

RESEARCH ARTICLE

Open Access

Mitochondrial DNA and the Y chromosome suggest the settlement of Madagascar by Indonesian sea nomad populations

Pradiptajati Kusuma^{1,4}, Murray P Cox², Denis Pierron¹, Harilanto Razafindrazaka¹, Nicolas Brucato³, Laure Tonasso¹, Helena Loa Suryadi⁴, Thierry Letellier¹, Herawati Sudoyo^{4,5} and François-Xavier Ricaut^{1*}

Abstract

Background: Linguistic, cultural and genetic characteristics of the Malagasy suggest that both Africans and Island Southeast Asians were involved in the colonization of Madagascar. Populations from the Indonesian archipelago played an especially important role because linguistic evidence suggests that the Malagasy language branches from the Southeast Barito language family of southern Borneo, Indonesia, with the closest language spoken today by the Ma'anyan. To test for a genetic link between Malagasy and these linguistically related Indonesian populations, we studied the Ma'anyan and other Indonesian ethnic groups (including the sea nomad Bajo) that, from their historical and linguistic contexts, may be modern descendants of the populations that helped enact the settlement of Madagascar.

Result: A combination of phylogeographic analysis of genetic distances, haplotype comparisons and inference of parental populations by linear optimization, using both maternal and paternal DNA lineages, suggests that Malagasy derive from multiple regional sources in Indonesia, with a focus on eastern Borneo, southern Sulawesi and the Lesser Sunda islands.

Conclusion: Settlement may have been mediated by ancient sea nomad movements because the linguistically closest population, Ma'anyan, has only subtle genetic connections to Malagasy, whereas genetic links with other sea nomads are more strongly supported. Our data hint at a more complex scenario for the Indonesian settlement of Madagascar than has previously been recognized.

Keywords: Ma'anyan, Indonesia, Madagascar, Sea nomad, Mitochondrial DNA, Y chromosome

Background

Prior to the European colonial period, Austronesian-speaking populations were the most widespread of any language family [1,2]. While most groups speaking Austronesian languages moved eastward, settling the Pacific Ocean, others moved westward through the Indian Ocean, reaching eastern Africa and Madagascar. Dispersing halfway around the world within the past two millennia, the Austronesian expansion is often considered the last substantial wave of migration in human prehistory [3-5].

Despite considerable research on the eastward Austronesian expansion, there is little equivalent research on the western edge, leaving major issues unresolved regarding the settlement of Madagascar. Although the exact nature and route of this movement is largely unknown, linguistic and anthropological evidence indicates strong Indonesian influences, as recorded in the vocabulary and socio-cultural life of Malagasy, the modern people of Madagascar [6-9]. Linguistic research suggests that the Malagasy language is derived from Southeast Barito (SEB), a subgroup of Austronesian languages, and is most closely related to the language spoken by the land-locked forest-dweller Ma'anyan in central and southeastern Kalimantan (Borneo) [6,10-13], one indigenous language among 73 others spoken in Borneo [14]. However, there is evidence of

* Correspondence: francois-xavier.ricaut@univ-tlse3.fr

¹Laboratoire d'Anthropologie Moléculaire et Imagerie de Synthèse UMR-5288, Université de Toulouse, Toulouse, France
Full list of author information is available at the end of the article

word borrowings from a small number of Austronesian languages spoken on other Indonesian islands as well [15,16]. This probably reflects multiple Austronesian arrivals to Madagascar from about 700 AD onward (although earlier dates cannot be completely excluded). One hypothesis is that earlier movements were perhaps linked to Southeast Barito speakers, with later arrivals during the 12th-15th centuries connected instead to the Srivijaya and Majapahit kingdoms of Southeast Asia [17,18].

These linguistic findings are broadly supported by genetic studies, which emphasize the shared Indonesian and African genetic heritage of Malagasy. A recent study of genome-wide SNP data suggests that the western and central regions of Indonesia (Java/Borneo/Sulawesi) have the closest genetic connections with Malagasy [19]. This is in agreement with previous studies of uniparental markers (mtDNA and the Y chromosome), which found genetic affinity between Malagasy and western Indonesian populations [20,21]. A key lineage linking Indonesia and Madagascar is the Polynesian motif (a mitochondrial DNA haplogroup, B4a1a1, characterized by the polymorphisms A14022G, T16217C, A16247G and C16261T) [22]. More recently, it has been recognized that Malagasy carry specific point mutation variants (mtDNA nucleotides 1473 and 3423), which together have been termed the Malagasy motif [23]. This Malagasy version of the Polynesian motif is distributed throughout Madagascar with frequencies in specific ethnic groups ranging from 11-50% [21,23-25]. While still debated, this relatively homogenous distribution has been interpreted as supporting the first arrival of the Polynesian motif during an early phase of Madagascar's settlement [26]. To date, the Malagasy motif has not been found in Indonesia [26], or anywhere else outside Madagascar. However, this may simply reflect the paucity of Indonesian populations available for study.

The westward Austronesian expansion was likely associated with trading activities of the Srivijaya empire, as suggested by the many Malay loanwords present in Malagasy [15,27,28], and this trade has been hypothesized to involve some sea nomad groups (i.e., the Orang Laut, Bajo and Bugis) [18,29,30]. This trading network was dominated by men, thus hinting at a potential male bias in the Indonesian contribution to Malagasy, in concordance with the standard matrilineal/matrilocal bias of traditional Austronesian society [31-34]. Contact between the Srivijaya empire and southeast Borneo may have stimulated the dispersal of Southeast Barito speakers to Madagascar – possibly at the same time as the dispersal of Sama-Bajaw speakers (a different subgroup of Barito languages) from the same area [35]. In this context, the Bajo are one sea nomad population of particular interest. Today, the Bajo live in several coastal communities around East Borneo, Sulawesi, the Lesser Sunda islands and the Maluku islands [36,37]. Because the Austronesian migration to Madagascar and the

Sama-Bajaw dispersal may be interrelated, we compare genetic data from recent seafaring populations, such as the Bajo, with the more settled Malagasy.

Similarly, whether Barito populations such as the Ma'anyan, the closest linguistic siblings to modern Malagasy, share close genetic lineages with the Malagasy also remains unanswered. For the first time, we report genetic data for the Ma'anyan and the Lebbo' (a population from Borneo with no presumed role in the settlement of Madagascar) to determine whether the Ma'anyan have an especially close genetic connection with Malagasy. We also include Bajo sea nomads from Sulawesi to determine whether there is a common genetic link based on their shared involvement in long distance maritime trading networks. A large data set of published and unpublished Indonesian populations is included for comparative analysis [38,39]. To investigate sex-specific genetic connections between Indonesia and Madagascar, we analyze both maternal (mtDNA) and paternal (Y chromosome) variation. We propose that the genetic connections of Malagasy to Indonesia are not restricted to Borneo, but instead include maternal and paternal lineages from a wide range of source populations from southern Sulawesi and the Lesser Sunda islands. We therefore propose that the settlement of Madagascar may have been mediated, at least in part, by sea nomad groups.

Results

Y chromosome and mitochondrial DNA classification

Based on analysis of 96 Y chromosome binary markers (Additional file 1: Table S1), the majority of men in the Ma'anyan, Lebbo' and Bajo carry haplogroups previously found in Southeast Asia, particularly C*, K*, and O* (Table 1). Only a few individuals carry Y chromosomes belonging to Western Eurasian haplogroups: R* (M207) [40,41] was found in four Bajo individuals, R1a (M17) [42,43] was found in one Ma'anyan individual, while the western Eurasian haplogroups L1a (M76) and T1a (M70) [44,45] were found in one and two Bajo individuals, respectively. Indian haplogroup R*, which includes R1a, has previously been identified in Bali, Java, Borneo, and Mandar (Additional file 2: Table S2), and thus could conceivably have transited through Indonesia (as opposed to a direct connection), but T1a and L1a have not been identified to date in any Indonesian population.

On the mitochondrial DNA (Table 2), the frequency distributions of haplogroups found in the Ma'anyan, Lebbo' and Bajo are broadly similar to, and consistent with, patterns of maternal lineages in Indonesia. Indeed, four main geographical/historical affiliations can be observed: mainland Asia, the Austronesian expansion, western Eurasia/India, and New Guinea. In brief, mainland Asian mtDNA haplogroups (such as B4c2, M73, M74, M12) are carried by a majority of individuals (64%), followed by haplogroups that have been putatively linked with an Austronesian

Table 1 Y chromosome haplogroup frequencies in the Ma'anyan, Lebbo' and Bajo

Haplogroups	Lebbo'	Ma'anyan	Bajo	Geographic origin	
C*	C-RPS4Y*	0.1333	0.5222	0.0370	SEA
K*	K-M9*	0.0667	–	–	SEA
KxLT	K-M526*	0.1333	0.0222	0.2222	SEA
O1a	O-M119*	–	0.0222	0.0370	SEA
O1a1	O-P203	–	0.1000	–	SEA
O2a1	O-M95*	0.3333	0.0778	–	SEA
O3	O-M122*	–	0.0111	0.0370	SEA
C1c	C-M38*	–	–	0.2222	SEA
O1a2	O-M110	0.2667	0.0222	–	SEA
O2a1a	O-M88	–	0.0111	–	SEA
O3a2	O-P201*	0.0667	0.1778	0.0741	SEA
O3a2b	O-M7	–	–	0.0370	SEA
M1a	M-186	–	–	0.0741	SEA
P*	P-M45*	–	0.0222	–	SEA
R*	M-207*	–	–	0.1481	WE
R1a	R-M17	–	0.0111	–	WE
T1a	T-M70	–	–	0.0741	WE
L1a	L-M76	–	–	0.0370	WE

Note.

SEA: Southeast Asian origin; WE: Western Eurasian origin; and paragroups are indicated using a "*" (star) suffix.

expansion out of Taiwan (such as B4a1a1, M7c1a4, F1*, E*; 32%). The remaining lineages likely derive from India and west Eurasia, and were only observed among the Ma'anyan (M2, M5a4, and M35a, ranging in frequency from 0.6-1.9%). The presence of Indian and other western Eurasian genetic traces has been observed previously in Borneo, as well as Sumatra, Java and Bali [46,47] (Additional file 3: Table S3). Indian haplogroups are restricted to western Indonesia, particularly in regions historically involved in the ancient trading networks of the Hindu kingdoms (such as Srivijaya and Majapahit). Among the Bajo, we also observed the M1a Y haplogroup and Q1 mitochondrial haplogroup, which likely traces its ancestry to New Guinea or eastern Indonesia [39,48-50]. These haplogroups represent a trace of Papuan genetic input. This is perhaps due to the extensive trading network of the Bajo eastward to New Guinea [30] and/or earlier westward expansions of Papuan speakers from New Guinea to eastern Indonesia [51].

Paternal lineage proximity to Malagasy

Shared lineages

Among the haplogroups shared between Malagasy and Indonesians (Additional file 4: Table S4), four originated in Island Southeast Asia (C, O1a, O1a2, O2a1*), while six have western Eurasian origins (J1, J2, J2b, T*, L* and R1a). The Ma'anyan and five other Indonesian groups,

Table 2 Mitochondrial haplogroup frequencies in the Ma'anyan, Lebbo' and Bajo

Haplogroups	Lebbo'	Ma'anyan	Bajo	Geographic origin
B4a	0.2105	0.0943	0.0741	MA
B4c1b	–	0.0252	0.0741	MA
B4c2	–	0.1887	–	MA
B5a	0.1579	0.0377	0.0370	MA
B4a2a	–	0.0440	–	MA
B4a4	–	–	0.0741	MA
B4b1	–	0.0818	–	MA
F3b1a	–	0.0189	–	MA
M12	–	0.0377	–	MA
M20	0.1579	0.0440	–	MA
M71a2	0.1579	–	–	MA
M73	–	0.0566	0.0370	MA
M74b1	–	0.1069	–	MA
N22	–	0.0189	–	MA
N9a6a	–	0.0252	–	MA
R22	–	–	0.0370	MA
R9b1a1a	0.1053	–	–	MA
X	–	–	0.0370	MA
Q1	–	–	0.0741	NG
B4a1a1	–	–	0.0370	Taiw
D4s	–	0.0629	–	Taiw
E1a	0.2105	–	0.0741	Taiw
F1a	–	0.0377	0.0370	Taiw
F1a1a	–	0.0377	–	Taiw
F1a3	–	–	0.0741	Taiw
F1a4	–	0.0252	–	Taiw
M7b1a1i	–	–	0.1481	Taiw
M7b1a2	–	0.0126	0.0741	Taiw
M7c1a4a	–	0.0126	0.1111	Taiw
M2	–	0.0063	–	WE
M35a	–	0.0189	–	WE
M5a4	–	0.0063	–	WE

Note.

MA: Mainland Asian origin; NG: New Guinea origin; Taiw: Out of Taiwan origin; WE: Western Eurasian origin.

all located around the Sulawesi sea (east Kalimantan Dayak, Java, Bali, Mandar and Sumba), share four of these Island Southeast Asian haplogroups. Importantly, Malagasy uniquely share just one subhaplogroup (O2a1a1-M88) with Ma'anyan, and this lineage has not been discovered in other regions of Indonesia. O2a1a1 may therefore be a marker of male genetic contributions from southern Borneo to Madagascar.

Shared Y chromosome haplogroups with a west Eurasian origin (J, T, L and R1a) (Additional file 4: Table S4) are also

present in Indonesian populations, but only in the south and west of the Sulawesi sea. They occur at low frequency (<0.1%) in Java, Bali, Mandar and Bajo, but R1a is the only west Eurasian haplogroup identified in southern Borneo (Ma'anyan and east Kalimantan Dayak). However, west Eurasian lineages in Indonesia and Madagascar may result from independent dispersal events. Indeed, Indian and Arab traders have been active on both side of the Indian Ocean within the last three and one thousand years, respectively [47,52-56]. Therefore, west Eurasian haplogroups shared between Malagasy and Indonesians may have originated from Indonesia, or alternatively, they may have been obtained directly from southwestern Eurasia (the Middle East or India).

Population cross-comparisons

F_{ST} values based on Y chromosome haplogroup frequencies (Additional file 5: Table S5) were visualized on a multidimensional scaling (MDS) plot (Figure 1). Due to their statistically supported genetic homogeneity, Malagasy groups were pooled. The MDS plot (Figure 1) shows that Malagasy Y chromosome lineages are an outlier to Indonesian populations, in a similar way to certain Indonesian population outliers (Mentawai, Nias, Besemah, Semende). Y chromosome F_{ST} values between

Malagasy and Indonesians are relatively high ($F_{ST} > 0.2$; Additional file 5: Table S5), mostly driven by the substantial African component of Malagasy (~65% of the paternal gene pool, but only ~30% of mtDNA [25,24]). No significant differences were observed to suggest specific genetic connections between Malagasy and eastern versus western Indonesians (Mann-Whitney U-test: $P = 0.06$). The Ma'anyan and other populations from Borneo cluster together with western Indonesian groups, including several population outliers (Mentawai, Nias, Besemah, Semende). The Bajo cluster with eastern Indonesian groups, consistent with the well-documented genetic division between western and eastern Indonesia broadly along the Wallace line.

When F_{ST} values are visualized with Surfer (Figure 2), the Indonesian populations with closest affinity to Malagasy (F_{ST} in the lower quartile of the range) are from regions near Wallace's line in the west and south of the Sulawesi sea (southern Sulawesi, eastern Borneo and the Lesser Sunda islands). Populations with highest affinity to Malagasy are Mandar (Sulawesi), Flores (Lesser Sunda), Bajo (Sulawesi), and east Kalimantan Dayak and Lebbo' (Borneo) (Additional file 5: Table S5). These results are supported by a linear optimization method, which aims to find the combination of Indonesian populations that most

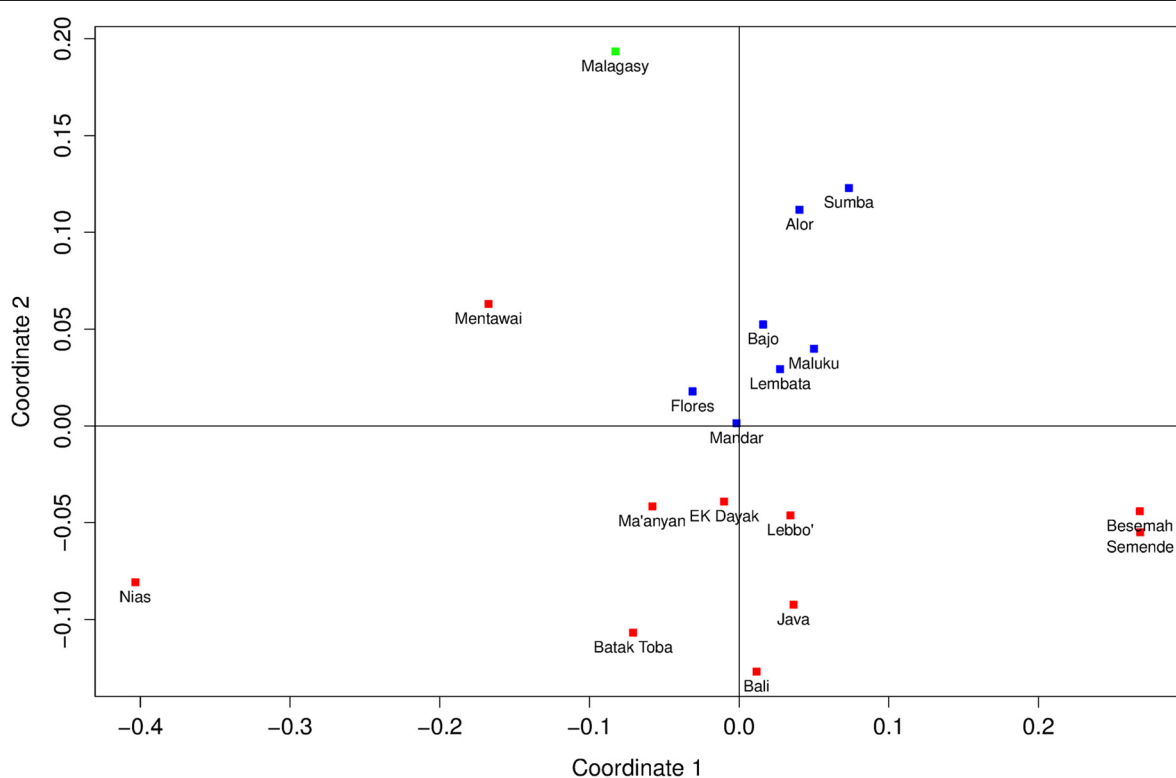
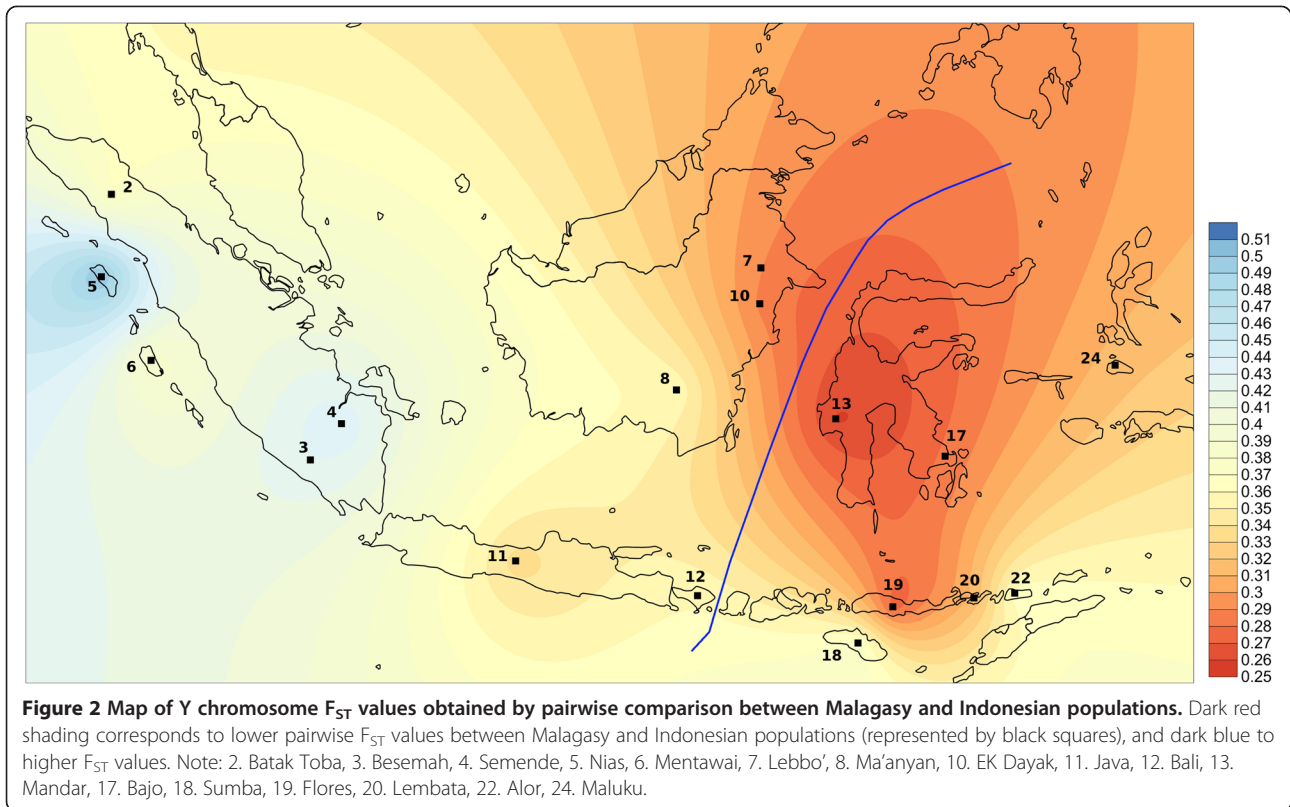


Figure 1 MDS plot showing F_{ST} values between Indonesian and Malagasy populations based on Y chromosome haplogroup frequencies (Kruskal stress: 0.149). Red: western Indonesians; blue: eastern Indonesians.



closely resembles the observed haplogroup diversity in Malagasy. This algorithm highlights two populations from the west and south of the Sulawesi sea, the Mandar (Sulawesi) and Lebbo' (Borneo), as populations that produce a Y chromosome genetic profile most closely resembling the observed pattern, while still accounting for the predominantly African genetic background found in Malagasy (Additional file 6: Figure S1).

These geographical regions comprised part of the trading sphere of the Srivijaya empire, including several Javanese kingdoms that played a crucial role in the region: Heluodan (5th century), Tarumanagara (5th century), Walaing (Chinese Heling, 7th-8th centuries), Kahuripan/Kediri (11th century), Singasari (13th century) and Majapahit (13th -15th centuries) [57]. This region also hosted several houseboat nomad groups (such as the Bajo), which had ample opportunities to incorporate men from a wider regional watershed. The Ma'anyan from southern Borneo do not show any privileged link with Malagasy (indeed, they have a relatively high F_{ST} value showing genetic differentiation), despite being the only Indonesian population that shares Y haplogroup O2a1a with Malagasy. This may indicate that the genetic contribution of Ma'anyan was limited, either due to the recent arrival of this lineage in Ma'anyan, or perhaps O2a1a has since been lost or is still undetected in other Indonesian populations.

Maternal lineage proximity to Malagasy

Shared lineages

Malagasy and Indonesians share mitochondrial haplogroups B4a1a, B4a1a1 (Polynesian motif), E1a1a, F3b, M7c1a4a, M32c and Q1 (Additional file 7: Table S6). Of these, B4a1a1, E1a1a and Q1 are found exclusively in eastern Indonesia. Conversely, F3b, B4a1a and M7c1a4a occur ubiquitously across both eastern and western Indonesia, and M32c has been observed in only one Javanese individual.

The Polynesian motif (B4a1a1) is considered strong evidence of Indonesian gene flow into Madagascar, where a variant is found at moderate frequency (11-50%). With the exception of Bali (Additional file 3: Table S3), B4a1a1 only occurs in eastern Indonesia. For the three new populations studied here, only the Bajo carry this Polynesian motif (just one of 27 individuals), and importantly, it was not found in any of our populations from Borneo. Furthermore, the specific Malagasy motif has not been found in Indonesia at all, including the new populations screened here. Considering the restricted geographic distribution of the Polynesian motif, it is most likely that this lineage from Madagascar traces back to eastern rather than western Indonesia.

Malagasy and Indonesians share ten haplotypes in seven haplogroups (Additional file 8: Table S7): two haplotypes each in B4a1a and B4a1a1; three haplotypes in M7c1a4a; and one haplotype in each of the other shared

haplogroups. As shown in Figure 3, eastern Indonesian populations tend to share more haplotypes with Malagasy than western Indonesian groups. Populations from Sumba share the highest number of haplotypes (n = 6), followed by North Maluku and Sulawesi Bugis (n = 5). In a recurring theme, Ma'anyan exhibit limited sharing with only two haplotypes in common.

Population cross-comparisons

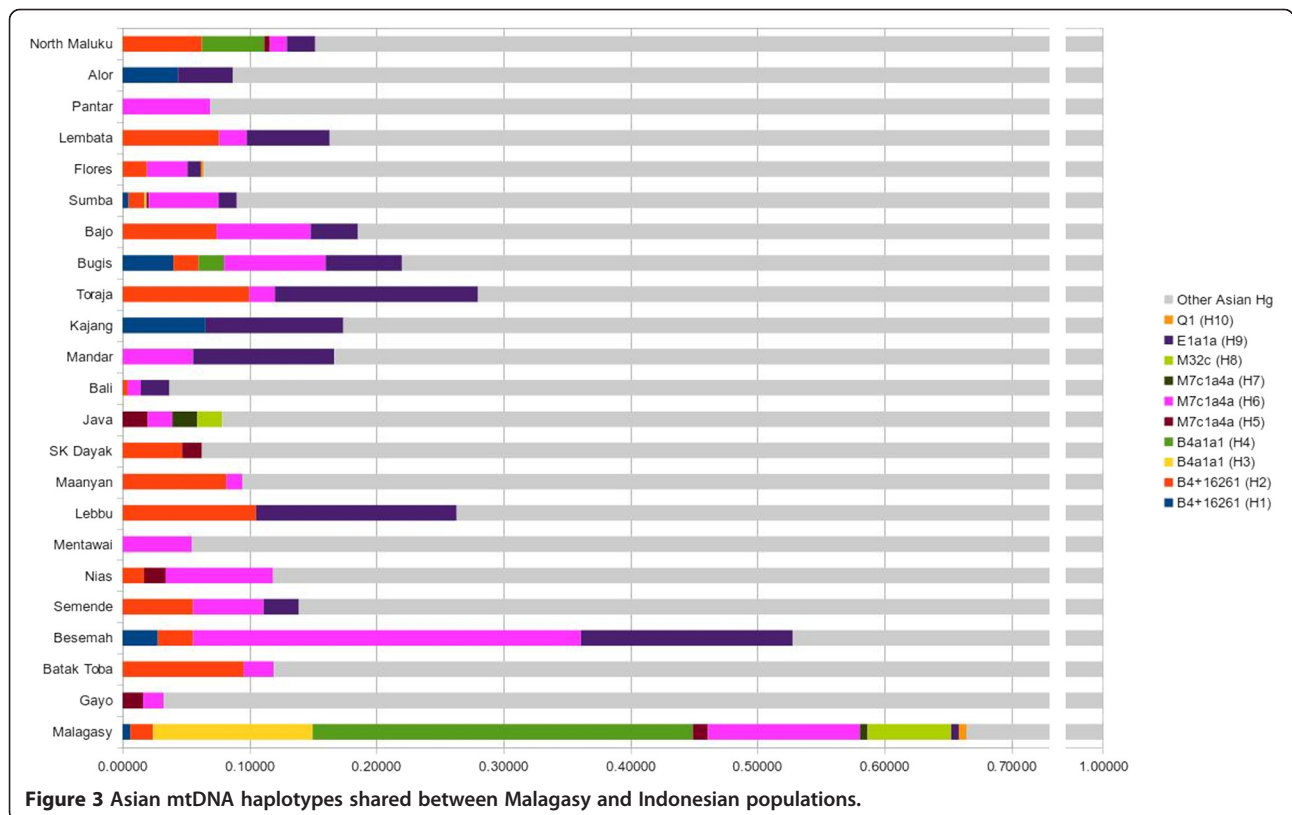
The MDS plot (Figure 4, Malagasy groups again pooled), which is based on F_{ST} values from mtDNA haplogroup frequencies (Additional file 3: Table S3), shows that Malagasy maternal lineages differ markedly from those of Indonesians, while paternal lineages appear relatively closer (Figure 1). Unlike the Y chromosome data, which favors both eastern and western Indonesian sources, Malagasy are closer to the mtDNA diversity of eastern rather than western Indonesians (Mann–Whitney U test: $P < 0.01$). F_{ST} values (Additional file 9: Table S8) visualized in Surfer (Figure 5) show that the regions with closest affinity occur to the south and east of Sulawesi, and support an eastern Indonesian connection. Populations with higher affinity to Malagasy (F_{ST} in the lower quartile of the range) include Sumba and Flores, the Maluku islands and the Bugis of south Sulawesi. As seen on the Y chromosome, the Bajo cluster with eastern Indonesian groups, while the Ma'anyan and other populations from Borneo cluster with western

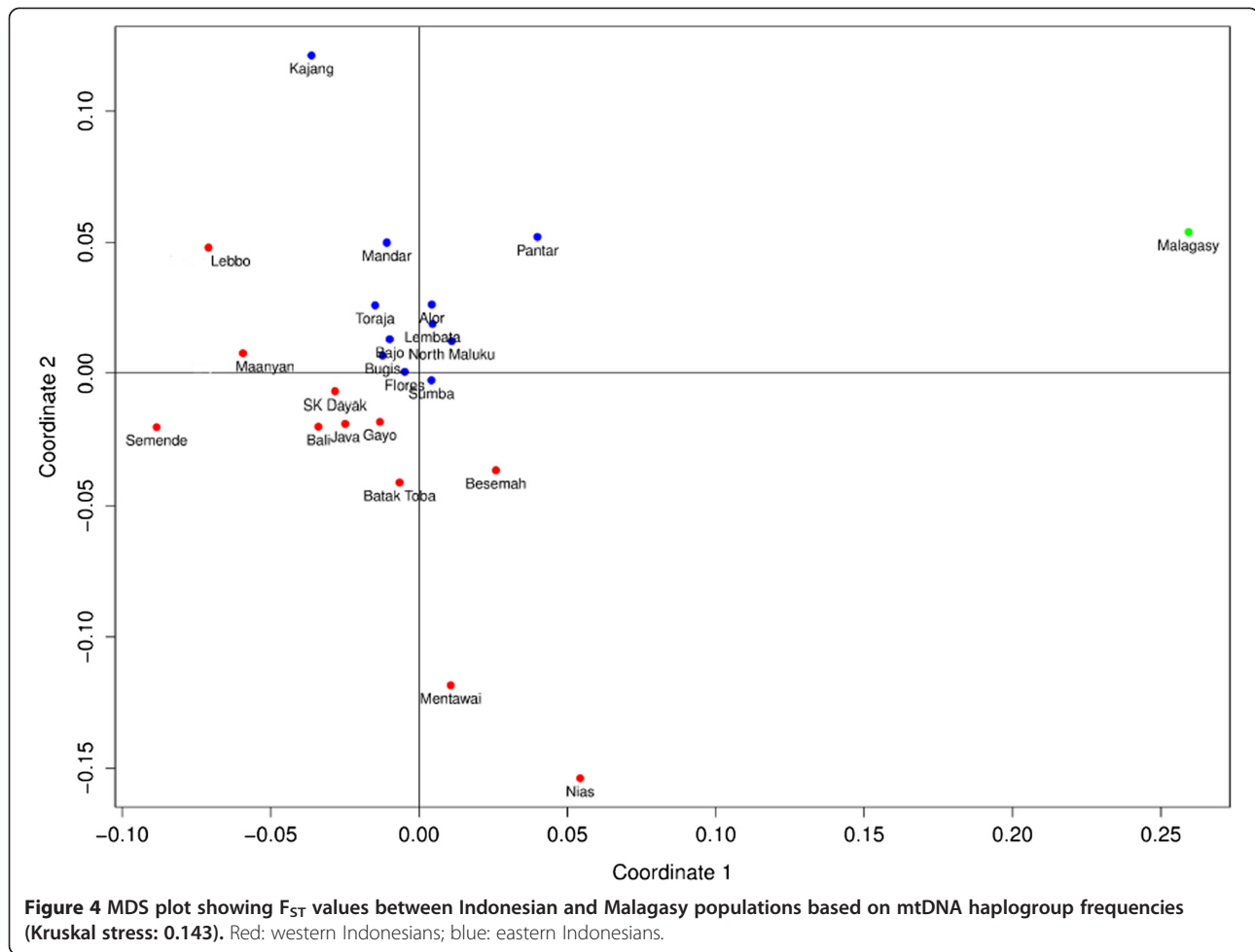
Indonesian groups. Linear optimization results broadly agree with the F_{ST} results: Malagasy are most likely derived from a combination of eastern Indonesian groups, such as North Maluku, Bugis and Bajo (Additional file 10: Figure S2).

These multiple lines of genetic evidence suggest that Malagasy women may have originated predominantly from eastern Indonesia. Key pieces of evidence include the restricted distribution of the Polynesian motif in eastern Indonesia and patterns of shared maternal lineages. This is consistent with the hypothesis that Austronesians borrowed the Polynesian motif – which perhaps arose in the Bismarck archipelago – from indigenous sources in eastern Indonesia [58,59]. It is worth noting that eastern Indonesian influences occur in Madagascar: for cultivated plants (i.e., myths of origins, rituals for yams, the ancient importance of taro) [18,30], and through the influence of the Orang Laut language on some Malagasy dialects (particularly the Vezo) [11]. Eastern Indonesia had no recorded sea-faring cultures involved in long-distance trading, except for the Bugis and Bajo. However, maritime foraging and trade were likely more common in earlier millennia as networks of long-distance sea-based interactions have been in place since at least the early Holocene [60].

Discussion

The Southeast Barito language subgroup includes two languages spoken by populations separated by the 7,500 km



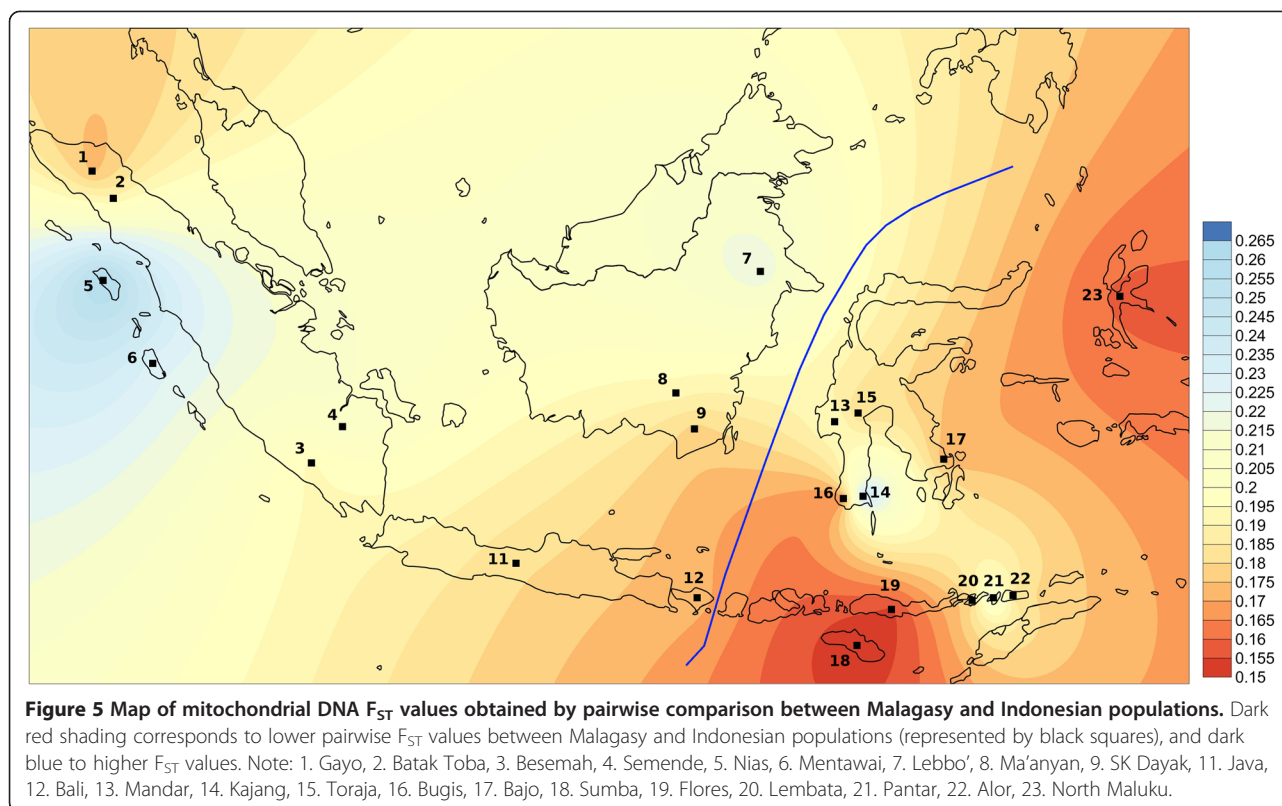


expanse of the Indian Ocean: the Malagasy of Madagascar, off the east coast of Africa, and the Ma'anyan of Borneo, an island in western Indonesia. Knowing this linguistic connection, we investigated genetic linkages between these two populations. Our results suggest that few genetic connections exist, either on the paternal Y chromosome or the maternal mtDNA. These results suggest that 1) the Ma'anyan groups sampled here are not directly related to the individuals who settled Madagascar, 2) subsequent demographic events have erased any genetic affinity between them, or 3) the Ma'anyan were just one population of many that contributed to the settlement of Madagascar (a possibility suggested by the exclusive sharing of Y chromosome haplogroup O2a1a). The first hypothesis might suggest that other Southeast Barito groups from southern Borneo (such as the Samihim or the Dusun Witu [61]) were involved instead.

These conclusions are drawn from a suite of complementary analyses, including phylogeography, haplotype sharing and linear optimization approaches. In combination, they paint a picture of the genetic dynamics between Indonesia and Madagascar. Although the geographic

distribution of Indonesian populations that most closely reflect Malagasy genetic diversity are remarkably convergent for both male and female lineages, it is noticeable that the regions suggested by these two systems do not overlap perfectly. Our analyses suggest that populations from the south and west of the Sulawesi Sea (east Borneo, south Sulawesi and the Lesser Sunda islands) best explain Y chromosome diversity, while populations from the south and east of Sulawesi (south Sulawesi, the Lesser Sunda islands and the Maluku islands), all in eastern Indonesia, best explain mtDNA diversity. We emphasize that parts of these two regions overlap, thus potentially explaining both paternal and maternal affinity.

For instance, the distribution of shared lineages favors different source populations for maternal and paternal loci. Y chromosome haplogroup O2a1a is found only in Ma'anyan, while mtDNA haplogroups B4a1a1 and Q1 are found exclusively in eastern Indonesia. (Suggesting yet more connections, M32c has only been found in Java). F_{ST} and linear optimization results also highlight different source regions for the Y chromosome and mtDNA. Together, these patterns suggest that multiple regions may



have contributed to the settlement of Madagascar, perhaps via one or a few admixed groups.

Sea nomads have been active traders along the eastern coast of Borneo, southern Sulawesi, the Lesser Sunda islands and the Maluku islands for at least the last few hundred years [62,63]. These mobile populations linked western and eastern Indonesia, and absorbed individuals from different regions. Sea nomads traveled with their families, even on long distance journeys [36,37]. Moreover, languages of the Sama-Bajaw group, as spoken by the sea nomad Bajo, form a subgroup of the Barito languages of southeast Kalimantan [35], although not the closest language subgroup to Malagasy. Their patterns of genetic diversity and lifestyle make them possible contenders for the Indonesian populations who helped enact the settlement of Madagascar, although a definite assignment remains elusive.

Conclusion

We propose that the settlement of Madagascar had an Indonesian source location around southern Sulawesi, the Lesser Sunda islands and eastern Borneo. The populations involved may be related to modern sea nomad groups and the ancient Malay Srivijaya trading network. The Indonesian ancestors of Malagasy certainly carried maternal lineages with greater putative contributions from eastern Indonesia, and paternal lineages from both eastern

and western Indonesia. The absence of any clear genetic connection between Malagasy and at least some populations speaking their most closely related language, Ma'anyan, raises important questions about the link between genes and language in the Indonesian dispersal across the Indian Ocean. Studying other Southeast Barito and sea nomad groups (such as the Orang Laut, who played a crucial role in the Srivijaya expansion) and the application of genome-wide genotyping technologies to provide additional genetic resolution promises to bring new insight to bear on these questions.

Methods

Population samples

All samples analyzed in this study were collected with informed consent from unrelated individuals. Subjects were surveyed for language affiliation, current residence, familial birthplaces, and a short genealogy of four generations to establish regional ancestry. A total of 205 DNA samples were analyzed from three ethnic groups: 159 Ma'anyan individuals were collected in Tamiang Layang (East Barito), Central Kalimantan, and Banjarmasin (South Kalimantan), representing the largest ethnically-defined population sample from Borneo to date; 19 Lebbo' in East Kalimantan; and 27 sea nomad Bajo in Kendari (Sulawesi). Collection and use of these samples was approved by the Research Ethics Commissions at both the

Eijkman Institute for Molecular Biology, Indonesia, and the University of Toulouse, France. We also included data for additional Indonesian populations from published [38,39,46] and unpublished sources (Gayo, North Maluku, and a mixed assemblage of other Dayak ethnic groups from the southern part of South Kalimantan province (“SK Dayak”) from the Eijkman Institute’s archived samples) (Additional file 11: Table S9). We also included published Malagasy data: seven Malagasy populations located in the southwest, southeast, and central highlands of Madagascar. The Malagasy were pooled for most analyses as these groups are genetically highly similar (between group $F_{ST} < 0.05$ and 95% of F_{ST} values non-significant ($P > 0.05$) for both mtDNA and the Y chromosome) [21,25,24]. In total, the mtDNA dataset comprises 529 Malagasy and 2,841 Indonesians, and the Y chromosome dataset comprises 371 Malagasy and 2,095 Indonesians.

DNA extraction and genotyping

We collected blood samples for the Ma’anyan, except the Lebbo’ and Bajo, for which saliva samples were collected using the Oragene DNA Collection kit (<http://dnagenotek.com>). DNA was extracted from blood using a standard salting-out procedure, and from saliva using the manufacturer’s standard protocol. For paternal lineage analysis, 96 binary markers on the non-recombining region of the Y chromosome were analyzed. We used a nanofluidic dynamic array (Fluidigm, USA) high-throughput genotyping system. This system is developed for SNP genotyping assays and able to perform 9,216 real-time polymerase chain reactions (PCRs) (96 primers \times 96 samples) on a single chip. The results were analyzed using the BioMark™ HD system (Fluidigm, USA) which integrated the Real-Time PCR Analysis software. Each haplogroup was assigned based on the updated ISOGG’s Y-DNA haplogroup tree [64] and the Y-Phylo tree [65]. The full list of markers is shown in Additional file 1: Table S1. The mtDNA hyper-variable region I was sequenced using primers F15989 (5’-CCCAAAGCTAAGATTCTAAT-3’) and R389 (5’-CTGGTTAGGCTGGTGTTAGG-3’). Sequences (GenBank accession numbers: KM590988-KM591192) were then analyzed and aligned against the revised Cambridge Reference Sequence (rCRS) [66] using the MAFFT aligner v.7 [67]. Mitochondrial haplogroups were determined with the Haplogrep program (<http://haplogrep.uibk.ac.at>) based on Phylo tree v.16 [68]. The Malagasy motif, defined by mitochondrial coding region polymorphisms at nucleotides 1,473 and 3,423, were typed on all individuals carrying haplogroup B4a* using the method previously described [23].

Statistical analysis

Pairwise F_{ST} distances between Indonesian and Malagasy populations were computed from haplogroup frequency

data using Arlequin v.3.5 [69] with 5,040 permutations. Multidimensional scaling (MDS) from F_{ST} values based on Y chromosome and mitochondrial DNA haplogroup frequencies (Additional files 2 and 3: Table S2 and Table S3) was performed to visualize inter-population relations. The nonparametric Mann–Whitney U-test was applied to analyze the statistical significance of genetic affinity between Malagasy and Western/Eastern Indonesian groups. This phylogeographical division was defined by Wallace’s line, in agreement with previous human genetic population studies [39,70,71]. F_{ST} values obtained for the pairwise comparison of maternal and paternal lineages between Malagasy and Indonesian populations were plotted geographically with Surfer v.12.0 using the Kriging method. To determine which linear combination of Indonesian populations produces the closest genetic profile to that observed in Malagasy, we employed a statistical analysis of least squares with equalities and inequalities (lsei) algorithm in the R package, limSolve [72]. To capture sampling variance and drift dynamics, the genetic data were resampled 5,000 times and the linear optimization results visualized with box plots using the R package, ggplot. This analysis used mitochondrial DNA and Y chromosome haplogroup frequency distributions for both Malagasy and Indonesian populations (Additional files 2 and 3: Table S2 and Table S3). An African reference group was used to represent the non-Asian contribution to Malagasy. This reference comprised African samples from populations in North-, East-, Central- and South Africa [73–79]. Sharing of mitochondrial haplotypes was ascertained using Arlequin v.3.5 [69]. For this analysis, sequences from Tofanelli [21] were excluded due to their short length (360 bp compared to 520 bp for the present study).

Additional files

Additional file 1: Table S1. List of Y chromosome SNP markers screened.

Additional file 2: Table S2. Y chromosome haplogroup frequencies in Indonesian and Malagasy populations.

Additional file 3: Table S3. Mitochondrial haplogroup frequencies in Indonesian and Malagasy populations.

Additional file 4: Table S4. Y chromosome haplogroup sharing frequencies between Malagasy and Indonesian populations.

Additional file 5: Table S5. Y chromosome F_{ST} distances between Malagasy and Indonesian populations.

Additional file 6: Figure S1. Population proportion based on Y chromosome haplogroups estimated with the lsei algorithm. Note: black dots represent mean values.

Additional file 7: Table S6. Mitochondrial haplogroup sharing frequencies between Malagasy and Indonesian populations.

Additional file 8: Table S7. Mitochondrial haplotype sharing between Malagasy and Indonesian populations.

Additional file 9: Table S8. Mitochondrial DNA F_{ST} distances between Malagasy and Indonesian populations.

Additional file 10: Figure S2. Population proportion based on mitochondrial DNA haplogroups estimated with the Isei algorithm. Note: black dots represent mean values.

Additional file 11: Table S9. List of populations used in this study.

Abbreviations

SEB: Southeast Barito; SNP: Single nucleotide polymorphism; mtDNA: mitochondrial DNA; MDS: Multidimensional scaling; SK Dayak: South Kalimantan Dayak; EK Dayak: East Kalimantan Dayak; Isei: Least squares with equalities and inequalities.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

PK, MPC, and F-XR designed the research, analyzed the data, and wrote the manuscript; PK, HLS, HS, and F-XR collected the samples, PK and LT carried out the molecular labwork, DP, HR, NB, and TL helped draft the manuscript. All authors gave final approval for publication.

Acknowledgments

We thank Stephanie Schiavinato and Julie Bosc for technical assistance, and to Alexander Adelaar for helpful discussion. We also thank the Eijkman Institute for Molecular Biology (Jakarta, Indonesia) for accommodating the sampling program; Phillipe Grangé, Charles Illouz, and Chandra Nuraini (Université de La Rochelle) for facilitating access to the Bajo population; Budi Amuranto (Dinas Kebudayaan dan Pariwisata, East Kutai Sangata, Indonesia) for facilitating access to the Lebbo' population; Bambang Sugiyanto (Balai Arkeologi Banjarmasin, Indonesia); Adhi Agus Oktaviana (National Research Center for Archaeology, Indonesia); Antonio Guerreiro (CNRS UMR-6571, Université Aix-Marseille); and especially, all our Indonesian participants. This research was supported by the French ANR grant number ANR-14-CE31-0013-01 (grant OcooAdapto to F-X.R), the French ANR-12-PDOC-0037-01 (grant GENOMIX to D.P), the Region Aquitaine of France (grant MAGE to D.P), the French Ministry of Foreign and European Affairs (French Archaeological Mission in Borneo (MAFBO) to F-X.R), the European Research Council (ERC 295918 MesAndLin (g) k to N.B.), and the French Embassy in Indonesia through its Cultural and Cooperation Services (Institut Français en Indonésie).

Author details

¹Laboratoire d'Anthropologie Moléculaire et Imagerie de Synthèse UMR-5288, Université de Toulouse, Toulouse, France. ²Statistics and Bioinformatics Group, Institute of Fundamental Sciences, Massey University, Palmerston North, New Zealand. ³Center for Linguistics, University of Leiden, Leiden, Netherlands. ⁴Genome Diversity and Diseases Laboratory, Eijkman Institute for Molecular Biology, Jakarta, Indonesia. ⁵Department of Medical Biology, Faculty of Medicine, University of Indonesia, Jakarta, Indonesia.

Received: 29 October 2014 Accepted: 24 February 2015

Published online: 17 March 2015

References

- Bellwood P, Fox JJ, Tryon D. The Austronesians in history: Common origins and diverse transformations. In: *The Austronesians: historical and comparative perspectives*. Canberra, Australia: ANU E Press; 1995. p. 1–16.
- Blust R. Austronesian: A sleeping giant? *Lang Linguist Compass*. 2011;5:538–50.
- Blench RM. The Pleistocene settlement of the rim of the Indian Ocean. Manila, Philippines: Paper presented on the 18th Congress of the Indo-Pacific Prehistory Association; 2006.
- Burney DA, Burney LP, Godfrey LR, Jungers WL, Goodman SM, Wright HT, et al. A chronology for late prehistoric Madagascar. *J Hum Evol*. 2004;47:25–63.
- Kumar A. "The single most astonishing fact of human geography": Indonesia's far west colony. *Indonesia*. 2011;92:59–96.
- Adelaar KA. The Austronesian languages of Asia and Madagascar: a historical perspective. In: *The Austronesian languages of Asia and Madagascar*, vol. 1. New York, USA: Routledge; 2005. p. 1–42.
- Blench R. Evidence for the Austronesian voyages in the Indian Ocean. In: Anderson A, Barrett JH, Boyle KV, editors. *The global origins and development of seafaring*. Cambridge, UK: McDonald Institute for Archeological Research; 2010. p. 239–48.
- Dewar RE, Wright HT. The culture history of Madagascar. *J World Prehistory*. 1993;7:417–66.
- Serva M, Petroni F, Volchenkov D, Wichmann S. Malagasy dialects and the peopling of Madagascar. *J R Soc Interface*. 2012;9:54–67.
- Adelaar KA. Borneo as a cross-roads for comparative Austronesian linguistics. In: *The Austronesians: historical and comparative perspectives*. Canberra, Australia: ANU E Press; 1995. p. 81–102.
- Dahl OC. Migration from Kalimantan to Madagascar. Oslo, Norway: Norwegian University [sic] Press: Institute for Comparative Research in Human Culture; 1991.
- Dahl OC. La subdivision de la famille Barito et la place du malgache. *Acta Orient*. 1977;38:77–134.
- Dahl OC. Malgache et maanja: une comparaison linguistique. Oslo, Norway: Edege-Intituttet; 1951.
- Lewis MP, Simons GF, Fennig CD (Eds): *Ethnologue: Languages of the World*. Seventeenth. Texas: SIL International. Online version: <http://www.ethnologue.com>; 2014.
- Adelaar KA. Loanwords in Malagasy. In: Haspelmath M, Tadmor U, editors. *Loanwords in the world's languages: a comparative handbook*. Berlin, Germany: De Gruyter Mouton; 2009. p. 717–46.
- Adelaar KA. The Indonesian migrations to Madagascar: making sense of the multidisciplinary evidence. In: *Austronesian diaspora and the ethnogenesis of people in Indonesian archipelago: Proceedings of the International Symposium*. Jakarta, Indonesia: LIPI Press; 2006. p. 205–32.
- Beaujard P. Les arrivées austronésiennes à Madagascar: vagues ou continuum?. *Études Océan Indien*. 2003;35-36:59–147.
- Beaujard P. *Les Mondes de L'océan Indien*. Vol. 1 : De La Formation de l'État Au Premier Système-Monde Afro-Eurasien (4e Millénaire Av. J.-C.-6e Siècle Apr. J.-C.). Paris, France: Armand Collin; 2012.
- Pierron D, Razafindrazaka H, Pagani L, Ricaut F-X, Antao T, Capredon M, et al. Genome-wide evidence of Austronesian-Bantu admixture and cultural reversion in a hunter-gatherer group of Madagascar. *Proc Natl Acad Sci*. 2014;111:936–41.
- 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.
- Tofanelli S, Bertoncini S, Castri L, Luiselli D, Calafell F, Donati G, et al. On the origins and admixture of Malagasy: New evidence from high-resolution analyses of paternal and maternal lineages. *Mol Biol Evol*. 2009;26:2109–24.
- Soodyall H, Jenkins T, Stoneking M. "Polynesian" mtDNA in the Malagasy. *Nat Genet*. 1995;10:377–8.
- Razafindrazaka H, Ricaut F-X, Cox MP, Mormina M, Dugoujon J-M, Randriamarolaza LP, et al. Complete mitochondrial DNA sequences provide new insights into the Polynesian motif and the peopling of Madagascar. *Eur J Hum Genet*. 2010;18:575–81.
- Razafindrazaka H. *Le peuplement humain de Madagascar: Anthropologie génétique de trois groupes traditionnels*. PhD Thesis. Université Paul Sabatier - Toulouse III, Faculté de Médecine; 2010.
- Capredon M, Brucato N, Tonasso L, Choismel-Cadamuro V, Ricaut F-X, Razafindrazaka H, et al. Tracing arab-islamic inheritance in Madagascar: Study of the Y-chromosome and mitochondrial DNA in the Antemoro. *PLoS One*. 2013;8:e80932.
- Cox MP, Nelson MG, Tumonggor MK, Ricaut F-X, Sudoyo H. A small cohort of Island Southeast Asian women founded Madagascar. *Proc R Soc B Biol Sci*. 2012;279:2761–8.
- Adelaar KA. Malay influence on Malagasy: linguistic and culture-historical implications. *Ocean Linguist*. 1989;28:1–46.
- Adelaar KA. Structural diversity in the Malayan subgroup. In: *The Austronesian Languages of Asia and Madagascar*. New York, USA: Routledge; 2005. p. 202–26.
- Pelras C. *The Bugis*. 1st ed. Oxford, UK: Wiley-Blackwell; 1997.
- Beaujard P. The first migrants to Madagascar and their introduction of plants: linguistic and ethnological evidence. *Azania*. 2011;46:169–89.
- Jordan FM, Gray RD, Greenhill SJ, Mace R. Matrilocal residence is ancestral in Austronesian societies. *Proc R Soc B Biol Sci*. 2009;276:1957–64.
- Lansing JS, Cox MP, de Vet TA, Downey SS, Hallmark B, Sudoyo H. An ongoing Austronesian expansion in Island Southeast Asia. *J Anthropol Archaeol*. 2011;30:262–72.
- Marck J. Proto Oceanic society was matrilineal. *J Polyn Soc*. 2008;117:345–82.

34. Peng M-S, Quang HH, Dang KP, Trieu AV, Wang H-W, Yao Y-G, et al. Tracing the Austronesian footprint in mainland Southeast Asia: a perspective from mitochondrial DNA. *Mol Biol Evol.* 2010;27:2417–30.
35. Blust R. The linguistic position of Sama-Bajaw. *Stud Philipp Lang Cult.* 2007;15:73–114.
36. Nuraini-Grangé C. Langue et production de récits d'une communauté Bajo des îles Kangean (Indonésie). PhD Thesis. Université de la Rochelle, Département de Langue et Civilisation Indonésienne; 2008.
37. Stacey N. Boats to Burn: Bajo fishing activity in the Australian fishing zone. Canberra, Australia: ANU E Press; 2007.
38. Karafet TM, Hallmark B, Cox MP, Sudoyo H, Downey S, Lansing JS, et al. Major east–west division underlies Y chromosome stratification across Indonesia. *Mol Biol Evol.* 2010;27:1833–44.
39. Tumonggor MK, Karafet TM, Hallmark B, Lansing JS, Sudoyo H, Hammer MF, et al. The Indonesian archipelago: an ancient genetic highway linking Asia and the Pacific. *J Hum Genet.* 2013;58:165–73.
40. Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, Hammer MF. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res.* 2008;18:830–8.
41. Hammer MF, Zegura SL. The human Y chromosome haplogroup tree: Nomenclature and phylogeography of its major divisions. *Annu Rev Anthropol.* 2002;31:303–21.
42. Pamjav H, Fehér T, Németh E, Pádár Z. Brief communication: New Y-chromosome binary markers improve phylogenetic resolution within haplogroup R1a1. *Am J Phys Anthropol.* 2012;149:611–5.
43. Underhill PA, Myres NM, Rootsi S, Metspalu M, Zhivotovskiy LA, King RJ, et al. Separating the post-Glacial coancestry of European and Asian Y chromosomes within haplogroup R1a. *Eur J Hum Genet.* 2010;18:479–84.
44. Mendez FL, Karafet TM, Krahn T, Ostrer H, Soodyall H, Hammer MF. Increased resolution of Y chromosome haplogroup T defines relationships among populations of the Near East, Europe, and Africa. *Hum Biol.* 2011;83:39–53.
45. Wang CC, Wang LX, Zhang M, Yao D, Jin L, Li H. Present Y chromosomes support the Persian ancestry of Sayyid Ajjal Shams al-Din Omar and Eminent Navigator Zheng He. *Con on C A.* 2014;3:8–10.
46. Gunnarsdóttir ED, Nandineni MR, Li M, Myles S, Gil D, Pakendorf B, et al. Larger mitochondrial DNA than Y-chromosome differences between matrilineal and patrilineal groups from Sumatra. *Nat Commun.* 2011;2:228.
47. Karafet T, Lansing J, Redd A, Reznikova S. 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.
48. Friedlaender JS, Friedlaender FR, Hodgson JA, Stoltz M, Koki G, Horvat G, et al. Melanesian mtDNA Complexity. *PLoS One.* 2007;2:e248.
49. Mona S, Grunz KE, Brauer S, Pakendorf B, Castri L, Sudoyo H, et al. Genetic admixture history of eastern Indonesia as revealed by Y-chromosome and mitochondrial DNA analysis. *Mol Biol Evol.* 2009;26:1865–77.
50. Kayser M, Brauer S, Weiss G, Schiefelhövel W, Underhill P, Shen P, 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.
51. Donohue M, Denham T. Farming and Language in Island Southeast Asia: Reframing Austronesian History. *Curr Anthropol.* 2010;51:223–56.
52. Donkin RA. Between east and west: The Moluccas and the traffic in spices up to the arrival of Europeans. Philadelphia, USA: Amer Philosophical Society; 2004.
53. Ferrand G. Les musulmans à Madagascar et aux îles Comores. Paris, France: E. Leroux; 1891.
54. Ardika I, Bellwood P. Sembiran: the beginnings of Indian contact with Bali. *Antiquity.* 1991;65:221–32.
55. Ardika IW, Bellwood P, Sutaba IM, Yulianti KC. Sembiran and the first Indian contacts with Bali: an update. *Antiquity.* 1997;71:193–4.
56. Calo A. Ancient trade between India and Indonesia. *Science.* 2014;345:1255.
57. Beaujard P. Les Mondes de L'océan Indien. Vol. 2: L'océan Indien, Au Cœur Des Globalisations de l'Ancien Monde (7e-15e Siècles). Paris, France: Armand Collin; 2014.
58. Oppenheimer S, Richards M. Fast trains, slow boats, and the ancestry of the Polynesian islanders. *Sci Prog.* 2001;84(Pt 3):157–81.
59. Soares P, Rito T, Trejaut J, Mormina M, Hill C, Tinkler-Hundal E, et al. Ancient voyaging and Polynesian origins. *Am J Hum Genet.* 2011;88:239–47.
60. Bulbeck D. An integrated perspective on the Austronesian diaspora: The switch from cereal agriculture to maritime foraging in the colonisation of Island Southeast Asia. *Aust Archaeol.* 2008;67:31–52.
61. Adelaar KA. L'importance du samihim (Bornéo du Sud) pour l'étymologie malgache. In: Champion B, editor. L'étrange intime. Mélanges offerts à Paul Ottino. Océan. France: Université de la Réunion; 1995.
62. Sather C. The Bajau Laut: adaptation, history, and fate in a maritime fishing society of South-Eastern Sabah. Oxford, UK: Oxford University Press; 1997.
63. Verheijen JAJ. The Sama/Bajau Language in the Lesser Sunda Islands. Canberra, Australia: Department of Linguistics, Research School of Pacific Studies, Australian National University; 1986.
64. ISOGG 2014 Y-DNA Haplogroup Tree, Version: 9.70. Date: 2 June 2014. [<http://www.isogg.org/tree/>]
65. Van Oven M, Van Geystelen A, Kayser M, Decorte R, Larmuseau MH. Seeing the wood for the trees: a minimal reference phylogeny for the human Y chromosome. *Hum Mutat.* 2014;35:187–91.
66. 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.
67. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol.* 2013;30:772–80.
68. Van Oven M, Kayser M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat.* 2009;30:E386–94.
69. Excoffier L, Lischer HEL. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour.* 2010;10:564–7.
70. Cox MP, Karafet TM, Lansing JS, Sudoyo H, Hammer MF. Autosomal and X-linked single nucleotide polymorphisms reveal a steep Asian–Melanesian ancestry cline in eastern Indonesia and a sex bias in admixture rates. *Proc R Soc B Biol Sci.* 2010;277:1589–96.
71. Xu S, Pugach I, Stoneking M, Kayser M, Jin L. Genetic dating indicates that the Asian–Papuan admixture through Eastern Indonesia corresponds to the Austronesian expansion. *Proc Natl Acad Sci.* 2012;109:4574–9.
72. Van den Meersche K, Soetaert K, Middelburg JJ. A Bayesian compositional estimator for microbial taxonomy based on biomarkers. *Limnol Oceanogr Methods.* 2008;6:190–9.
73. Salas A, Richards M, De la Fe T, Lareu MV, Sobrino B, Sánchez-Diz P, et al. The making of the African mtDNA landscape. *Am J Hum Genet.* 2002;71:1082–111.
74. Salas A, Richards M, Lareu MV, Scozzari R, Coppa A, Torroni A, et al. The African diaspora: mitochondrial DNA and the Atlantic slave trade. *Am J Hum Genet.* 2004;74:454–65.
75. Harich N, Costa MD, Fernandes V, Kandil M, Pereira JB, Silva NM, et al. The trans-Saharan slave trade - clues from interpolation analyses and high-resolution characterization of mitochondrial DNA lineages. *BMC Evol Biol.* 2010;10:138.
76. Tishkoff SA, Gonder MK, Henn BM, Mortensen H, Knight A, Gignoux C, et al. History of click-speaking populations of Africa inferred from mtDNA and Y chromosome genetic variation. *Mol Biol Evol.* 2007;24:2180–95.
77. Rosa A, Ornelas C, Brehm A, VILLEMS R. Population data on 11 Y-chromosome STRs from Guine-Bissau. *Forensic Sci Int.* 2006;157:210–7.
78. Cruciani F, La Fratta R, Santolamazza P, Sellitto D, Pascone R, Moral P, et al. Phylogeographic analysis of haplogroup E3b (E-M215) y chromosomes reveals multiple migratory events within and out of Africa. *Am J Hum Genet.* 2004;74:1014–22.
79. Wood ET, Stover DA, Ehret C, Destro-Bisol G, Spedini G, McLeod H, et al. Contrasting patterns of Y chromosome and mtDNA variation in Africa: evidence for sex-biased demographic processes. *Eur J Hum Genet.* 2005;13:867–76.