Genomic landscape of human diversity across Madagascar

Denis Pierron[a], Margit Heiske[a], Harilanto Razafindrazaka[a], Ignace Rakoto[b,1], Nelly Rabetokotany[b,2], Bodo Ravolomanga[b], Lucien M.-A. Rakotozafy[b], Mireille Mialy Rakotomalala[b], Michel Razafiarivony[b], Bako Rasoarifetra[b], Miakabola Andriamampianina Raharijesy[b], Lolona Razafindralambo[b], Ramilisonina[b], Fulgence Fanony[b], Sendra Lejamble[c], Olivier Thomas[c], Ahmed Mohamed Abdallah[c], Christophe Rocher[c], Amal Arachiche[c], Laure Tonaso[a], Veronica Pereda-loth[a], Stéphanie Schiavinato[a], Nicolas Brucato[a], Francois-Xavier Ricaut[a], Pradiptajati Kusuma[a,d,e], Herawati Sudoyo[d,e], Shengyu Ni[f], Anne Boland[g], Jean-Francois Deleuze[g], Philippe Beaujard[h], Philippe Grange[i], Sander Adelaar[j], Mark Stoneking[f], Jean-Aimé Rakotoarisoa[b,3], Chantal Radimilahy[b,3], and Thierry Letellier[a,3]

[a]Laboratoire d'Anthropologie Moléculaire et Imagerie de Synthèse, Equipe de Médecine Évolutive, UMR 5288 CNRS, Université de Toulouse, 31073 Toulouse, France; [b]Institut de Civilisations/Musée d'Art et d'Archéologie, Université d'Antananarivo, Antananarivo 101, Madagascar; [c]Monitoring of Monoclonal Antibodies Group in Europe (MAGE) Consortium, INSERM U688, Université Bordeaux 2, 33000 Bordeaux, France; [d]Genome Diversity and Diseases Laboratory, Eijkman Institute for Molecular Biology, Jakarta 10430, Indonesia; [e]Department of Medical Biology, Faculty of Medicine, University of Indonesia, Jakarta 10430, Indonesia; [f]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, D-04103 Leipzig, Germany; [g]Commissariat à l'Energie Atomique, Institut Génomique, Centre National de Génotypage, 91000 Evry, France; [h]L'Institut des Mondes Africains (IMAF)–CNRS, 94200 Ivry-sur-Seine, France; [i]Département Langues Etrangères Appliquées, Université de La Rochelle, 17042 La Rochelle cedex 1, France; and [j]Asia Institute, University of Melbourne, Parkville VIC 3010, Australia

Contributed by Jean-Aimé Rakotoarisoa, June 15, 2017 (sent for review March 30, 2017; reviewed by Andres Ruiz-Linares and Antonio Torroni)

Although situated ~400 km from the east coast of Africa, Madagascar exhibits cultural, linguistic, and genetic traits from both Southeast Asia and Eastern Africa. The settlement history remains contentious; we therefore used a grid-based approach to sample at high resolution the genomic diversity (including maternal lineages, paternal lineages, and genome-wide data) across 257 villages and 2,704 Malagasy individuals. We find a common Bantu and Austronesian descent for all Malagasy individuals with a limited paternal contribution from Europe and the Middle East. Admixture and demographic growth happened recently, suggesting a rapid settlement of Madagascar during the last millennium. However, the distribution of African and Asian ancestry across the island reveals that the admixture was sex biased and happened heterogeneously across Madagascar, suggesting independent colonization of Madagascar from Africa and Asia rather than settlement by an already admixed population. In addition, there are geographic influences on the present genomic diversity, independent of the admixture, showing that a few centuries is sufficient to produce detectable genetic structure in human populations.

Indian Ocean | proto-globalization | genetics | Malagasy origins | genome-wide data

Ancient long-distance voyaging between continents stimulates the imagination, raises questions about the circumstances surrounding such voyages, and reminds us that globalization is not a recent phenomenon. Moreover, populations which thereby come into contact can exchange genes, goods, ideas and technologies (1). One of the most remarkable examples of such ancient intercontinental contact is the Malagasy, the Austronesian-speaking population that occupies Madagascar. Numerous theories have been proposed to explain the origin of the human diversity observed in Madagascar (summarized in ref. 2). Although historical, linguistic, ethnographic, archeological, and genetic studies confirm the dual African and Asian influences (3–11), no consensus exists regarding how, where, and when the two worlds met and merged. The lack of written history and the limited archeological evidence make it difficult to differentiate (*i*) founding myths and oral history, (*ii*) scientific hypothesis (*iii*), and pure speculation sometimes spread with political agenda. Because the ancestor "cult" is a fundamental aspect of Malagasy society, the roots of Malagasy population are a heated topic around the country. For instance, whether the Malagasy are of mainly African or Asian ancestry is still vigorously debated. Along with African and Austronesian connections (11), contributions from Arabic, Indian, Papuan, and/or Jewish populations have been suggested for a long time (12), as have the existence and heritage of the legendary first settlers of Madagascar, namely hunter-gatherers called variously "Vazimba," "Kimosy," or "Gola" (13).

Genetic data can illuminate population histories but are still limited and puzzling regarding Madagascar. Early studies in 1995 detected heterogeneity in Austronesian and Bantu ancestry

Significance

The origins of the Malagasy raise questions about ancient connections between continents; moreover, because ancestors are fundamental to Malagasy society, Malagasy origins is also a heated topic around the country, with numerous proposed hypotheses. This study provides a comprehensive view of genomic diversity (including maternal lineages, paternal lineages, and genome-wide data) based on a sampling of 257 villages across Madagascar. The observed spatial patterns lead to a scenario of a recent and sex-biased admixture between Bantu and Austronesian ancestors across the island. Moreover, we find geographical influences creating subtle signals of genetic structure that are independent of the Bantu/Austronesian admixture, suggesting that recent history has a role in the genomic diversity of the Malagasy.

Author contributions: D.P., M.H., I.R., N.R., B. Ravolomanga, L.M.-A.R., M.M.R., M.R., B. Rasoarifetra, M.A.R., L.R., R., F.F., O.T., J.-A.R., C.R., and T.L. designed research; D.P., M.H., H.R., I.R., N.R., B. Ravolomanga, L.M.-A.R., M.M.R., M.R., B. Rasoarifetra, M.A.R., L.R., R., F.F., S.L., O.T., C.R., L.T., V.P.-l., S.S., S.N., A.B., J.-F.D., M.S., J.-A.R., C.R., and T.L. performed research; I.R., N.R., B. Ravolomanga, L.M.-A.R., M.M.R., M.R., B. Rasoarifetra, M.A.R., L.R., R., F.F., S.L., O.T., A.M.A., C.R., A.A., V.P.-l., N.B., F.-X.R., P.K., H.S., P.B., P.G., S.A., M.S., J.-A.R., C.R., and T.L. contributed new reagents/analytic tools; D.P., M.H., H.R., A.A., M.S., C.R., and T.L. analyzed data; and D.P., M.H., H.R., M.S., J.-A.R., C.R., and T.L. wrote the paper.

Reviewers: A.R-L., University College London; and A.T., Università di Pavia.

The authors declare no conflict of interest.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. MF055747–MF058597) from the European Genome-Phenome Archive (EGAS00001002549).

[1]Deceased December 22, 2013.

[2]Deceased February 25, 2016.

[3]To whom correspondence may be addressed. Email: jarakoto@gmail.com, radimilahych@gmail.com, or thierry.letellier@inserm.fr.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1704906114/-/DCSupplemental.

according to the DNA loci studied (14, 15). In 2005, a study of mtDNA and Y chromosome variation in 37 Malagasy individuals identified approximately equal African and Indonesian contributions to both paternal and maternal Malagasy lineages (16). However, a later study of southeast Madagascar observed a strong sex bias, with the Austronesian ancestry appearing more conserved in the female gene pool, and a strong regional heterogeneity (17). Additionally, the first whole sequencing of mtDNA revealed that one maternal lineage thought to be of Austronesian origin (the M lineage) was actually a haplogroup (M23) specific to Madagascar, raising the possibility of a maternal heritage from hypothesized pre-Austronesian and pre-Bantu populations such as the Vazimba (13). However, the first study of genome-wide SNP data from southern Madagascar did not find any evidence of Vazimba heritage and instead proposed a recent (during the last millennium) admixture between Bantu-speaking and Austronesian-speaking groups (18). This study also argued for the predominance of African ancestry, in contradiction to a model-based study that suggested a maternal ancestry coming mainly from Asia (19).

These conflicting results suggest that the ancestry across Malagasy genomes may be highly heterogeneous according to the genetic loci and the geographical locations studied. Accordingly, a global study of maternal, paternal, and autosomal genetic variation across all of Madagascar is necessary to understand the settlement process fully. At present, the genetic ancestry of most of Madagascar remains unknown, and, given that the island is almost three times the size of Great Britain, settlement processes and admixture timing might have varied across the island. Thus, more comprehensive studies are needed to investigate the possibility of contributions of putative autochthonous populations and/or the existence of multiple waves of migration to Madagascar.

Results

Sampling Across 257 Villages. To address these issues, we present here a comprehensive study of Madagascar's population based on a grid-based sampling encompassing 257 villages across the island (Fig. 1*A*). We collected data on genetic diversity based on maternal mtDNA (full sequences from 2,691 individuals, 10.5 ± 3.5 individuals per village), paternal Y chromosome (genotyping of 1,554 male individuals, 6.0 ± 2.8 individuals per village), and genome-wide SNP data (700 individuals genotyped for 2.5 M SNPs, 2.8 ± 0.70 individuals per village).

Origin of Malagasy Genomic Diversity. Phylogenetic analysis of mitochondrial lineages shows that, with the exception of M23, all other lineages have been reported outside of Madagascar and have origins in either East Asia or Africa (Fig. 2*A* and Fig. S1). We find no evidence of maternal gene flow from Europe or the Middle East in our sample of 2,691 mtDNA sequences. Although the ratio of Asian/African maternal lineages varies across the island, the overall frequency of East Asian and African mtDNA lineages are roughly equal. All African lineages with a frequency >1% (Fig. 2*A*) are associated with Bantu-speaking groups (20–22), with the exception of haplogroup L2a1b1a, which is classified by previous studies as an East African haplogroup (but still could have been brought to Madagascar by Bantu-speaking groups). Although M23 has been found only in Madagascar so far, suggesting that it arose there, M23 has a recent origin (1,200 ± 300 y BP) (Fig. S1). Hence, M23 diversity does not support an ancient settlement of Madagascar by a putative ancient pre-Bantu/pre-Austronesian population such as the hypothesized Vazimba (13).

Overall, Y chromosome lineages of African origin are much more frequent in Madagascar than are lineages of East Asian origin (70.7 vs. 20.7%), in contrast to the mtDNA lineages (42.4 African origin vs. 50.1% East Asian origin) (Fig. 2*B* and Fig. S1). Other Y chromosome lineages with uncertain origins are also present; some of these (R1a, J2, T1, G2) are also present in the Middle East and may reflect the Muslim influence on Madagascar and the Comoros (23, 24). Haplogroup R1b, characteristic of western Europeans, is present in low frequency (0.9%), suggesting a limited paternal contribution from western Europeans.

Admixture analyses (25) of the genome-wide data, based on various datasets (Dataset S1), confirm Southeast Asian and East African Bantu groups as the major contributors and exclude any major influence from other parts of the world (Fig. 3). Admixture results at k = 3 from the high-density panel produced a clear distinction between African, East Asian, and West Eurasian populations, permitting us to estimate the frequency of each ancestry component across Madagascar (Fig. S2). On average, the African

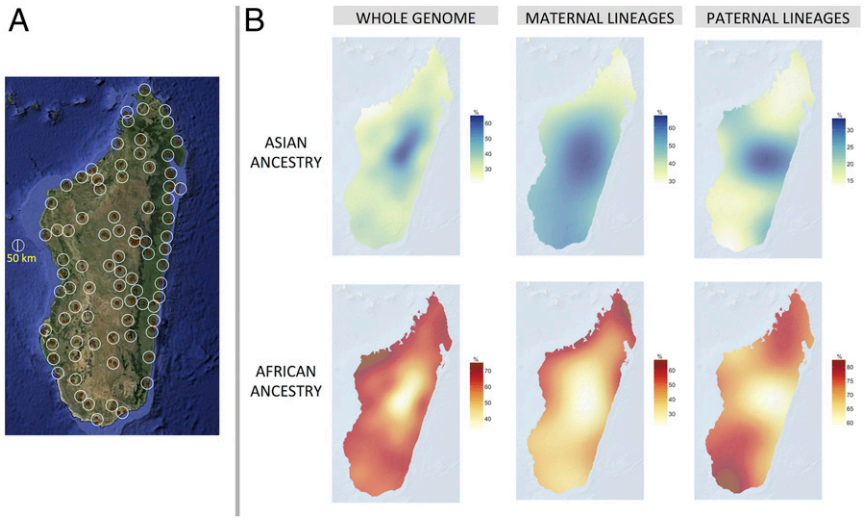

Fig. 1. Geographic distribution of Asian and African genetic ancestry across Madagascar. (*A*) Sampling grid across Madagascar: Three to four villages were sampled in each of 82 spots that are each 50 km in diameter. Image courtesy of Google Earth ©2016 TerraMetrics. (*B*) Exponential kriging interpolation of the ancestry across the Madagascar landscape based on the frequency of mtDNA lineages and Y chromosome lineages in each village and the average of Admixture (k = 3) analysis for genome-wide data based on the high-density panel.

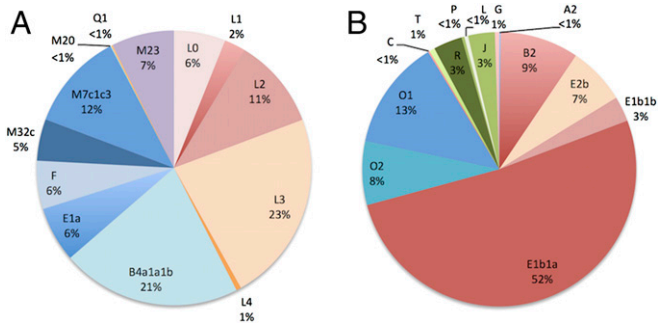

Fig. 2. Uniparental lineages. (*A*) Distribution of mtDNA lineages according to continental origin. Asian lineages are in blue, African lineages are in red, and M23 (unknown origin) is in purple. (*B*) Distribution of Y chromosome lineages according to continental origin. Asian lineages are in blue, African lineages are in red, and Eurasian lineages are in green.

component is 59.4 ± 0.4%, and the Asian component is 36.6 ± 0.4%, whereas the West-Eurasian component is only 3.9 ± 0.1%. All studied individuals present a similar pattern associating African and East Asian components, but with considerable variation (with African ancestry ranging from 26.1 to 92.6%).

Ancestral contributions were also assessed by computing the chromosome fragments shared between Malagasy and other populations (Fig. 3 and Fig. S3). Using reference populations sampled from across the world, we found that Bantu and Indonesian populations share most of the large fragments identical by descent (IBD >2 cM) with Malagasy individuals. These analyses exclude other (e.g., Indian, Ethiopian, or Somali) populations as putative major contributors (Fig. S3). Nevertheless, there are potential minor contributions; for example, one Malagasy individual shared on average one fragment with the French Basque population. Using a second set of reference populations more centered on the Indian Ocean, we confirmed the close link with Bantu and Indonesian populations. On average Malagasy individuals share 4.32 ± 0.04 fragments with Bantu populations in general and share even more (5.5 ± 0.05 fragments) with the Bantu populations from southeast Africa (Fig. S3). On the Asian side, populations from Indonesia, especially south Borneo, share the highest number of fragments (1.23 ± 0.01), suggesting that, among the populations sampled from outside Africa, they are the closest link with the Malagasy. Based on IBD sharing, demographic simulation of the split between the Malagasy and source populations from south Borneo led to two scenarios with similar likelihoods: (*i*) a hard split 2,500 y BP and (*ii*) a slow divergence with less and less migration between 3,000–2,000 y BP (Fig. S4). The split from south African Bantu groups seems to have occurred much more recently, around 1,500 y BP.

All analyses thus converge on two main ancestries for the entire Malagasy population, namely Bantu from southeast Africa and Austronesians from Indonesia (in particular, south Borneo), with a very limited contribution from Europe and the Middle East. The IBD-based model suggests that the split between Malagasy and south Borneo is older than the split between Malagasy and southeast African Bantu, indicating that Indonesians populations might have arrived before African populations.

Geography of Malagasy Genomic Diversity. We next investigated the geographical distribution of African and Asian ancestry across Madagascar and found significant differences (Fig. 1*B*). All three genomic components (mtDNA, Y chromosome, and genome-wide) are highly correlated with geography [Moran's autocorrelation coefficient (*I*) for all analyses: $P < 10^{-5}$]. Maternal African lineages are present mostly in the north of the island and are even in the majority in the extreme north of Madagascar, whereas maternal lineages from Asia are in higher frequency in the center and the south of the island. In contrast, Asian paternal lineages are much lower in frequency, reaching only 30% in the center, and African paternal lineages are present mostly on the coast and in the north of Madagascar. The distribution of ancestral components based on genome-wide data indicates that people in the highlands in the center of the island have mostly Asian ancestry (>65%), whereas people from the coastal regions have higher African ancestry (>65%).

To test the existence of further genetic structure linked to geography, we performed hierarchical clustering of the genome-wide data via fineSTRUCTURE (Fig. S5) (26). Because the fineSTRUCTURE algorithm does not use geographical information, the correlation between the geographical position and genetic cluster assignment of individuals suggests the existence of an effect of geography on Malagasy diversity (27). Examining the different levels of clustering indicates how human genetic structure varies across the Madagascar landscape (27). At the lowest level (k = 2) (Fig. S5*D*), the distributions of the two clusters are significantly correlated with geography (Moran's *I* all $P < 10^{-5}$) and distinguish highland from coastal regions, similar to the geographical pattern observed in the admixture analysis. Furthermore, the ratio of African/Asian ancestry in these two clusters differs significantly: The cluster in the center of the island has a mean Asian ancestry of 68%, vs. 38% in the other cluster (Wilcoxon tests; $P < 10^{-16}$). This result means that the main genetic structure of the present Malagasy population reflects variation in the amount of African vs. Asian ancestry. However, increasing the level of clustering produces new clusters that are also correlated with geography: At k = 3 a new cluster appears in the center; at k = 4 and k = 5 new clusters appear in the north; and from k = 6 to k = 8 new clusters appear in the south (Fig. S5*D*). For all clusters at all levels of analysis there is a significant correlation with geography (Moran's *I*, all $P < 10^{-5}$). The spatial distribution at each level of genetic clustering thus reveals a strong effect of geography on present genomic diversity.

This result also suggests that the genome-wide diversity of Malagasy populations is not structured solely by a simple dichotomy of African vs. Asian ancestry. To study the role of ancestry in the present genetic structure in more detail, we analyzed a division of the fineSTRUCTURE tree into 10 genetic groups (g1–g10) (Fig. 4). Although there is no optimal level of clustering, and all levels of clustering are informative (27), this level of clustering has the advantages of giving a fairly large number of clusters to investigate fine-scale differences among clusters along with enough individuals per cluster (between 50–100 individuals) that differences between clusters are likely to be real and not an artifact of small sample sizes. Fixation index (F_{st}) distances between these genetics groups are low (F_{st} = 0.0075) (Fig. S6), similar to F_{st} values between populations living in Great Britain (27), suggesting the same level of differentiation.

As seen with lower levels of clustering, the amount of African vs. Asian ancestry varies significantly across the genetic groups (*F* value 551.6, $P < 2\ 10^{-16}$; ANOVA) (Fig. S6). For example, the Asian component is dominant in the highland cluster g01 (65 ± 0.5%), whereas it is present at only 22.6 ± 0.7% in the northern cluster g03. However, differences in African vs. Asian ancestry cannot explain all the observed structure, because several genetic groups do not differ significantly in terms of African vs. Asian ancestry. For example, although there is a marked difference in geographic distribution between genetic groups 7 and 8, their percentages of African and Asian ancestry are nearly identical [$P > 0.99$; Tukey's honestly significant difference (HSD)] (Fig. S6*C*).

Because it has been suggested that different source populations were involved in the settlement of different regions of Madagascar (28), we tested whether different populations in Indonesia or Africa might be closer to different Malagasy genetic groups by performing IBD analysis group by group (Fig. S3 *C* and *D*). We

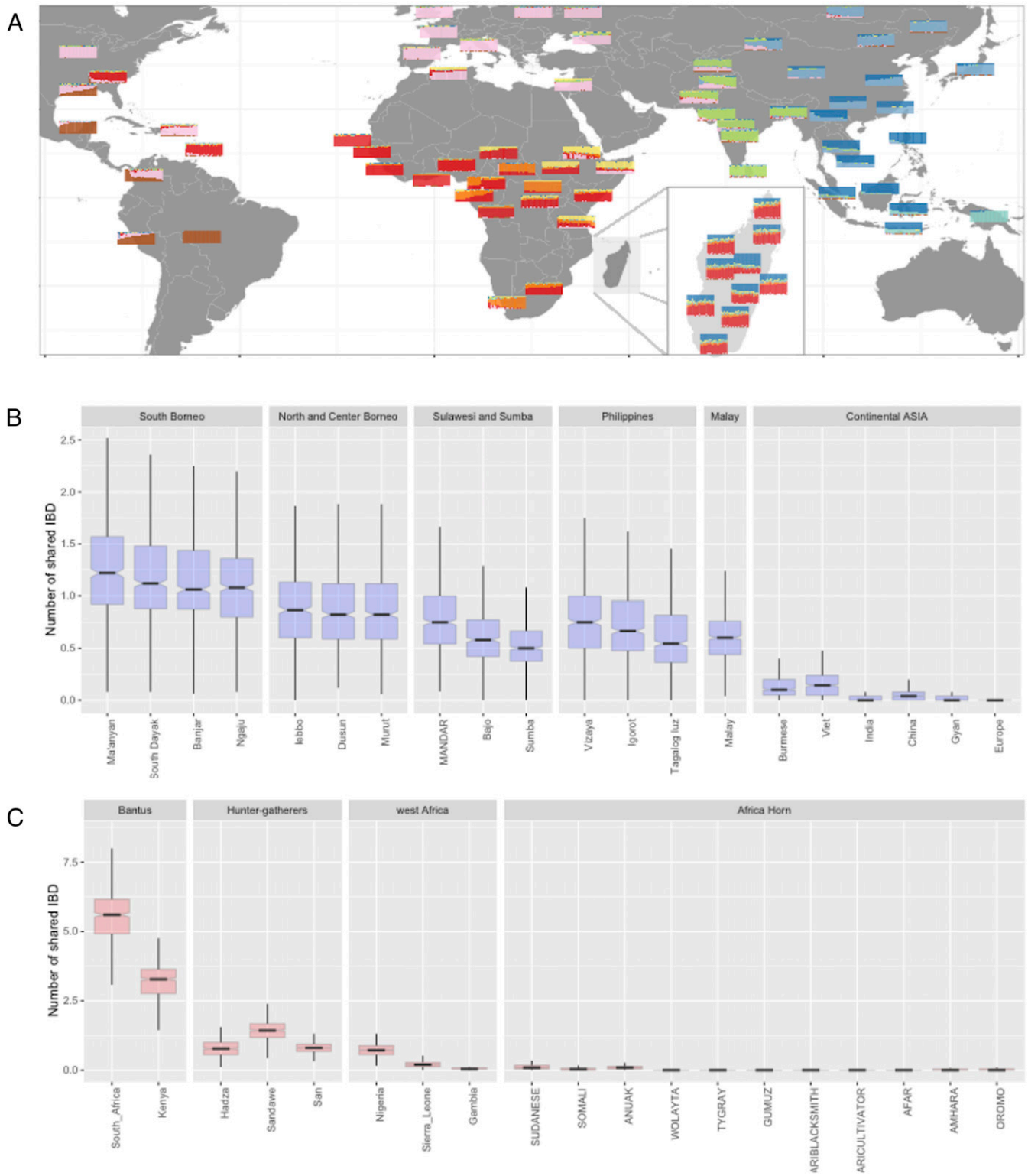

Fig. 3. Genome-wide analysis: Admixture plots based on a worldwide population panel (Dataset S1: low density) for k = 9. (*A*) Boxplot of the distribution of shared IBD segments between any Malagasy and any individual from the reference population in high-density panel 2 from Eurasia in blue (*B*) or Africa in red (*C*).

obtained very similar results for all populations; the highest number of shared IBD is always with south African Bantu, and on the Asian side the highest is always with south Borneo. Moreover, based on PCAdmix it is possible to deconvolute ancestry at each locus in the genome and then analyze the African and Asian ancestry in each Malagasy genetic group separately. Based on this deconvolution, TreeMix analyses run separately with African and Asian populations confirm that all Malagasy genetic groups share the same origin; all diverge from both south Borneo and Bantu populations (Fig. S6). Both analyses thus indicate that differential

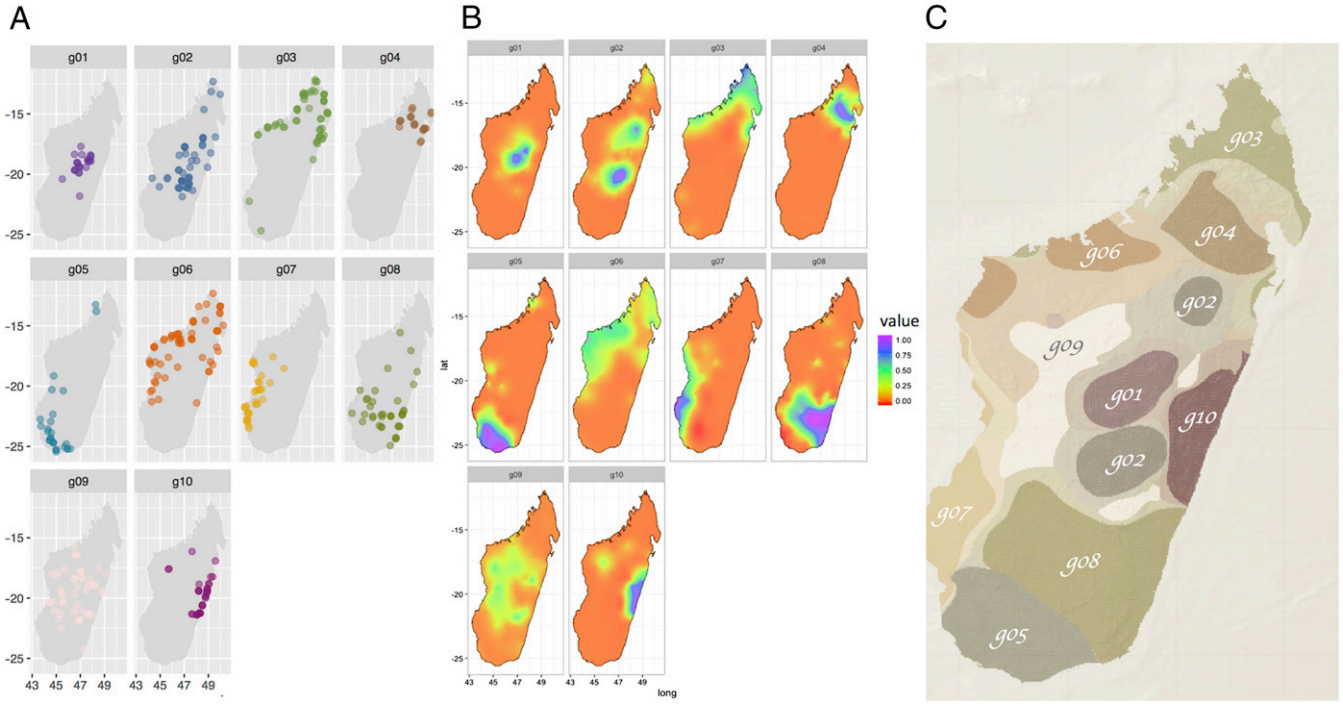

Fig. 4. Genetic groups. (*A*) Geographic distribution of genetic groups in Madagascar. Each dot represents a village, and the intensity of the color corresponds to the relative presence of individuals of each group. (*B*) Kriging model of the spatial distribution of genetic groups based on the frequency of each group in each sampled village. (*C*) Superposition of all genetic groups distributions (based on kriging model). Colors were assigned according to the dominant genetic cluster present in a given area. Plain colors were used for locations where the majority of people (>50%) belong to this cluster.

origin does not explain the observed differences among the genetic groups and thus suggest that the detectable genomic structure of the Malagasy population that is correlated with geography is not solely the result of the admixture and settlement process but might also reflect the later history of Madagascar.

Demographic History. To study the tempo of the settlement of Madagascar, we computed the admixture time and inferred the demographic history of the genetic groups defined by fineSTRUCTURE. We computed the time since admixture using two different methods, namely GLOBEtrotter and ALDER. The GLOBEtrotter analysis indicates that all genetic groups are from an admixture with two sources: south African Bantus and a south Borneo population, i.e., Benjar for g01 and south Dayak for the others (for all analyses, $r^2 > 0.99$, $P < 0.01$). For all populations but one, the GLOBEtrotter results suggest a single admixture event rather than multiple admixture events. GLOBEtrotter suggests two admixture events only for genetic group g07, with the first admixture occurring 28 generations ago and the second occurring four generations ago. Both GLOBEtrotter and ALDER analysis date the single admixture event to between 500 and 900 y BP. The oldest admixture occurred 800 ± 25 y BP in the eastern populations (g10, g08, and g04 based on GLOBEtrotter) (Fig. 5*B*), whereas the most recent admixture events (665 ± 19 y BP) involve genetic group g03, which is the most northern genetic group and also has the most African ancestry (Fig. S7 and Dataset S2). The significant differences in admixture dates and in the percentage of African/Asian ancestry between the genetic groups suggest independent admixture events across Madagascar rather than settlement by an already admixed population.

Demographic inference for the entire Malagasy population based on IBD sharing (*Materials and Methods* and Fig. 5*C*) suggests a population expansion beginning between 1,250 and 1,000 y BP. Separate analyses for each genetic group present similar patterns, with expansions between 1,250 and 750 y BP. The earliest expansion is g03, from the north of the island; g01 (in the center) underwent a strong bottleneck, with a reduction in population size to a few hundred people between 1,000 and 800 y BP. We also observed a decrease in the size of g05 (from the south) between 500 and 250 y BP (Fig. S7).

Discussion

This study presents an extensive overview of the genetic diversity across Madagascar, providing comprehensive insights into the settlement of the island Fig. 6). The present Malagasy population shares recent common ancestors with Bantu and Austronesian populations now living 8,000 km apart (Fig. 6*A*). The distribution of African and Asian ancestry across the island reveals that the admixture was sex biased and happened heterogeneously across Madagascar, suggesting independent colonization of Madagascar from African and Asians populations (Fig. 6*B*). After the admixture, further events led to a finer-scale genetic structure (Fig. 6*C*), despite the recent internal migration reported by historians (Fig. 6*D*).

Our results indicate that across the entire country all Malagasy individuals share recent Austronesian and Bantu ancestry (Fig. 6). We identified a recent split of the proto-Malagasy population from southern African Bantus around 1,500 y BP and an older split from south Borneo between 3,000 and 2,000 y BP. This result suggests that Indonesians populations may have arrived on Madagascar before African populations. However, these dates reflect the age of the oldest possible common ancestors between Malagasy and the African/Indonesian sampled populations, meaning that the departure to Madagascar is not later but could be earlier than these dates. Our large sampling across Madagascar indicates a link to the south Borneo region (confirming a link to Ma'anyan-related populations) and does not support specific genetic connections with Sulawesi or Malays (29, 30). However, it is possible that more closely related populations exist in regions for which we lack data (e.g., Java or Mozambique),

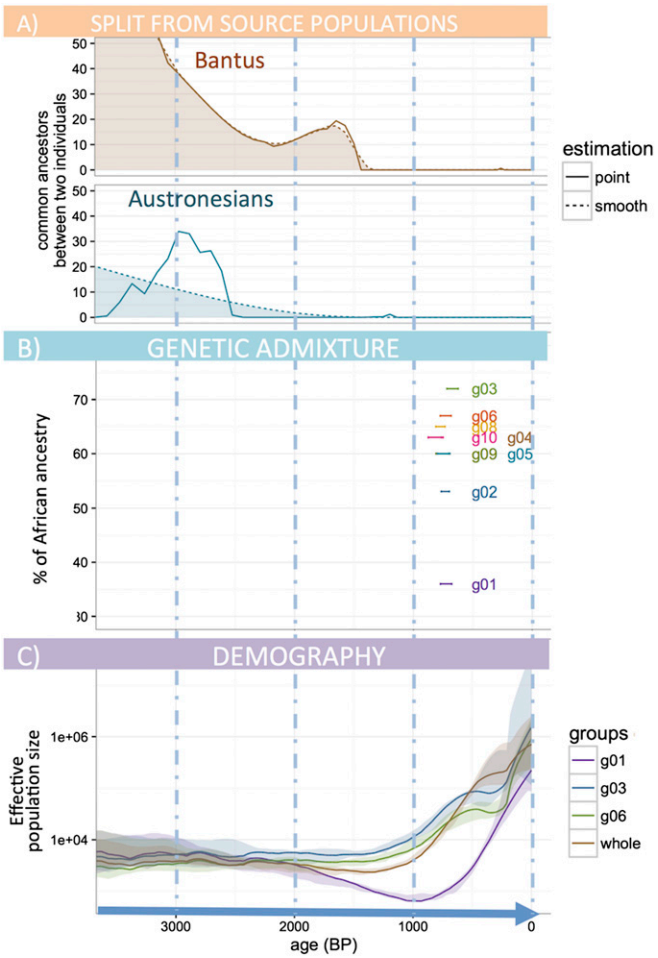

Fig. 5. Demographic model of the settlement of Madagascar. The plots in the three panels share the same *x* axis. (*A*) Timeline of the point and smoothed estimation of the average number of common ancestors shared between Malagasy and Borneo (in blue) and Malagasy and Bantu (in red) populations, as estimated by shared IBD segments from genome-wide data. (*B*) Estimation of percentage and date of admixture for each genetic group of Madagascar. The straight line represents the uncertainty (±1 SD) of admixture dates estimated from GLOBEtrotter. Some populations overlap. (More detailed results are provided in Fig. S7*B*). (*C*) Estimation of changes in the effective population size across time for selected genetic groups and for the whole Malagasy population, estimated by shared IBD genome-wide.

and it is also possible that the present-day locations of the putative source populations are not where they were in the past.

We also detected a small contribution from Middle East and European populations (Fig. 6*A*), probably connected to Swahili populations and the Arab world. However, we did not detect substantial genetic ancestry from ancient pre-Bantu/pre-Austronesian populations, as previously proposed based on the mtDNA M23 haplogroup (13), or any genetic contribution from Indian populations. However, minor contributions might have happened, and further work based on genome-wide sequencing and/or additional reference populations might be able to identify such small contributions. It is also possible that there were pre-Austronesian/Bantu people in Madagascar but that they did not contribute any genetic ancestry to the present Malagasy population; hence future work involving ancient DNA will be crucial to test such possibilities.

The dates for the split from the source populations, along with the admixture dates and demographic analyses, all indicate that the settlement of Madagascar by the ancestors of the present Malagasy populations was a recent and rapid process, with the admixture happening within the last 1,500 y (Fig. 6). That the African ancestry on Madagascar falls within the present Bantu genetic diversity and the high number of African mtDNA haplogroups also suggest recent settlement.

The significant differences in admixture dates for the genetic groups suggest independent admixture events across Madagascar (Fig. 6). This assumption is reinforced by the variation in amounts of African and Asian ancestry across the island and by the heterogeneity in sex bias across Madagascar. These observations strongly support the separate arrival of Asians and Africans across Madagascar, with subsequent genetic mixing occurring independently across the island. North Madagascar appears to be the principal landing zone for African populations, with demographic expansion beginning before 1,000 y BP. Northern populations have the highest amount of paternal, maternal, and autosomal African ancestry, and their identification as the major substrate for the African ancestry across the rest of the island is supported by the following evidence: the TreeMix analysis of African ancestry places the northern population as the root of the tree (Fig. S6*D*); the earliest demographic expansion occurred in the north (Fig. S7*A*), which is the location with the highest mtDNA diversity (Fig. S1*D*); and ChromoPainter analysis showed that all other genetic groups located in the coastal area of Madagascar share the most genetic ancestry with the northern genetic group g03 (Fig. S5*A*). One potential complication with this scenario is that genetic group g03 (located in the north and with the most African ancestry) has the most recent date for admixture, suggesting a later and limited diffusion of Asian gene flow in this population. Another complication with this interpretation of the data would be the existence of a possible recurrent genetic exchange between Africa and Madagascar, even if it is not detected by ChromoPainter. More genomic data from east Africa would be needed to explore this possibility. Nevertheless, the scenario of an initial colonization in the north is also in agreement with archaeological evidence indicating permanent occupation in northern Madagascar at the site of Mahilaka (31). Other archaeological sites in the north, such as Lakaton'I Anja, Irodo, and Iharana, demonstrate trade connections with East Africa (6, 32, 33, 34). However, in arid southern Madagascar there are trade centers that are contemporary with but much smaller than Mahilaka (such as the archaeological sites of Andranosoa, Mahirane, and Andaro), and these also indicate permanent occupation and trade connections with East Africa (11, 35, 36).

Although northern populations present a majority of both maternal and paternal lineages from Africa, across the other

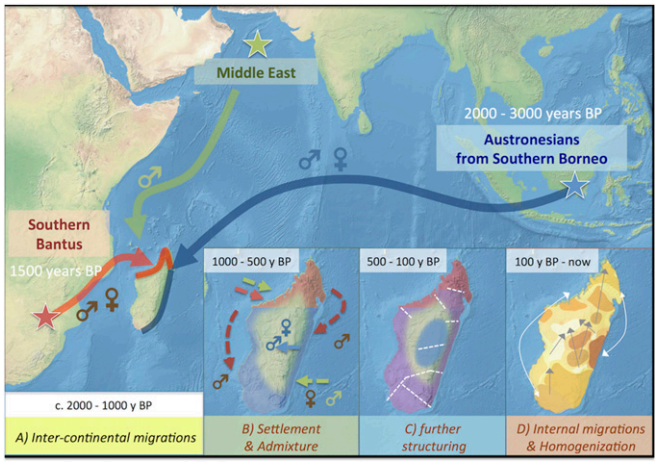

Fig. 6. Overview of the inferred history of Madagascar. Descriptions and dates are given in *A*–*D*.

parts of Madagascar the maternal lineages are predominantly from Island Southeast Asia, whereas paternal lineages are mainly from Africa (Figs. 3 and 6). This difference suggests a strong sex bias involving contributions from Bantu males diffusing from the north to the south. The earliest admixture dates are around 800 y BP on the Madagascar eastern coast, suggesting that the southeast was already settled by Austronesians (men and women) before the arrival of Africans (primarily men). The TreeMix analysis of Asian ancestry is compatible with this scenario because the root of the Asian ancestry is on the northeast coast, with a rapid diversification of other populations to the south. The hypothesis that Austronesians were the first to settle Madagascar before an African paternal wave is supported by the earlier split of Malagasy from Indonesian source populations and explains the predominance of both Austronesian maternal lineages and the Austronesian linguistic background. The known archeological sites of fishermen and rice cultivators in the south, such as Maliovola, Mokala, and Ambinanibe, dated to the ninth century AD onwards (37), might be related to the Austronesian colonists; future paleogenetic studies might be informative.

Our analyses also reveal a singular history for the Central Highlands: Contemporaneous with the admixture progressing across Madagascar, there was a drastic decrease in the effective size (down to a few hundred persons) of the population now located in the Central Highlands (Fig. S7*A*). Further, the Bantu contribution to this population was limited (~32% based on genome-wide SNPs, 8.8% for maternal lineages, and 55% for paternal lineages). Because the effective population size reflects the number of breeding individuals in a population, this decrease is not necessarily representative of a dramatic event such as disease or famine but instead likely reflects a demographic event such as the migration of a small founding population. It appears that there was a late founder effect in the settlement of the Central Highlands by a small number of individuals (mainly with genetic ancestry from Borneo) while admixture was happening across the rest of the country.

Our study shows a strong correlation between geography and genomic diversity across Madagascar. To be sure, the genetic groups we identified are based on arbitrary criteria, and there is no method for identifying the "true" number of genetic groups. However, the distributions of the 10 genetic groups we analyzed are strongly influenced by geography, suggesting that they reflect in some sense the past demographic history of the Malagasy. Interestingly, many of these genetic groups overlap populations presented in the various controversial ethnographic descriptions made by explorers and other scholars in the 20th century (38, 39), and those descriptions may, in turn, reflect the influence of ancient kingdoms across Madagascar (40). In agreement, our study attests that the genetic structure is young and not necessarily due to the result of different population sources. Undoubtedly other factors have influenced the genetic structure of Madagascar; nevertheless, our study shows that in the few centuries since admixture, these factors have produced a subtle but nonetheless detectable structure in the Malagasy that is independent of the African/Asian admixture, even despite the higher levels of internal migration reported during the last century (Fig. 6) (38).

Materials and Methods

Sampling. The samples analyzed in this study were collected during 2007–2014 with ethical approval by the Human Subjects' Ethics Committees of the Health Ministry of Madagascar and by French committees (Ministry of Research, National Commission for Data Protection and Liberties and Persons Protection Committee). Individuals were given detailed information about the study, and all gave written consent before the study. DNA was purified from saliva using the Oragen Kit (DNA Genotek Inc.). The extensive sampling was based on a grid sampling approach, in which 82 "spots" 50 km in diameter were placed all over Madagascar (taking into account population density data), and three to four villages were sampled in each spot (Fig. 1). Sampled villages were founded before 1900, and sampled individuals were 61 ± 15 y old, with the maternal grandmother and paternal grandfather born within a 50-km radius of the sampling location. Subjects were surveyed for current residence, familial birthplaces, and a genealogy of three generations to establish lineage ancestry and to select unrelated individuals. A total of 2,704 individuals from 257 villages were sampled (10.5 ± 3.5 individuals per village). Global Positioning System locations were obtained during sampling.

Uniparental Markers. Whole mtDNA genome sequences were obtained from 2,691 individuals from 256 villages (10.5 ± 3.5 individuals per village). Multiplex sequencing libraries were constructed and enriched for mtDNA sequences as described previously (41). A double-indexed Illumina sequencing library, with barcodes specific for each sample, was prepared from each extract. Up to 250 libraries were pooled in an equimolar ratio, and mtDNA sequences were enriched via in-solution capture (42). The capture-enriched library pools then were pooled in an equimolar ratio into a single pool, which then was sequenced on eight lanes on an Illumina HiSeq2000 platform with 95-bp paired-end reads. Base-calling was performed using freeIbis (43). Reads then were mapped to the revised Cambridge reference sequence (rCRS) (44) and assembled as described previously (41). Duplicate reads were removed along with reads with a mapping quality score lower than 20 and a base quality score lower than 20. The average coverage per position per individual was 543.2 ± 9.2 reads. The dataset is available from GenBank (accession nos. MF055747–MF058597). Samples then were aligned with Clustal to the rCRS. Haplogroups were assigned to consensus sequences for each sample with the HaploGrep webtool (45) and PhyloTree Build 15 (46). For subsequent analysis only sequences lacking gaps and with a minimum coverage of 15× per position were retained (i.e., 2,409 genomes). For all analyses except haplogroup assignment, the poly-C regions (positions 303–315 and 16,182–16,193) were removed from all sequences. To reconstruct the M23 phylogeny, other sequences belonging to haplogroup M23 reported in PhyloTree Build 15 (46) were also aligned, and a maximum parsimony tree was constructed based on all positions and the MJ algorithm (47); the root age of the haplogroup was computed using the ρ statistic and a mutation rate of one synonymous mutation per 7,884 y (48).

Y chromosome haplogroups were determined for 1,554 male individuals (6.7 ± 2.6 individuals per village) by following a previously described method (30). Briefly, 96 binary markers (all located on the nonrecombining region of the Y chromosome) (Dataset S3) were analyzed with a high-throughput genotyping system (nanofluidic Dynamic Array; Fluidigm), and the results were analyzed using the BioMark HD system (Fluidigm), which integrated the real-time PCR Analysis software. Each haplogroup was assigned according to the updated Y-PhyloTree (49).

Estimating Population Sources from the Genome-Wide Dataset. To identify the ancestral source populations and to estimate the admixture fractions of the Malagasy population, we performed a structurelike analysis using the Admixture software (25) after thinning the marker sets for linkage disequilibrium. We used the Plink software to remove each SNP with an r^2 value greater than 0.1 with any other SNP within a 50-SNP sliding window (advanced by 10 SNPs each time) (50). Admixture was run using the projection mode, i.e., by projecting the Malagasy individuals onto the reference dataset. We performed these analyses using three reference datasets (Dataset S1), keeping only overlapping sets of compatible SNPs after correcting for strand consistencies: (*i*) a worldwide analysis based on the Centre d'Étude du Polymorphisme Humain Human Genome Diversity Panel (CEPH-HGDP) (51) and the1,000 Genomes Project (52) populations, to which we added several African populations (6,637 SNPs after pruning; populations are listed in Dataset S1); (*ii*); an analysis focused on the Asian ancestry using the Pan-Asian dataset (53) (12,689 SNPs after pruning) along with one east African population from the 1,000 Genomes Project; and (*iii*) a high-density panel with more SNPs (high-density 1) with the 1,000 Genomes Project populations, Khoisan-speaking African populations, and a high-density Indonesian dataset (54) (184,658 SNPs before pruning and 75,410 after pruning; populations are listed in Dataset S1). Admixture results at k = 3 from the high-density panel produced a clear separation between African, East Asian, and West Eurasian populations (Fig. S2); we therefore used the results for k = 3 to estimate each ancestry level across Madagascar. The geographic distribution of African and Asian ancestry was analyzed by computing Moran's *I* using the Analysis of Phylogenetics and Evolution (ape) package from R (55) and gradient plots computed using the exponential kriging model in the package geoR (56).

Population Structure. The genome-wide dataset was generated for 700 individuals from 253 villages (2.8 ± 0.7 individuals per village) using the Illumina

Human Omni 2.5-8 (Omni 2.5) BeadChip array. Analyses were performed using Plink 1.9 (57). All genotyped individuals passed the quality filters, i.e., had genotype call rates higher than 95%, and were not close relatives (identity by descent estimation under the threshold of 0.25). Analyses were performed on 2,268,323 SNPs. The dataset is available from the European Genome-Phenome Archive (ega-box-658).

Population structure was analyzed using the fineSTRUCTURE approach (26). The first step of this method (ChromoPainter) examines each segment of the autosomal genome of one individual and determines which specific individual in the rest of the population shares the most homologous fragment. By assuming that the number and the size of shared fragments between two individuals depend on the ancestors shared by these two individuals, this step provides a coancestry matrix between all pairs of individuals. For this purpose the autosomal haplotypes were inferred, and IBD was searched for all individuals by phasing using Beagle version 4.1 (58) (ibdlod = 3; ibdtrim = 40, Grch37 genetics maps) followed by analysis with the ChromoPainter program (26). Following the authors' instructions, we first ran ChromoPainter on chromosomes 3, 7, 8 and 10, weighting each chromosome by their relative size, on a subset of individuals and using 10 iterations of the expectation-maximization algorithm to infer the genome-wide average switch and global emission rates. Then, using these inferred values, we ran ChromoPainter on all individuals and chromosomes to produce the counts and lengths of fragments shared between individuals.

In the second step we ran the fineSTRUCTURE program on the coancestry matrix based on counts of shared fragments. fineSTRUCTURE is a model-based statistical algorithm that uses a Markov chain Monte Carlo (MCMC) approach (26). Initially all individuals were set into a single cluster at iteration 0. Following 10 million burn-in iterations, we sampled values every 10,000 iterations for 10 million MCMC iterations. At the end, fineSTRUCTURE provided 61 clusters of individuals and the cluster membership of each individual. Then similar clusters were merged hierarchically to give a tree, which can be used to describe population structure at different levels. Finally, we improved individual clustering, as described elsewhere (27). The tree should not be seen as a phylogenetic tree, and all levels of the tree are informative (27). We defined clusters for further analysis as the highest-level monophyletic groups with less than 100 individuals, thus leading to 10 genetic groups with sample sizes of at least 50. The geographic distribution of each of these genetic groups was analyzed by computing Moran's *I* statistic, which measure the spatial autocorrelation, using the R package ape (55). The geographic distribution of genetic groups can be considered as post hoc evidence of true clustering (27).

Gradients of the distribution of each genetic group were computed in R using the exponential kriging model in the package geoR (56); all gradients were merged into a single figure. For this purpose, each location on the final map was colored by the color of the main cluster, and the color was attenuated if the principal cluster represented less than 50% of the individuals. All graphs were produced with the ggplot2 package (59). F_{st} values were computed between each genetic group using Plink (57).

We tested if the genetic admixture was significantly different across the genetic groups using ANOVA and the Tukey HSD statistic from the package stats in R.

IBD Statistics. The number of IBD segments shared between each pair of individuals was estimated from the phased SNP high-density dataset by the Refined IBD algorithm in Beagle v4.0 (58, 60), filtering for detected fragments with a logarithm of odds ratio >3 (ibdlod = 3; ibdtrim = 40). From these results the number of shared IBD segments >2 cM was used to compute the distribution between each Malagasy genetic group and the population from the two high-density panels, using scripts in R. The distribution of shared IBD segments was used to compute the number of shared ancestors across the past 2,000 y between populations, using a generation time of 30 y, as done previously (61). A refined IBD algorithm also was performed on the Malagasy individuals alone to estimate the history of population-size changes in the Malagasy population using nonparametric estimation (62). We computed the demographic history of the Malagasy population as a whole and by genetic group.

TreeMix. To analyze the Asian and African ancestry in the Malagasy populations separately, we identified local ancestry using PCAdmix (63) and a window of 1 cM for all Malagasy individuals phased with the high-density panel. We used two parental populations: one African metapopulation grouping Somali, San, south African Bantu, and Luhya, and one Asian metapopulation grouping Han Chinese, Igorot, Ma'anyan, south Dayak, Mandar, and Malay groups. Then for each SNP and each Malagasy genetic group (determined previously by fineSTRUCTURE) we computed the frequency of each haplotype for the Asian component and for the African component. We then performed two analyses with TreeMix v1.12 (64): one with the Asian component of the Malagasy populations with all Asian populations plus one African population as an outgroup, and one with the African component of the Malagasy populations with all African populations plus one Asian population as an outgroup. TreeMix was run to build the maximum likehood tree with blocks of 2,000 SNPs to account for linkage disequilibrium, and the tree was drawn using FigTree (tree.bio.ed.ac.uk/software/figtree/).

Admixture Date and Scenario. To estimate the admixture scenario and corresponding date, we performed a GLOBEtrotter analysis (65) based on the high-density panel 1 data phased by Beagle. We used the ChromoPainter program to produce the coancestry matrix using linkage disequilibrium. We initially ran the model using 10 iterations to estimate the recombination scaling constant and mutation probabilities. Using these parameters, we ran ChromoPainterv2 using all populations except the Malagasy as donor populations and all individuals as recipients. GLOBEtrotter (65) then was used following the authors' instructions: For each Malagasy genetic group from the fineSTRUCTURE analysis we first ran GLOBEtrotter, with each genetic group as the target population to determine the *P* value for evidence of any detectable admixture. We then estimated the date of admixture and confidence intervals, as well as the evidence for simple admixture involving two admixing sources coming together at a single time, versus a complex admixture involving multiple sources and/or multiple dates of admixture. Finally we also estimated the admixture date using ALDER (66) and all Asian and African populations from the high-density panel.

ACKNOWLEDGMENTS. We thank all participants in the study; the students and staff from the Institut de Civilisations/Musée d'Art et d'Archéologie who have contributed to the sampling across Madagascar (T. Ramaharoarivo, M. Ramaheninjohary, M. Felantsoa, S. Rakotonandrasana, F. Andrianjafy, N. Randrianarivelo, A. Jao, A. Rakotoarison, T. Rakotovao, G. Herisoa, A. Fanomezantsoa, N. Ratsimbazafy, S. Herifanomezanjo, N. Rakotondrasoa, J. Raharinoro, L. Rabearisoa, S. Razafiarivony, O. Noarizafy, H. Rasolonantenaina, N. Haifetra, H. Razafindrainibe, R. Rakotomalala, Ismael, T. Rasolondrainy, C. Rasolonirina, D. Razanatolonjanaharitovo, V. Razanatovo, Zova, and P. Ratsimiveh); R. Schröder for technical assistance; and E. Crubezy for his support. This research was funded by Région Aquitaine "Projet MAGE" (Monitoring of Monoclonal Antibodies Group in Europe); French National Research Agency (ANR) Grants ANR-12-PDOC-0037-01 "GENOMIX" and ANR-14-CE31-0013-01 "OceoAdapo"; and Le Service de Coopération et d'Action Culturelle (SCAC) from l'Ambassade de France à Antananarivo, Région Midi-Pyrénées.

1. Crowther A, et al. (2016) Ancient crops provide first archaeological signature of the westward Austronesian expansion. *Proc Natl Acad Sci USA* 113:6635–6640.
2. Beaujard P (2012) *Les mondes de l'océan Indien* (Armand Colin, Paris).
3. Adelaar K (1989) Malay influence on Malagasy; historical and linguistic inferences. *Ocean Ling* 28:1–46.
4. Allibert C (2007) Migration austronésienne et mise en place de la civilisation malgache. Lectures croisées: Linguistique, archéologie, génétique, anthropologie culturelle. *Diogène* 2:6–17.
5. Beaujard P (2011) The first migrants to Madagascar and their introduction of plants: Linguistic and ethnological evidence. *Azania* 46:169–189.
6. Dewar RE, et al. (2013) Stone tools and foraging in Northern Madagascar challenge Holocene extinction models. *Proc Natl Acad Sci USA* 110:12583–12588.
7. Ferrand G (1909) L'Origine Africaine des Malgaches. *Bull Mem Soc Anthropol Paris* 10:22–35.
8. Grandidier A (1901) *L'Origine des Malgaches* (Imprimerie Nationale, Paris).
9. Kent RK (1970) *Early Kingdoms in Madagascar, 1500-1700* (Holt, Rinehart and Winston, New York).
10. Verin P (1990) *Madagascar* (Karthala, Paris), p 247.
11. Ekblom A, et al. (2016) Migration and interaction between Madagascar and Eastern Africa, 500 BC–1000 AD: An archeological perspective. *Early Exchange Between Africa and the Wider Indian Ocean World*, Palgrave Series in Indian Ocean World Studies (Springer International Publishing, Cham, Switzerland), pp 195–230.
12. Vidal de la Blache P (1902) L'origine des Malgaches par Mr A. Grandidier: Histoire physique, naturelle et politique de Madagascar. Ethnographie. *Ann Geogr* 11:171–173.
13. Ricaut FX, et al. (2009) A new deep branch of eurasian mtDNA macrohaplogroup M reveals additional complexity regarding the settlement of Madagascar. *BMC Genomics* 10:605.
14. Soodyall H, Jenkins T, Stoneking M (1995) 'Polynesian' mtDNA in the Malagasy. *Nat Genet* 10:377–378.
15. Hewitt R, Krause A, Goldman A, Campbell G, Jenkins T (1996) Beta-globin haplotype analysis suggests that a major source of Malagasy ancestry is derived from Bantu-speaking negroids. *Am J Hum Genet* 58:1303–1308.
16. 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.

17. Tofanelli S, et al. (2009) On the origins and admixture of Malagasy: New evidence from high-resolution analyses of paternal and maternal lineages. *Mol Biol Evol* 26: 2109–2124.
18. Pierron D, et al. (2014) Genome-wide evidence of Austronesian-Bantu admixture and cultural reversion in a hunter-gatherer group of Madagascar. *Proc Natl Acad Sci USA* 111:936–941.
19. Cox MP, Nelson MG, Tumonggor MK, Ricaut FX, Sudoyo H (2012) A small cohort of Island Southeast Asian women founded Madagascar. *Proc Biol Sci* 279:2761–2768.
20. Soares P, et al. (2012) The Expansion of mtDNA Haplogroup L3 within and out of Africa. *Mol Biol Evol* 29:915–927.
21. Rito T, et al. (2013) The first modern human dispersals across Africa. *PLoS One* 8: e80031.
22. Chan EK, et al. (2015) Revised timeline and distribution of the earliest diverged human maternal lineages in Southern Africa. *PLoS One* 10:e0121223.
23. Msaidie S, et al. (2011) Genetic diversity on the Comoros Islands shows early seafaring as major determinant of human biocultural evolution in the Western Indian Ocean. *Eur J Hum Genet* 19:89–94.
24. Capredon M, et al. (2013) Tracing Arab-Islamic inheritance in Madagascar: Study of the Y-chromosome and mitochondrial DNA in the Antemoro. *PLoS One* 8:e80932.
25. Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 19:1655–1664.
26. Lawson DJ, Hellenthal G, Myers S, Falush D (2012) Inference of population structure using dense haplotype data. *PLoS Genet* 8:e1002453.
27. Leslie S, et al.; Wellcome Trust Case Control Consortium 2; International Multiple Sclerosis Genetics Consortium (2015) The fine-scale genetic structure of the British population. *Nature* 519:309–314.
28. Beaujard P (2003) Les arrivées austronésiennes à Madagascar: Vagues ou continuum? *Études Océan Indien* 35–36:59–147.
29. Kusuma P, et al. (2016) Contrasting linguistic and genetic origins of the Asian source populations of Malagasy. *Sci Rep* 6:26066.
30. Kusuma P, et al. (2015) Mitochondrial DNA and the Y chromosome suggest the settlement of Madagascar by Indonesian sea nomad populations. *BMC Genomics* 16:191.
31. Radimilahy C (1998) *Mahilaka: An Archaeological Investigation of an Early Town in Northwestern Madagascar* (Dept. of Archaeology and Ancient History, Uppsala University, Uppsala).
32. Radimilahy C (2011) Réflexions sur la production pré-européenne du textile dans le Nord de Madagascar Vohémar, cité-état malgache. *Études Océan Indien* 46–47: 162–176.
33. Dewar RE, Rakotovololona S (1992) La chasse aux subfossiles: Les preuves du XIeme siecle au XIIIeme siecle. *Taloha* 11:4–15.
34. Rasoarifetra B (2012) Les perles de Vohémar, origine et marqueurs culturels. Vohémar, cité-état malgache. *Études Océan Indien* 46–47:177–193.
35. Radimilahy C (1981) Archéologie de l'Androy: Sud de Madagascar. *Rech Pédagog et Cult* 9:62–65.
36. Parker PM (2010) *Pastoralists, Warriors and Colonists: The Archaeology of Southern Madagascar* (Archaeopress, Oxford, UK).
37. Rakotoarisoa JA (1998) *Mille Ans d'Occupation Humaine dans le Sud-Est de Madagascar: Anosy, une Ile au Milieu des Terres* (L'Harmattan, Paris).
38. Deschamps H (1959) *Les Migrations Interieures à Madagascar* (Berger-Levrault, Nancy, France).
39. Decary R (1940) *Carte Ethnographique et Démographique de Madagascar au 1:1 000 000* (Service Géographique de Madagascar, Tananarive, Madagascar).
40. Lombard J (1988) *Le Royaume Sakalava du Menabe - Essai d'Analyse d'un Système Politique à Madagascar du XVIIème au XXème Siècle* (Office de la Recherche Scientifique et Technique Outre-Mer, Paris).
41. Li M, Schröder R, Ni S, Madea B, Stoneking M (2015) Extensive tissue-related and allele-related mtDNA heteroplasmy suggests positive selection for somatic mutations. *Proc Natl Acad Sci USA* 112:2491–2496.
42. Maricic T, Whitten M, Pääbo S (2010) Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS One* 5:e14004.
43. Renaud G, Kircher M, Stenzel U, Kelso J (2013) freeIbis: An efficient basecaller with calibrated quality scores for Illumina sequencers. *Bioinformatics* 29:1208–1209.
44. Andrews RM, et al. (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 23:147.
45. Weissensteiner H, et al.HaploGrep 2: Mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res* 44:W58–W63.
46. van Oven M, Kayser M (2009) Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat* 30:E386–E394.
47. Bandelt HJ, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16:37–48.
48. Soares P, et al. (2009) Correcting for purifying selection: An improved human mitochondrial molecular clock. *Am J Hum Genet* 84:740–759.
49. van Oven M, Van Geystelen A, Kayser M, Decorte R, Larmuseau MH (2014) Seeing the wood for the trees: A minimal reference phylogeny for the human Y chromosome. *Hum Mutat* 35:187–191.
50. Purcell S, et al. (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81:559–575.
51. Li JZ, et al. (2008) Worldwide human relationships inferred from genome-wide patterns of variation. *Science* 319:1100–1104.
52. Auton A, et al.; 1000 Genomes Project Consortium (2015) A global reference for human genetic variation. *Nature* 526:68–74.
53. Ngamphiw C, et al.; HUGO Pan-Asian SNP Consortium (2011) PanSNPdb: The Pan-Asian SNP genotyping database. *PLoS One* 6:e21451.
54. Brucato N, et al. (2016) Malagasy genetic ancestry comes from an historical Malay trading post in Southeast Borneo. *Mol Biol Evol* 33:2396–2400.
55. Paradis E, Claude J, Strimmer K (2004) APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* 20:289–290.
56. Ribeiro P, Diggle P (2001) geoR: A package for geostatistical analysis. *R News* 1:15–18.
57. Chang CC, et al. (2015) Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* 4:7.
58. Browning SR, Browning BL (2007) Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet* 81:1084–1097.
59. Ginestet C (2011) ggplot2: Elegant graphics for data analysis. *J R Stat Soc Ser A Stat Soc* 174:245–246.
60. Browning BL, Browning SR (2013) Improving the accuracy and efficiency of identity-by-descent detection in population data. *Genetics* 194:459–471.
61. Ralph P, Coop G (2013) The geography of recent genetic ancestry across Europe. *PLoS Biol* 11:e1001555.
62. Browning SR, Browning BL (2015) Accurate non-parametric estimation of recent effective population size from segments of identity by descent. *Am J Hum Genet* 97: 404–418.
63. Brisbin A, et al. (2012) PCAdmix: Principal components-based assignment of ancestry along each chromosome in individuals with admixed ancestry from two or more populations. *Hum Biol* 84:343–364.
64. Pickrell JK, Pritchard JK (2012) Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet* 8:e1002967.
65. Hellenthal G, et al. (2014) A genetic atlas of human admixture history. *Science* 343: 747–751.
66. Loh PR, et al. (2013) Inferring admixture histories of human populations using linkage disequilibrium. *Genetics* 193:1233–1254.