



HAL
open science

The loss of biodiversity in Madagascar is contemporaneous with major demographic events

Omar Alva, Anaïs Leroy, Margit Heiske, Veronica Pereda-Loth, Lenka Tisseyre, Anne Boland, Jean-François Deleuze, Jorge Rocha, Carina Schlebusch, Cesar Fortes-Lima, et al.

► To cite this version:

Omar Alva, Anaïs Leroy, Margit Heiske, Veronica Pereda-Loth, Lenka Tisseyre, et al.. The loss of biodiversity in Madagascar is contemporaneous with major demographic events. *Current Biology - CB*, 2022, 32 (23), pp.4997-5007.e5. 10.1016/j.cub.2022.09.060 . halshs-03923549

HAL Id: halshs-03923549

<https://shs.hal.science/halshs-03923549v1>

Submitted on 4 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The loss of biodiversity in Madagascar is contemporaneous with major demographic events

Authors:

Alva Omar¹, Leroy Anaïs¹, Heiske Margit¹, Pereda-Loth Veronica¹, Tisseyre Lenka¹, Boland Anne², Deleuze Jean-François², Rocha Jorge^{3,4,5}, Schlebusch Carina⁶, Fortes-Lima Cesar⁶, Stoneking Mark^{7,8}, Radimilahy Chantal⁹, Rakotoarisoa Jean-Aimé⁹, Letellier Thierry¹, Pierron Denis^{1*}

Affiliations:

1- Équipe de Médecine Evolutive, EVOLSAN faculté de chirurgie dentaire, Université Toulouse III, Toulouse, France.

2-Commissariat à l'Energie Atomique, Institut Génomique, Centre National de Génotypage, 91000 Evry, France.

3-CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Campus de Vairão, Universidade do Porto, 4485-661 Vairão, Portugal

4- Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, 4099-002 Porto, Portugal

5- BIOPOLIS Program in Genomics, Biodiversity and Land Planning, CIBIO, Campus de Vairão, 4485-661 Vairão, Portugal

6- Human Evolution, Department of Organismal Biology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18C, 75236 Uppsala, Sweden

7-Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, D-04103 Leipzig, Germany

8- Université Lyon 1, CNRS, Laboratoire de Biométrie et Biologie Evolutive, UMR 5558, Villeurbanne, France

9- Musée d'Art et d'Archéologie, University of Antananarivo, Antananarivo, Madagascar.

* **Corresponding authors:** Pierron Denis

denis.pierron@univ-tlse3.fr; Tél. +33 (0)5 62 17 29 29; Fax: +33(0)5 61 25 47 1

Twitter handle : @Evol_San

SUMMARY

Only 400 km off the coast of East Africa, the island of Madagascar is one of the last large land masses to have been colonized by humans. While many questions surround the human occupation of Madagascar, recent studies raise the question of human impact on endemic biodiversity and landscape transformation. Previous genetic and linguistic analyses have shown that the Malagasy population has emerged from an admixture that happened during the last millennium, between Bantu-speaking African populations and Austronesian-speaking Asian populations. By studying the sharing of chromosome segments between individuals (IBD determination), local ancestry information and simulated genetic data, we inferred that the Malagasy ancestral Asian population was isolated for more than 1000 years with an effective size of just a few hundred individuals. This isolation ended around 1000 years before present (BP) by admixture with a small African population. Around the admixture time, there was a rapid demographic expansion due to intrinsic population growth of the newly admixed population, which coincides with extensive changes in Madagascar's landscape and the extinction of all endemic large-bodied vertebrates. Therefore, our approach can provide new insights into past human demography and associated impacts on ecosystems.

INTRODUCTION

High place of biodiversity, Madagascar is regarded as one of the last large land masses to have been colonized by humans¹. Although this island is only 400 km off the coast of East Africa, the question of how and when human populations arrived there is still unresolved. And recent studies raise the question of the impact of human colonization on endemic biodiversity. The last two millennia have seen on the island of Madagascar a decline in biodiversity, that ultimately resulted in the extinction of all endemic large-bodied vertebrates, such as the giant lemur, elephant bird, turtles, and hippopotami¹⁻³. These extinctions have been variously attributed to climate change, the arrival of human populations, a cultural transition after human settlement, or a combination of these factors^{4,5}. Resolving the history of Malagasy human populations over the last two millennia is crucial to understand the reasons for the collapse of biodiversity. While the earliest human artefacts or features may date back to 4,000 years before present (BP)⁶, the human presence since more than 1,000 years BP is debated⁷⁻⁹. The settlement of the island has been addressed by multiple disciplines, such as archaeology, anthropology and linguistics¹⁰⁻¹². For several centuries, the settlement process of Madagascar has been the subject of much debate. The first European manuscripts mention populations of African and Arab descent¹³, and very early on, an Asian origin was also suspected¹⁴. At the beginning of the twentieth century, more than 50 authors had already proposed the most diverse origins^{15,16}. During the last century, a consensus has emerged in linguistics suggesting that the Malagasy language evolved from an Austronesian background with Bantu and Arab-Swahili contributions¹⁷⁻²⁰. This consensus around ancestors from Arabo-swahili, Bantu and Austronesian speaking populations expanded from linguistics and is now widely accepted on the basis of multi and interdisciplinary approaches²¹⁻²³. Genetic analyses showed that present-day Malagasy populations are the result of an admixture event at the beginning of the last millennium, between Bantu-speaking African and Austronesian-speaking Asian populations^{24,25}. But there is still no consensus about the past of the ancestral populations before the admixture event. A better understanding of the demographics of early Malagasy human populations may provide important clues regarding their possible impact on the island's ecosystem.

It has recently been shown that the demographic history of populations over the past two millennia can be addressed by studying the sharing of genomic fragments within a population²⁶⁻²⁸. The number and length of shared genetic fragments, called Identity-By-Descent haplotypes (IBD), are dependent on the coalescence probabilities and are therefore markers of fluctuations in effective population size (N_e) over time. Importantly, through local ancestry inference methods, it is possible to reliably assess the characteristics and demographic parameters of each ancestral population prior to an admixture event²⁶.

In the present study, we show that there are many short IBD segments of Asian ancestry that are shared between Malagasy individuals. Based on extensive demographic simulations, we show that the distribution of Asian IBDs supports only scenarios where the ancestral Austronesian-speaking population genetically isolated itself around 2,000 years BP, remaining as a small population ($N_e \approx 500$) for approximately 1,000 years. The smaller number of large IBD fragments that are shared between all Malagasy individuals sampled shows that the population underwent a major demographic expansion around 1,000 years BP, when admixture with the ancestral Bantu-speaking African population is detected. Importantly, this latter time period coincides with the disappearance of the last large-bodied vertebrates in Madagascar.

RESULTS

IBD-sharing in Madagascar populations

To better understand the history of Malagasy ancestral populations, we studied the distribution of shared IBD segments from genomic data of 700 individuals sampled all across the island of Madagascar, as well as from genomic data of 3,464 individuals sampled in other populations. (Methods S1A). Using RefineIBD²⁹, we first performed an inter-population analysis to determine the number of IBD segments (> 2 cM) shared between Malagasy individuals and individuals from other populations. This analysis confirmed that the populations most closely related to the Malagasy are the Bantu-speaking populations of Eastern Africa and the Austronesian-speaking populations of southern Borneo (Methods S1B, Fig. S1). We examined the number of admixture pulses using MALDER³⁰ and found evidence for a single event of admixture, happening around 25 generations BP. As previously shown using different software^{24,30-32}, Malagasy individuals can be grouped into ten clusters based on genetic similarity, and the date of admixture varies (between 20-30 generations BP, Methods S1F) according to the Malagasy genetic cluster. Next, we performed an intra-population analysis to determine the number of IBD segments shared between individuals belonging to the same population. By applying local ancestry inference with RFMix³³, we characterized the continental origin of each IBD segment (Asian, African or heterogeneous). We found that on average, two Malagasy individuals share 11.68 ± 2.19 IBD segments, of which 48.67% are of Asian origin, 40.68% of African origin, and 10.65% of heterogeneous origin (Fig. S2). Considering the overall level of ancestry (60.64% African and 39.36% Asian), the number of IBD segments shared between two Malagasy individuals is proportionally higher for those of Asian origin. Specifically, when we focus on IBDs older than the Asian/African admixture (i.e., those shorter than 5 cM and corresponding to 30 generations BP, according to equation s19 of reference³⁴, see methods), and normalize for the reduced fraction of the ancestry-specific genome considered for each individual's genome (similar to previous studies^{26,35}), two Malagasy individuals share about 31.35 ± 4.47 Asian IBD segments, almost three times more than those of African origin (Fig. 1A, 1B). Interestingly, the high number of Asian IBD segments is present in all Malagasy genetic clusters previously identified²⁴ (Methods S1C).

In contrast, the overall number of IBD segments observed in the most closely related populations is much lower, i.e., populations in southern Borneo or Mozambique, where individuals share 7.43 ± 2.35 and 9.21 ± 1.80 IBD segments, respectively (Fig.1A, 1B). These results were replicated using two independent IBD and local ancestry detection methods (Methods S1D), confirming that the detection of a high number of Asian IBDs smaller than 5 cM is not an artefact of a particular method.

This result suggests that some demographic or selection event strongly impacted the Asian ancestral population after their separation from the Bornean Austronesian-speaking populations, which are the closest match to the Asian ancestral population in our dataset. We further investigated possible causes of this observation by first interrogating if the signal was present across the genome. As we detected the same signal on all 22 chromosomes (Methods S1E), we ruled out the possibility that it was due to a specific selection pressure on particular genes or pathways, suggesting a global demographic event that impacted the whole genome. To characterize this demographic phenomenon, we studied the IBD-sharing distribution based on the age of haplotypes (Fig.1C, 1D). The excess of Asian IBDs is mainly due to old haplotypes, dating from 45 to 75 generations BP (Fig.1C, 1D). The distribution of recent IBDs (less than 10 generations) is similar to other populations. Given that the numbers and lengths of IBD segments contain information about coalescence probabilities and thus effective population sizes^{26,36,37}, this result suggests that the small Asian ancestral population might have drastically increased in size around the time of admixture (30 generations BP).

Malagasy ancestry-specific demographic history

We sought to better characterize the recent demographic history of African and Asian ancestral populations using the ancestry-specific IBDNe pipeline²⁶, which allows us to track demographic trajectories from the last 150 to 2250 years BP. We observed that the admixture event involved two small ancestral populations, followed by a large increase in N_e around the time of admixture (30 generations or 900 years BP) (Fig. 2). Notably, the observed increase in effective size is much greater than what would be produced by the simple addition of effective sizes from both populations, showing an intrinsic demographic expansion in addition to the admixture (Methods S1F). Prior to admixture, the size of the

ancestral African population appears relatively stable (Fig. 2), with only a small reduction ($N_e = 2,740 - 3,170$) around 1,200 years BP; regarding the Asian component (Fig. 2), the estimated Asian-specific effective size presents a drastic decline for several generations before the time of admixture, reaching the lowest N_e at some hundreds of individuals ($N_e = 361 - 444$) around 1,020 years BP. Consistent with the results observed from the Asian IBD-sharing distribution, the pattern of a small population increasing strongly during admixture is common to all Malagasy populations (Methods S1F), and it does not correspond to N_e fluctuations observed in the closest Austronesian-speaking populations from Borneo (Fig. S3). Results from IBDNe inferences seem to be robust, as we analyzed a large sample size (700 individuals) and moreover they were constant through the different genetic clusters described elsewhere²⁴.

As pointed out previously, IBDNe estimation tends to smooth over sudden changes in effective size^{26,38}, thus, the gradual decline observed between 2,250 and 1,000 years BP for the Asian ancestral population might represent an artefact of the method. Therefore, we used a coalescent-based simulator to model alternative scenarios for the demographic history of the Asian ancestral population³⁹, in order to explore their consequences for the expected Asian IBD-sharing distribution. Taking into account the ascertainment bias present in our empirical dataset (see Methods), we compared the impact of multiple parameters on the distribution of IBDs, such as the strength of founder events, the duration of bottleneck events, the amount of gene flow, and the influence of demographic changes during the bottleneck (N_e stable, in expansion or declining) (Methods S1G,H,I). As expected, all of the 784 tested scenarios produced similar results to those observed in Madagascar in terms of allele frequency spectrum, global ancestry, and the distribution of shared African IBD segments (Methods S1N,O,P). In contrast, the studied parameters heavily influenced the Asian IBD-sharing distribution (Methods S1I, S1J). All scenarios presenting the smallest distance metric compared to the observed Asian IBD-sharing distribution were those in which an Asian ancestral population remained small for approximately 1000 years following divergence from other Asian populations, receiving minimal or no gene flow (Fig. S4, Data S1).

Furthermore, we demonstrated that several previously plausible demographic scenarios are in fact incompatible with the observed distribution of Asian IBD segments.

Indeed, the observed distribution of Asian IBD segments is outside five standard deviations of the expected distribution of Asian IBD segments for such scenarios. Thus these scenarios have a probability of less than 10^{-6} of producing the observed curves (Fig. 3, Methods S1H). For example, the high level of shared Asian IBD segments excludes that the population expansion in Madagascar was due to gene flow occurring over several generations (10 generations, between 900 to 600 years BP) from an ancestral Southern Borneo population (Fig. 3: Multiple Asian waves scenario). Scenarios with strong founder events without a bottleneck period before admixture have virtually no probability of yielding the large number of shared Asian IBD segments observed in Madagascar data (Fig. 3: Asian founder event and Strong Asian founder event scenarios). Similarly, this distribution excludes a gradual decrease in the Asian population size before admixture, as suggested by IBDNe results (Fig. 3: Slow decrease of a large population scenario). Interestingly, we show that a scenario presenting a 300 years long bottleneck (duration=10 generations) with a stringent founder event ($N_e=50$), followed by a rapid expansion (reaching $N_e=900$ at the admixture time), could produce a similar level of Asian IBDs, nevertheless IBD segments are larger (hence younger) than observed (Fig. 3: Bottleneck scenario). Finally, the observed Asian IBD-sharing distribution is only consistent with simulated scenarios where a small Asian population ($N_e=500$) has been isolated for at least several centuries (Fig. 3: Long-term bottleneck scenario). Importantly, identical results are obtained with or without considering ascertainment bias (Methods S1K).

Notably, when including in the model a moderate migration rate from Asian populations (5 or 10 individuals per generation), the number of IBD segments decreased drastically, rendering the scenario of a highly connected ancestral population implausible (Methods S1G: Long-term bottleneck scenarios). Furthermore, we show that gene flow from African populations starting before 30 generations BP would produce a date of admixture incompatible with the timing estimated by MALDER (Methods S1G). We also confirmed the strong probability of a bottleneck event for the Asian ancestral population using an ABC approach (Approximate posterior probabilities: Founder effect = 0.033; Multiple Asian Waves= 0.00; Bottleneck=0.967, Methods S1G). ABC parameter estimations also support a long bottleneck scenario with limited migration flow based on Neural Network, Local linear regression and rejection methods (Table S1). For instance,

based on local linear regression, a tolerance rate of 5 %, we estimated a duration of bottleneck 65 generations (95% CI: 46-74) and the migration rate between 0-0,02. Interestingly ABC parameter estimations suggest a N_e at the start of the bottleneck around 930 (95% CI: 109-3560) and N_e at the end: 615 (95% CI: 505-790). This show that the population size of the Asian ancestors of the current Malagasy populations just before admixture is of the same order of magnitude as at the beginning of the bottleneck ($N_{e_startBOT}$ vs N_{e_endBOT} , Table S1). Notably these sizes are also very small compared to earlier (before bottleneck) and later periods (after admixture). This confirms the IBDne estimate of a contemporary population expansion of the admixture.

Therefore, our modelling suggests that the current Malagasy population originated from an Austronesian-speaking population who was genetically isolated between 2000 and 1000 years. This isolation ended for the Malagasy Asian ancestors around 1000 years BP, when they admixed with the Malagasy African ancestors. At about the same time, a rapid, large and intrinsic demographic expansion happened, a date which coincides with extensive changes in the landscape, flora, and fauna (Fig. 4).

DISCUSSION

By the analysis of genetic fragments shared across a large representative sample of Madagascar's inhabitants, we show that today's Malagasy populations originate from an admixture between two small populations, and that the size of this population grew rapidly during admixture. The large number of short Asian IBDs shows that the admixture event ended a long period of bottleneck and isolation for the Asian ancestral population.

This conclusion is based on the use of deconvolution techniques of the genomes of admixed individuals that allow to identify IBDs and thanks to proxy populations, identify the portions of genomes that come from each ancestral population before admixture. While these techniques represent a major advance, several major points must be evaluated. A first point to consider is the assumption that if the proxies used are not equally close to the populations that were mixed, this could result in different power to detect African and

Asian ancestry in the Malagasy genomes. To limit any possibility of bias in proxy selection, we did not use specific populations as proxies but rather a group of East Asian and African populations (see methods). Moreover, we have shown previously that for the Malagasy population, the estimation of local ancestry is very stable whatever the combination of proxies and software used⁴⁰. This stability in the attribution of local ancestry is due to the facts that the Malagasy population comes from a admixture of populations that had almost no gene exchange before or after the admixture (African and Asian) and that the separation between the two populations is old (>60,000 years BP) compared to the admixture (1000 years BP). Therefore, very similar local ancestry results are obtained using multiple pairs: (a) Esan in Nigeria and Beijing Chinese; (b) Esan and Mandar from Sulawesi; (c) Luhya in Webuye, Kenya and Mandar; (d) Yoruba in Ibadan, Nigeria and Mandar; (e) a pool of all African individuals in the 1000 genome dataset as the African source and a pool of all Asian individuals in the 1000 genome dataset as the Asian source⁴⁰. The results are also similar when switching local ancestry software, as a result, it is unlikely that the overrepresentation of small Asian IBDs in our dataset is due to misattribution by local ancestry software.

A second point to consider is the distortion of the relationship between age of IBDs and their size during population fluctuations. We have shown that the high number of short Asian IBD cannot be explained by any other phenomenon than a long period of bottleneck and isolation. This excess of retained IBD, is indeed due to the fact that in a small isolated population, IBDs tend to accumulate and disappear less frequently than in a large population. We have observed that the time/length relationship is almost absent for IBD accumulated more than 30 generations ago (Methods S1Q). While the ABC estimations does not rely on this relationship, this is nevertheless a limitation to fine estimation of parameters such as bottleneck duration or strength (Methods S1I). Thus, it is possible to consider that the bottleneck lasted between 30 and 70 generations and that the effective population size may have been several hundred individuals (Table S1). Further work is needed to clarify the duration of this long bottleneck.

From an anthropological point of view, an important question is whether the start of the bottleneck corresponds to the arrival in Madagascar of the Asian ancestral population around or before 2000 years BP or if alternatively, the migration across the Indian ocean

was a later event during the long-term isolation (two-step scenario), with the Asian ancestral population possibly even arriving just before the admixture. The long-term bottleneck might be more compatible with the settlement and isolation on Madagascar by a small Asian ancestral population, as genetic bottlenecks have been regularly associated with transcontinental migrations^{26,41,42}. But bottlenecks have also been reported in Island South East Asia (ISEA), including in other Austronesian populations^{35,43-46}. Therefore, a period of genetic isolation before the transcontinental migration event is also possible. Notably, the onset of the bottleneck could mark a geographic migration but not necessarily to Madagascar, with the possibility of isolation in an unknown location (Indonesia archipelago, Maldivian archipelago, Eastern African shore). Noteworthy, in the two-step scenario, present and previous results exclude any significant genetic exchange with any other neighboring populations (Asian or African)²⁴.

The journey of the Austronesians to reach Madagascar is much discussed and there is a considerable body of work on the process of colonization of Madagascar although there is no historical record of this phenomenon. Interestingly, an early presence of an Austronesian population in east coast of Africa before or around 2000 BP is increasingly accepted^{47 48 49}. However, many authors doubt that these contacts led to the settlement of Madagascar⁵⁰. The dates most often cited for this settlement vary between 1600 and 1200 BP²³. These dates are supported in particular (but not only) by linguistic works which are based on the presence in the Austronesian language of Sanskrit words⁵¹. Moreover, the date of the 7th-8th century (1200 years BP) is often retained because it would correspond to the emergence of the Srivijaya empire in Southeast Asia. The expansion of this maritime empire could have caused the migration of people to Madagascar⁵¹. But the reasons for the migration of Austronesians to Madagascar remain unknown as well as the route taken through the Indian Ocean⁵². Several authors propose a scenario where the Austronesian population would have arrived in East Africa (or the Comoros) and where a first contact with the Africans would have taken place^{9,47}. From the 8th century AD (1200 years BP), an admixed population speaking Malagasy would have entered Madagascar from the west or northwest. It is interesting to note that we find no evidence of ancient genetic admixture (before the 8th century) between African and Asian populations. This emphasizes once again that genetic admixture and linguistic mixing appears to be disconnected in Madagascar. For

instance, while linguistics show an origin predominantly austronesian⁵³, genetics suggests that Asian ancestry is only about 40%. Another example of such disconnection is the process of the arrival of Austronesians on the island as waves or as a continuum over several centuries which is discussed⁵⁴. Our work emphasizes that if these contacts have existed, they have been mainly cultural and have not left any substantial trace on the genome of the current populations. In particular, our work excludes arrivals from different Austronesian populations because this would have greatly reduced the number of IBD observed. Like admixture phenomena, the history of cultural and genetic connections across the Indian ocean might be only loosely related. Future ancient DNA evidence and a more extensive sampling in ISEA might shed more light on the Asian ancestral population and the timing of their arrival in Madagascar.

Our estimations suggest that the effective population size of the Asian ancestral population was around 500 individuals. It should be noted that estimates of effective population size based on the coalescence probabilities provide a smaller number than the census population size³⁶. Therefore, our work implies that the ancestral Asian population may have reached a few thousand individuals that were genetically isolated over a thousand years, as they should have experienced minimal or no gene flow (Migration rate < 0.02 , Fig. S5). The migration rate in the isolation event is comparable with other previously simulated populations that were separated about 5,000 km (Methods S1G). This isolation, associated with a long-term bottleneck, explains the large genetic distance between the Malagasy population and the closest Austronesian-speaking populations, observed in multiple analyses (IBD sharing, *F_{st}*, treemix, mitochondrial haplotype diversity)²⁴. Interestingly, this long-term bottleneck might solve one inconsistency in the search for the most related populations of Malagasy in East-Asia. Indeed, while Y-chromosome and whole genome analyses point to Borneo populations, mitochondrial DNA diversity seems to point to the islands of eastern Indonesia. In particular, there is a high frequency of the so-called "Polynesian motif" in mitochondrial DNA observed among Malagasy maternal lines⁵⁵, whereas this motif is virtually absent in Borneo populations. Our current work opens the possibility that this difference in frequency for the "Polynesian pattern" observed today between Borneo and Madagascar populations is caused by an ancient split between

these two populations and a genetic drift associated with the long-term bottleneck experienced by Malagasy ancestors.

The genetic isolation of the ancestral Austronesian population terminated with the gene flow from Bantu-speaking populations around 1000 years BP. Based on the absence of evidence of similarly admixed populations on African mainland or in South-East Asia and the heterogeneity of admixture signals across Malagasy population²⁴, it is likely that this event happened in Madagascar. The distribution of African IBD fragments suggests that at around this time, the African ancestral population underwent a moderate demographic reduction before a rapid increase. This fluctuation is consistent with the estimated age of separation between the African population ancestral to the Malagasy and other Bantu-speaking populations, as it coincides with the Bantu-speaking populations expansion across Africa^{24,56–58}. It should be noted that this does not exclude the presence of older African populations in Madagascar that would not have contributed significantly to the current genetic pool, as has been postulated²⁴.

One contribution of the present study is the evidence that both ancestral populations were of limited size a few generations before the start of admixture. In contrast, during the admixture, the Malagasy effective population size increased by a factor of 100 between 1000 to 300 years BP. In parallel to this human expansion, from 500 to 1300 years BP^{3,5} mega faunal extinctions associated with landscape transformations happened on Madagascar (Fig. 4). While various explanations have been proposed to explain the extinctions, recent studies linked these changes to an increase of the human population on the island. But the origin of this increase of human presence is unknown. The present analysis shows that this expansion is not due to the massive arrival of external populations, but instead shows a demographic increase of the newly admixed populations. The high number of 5 centimorgan IBD (African and Asian) suggest that the genetic contribution of any new population arriving on Madagascar island after 1000 years BP is at most limited. In other words, all the current Malagasy descend from the admixture of a small number of individuals of Asian and African origin present on Madagascar around 1000 years BP.

One may therefore ask why the admixture of these populations is linked to such demographic growth around 1000 years BP. The question is particularly acute considering that the Asian ancestral population's effective size was limited to a few hundred individuals

for ~1000 years. A first possibility could be that genetic admixture helped the population to overcome new environmental constraints. As previously shown⁴⁰, the African alleles on chromosome 1 were strongly selected for, probably because of the protection against vivax malaria conferred by the Duffy null allele. The selection signal observed is one of the strongest detected in any human population, suggesting that malaria was a heavy burden for the colonizing Asian ancestral population. But beyond genes, the demographic expansion might also be due to the interchanges of cultural practices between populations of Asian and African origins. Genetic admixture started at different times (between 600-1000 years BP) in different regions of Madagascar. In this period of time, diverse regions presented a transition from forest to grassland^{59,60}, and also evidence of a population subsistence strategy shift³, coupled with the introduction of domesticated animals, such as cattle⁴, dogs⁶¹ and rats⁶², as well as cultivated plants, like rice⁹. Importantly, iron smelting practices in the north-eastern part shows human occupation of some importance since 800 years BP^{63,64}. In consequence, the end of the ancestral Austronesian genetic isolation happened in a period where Malagasy societies seem to have undergone a demographic expansion, marked by the development of agro-pastoralism.

Finally, these changes probably deeply impacted the human demography but also the island. Major landscape transformations were induced due to the usage of fire by humans; it have been proposed that the introduction and expansion of cattle indirectly facilitated the hunting of endemic animals in a context of challenging climatic conditions⁵; also that the introduced herbivores competed with the endemic fauna and ultimately caused a disruption of natural vegetation^{2,3,65}. It is now clear that environmental, cultural and genetic evidence should be considered together to understand this crucial shifting period around 1,000 years BP in Madagascar.

In conclusion, we show that genetic modelling is a powerful tool for testing the plausibility of a historical scenario. Specifically, the demographic model based on IBD distribution and local ancestry can accurately detect periods of isolation or population expansion due to external migrations. The comparison of these models with environmental and cultural data seems to be particularly insightful. A limitation of this strategy is that the chronology obtained from genetic analyses is different from that obtained in other disciplines. Genetic analysis depends on estimates of recombination and mutation rates and

generation times, which have not remained constant through time and may differ between populations⁶⁶. For instance, early extinctions appear to precede genetic mixing and population expansion (Fig. 4). This slight lag may be due to imprecise dating estimates, but it may be also an important signal to explore.

Another limitation of this strategy is that Ne estimation does not inform about the presence of individuals and populations that could have potentially been there also, but not contributed to the current studied population. Keeping these limitations in mind, we propose that for the study of the recent past of populations, the strategy used here, based on the IBD-sharing distribution and simulated genetic data, should be useful for other investigations of human demography and their associated ecosystem impact.

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. This research was funded by Région Aquitaine "Projet MAGE" (Madagascar Génétique et Ethnolinguistique); French National Research Agency (ANR) Grants "MADEOGEN".

AUTHOR CONTRIBUTIONS

Conceptualization, R.C., L.T., A.O. and P.D.; Methodology, A.O. and P.D.; Software, A.O.; Formal analyses, A.O.; Investigation L.A., H.M., P.L.V., T.L., A.O and P.D.; Writing - Original draft, A.O and P.D.; Writing - Review and Editing, L.T., R.K.A., A.O, P.D. and S.M.; Resources, B.A. D.J.F., R.J., S.C., F.L.C, R.C., R.J.A., L.T. and P.D.; Supervision, S.M., R.C., R.J.A., L.T. and P.D.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

DECLARATION OF INTERESTS

The authors declare no competing interests.

MAIN FIGURE TITLES AND LEGENDS

Figure 1. Pairwise IBD sharing in Madagascar, according to inferred local ancestry. (a and b) Boxplot of the distribution of shared IBD segments in Madagascar and their closest genetic population in Africa (A) and Asia (B). The number of segments shared per effective pair (y axis) is shown according to their genetic length (before admixture: $< 5\text{cM}$; after admixture: $\geq 5\text{cM}$) and their ancestral component. (C and D) IBD-sharing distribution according to continental origin and length in centiMorgans (cM), along with the distribution detected in the Malagasy's closest genetic populations inhabiting Africa (C) and Asia (D). The x-axis represents haplotype length cM, and the age of IBD haplotypes (in generations before present) are labelled on the superior horizontal axis. The number of segments shared per effective pair is shown on the vertical axis (y axis). The yellow dashed line shows the time since admixture in Madagascar estimated by MALDER³⁰(See also Figure S1, S2 and Methods S1).

Figure 2. Ancestry-specific effective population size (Ne) estimated for the Malagasy. Left panel, ancestry-specific effective size is shown for African ancestry in the Malagasy populations (red line), with the estimated effective size for Mozambique populations (black line). Right panel, ancestry-specific estimated Ne is shown for the Asian ancestry in the Malagasy population (blue line), as well as the estimated effective size for South Borneo (black line). The y-axes show ancestry-specific population size (Ne), plotted on a log scale. The solid lines show estimated effective sizes, and the colored regions show 95% bootstrap confidence intervals. The x-axes show years before present, while generations before present are labelled on the superior horizontal axis. (See also Figure S3 and Methods S1).

Figure 3. Asian ancestral demographic history simulated scenarios. Distribution of IBD segments of Asian origin shared per generation in Madagascar and in simulated data. Each panel shows the result for simulated scenarios (see Supplementary Methods S1H), according to the strength of founder events (red line) and the duration of bottleneck events (shaded area). Admixture was simulated between Austronesian-speaking and Bantu-speaking populations (see Methods S1H). The x-axis represents the age of IBD haplotypes (in years before present). The number of segments shared per pair is shown on the y-axis. Yellow curves show the mean pairwise IBD-sharing \pm 5 standard deviations at each generation, based on 3 simulated replicas. Blue lines in each panel represent the observed pairwise IBD-sharing of Asian segments from the Malagasy population. (See also Figure S4, Table S1, Data S1 and Methods S1).

Figure 4. Major demographic events for the Malagasy and extensive changes in Madagascar biodiversity during the last 5000 years. Settlement evidence and megafaunal extinction windows are illustrated according to the following numbers (y-axis): (1) Isolation of Asian ancestors (this study), with an effective size of just a few hundred individuals; (2) Split from African populations (southern African Bantus) around 1,500 years BP²⁴; (3) Admixture event happening heterogeneously across the island, between African and Asian populations²⁴; (4) Demographic expansion (this study), due to intrinsic population growth of the newly admixed population; (5) Megafaunal extinction window^{3,5}, with bone icons representing the evidence of anthropogenic modifications on large-bodied vertebrates dating from before 1,300 years BP; (6) Transition from forest to grass⁴⁰, as result of a dramatic increase in the use of fire by humans; (7) Presence of domesticated and commensal animals^{4,61,62}, where radiocarbon dates confirm that introduced animals were established in Madagascar between 1,200 and 700 years ago; (8) Presence of cultivated plants (rice)⁹, brought by migrating people on Madagascar; and (9) the start of urbanization, marked by iron smelting practices^{63,64}.

STAR METHODS

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Denis Pierron (denis.pierron@univ-tlse3.fr).

Materials availability

This study did not generate new unique reagents.

Data and code availability

This study did not generate new unique dataset. The Malagasy genetic data described in this article are available from EGA (EGAS00001002549). Codes and scripts for Msprime simulations can be found in Supplementary Methods 1.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Sample collection

The genome wide SNP data from 700 individuals analyzed here were produced by the project MAGE²⁴. Samples were collected from 2007–2014 utilizing an extensive grid-based approach, in which individuals were sampled from 257 villages (2.8 ± 0.7 individuals per

village) from all over Madagascar. All individuals were unrelated and all of them received detailed information on the study and gave written consent prior to the donation of their sample. This study was approved by the Human Subjects' Ethics Committee of the Health Ministry of Madagascar as well as by French Ethics Committees (Committees of Protection of Persons, and French National Commission on Informatics and Liberty).

METHOD DETAILS

Genotyping

Genome-wide SNP data (>2.3 million markers) was generated using the Illumina Human Omni 2.5-8 (Omni 2.5) BeadChip. Filtering and quality controls were performed using VCFTools⁶⁷, with SNPs excluded that failed the Hardy-Weinberg exact (HWE) test (P-value < 1×10^{-5}) for the 700 Malagasy individuals or had missing data rates >0.10 across all samples.

QUANTIFICATION AND STATISTICAL ANALYSES

Integrating Malagasy data with SNP array datasets

We merged the Malagasy genome wide SNP data with the following datasets:

- 1000 Genomes Project⁶⁸.
- Asian individuals previously genotyped by different studies^{24,40,44,69-72}.
- African individuals from different studies^{58,73-80}.

From the 1000 Genomes Project dataset, we selected individuals of African (ESN, YRI, GWD and LWK) and Asian (CHB, CDX, CHS and KHV) ancestries. Before merging population genetic datasets, we performed the following steps in order to reduce possible batch effects produced by different genotyping platforms. First, we used bcftools for checking the correspondence between the reference allele (from human genome assembly GRCh37) and the presentation of alleles in genotype calls. Then, we performed a quality control procedure on the original datasets, removing all SNPs with missing call rates and

sites failing Hardy-Weinberg equilibrium (HWE) analyses (p -value $< 1e-5$). Following this step, we performed a missing genotype filter of 90% for the merged dataset. The datasets were then merged with high quality genotyping data. Finally, we applied a relatedness filter using KING⁸¹ for removal of possible duplicated samples between original datasets. We removed 15 individuals presenting kinship coefficients above 0.354 (putative twin/duplicated samples).

We generated two datasets (Methods S1A) considering the density of SNPs and samples sizes:

- Density Dataset (311,411 SNPs and 3168 individuals).
- Diversity Dataset (78,906 SNPs and 4212 individuals).

Haplotype Inference - Phasing process

For both datasets, imputation of missing genotypes and estimation of phased haplotypes were obtained with BEAGLE v5.1⁸², using the HapMap GRCh37 genetic map.

Relatedness analysis

We identified the maximal set of unrelated individuals using kinship analyses implemented in KING⁸¹ for removal of possible related individuals. We considered that two individuals were related (third degree or closer) if the kinship coefficient was greater than 0.0442. In order to apply the IBDNe Ancestry-specific method, we followed the same steps as in the original publication. More precisely, we discarded all IBD segments belonging to 114 and 268 couples from the Density and Diversity datasets, respectively. From these couples, 27 and 26 belonged to Malagasy in the Density and Diversity data set, respectively. In total, we removed 4,234 IBD from the Density dataset and 13,330 IBD from the Diversity dataset where respectively 25.88% and 6.82% were from Malagasy couples. The resulting IBD segments were used for IBD-sharing analyses (using both datasets), IBDNe and IBDNe-AS estimation (using only the Density dataset).

Admixture Dates estimation

We inferred the time of admixture in the Malagasy population using MALDER³⁰ with default parameters. The African reference panel is composed of 419 individuals (ESN, YRI, GWD and LWK) and the east-Asian reference panel of 400 individuals (CHB, CDX, CHS and KHV). These 819 samples from the 1000 Genomes Consortium Data were used as a proxy for the Malagasy source populations in all following analyses.

Local Ancestry Inference

Local ancestry analyses were performed on admixed individuals using the software package RFMix³³ and LOTER⁸³. RFMix results were produced using one expectation-maximization (EM) step and a window size of 0.2 cM. For both methods, the time since admixture was set at 30 generations before present. We used as a reference panel the African (ESN, YRI, GWD and LWK) and Asian (CHB, CDX, CHS and KHV) population from the 1000 Genomes Project.

Identical-by-Descent segment determination

We used RefinedIBD²⁹ for Identity-By-Descent (IBD) detection. We filled a gap between two detected IBD segments if the length of the gap was less than 0.6 cM, allowing a number of discordant sites = 1²⁶. For confirmation purposes, we also inferred IBD haplotypes using Germline⁸⁴ with the following parameters: bits = 32, err_hom = 1, err_het = 1 as in Browning, et al²⁹. Only IBD segments greater than 2 cM were used in the experiments and IBD segments from related individuals were removed from the analyses. The R programming language⁸⁵ was used for performing the following analyses on detected IBD segments:

1. We approximated the age of IBD segments using equation s19 from³⁴, under the assumption the population is sufficiently large:

$$E = 75 \times (2/L), \text{ where}$$

E: is the time in generations to the most recent ancestor

L: is the length of the IBD segment in centiMorgans.

2. We estimated the ancestry of each IBD segment by mapping its coordinates to local ancestry assignments from the sample haplotypes. As previous studies^{24,40,69,71,80} identified two main ancestries contributing to the admixture (Bantu from southeast

Africa and Austronesian from Indonesia), ancestry proportions were calculated using the number of genotyped locus (SNP) assigned as Asian (ASI) or African (AFR) divided by the total number of genotyped locus spanning the IBD region in both haplotypes. IBD segments were classified as African (AFR locus $\geq 90\%$), Heterogeneous ($10\% < \text{ASI locus} < 90\%$) or Asian (ASI locus $\geq 90\%$).

3. After IBD ancestry assignment, we filtered out admixed segments showing discordance between the origins of haplotypes. We kept segments with $>50\%$ concordance, based on:

$$\text{Concordance} = 1 - [|(n - m)| / s], \text{ where}$$

n: number of sites of ANC origin on haplotype 1

m: number of sites of ANC origin on haplotype 2

s: number of sites covering the IBD segment.

4. We compared the level of IBD sharing within populations using the mean number of segments shared between individuals. To calculate this mean for each population of size N, the denominator was the number of pairs in that group, specified by the equation $(N^2 - N)/2$.
5. We compared the level of ancestry-specific IBD sharing within populations using the mean number of segments of a particular ancestral origin shared between individuals. We normalized this mean using as the denominator the ancestry-adjusted number of effective pairs, as described previously^{26,35}. If there are N individuals, then the ancestry adjusted number of pairs of haplotypes is defined by

$$\sum_{i=1}^{N-1} \sum_{j=i+1}^N p_i p_j, \text{ where}$$

p_i = the global ancestry of individual i

p_j = the global ancestry of individual j

Global and Ancestry-specific effective size estimation

To estimate whether the Malagasy ancestral populations came from small or large populations, we ran IBDNe Ancestry-Specific (AS) for the population following the pipeline described in Browning *et al*²⁶. This method allows inferring ancestry-specific recent effective population size (Ne) by integrating inferred IBD segments and local

ancestry calls. Based on this information, IBD segments can be classified based on ancestry and then used to infer the ancestry-specific N_e history, based on the relation between the number and length of IBD segments and the probability of coalescence. Following Browning *et al*²⁶, we implemented the IBDNe-AS method using IBD segments greater than 2cM.

To better place the demographic history of Madagascar in a global context, we calculated the overall recent demographic history with the software IBDNe³⁶ for populations across the Indian Ocean, using only the inferred IBD segments greater than 4cM. IBDNe authors recommend a cutoff between 3-6 cM for SNP array data, and 4 cM has been used for other populations in order to avoid the underestimation of effective population size²⁷. As the density of SNPs is an important factor for these inferences, we only applied these methods to the Density dataset. All effective size analyses were limited to populations with sample sizes greater than 39 individuals and we inferred N_e only between 4 and 100 generations before the present, as in Borda *et al*³⁸.

Demographic model

We used Msprime³⁹ to simulate three continental populations (Africa, Europe and Asia), two Bantu populations (Bantu, Mozambique), two Austronesian populations (Austronesian and South Borneo) and an admixed population (Madagascar). Our pre-admixture model for the three continental populations is based on a published model inferred from the 1000 Genomes project data⁴¹ and also described in Browning *et al*²⁶. This model simulates the origin of modern humans, the Out of Africa (OoA) event, the Asian/European split from the OoA population, and finally, the split between Asian and European populations.

The dispersal of Bantu lineages is the result of a population expansion that started in the Nigeria-Cameroon borderlands around 4,000 to 5,000 YBP⁵⁸. Regarding the pre-admixture model for the Bantu expansion, we simulated an initial Bantu population (BAN) that diverged from the African continental population at 167 generations before present, starting with 2,200 individuals⁸⁶. There are diverse hypotheses about the migration route(s) of the Bantu expansion; nevertheless, our aim in this study is to evaluate the demographic dynamics of the African ancestral population of Madagascar. Then, we split a second Bantu population (MOZ) representing the Mozambique population; which was sampled and is

accessible through the publication⁵⁸. Using the method euroIBD⁸⁷, this date was estimated to be 60 generations BP (~ 1,800 years BP) (Methods S1L); IBDNe analyses showed an initial population size at this time of 3,000 individuals. Migration rates between African and Bantu populations were set to 0.0037 as in⁸⁶. All African simulated populations were programmed to reach an effective size of 14,000 individuals at present.

For the Austronesian expansion event, evidence from linguistics and archaeology indicates that the expansion began around 4,000 to 5,000 years BP^{88,89} with an initial effective size of 5,000 individuals⁸⁹. Thus, we simulated an Austronesian population (AUS) that diverged from Mainland Southeast Asia (MSEA) based on these parameters: divergence from MSEA populations at 166 generations BP (~ 5,000 years BP, assuming a generation time of 30 years), with an initial Ne of 5,000 individuals. It should be noted that recent evidence suggest that Austronesians arrived in Taiwan in the north ~6,000 years BP^{90,91}, nevertheless, this date would not affect our more recent inferences. Then, we approximated a South Borneo population (SBO) by simulating a population that diverged from the Austronesian population (AUS) 100 generations before present, as the euroIBD results over our Density dataset show a split between South Borneo and Philippines at this time (Methods S1L). South Borneo population correspond to an admixture of the Austronesian population and Mainland Southeast Asia. Based on IBDNe analysis, we set the Ne initial for South Borneo at 10,000 individuals. Migration rates between Asian and Austronesian populations were set to 0.0016⁹². All Asian simulated populations must reach an effective size of 45,000 individuals at present.

Finally, multiple scenarios were created and evaluated to simulate the history of the Malagasy population. We created this admixed population under two different admixture models – single pulse of admixture vs. multiple pulses of admixture (representing admixture occurring over several generations)⁹³. Regarding the single pulse model, we simulated an admixture event occurring at 20, 25 and 30 generations before present, between populations from MOZ and SBO at a proportion of 62.5% and 37.5%, respectively. The initial Ne was set at 2500 individuals. We also simulated multiples pulses of admixture happening between the periods 20-25, 20-30 and 20-40 generations. Under this scenario, the admixed population begins with 900 individuals of AUS origin who then receive MOZ gene flow at a constant rate during 5, 10 or 20 generations, in order to reach the global

ancestry proportions observed in Madagascar of $\sim 3/8$ Asian ancestry and $\sim 5/8$ African ancestry. Based on IBDne results, we simulated for the admixed populations a current N_e of 150,000 individuals. We consider this N_e realistic taking into account the whole Malagasy population over the past 100 years; in contrast to long-term N_e sizes that are much smaller, this N_e is in the range previously estimated for human populations in the Holocene period⁹⁴.

Simulated genetic data

We simulated sequence data with a mutation rate of 1.25×10^{-8} per base pair per meiosis, using the HapMap GRCh37 genetic map for each autosomal chromosome. To replicate the real data available, we sampled 700 individuals from the admixed population, 161 from the Mozambique population, 91 from the South Borneo population, 100 from the Bantu population, 100 from the Austronesian population, 400 from East-Asia, 400 from Europe and 419 from Africa.

Given that Msprime reports the ancestral recombination graph for each individual's haplotypes, the output genetic data represent the ancestral and derived states with certainty of the haplotypes involved. Thus, as the output VCF report phased data, we did not directly perform a second phasing step with Beagle.

Simulation of empirical ascertainment bias in simulated genetic data

As Msprime produce a VCF file covering all the positions declared in the HapMap genetic map, we extracted all sites that lie in the genomic regions covered by our Density dataset. Particularly, we defined the genomic region covered by the Density dataset as the coordinates produced by the first SNP in the dataset and the last one for each chromosome. In this manner, we subset all the simulated positions declared in the HapMap genetic map for reducing the computation time of following steps. It has been observed that ascertainment bias can influence analyses based on individual SNPs, such as PCA and F_{st} analyses⁹⁵, and thus can skew comparisons between empirical and simulated data. However, we have not found studies reporting the possible effect of ascertainment bias on IBD analyses. In principle, IBD analyses are based on genomic segments whose

identification should not be skewed by SNP ascertainment bias. Nevertheless, we reproduced the observed ascertainment bias in our empirical datasets and confirmed that SNP ascertainment bias does not influence the results (Methods S1K). More specifically, we applied diverse filters based on minor allele frequency (MAF) in order to match the allele frequency spectrum between the empirical and simulated data (as done previously⁹⁵). First, we applied different MAF filters (0.0, 0.01, 0.03 and 0.05) based on control or reference populations (African, European and Asian continental populations, from the 1000 Genomes panel) that were external to the history of Malagasy populations. Next, we obtained the union or the intersection between the set of filtered SNPs according to each reference population and MAF filter. By analysing 108 possible MAF filter combinations, we found that applying a $MAF \geq 0.03$ and $MAF \geq 0.01$ in Africa and Europe simulated populations, respectively, could reproduce the allele frequency spectrum observed in the African and Asian populations from 1000 Genomes in our empirical dataset (Methods S1M). Finally, we randomly selected a certain number of SNPs in a 10kb window, in order to match the SNP density of our Density Dataset. We note that even when ascertainment bias is neglected, we still found the same results for IBD detection using the scenarios from Fig. 3 (Methods S1K). This result is in accordance with a similar demonstration performed by Gopalan et al.⁹⁶.

Demographic conditions for the Malagasy's Asian ancestral population

Given that we found an excess of IBD of Asian origin dated before the admixture, we hypothesize that this overrepresentation is possibly the result of the high coalescence rate between haplotypes due to founder or bottleneck events happening before the admixture. First, we compared the impact of multiple parameters on the distribution of IBDs, such as the strength of founder events (N_e at the beginning of bottleneck: $N_{e_startBOT}$), the duration of bottleneck events (T_{dur_BOT}), the amount of gene flow (Migration rate during the bottleneck: Mig_BOT), and the influence of demographic changes during the bottleneck i.e stable, in expansion or declining (N_e at the beginning of admixture : N_{e_endBOT}).

Due to similar results obtained using the whole genome or a subset of chromosomes (Methods S1E), we explored the Asian IBD distribution using only chromosomes 22 and 20, allowing us to run each of the 784 different scenarios three times, in order to compute a

standard deviation and probability distribution. More specifically, we interrogated this signal by simulating a scenario of single-pulse admixture (happening at 25 generation BP) and other scenarios of multiple-waves of admixture (happening during 10 generations, between 20 to 30 generations BP). For both scenarios, Ne trajectories for the Asian ancestral population of Malagasy were simulated, varying the initial effective population size ($Ne_{startBOT} = 50, 100, 300, 500, 1000, 5000$ or 10000 individuals), the time of isolation ($Tdur_{BOT} = 1, 5, 10, 20, 30, 40, 50, 70$ generations before the admixture), the influence of gene flow ($Mig_{BOT} = 1, 5$ or 10 migrants per generation) and the possibility of demographic expansion/decline (reaching an $Ne_{endBOT} = 900$ individuals).

The maximal time of isolation and bottleneck tested is 70 generations (4,000 y. BP). This limit has been selected because it corresponds to the arrival of Austronesians in the Southeast Asian island and the admixture between AUS and MSEA⁹⁷. Present Malagasy and present Borneo populations share this common admixture signal but not the bottleneck signal. This allows the maximal time of isolation/bottleneck as the split time between AUS and SBO (common ancestral population of simulated Malagasy and Borneo populations). Also, in all scenarios, bottleneck and isolation periods stop simultaneously at the beginning of the admixture with MOZ. By definition the isolation stops at the admixture. IBDne and IDBNe-AS showed that the admixture happened between two small populations. Since our method would not detect a limited expansion a few generations before the admixture, we use the start of admixture as the final point for bottleneck which allows us to keep constant the african demographic parameters (estimated through IBDne-AS).

We evaluated different demographic scenarios by calculating an indicator based on the difference between the number of observed and simulated IBD segments. Specifically, for each of the last 75 generations, we calculated the absolute difference between the number of observed and simulated IBD, and then computed the mean to obtain the final indicator.

Based on this exploratory analysis, we simulated all 22 chromosomes to implement IBD-sharing analyses over a set of specific demographic scenarios, which were chosen based on anthropological hypotheses (Methods S1H). The objective of this step is to assess the likelihood of each scenario according to the Asian IBD-sharing distribution. More precisely, we considered the following scenarios for the Asian population ancestral to the Malagasy:

- Split from Southern Borneo at 25 generations BP through a founder event followed immediately by the admixture ($Ne_{endBOT}=900$, $Ne_{startBOT}=900$, $Tdur_{BOT} =0$, $Mig_{bot}=0$).
- Split from Southern Borneo at 26 generation BP through a strong founder event followed by admixture one generation after ($Ne_{endBOT}=900$, $Ne_{startBOT} =50$, $Tdur_{BOT} =1$, $Mig_{BOT} =0$).
- Admixture through multiple pulses of gene flow ($Ne=70$ to 300) coming from an ancestral Southern Borneo population, between 20-30 generations before present.
- Split from Southern Borneo through a founder event of $Ne_{startBOT} =5,000$ individuals, followed by a slow decrease of Ne during 50 generations ($Tdur_{BOT} =50$, $Mig_{BOT} =0$) until reaching $Ne_{endBOT}=900$ (before the admixture at 25 generations BP).
- Split from Southern Borneo through a strong founder event of $Ne_{startBOT} =50$, with effective population size expansion during 10 generations ($Tdur_{BOT} =10$, $Mig_{BOT} =0$), reaching $Ne_{endBOT}=900$ (before the admixture at 25 generations BP).
- Split from Southern Borneo through a founder event of $Ne_{startBOT} =500$, with effective population size constant ($Ne_{endBOT}=500$) during 40 generations ($Tdur_{BOT} =40$, $Mig_{BOT} =0$), receiving no gene flow before the admixture. It is important to notice that this was the best candidate based on the exploratory simulation analysis (Methods S1H, S1I and Data S1).

Analysis of demographic simulations

We analyzed all demographic scenarios with an automated bioinformatics pipeline, implementing the same methods and parameters used for the observed data. More precisely, we performed local ancestry inference using the software package RFMix. The time since admixture was set to 30 generations before present and a window size of 0.2 cM. We used as a reference panel the AFR and MSEA populations from the simulated data. Regarding the IBD segment determination, we used RefinedIBD and filled the gaps between segments with the merge-ibd script, specifying a gap's length of 0.6 cM and a number of discordant sites =1. In order to assess the relationship between age of IBDs and their size, we also recovered the IBD segments shared between simulated Malagasy individuals from Msprime simulated tree sequences (chromosome 22 and 20, scenario 6, Methods S1Q).

Effect of the coalescent simulators

To control the effect of the coalescent simulators, we simulated the same demographic scenario ($Ne_{startBOT}=500$; $Ne_{endBOT}=500$; $Tdur_BOT=40g$, $Mig_BOT=0.0$) under the Standard Coalescent model, the discrete time Wright-Fisher model and a combination of these two (DFTW during the first 500 generations and the Standard Coalescent until the end of simulation) on chromosome 22 and 20. All coalescence models produce similar results, that is a similar accumulation of short IDB in a long bottleneck scenario (Methods S1R).

Approximate Bayesian Computation (ABC)

We implemented the Approximate Bayesian Computation (ABC) for estimating confidence intervals. We defined three alternative models of genetic and demographic history: demographic model A: a founder effect for the Asian ancestral population, demographic model B : multiple founder effects through multiple Asian waves of gene flow, demographic model C: bottleneck event for the Asian ancestral population (Methods S1S). For each model, 500 simulations were generated, and for each simulation different parameter values were randomly chosen, as shown in Supplementary Table 1. We choose as summary statistics the pairwise IBD-sharing, as it should be informative for distinguishing among competing scenarios that are under consideration, and should be strongly influenced by the parameters of interest. Model choice was performed using a rejection method (package `abc` from R) based on 500 simulations for each studied demographic scenario. First, we performed a cross-validation step based on 100 pseudo-observed simulations with the rejection method, considering a tolerance rate of 5% (Methods S1S). Next, we estimated the posterior probability of each demographic scenario, using the rejection method and retaining 10% of the simulations with the closest summary statistic values. Based on the best supported demographic model, we performed parameter estimation with ABC from 5668 simulations. In particular, we estimated the $Ne_{startBOT}$, Ne_{endBOT} , $Tdur_BOT$ and Mig_BOT parameters, using the neural network, local linear

regression and rejection algorithms, implemented in the abc package from R, under a tolerance rate of 0.05.

SUPPLEMENTAL INFORMATION TITLES AND LEGENDS

SUPPLEMENTARY FIGURES

Figure S1. Boxplot of the distribution of the average number of IBD segments shared, in the density dataset, between a Malagasy individual and an individual from Africa in red (*upper panel*) or from Asia in blue (*lower panel*). Related to Figure 1. For each population IBDs are divided into 2 length bin $<5\text{cm}$ (*left*) and $>5\text{cm}$ (*right*). Comparisons were made using the density dataset, comprising 311,411 SNPs and 3168 individuals. In this study, we classified the Khoe-San populations based on the designations previously proposed, when referring collectively to the traditional hunter-gatherers of southern Africa as San; and when referring to traditional pastoralist groups as Khoe. Regarding the linguistic component of these populations, we notice that the Khwe population speak a click language from the Khoe subfamily.

Figure S2. Distribution of the average number of shared IBD segments between two Malagasy individuals, according to the inferred continental ancestry of each IBD segment. A) boxplot of the global number of shared IBD B) Distribution of shared IBD according their size. Related to Figure 1. We estimated the ancestry of each IBD segment by mapping its coordinates to local ancestry assignments from sample haplotypes. The ancestry proportion was calculated using the number of Asian (ASI) sites divided by the total number of SNPs spanning the IBD region in both haplotypes. IBD segments were classified as African (AFR SNPs $\geq 90\%$), Heterogenous ($10\% < \text{ASI SNPs} < 90\%$) or Asian (ASI SNPs $\geq 90\%$).

Figure S3. Estimated effective population size (N_e) over the last 100 generations for the African populations (A) and Asian populations (B) analysed in this study. Related to Figure 2. The line in each plot shows the estimated effective population size, and the coloured regions are bootstrap 95% confidence intervals. The y-axes (effective population size) are plotted on a logarithmic scale.

Figure S4. Effect of gene flow during the long-term bottleneck on the IBD segments of Asian origin shared in Madagascar. Related to Figure 3. Each panel shows the result for a Long-term bottleneck scenario according to different simulated migration rates. The x-axis represents the age of IBD haplotypes (in years before present). The number of segments shared per pair is shown on the y-axis. Yellow curves show the mean pairwise IBD-sharing at each generation ± 5 standard deviations, based on 3 simulated replicas. Blue lines in each panel represent the observed pairwise IBD-sharing of Asian segments in the Malagasy population.

SUPPLEMENTARY TABLES

Table S1. Estimated demographic parameters under the Bottleneck event model based on ABC method (tolerance rate of 5 %). Related to Figure 3.

SUPPLEMENTARY DATA

Data S1. Average reported scores from Asian IBD-sharing absolute differences between observed and simulated data. Related to STAR Methods and Figure 3.

SUPPLEMENTARY METHODS

METHODS S1: Model construction and impact of potential confounding variables and (Related to STAR Methods and Figures 1, 2 and 3).

(A) Sampling locations included in the study. [A.1] Map showing the locations of samples used in the Density dataset. [A.2] Map showing the locations of samples used in the Diversity dataset. Colors indicate regional affiliation of populations. [A.3] Detailed sample information.

(B) Boxplot of the distribution of the average number of IBD segments shared, in the diversity dataset between a Malagasy individual and an individual from Africa in red (upper panel) or from Asia in blue (lower panel). For each population IBDs are divided into 2 length bin $<5\text{cm}$ (left) and $>5\text{cm}$ (right). Comparisons were made using the diversity dataset, comprising 78,906 SNPs and 4,164 individuals. In this study, we classified the Khoe-San populations based on the designations previously proposed, when referring collectively to the traditional hunter-gatherers of southern Africa as San; and when referring to traditional pastoralist groups as Khoe. Regarding the linguistic component of these populations, we notice that the Khwe population speak a click language from the Khoe subfamily.

(C) Average number of shared IBD segments between two Malagasy individuals according to ancestral origin and genetic clustering across Madagascar. Each panel depicts the location of a genetic grouping of individuals as identified previously¹², and boxplots of the distribution of shared IBD segments per effective pair of individuals, according to their continental ancestry (African or Asian) and genetic length (IBD $< 5\text{cM}$ or IBD $\geq 5\text{cM}$). We normalized the ancestry-specific numbers of shared IBD fragments by taking into account the reduced fraction of the genome represented by the specific ancestry.

(D) Replication analysis of pairwise IBD sharing in Madagascar, according to inferred local ancestry based on LOTER and Germline methods. (1 and 2) Boxplot of the distribution of shared IBD segments in Madagascar and the closest genetic population in Africa and Asia (left and right panel, respectively) according to genetic length ($< 5\text{cM}$ or $\geq 5\text{cM}$). **(3 and 4)** IBD-sharing distribution according to continental origin and length in centiMorgans (cM), along with the distribution detected in the Malagasy's closest genetic populations inhabiting Africa and Asia (right and left panel, respectively). The x-axis represents haplotype length cM, and the age of IBD haplotypes (in generations before present) are labelled on the superior horizontal axis. The number of segments shared per effective pair is shown on the vertical axis (y axis). The yellow dashed line shows the time since admixture in Madagascar estimated by MALDER. Germline and LOTER were used for IBD determination and local ancestry inference, respectively.

(E) Distribution of IBD segments across the genome of Malagasy individuals. 1) global coverage of shared IBD by genomic locus according to continental ancestry (African in red and Asian in blue). Using the R package "Genomic ranges", we estimated for each locus (x-axis) the ancestry-specific coverage from detected IBD segments (y-axis). Upper and lower

black dotted lines represent the 0.99 and 0.01 quantiles of data, respectively. 2) Average number of shared IBD segments between two Malagasy individuals according to ancestral origin and for each chromosome. Each panel represents a chromosome, showing the boxplot of the distribution of shared IBD segments per effective pair of individuals, according to continental ancestry (African or Asian) and genetic length (IBD < 5cM or IBD \geq 5cM). We normalized the ancestry-specific number of shared IBD fragment by taking into account the reduced fraction of the genome represented by each specific ancestry.

(F) Estimated ancestry-specific effective population size (Ne) for Madagascar over the last 100 generations. Related Malagasy samples ($k > 0.0442$) were removed from the data. Using IBDNe, ancestry specific and overall effective population sizes were estimated. The y-axes show Ne values, plotted on a log scale. The x-axes show generations before present. **(1)** Historical effective population size trajectory for Madagascar. **(2)** Historical Ne trajectories for individual genetic groups described previously¹². For both (F.1 and F.2), the lines show estimated effective population sizes. The colored regions represent the 95% bootstrap confidence interval. **(3)** IBDNe maximum and minimum estimated effective population sizes at the inferred date of admixture (in generations before present, according to MALDER). Growth rate was calculated using the Ne during a period of 10 generations after the admixture date

(G) Demographic parameters used for Malagasy demographic history. We used Msprime²⁷ (Kelleher et al., 2016) to simulate three continental populations (AFR, EUR and ASI), two Bantu populations (BAN, MOZ), two Austronesian populations (AUS and SBO) and an admixed population (MAD). Our pre-admixture model for the African (AFR), European (EUR) and Asian (ASI) continental populations is based on a published model inferred from the 1000 Genomes project data²⁸ and also described elsewhere¹⁴. This model simulates the advent of modern humans, the Out of Africa (OoA) event, and the Eurasian bottleneck after the OoA and the split between Asian and European populations. Regarding the Bantu-speaking and Austronesian-speaking expansion, we simulated an initial Bantu population (BAN) that diverged from the African continental population and an Austronesian population (AUS) diverging from the Asian continental population (ASI). We approximated a South Borneo population (SBO) by simulating a population that diverged from AUS at 100 generations BP; we approximated the time of Bantu occupation of eastern Africa by simulating a Mozambique population (MOZ) splitting from BAN at 60 generations BP. Finally, we simulated an admixture event between MOZ and SBO populations at a proportion of 5/8 and 3/8, in order to reach the global ancestry proportions observed in Madagascar: 1/3 Asian ancestry and 2/3 African ancestry. G.1) illustration of the model in a direct admixture case, G.2) illustration of the model in a post bottleneck admixture case. G.3) Migration rates used in previous models of human demographic history. G.4) Inferred admixtures dates on simulated data according to different admixture models.

(H) Graphical summary of simulated demographic scenarios for the Malagasy ancestral Asian population. H.1) Exploratory step with population with constant size H.2) Exploratory step with population in expansion or decline H.3) Compatibility step. Given that the effective size of the Malagasy's ancestral Asian population underwent a severe decline for several generations, we hypothesize that many of the possible coalescences between haplotypes occurred in the pre-admixture period through founder effect or bottleneck events. In the exploratory step, we simulated an admixture event (immediate admixture at 25 generations BP or continuous gene flow between 20-30 generations BP) for

particular scenarios regarding the Asian ancestral population. We varied the strength of founder events ($Ne_{startBOT} = 50, 100, 300, 500, 1000, 5000$ or 10000 individuals) at the different split dates from South Borneo, the duration of the bottleneck ($T_{dur_bot} = 1, 5, 10, 20, 30, 40, 50,$ or 70 generations before the admixture), and the influence of gene flow ($1, 5$ or 10 migrants per generation). We allowed for the possibility of changes in demography during the bottleneck by adding a constant Ne ($Ne_{endBOT} = Ne_{startBOT}$) or an expansion/decline in Ne ($Ne_{endBOT} = 900$). In the compatibility step, we simulated 6 scenarios 1: Funder event ($Ne_{endBOT}=900, Ne_{startBOT}=900, T_{dur_bot}=0, Mig_bot=0$), 2: Strong funder event ($Ne_{endBOT}=900, Ne_{startBOT} =50, T_{dur_bot}=1, Mig_bot =0$), 3: Multiple Asian waves, 4: Slow decrease of a large population ($Ne_{startBOT} =5,000, T_{dur_bot}=50, Mig_bot =0, Ne_{endBOT}=900$), 4: Bottleneck ($T_{dur_bot}=10, Mig_bot =0$), reaching $Ne_{endBOT}=900$), 5: long-term bottleneck ($Ne_{startBOT} =500, Ne_{endBOT}=500, T_{dur_bot}=40, Mig_bot=0$). H.4) Description of Asian ancestral simulated demographic scenarios and Malagasy settlement H.5) Migration rates used in previous models of human demographic history

(I). Effect of different demographic scenarios on the simulated IBD segments of Asian origin shared in Madagascar. Each panel shows the result for simulated scenarios, according to the strength of founder events ($Ne = 50, 100, 300, 500, 1000, 5000$ or 10000 individuals) and the duration of the bottleneck ($T = 1, 5, 10, 20, 30, 40, 50,$ or 70 generations before the admixture) for the Malagasy Asian ancestral population. The x-axis represents the age of IBD haplotypes (in generations before present). The number of segments shared per pair is shown on the y axis. Coloured lines show the pairwise IBD-sharing from simulations at each generation, while black lines represent the observed pairwise IBD-sharing. The average score and standard deviation using 3 replicates are shown in the upper-right legend. **(I.1 to I.4).** Before the admixture event, we simulated an isolated Austronesian population with stable Ne during the bottleneck. Migration rates to the isolated population were set to 0 (I.1), 1 (I.2), 5 (I.3) and 10 (I.4) individuals per generation. **(I.5 to I.8)** We simulated an isolated Austronesian population with changing Ne during the bottleneck. Migration rates to the isolated population were set to 0 (I.5), 1 (I.6), 5 (I.7) and 10 (I.8) individuals per generation. **(I.9 and I.10)** We simulated a South Bornean population with initial $Ne=5,000$ and growth rate of 0.0219 . From this population, an Austronesian population diverged with stable Ne during a bottleneck event. Migration rates to the isolated population were set to 0 or 1 individuals per generation (I.9 and I.10, respectively). **(I.11 and I.12)** We simulated a South Bornean population with initial $Ne=5,000$ and growth rate of 0.0219 . From this population, an Austronesian population diverged with changing Ne during a bottleneck event. Migration rates to the isolated population were set to 0 or 1 individuals per generation (I.11 and I.12, respectively). **(I.13 and I.14)** Admixture was simulated under the Continuous Gene Flow model, with the Austronesian ancestral population receiving African gene flow during the period 20-30 generations BP. Before the admixture event, we simulated a bottleneck event for the Austronesian population with stable (I.13) or changing (I.14) Ne during the bottleneck. Migration rates to the isolated population were set to 0 individuals per generation.

(J). Distribution of the absolute mean differences scores between simulated and observed IBD-sharing of Asian origin in Madagascar, according to different simulated demographic scenarios. Related to Figure 3. We evaluated different demographic scenarios using the average absolute difference between the number of observed and

simulated Asian IBD segments during the last 75 generations. **J.1)** Distribution of scores according to different bottleneck durations. **J.2)** Distribution of scores according to different strengths in founder events. **J.3)** Distribution of scores according to different migration rates. The y-axes show the number of scenarios simulated. The x-axes show generations before present.

K. Asian ancestral demographic history simulated scenarios and Malagasy settlement without inclusion of ascertainment bias.. Each panel shows the result for simulated scenarios, according to the strength of founder events (red line) and the duration of bottleneck events (shaded area). The x-axis represents the age of IBD haplotypes (in years before present). The number of segments shared per pair is shown on the y-axis. Yellow curves show 5 standard deviations from the mean pairwise IBD-sharing at each generation across 3 simulated replicas **with ascertainment bias**. Green curves show 5 standard deviations from the mean pairwise IBD-sharing at each generation across 3 simulated replicas **without ascertainment bias**.

L. Estimated average number of most recent genetic common ancestors per generation back through time according to different pairs of populations. Timeline of the point (dotted line) and smooth (solid line) estimations of the average number of common ancestors shared between Malagasy and South Borneo, Malagasy and Mozambique, Mozambique and Angola populations, and South Borneo and Philippines, as estimated by shared IBD segments from genome-wide data.

M. Effect of minor allele frequency (MAF) filters on allele frequency spectrum of simulated reference populations. Coloured bars show the allele frequency distribution from simulations in African (M.1) and Asian (M.2) reference populations, while red lines represent the observed allele frequency distribution in our density dataset. M.3) Score.

N. Effect of different demographic scenarios on the simulated IBD segments of African origin shared in Madagascar. Each panel shows the result for simulated scenarios, according to the strength of founder events ($N_e = 50, 100, 300, 500, 1000, 5000$ or 10000 individuals) and the duration of bottleneck ($T = 1, 5, 10, 20, 30, 40, 50,$ or 70 generations before the admixture) for the Asian ancestral population. The x-axis represents the age of IBD haplotypes (in generations before present). The number of segments shared per pair is shown on the vertical axis (y axis). Coloured lines show the pairwise IBD-sharing from simulations at each generation, while black lines represent the observed pairwise IBD-sharing. The average score and standard deviation using 3 replicates are shown in the upper-right legend. (**N1 to N4**). Before the admixture event, we simulated an isolated Austronesian population with stable N_e during the bottleneck. Migration rates to the isolated population were set to 0 (N1), 1 (N2), 5 (N3) and 10 (N4) individuals per generation. (**N5 to N8**) We simulated an isolated Austronesian population with changing N_e during the bottleneck. Migration rates to the isolated population were set to 0 (N5), 1 (N6), 5 (N7) and 10 (N8) individuals per generation. (**N9 and N10**) We simulated a South Bornean population with initial $N_e=5,000$ and growth rate of 0.0219. From this population, we an Austronesian population diverged with stable N_e during a bottleneck event. Migration rates to the isolated population were set to 0 (N9) or 1 (N10) individuals per generation. (**N11 and N12**) We simulated a South Bornean population with initial $N_e=5,000$ and growth rate of 0.0219. From this population, an Austronesian population diverged with changing N_e during a bottleneck event. Migration rates to the isolated population were set to 0 and 1 individual per generation (N11 and N12, respectively). (**N13**

and N14) Admixture was simulated under the Continuous Gene Flow model, with the Austronesian ancestral population receiving African gene flow during the period 20-30 generations BP. Before the admixture event, we simulated a bottleneck event for the Austronesian population with stable (M) or changing (N) N_e during the bottleneck. Migration rates to the isolated population were set to 0 individuals per generation.

O. Effect of different demographic scenarios on the simulated allele frequency spectrum in Madagascar. Each panel shows the result for simulated scenarios. The details of these correspond to the legend from Fig. N.

P. Effect of different demographic scenarios on the simulated global ancestry profile in Madagascar. Each panel shows the result for simulated scenarios. The details of these correspond to the legend from Fig. N.

Q. Relationship between age of IBDs and their size detected (Q.1) and age estimated using Al-Asadi et al. model (Q.2) during a long bottleneck scenario (scenario 6). The limit of 5 centimorgan is marked by a red line. All IBD over 45 generations appear to be of similar size. The ratio of pre-, during and post-admixture IBDs is plotted in Panel Q.3 according to their estimated age using the formula of Al-Asadi et al. (expressed in generations before present).

R. Effect of different simulation ancestry models on the Asian IBD-sharing distribution. We simulated the same demographic scenario ($N_{e_startBOT}=500$; $N_{e_endBOT}=500$; $T_{dur_BOT}=40g$, $Mig_BOT=0.0$) under the Standard Coalescent model, the discrete time Wright-Fisher model and a combination of these two (DTWF during the first 500 generations and the Standard Coalescent until the end of simulation).

S. ABC model misclassification. S.1) Prior distributions of the parameters for the studied demographic models. S.2) Graphical representation of the three demographic models tested. S.3) Model choice step in the ABC approach. Confusion matrix for three models. The colors from dark to light gray correspond to models Bottleneck, Founder effect, Multiple Asian waves of admixture, respectively.

T. Codes used in this study.

REFERENCES

1. Douglass, K., Hixon, S., Wright, H.T., Godfrey, L.R., Crowley, B.E., Manjakahery, B., Rasolondrainy, T., Crossland, Z., and Radimilahy, C. (2019). A critical review of radiocarbon dates clarifies the human settlement of Madagascar. *Quaternary Science Reviews* 221, 105878.
2. Burney, D. (2004). A chronology for late prehistoric Madagascar. *Journal of Human Evolution* 47, 25–63.
3. Godfrey, L.R., Scroxton, N., Crowley, B.E., Burns, S.J., Sutherland, M.R., Pérez, V.R., Faina, P., McGee, D., and Ranivoharimanana, L. (2019). A new interpretation of Madagascar's megafaunal decline: The "Subsistence Shift Hypothesis." *Journal of Human Evolution* 130, 126–140.
4. Hixon, S.W., Douglass, K.G., Crowley, B.E., Rakotozafy, L.M.A., Clark, G., Anderson, A., Haberle, S., Ranaivoarisoa, J.F., Buckley, M., Fidiarisoa, S., et al. Late Holocene spread of pastoralism coincides with endemic megafaunal extinction on Madagascar. 10.
5. Li, H., Sinha, A., Anquetil André, A., Spötl, C., Vonhof, H.B., Meunier, A., Kathayat, G., Duan, P., Voarintsoa, N.R.G., Ning, Y., et al. (2020). A multimillennial climatic context for the megafaunal extinctions in Madagascar and Mascarene Islands. *Sci. Adv.* 6, eabb2459.
6. Dewar, R.E., Radimilahy, C., Wright, H.T., Jacobs, Z., Kelly, G.O., and Berna, F. (2013). Stone tools and foraging in northern Madagascar challenge Holocene extinction models. *Proceedings of the National Academy of Sciences* 110, 12583–12588.
7. Anderson, A., Clark, G., Haberle, S., Higham, T., Nowak-Kemp, M., Prendergast, A., Radimilahy, C., Rakotozafy, L.M., Ramilisonina, Schwenninger, J.-L., et al. (2018). New evidence of megafaunal bone damage indicates late colonization of Madagascar. *PLoS ONE* 13, e0204368.
8. Mitchell, P. (2020). Settling Madagascar: When Did People First Colonize the World's Largest Island? *The Journal of Island and Coastal Archaeology* 15, 576–595.
9. Crowther, A., Lucas, L., Helm, R., Horton, M., Shipton, C., Wright, H.T., Walshaw, S., Pawlowicz, M., Radimilahy, C., Douka, K., et al. (2016). Ancient crops provide first archaeological signature of the westward Austronesian expansion. *Proc Natl Acad Sci USA* 113, 6635–6640.
10. Beaujard, P. (2011). The first migrants to Madagascar and their introduction of plants: linguistic and ethnological evidence. *Azania: Archaeological Research in Africa* 46, 169–189.

11. Hurles, M.E., Sykes, B.C., Jobling, M.A., and Forster, P. (2005). The Dual Origin of the Malagasy in Island Southeast Asia and East Africa: Evidence from Maternal and Paternal Lineages. *The American Journal of Human Genetics* 76, 894–901.
12. Serva, M. (2012). The Settlement of Madagascar: What Dialects and Languages Can Tell Us. *PLoS ONE* 7, e30666.
13. Grandidier, A., Charles-Roux, J., Delhorbe, C., Froidevaux, H., and Grandidier, G. (1903). *Collection des ouvrages anciens concernant Madagascar: Ouvrages ou extraits d'ouvrages portugais, hollandais, anglais, français, allemands, italiens, espagnols et latins relatifs à Madagascar (1500 à 1613) Comité de Madagascar.*
14. Pyrard, F. (1603). *Discours du voyage des François aux Indes Orientales.* In (David Le Clerc), p. 371.
15. Grandidier, A., and Grandidier, G. *Histoire physique, naturelle et politique de Madagascar* Imprimerie Nationale.
16. Ferrand, G. (1909). L'Origine africaine des Malgaches. *Bul. Mém. Soc. Anthropol. Paris*, 22–35.
17. Dahl, O.C. (1988). Bantu substratum in Malagasy *Etudes ocean indien.* 91–132.
18. Dahl, O.C. (1977). La subdivision de la famille Barito et la place du Malgache. *Acta Orientalia*, 77–134.
19. Adelaar, K.A. (1989). Les langues austronésiennes et la place du malagasy dans leur ensemble. *Arch. Androl.*, 25–52.
20. Simon, P. *La langue des ancêtres.* Ny Fitenin-dRazana L'Harmattan.
21. Blench, R.M. (2015). Interdisciplinary approaches to stratifying the peopling of madagascar. In.
22. Adelaar, A. (2016). Austronesians in Madagascar: A Critical Assessment of the Works of Paul Ottino and Philippe Beaujard. In *Early Exchange between Africa and the Wider Indian Ocean World*, G. Campbell, ed. (Springer International Publishing), pp. 77–112.
23. Beaujard, P. (2011). The first migrants to Madagascar and their introduction of plants: linguistic and ethnological evidence. *Azania: Archaeological Research in Africa* 46, 169–189.
24. Pierron, D., Heiske, M., Razafindrazaka, H., Rakoto, I., Rabetokotany, N., Ravololomanga, B., Rakotozafy, L.M.-A., Rakotomalala, M.M., Razafiarivony, M., Rasoarifetra, B., et al. (2017). Genomic landscape of human diversity across Madagascar. *Proc Natl Acad Sci USA* 114, E6498–E6506.

25. Serva, M., and Pasquini, M. (2020). Dialects of Madagascar. *PLoS ONE* *15*, e0240170.
26. Browning, S.R., Browning, B.L., Daviglus, M.L., Durazo-Arvizu, R.A., Schneiderman, N., Kaplan, R.C., and Laurie, C.C. (2018). Ancestry-specific recent effective population size in the Americas. *PLoS Genet* *14*, e1007385.
27. Castro e Silva, M.A., Nunes, K., Lemes, R.B., Mas-Sandoval, À., Guerra Amorim, C.E., Krieger, J.E., Mill, J.G., Salzano, F.M., Bortolini, M.C., Pereira, A. da C., et al. (2020). Genomic insight into the origins and dispersal of the Brazilian coastal natives. *Proc Natl Acad Sci USA* *117*, 2372–2377.
28. Yunusbaev, U., Ionusbaev, A., Han, G., and Kwon, H.W. (2020). Recent effective population size in Eastern European plain Russians correlates with the key historical events. *Sci Rep* *10*, 9729.
29. Browning, B.L., and Browning, S.R. (2013). Improving the Accuracy and Efficiency of Identity-by-Descent Detection in Population Data. *Genetics* *194*, 459–471.
30. Pickrell, J.K., Patterson, N., Loh, P.-R., Lipson, M., Berger, B., Stoneking, M., Pakendorf, B., and Reich, D. (2014). Ancient west Eurasian ancestry in southern and eastern Africa. *Proceedings of the National Academy of Sciences* *111*, 2632–2637.
31. Lawson, D.J., Hellenthal, G., Myers, S., and Falush, D. (2012). Inference of Population Structure using Dense Haplotype Data. *PLoS Genet* *8*, e1002453.
32. Hellenthal, G., Busby, G.B.J., Band, G., Wilson, J.F., Capelli, C., Falush, D., and Myers, S. (2014). A Genetic Atlas of Human Admixture History. *Science* *343*, 747–751.
33. Maples, B.K., Gravel, S., Kenny, E.E., and Bustamante, C.D. (2013). RFMix: A Discriminative Modeling Approach for Rapid and Robust Local-Ancestry Inference. *The American Journal of Human Genetics* *93*, 278–288.
34. Al-Asadi, H., Petkova, D., Stephens, M., and Novembre, J. (2019). Estimating recent migration and population-size surfaces. *PLoS Genet* *15*, e1007908.
35. Ioannidis, A.G., Blanco-Portillo, J., Sandoval, K., Hagelberg, E., Barberena-Jonas, C., Hill, A.V.S., Rodríguez-Rodríguez, J.E., Fox, K., Robson, K., Haoa-Cardinali, S., et al. (2021). Paths and timings of the peopling of Polynesia inferred from genomic networks. *Nature* *597*, 522–526.
36. Browning, S.R., and Browning, B.L. (2015). Accurate Non-parametric Estimation of Recent Effective Population Size from Segments of Identity by Descent. *The American Journal of Human Genetics* *97*, 404–418.

37. Palamara, P.F., Lencz, T., Darvasi, A., and Pe'er, I. (2012). Length Distributions of Identity by Descent Reveal Fine-Scale Demographic History. *The American Journal of Human Genetics* *91*, 809–822.
38. Borda, V., Alvim, I., Mendes, M., Silva-Carvalho, C., Soares-Souza, G.B., Leal, T.P., Furlan, V., Scliar, M.O., Zamudio, R., Zolini, C., et al. (2020). The genetic structure and adaptation of Andean highlanders and Amazonians are influenced by the interplay between geography and culture. *Proc Natl Acad Sci USA* *117*, 32557–32565.
39. Kelleher, J., Etheridge, A.M., and McVean, G. (2016). Efficient Coalescent Simulation and Genealogical Analysis for Large Sample Sizes. *PLoS Comput Biol* *12*, e1004842.
40. Pierron, D., Heiske, M., Razafindrazaka, H., Pereda-Ioth, V., Sanchez, J., Alva, O., Arachiche, A., Boland, A., Olaso, R., Deleuze, J.-F., et al. (2018). Strong selection during the last millennium for African ancestry in the admixed population of Madagascar. *Nat Commun* *9*, 932.
41. Gravel, S., Henn, B.M., Gutenkunst, R.N., Indap, A.R., Marth, G.T., Clark, A.G., Yu, F., Gibbs, R.A., The 1000 Genomes Project, Bustamante, C.D., et al. (2011). Demographic history and rare allele sharing among human populations. *Proceedings of the National Academy of Sciences* *108*, 11983–11988.
42. Pedersen, C.-E.T., Lohmueller, K.E., Grarup, N., Bjerregaard, P., Hansen, T., Siegismund, H.R., Moltke, I., and Albrechtsen, A. (2017). The Effect of an Extreme and Prolonged Population Bottleneck on Patterns of Deleterious Variation: Insights from the Greenlandic Inuit. *Genetics* *205*, 787–801.
43. Cox, M.P., Hudjashov, G., Sim, A., Savina, O., Karafet, T.M., Sudoyo, H., and Lansing, J.S. (2016). Small Traditional Human Communities Sustain Genomic Diversity over Microgeographic Scales despite Linguistic Isolation. *Mol Biol Evol* *33*, 2273–2284.
44. Hudjashov, G., Karafet, T.M., Lawson, D.J., Downey, S., Savina, O., Sudoyo, H., Lansing, J.S., Hammer, M.F., and Cox, M.P. (2017). Complex Patterns of Admixture across the Indonesian Archipelago. *Molecular Biology and Evolution* *34*, 2439–2452.
45. Larena, M., Sanchez-Quinto, F., Sjödin, P., McKenna, J., Ebeo, C., Reyes, R., Casel, O., Huang, J.-Y., Hagada, K.P., Guilay, D., et al. (2021). Multiple migrations to the Philippines during the last 50,000 years. *Proc Natl Acad Sci USA* *118*, e2026132118.
46. Tumonggor, M.K., Karafet, T.M., Hallmark, B., Lansing, J.S., Sudoyo, H., Hammer, M.F., and Cox, M.P. (2013). The Indonesian archipelago: an ancient genetic highway linking Asia and the Pacific. *J Hum Genet* *58*, 165–173.
47. Allibert, C. (2012). Les réseaux de navigation du début de l'ère chrétienne au XVI^e siècle. Rencontre de populations, échanges commerciaux et matrimoniaux, concurrence

à l'ouest et à l'est de Madagascar. In *Topoi. Orient-Occident. Supplément* 11, 2012. *Autour du Périples de la mer Érythrée. Topoi. Orient-Occident.*

48. Beaujard, P. (2013). Madagascar and Africa, Austronesian migration. In *The Encyclopedia of Global Human Migration*, I. Ness, ed. (Wiley).

49. Blench, R.M. (2008). The Austronesians in Madagascar and their interaction with the Bantu of the East African coast: Surveying the linguistic evidence for domestic and translocated animals. *Philippines Journal of Linguistics* 18, 18–43.

50. Blench, R., and Dendo, M. (2006). The Austronesians in Madagascar and on the East African coast: surveying the linguistic evidence for domestic and translocated animals. In *International Conference on Austronesian Languages Palawan: 17-20 January 2006*.

51. Adelaar, A. (2006). 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*, pp. 205–232.

52. Fitzpatrick, S.M., and Callaghan, R. (2008). Seafaring simulations and the origin of prehistoric settlers to Madagascar. *Islands of inquiry: Colonisation, seafaring and the archaeology of maritime landscapes*, 47–58.

53. Verin, P., Kottak, C.P., and Gorlin, P. (1969). The Glottochronology of Malagasy Speech Communities. *Oceanic Linguistics* 8, 26.

54. Beaujard, P. (2003). Les arrivées austronésiennes à Madagascar: vagues ou continuum? *Études Océan Indien*, 59–147.

55. Razafindrazaka, H., Ricaut, F.-X., Cox, M.P., Mormina, M., Dugoujon, J.-M., Randriamarolaza, L.P., Guitard, E., Tonasso, L., Ludes, B., and Crubézy, E. (2010). Complete mitochondrial DNA sequences provide new insights into the Polynesian motif and the peopling of Madagascar. *Eur J Hum Genet* 18, 575–581.

56. Boivin, N., Crowther, A., Helm, R., and Fuller, D.Q. (2013). East Africa and Madagascar in the Indian Ocean world. *J World Prehist* 26, 213–281.

57. Choudhury, A., Sengupta, D., Ramsay, M., and Schlebusch, C. (2021). Bantu-speaker migration and admixture in southern Africa. *Human Molecular Genetics* 30, R56–R63.

58. Semo, A., Gayà-Vidal, M., Fortes-Lima, C., Alard, B., Oliveira, S., Almeida, J., Prista, A., Damasceno, A., Fehn, A.-M., Schlebusch, C., et al. (2020). Along the Indian Ocean Coast: Genomic Variation in Mozambique Provides New Insights into the Bantu Expansion. *Molecular Biology and Evolution* 37, 406–416.

59. Burns, S.J., Godfrey, L.R., Faina, P., McGee, D., Hardt, B., Ranivoharimanana, L., and Randrianasy, J. (2016). Rapid human-induced landscape transformation in

Madagascar at the end of the first millennium of the Common Era. *Quaternary Science Reviews* 134, 92–99.

60. Voarintsoa, N.R.G., Railsback, L.B., Brook, G.A., Wang, L., Kathayat, G., Cheng, H., Li, X., Edwards, R.L., Rakotondrazafy, A.F.M., and Madison Razanatseheno, M.O. (2017). Three distinct Holocene intervals of stalagmite deposition and nondeposition revealed in NW Madagascar, and their paleoclimate implications. *Clim. Past* 13, 1771–1790.

61. Hixon, S.W., Douglass, K.G., Godfrey, L.R., Eccles, L., Crowley, B.E., Rakotozafy, L.M.A., Clark, G., Haberle, S., Anderson, A., Wright, H.T., et al. (2021). Ecological Consequences of a Millennium of Introduced Dogs on Madagascar. *Front. Ecol. Evol.* 9, 689559.

62. Crowley, B.E. (2010). A refined chronology of prehistoric Madagascar and the demise of the megafauna. *Quaternary Science Reviews* 29, 2591–2603.

63. Radimilahy Chantal (1985). Contribution à l'étude de l'ancienne métallurgie du fer à Madagascar. *Musée D'Art et d'Archéologie de l'Université de Madagascar, Travaux et Documents XXV*. Tananarive.

64. Clist Bernard-Olivier (1995). New Field Data on the Ancient Iron Metallurgy of Madagascar. *NYAME AKUMA* 43, 23–27.

65. Hansford, J.P., Lister, A.M., Weston, E.M., and Turvey, S.T. (2021). Simultaneous extinction of Madagascar's megaherbivores correlates with late Holocene human-caused landscape transformation. *Quaternary Science Reviews* 263, 106996.

66. Coop, G., and Przeworski, M. (2007). An evolutionary view of human recombination. *Nat Rev Genet* 8, 23–34.

67. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158.

68. The 1000 Genomes Project Consortium, Corresponding authors, Auton, A., Abecasis, G.R., Steering committee, Altshuler, D.M., Durbin, R.M., Abecasis, G.R., Bentley, D.R., Chakravarti, A., et al. (2015). A global reference for human genetic variation. *Nature* 526, 68–74.

69. Brucato, N., Kusuma, P., Cox, M.P., Pierron, D., Purnomo, G.A., Adelaar, A., Kivisild, T., Letellier, T., Sudoyo, H., and Ricaut, F.-X. (2016). Malagasy Genetic Ancestry Comes from an Historical Malay Trading Post in Southeast Borneo. *Mol Biol Evol* 33, 2396–2400.

70. Mörseburg, A., Pagani, L., Ricaut, F.-X., Yngvadottir, B., Harney, E., Castillo, C., Hoogervorst, T., Antao, T., Kusuma, P., Brucato, N., et al. (2016). Multi-layered population structure in Island Southeast Asians. *Eur J Hum Genet* 24, 1605–1611.

71. Pierron, D., Razafindrazaka, H., Pagani, L., Ricaut, F.-X., Antao, T., Capredon, M., Sambo, C., Radimilahy, C., Rakotoarisoa, J.-A., Blench, R.M., et al. (2014). Genome-wide evidence of Austronesian-Bantu admixture and cultural reversion in a hunter-gatherer group of Madagascar. *Proceedings of the National Academy of Sciences* *111*, 936–941.
72. Teo, Y.-Y., Sim, X., Ong, R.T.H., Tan, A.K.S., Chen, J., Tantoso, E., Small, K.S., Ku, C.-S., Lee, E.J.D., Seielstad, M., et al. (2009). Singapore Genome Variation Project: A haplotype map of three Southeast Asian populations. *Genome Research* *19*, 2154–2162.
73. Busby, G.B., Band, G., Si Le, Q., Jallow, M., Bougama, E., Mangano, V.D., Amenga-Etego, L.N., Enimil, A., Apinjoh, T., Ndila, C.M., et al. (2016). Admixture into and within sub-Saharan Africa. *eLife* *5*, e15266.
74. May, A., Hazelhurst, S., Li, Y., Norris, S.A., Govind, N., Tikly, M., Hon, C., Johnson, K.J., Hartmann, N., Staedtler, F., et al. (2013). Genetic diversity in black South Africans from Soweto. *BMC Genomics* *14*, 644.
75. Montinaro, F., Busby, G.B.J., Gonzalez-Santos, M., Oosthuizen, O., Oosthuizen, E., Anagnostou, P., Destro-Bisol, G., Pascali, V.L., and Capelli, C. (2017). Complex Ancient Genetic Structure and Cultural Transitions in Southern African Populations. *Genetics* *205*, 303–316.
76. Pagani, L., Kivisild, T., Tarekegn, A., Ekong, R., Plaster, C., Gallego Romero, I., Ayub, Q., Mehdi, S.Q., Thomas, M.G., Luiselli, D., et al. (2012). Ethiopian Genetic Diversity Reveals Linguistic Stratification and Complex Influences on the Ethiopian Gene Pool. *The American Journal of Human Genetics* *91*, 83–96.
77. Patin, E., Siddle, K.J., Laval, G., Quach, H., Harmant, C., Becker, N., Froment, A., Régault, B., Lemée, L., Gravel, S., et al. (2014). The impact of agricultural emergence on the genetic history of African rainforest hunter-gatherers and agriculturalists. *Nat Commun* *5*, 3163.
78. Schlebusch, C.M., Skoglund, P., Sjödin, P., Gattepaille, L.M., Hernandez, D., Jay, F., Li, S., De Jongh, M., Singleton, A., Blum, M.G.B., et al. (2012). Genomic Variation in Seven Khoe-San Groups Reveals Adaptation and Complex African History. *Science* *338*, 374–379.
79. Schlebusch, C.M., Prins, F., Lombard, M., Jakobsson, M., and Soodyall, H. (2016). The disappearing San of southeastern Africa and their genetic affinities. *Hum Genet* *135*, 1365–1373.
80. Brucato, N., Fernandes, V., Mazières, S., Kusuma, P., Cox, M.P., Ng'ang'a, J.W., Omar, M., Simeone-Senelle, M.-C., Frassati, C., Alshamali, F., et al. (2018). The Comoros Show the Earliest Austronesian Gene Flow into the Swahili Corridor. *The American Journal of Human Genetics* *102*, 58–68.

81. Manichaikul, A., Mychaleckyj, J.C., Rich, S.S., Daly, K., Sale, M., and Chen, W.-M. (2010). Robust relationship inference in genome-wide association studies. *Bioinformatics* 26, 2867–2873.
82. Browning, S.R., and Browning, B.L. (2007). Rapid and Accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies By Use of Localized Haplotype Clustering. *The American Journal of Human Genetics* 81, 1084–1097.
83. Dias-Alves, T., Mairal, J., and Blum, M.G.B. (2018). Loter: A Software Package to Infer Local Ancestry for a Wide Range of Species. *Molecular Biology and Evolution* 35, 2318–2326.
84. Gusev, A., Lowe, J.K., Stoffel, M., Daly, M.J., Altshuler, D., Breslow, J.L., Friedman, J.M., and Pe'er, I. (2008). Whole population, genome-wide mapping of hidden relatedness. *Genome Research* 19, 318–326.
85. R Core Team (2021). R: A language and environment for statistical. R Foundation for Statistical Computing.
86. Li, S., Schlebusch, C., and Jakobsson, M. (2014). Genetic variation reveals large-scale population expansion and migration during the expansion of Bantu-speaking peoples. *Proc. R. Soc. B.* 281, 20141448.
87. Ralph, P., and Coop, G. (2013). The Geography of Recent Genetic Ancestry across Europe. *PLoS Biol* 11, e1001555.
88. Lipson, M., Loh, P.-R., Patterson, N., Moorjani, P., Ko, Y.-C., Stoneking, M., Berger, B., and Reich, D. (2014). Reconstructing Austronesian population history in Island Southeast Asia. *Nat Commun* 5, 4689.
89. Soares, P.A., Trejaut, J.A., Rito, T., Cavadas, B., Hill, C., Eng, K.K., Mormina, M., Brandão, A., Fraser, R.M., Wang, T.-Y., et al. (2016). Resolving the ancestry of Austronesian-speaking populations. *Hum Genet* 135, 309–326.
90. Ko, A.M.-S., Chen, C.-Y., Fu, Q., Delfin, F., Li, M., Chiu, H.-L., Stoneking, M., and Ko, Y.-C. (2014). Early Austronesians: Into and Out Of Taiwan. *The American Journal of Human Genetics* 94, 426–436.
91. Choin, J., Mendoza-Revilla, J., Arauna, L.R., Cuadros-Espinoza, S., Cassar, O., Larena, M., Ko, A.M.-S., Harmant, C., Laurent, R., Verdu, P., et al. (2021). Genomic insights into population history and biological adaptation in Oceania. *Nature* 592, 583–589.
92. Xu, S., Pugach, I., Stoneking, M., Kayser, M., Jin, L., and The HUGO Pan-Asian SNP Consortium (2012). Genetic dating indicates that the Asian-Papuan admixture through Eastern Indonesia corresponds to the Austronesian expansion. *Proceedings of the National Academy of Sciences* 109, 4574–4579.

93. Zhou, Y., Qiu, H., and Xu, S. (2017). Modeling Continuous Admixture Using Admixture-Induced Linkage Disequilibrium. *Sci Rep* 7, 43054.
94. Gignoux, C.R., Henn, B.M., and Mountain, J.L. (2011). Rapid, global demographic expansions after the origins of agriculture. *Proceedings of the National Academy of Sciences* 108, 6044–6049.
95. Albrechtsen, A., Nielsen, F.C., and Nielsen, R. (2010). Ascertainment Biases in SNP Chips Affect Measures of Population Divergence. *Molecular Biology and Evolution* 27, 2534–2547.
96. Gopalan, S., Berl, R.E.W., Myrick, J.W., Garfield, Z.H., Reynolds, A.W., Bafens, B.K., Belbin, G., Mastoras, M., Williams, C., Daya, M., et al. (2022). Hunter-gatherer genomes reveal diverse demographic trajectories during the rise of farming in Eastern Africa. *Current Biology* 32, 1852-1860.e5.
97. Hudjashov, G., Karafet, T.M., Lawson, D.J., Downey, S., Savina, O., Sudoyo, H., Lansing, J.S., Hammer, M.F., and Cox, M.P. (2017). Complex Patterns of Admixture across the Indonesian Archipelago. *Molecular Biology and Evolution* 34, 2439–2452.