npg

ARTICLE

Complete mitochondrial DNA sequences provide new insights into the Polynesian motif and the peopling of Madagascar

Harilanto Razafindrazaka[1,6], François-X Ricaut[*,1,6], Murray P Cox[2], Maru Mormina[3], Jean-Michel Dugoujon[1], Louis P Randriamarolaza[4], Evelyne Guitard[1], Laure Tonasso[1], Bertrand Ludes[1,5] and Eric Crubézy[1]

More than a decade of mitochondrial DNA (mtDNA) studies have given the 'Polynesian motif' renowned status as a marker for tracing the late-Holocene expansion of Austronesian speaking populations. Despite considerable research on the Polynesian motif in Oceania, there has been little equivalent work on the western edge of its expansion – leaving major issues unresolved regarding the motif's evolutionary history. This has also led to considerable uncertainty regarding the settlement of Madagascar. In this study, we assess mtDNA variation in 266 individuals from three Malagasy ethnic groups: the Mikea, Vezo, and Merina. Complete mtDNA genome sequencing reveals a new variant of the Polynesian motif in Madagascar; two coding region mutations define a Malagasy-specific sub-branch. This newly defined 'Malagasy motif' occurs at high frequency in all three ethnic groups (13–50%), and its phylogenetic position, geographic distribution, and estimated age all support a recent origin, but without conclusively identifying a specific source region. Nevertheless, the haplotype's limited diversity, similar to those of other mtDNA haplogroups found in our Malagasy groups, best supports a small number of initial settlers arriving to Madagascar through the same migratory process. Finally, the discovery of this lineage provides a set of new polymorphic positions to help localize the Austronesian ancestors of the Malagasy, as well as uncover the origin and evolution of the Polynesian motif itself.
European Journal of Human Genetics (2010) **18**, 575–581; doi:10.1038/ejhg.2009.222; published online 23 December 2009

INTRODUCTION

The 'Polynesian motif', popularly named for its high frequency among Polynesians, is characterized by a well known series of mitochondrial DNA (mtDNA) polymorphisms that now define haplogroup B4a1a1[1,2] (A14022G, T16217C, A16247G, and C16261T[1,3,4]). This lineage probably developed in eastern Island Southeast Asia or Near Oceania[1,2] during the mid- to late-Holocene, with recent dates suggesting an origin around 6200–10 900 years before present (YBP) (95% confidence intervals: 2650–18 800 YBP).[5] The haplogroup's immediate precursor (lacking only A16247G) has been found in Taiwanese aboriginal groups with an estimated age of 13 200 YBP (95% confidence interval: 9400–17 000).[1] This incremental series of dates are consistent with a model whereby Austronesian speaking populations expanded out of Taiwan during the mid- to late-Holocene (but see ongoing discussions surrounding this model[2,6]). Ultimately, the Austronesian expansion spread the immediate ancestor of the Polynesian motif, and later the motif itself, over a vast geographical area – from Taiwan in the north, New Zealand in the south, remote Polynesia in the east, and finally, Madagascar in the far west.[2,5,7–10]

The Polynesian motif is currently found at highest frequency in Polynesia, where it approaches fixation in some populations.[4,11] It is also common in Micronesia and parts of Near Oceania,[2,3,5,12–18] where

it is not necessarily restricted to Austronesian speaking populations, but also occurs in some rare Papuan speaking groups.[2,18,19] The motif is much less frequent in Island Southeast Asia, although it has been found sporadically in both central and eastern Indonesia.[4,6,7,11,20,21] In Madagascar – the western edge of the Austronesian expansion – the Polynesian motif reaches a frequency of around 20%,[8,9,20,22] thus leading to proposals that the island was settled by an Indonesian population, which later colonized the Pacific Islands,[8,22] or even more speculatively, by direct migration from Polynesia itself.[22] However, modern Malagasy should carry both maternal and paternal lineages that trace back to their ancestral population(s), and the latter hypothesis was discounted when Malagasy where shown not to carry the predominant Y chromosome haplogroups found in Polynesia (eg, C and O3).[9,20] Furthermore, these studies revealed that Indonesians have a major role in the colonization of Madagascar, and highlighted Borneo as a likely source of the Asian-derived Y chromosomes found in Malagasy today. This is consistent with linguistic evidence suggesting that the Malayo–Polynesian language spoken by Malagasy is related to the Barito language of southern Borneo.[23–25] Currently, our best model for the settlement of Madagascar suggests that the first settlers reached the island ∼1500–2000 years ago, when there is clear archeological and paleoecological evidence of their

[1]CNRS FRE 2960, Laboratoire d'Anthropobiologie, Université de Toulouse, Toulouse III Paul Sabatier, Toulouse, France; [2]The Institute of Molecular BioSciences, the Allan Wilson Centre for Molecular Ecology and Evolution, and the Bio-Protection Centre, Massey University, Palmerston North, New Zealand; [3]Leverhulme Centre for Human Evolutionary Studies, University of Cambridge, Cambridge, UK; [4]Laboratoire d'Anthropologie Culturelle, Université Antananarivo, Antananarivo, Madagascar; [5]Laboratoire d'Anthropologie Moléculaire, Institut de Médecine Légale, Strasbourg, France
*Correspondence: Dr F-X Ricaut, Laboratoire d'Anthropobiologie, CNRS FRE 2960, Université de Toulouse, 37 allées Jules Guesdes, Toulouse cedex 3 31073, France. Tel: +33 05 61 55 80 84; Fax: +33 05 61 14 59 79; E-mail: fx.ricaut@infonie.fr
[6]These authors contributed equally to this work.
Received 16 April 2009; revised 2 November 2009; accepted 26 November 2009; published online 23 December 2009

occupation.[26,27] Ultimately, a complex – and largely unknown – genetic and linguistic admixture process between populations of African and Southeast Asian descent produced the Malagasy we recognize today.[8,9,20,23,28]

However, several outstanding questions remain. What is the spatial and temporal origin of the Polynesian motif in Island Southeast Asia? What is the manner of its arrival in Madagascar? And what is its distribution on the island? To address these questions, we analyzed Polynesian motif carriers in Madagascar using complete mtDNA sequencing in the largest Malagasy sample available to date.

MATERIALS AND METHODS

Samples

The samples analyzed in this study form part of our Madagascar assemblage collected in field seasons 2007–2008. They comprise buccal cell and peripheral blood samples collected from unrelated individuals in EDTA Vaccutainer tubes. Information on survey subjects includes languages spoken, current residence, familial birthplaces, and a short genealogy of four generations to establish regional ancestry. A total of 266 DNA samples were analyzed here, which includes individuals from three ethnic groups: 127 Mikea (hunter–gatherers in the southwest), 101 Vezo (semi-nomadic fisherman also in the southwest), and 38 Andriana Merina (individuals from the central highlands). Culturally, the Andriana Merina who are strongly endogamous, have been identified as being the primary descendents of Island Southeast Asian migrants.[29,30] All samples were obtained with informed consent, and this study was approved by the appropriate ethical committees both within Madagascar, and at the University of Toulouse.

DNA extraction, amplification, and sequencing

DNA was extracted from cheek swabs and blood samples using a standard phenol–chloroform protocol, followed by purification with CleanMix (Talent, Trieste, Italy) according to the manufacturer's instructions.

The mtDNA analysis was carried out in four phases (see Table 1 for summary):

(i) Hypervariable segments 1 and 2 (HVS1/HVS2) of the mtDNA control region were sequenced first. This region was amplified using primers L15973 (5′-AACTCCACCATTAGCACCCA-3′) and H296 (5′-TCTGT AGTATTGTTTTTAAAGG-3′). PCR amplification was carried out in 50 μl of reaction mixture containing 3 mM of MgCl$_2$, 0.2 mM of dNTPs, 0.2 μM of each primer, 1× AmpliTaq Gold reaction buffer (PE Applied Biosystems, Foster City, CA, USA), 0.5 U of Taq HotGoldstar DNA polymerase (PE Applied Biosystems), and 1 μl of DNA extract. Amplification was performed using a T3 Thermocycler (Biometra, Archamps, France). Thermal cycling conditions were: pre-denaturation at 95 °C for 10 min; followed by 35 cycles at 95 °C for 60 s, 58 °C for 60 s, and 72 °C for 90 s; and final extension at 72 °C for 5 min. PCR products were visualized on a 2% agarose gel, and purified with QIAquick (Qiagen GmbH, Hilden, Germany). Sequencing reactions were carried out on both strands with ABI Prism BigDye Terminator v3.1 Cycle Sequencing (PE Applied Biosystems) using the same primers used for PCR amplification. Sequencing reaction products were analyzed on an ABI Prism 3730 Genetic Analyser (PE Applied Biosystems). For samples not definitively assigned to a known haplogroup from HVS1 and HVS2 sequences, phylogenetic status was clarified by screening RFLPs known to identify haplogroups L3 (−10 871 MnlII), M (+10 397 AluI; +10 394 DdeI), N (−10 397 AluI; −10 394 DdeI), M7 (+9824 HinfI), E (−7598 HhaI), F (−10 306 BspM1), and F3 (+10 319 Tsp509I). Finally, haplogroup M7c3 was determined by direct sequencing of nucleotide position A3606G.

(ii) For samples carrying control region mutations characteristic of the Polynesian motif (ie, nucleotides 16217C, 16247G and 16261T), or the immediate Polynesian motif ancestor (ie, nucleotides 16217C and 16261T), two additional diagnostic mutations were analyzed to confirm affiliation to haplogroup B4a1a1. First, the presence of the 9-base pair (bp) intergenic region V deletion[31] was determined by amplifying

Table 1 Strategy followed for sample analysis, and diversity of the lineages affiliated to haplogroup B4a1a1a and their distribution in three Malagasy populations

Population (n)	Haplotype	n	%	First		Second			Third[a]	Fourth[b]		Final haplogroup
										−1473	−3423	
				HVS1	HVS2	9-bp del	+6719 NlaIII	A14022G	Whole mtDNA sequence	HhaI	AcII	
Merina (38)	H1	17	44.7	16182C, 16183C, 16189, 16217, 16247, 16261, 16519	73 146 263	+	+			+	+	B4a1a1a Malagasy motif
	H2	2	5.3	16182C, 16183C, 16189, 16217, 16247, 16261, 16357	73 146 263	+	+		+	+	+	B4a1a1a Malagasy motif
Vezo (101)	H1	18	17.8	16182C 16183C 16189 16217 16247 16261 16519	73 146 263	+	+			+	+	B4a1a1a Malagasy motif
	H3	1	1	16182C, 16183C, 16189, 16217, 16247, 16261, 16368	73 146 263	+	+		+	+	+	B4a1a1a Malagasy motif
	H4	3	3	16182C, 16183C, 16189, 16217, 16261, 16519	73 146 263	+	+	+		+	+	B4a1a1a Malagasy motif
Mikea (127)	H1	17	13.4	16182C, 16183C, 16189, 16217, 16247, 16261, 16519	73 146 263	+	+		+	+	+	B4a1a1a Malagasy motif

[a]Sequencing of the complete mtDNA genome was performed using one sample from each different mtDNA haplotypes affiliated to B4a1a1a haplogroup. Furthermore, one sample from each of the three Malagasy groups was used for complete sequencing (haplotype H1 from Mikea; haplotype H2 from Merina; and haplotype H3 from Vezo).
[b]Two coding region mutations (C1473T and T3423A) revealed by complete mtDNA sequences are shared by all the Malagasy samples affiliated to haplogroup B4a1a1a and consequently define the new 'Malagasy motif'.

a fragment of ~120 bp, including mtDNA region V, using primers L8196 and H8297.[32] PCR conditions were the same as for the control region analysis, and PCR products were visualized on a 2% agarose gel. Second, an informative RFLP site (+6719 NlaIII) was screened using primers F6120 and R7013 as described by Rieder et al.[33] Fragments were visualized on a 2% agarose gel. In addition, because reversion at the hypermutable nucleotide 16 247 sometimes leads to incorrect haplogroup assignment of samples affiliated to B4a1a1a,[2,5] we verified that all samples harboring nucleotides 16217C and 16261T have a transition at 14 022 that defines haplogroup B4a1a1.[1] We applied a mini-sequencing strategy using SNaPshot mini-sequencing reactions (PE Applied Biosystems), following a previous published protocol.[34] We designed primers F13957 (5′-GGCCTTCTTACGAGCCAAAA-3′) and R14257 (5′-TATTGGTGCGGGGGCTTTGTATAA-3′), and performed a Touchdown PCR: seven cycles at 94°C for 20 s, 62°C for 30 s, and 70°C for 30 s with annealing temperature decreasing 1°C per cycle; followed by 31 cycles with an annealing temperature at 55°C. A final extension period of 5 min at 72°C was also applied.

(iii) Complete mtDNA genomes of one sample from each different mtDNA haplotype affiliated to B4a1a1a were then sequenced. One sample from each of the three Malagasy groups was chosen for geographical coverage. Complete sequencing was performed by the Genomic Analysis Technology Core at the University of Arizona (http://gatc.arl.arizona.edu/) using 28 pairs of primers, which allowed amplification and sequencing of overlapping fragments for both forward and reverse DNA strands (http://bcf.arl.arizona.edu/). Sequences were edited and aligned against the revised Cambridge reference sequence (rCRS[35]) using BioEdit 7.0.9,[36] and deviations from the rCRS were confirmed by manual checking of electropherograms.

(iv) New diagnostic mutations ascertained through complete mtDNA sequencing were screened for all Malagasy samples assigned to the B4a1a1a haplogroup. RFLP screening of these new diagnostic mutations was performed for nucleotides 1473 (−1473 HhaI; primers F1166 and R1607), and 3423 (−3423 AciI; primers F3200 and R3693), using the methodology described by Torroni et al.[37] The resulting fragments were visualized on a 2% agarose gel.

Sequences were classified into mtDNA haplogroups based on accepted nomenclature (eg, publications[1,2,5,6,17,38–42]).

Statistical analysis: phylogeny reconstruction and age estimation

A mtDNA phylogeny was constructed from 18 Polynesian B4a1a1a complete mtDNA sequences, including all previously published sequences ($n=15^{5,43–46}$), and the three complete Malagasy sequences obtained in this study.

Time estimates were calculated in a number of different ways. First, using the rho (ρ) statistic[47] with three previously described mutation rates based on coding region mutations. Two mutation rates were calculated using estimated substitution rates for protein-coding synonymous changes of 3.5×10^{-8} mutations/site/year, which yields 6764 years per synonymous transition,[40] and of one synonymous mutation (ie, transition or transversion) in every 7884 years from the recently improved mitochondrial molecular clock published by Soares and colleagues.[48] The MAMMAG website (http://mammag.web.uci.edu/bin/view/Main/WebHome) was used to determine synonymous transitions. The third mutation rate was based on substitutions for the entire coding region; 1.26×10^{-8} mutations/site/year, which yields 5139 years per mutation between positions 577 and 16 023 of the rCRS.[49] Dates estimated from synonymous changes are presumed to be the most robust, as these changes are more likely to be selectively neutral.[40] The variances of these rho-based dating estimates were calculated as per Saillard et al.[47]

In addition, in some cases the only mutations that contribute to the age estimate of a group of sequences are located in the control region of the mtDNA and are not taken into account by the previous estimates based on coding region mutations. Therefore, we also calculated age estimates from the control region using (i) the most widely used mutation rate for HVS1 of 1.80×10^{-8} transitions/site/year or one mutation per 20 180 years for the region between positions 16 090 and 16 365,[50] and (ii) the recently improved mutation

rates by Soares et al.[48] of one mutation every 18 845 years for the HVS1 segment (positions 16 090–16 365); one mutation every 16 677 years for the HVS1 segment (positions 16 051–16 400) and one mutation every 9058 years for the whole control region.

We also applied a second dating method broadly based on the coalescent concept of waiting time. We assumed a mutation rate of 1.16×10^{-4} per year over the hypervariable region from nucleotides 16 024–00 300, which yields one mutation approximately every 8642 years.[40] Using custom simulation code written in R, we calculated the waiting time to observe three lineages – two each carrying a single new mutation, and one carrying no new mutations. Our simulation returned a likelihood surface from which we determined a best estimate of the TMRCA together with 95% confidence intervals. It is worth noting that because of the ongoing debate regarding the true mutation rate,[40,51] limitations of dating methods – including the rho (ρ) statistic,[52] the conversion of molecular dates into chronological dates, and the estimation of associated error values – the values provided here are intended only as approximations. All molecular dates should be interpreted cautiously. We intend these dates only to be used as ballpark measures.

RESULTS

Analysis of mtDNA from 266 Malagasy individuals (Supplementary Table S1) is broadly consistent with previous genetic studies.[8,9,20,22] We see a combination of Southeast Asian and African lineages that can be linked to settlement of the island around 1500–2000 years ago. We observed the Polynesian motif at relatively high frequency in all three Malagasy groups: 50.0% in Merina, 21.8% in Vezo, and 13.4% in Mikea (Table 1). Indeed, the first and second phases of our analysis revealed that 58 of the 266 Malagasy shared a set of mutations (9-bp deletion, 6719C, 14022G, 16217C, and 16261T), which assign them to haplogroup B4a1a1. Although most sequences ($n=55$) also harbored the HVS1 transition at nucleotide 16 247, which traditionally defines the Polynesian motif, three other individuals who lacked this mutation at 16 247 could also be affiliated to the lineage after other analyses were completed (Table 1). The high incidence of the Polynesian motif in the Merina central highlanders can be partly explained by the high endogamy of this group; its lower frequency in the other two groups (Vezo and Mikea) is in the range observed in previous studies.[8,9,20]

The diversity of the Polynesian motif sequences is low, as only four different haplotypes (based on control region mutations) have been observed among the three Malagasy groups (Table 1). Haplotype H1, which represents the root of the Polynesian motif, is the only haplotype shared by all three groups, and is also the most frequent haplotype in each group (44.7% in Merina, 17.8% in Vezo, and 13.4% in Mikea). The other three haplotypes differ by one mutation from haplotype H1, and each is specific to a single ethnic group (H2 is only observed at 5.3% in Merina, and H3 and H4 only at 1 and 3% in Vezo).

Complete mtDNA sequencing of the three haplotypes (H1, H2, and H3) that carry the full Polynesian motif, and phylogenetic analysis with the 15 published Polynesian motif complete mtDNA sequences, reveals an interesting substructure for this group (Figure 1). A major subgroup within the Polynesian motif lineage can be defined using two coding region mutations at nucleotides 1473 and 3423. These two polymorphisms are shared by all 58 Malagasy samples belonging to haplogroup B4a1a1a, and thus define a 'Malagasy motif'. Although only a few complete mtDNA genomes are available from Melanesia, Polynesia, and Island Southeast Asia, these additional mutations have not been found in any other B4a1a1a individuals sequenced to date (Figure 1; Martin Richards and Pedro Soares, Institute of Integrative and Comparative Biology, University of Leeds, UK; personal communication). The TMRCA of this 'Malagasy motif' has been estimated at 6000 YBP (95% confidence interval: 0–14 300) using a recently
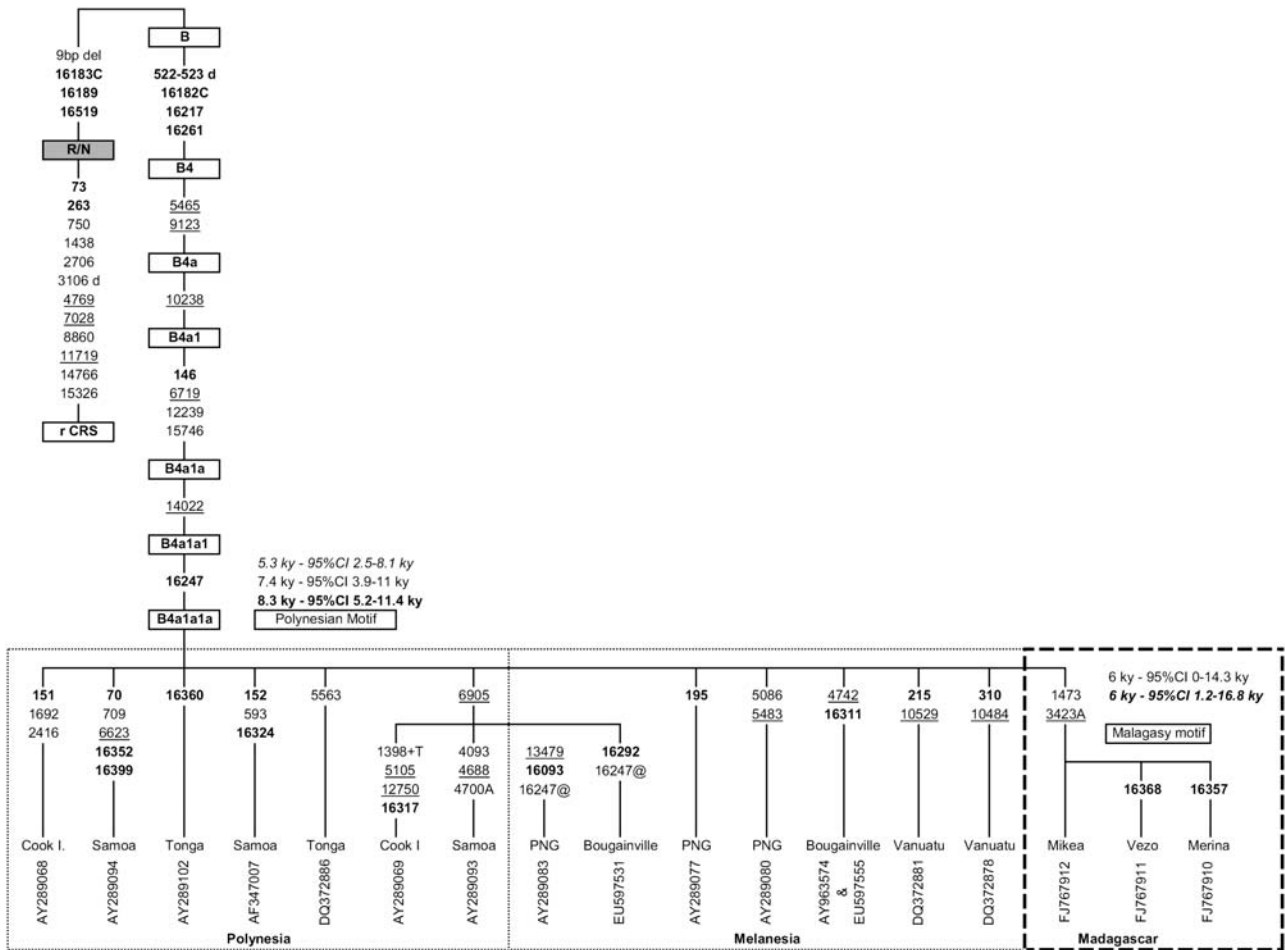
npg

Figure 1 Phylogenetic tree constructed from complete mtDNA sequences from 3 Malagasy individuals and 15 previously published samples. Mutations were scored relative to the rCRS.[34] Length variations in the poly-C region from nucleotides 303–315 are not shown. Numbers along links refer to nucleotide positions. Suffixes A, C, G, and T indicate transversions; 'd' signifies a deletion, a plus sign (+) an insertion. Recurrent mutations in the phylogeny are underlined. The prefix '@' indicates back mutations. Control region mutations are shown in bold, and synonymous transitions are underlined. TMRCA estimates, calculated as per references[40,48,49] and using the waiting time dating method are presented in italic, bold, regular, and bold italic, respectively. Values are given in thousands of years before present. The 15 non-Malagasy complete mtDNA sequences used, in addition to the three Malagasy complete mtDNA sequences analyzed in this study, were reported by Hartmann et al.[46] (EU597531 and EU597555), Ingman and Gyllensten[44] (AY289068, AY289094, AY289102, AY289069, AY289093, AY289083, AY289077, and AY289080), Ingman et al.[43] (AF347007), Pierson et al.[5] (DQ372886, DQ372881, and DQ372878), and Macaulay et al.[45] (AY963574).

improved control region mutation rate[48] (Table 2), in broad agreement with dates obtained from other control region mutation rates (Supplementary Table S2). Using an alternative approach based on waiting times, we also infer a date of 6000 YBP (95% confidence interval: 1200–16 800) (Supplementary Figure S1). Remembering the limitations of molecular dating, and the fact that any demographic expansion may have predated geographic expansion, we only claim that these dates are broadly consistent with archeological and paleoecological estimates for the first settlement of Madagascar around 1500–2000 years ago.[26,27]

DISCUSSION

Sampling three very different Malagasy ethnic groups (coastal fishermen, forest hunter–gatherers, and traditionally agriculturalist highlanders), together with HVS1/HVS2 sequencing, whole mtDNA sequencing and RFLP typing for several Polynesian motif lineages, provides important new information regarding the settlement of

Madagascar. Our study reveals that (i) between 13 and 50% of all Malagasy lineages belong to the Polynesian motif B4a1a1a haplogroup, whereas none belong to its immediate precursor, or to other B subhaplogroups; (ii) all Polynesian motif B4a1a1a sequences on Madagascar share two coding region mutations that together define a 'Malagasy motif'; (iii) the low genetic diversity of this motif (based on control region mutations) is consistent with genetic patterns expected if the island were colonized only recently;[9,40] and (iv) the 'Malagasy motif' founding sequence (defined by mutations 1473 and 3423) is distributed across all three Malagasy ethnic groups, while derived sequences occur only in individual groups. This suggests that the founding haplotype of the 'Malagasy motif' expanded in a relatively short time after its appearance in Madagascar. Subsequently, the lineage has accumulated distinct mutations in each Malagasy population, which is consistent with long-standing isolation among these groups. Interestingly, a similar pattern is observed for all the Malagasy mtDNA haplogroups originating from the eastern part of

Table 2 Age estimates for B4a1a1 haplogroup, Polynesian motif, and Malagasy motif using the rho (ρ) statistic dating method

Haplogroup	n	Coding region nt 577–16 023					Synonymous mutations					Control region nt 16 024–576				
		ρ	σ	TMRCA years	95% CI	Mutation rate	ρ	σ	TMRCA years	95% CI	Mutation rate	ρ	σ	TMRCA years	95% CI	Mutation rate
B4a1a1	23	1.57	0.26	8.100	5400–10800	(1)	0.70	0.17	4.700	2500–7000	(2)	1.57	0.26	14.200	9600–18800	(4)
							0.83	0.19	6.500	3600–9500	(3)	—	—	—	—	—
Polynesian motif	18	1.61	0.30	8.300	5200–11400	(1)	0.78	0.21	5.300	2500–8100	(2)	0.89	0.22	8.000	4200–12000	(4)
							0.94	0.23	7.400	3900–11000	(3)	—	—	—	—	—
Malagasy motif	3	0	—	—	—	—	0	—	—	—	—	0.66	0.47	6.000	0–14300	(4)

Note. Sequences used for molecular dating of the Malagasy motif and the Polynesian motif are given in Figure 1. The sequences used for molecular dating of the B4a1a1 haplogroup included, in addition to the 15 complete mtDNA sequences harboring the Polynesian motif and the three sequences harboring the Malagasy motif reported in Figure 1, five complete mtDNA sequences reported by Pierson *et al*.[5] (DQ372873, DQ372874, DQ372875, and DQ372877) and Ingman and Gyllensten[44] (AY289076). Mutation rate references: (1) One substitution per 5139 years;[49] (2) One synonymous transition per 6764 years;[40] (3) One synonymous mutation (transition or transversion) per 7884 years;[48] (4) One mutation per 9058 years;[48] Age estimates of the Malagasy motif calculated by using other control region mutation rates are provided in Supplementary Table S2.

the Indian Ocean (F3b, M46, E1a1a, M7c3): a founding sequence for each haplogroup is distributed across all three Malagasy ethnic populations, while derived sequences occur only in individual groups (Supplementary Table S1).

Together, these elements are consistent with some degree of isolation between groups. Although these results are best explained by a relatively small number of initial settlers arriving to Madagascar through the same migratory process, our data alone does not allow us to clearly favor or reject alternative hypotheses regarding the number of waves of migration that reached Madagascar through this process, the source population(s) involved, or the duration of this process (for discussion, see[9,53–55]).

Although all molecular dating is uncertain, the TMRCA of the 'Malagasy motif' is in broad agreement with the archeological time-frame suggested for the settlement of Madagascar.[26,27] However, the lineage's geographic origins warrant further discussion, as they may trace back to three quite different regions: Polynesia, Melanesia, or Island Southeast Asia.

It is unlikely that the Polynesian motif came to Madagascar directly from Polynesia, because Polynesian Y chromosome haplogroups (eg, O3 and C2a1, which predominate in Polynesians[9,16,17,56–58]) have not been found among Malagasy paternal lineages ([9,20] and authors' unpublished data). In contrast, the Y haplogroups O1b and O2a – the former showing its highest frequency in Taiwan and possessing a frequency distribution possibly suggesting dispersal in association with the Austronesian expansion,[17,59] and the latter having its highest frequency in Southeast Asia,[16,60,61] – were observed in Madagascar at moderate frequency ([9] and authors' unpublished data). These same lineages are nearly absent in Polynesia.[16,17,61] Austronesian communities were matriarchal and matrilocal[62] and experienced serious bottlenecks during their expansion process, but this would have most likely only reduced the number and diversity of Y chromosome lineages. Some outside possibilities remain, both rather unlikely. Either the Y chromosome pool of early Polynesian migrants lost these lineages as they migrated to Madagascar, or they are so exceedingly rare in Madagascar that they have not yet been detected ([9] and authors' unpublished data).

Our study partially addresses these latter hypotheses by providing additional mtDNA evidence that does not support Polynesia as a potential ancestral region for the colonization of Madagascar. Indeed, none of the seven published B4a1a1a complete mtDNA sequences from Polynesia (Figure 1) harbor the coding region mutations (nucleotides 1473 and 3423) that all Malagasy B4a1a1a lineages share. However, the same argument holds true for B4a1a1a lineages from Melanesia, and from Island Southeast Asia (Martin Richards and Pedro Soares, personal communication), despite linguistic and Y chromosome evidence[9,20,23,25] pinpointing this latter region as the most likely origin of the 'Asian' migration to Madagascar. Importantly, we acknowledge that the Polynesian motif is extremely rare in Island Southeast Asia[4,6,7,11,20,21] and the relatively small number of complete mtDNA sequences carrying the Polynesian motif, which have been tested for the Malagasy motif are even smaller. Therefore, we cannot yet determine conclusively from which region the Malagasy motif may have originated.

Alternatively, these results may indicate the absence of the Malagasy motif outside Madagascar, which would suggest that it originated *in situ* after the arrival of the Polynesian motif carriers. This hypothesis is still unlikely as it involves that (i) two coding region mutations (nucleotides 1473 and 3423) appeared in Madagascar in the last 1500–2000 years, (ii) diffused across the entire island, including to diverse populations (Mikea hunter–gatherers, Vezo semi-nomadic fisherman and central highlanders), and (iii) the immediate Malagasy motif

precursor (ie, the Polynesian motif) subsequently disappeared from Malagasy populations.

Nevertheless, these speculations are constrained by several factors, such as the effect of genetic drift in small island populations, and by the poor coverage of parental source regions (Island Southeast Asia, Melanesia, and Polynesia). Only analysis of additional B4a1a1a sequences, especially through microgeographic sampling of well-defined populations (eg, linguistically and culturally), can provide more comprehensive insights into the geographic origin of the 'Malagasy motif'.

CONFLICT OF INTEREST
The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

1 Trejaut JA, Kivisild T, Loo JH *et al*: Traces of archaic mitochondrial lineages persist in Austronesian-speaking Formosan populations. *PLoS Biol* 2005; **3**: e247.
2 Friedlaender JS, Friedlaender FR, Hodgson JA *et al*: Melanesian mtDNA complexity. *PLoS ONE* 2007; **2**: e248.
3 Lum JK, Rickards O, Ching C, Cann RL: Polynesian mitochondrial DNAs reveal three deep maternal lineage clusters. *Hum Biol* 1994; **66**: 567–590.
4 Redd AJ, Takezaki N, Sherry ST, McGarvey ST, Sofro AS, Stoneking M: Evolutionary history of the COII/tRNALys intergenic 9 base pair deletion in human mitochondrial DNAs from the Pacific. *Mol Biol Evol* 1995; **12**: 604–615.
5 Pierson MJ, Martinez-Arias R, Holland BR, Gemmell NJ, Hurles ME, Penny D: Deciphering past human population movements in Oceania: provably optimal trees of 127 mtDNA genomes. *Mol Biol Evol* 2006; **23**: 1966–1975.
6 Hill C, Soares P, Mormina M *et al*: A mitochondrial stratigraphy for island Southeast Asia. *Am J Hum Genet* 2007; **80**: 29–43.
7 Cox MP: Indonesian mitochondrial DNA and its opposition to a Pleistocene era origin of proto-Polynesians in island Southeast Asia. *Hum Biol* 2005; **77**: 179–188.
8 Soodyall H, Jenkins T, Stoneking M: Polynesian' mtDNA in the Malagasy. *Nat Genet* 1995; **10**: 377–378.
9 Hurles ME, Sykes BC, Jobling MA, Forster P: The dual origin of the Malagasy in Island Southeast Asia and East Africa: evidence from maternal and paternal lineages. *Am J Hum Genet* 2005; **76**: 894–901.
10 Gray RD, Drummond AJ, Greenhill SJ: Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science* 2009; **323**: 479–483.
11 Richards M, Oppenheimer S, Sykes B: mtDNA suggests Polynesian origins in Eastern Indonesia. *Am J Hum Genet* 1998; **63**: 1234–1236.
12 Melton T, Peterson R, Redd AJ *et al*: Polynesian genetic affinities with Southeast Asian populations as identified by mtDNA analysis. *Am J Hum Genet* 1995; **57**: 403–414.
13 Friedlaender J, Schurr T, Gentz F *et al*: Expanding Southwest Pacific mitochondrial haplogroups P and Q. *Mol Biol Evol* 2005; **22**: 1506–1517.
14 Ohashi J, Naka I, Tokunaga K *et al*: Brief communication: mitochondrial DNA variation suggests extensive gene flow from Polynesian ancestors to indigenous Melanesians in the northwestern Bismarck Archipelago. *Am J Phys Anthropol* 2006; **130**: 551–556.
15 Kayser M, Brauer S, Weiss G *et al*: Reduced Y-chromosome, but not mitochondrial DNA, diversity in human populations from West New Guinea. *Am J Hum Genet* 2003; **72**: 281–302.
16 Kayser M, Brauer S, Cordaux R *et al*: Melanesian and Asian origins of Polynesians: mtDNA and Y chromosome gradients across the Pacific. *Mol Biol Evol* 2006; **23**: 2234–2244.
17 Kayser M, Choi Y, van Oven M *et al*: The impact of the Austronesian expansion: evidence from mtDNA and Y chromosome diversity in the Admiralty Islands of Melanesia. *Mol Biol Evol* 2008; **25**: 1362–1374.
18 Ricaut FX, Thomas T, Arganini C *et al*: Mitochondrial DNA variation in Karkar Islanders. *Ann Hum Genet* 2008; **72**: 349–367.
19 Merriwether DA, Friedlaender JS, Mediavilla J, Mgone C, Gentz F, Ferrell RE: Mitochondrial DNA variation is an indicator of austronesian influence in Island Melanesia. *Am J Phys Anthropol* 1999; **110**: 243–270.
20 Cox MP: *Genetic Patterning at Austronesian Contact Zones*. University of Otago: Dunedin, New Zealand, 2003.
21 Mona S, Grunz KE, Brauer S *et al*: Genetic admixture history of Eastern Indonesia as revealed by Y-chromosome and mitochondrial DNA analysis. *Mol Biol Evol* 2009; **26**: 1865–1877.
22 Soodyall H, Jenkins T, Hewitt R, Krause A, Stoneking M: The peopling of Madagascar; in Boyce J, Mascie-Taylor CGN (eds): *Molecular Biology and Human diversity*. Cambridge: Cambridge University Press, 1996, pp 156–170.
23 Dahl OC: *Malgache et Maanjan. Une comparaison linguistique*. Oslo: Egede-Institutet, 1951, pp 408.
24 Dahl OC: La subdivision de la famille barito et la place du malgache. *Copenhagen, Acta Orientalia* 1977; **38**: 77–134.
25 Blench RDM: The Austronesians in Madagascar and on the East African coast: surveying the linguistic evidence for domestic and translocated animals. Paper presented at the International Conference on Austronesian Languages Palawan 2006.
26 Dewar RE, Wright HT: The culture history of Madagascar. *J World Prehistory* 1993; **7**: 417–466.
27 Burney DA, Burney LP, Godfrey LR *et al*: A chronology for late prehistoric Madagascar. *J Hum Evol* 2004; **47**: 25–63.
28 Adelaar AK: 'Asian roots of the Malagasy: a linguistic perspective'. *Leiden, Bijdragen tot de Taal-, Land- en Volkenkunde* 1995; **151**: 325–356.
29 Deschamps H: *Histoire de Madagascar*. Paris: Berger-Lgevault, 1960.
30 Ottino P: *Madagascar, les Comores et le Sud-ouest de l'Océan Indien*. Publications du Centre d'Anthropologie Culturelle en Sociale, Université de Madagascar: Antananarivo, Madagascar, 1974.
31 Cann RL, Wilson AC: Length mutations in human mitochondrial DNA. *Genetics* 1983; **104**: 699–711.
32 Handt O, Krings M, Ward RH, Paabo S: The retrieval of ancient human DNA sequences. *Am J Hum Genet* 1996; **59**: 368–376.
33 Rieder MJ, Taylor SL, Tobe VO, Nickerson DA: Automating the identification of DNA variations using quality-based fluorescence re-sequencing: analysis of the human mitochondrial genome. *Nucleic Acids Res* 1998; **26**: 967–973.
34 Salas A, Quintans B, Alvarez-Iglesias V: SNaPshot typing of mitochondrial DNA coding region variants. *Methods Mol Biol* 2005; **297**: 197–208.
35 Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N: Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 1999; **23**: 147.
36 Hall TA: BioEdit: a user-friendly biological sequence alignement editor and analysis program for windows 95/98/NT. *Nucleic Acids Symp Ser* 1999; **41**: 95–98.
37 Torroni A, Petrozzi M, D'Urbano L *et al*: Haplotype and phylogenetic analyses suggest that one European-specific mtDNA background plays a role in the expression of Leber hereditary optic neuropathy by increasing the penetrance of the primary mutations 11778 and 14484. *Am J Hum Genet* 1997; **60**: 1107–1121.
38 Kong QP, Bandelt HJ, Sun C *et al*: Updating the East Asian mtDNA phylogeny: a prerequisite for the identification of pathogenic mutations. *Hum Mol Genet* 2006; **15**: 2076–2086.
39 Van Oven M, Kayser M: Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat* 2009; **30**: E386–E394.
40 Kivisild T, Shen P, Wall DP *et al*: The role of selection in the evolution of human mitochondrial genomes. *Genetics* 2006; **172**: 373–387.
41 Behar DM, Villems R, Soodyall H *et al*: The dawn of human matrilineal diversity. *Am J Hum Genet* 2008; **82**: 1130–1140.
42 Soares P, Trejaut JA, Loo JH *et al*: Climate change and postglacial human dispersals in southeast Asia. *Mol Biol Evol* 2008; **25**: 1209–1218.
43 Ingman M, Kaessmann H, Paabo S, Gyllensten U: Mitochondrial genome variation and the origin of modern humans. *Nature* 2000; **408**: 708–713.
44 Ingman M, Gyllensten U: Mitochondrial genome variation and evolutionary history of Australian and New Guinean aborigines. *Genome Res* 2003; **13**: 1600–1606.
45 Macaulay V, Hill C, Achilli A *et al*: Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science* 2005; **308**: 1034–1036.
46 Hartmann A, Thieme M, Nanduri LK *et al*: Validation of microarray-based resequencing of 93 worldwide mitochondrial genomes. *Hum Mutat* 2009; **30**: 115–122.
47 Saillard J, Forster P, Lynnerup N, Bandelt HJ, Norby S: mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. *Am J Hum Genet* 2000; **67**: 718–726.
48 Soares P, Ermini L, Thomson N *et al*: Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am J Hum Genet* 2009; **84**: 740–759.
49 Mishmar D, Ruiz-Pesini E, Golik P *et al*: Natural selection shaped regional mtDNA variation in humans. *Proc Natl Acad Sci USA* 2003; **100**: 171–176.
50 Forster P, Harding R, Torroni A, Bandelt HJ: Origin and evolution of Native American mtDNA variation: a reappraisal. *Am J Hum Genet* 1996; **59**: 935–945.
51 Bandelt HJ, Kong QP, Richards M, Macaulay V: Estimation of mutation rates and coalescence times: Somes caveats; in Bandelt HJ, Macaulay, V, Richards M (eds): *Human Mitochondrial DNA and Evolution of Homo Sapiens*. Berlin: Springer, 2006, pp 47–90.
52 Cox MP: Accuracy of molecular dating with the rho statistic: deviations from coalescent expectations under a range of demographic models. *Hum Biol* 2008; **80**: 335–357.
53 Beaujard MP: Les arrivées austronésiennes à Madagascar: vagues ou continuum?(-Partie 1 et 2). *tudes Océan Indien* 2003; **35–36**: 59–147.
54 Blench MR: New palaeo-zoogeographical evidence for the settlement of Madagascar. *Azania XLII: J Brit Inst Eastern Africa* 2007; **42**: 69–82.
55 Adelaar A: Towards an integrated theory about the Indonesian migrations to Madagascar; in PN. Peregrine IP, Feldman M (eds): *Ancient Human Migrations: An Interdisciplinary Approach*. Salt Lake City: University of Utah Press, 2009.
56 Kayser M, Brauer S, Weiss G *et al*: Melanesian origin of Polynesian Y chromosomes. *Curr Biol* 2000; **10**: 1237–1246.

57 Su B, Jin L, Underhill P *et al*: Polynesian origins: insights from the Y chromosome. *Proc Natl Acad Sci USA* 2000; **97**: 8225–8228.

58 Cox MP, Redd AJ, Karafet TM *et al*: A Polynesian motif on the Y chromosome: population structure in remote Oceania. *Hum Biol* 2007; **79**: 525–535.

59 Karafet TM, Lansing JS, Redd AJ *et al*: Balinese Y-chromosome perspective on the peopling of Indonesia: genetic contributions from pre-neolithic hunter-gatherers, Austronesian farmers, and Indian traders. *Hum Biol* 2005; **77**: 93–114.

60 Underhill PA: A synopsis of extant Y chromosome diversity in East Asia and Oceania; in Sagart L BR, Sanchez-Mazas A (eds): *The Peopling of East Asia: Putting Together Archaeology, Linguistics and Genetics.* London: Routledge Curzon, 2004,, pp 301–319.

61 Scheinfeldt L, Friedlaender F, Friedlaender J *et al*: Unexpected NRY chromosome variation in Northern Island Melanesia. *Mol Biol Evol* 2006; **23**: 1628–1641.

62 Hage P, Marck J: Matrilineality and the melanesian origin of Polynesian Y chromosomes. *Curr Anthropol* 2003; **44**: S121–S127.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (http://www.nature.com/ejhg)