu via the Indian Ocean much earlier of the first archaeological evidence of human settlements in Madagascar (2,300–2,200 yBP, Hedges et al. 1997; Burney et al. 2003, 2004). Thus, Malagasy admixture could have had a history in East Africa before it crossed the Mozambique Channel, even though genetic signatures of these first mainland contacts are still missing.

Unfortunately, neither Y chromosomal nor mitochondrial sublineages, unambiguously discriminating between a southeastern and a central-eastern Bantu ancestry, are available yet, but the dissection of E1b1b1a Y variability (fig. 3) and complete mtDNA sequencing (Gonder et al. 2007) open optimistic perspectives. Other insights should come from future genetic researches in populations from the Swahili coast and in a more relevant Indonesian population sample.

Supplementary Material

Supplementary tables S1–S3 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/)

Acknowledgments

We warmly thank all sample donors and the Malagasy who helped with the sample collection. Thanks are also due to Luca Taglioli, Laura Caciagli, Cristina Mela (University of Pisa), Davide Merlitti (Scuola Normale Superiore, Pisa), Cristina Fabbri, and Antonella Useli (University of Bologna) for technical support. We especially acknowledge Matt Hurles and Martin Richards for details of published mtDNA data, Brigitte Holt for her participating comments. The research was supported by a grant from the University of Pisa to G.P.

Literature Cited

Adelaar A. 1995a. Borneo as a cross-roads for comparative Austronesian linguistics. In: Bellwood P, Fox J, Tryon D, editors. The Austronesians. Historical and comparative perspectives. Canberra (Australia): RSPAS, Department of Anthropology (Australian National University). p. 81–102.

Adelaar A. 1995b. Asian roots of the Malagasy: a linguistic perspective. Bijdr Taal-Land-V. 151:325–356.

Adelaar A. 2006. The Indonesian migrations to Madagascar: making sense of the multidisciplinary evidence. In: Simanjuntak T, Pojoh IHE, Hisyam M, editors. Austronesian diaspora and the ethnogenesis of people in Indonesian archipelago. LIPI press. p. 1–23.

Adelaar A. 2009. Towards an integrated theory about the Indonesian migrations to Madagascar. In: Peregrine PN, Peiros I, Feldman M, editors. Ancient human migrations: an interdisciplinary approach. Salt Lake City (UT): University of Utah Press.

Alves C, Gusmão L, Barbosa J, Amorim A. 2003. Evaluating the informative power of YSTRs: a comparative study using European and new African haplotype data. Forensic Sci Int. 134:126–133.

Arroyo-Pardo E, Gusmão L, López-Parra AM, Baezam C, Mesa MS, Amorim A. 2005. Genetic variability of 16 Y-chromosome STRs in a sample from Equatorial Guinea Central Africa. Forensic Sci Int. 149:109–113.

Athey WT. 2006. Haplogroup prediction from Y-STR values using a Bayesian-allele-frequency approach. J Genet Geneal. 2:34–39.

Banerjee J, Trivedi R, Kashyap VK. 2005. Y-STR haplotypes in autochthonous tribal population of Chotanagpur Plateau India. J Forensic Sci. 50:739–744.

Barrot C, Sánchez C, Xifró A, Ortega M, Mas J, Huguet E, Corbella J, Gené M. 2007. Data for Y-chromosome haplotypes in Fang and Bubi populations from Bioko Equatorial Guinea. Forensic Sci Int. 168:e10–e12.

Bekaert B, Zainuddin Z, Hadi S, Goodwin W. 2006. A comparison of mtDNA and Y chromosome diversity in Malay populations. Int Congr Ser. 1288:252–255.

Beleza S, Gusmão L, Amorim A, Carracedo A, Salas A. 2005. The genetic legacy of western Bantu migrations. Hum Genet. 117:366–375.

Bellwood P. 1995. Austronesian prehistory in Southeast Asia: homeland, expansion and transformation. In: Bellwood P, Fox J, Tryon D, editors. The Austronesians. Historical and comparative perspectives. Canberra (Australia): RSPAS, Department of Anthropology (Australian National University). p. 81–102.

Blench RP. 1996. The ethnographic evidence for long-distance contacts between Oceania and East Africa. In: Reade J, editor. The Indian Ocean in antiquity. London: Kegan Paul. p. 417–438.

Blench R. 2006. The Austronesians in Madagascar and on the East African coast: surveying the linguistic evidence for domestic and translocated animals. Paper presented at: Tenth International Conference on Austronesian Linguistics; 2006 January 17–20; Palawan, Philippines. Linguistic Society of the Philippines and SIL International. http://www.sil.org/asia/Philippines/ical/papers.html.

Blench R. 2007. New palaeo-zoogeographical evidence for the settlement of Madagascar. Azania: J Brit Inst East Africa. 42:69–82.

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

Bosch E, Calafell F, González-Neira A, et al. (13 co-authors). 2006. Paternal and maternal lineages in the Balkans show a homogeneous landscape over linguistic barriers except for the isolated Aromuns. Ann Hum Genet. 70:459–487.

Bowler JM, Johnston H, Olley JM, Prescott JR, Roberts RG, Shawcross W, Spooner NA. 2003. New ages for human occupation and climatic change at Lake Mungo Australia. Nature. 421:837–840.

Brandstätter A, Peterson CT, Irwin JA, Mpoke S, Koech DK, Parson W, Parsons TJ. 2004. Mitochondrial DNA control region sequences from Nairobi Kenya: inferring phylogenetic parameters for the establishment of a forensic database. Int J Legal Med. 118:294–306.

Brehm A, Pereira L, Bandelt HJ, Prata MJ, Amorim A. 2002. Mitochondrial portrait of the Cabo Verde archipelago: the Senegambian outpost of Atlantic slave trade. Ann Hum Genet. 66:49–60.

Burney DA. 1987. Pre-settlement vegetation changes at Lake Tritrivakely, Madagascar. Palaeoecol Africa. 18:357–381.

Burney DA, Burney LP, Godfrey LR, Jungers WL, Goodman SM, Wright HT, Jull AJ. 2004. A chronology for late prehistoric Madagascar. J Hum Evol. 47:25–63.

Burney DA, Robinson GS, Burney P. 2003. Sporormiella and the late Holocene extinctions in Madagascar. Proc Natl Acad Sci USA. 100:10800–10805.

Campbell G. 1981. Madagascar and the Slave Trade, 1810–1895. J African Hist. 22:203–227.

Castrì L, Tofanelli S, Garagnani P, Bini C, Fosella X, Pelotti S, Paoli G, Pettener D, Luiselli D. 2009. MtDNA variability in two Bantu-speaking populations (Shona and Hutu) from eastern Africa: implications for peopling and migration patterns in sub-Saharan Africa. Am J Phys Anthropol. doi: 10.1002/ajpa.21070.

Cavalli-Sforza LL, Edwards AWF. 1967. Phylogenetic analysis: models and estimation procedures. Am J Hum Genet. 19:233–257.

Cerný V, Hájek M, Cmejla R, Brezek J, Brdicka R. 2004. mtDNA sequences of Chadic-speaking populations from northern Cameroon suggest their affinities with eastern Africa. Ann Hum Biol. 31:554–569.

Chami F. 2001. Chicken bones from a Neolithic limestone cave site, Zanzibar: contact between East Africa and Asia. In: Chami F, Pwiti G, Radimilahy C, editors. People, contacts and the environment in the African past. Dar-es-Salaam (Tanzania): University of Dar-es-Salaam Press. p. 84–97.

Chang YM, Perumal R, Keat PY, Kuehn DL. 2007. Haplotype diversity of 16 Ychromosomal STRs in three main ethnic populations Malays Chinese and Indians in Malaysia. Forensic Sci Int. 167:70–76.

Chang YM, Swaran Y, Phoon YK, Sothirasan K, Sim HT, Lim KB, Kuehn D. 2009. Haplotype diversity of 17 Y-chromosomal STRs in three native Sarawak populations (Iban, Bidayuh and Melanau) in East Malaysia. Forensic Sci Int Genet. 3:77–80.

Coia V, Destro-Bisol G, Verginelli F, Battaglia C, Boschi I, Cruciani F, Spedini G, Comas D, Calafell F. 2005. Brief communication: mtDNA variation in North Cameroon: lack of Asian lineages and implications for back migration from Asia to sub-Saharan Africa. Am J Phys Anthropol. 128:678–681.

Dahl OC. 1951. Malgache et Maanyan: une comparison linguistique. Oslo (Norway): Egede Intitutett.

Dahl OC. 1988. Bantu substratum in Malagasy. Etudes Océan Indien. 9:91–132.

Dahl O. 1991. Migration from Kalimantan to Madagascar. Oslo (Norway): The Institute for Comparative Research in Human Culture, Norwegian University Press.

Deschamps M. 1960. In: Levrault B, editor. Histoire de Madagascar. Col. Mondes d'outre-Mer. Paris: Berger-Levrault.

Destro-Bisol GH, Coia V, Boschi I, Verginelli F, Cagliá A, Pascali V, Spedini G, Calafell F. 2004. The analysis of variation of mtDNA hypervariable region 1 suggests that Eastern and Western Pygmies diverged before the Bantu expansion. Am Nat. 163:212–226.

Dewar R, Wright H. 1993. The culture history of Madagascar. J World Prehist. 7:417–466.

Dobashi Y, Kido A, Fujitani N, Susukida R, Hara M, Oya M. 2005. Y-chromosome STR haplotypes in a Bangladeshi population. Leg Med Tokyo. 7:122–126.

Excoffier L, Laval G, Schneider S. 2005. Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evol Bioinform Online. 1: 47–50.

Francalacci P, Sanna D. 2008. History and geography of human Y-chromosome in Europe: a SNP perspective. J Anthropol Sci. 86:59–89.

Gasse F, Van Campo E. 1998. A 40,000-yr pollen and diatom record from Lake Tritrivakely, Madagascar, in the southern tropics. Quat Res. 46:299–311.

Gibert M, Touinssi M, Reviron D, Mercier P, Boëtsch G, Chiaroni J. 2006. HLA–DRB1 frequencies of the Comorian population and their genetic affinities with Sub-Saharan African and Indian Oceanian populations. Ann Hum Biol. 33:265–278.

Gonder MK, Mortensen HM, Reed FA, de Sousa A, Tishkoff SA. 2007. Whole-mtDNA genome sequence analysis of ancient African lineages. Mol Biol Evol. 24: 757–768.

Graven L, Passarono G, Semino O, Boursot P, Santachiara-Benerecetti S, Langaney A, Excoffier L. 1995. Evolutionary correlation between control region sequence and restriction polymorphisms in the mitochondrial genome of a large Senegalese Mandenka sample. Mol Biol Evol. 12:334–345.

Gray RD, Drummond AJ, Greenhill SJ. 2009. Language phylogenies reveal expansion pulses and pauses in Pacific settlement. Science. 323:479–483.

Gusmão LB, Sánchez-Diz P, Calafell F, et al. (40 co-authors). 2005. Mutation rates at Y chromosome specific microsatellites. Hum Mutat. 26:520–528.
Hallenberg C, Simonsen B, Sanchez J, Morling N. 2005. Y-chromosome STR haplotypes in Somalis. Forensic Sci Int. 151:317–321.
Hammer MF. 1994. A recent insertion of an alu element on the Y chromosome is a useful marker for human population studies. Mol Biol Evol. 11:749–761.
Hedges REM, Pettitt PB, Bronk-Ramsey C, van Klinken GJ. 1997. Radiocarbon dates from the Oxford Ams System: archaeometry datelist 23. Archaeometry. 39:247–262.
Henke J, Henke L, Chatthopadhyay P, Kayser M, Dülmer M, Cleef S, Pöche H, Felske-Zech H. 2001. Application of Y-chromosomal STR haplotypes to forensic genetics. Croatian Med J. 42:292–297.
Hewitt R, Krause A, Goldman A, Campbell G, Jenkins T. 1996. β-globin haplotype analysis suggests that a major source of Malagasy ancestry is derived from Bantu-speaking Negroids. Am J Hum Genet. 58:1303–1308.
Hill C, Soares P, Mormina M, et al. (12 co-authors). 2006. Phylogeography and ethnogenesis of aboriginal Southeast Asians. Mol Biol Evol. 23:2480–2491.
Hill C, Soares P, Mormina M, et al. (11 co-authors). 2007. A mitochondrial stratigraphy for island southeast Asia. Am J Hum Genet. 80:29–43.
Howell N, Smejkal CB, Mackey DA, Chinnery PF, Turnbull DM, Herrnstadt C. 2003. The pedigree rate of sequence divergence in the human mitochondrial genome: there is a difference between phylogenetic and pedigree rates. Am J Hum Genet. 72:659–670.
Hu SP. 2006. Polymorphism of Y-chromosomal STR haplotypes in the Chaoshan Han Chinese in South China. Forensic Sci Int. 158:80–85.
Huang TY, Hsu YT, Li JM, Chung JH, Shun CT. 2008. Polymorphism of 17 Y-STR loci in Taiwan population. Forensic Sci Int. 174:249–254.
Hurles ME, Nicholson J, Bosch E, Renfrew C, Sykes BC, Jobling MA. 2002. Y chromosomal evidence for the origins of oceanic-speaking peoples. Genetics. 160:289–303.
Hurles ME, Sykes BC, Jobling MA, Forster P. 2005. The dual origin of the Malagasy in Island Southeast Asia and East Africa: evidence from maternal and paternal lineages. Am J Hum Genet. 76:894–901.
Jackson BA, Wilson JL, Kirbah S, Sidney SS, Rosenberger J, Bassie L, Alie JA, McLean DC, Garvey WT, Ely B. 2005. Mitochondrial DNA genetic diversity among four ethnic groups in Sierra Leone. Am J Phys Anthropol. 128:156–163.
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.
Kent R. 1962. From Madagascar to the Malagasy Republic. New York: F.A. Praeger.
Kent R. 1970. Early kingdoms in Madagascar 1500–1700. New York: Holt, Rinehart and Winston.
Kido A, Dobashi Y, Susukida R, Fujitani N, Hara M, Masakazu O. 2005. Y-chromosomal STR haplotypes in Indonesians. J Forensic Sci. 50:962–964.
Kivisild T, Reidla M, Metspalu E, Rosa A, Brehm A, Pennarun E, Parik J, Geberhiwot T, Usanga E, Villems R. 2004. Ethiopian mitochondrial DNA heritage: tracking gene flow across and around the gate of tears. Am J Hum Genet. 75:752–770.
Kivisild T, Tolk HV, Parik J, Wang Y, Papiha S, Bandelt HJ, Villems R. 2002. The emerging limbs and twigs of the East Asian mtDNA tree. Mol Biol Evol. 19:1737–1751.
Knight A, Underhill PA, Mortensen HM, Zhivotovsky LA, Lin AA, Henn BM, Louis D, Ruhlen M, Mountain JL. 2003. African Y chromosome and mtDNA divergence provides insight into the history of click languages. Curr Biol. 13:464–473.
Krings M, Salem AE, Bauer K, et al. (13 co-authors). 1999. mtDNA analysis of Nile River Valley populations: a genetic corridor or a barrier to migration? Am J Hum Genet. 64:1166–1176.
Kwak KD, Jin HJ, Shin DJ, Kim JM, Roewer L, Krawczak M, Tyler-Smith C, Kim W. 2005. Y-chromosomal STR haplotypes and their applications to forensic and population studies in east Asia. Int J Legal Med. 119:195–201.
Leat N, Benjeddou M, Davison S. 2004. Nine-locus Y-chromosome STR profiling of Caucasian and Xhosa populations from Cape Town South Africa. Forensic Sci Int. 144:73–75.
Lecerf M, Filali M, Grésenguet G, Ndjoyi-Mbiguino A, Le Goff J, de Mazancourt P, Bélec L. 2007. Allele frequencies and haplotypes of eight Y-short tandem repeats in Bantu population living in Central Africa. Forensic Sci Int. 171:212–215.
Lejju BJ, Robertshaw P, Taylor D. 2006. Africa's earliest bananas? J Arch Sci. 33:102–113.
Li KL, Kai L, Yuming M. 2007. Y chromosome STR haplotypes of Tibetan Living Tibet Lassa. Forensic Sci Int. 172:79–83.
Luis JR, Rowold DJ, Regueiro M, Caeiro B, Cinnioğlu C, Roseman C, Underhill PA, Cavalli-Sforza LL, Herrera RJ. 2004. The Levant versus the Horn of Africa: evidence for bidirectional corridors of human migrations. Am J Hum Genet. 74:532–544.
MacPhee R, Burney D. 1991. Dating of modified femora of extinct dwarf Hippopotamus from southern Madagascar: implications for constraining human colonization and vertebrate extinction events. J Archaeol Sci. 18:695–706.
Magnavita C. 2006. Ancient humped cattle in Africa: a view from the Chad Basin. African Archaeol Rev. 23:55–84.
Mateu E, Comas D, Calafell F, Pérez-Lezaun A, Abade A, Bertranpetit J. 1997. A tale of two islands: population history and mitochondrial DNA sequence variation of Bioko and São Tomé Gulf of Guinea. Ann Hum Genet. 61:507–518.
Mbida C, van Neer W, Doutrellepont H, Vrydaghs L. 2000. Evidence for banana cultivation and animal husbandry during the first millennium B.C. in the forest of Southern Cameroon. J Arch Sci. 27:151–162.
Mellars P. 2006. Why did modern human populations disperse from Africa ca 60000 years ago? A new model. Proc Natl Acad Sci USA. 103:9381–9386.
Metspalu M, Kivisild T, Metspalu E, et al. (16 co-authors). 2004. Most of the extant mtDNA boundaries in south and southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans. BMC Genet. 5:26.
Michalakis Y, Excoffier L. 1996. A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. Genetics. 142: 1061–1064.
Morwood MJ, Soejono RP, Roberts RG, et al. (14 co-authors). 2004. Archaeology and age of a new hominin from Flores in eastern Indonesia. Nature. 431:1087–1091.
Nagy M, Henke L, Henke J, Chatthopadhyay PK, Völgyi A, Zalán A, Peterman O, Bernasovská J, Pamjav H. 2007. Searching for the origin of Romanies: slovakian Romani Jats of Haryana and Jat Sikhs Y-STR data in comparison with different Romani populations. Forensic Sci Int. 169:19–26.
Nei M. 1987. Molecular evolutionary genetics. New York: Columbia University Press.

Nei M, Li WH. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci. 76:5269–5273.

Pawley A. 2002. Examining the farming/language dispersal hypothesis. Cambridge (United Kingdom): McDonald Institute for Archaeological Research. p. 251–274.

Pereira L, Macaulay V, Torroni A, Scozzari R, Prata MJ, Amorim A. 2001. Prehistoric and historic traces in the mtDNA of Mozambique: insights into the Bantu expansions and the slave trade. Ann Hum Genet. 65:439–458.

Perez VR, Burney DA, Godfrey LR, Nowak-Kemp M. 2003. Butchered sloth lemurs. Evol Anthropol. 12:260.

Pinto F, González AM, Hernández M, Larruga JM, Cabrera VM. 1996. Genetic relationship between the Canary Islanders and their African and Spanish ancestors inferred from mitochondrial DNA sequences. Ann Hum Genet. 60:321–330.

Piry S, Alapetite A, Cornouet JM, Paetkau D, Baudouin L, Estoup A. 2004. GeneClass2: a software for genetic assignment and first-generation migrant detection. J Hered. 95:536–539.

Plaza S, Salas A, Calafell F, Côrte-Real F, Bertranpetit J, Carracedo A, Comas D. 2004. Insights into the western Bantu dispersal: mtDNA lineage analysis in Angola. Hum Genet. 115:439–447.

Rando JC, Pinto F, González AM, Hernández M, Larruga JM, Cabrera VM, Bandelt HJ. 1998. Mitochondrial DNA analysis of northwest African populations reveals genetic exchanges with European near-eastern and sub-Saharan populations. Ann Hum Genet. 62:531–550.

Rannala B, Mountain JL. 1997. Detecting immigration by using multilocus genotypes. Proc Natl Acad Sci USA. 94:9197–9201.

Rosa A, Ornelas C, Brehm A, Villems R. 2006. Population data on 11 Y-chromosome STRs from Guiné-Bissau. Forensic Sci Int. 157:210–217.

Salas A, Richards M, De la Fé T, Lareu MV, Sobrino B, Sánchez-Diz P, Macaulay V, Carracedo A. 2002. The making of the African mtDNA landscape. Am J Hum Genet. 71:1082–1111.

Salas A, Richards M, Lareu MV, Scozzari R, Coppa A, Torroni A, Macaulay V, Carracedo A. 2004. The African diaspora: mitochondrial DNA and the Atlantic slave trade. Am J Hum Genet. 74:454–465.

Schraeder PJ. 1995. Madagascar. In: Metz HC, editor. A country study: Madagascar. Washington (DC): GPO for the Library of Congress.[Internet]. [cited 2009 May 7]. Available from: http://www.loc.gov.

Sengupta S, Zhivotovsky LA, King R, Mehdi SQ, et al. (14 co-authors). 2006. Polarity and temporality of high-resolution y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of Central Asian pastoralists. Am J Hum Genet. 78:202–221.

Singh A, Trivedi R, Kashyap VK. 2006. Haplotype diversity at 20 Y-chromosome microsatellites in an Indo-Caucasian population of Bengal India. J Forensic Sci. 51:1434–1437.

Soodyall H, Jenkins T, Stoneking M. 1995. 'Polynesian' mtDNA in the Malagasy. Nat Genet. 10:377–378.

Souto L, Gusmaõ L, Ferriera E, Amorim A, Côrte-Real F, Vieira DN. 2006. Y-chromosome STR haplotypes in East Timor: forensic evaluation and population data. Forensic Sci Int. 156:261–265.

Spriggs M. 2003. Chronology of the Neolithic transition in Island Southeast Asia and the western Pacific: a view from 2003. Rev Archaeol. 24:57–80.

Tang JS, Wong HY, Syn CK, Tan-Siew WF, Chow ST, Budowle B. 2006. Population study of 11 Y-chromosomal STR loci in Singapore Chinese. Forensic Sci Int. 158:65–71.

Trejaut JA, Kivisild T, Loo JH, Lee CL, He CL, Hsu CJ, Lee ZY, Lin M. 2005. Traces of archaic mitochondrial lineages persist in Austronesian-speaking Formosan populations. PLoS Biol. 3:e247.

Trovoada MJ, Pereira L, Gusmão L, Abade A, Amorim A, Prata MJ. 2004. Pattern of mtDNA variation in three populations from São Tomé e Príncipe. Ann Hum Genet. 68:40–54.

Van, Oven M, Kayser M. 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum Mutat. 30:E386–E394.

Vigilant L, Pennington R, Harpending H, Kocher TD, Wilson AC. 1989. Mitochondrial DNA sequences in single hairs from a southern African population. Proc Natl Acad Sci USA. 86:9350–9354.

Wang X, Sawaguchi T. 2006. Analysis of Y-STR loci in a population sample from northeast China. J Forensic Sci. 51:195–198.

Watson E, Bauer K, Aman R, Weiss G, von Haeseler A, Pääbo S. 1996. mtDNA sequence diversity in Africa. Am J Hum Genet. 59:437–444.

Watson E, Forster P, Richards M, Bandelt HJ. 1997. Mitochondrial footprints of human expansions in Africa. Am J Hum Genet. 61:691–704.

Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. Evolution. 38:1358–1370.

Wong HY, Tang JS, Budowle B, Allard MW, Syn CK, Tan-Siew WF, Chow ST. 2007. Sequence polymorphism of the mitochondrial DNA hypervariable regions I and II in 205 Singapore Malays. Leg Med Tokyo. 9:33–37.

Yang B, Gu M, Wang G, Li X, Liu Y, Yang W. 2006. Population data for 11 Y-chromosome STRs in northeast China Han. Forensic Sci Int. 164:65–71.

Yeung SM, Wong LM, Cheung BK, To KY. 2006. Allele frequencies and haplotypes of 12 Y-STR loci for the local Chinese population in Hong Kong. Forensic Sci Int. 162:55–63.

Yong RY, Lee LK, Yap EP. 2006. Y-chromosome STR haplotype diversity in three ethnic populations in Singapore. Forensic Sci Int. 159:244–257.

Zhu B, Deng Y, Zhang F, et al. (12 co-authors). 2006. Genetic analysis for Y chromosome short tandem repeat haplotypes of Chinese Han population residing in the Ningxia province of China. J Forensic Sci. 51:1417–1420.

Zhu B, Liu S, Ci D, et al. (11 co-authors). 2006. Population genetics for Y-chromosomal STRs haplotypes of Chinese Tibetan ethnic minority group in Tibet. Forensic Sci Int. 161:78–83.

Zhu B, Shen C, Qian G, et al. (11 co-authors). 2005. Genetic polymorphisms for 11 Y-STRs haplotypes of Chinese Yi ethnic minority group. Forensic Sci Int. 158:229–233.

Zhu BF, Wang ZY, Yang CH, Li XS, Zhu J, Yang G, Huang P, Liu Y. 2005. Y-chromosomal STR haplotypes in Chinese Uigur ethnic group. Int J Legal Med. 119:306–309.

Connie Mulligan, Associate Editor

Accepted June 11, 2009