

The rise of baobab trees in Madagascar

<https://doi.org/10.1038/s41586-024-07447-4>

Received: 20 May 2023

Accepted: 19 April 2024

Published online: 15 May 2024

Open access

 Check for updates

Jun-Nan Wan^{1,2,12}, Sheng-Wei Wang^{1,2,12}, Andrew R. Leitch^{3,12}, Ilia J. Leitch^{4,12}, Jian-Bo Jian^{5,6,12}, Zhang-Yan Wu⁵, Hai-Ping Xin¹, Mijoro Rakotoarinivo⁷, Guy Eric Onjalalaina⁷, Robert Wahiti Gituru^{2,8}, Can Dai⁹, Geoffrey Mwachala¹⁰, Ming-Zhou Bai^{5,6}, Chen-Xi Zhao⁵, Hong-Qi Wang⁵, Sheng-Lan Du¹, Neng Wei^{1,2}, Guang-Wan Hu^{1,2}, Si-Chong Chen^{1,2}, Xiao-Ya Chen^{2,11}, Tao Wan^{1,2}✉ & Qing-Feng Wang^{1,2}✉

The baobab trees (genus *Adansonia*) have attracted tremendous attention because of their striking shape and distinctive relationships with fauna¹. These spectacular trees have also influenced human culture, inspiring innumerable arts, folklore and traditions. Here we sequenced genomes of all eight extant baobab species and argue that Madagascar should be considered the centre of origin for the extant lineages, a key issue in their evolutionary history^{2,3}. Integrated genomic and ecological analyses revealed the reticulate evolution of baobabs, which eventually led to the species diversity seen today. Past population dynamics of Malagasy baobabs may have been influenced by both interspecific competition and the geological history of the island, especially changes in local sea levels. We propose that further attention should be paid to the conservation status of Malagasy baobabs, especially of *Adansonia suarezensis* and *Adansonia grandidieri*, and that intensive monitoring of populations of *Adansonia za* is required, given its propensity for negatively impacting the critically endangered *Adansonia perrieri*.

The genus *Adansonia*, better known as the baobabs and ‘mother of the forest’¹, has captivated botanists, tourists, naturalists and passers-by for centuries. Probably the earliest record of humans marvelling at these amazing trees can be traced back to the Ancient Egyptians, around 2,300 BC (ref. 1). With their grotesque appearance, enormous size¹, reputed longevity⁴ and diverse uses⁵, baobabs have become one of the most charismatic species on our planet. Embedded in folklore and tradition, baobabs have inspired innumerable pieces of art and have been associated with human settlements and cultures over millennia^{6,7}.

Adansonia (family Malvaceae, subfamily Bombacoideae) comprises eight morphologically distinct species^{8,9}. The extant geographic distribution of baobabs is unusual, with one tetraploid species, *Adansonia digitata* ($2n = 4x = 168$), which is widespread across continental Africa, one diploid species, *Adansonia gregorii* ($2n = 2x = 88$), restricted to Northwestern Australia, and six diploid species ($2n = 2x = 84, 88$) which are endemic to Madagascar³ (Fig. 1a). A substantial body of research has detailed the natural history of *Adansonia*, including its taxonomy, ethnobotany, ecology and physiology^{1,8}. All but *A. digitata* are listed in The IUCN Red List of Threatened Species, 2023, with three of the Malagasy species threatened with extinction (*Adansonia perrieri* classified as critically endangered and *Adansonia grandidieri* and *Adansonia suarezensis* as endangered). The remaining Malagasy species are now listed as being of least concern but their declining populations indicate that more rigorous conservation strategies are required to ensure the long-term survival of these culturally and globally important species. For that to happen, a detailed understanding of the genetics of baobabs is urgently needed.

Here we present high-quality, chromosomal-level genome assemblies for all eight baobab species (Supplementary Fig. 1). Using these data together with ecological analyses, we uncover the evolutionary history of this genus, including its origin, diversification, population history and patterns of interspecific hybridization. Our findings provide valuable insights for developing effective and biologically informed conservation strategies for these spectacular trees in a changing world¹⁰.

The genetic landscape of baobab genomes

Clustering of scaffolds from Hi-C (high-throughput chromosome conformation capture) libraries validated cytological data showing that baobab karyotypes are comprised of numerous small chromosomes (Supplementary Fig. 2). The base chromosome number of $n = 44$ was conserved in all diploid *Adansonia* species, except *A. perrieri*, consistent with previous counts for *Adansonia*¹¹. The chromosome count of *A. perrieri* ($n = 42$), was validated by the Hi-C contact map and syntenic alignment to *A. grandidieri* ($n = 44$) (Supplementary Fig. 3).

Macrosynteny alignments revealed high levels of synteny between species (Supplementary Fig. 4a and Supplementary Table 1) and suggest that chromosomal rearrangements and fusions occurring in the lineage leading to *A. perrieri* (Supplementary Fig. 4b) gave rise to its reduced chromosome number. Alignments of *A. perrieri* to outgroup Malvaceae species *Theobroma cacao* ($2n = 20$) and *Gossypium raimondii* ($2n = 26$) showed several scattered syntenic blocks in *A. perrieri* (Supplementary Fig. 4c).

¹State Key Laboratory of Plant Diversity and Specialty Crops, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan, China. ²Sino-Africa Joint Research Centre, Chinese Academy of Sciences, Wuhan, China. ³School of Biological and Chemical Sciences, Queen Mary University of London, London, UK. ⁴Royal Botanic Gardens, Kew, London, UK. ⁵BGI Genomics, BGI-Shenzhen, Shenzhen, China. ⁶Department of Biotechnology and Biomedicine, Technical University of Denmark, Lyngby, Denmark. ⁷University of Antananarivo, Antananarivo, Madagascar. ⁸Department of Botany, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya. ⁹School of Resources and Environmental Science, Hubei University, Wuhan, China. ¹⁰East African Herbarium, National Museums of Kenya, Nairobi, Kenya. ¹¹Shanghai Chenshan Botanical Garden, Shanghai, China. ¹²These authors contributed equally: Jun-Nan Wan, Sheng-Wei Wang, Andrew R. Leitch, Ilia J. Leitch, Jian-Bo Jian. ✉e-mail: wantao@wbgcas.cn; qfwang@wbgcas.cn

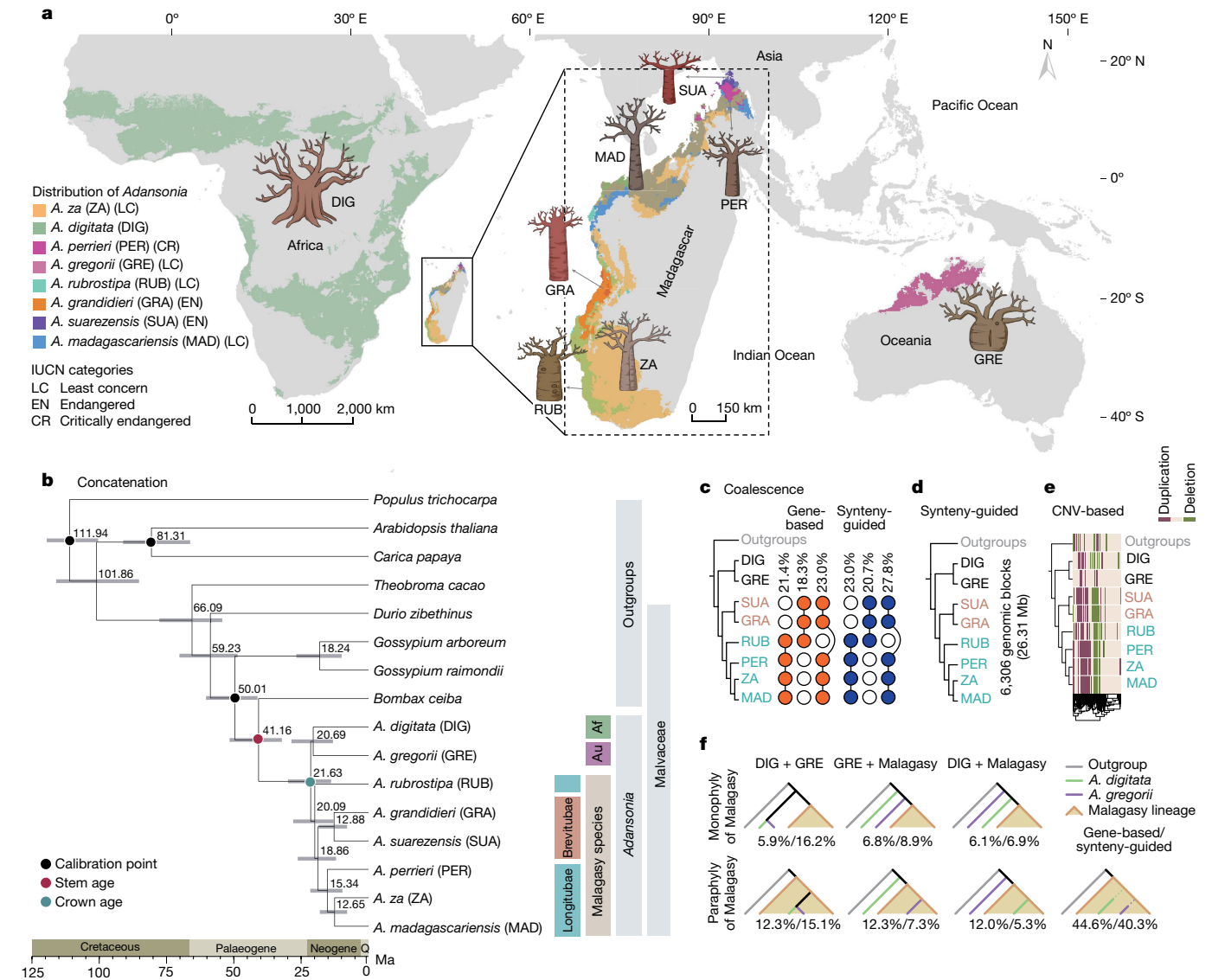


Fig. 1 | Geographic distribution and phylogenetic relationships between species of *Adansonia*. **a**, The distribution of eight baobab species and their conservation status in IUCN. The coloured patches represent the occurrence of each baobab species. **b**, A maximum-likelihood tree from the concatenation of 999 SCN genes. Median ages of the nodes are shown and a timescale is provided at the bottom; grey horizontal bars show 95% CIs of estimated divergence dates between lineages. The geographic occurrence of the *Adansonia* species is designated: Af, African lineage; Au, Australian lineage and Malagasy species. **c**, Coalescence analyses of 999 concatenated SCN genes. Summary of the proportion of gene tree topologies using either SCN genes (that is, gene-based) or syntenic-guided genomic blocks. **d**, Maximum-likelihood phylogenetic

relationships of *Adansonia* inferred from analysing concatenated syntenic-guided genomic blocks. **e**, The *Adansonia* phylogeny based on CNVs (calculated as a pairwise edit distance of duplications and deletions). *A. gregorii* was selected for the identification of CNVs to circumvent any potential bias caused by using species with a closer phylogenetic relationship as the reference. **f**, The proportions of conflicting maximum-likelihood gene tree topologies. The acronyms used for the species are as in **a**. Numbers represent the percentage of trees with the given topology (SCN gene dataset/syntenic-guided genomic blocks). The topologies are sketched to show the different placements of each lineage, regardless of the branch length.

Analysis of total coverage of *K*-mer pairs indicates that *A. digitata* is an autotetraploid (Supplementary Fig. 5). Although analyses of synonymous substitutions per synonymous site (*K_s*) in paralogous gene pairs failed to identify any further whole-genome duplications (WGD) with the divergence of *Adansonia*, evidence of a shared ancient WGD that is either Malvaceae-specific or specific to the common ancestor of Bombacoideae/Malvoideae (Supplementary Fig. 6), confirmed earlier findings¹². The higher chromosome numbers in Bombacoideae compared with most other species in Malvaceae^{13,14} may reflect this ancestral WGD, with reduced numbers in Malvoideae arising from fusions in the lineage leading to that tribe. However, chromosome rearrangements and fissions in the lineage leading to Bombacoideae may also have given

rise to their high chromosome numbers. See Supplementary Note 1 for a summary of genome size diversity in *Adansonia* and Malvaceae.

The *Adansonia* phylogeny

The eight extant baobab species form three geographic lineages, *A. digitata* on continental Africa, *A. gregorii* in Northwestern Australia and the remainder in Madagascar³. Owing to the limited number of loci used in previous phylogenetic studies³, the interspecific relationships between *Adansonia* species have remained controversial, making it difficult to infer patterns of geographic divergence and the centre of origin for the three extant lineages.

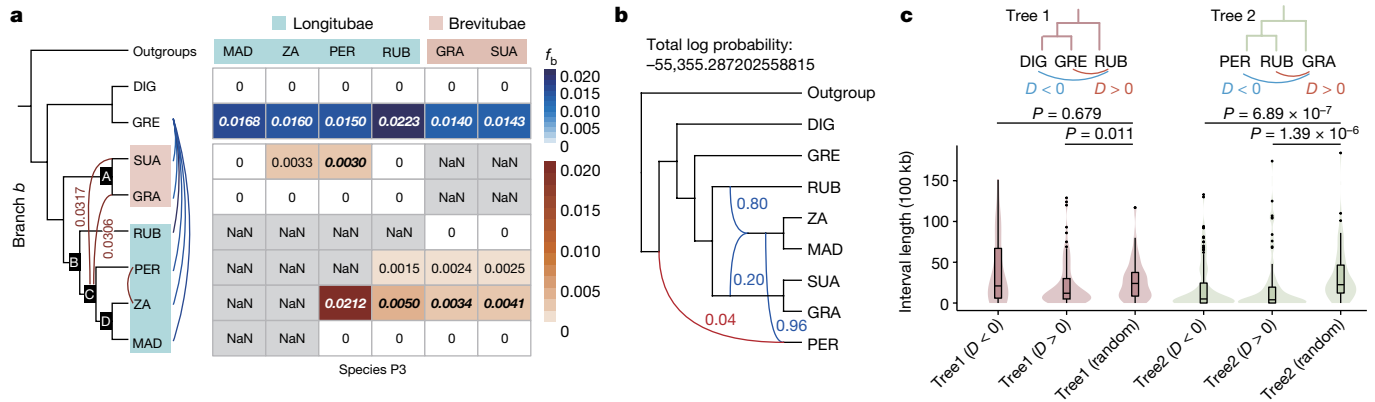


Fig. 2 | Test for introgression. **a**, D -statistics calculated between *Adansonia* species branches or internal tree branches (branch b , ordered A–D) and baobabs species (P3; acronyms as in Fig. 1a). The associated table is in two parts, the upper part (*Bombax ceiba* as the outgroup; *A. rubrostipa* as the reference genome) and lower part (*A. gregorii* as the outgroup and the reference genome) have cells coloured according to their f_b -branch score (f_b) and show gene flow between each species tree branch and the species P3. C–*A. suarezensis* and C–*A. grandidieri* had high-value f_b ('0' indicates a zero score; italics depicts Z-score greater than 3; 'NaN' indicates an unperformable test). **b**, The best-supported network derived from analyses of 999 concatenated SCN gene trees

To dissect the evolutionary history of baobabs, we analysed our whole-genome sequence assemblies for the eight baobabs with eight outgroup species (five species from Malvaceae and three from other families; Supplementary Table 2). Concatenated datasets of 999 single-copy nuclear (SCN) genes were extracted to establish a phylogenetic framework for *Adansonia*. The highest proportion of maximum-likelihood trees indicated that *A. digitata* and *A. gregorii* form a sister clade to the Malagasy lineage (Fig. 1b and Supplementary Fig. 7a). Secondary calibrated molecular clocks indicated a stem lineage origin of *Adansonia* at about 41.1 (95% confidence interval (CI) 52.0–32.4) million years ago (Ma), far younger than the break-up of Gondwana (about 160 Ma)¹⁵ (Fig. 1b), supporting previous analyses². The diversification of Malagasy baobabs probably occurred over the last 20.6 (95% CI 28.8–13.3) to 12.6 (95% CI 18.2–8.0) Myr (Fig. 1b). There was partial support for the two morphologically defined sections Brevitubae and Longitubae⁸, but the results placed *Adansonia rubrostipa* as sister to the two sections, indicating paraphyly of Longitubae (Fig. 1b). However, in coalescent-based species trees of the 999 SCN genes by ASTRAL (Supplementary Fig. 8a), monophyly of Longitubae was observed in 21.4% of the trees recovered, occurring without topological consensus for the placement of *A. rubrostipa* (Fig. 1c).

To circumvent potential bias brought about by natural selection on non-neutral gene sites, we also conducted a synteny-guided phylogenomic analysis of all baobab species and *G. raimondii*, an approach that incorporated more non-coding DNA. The analysis used a concatenated block of 6,306 highly conserved syntenic segments (≥ 3 kilobases (kb), synteny-guided genomic blocks) which were assembled into a 26.31 megabase (Mb) fragment. The maximum-likelihood tree supported a topology with *A. digitata* + *A. gregorii* being sister to the rest of the baobabs and monophyly of sections Brevitubae and Longitubae (Fig. 1d and Supplementary Fig. 7b). However, monophyly of Longitubae was not recovered in coalescent-based species trees using these synteny-guided genomic blocks (Supplementary Fig. 8b). Similar results were observed when using small (2 kb) alignment blocks to reduce any potential influence from intralocus recombination breakpoints (Supplementary Fig. 9). We extended our phylogenetic analyses to use copy number variations (CNVs) (Supplementary Table 3) which are less likely to occur independently in separate lineages than DNA polymorphisms (for example, single-nucleotide polymorphisms)

in PhyloNet in runs with a 50% bootstrap threshold. Reticulation edges are coloured blue and the minor inheritance in *A. perrieri* is highlighted in red. **c**, Kernel density plots of the interval length between adjacent windows of introgressed genomic blocks (100 kb). The type of introgression is shown in **c** and for each tree topology, 100 windows were sampled randomly as a control. One-sided Wilcoxon test; left to right, $n = 43, 109, 63, 136, 71, 62$. In box plots, the centre line is the median, the box limits show the interquartile range (25th to 75th percentile), whiskers extend to quartiles $\pm 1.5 \times$ interquartile range, and dots show potential outliers.

and are valuable markers for phylogenetic reconstructions¹⁶. The neighbour-joining tree of CNV genotypes (Fig. 1e) recapitulated the phylogenetic relationships seen with the concatenated synteny-guided genomic blocks (Fig. 1d) and many of the coalescence gene trees (Fig. 1c), leading to an emerging consensus of the *Adansonia* phylogeny comprising monophyly of sections Brevitubae and Longitubae. Furthermore, we revisited the plastid phylogeny⁸ and showed a well-supported Longitubae clade (Extended Data Fig. 1) but with some discordance between nuclear and plastid phylogenies in the placement of *A. suarezensis*, perhaps because the latter retains a different ancestral plastome than the other species.

Despite an overall consensus on the monophyly of the Malagasy lineage, we observed a large proportion of SCN gene trees (81.2%) and synteny-guided genomic block trees (68.1%) which showed paraphyly of the Malagasy baobabs (Fig. 1f). We also noticed that more trees supported a clade containing *A. gregorii* and the Malagasy baobabs as sister to *A. digitata* than a clade containing *A. digitata* and Malagasy baobabs as sister to *A. gregorii* (19.1% versus 18.1% in SCN gene trees and 16.2% versus 12.2% in synteny-guided genomic block trees) (Fig. 1f). Such genome-wide phylogenetic discordances indicate a complex scenario of reticulations in baobabs³, involving introgression and/or incomplete lineage sorting (ILS) of ancestral variation^{17,18}.

Gene flow among baobab species

To explore the impact of introgression and ILS on the *Adansonia* phylogeny, we used three approaches: (1) Patterson's D -statistics and f -branch scores to identify gene flow among baobab species; (2) the tree-based method QuIBL (quantifying introgression by means of branch lengths) designed to distinguish introgression from ILS¹⁷; and (3) PhyloNet¹⁹ to recover putative reticulate evolution.

Patterson's D -statistics detected an excess of shared derived alleles between *A. gregorii* and each of the Malagasy species, providing strong evidence of introgression (Fig. 2a and Extended Data Fig. 2a). To account for potential bias caused by mapping resequenced data to a reference genome using D -statistics²⁰, we replicated the comparisons by using each baobab species as the reference. Similar results were found with all reference genomes, except when *A. gregorii* was the reference (Extended Data Fig. 2a, Supplementary Fig. 10 and Supplementary Table 4). For the

Malagasy species, the frequent occurrence of shared alleles between non-sister lineages is probably a result of interspecific hybridization (Extended Data Fig. 2a and Supplementary Table 4). For example, analysis of f -branch score f_b values indicates extensive gene flow between *Adansonia za* and *A. perrieri* (high f_b value of 0.0212; Fig. 2a), a result that is supported by admixed microsatellite genotypes in individuals of *A. perrieri* and *A. za* occurring in sympatry and morphologically intermediate individuals²¹. It suggests that hybridization is continuing because of the overlapping distributions and flowering times of the two species²¹. Similarly, gene flow is predicted between the lineages leading to *A. za* and *A. rubrostipa* ($f_b = 0.0050$). Importantly, f_b values indicate gene flow between the ancestors of each Brevitubae species and the most recent common ancestor (MRCA) of Longitubae, excluding *A. rubrostipa* (Fig. 2a and Extended Data Fig. 2b). Such gene flow could have occurred before a shift in flowering phenology of the two sections (Brevitubae flower in the dry season and Longitubae in the rainy season⁸). As further evidence of introgression, we compared the topology of SCN gene trees in genomic regions with different introgression signals ($D > 0$ and $D < 0$). The result showed that genes in introgressed regions involving branches leading to Brevitubae and core Longitubae ($D < 0$) contributed significantly to the variable phylogenetic placement of *A. rubrostipa* (Extended Data Fig. 2c), an interpretation that is contrary to past assertions of introgression from *A. digitata* to Brevitubae³.

Evidence of introgression is supported by QuIBL, which shows that a tree topology ((*A. gregorii*, Malagasy species that is *A. rubrostipa*), *A. digitata*) has a lower Bayesian information criterion (BIC) value for an 'introgression + ILS' model than an 'ILS-only' model (Supplementary Table 5), even though less than 0.6% of loci were considered as introgressed (Extended Data Fig. 2d). This low estimation of introgression has probably been obscured by ILS among the three geographic lineages which have a short internal branch length in the phylogenetic tree (Fig. 1b) and similar proportions of variant tree topologies (that is, ((*A. digitata*, *A. gregorii*), Malagasy lineage), ((*A. digitata*, Malagasy lineage), *A. gregorii*), ((*A. gregorii*, Malagasy lineage), *A. digitata*)) (Fig. 1f). Given a sister relationship between *A. digitata* and *A. gregorii* (Fig. 1b) and gene flow between *A. gregorii* and the Malagasy baobabs (Fig. 2a), we suggest that *A. gregorii* was once in sympatry with the Malagasy baobabs, meaning that gene flow occurred after the split separating the Malagasy clade from the other two species. In addition, both D -statistics and QuIBL analysis showed signatures of gene flow between *A. digitata* and Malagasy baobabs, suggesting that *A. digitata* or the common ancestors of *A. digitata* and *A. gregorii* may have been in sympatry with the Malagasy baobabs (Extended Data Fig. 2a, d and Supplementary Table 5).

Further support for introgression was seen by analysing the dataset of 999 SCN genes using PhyloNet, implementing six search networks and allowing for zero to five reticulation events. The best-supported species networks showed that Longitubae, including *A. rubrostipa*, form a clade with a major inheritance probability (80%), whereas the placement of *A. rubrostipa* as sister to the entire Malagasy clade occurs with a minor inheritance probability (20%) (Fig. 2b). Given a consensus of monophyly of Longitubae in the *Adansonia* phylogeny, it is most likely that the 20% of loci are derived by introgression. The phylogenetic network is best interpreted as the product of historical hybridization between core Longitubae and the MRCA of Brevitubae, a result that is compatible with the D -statistics analyses (Fig. 2a). This reticulation event also explains the lack of topological consensus for the placement of *A. rubrostipa* in gene trees (Fig. 1c and Supplementary Figs. 7 and 8). The closer placement of *A. gregorii* to the Malagasy baobabs than between those species and *A. digitata* may also have resulted from introgression between the Malagasy baobabs and *A. gregorii*. Notably, PhyloNet species networks consistently identified about 4% admixture from a distantly related lineage into *A. perrieri* only. Although values of less than 5% admixture should be treated with caution, the stability of the signal suggests that it may be a biological feature reflecting the residual retention of ancestral genomic characters, possibly as a result

of an extremely small population size²² (Fig. 2b and Extended Data Fig. 3). Small population sizes can fix introgressed segments rapidly by genetic drift, although of course they can also be lost rapidly, which might explain why this signal is not observed in other baobab species. Nevertheless, the possibility of the edge in *A. perrieri* as an artefact cannot be ruled out, as perhaps suggested by the lack of an *A. gregorii* + *A. digitata* clade on the species tree network.

The timing of proposed introgression events can be inferred by analysing the consecutiveness of introgressed regions in the genome, given that recent reticulation events are expected to result in longer stretches of introgressed segments than are more ancient events²³. We investigated the length of the interval between two adjacent genomic windows of 100 kb in length that carry a significant signal of introgression (that is, D values deviated significantly from zero, $D > 0$ and $D < 0$). We compared these interval lengths against those returned from 100 random windows of 100 kb (Fig. 2c and Supplementary Fig. 11). Many windows with significant D values were found in the Malagasy species with interval lengths of 0 bp, indicating large and recent introgressed regions. By contrast, for the three geographic lineages, only *A. gregorii*-Malagasy species genomic windows had significant D values ($D > 0$), showing signatures of consecutive introgressed regions (Fig. 2c), which indicates earlier introgressions than those that occurred in Malagasy baobabs. As for gene flow involving *A. digitata*, the distribution pattern of introgressed windows suggests more ancient introgression, which could have occurred before polyploidy, followed by considerable recombination, breaking up introgressed segments as the polyploid genome diverged²⁴.

Genetic diversity and population dynamics

To investigate the genetic diversity of baobabs, we characterized genome-wide heterozygosity (GWH) and runs of homozygosity (ROH) in baobab genomes (*A. digitata* was excluded as it is autopolyploid) (Fig. 3a and Supplementary Tables 6 and 7). Both of the two endangered species, *A. suarezensis* and *A. grandidieri*, had lower GWH and numerous segments of long ROH (more than 2 Mb), which suggest high levels of recent inbreeding^{25,26} (Fig. 3a and Extended Data Fig. 4). Unexpectedly, the genome of the critically endangered *A. perrieri* exhibited high GWH (1.33%) despite the presence of abundant long ROHs (Fig. 3a). Compared to *A. suarezensis* and *A. grandidieri*, the higher GWH of *A. perrieri* could have arisen from gene flow with other Longitubae species (that is, *A. za*) (Fig. 2a). The abundant long ROHs may also have arisen by inbreeding between a few individuals in a constrained geographical range (Fig. 1a). *Adansonia madagascariensis*, *A. rubrostipa* and *A. za* had more short ROHs than other species, which may represent genomic remnants of ancient inbreeding combined with recent outcrossing (Fig. 3b). In contrast to the Malagasy species, the Australian species *A. gregorii* on average had shorter and fewer ROH segments and relatively low GWH (1.04%) (Fig. 3a), which probably arose from the 'bottleneck' caused by its dispersal to Australia²⁷. The small number of ROH in *A. gregorii* could then be interpreted as a genetic consequence of its rapid radiation and high intraspecific gene flow⁷ in Australia.

We estimated changes in effective population size (N_e) through the Pleistocene using the pairwise sequentially Markovian coalescent (PSMC) model²⁸. Analysis revealed an exponential increase in N_e for all Malagasy baobabs over the period 10–1 Ma (Fig. 4a). Then, from about 1 Ma, two distinct population dynamics emerged, with *A. grandidieri*, *A. suarezensis* and *A. perrieri* showing a decrease in N_e , whereas the other Malagasy species (*A. za*, *A. madagascariensis* and *A. rubrostipa*) exhibited a lag phase before a decrease in N_e lasting about 0.3 Myr and then a significant recovery in N_e around 0.1 Ma (Fig. 4a). Inferences from coupled logistic models for species experiencing ecological competition²⁹ are consistent with these data when models are perturbed by a dramatic and asynchronous population size decrease (Methods; Fig. 4b). It is likely that geological and environmental events

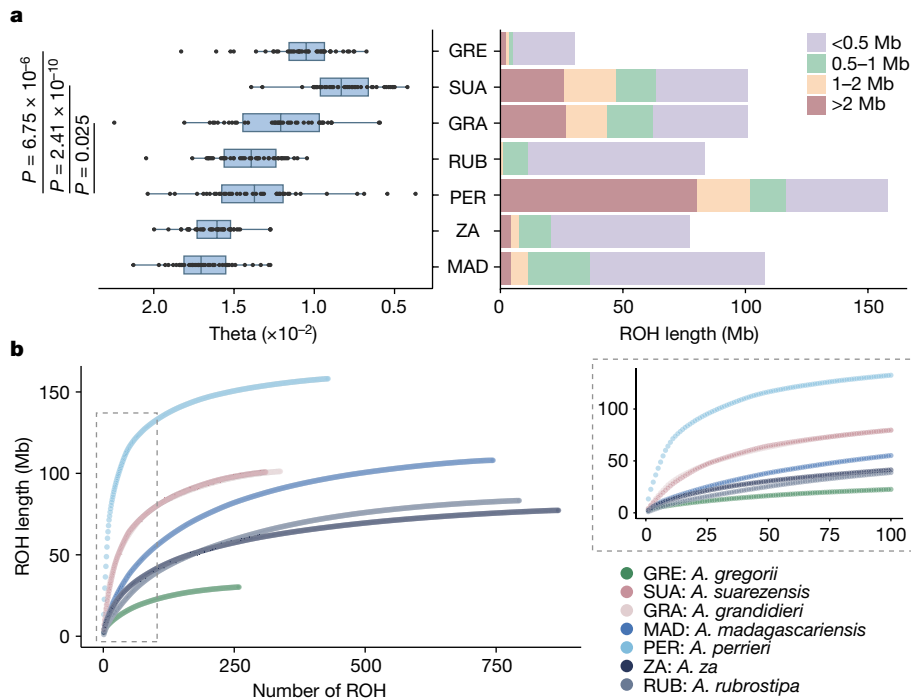


Fig. 3 | The genomic footprints of inbreeding and hybridization in diploid baobabs. **a**, GWH (left) and ROH (right) in seven baobab species (*A. digitata* is excluded because it is an autopolyploid). The heterozygosity statistics for all chromosomes of each species are presented as a box plot. The two-sided Wilcoxon test showed significantly higher levels of GWH in *A. perrieri* ($n = 42$)

in Madagascar (for example, uplift of mountains, mantle upwelling and volcanism)^{10,30,31}, along with severe climate oscillations in East Africa around 1 Ma (ref. 32) may have contributed to these population dynamics. Furthermore, estimated inverse coalescence rates are a reflection of not only the census population size but also historical gene flow and population structures³³. For instance, the significant increase of N_e in *A. za* at about 0.1 Ma could have arisen from population growth and extensive gene flow with other Malagasy baobabs, whereas the small N_e of *A. perrieri* over the last approximately 1 Ma may be due to high levels of inbreeding caused by low numbers of individuals and/or population fragmentation.

Influences of geology and the environment

To understand the demography of Malagasy taxa and to infer past population dynamics, integrated ecological analyses were performed using present distributions (Supplementary Table 8) and 19 present- and palaeo-bioclimatic variables and altitude data from the WorldClim database v.2.0 (www.worldclim.org)³⁴ (Supplementary Table 9).

Ecological valence analysis revealed that *A. za* can tolerate the widest range of environmental conditions, encompassing many of the ecological niches occupied by the other species (Fig. 4c and Supplementary Table 10). Potentially *A. za* is strongly competitive for habitat space³⁵ and able to replace more locally distributed species when in sympatry (for example, *A. suarezensis*, *A. grandidieri* and *A. perrieri*). Niche equivalency and similarity statistical tests also demonstrated that the ecological niche of all Longitubae species overlapped, with *A. za* having the highest coefficient of overlap (Schoener's $D = 0.411419$) (Fig. 4d). The two endangered species, *A. suarezensis* and *A. grandidieri*, despite their close phylogenetic distance, have very different ecological niches, which suggests local adaptation after geographic isolation³⁶ (Fig. 4d). Principal component analysis of environmental drivers of ecological niche differentiation indicated that the spatial heterogeneity

compared to species with, on average, longer ROHs ($n = 44$). In box plots, the centre line is the median, the box limits show the interquartile range (25th to 75th percentile) and whiskers extend to quartiles $\pm 1.5 \times$ interquartile range. **b**, The relationship between the number and lengths of ROH. The dashed box shows the distribution pattern of the top 100 longest ROH in each species.

of temperature, altitude and precipitation in Madagascar (Extended Data Fig. 5) could be the principal environmental factors leading to the present distribution of baobabs, with *A. madagascariensis* and *A. rubrostipa* mainly influenced by temperature and altitude, whereas the remaining species were mainly affected by precipitation and altitude (Extended Data Fig. 6).

The varied environmental drivers of ecological niche differentiation in Malagasy baobabs, encouraged us to project their historical distribution using species distribution modelling (MaxEnt)³⁷ (Extended Data Fig. 7) and spatial superimposition modelling (Extended Data Fig. 8). Both models indicated that almost all species previously had considerably larger ranges on the western coast of the island during the last glacial maximum (LGM; about 22,000 years before present; Extended Data Figs. 7 and 8), when global sea levels were low³⁸. The results also imply that *A. suarezensis* and *A. perrieri* had more suitable habitats in the north of Madagascar at that time, suggesting that their distribution may have been continuous rather than fragmented as it is today.

We examined the likely effects of global mean sea level (GMSL)³⁸ changes over the last 10 Myr on the distribution of Malagasy baobabs (Fig. 4e). The results indicated that about 4 Ma the island was approximately 11% smaller than today and 19% larger around 0.02 Ma (Supplementary Table 11). Given that past local sea levels were also influenced by the Earth's viscoelastic structure and global ice sheet history³⁹, the relative sea level (RSL) changes along the west coast of Madagascar from 0.12 Ma to the present were predicted, adopting from ref. 40 (Supplementary Fig. 12). The results were generally consistent with the predictions from GMSL changes (Fig. 4e). The present area of the island is likely to be relatively small, having declined from its largest extent during the LGM (when RSL reached about -120 m; Fig. 4e). During times of relatively low sea level, vast areas of Madagascar were probably suitable for baobab population expansion and dispersal, whereas periods of high sea level led to smaller suitable areas, population fragmentation, species isolation and reduced gene flow. From 10 Ma to present,

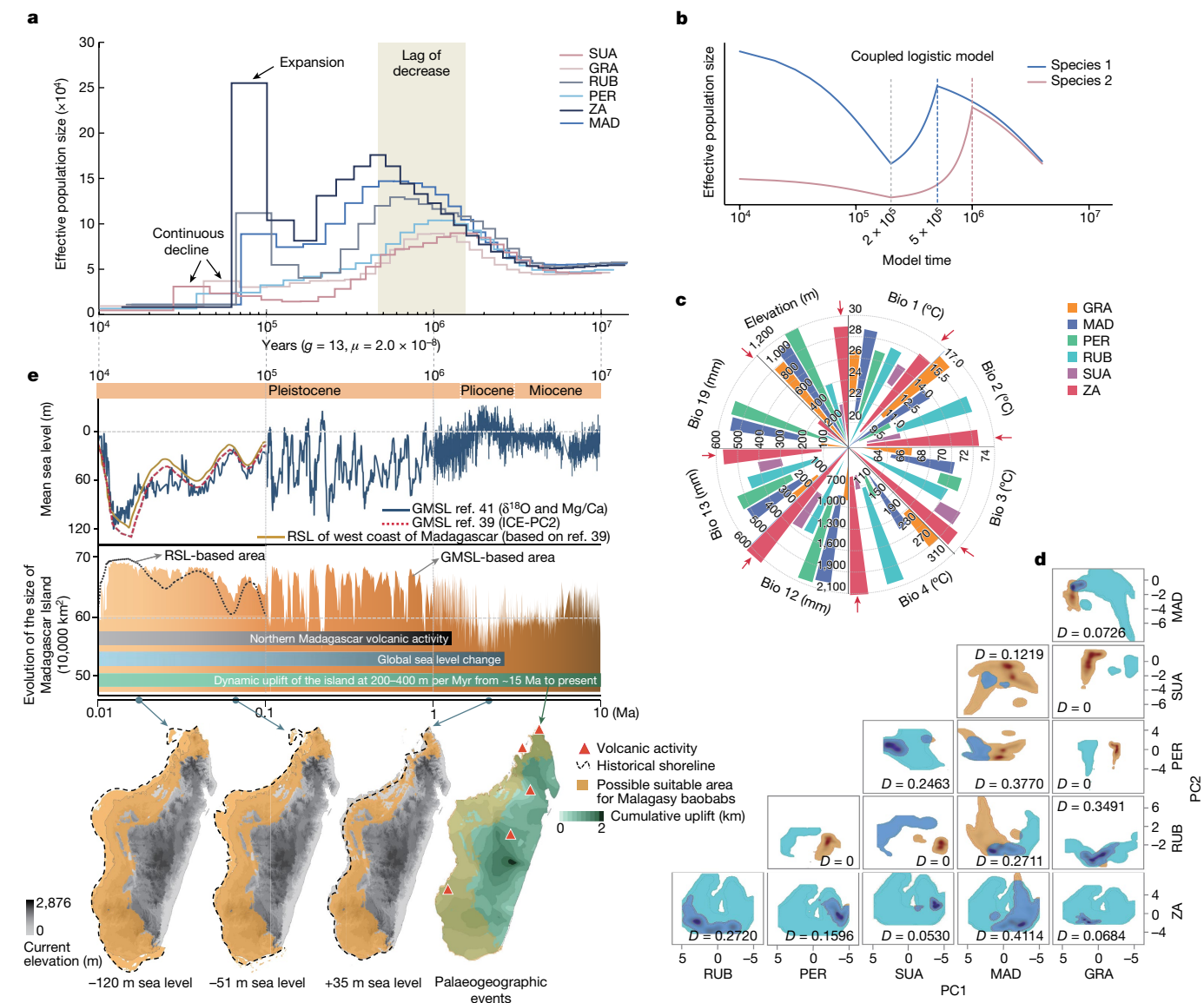


Fig. 4 | The ecological drivers of population dynamics in Malagasy baobabs. **a**, A PSMC model inferring the effective population sizes of Malagasy baobabs. A generation time of 13 years was used to translate substitution rates to ages. **b**, A coupled logistic model supporting the asynchronous population size decreases from around 1 Ma apparent in the PSMC model (Methods). **c**, Variation in ecological valence of six Malagasy baobabs to different environmental factors. The red arrows highlight the widest ecological range observed for *A. za*. **d**, Niche equivalency and similarity statistical tests and patterns of niche overlaps between baobabs. Coloured patches represent ecological niche differentiation of species pairs (pale blue and brown with areas of highest

density being most intense; x axis, PC1; y axis, PC2). Areas of niche overlap are shown in blue. **e**, Geological events impacting the distribution of Malagasy baobabs. Changing sea levels over time (top, GMSL, RSL)^{39,48} was inferred by different models (Methods). The dynamics of the land area of Madagascar (middle) were derived from a calculation using the GMSL (orange patch) and RSL (dash line). The timescale of reported geological events^{31,45,49} is indicated by coloured strips. The predicted suitable areas of habitat for baobabs were mapped onto the simulated territories of the island under different sea-level scenarios (bottom). The acronyms used for species are as in Fig. 1a.

mountain uplift (total 1–2 km at 200–400 m per Myr from about 15 Ma to present) and volcanic activity in the north of Madagascar^{30,41} (Fig. 4e) have also occurred^{10,42} and this also probably influenced areas suitable for baobabs (Fig. 4e). Changes in sea level and palaeogeographic events probably explain the limited distributions of *A. suarezensis* and *A. perrieri* in the north today, arising from a probable wider population range in the past (Fig. 4e). Those species with wider ecological valence (that is, *A. za* and *A. madagascariensis*) may have had more opportunities for population expansion, generating a dynamic that could have contributed to population declines of other baobab species. These interfered population dynamics are corroborated by the genomic signatures observed in the PSMC model. For instance, a remarkable decrease in population size probably occurred in *A. suarezensis* and

A. perrieri over the last 1 Myr, a time when the northern area was impacted by sea level rises and volcanic activity (Fig. 4e).

The continental origin of baobabs

In the absence of fossils, it has been difficult to determine the ancestral location of origin for the stem group of *Adansonia* (Supplementary Note 2). However, the results presented here help dissect three hypotheses about the centre of origin for the crown group of baobabs (Extended Data Fig. 9). Critically, evidence of introgression between species indicates that *A. gregorii* is likely to have been in sympatry with the substantially diverged Malagasy baobabs during the diversification of the genus. Assuming so, there are three scenarios which explain the

data from the perspective of an out-of-Africa hypothesis. One scenario is that the Malagasy baobabs must have colonized Madagascar by means of several independent dispersal events before going extinct in Africa. This assumes that the diversification of the Malagasy lineage occurred on the continent of Africa before these independent dispersals, which seems improbable. In a second scenario, if an ancestor of the Malagasy baobabs colonized Madagascar and diversified there (requiring only one dispersal event from Africa), then it necessitates transoceanic gene flow between the Australia-residing *A. gregorii* and each of the species in Madagascar independently. This also seems improbable because there are no validated instances for such transoceanic gene flow. In the third out-of-Africa scenario, there could have been two transoceanic dispersals from Africa to Madagascar, one being the ancestor of the Malagasy baobabs which diversified to form the Malagasy baobabs and another being the *A. gregorii* lineage which hybridized with the Malagasy baobabs before going extinct in Madagascar, requiring also an extra transoceanic dispersal of *A. gregorii* to Australia. Such a complex migration and extinction pattern for *A. gregorii* is not as parsimonious as other hypotheses.

With respect to the 'Australia-origin' hypothesis, as with the 'Africa-origin' hypothesis above, it would require either several independent dispersals or high levels of transoceanic gene flow. Beyond that, the out-of-Australia hypothesis is challenged by a probable bottleneck in the ancestry of *A. gregorii*, which is an expected product of long-distance dispersal.

Compared with the above two hypotheses, the 'Madagascar-origin' hypothesis offers the most reasonable explanation of the present data. First, the Madagascar-origin hypothesis requires only one dispersal event for *A. gregorii* from Madagascar to Australia, alleviating the requirements of unlikely transoceanic gene flow or several migrations. Second, only the Madagascar-origin allows for the assumption that geographical isolation between the Malagasy lineage and the other two baobab species started after the diversification of the Malagasy baobabs, enabling gene flow among all baobabs. This scenario is compatible with the gene flow detected between Malagasy baobabs and *A. gregorii* or *A. digitata*. Moreover, the hypothesis also explains the unexpectedly high proportion of gene trees (81.2% in gene-based trees and 68.1% in syntenic-guided trees) supporting the admixture of the Malagasy lineage and the other two lineages. Assuming a Madagascar-origin, the minor yet stable inheritance of ancestral *Adansonia* genes in *A. perrieri* was derived from long-term sympatry of *Adansonia* on the island. A stronger inference for a Madagascar-origin will be possible if/when relevant fossils and other robust historical information become available.

Building on the Madagascar-origin hypothesis, it is possible that the autopolyploid *A. digitata* arose in Madagascar, then dispersed to east Africa where it spread across the entire continent. However, that scenario is apparently compromised by the occurrence of most ancestral plastid haplotypes of *A. digitata* in West Africa⁴³. Such haplotypes are indicative of an early autopolyploid colonizer being of West African origin, before it later spread eastwards and southwards across much of savannah Africa. To reconcile the conflict, a diploid progenitor of *A. digitata* may have dispersed from Madagascar to Africa (Fig. 5a). Then, probably in west Africa, autopolyploidy occurred leading to the extant tetraploid cytotype which replaced the diploid progenitor. There are several examples showing that autopolyploids can differ in their ecological tolerances from their diploid progenitors, perhaps arising from polysomic inheritance, increased heterozygosity (heterosis) and altered gene expression, all of which can contribute to an increased competitive advantage and range shifts⁴⁴. Potentially, *A. digitata* benefited substantially from the neo-polyploidy event, which led to its present wide distribution across Africa (Fig. 5a).

Madagascar is renowned for its high rates of endemism¹⁰, including *Adansonia*. *Adansonia*'s phenology shifts coupled with geographic isolation and interspecific hybridization have collectively contributed to the diversity of Malagasy baobabs. For the two phylogenetic sections

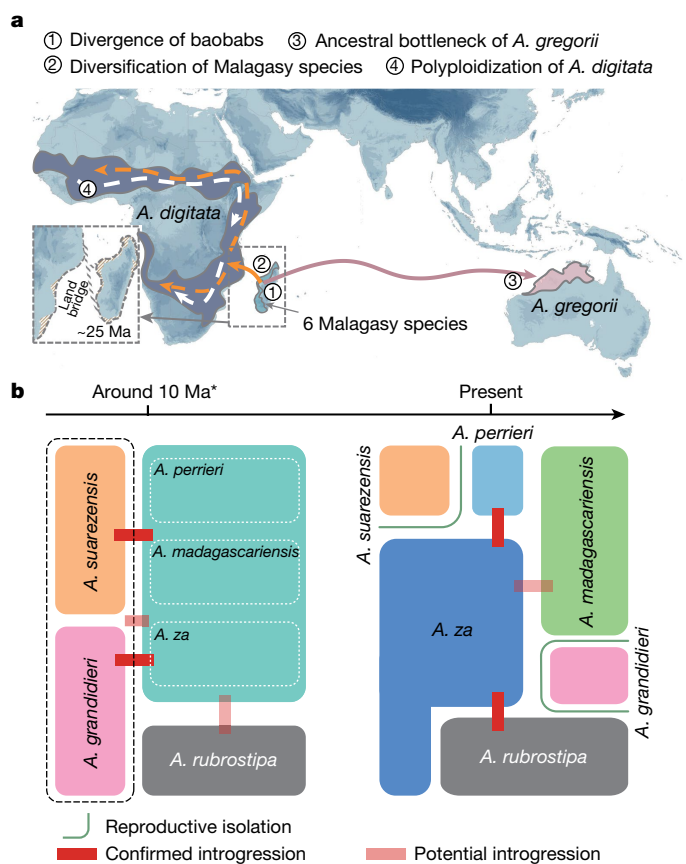


Fig. 5 | The evolutionary history of baobab trees. **a**, A summary of the predicted evolutionary history of baobabs under the Madagascar-origin hypothesis showing the probable migration routes of *A. digitata* and *A. gregorii*. The location of a controversial ancient land bridge^{49,50} between Africa and Madagascar is shown (dotted box) along with key evolutionary events. **b**, A summary of the diversification of Malagasy baobab species, including dynamic changes in population sizes, recurrent introgression and geographic isolation from 10 Ma to present. The size of each species block around 10 Ma corresponds to the estimated population size based on a PSMC model and predicted ecological inferences, and the thick red lines show proposed gene flow based on *f*-branch analysis (translucent red lines indicate proposed introgression that is undetectable using present methods). *Indicates the time approximation given by the PSMC model.

of Malagasy baobabs (Brevitubae and Longitubae), we predict that the fixation of diverged flowering time was a critical step in establishing reproductive barriers (more discussion on baobab pollination ecology is provided in Supplementary Note 3). We also predict that global changes in sea level and local palaeogeographic events, resulting in population isolation and expansion, were important in shaping their population dynamics and facilitated local adaptation. Hybridization between Brevitubae and core Longitubae is also likely to have contributed to the diversification of Malagasy baobabs⁴⁵ (Fig. 5b). These processes are continuing, as indicated by reports of recent hybridization between *A. rubrostipa* and *A. za*²¹. Although diversification of *Adansonia* in Madagascar seems most likely, there is potentially an unresolved story relating to the dispersal of the lineage leading to *Adansonia*, given that the present distribution of the closest relatives to *Adansonia* are South America⁴⁶.

The conservation status of baobabs

The genomic and ecological analyses presented here enable a re-evaluation of the conservation status of baobabs. According to the

IUCN Red List criteria, the two Brevitubae species are now listed as endangered (with *A. suarezensis* coded Blab+2ab and *A. grandidieri* A4c, IUCN Red List, 2023). Our predictions, based on PSMC analyses and the identification of potentially suitable habitats, show considerable population size reductions for these species from 1 to 0.1 Ma and from the LGM (about 22 thousand years ago (ka)) to the present (Supplementary Fig. 13). Despite these reductions being less than 80% over this period (that is, below the threshold for assessment as critically endangered), the large amounts of long ROH in *A. suarezensis* and *A. grandidieri* and the small N_e (based on the PSMC analyses) are features that indicate constrained population sizes²⁶. These features plus their distinct ecological niches, narrow ecological valence and low genetic diversity²² suggest that the two species are likely to have reduced resilience to ecological perturbations and habitat fragmentation³⁶. In addition, coupled logistic models for species experiencing ecological competition fit the past population dynamics of species in Brevitubae and Longitubae (*A. perrieri* excluded) (Fig. 4b), whereas previous studies modelling the distribution of *A. suarezensis* and *A. grandidieri* forward in time to 2050 and 2080 (ref. 47) suggest that climate change will pose severe threats for *A. suarezensis*, leading to its extinction before 2080. We therefore propose the conservation status of *A. suarezensis* be moved from endangered to critically endangered on the basis of criterion A3ce of the IUCN Red List. Beyond that, only three of the six subpopulations of *A. suarezensis* are now included in the protected area network in Madagascar (IUCN Red List, 2023). We therefore recommend that conservation efforts should be focused on the three subpopulations outside the protected area network.

For *A. grandidieri*, given the interspecific competition indicated, we recommend that the IUCN rejustify its endangered status regarding criterion A4ce. More conservation measures and studies are required, especially the evaluation of its population dynamics with respect to other sympatric baobab species. In particular, its suitable habitat completely overlaps with *A. za* (Fig. 4d), a species predicted to have undergone large population expansions in the past (Fig. 4a) and characterized by the widest ecological valence amongst baobabs (Fig. 4c).

Despite *A. perrieri* (critically endangered) having relatively high GWH, mediated by introgression from other baobabs, especially from *A. za* (Fig. 2a), the numerous long ROHs suggest inbreeding and a very small N_e . In addition, PSMC analysis shows that its N_e has been continuously decreasing for about a million years (Fig. 4a). It is likely that *A. za* has been genetically swamping *A. perrieri*, rather than enabling its genetic rescue from the disadvantages of a small N_e . Moreover, hybridization between *A. perrieri* and *A. za* seems to be continuing, given the occurrence of morphologically intermediate individuals²¹. Hence, we highlight the potentially negative genetic influence of *A. za* on *A. perrieri* and recommend that individuals in sympatry should be discouraged if species distinctiveness is desirable.

Further research into the population genetics of baobabs is needed, especially for *A. rubrostipa* and *A. madagascariensis*, which overlap with *A. za* in their distribution and flowering time. Similarly, the present genetic understanding of baobabs would benefit from broader sampling to bridge the gaps in the evolutionary history of baobabs. The key priority for safeguarding *Adansonia*, especially the two species in section Brevitubae and *A. perrieri*, should be focused on providing data to guide both in situ and ex situ conservation programmes, to enable these iconic species to recover and thrive.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-024-07447-4>.

- Wickens, G. E. *The Baobabs: Pachycauls of Africa, Madagascar and Australia* (Springer, 2008).
- Baum, D. A., Small, R. L. & Wendel, J. F. Biogeography and floral evolution of baobabs (*Adansonia*, Bombacaceae) as inferred from multiple data sets. *Syst. Biol.* **47**, 181–207 (1998).
- Karimi, N. et al. Reticulate evolution helps explain apparent homoplasy in floral biology and pollination in baobabs (*Adansonia*; Bombacoideae; Malvaceae). *Syst. Biol.* **69**, 462–478 (2020).
- Patrut, A. et al. Radiocarbon dating of a very large African baobab. *Tree Physiol.* **27**, 1569–1574 (2007).
- Gebauer, J., El-Siddig, K. & Ebert, G. Baobab (*Adansonia digitata* L.): a review on a multipurpose tree with promising future in the Sudan. *Eur. J. Hort. Sci.* **67**, 155–160 (2002).
- Duval, C. S. Human settlement and baobab distribution in south-western Mali. *J. Biogeogr.* **34**, 1947–1961 (2007).
- Rangan, H. et al. New genetic and linguistic analyses show ancient human influence on baobab evolution and distribution in Australia. *PLoS ONE* **10**, e0119758 (2015).
- Baum, D. A. The comparative pollination and floral biology of baobabs (*Adansonia*—Bombacaceae). *Ann. Mo. Bot. Gard.* **82**, 322–348 (1995).
- Baum, D. A. A systematic revision of *Adansonia* (Bombacaceae). *Ann. Mo. Bot. Gard.* **82**, 440–471 (1995).
- Antonelli, A. et al. Madagascar's extraordinary biodiversity: evolution, distribution and use. *Science* **378**, eabf0869 (2022).
- Baum, D. A. & Oginuma, K. A review of chromosome numbers in Bombacaceae with new counts for *Adansonia*. *Taxon* **43**, 11–20 (1994).
- Conover, J. L. et al. A Malvaceae mystery: a mallow maelstrom of genome multiplications and maybe misleading methods? *J. Integr. Plant Biol.* **61**, 12–31 (2019).
- Marinho, R. C. et al. Do chromosome numbers reflect phylogeny? New counts for Bombacoideae and a review of Malvaceae s.l. *Am. J. Bot.* **101**, 1456–1465 (2014).
- Cvetković, T. et al. Phylogenomics resolves deep subfamilial relationships in Malvaceae s.l. *Genes Genomes Genet.* **11**, jkab136 (2021).
- Yoder, A. D. & Nowak, M. D. Has vicariance or dispersal been the predominant biogeographic force in Madagascar? Only time will tell. *Annu. Rev. Ecol. Syst.* **37**, 405–431 (2016).
- Peter, H. et al. Evolution and diversity of copy number variation in the great ape lineage. *Genome Res.* **23**, 1373–1382 (2013).
- Edelman, N. B. et al. Genomic architecture and introgression shape a butterfly radiation. *Science* **366**, 594–599 (2019).
- Meleshko, O. et al. Extensive genome-wide phylogenetic discordance is due to incomplete lineage sorting and not ongoing introgression in a rapidly radiated bryophyte genus. *Mol. Biol. Evol.* **38**, 2750–2766 (2021).
- Wen, D. Q., Yu, Y., Zhu, J. F. & Nakhleh, L. Inferring phylogenetic networks using PhyloNet. *Syst. Biol.* **67**, 735–740 (2018).
- Green, R. E. A draft sequence of the Neandertal genome. *Science* **328**, 710–722 (2010).
- Leong Pock-Tsy, J. M. et al. Nuclear microsatellite variation in Malagasy baobabs (*Adansonia*, Bombacoideae, Malvaceae) reveals past hybridization and introgression. *Ann. Bot.* **112**, 1759–1773 (2013).
- Whitlock, M. C. Fixation of new alleles and the extinction of small populations: drift load, beneficial alleles and sexual selection. *Evolution* **54**, 1855–1861 (2000).
- Barlow, A. et al. Partial genomic survival of cave bears in living brown bears. *Nat. Ecol. Evol.* **2**, 1563–1570 (2018).
- del Pozo, J. C. & Ramirez-Parra, E. Whole genome duplications in plants: an overview from *Arabidopsis*. *J. Exp. Bot.* **66**, 6991–7003 (2015).
- Palkopoulou, E. et al. Complete genomes reveal signatures of demographic and genetic declines in the woolly mammoth. *Curr. Biol.* **25**, 1395–1400 (2015).
- Ceballos, F. C., Hazellhurst, S. & Ramsay, M. Assessing runs of homozygosity: a comparison of SNP array and whole genome sequence low coverage data. *BMC Genomics* **19**, 106 (2018).
- Bell, K. L. et al. Genetic diversity and biogeography of the baobab *Adansonia gregorii* (Malvaceae: Bombacoideae). *Aust. J. Bot.* **62**, 164–174 (2014).
- Li, H. & Durbin, R. Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493–496 (2011).
- Sepkoski, J. J. Jr., McKinney, F. K. & Lidgard, S. Competitive displacement among post-Paleozoic cyclostome and cheilostome bryozoans. *Paleobiology* **26**, 7–18 (2000).
- Roberts, G. G., Paul, J. D., White, N. & Winterbourne, J. Temporal and spatial evolution of dynamic support from river profiles: a framework for Madagascar. *Geochem. Geophys. Geosyst.* **13**, Q04004 (2012).
- Stephenson, S. N. et al. Cenozoic dynamic topography of Madagascar. *Geochem. Geophys. Geosyst.* **22**, e2020GC009624 (2021).
- Johnson, T. C. et al. A progressively wetter climate in southern East Africa over the past 1.3 million years. *Nature* **537**, 220–224 (2016).
- Chikhi, L. et al. The IICR (inverse instantaneous coalescence rate) as a summary of genomic diversity: insights into demographic inference and model choice. *Heredity* **120**, 13–24 (2018).
- Fick, S. E. & Hijmans, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land area. *Int. J. Climatol.* **37**, 4302–4315 (2017).
- Whittaker, R. H. Dominance and diversity in land plant communities: numerical relations of species express the importance of competition in community function and evolution. *Science* **147**, 250–260 (1965).
- Aguilar, R. et al. Habitat fragmentation reduces plant progeny quality: a global synthesis. *Ecol. Lett.* **22**, 1163–1173 (2019).
- Phillips, S. J., erson, R. P. & Schapire, R. E. Maximum entropy modeling of species geographic distributions. *Ecol. Model.* **190**, 231–259 (2006).
- Fernández-Palacios, J. M. et al. Towards a glacial-sensitive model of island biogeography. *Glob. Ecol. Biogeogr.* **25**, 817–830 (2016).

39. Pico, T., Creveling, J. R. & Mitrovica, J. X. Sea-level records from the US mid-Atlantic constrain Laurentide Ice Sheet extent during Marine Isotope Stage 3. *Nat. Commun.* **8**, 15612 (2017).
40. Borreggine, M. et al. Sea-level rise in Southwest Greenland as a contributor to Viking abandonment. *Proc. Natl Acad. Sci. USA* **120**, e2209615120 (2023).
41. Miller, K. G. et al. Cenozoic sea-level and cryospheric evolution from deep-sea geochemical and continental margin records. *Sci. Adv.* **6**, eaaz1346 (2020).
42. Liu, Y., Wang, Y., Willett, S. D., Zimmermann, N. E. & Pellissier, L. Escarpment evolution drives the diversification of the Madagascar flora. *Science* **383**, 653–658 (2024).
43. Leong Pock-Tsy, J. et al. Chloroplast DNA phylogeography suggests a West African centre of origin for the baobab, *Adansonia digitata* L. (Bombacoideae, Malvaceae). *Mol. Ecol.* **18**, 1707–1715 (2009).
44. Parisod, C., Holderegger, R. & Brochmann, C. Evolutionary consequences of autopolyploidy. *New Phytol.* **186**, 5–17 (2010).
45. Marques, D. A., Meier, J. I. & Seehausen, O. A combinatorial view on speciation and adaptive radiation. *Trends Ecol. Evol.* **34**, 531–544 (2019).
46. Carvalho-Sobrinho, J. G. et al. Revisiting the phylogeny of Bombacoideae (Malvaceae): novel relationships, morphologically cohesive clades and a new tribal classification based on multilocus phylogenetic analyses. *Mol. Phylogenet. Evol.* **101**, 56–74 (2016).
47. Vieilledent, G. et al. Vulnerability of baobab species to climate change and effectiveness of the protected area network in Madagascar: towards new conservation priorities. *Biol. Conserv.* **166**, 11–22 (2013).
48. Goodman, S. M. *The New Natural History of Madagascar* (Princeton Univ. Press, 2022).
49. Masters, J. C. et al. Biogeographic mechanisms involved in the colonization of Madagascar by African vertebrates: rifting, rafting and runways. *J. Biogeogr.* **48**, 492–510 (2021).
50. Ali, J. R. & Hedges, S. B. A review of geological evidence bearing on proposed Cenozoic land connections between Madagascar and Africa and its relevance to biogeography. *Earth Sci. Rev.* **232**, 104103 (2022).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024

Plant materials and whole-genome sequencing

Young leaves of *A. digitata* and *A. gregorii* were collected from trees in wild populations. The ripe fruits of Malagasy baobabs were collected from natural populations and seeds were further germinated for fresh young leaves at a local nursery garden in Antananarivo (Madagascar). Information on voucher specimens and the locality of the individuals sampled are provided in Supplementary Fig. 14. Genomic DNAs were extracted locally using the DNeasy Plant Mini kit (Qiagen) with a modified protocol. A washing step was included before CTAB buffer extraction to exclude the secondary metabolites in baobab leaves. The washing buffer contained 50 mM Tris-HCl, 5 mM EDTA-2Na, 0.35 M D-sorbitol, 1% (w/v) polyvinylpyrrolidone (PVP-K30) and 1% beta-mercaptoethanol. The RNA was isolated using TRIzol (Invitrogen) and further treated with RNase-free DNase I (Promega).

Next-generation sequencing (NGS) sequence datasets were generated (DNBSEQ short reads, PacBio long reads and Hi-C short reads) for genome assemblies. The fragmented reads (2×150 bp) with insert sizes of 300–500 bp for paired-end reads were produced using the DNBSEQ platforms. For the PacBio long reads, the high-molecular-weight gDNA with an insert size of approximately 20 kb was collected using BluePippin (Sage Science) and was sequenced on PacBio Sequel I/ Sequel II platform. The DNA libraries were further sequenced on PacBio Sequel I SMRT cells with continuous long reads or PacBio Sequel II SMRT cells in circular consensus sequence (CCS) mode. Fresh young leaves were fixed in a 2% formaldehyde solution for Hi-C library construction. The nuclei/ chromatin was digested with MboI (New England Biolabs) as previously described⁵¹. Finally, the Hi-C libraries were sequenced with DNBSEQ to produce 150 bp paired-end reads.

K-mer analysis for genome size estimation

The short reads were first filtered by SOAPnuke (v.1.5.6) (-n 0.01 -l 15 -q 0.4 -i -Q 2 -G -M 2 -A 0.5 -d)⁵². The clean data were then used for K-mer analysis to estimate genome size using jellyfish (v.2.3.0)⁵³ and Genomescope1.0 (ref. 54). The ploidy level of *A. digitata* was further analysed using Genomescope 2.0 and Smudgeplot (v.0.2.5)⁵⁵.

Genome assembly and contigs anchoring

Falcon v.0.5 (ref. 56), a hierarchical, haplotype-aware genome software, was used to generate the initial contigs. Meanwhile, the pa_HPC-daligner_option with parameters set to '-v -D24 -B8 -M32 -h300 -w8 -e.75 -l2000 -s1000 -k18' and the ovlp_HPCdaligner_option with '-v -D251 -B24 -M36 -h600 -e.96 -l1000 -s1000 -w6 -k25 -T8' were used for pre-assembly and read corrections. The initial contigs were further polished with all sequenced PacBio long reads using the Arrow algorithm (<https://github.com/lixuenan/GenomicConsensus>). Then, the polished genome sequences were error-corrected using the short insert size reads by Pilon (v.1.23)⁵⁷ with default settings. The whole-genome assemblies were further processed with purge_haplotigs (https://bitbucket.org/mroachawri/purge_haplotigs/src/master/) to remove redundant sequences.

For the PacBio CCS HiFi data, the primary contigs were assembled using Hifiasm (v.0.15.1)⁵⁸ with default parameters. To anchor the contigs to chromosomes, the Hi-C cleaned reads were aligned to the assembled genome. The global mapped reads were obtained using bwa (v.0.7.12). The Juicer pipeline (v.1.5)⁵⁹ was applied to process the alignments of Hi-C reads. The 3d-dna (v.180922)⁶⁰ was used to link the contig assemblies into chromosome-length scaffolds guided by linking information. The Hi-C contact maps were processed with 500 kb using the visualized module of 3d-dna and reviewed in JuiceBox (v.1.9.0) (<https://github.com/aidenlab/Juicebox>).

Genome annotation and assessment

Two strategies, including homology-based and de novo methods, were applied for repetitive sequence annotation. The homology-based method was performed using the database of repeat sequences in RepBase (v.21.12) (<http://www.girinst.org/repbase>)⁶¹. On the basis of Repbase, RepeatModeler (v.2.0)⁶² (<http://www.repeatmasker.org/RepeatModeler/>) and LTR_FINDER (v.1.07)⁶³ (http://tlife.fudan.edu.cn/ltr_finder/) were also used to identify repetitive sequences in baobab genomes. For the de novo method, we first generated a repetitive sequence database using de novo prediction based on the features of sequence alignment and transposable element structure. Then, the repetitive sequences were refined and verified by REPEATMASKER (v.3.3.0; <http://www.repeatmasker.org>). In addition, tandem repeats were identified using Tandem Repeats Finder (v.4.09)⁶⁴ (<http://tandem.bu.edu/trf/trf.html>).

Three different algorithms, transcriptome data-based prediction, homology-based prediction and ab initio prediction, were collectively applied for gene structure prediction and annotation. The following pipeline was used for each baobab species studied here. (1) RNA-seq reads from root, stem and leaf were, respectively, mapped to the assembled genome using hist2 (v.2.1.0)⁶⁵. The transcripts were generated by stringtie (v.2.1.6)⁶⁶ on the basis of the aligned results. (2) The protein sequences of *Arabidopsis thaliana*⁶⁷, *Gossypium arboreum*⁶⁸, *Durio zibethinus*⁶⁹, *T. cacao*⁷⁰ and *G. raimondii*⁷¹ were aligned to the baobab genomes using GeMoMa (v.1.6.4)⁷² with default parameters. On the basis of the results of homology prediction, the genes with complete integrity of structure were preserved to train the gene model analysis which was then used with AUGUSTUS v.3.2.3 (ref. 73) and SNAP (v.2006-07-28)⁷⁴. (3) For de novo prediction, two gene prediction programmes, AUGUSTUS (v.3.2.1)⁷³ and SNAP were used to predict coding regions in the repeat-masked genome. Finally, the information of repetitive sequences, transcripts and homology-based gene predictions were integrated into the Maker-P (v.2.31.11)⁷⁵ with the following parameters 'max_dna_len=100000, split_hit=25000, always_complete=1, split_hit=25000, est2genome=0, protein2genome=0'.

Historical activity of LTR retrotransposons

The LTR retrotransposon candidates (LTR-RTs) were de novo detected in the genomes of the eight baobab species through LTR_Finder (v.1.0.7)⁶³ and LTRharvest incorporated in Genome Tools (v.1.5.8)⁷⁶ with default parameters. The two datasets were integrated to remove the false positives using the LTR_retriever package (v.1.9) with default parameters. To further classify LTR-RTs into detailed clades, the TESorter (v.1.3.0)⁷⁷ (<https://github.com/zhangrengang/TEsorter>) hidden Markov model profile-based classifier was used with default settings, taking as reference the protein domains found in the REXdb Viridiplantae version of the database⁷⁸. Meanwhile, LTR_retriever (v.1.9)⁷⁹ was also applied for estimating the insertion times of LTR-RTs. The rice mutation rate of 1.3×10^{-8} mutations per site per year was adopted in this study.

Syntenic analysis and whole-genome duplication detection

The syntenic gene pairs between species were identified using JCVI (v.1.1.11, a Python version of MCScan, <https://github.com/tanghaibao/jcvi/wiki/MCScan-Python-version>). By using the coding sequences (CDS) and annotation gff3 files as the input data, the syntenic blocks for each paired species were detected using 'jcvi.compara.catalog.ortholog' with the parameter '-cscore=0.7'. The syntenic blocks were filtered by 'jcvi.compara.synteny.screen' with parameters '-minspan=90 -simple'. The MCScanX-170908 (ref. 80) was used to construct chromosomal syntenic blocks at interspecific and intraspecific levels (parameters: '-a -e 1e-5 -s 5'). With the HKY substitution model⁸¹, the paralogous and orthologous gene pairs in syntenic blocks were used to calculate fourfold synonymous third-codon transversion rate (4dTv) distances to look for evidence of WGDs⁸². The K_2 distribution

was calculated using the Nei–Gojobori method implemented in yn00 programme of PAML (v.4.5)⁸³.

Orthologous gene set construction

A total of 15 species were used for orthogroup sorting, including the whole *Adansonia* genus (eight baobab species), four other Malvaceae species (*T. cacao*, *D. zibethinus*, *G. arboretum* and *G. raimondii*) and three outgroup species (*Populus trichocarpa*, *A. thaliana* and *Carica papaya*). The protein sequence references were collected from the available data sources (Supplementary Table 2). The longest transcripts were chosen for those genes with alternative splicing variants. The similarities between protein sequences were calculated using BLASTP (2.9.0+) with an e-value threshold of 1×10^{-5} . Then, OrthoMCL (v.1.4)⁸⁴ was used to identify gene families on the basis of the similarities of the genes. Markov chain clustering (MCL) was used with default parameter settings except with a main inflation value of 1.5.

Phylogenetic inference and molecular dating

A total of 1,086 single-copy orthologous genes were obtained from the above orthogroup sorting and were further aligned by MUSCLE (v.3.8.31)⁸⁵. The protein alignments were converted to corresponding CDS. Each amino acid was substituted for the corresponding triplet bases from its CDS according to the same ID information. These coding sequences for each single-copy gene family were concatenated to form one super gene for each species. The secondary nucleotides of each codon with phasing alignment were extracted and used for constructing the phylogenetic tree using MrBayes (v.3.1.2)⁸⁶ under the GTR model. The divergence time between species was estimated using the MCMCTREE algorithm from the PAML package (v.4.5)⁸³, using an autocorrelated clock model based on the inferred phylogenetic relationships. Two calibration times from the TimeTree database (<http://www.timetree.org/>) for (1) the divergence between *A. thaliana* and *C. papaya* (54–90 Ma), (2) *A. thaliana* and *P. trichocarpa* (100–120 Ma) and (3) the bombacoid–malvoid split⁸⁷ (66.3–42.6 Ma) were used for the molecular dating. Coalescence gene trees were constructed with each gene applied using RAXML-8.2.11. Then coalescence gene trees were reconstructed using ASTRAL-5.6.1 (ref. 88) and each topology was calculated.

For the syntenic-based phylogenetic analysis, the eight baobab genomes were aligned to the *G. raimondii* ‘D5-4’ reference genome (https://www.cottongen.org/species/Gossypium_raimondii/NSF-D5) using LASTZ1.04 (<https://github.com/lastz/lastz/>) with default parameters. The pairwise alignments were systematically chained and integrated to generate high-quality blocks. Subsequently, the non-syntenic regions were excluded and pairwise local alignments were merged to generate a final multispecies whole-genome alignment using MULTIZ (v.11.2)⁸⁹. Syntenic fragments of more than 3 kb were considered for further concatenation. The concatenated fragment represented by *G. raimondii* was subjected to a sliding window-based phylogenetic analysis with a window size of 20 kb. Each individual window was used to construct a phylogenetic tree using RAXML-8.2.11. The likelihoods of each potential topology were input into CONSEL (v.1.20)⁹⁰ and an approximately unbiased test was conducted with a significance level of $P > 0.95$.

De novo assembly of the plastome sequence of each baobab species was performed by NOVOPlasty 4.3.1 (ref. 91). After alignment with MUSCLE, we used RAXML to infer maximum-likelihood trees using the GTR-Γ model with 1,000 bootstrap replicates. The *Bombax ceiba* plastome sequence was set as the outgroup.

Variant calling and filtering

For each baobab species, the clean paired-end reads from NGS were mapped back onto its assembled genome with bwa (v.0.7.17)⁹² using default parameters. Samtools (v.1.9)⁹³ was then used to sort the aligned reads and Picard (v.2.21.6)⁹⁴ was used to remove redundant reads.

Variant detection and filtering were carried out using GATK (v.4.1.2.0)⁹⁵ with the filtering parameters set to ‘QD < 2.0, QUAL < 30.0, SOR > 3.0, FS > 60.0, MQ < 40.0, MQRankSum < -12.5, ReadPosRankSum < -8.0’. The output VCF file set was named as ‘align_SELF’. To avoid any probable conflicting results caused by choice of reference in downstream analyses, we aligned the cleaned NGS reads of each baobab species to the assembled genome of all other baobab genomes. After variant calling and filtering, all output files based on the same species’ genome reference formed a dataset and named as ‘align_REF’.

Identification of copy number variations

To detect CNVs in all baobab species and the outgroup *B. ceiba*⁹⁶, the sorted BAM file of align_REF was processed by CNVpytor (v.1.2.1)⁹⁷ for genome alignment based on the genome of *A. gregorii*. The CNV calling was carried out by setting a bin size of 1,000 bp and the following criteria were set to filter false positive candidates: the CNV calls with $P < 0.0001$, sizes greater than 5 kb, a fraction of mapped reads with zero quality (q_0) > 50% and the distance to the nearest gap in reference genome (d_c) > 10,000. Duplication and deletion information was extracted from the CNV calls and then used to reconstruct the neighbour-joining tree with the R package ape (v.5.6-2).

Gene flow

The Patterson’s D and f_b admixture ratio for all possible combinations of baobab species were calculated using Dtrios in Dsuite (v.0.5 r47)⁹⁸ on the basis of the align_REF dataset. For the calculation of all baobab species, *B. ceiba* was chosen as the outgroup, whereas, for that of the Malagasy baobabs, *A. gregorii* was chosen as the outgroup. To further specify local introgressed genomic regions, Patterson’s D was calculated for the whole genome using ABBA-BABAwindows.py in genomics_general-master2 (v.0.4) with a sliding window of 100,000 bp. We defined the introgressed regions that were ranked in the top 1%, top 5% and top 10% of the absolute D value, respectively.

The tree-based method QuIBL (<https://github.com/miriammyagi/QuIBL>)¹⁷ was used to differentiate introgression from ILS. Because only four lineages can be involved in a single run of QuIBL, we focused on two previous trees: ((*A. digitata*, *A. gregorii*), *A. rubrostipa*), *B. ceiba*) (*B. ceiba* as outgroup) and ((*A. perrieri*, *A. rubrostipa*), *A. grandidieri*), *A. gregorii*) (*A. gregorii* as outgroup). To carry out the QuIBL analysis, we used RAXML to construct the input files of QuIBL on the basis of the fasta alignments of the 999 single-copy orthologous genes. We distinguished between an ILS-only model and an introgression + ILS model using the BIC test with a strict cutoff of $\Delta\text{BIC} > 10$.

To construct the species networks, we used the tree-based maximum-pseudolikelihood technique implemented in PhyloNet v.3.8.2 (refs. 99,100). The set of 999 SCN gene trees was used as the input, the same as that in ASTRAL. Twenty-seven parallel network searches were carried out by allowing zero to five reticulation events. The number of runs of the search (-x) was set as 50. To identify the potential influence of edges with low supporting rate, we ran it three times with no bootstrap threshold, with a threshold of 50 (-b 50) and with a threshold of 70 (-b 70). All other parameters remained as default settings. To estimate the best network among models with different numbers of reticulation events, we performed model selection using the Akaike information criterion and the BIC.

Genome-wide heterozygosity estimation

We characterized the GWH for all baobab species in theta. The theta of each chromosome was calculated in MIRho (v.2.9)¹⁰¹ on the basis of the align_SELF dataset of each diploid species.

Demographic inference and ROH estimation

We applied the pairwise sequentially Markovian coalescent (PSMC)²⁸ method to explore the demographic history of all diploid baobab species. On the basis of the align_SELF dataset for each species, PSMC

Article

(v.0.6.5-r67) analyses were conducted with parameters $\tau = N_{25} - t_{15} - r_5 - p \cdot 4 + 25 \cdot 2 + 4 + 6$. A generation time (g) of 13 years and a substitution rate (μ) of 2×10^{-8} were used to estimate historical changes of effective population size over time. The consistency of the demographic predictions was tested by performing 100 bootstrap replicates. ROH segments were characterized by genome regions with a low best-fit K value (less than or equal to 2) following the method suggested in refs. 25,102.

Coupled logistic modelling

We demonstrate a solution to a pair of coupled logistic equations to simulate population size changes of two species with ecological competition²⁹ using the R package deSolve (v.1.35):

$$\begin{aligned}dX &= a \times X \times (1 - (X + Y)/m) \\dY &= b \times Y \times (1 - (X + Y)/n)\end{aligned}$$

where X and Y are the population size of either species 1 or 2; a and b are the constant growth rate of population size of each species in the absence of interactions with other species; and m and n are the equilibrium population size of each species.

Three stages were divided in the simulation: at the first and third stages, two species were parameterized with the same settings to simulate population growth; a perturbation was simulated at the second stage in both species, with the perturbation to species 1 delayed for 5×10^5 unit of model time.

Species distribution data

The present distribution data (in both longitude and latitude) for the six Malagasy species were obtained from publically available sources, including the Global Biodiversity Information Facility (<https://www.gbif.org/>; 28 November 2022, <https://doi.org/10.15468/dl.ajuz23>), as well as our own field surveys made in 2018 (ref. 103). Repetitious and erroneous records were removed from the integrated dataset giving a total of 742 occurrence points (90 for *A. grandidieri*, 49 for *A. suarezensis*, 87 for *A. madagascariensis*, 41 for *A. perrieri*, 256 for *A. rubrostipa* and 219 for *A. za*) which were kept for the subsequential analyses.

Environmental variables

We extracted the present- and palaeo-climate data for 19 bioclimatic variables from the WorldClim database v.2.0 (ref. 34). The last interglacial (LIG; about 120,000–140,000 years before present) and the LGM (about 21,000 years ago) were defined as the palaeo-climate in this study. The elevation data were also collected from the WorldClim database. All variables were downloaded at a spatial resolution of 2.5 min (about 5 km resolution at the equator). The LIG variables were downloaded at a resolution of 30 arcsec and resampled in ArcGIS 10.2 at a spatial resolution of 2.5 min. To circumvent the effect of highly correlated variables, a Pearson correlation analysis (Pearson $r < 0.8$)¹⁰⁴ for all 19 bioclimatic variables was conducted and eight bioclimatic variables with the lowest correlations were selected for subsequent analysis, these being: annual mean temperature (Bio 1), mean diurnal range (Bio 2), isothermality (Bio 3), temperature seasonality (Bio 4), annual precipitation (Bio 12), precipitation of the wettest month (Bio 13) and precipitation of the coldest quarter (Bio 19). More details of the variables are given in Supplementary Table 9.

Species distribution modelling

The MaxEnt 3.4.1 program (<http://biodiversityinformatics.amnh.org>)³⁷ was used to model the present- and palaeo-distribution of the six Malagasy baobab species. Auto feature function was used for the modelling, 75% of the point occurrences were randomly selected for the model training and cross-validation. Meanwhile, 25% of the data was set aside for model testing and independent validation when the test sample was chosen at random. Cross-validation was used as the replicated run type and other parameters were set to default values.

We computed the area under the curve in test data¹⁰⁵ to validate the specificity and sensitivity of our model in this study.

Palaeo-distribution prediction by ecological niche consistency

The ecological amplitude refers to the tolerance range of a plant to environmental factors, that is, the range between the highest and lowest values tolerated for each environmental factor documented. The ecological amplitude threshold of a species under each environmental factor was extracted through ArcGIS on the basis of its present distribution. Subsequently, the spatial pattern of the ecological amplitude threshold for each environment factor in Madagascar at different time periods was determined spatially by ArcGIS. Correspondingly, the eight environmental factors were superimposed on the layers and the areas with the most overlapping layers were considered to correspond to the potential habitat of a species.

The analysis of niche overlap

The statistical tests for niche equivalency and similarity were computed from the density in environmental niche space¹⁰⁶. The niche equivalency can assess whether the niches of two species being compared are equal (show constant overlap), moderately similar (show some overlap) or significantly different (show no overlaps) when occurrences are randomly shuffled across the ranges. We performed the niche equivalence test by comparing the niche overlap (Schoener's D) values for sets of two species under the overlap of a null distribution. The null distribution was created by extracting the climatic variables from a randomly selected set of coordinates across the background region of the study area. Correspondingly, Schoener's D varied from 0 (no overlap) to 1 (complete overlap). The comparative computing was replicated 1,000 times¹⁰⁷. The niche similarity test was performed to assess the statistical significance of a measured niche difference against niches of the null model taken randomly in a given background area. These analyses were carried out in R software with the package ecospat (v.3.4)¹⁰⁸.

The geological history of Madagascar

The GMSL data are mainly from results obtained from the Pacific benthic foraminiferal $\delta^{18}\text{O}$ and Mg/Ca records⁴¹. The RSL simulation we have adopted is taken from ref. 40. A total of 30 sample sites were randomly selected along the west coast of Madagascar for the assessment. The historical area of island and palaeo-coastline under GMSL and RSL were reconstructed by the Global Elevation Layer (<https://www.gmrt.org/>) using ArcGIS, respectively. The topography map of Madagascar at 5–15 Ma was referenced from ref. 49.

The simulation of potential distribution areas of baobabs (by MaxEnt) in the LIG (Extended Data Fig. 7) was applied when reconstructing the suitable areas under different sea levels. Considering that elevation is one of the main factors contributing to the distribution of almost all Malagasy baobabs (Extended Data Fig. 6), we adjusted the easternmost boundary of baobab distribution according to sea-level changes whereas the westernmost boundary coincided with sea level. The cumulative uplift history and palaeogeographic events are introduced according to the results of refs. 31,49.

Ethics and inclusion statement

This study was performed with full participation of universities and institutes in Africa. The local researchers were involved in research design, sample collection and relevant data analyses. All participants who met the criteria have been included as co-authors, whereas those who made a partial contribution have all been included in the Acknowledgements. The sampling of endangered species (IUCN) in this study was carried out under the permission of the Malagasy government. To protect the endangered species, only fallen fruits were collected and seeds were further germinated to obtain leaves for further genomic studies. Local guides received fees for their services on sample collection and field surveys. Four Madagascar students were supported by

this research programme as PhD or Master candidates at the University of Chinese Academy of Science.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The raw sequencing data reported in this paper are publicly available and have been deposited under accession number CRA015211 in the Genome Sequence Archive¹⁰⁹ of the National Genomics Data Center¹¹⁰, China National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences (GSA: CRA015211), which are publicly accessible at <https://ngdc.cnbc.ac.cn/gsa>. Data of the assembly and annotation of eight baobab species are deposited in figshare at <https://doi.org/10.6084/m9.figshare.25422502.v2> (ref. 111). The distribution data of baobabs used in our analyses are freely available in GBIF (<https://doi.org/10.15468/dl.ajuz23>).

51. Xin, H. P. et al. A genome for *Cissus* illustrates features underlying the evolutionary success in dry savannas. *Hort. Res.* **9**, uhac208 (2022).
52. Chen, Y. et al. SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *GigaScience* **7**, gix120 (2018).
53. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers. *Bioinformatics* **27**, 764–770 (2011).
54. Vurture, G. W. et al. GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* **33**, 2202–2204 (2017).
55. Ranallo-Benavidez, T. R., Jaron, K. S. & Schatz, M. C. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat. Commun.* **11**, 1432 (2020).
56. Chin, C. S. et al. Phased diploid genome assembly with single-molecule real-time sequencing. *Nat. Methods* **13**, 1050–1054 (2016).
57. Walker, B. J. et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **9**, e112963 (2014).
58. Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* **18**, 170–175 (2021).
59. Durand, N. C. et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98 (2016).
60. Dudchenko, O. et al. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).
61. Bao, W., Kojima, K. K. & Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA* **6**, 11 (2015).
62. Flynn, J. M. et al. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl Acad. Sci. USA* **117**, 9451–9457 (2020).
63. Xu, Z. & Wang, H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* **35**, W265–W268 (2007).
64. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999).
65. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
66. Kovaka, S. et al. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol.* **20**, 278 (2019).
67. Cheng, C. Y. et al. Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. *Plant J.* **89**, 789–804 (2017).
68. Du, X. et al. Resequencing of 243 diploid cotton accessions based on an updated A genome identifies the genetic basis of key agronomic traits. *Nat. Genet.* **50**, 796–802 (2018).
69. Teh, B. T. et al. The draft genome of tropical fruit durian (*Durio zibethinus*). *Nat. Genet.* **49**, 1633–1641 (2017).
70. Argout, X. et al. The cacao Criollo genome v2. 0: an improved version of the genome for genetic and functional genomic studies. *BMC Genomics* **18**, 730 (2017).
71. Udall, J. A. et al. De novo genome sequence assemblies of *Gossypium raimondii* and *Gossypium turneri*. *Genes Genomics Genet.* **9**, 3079–3085 (2019).
72. Keilwagen, J. et al. Using intron position conservation for homology-based gene prediction. *Nucleic Acids Res.* **44**, e89–e89 (2016).
73. Stanke, M., Schöffmann, O., Morgenstern, B. & Waack, S. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinform.* **7**, 62 (2006).
74. Korf, I. Gene finding in novel genomes. *BMC Bioinform.* **5**, 59 (2004).
75. Campbell, M. S., Holt, C., Moore, B. & Yandell, M. Genome annotation and curation using MAKER and MAKER-P. *Curr. Protoc. Bioinform.* **48**, 4–11 (2014).
76. Gremme, G., Steinbiss, S. & Kurtz, S. GenomeTools: a comprehensive software library for efficient processing of structured genome annotations. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **10**, 645–656 (2013).
77. Zhang, R. G. et al. TEsor: an accurate and fast method to classify LTR-retrotransposons in plant genomes. *Hort. Res.* **9**, uhac017 (2022).
78. Neumann, P., Novak, P., Hostakova, N. & Macas, J. Systematic survey of plant LTR-retrotransposons elucidates phylogenetic relationships of their polyprotein domains and provides a reference for element classification. *Mobile DNA* **10**, 1 (2019).
79. Ou, S. & Jiang, N. LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiol.* **176**, 1410–1422 (2018).
80. Tang, H. et al. Synteny and collinearity in plant genomes. *Science* **320**, 486–488 (2008).
81. Hasegawa, M., Kishino, H. & Yano, T. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **22**, 160–174 (1985).
82. Huang, S. et al. The genome of the cucumber, *Cucumis sativus* L. *Nat. Genet.* **41**, 1275–1281 (2009).
83. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
84. Li, L., Stoeckert, C. J. Jr. & Roos, D. S. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13**, 2178–2189 (2003).
85. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
86. Huelsenbeck, J. P. & Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754–755 (2001).
87. Hernández-Gutiérrez, R. & Magallón, S. The timing of Malvales evolution: incorporating its extensive fossil record to inform about lineage diversification. *Mol. Phylo. Evol.* **140**, 106606 (2019).
88. Mirarab, S. et al. ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics* **30**, i541–i548 (2014).
89. Blanchette, M. et al. Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res.* **14**, 708–715 (2004).
90. Shimodaira, H. & Hasegawa, M. CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* **17**, 1246–1247 (2001).
91. Dierckx, N., Mardulyn, P. & Smits, G. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* **45**, e18 (2017).
92. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
93. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *GigaScience* **10**, giab008 (2021).
94. Broad Institute. Picard Toolkit. *GitHub* <https://broadinstitute.github.io/picard/> (2019).
95. van der Auwera, G. A. From FastQ data to high confidence variant calls: the genome analysis toolkit best practices pipeline. *Curr. Protoc. Bioinform.* **43**, 11.10.1–11.10.33 (2013).
96. Gao, Y. et al. De novo genome assembly of the red silk cotton tree (*Bombax ceiba*). *Gigascience* **7**, giy051 (2018).
97. Suvakov, M., Panda, A., Diesh, C., Holmes, I. & Abyzov, A. CNVpytor: a tool for copy number variation detection and analysis from read depth and allele imbalance in whole-genome sequencing. *GigaScience* **10**, giab074 (2021).
98. Malinsky, M., Matschiner, M. & Svardal, H. Dsuite—fast D-statistics and related admixture evidence from VCF files. *Mol. Ecol. Resour.* **21**, 584–595 (2021).
99. Than, C., Ruths, D. & Nakhleh, L. PhyloNet: a software package for analyzing and reconstructing reticulate evolutionary relationships. *BMC Bioinform.* **9**, 322 (2008).
100. Yu, Y. & Nakhleh, L. A maximum pseudo-likelihood approach for phylogenetic networks. *BMC Genomics* **16**, S10 (2015).
101. Haubold, B., Pfaffelhuber, P. & Lynch, M. mlRho—a program for estimating the population mutation and recombination rates from shotgun-sequenced diploid genomes. *Mol. Ecol.* **19**, 277–284 (2010).
102. Liu, S. et al. Ancient and modern genomes unravel the evolutionary history of the rhinoceros family. *Cell* **184**, 4874–4885 (2021).
103. Wan, J. N. et al. Modeling impacts of climate change on the potential distribution of six endemic baobab species in Madagascar. *Plant Divers.* **43**, 117–124 (2021).
104. Pearson, K. Notes on the history of correlation. *Biometrika* **13**, 25e45 (1920).
105. Vargas-Jaimes, J. et al. Impact of climate and land cover changes on the potential distribution of four endemic salamanders in Mexico. *J. Nat. Conserv.* **64**, 126066 (2021).
106. Broennimann, O. et al. Measuring ecological niche overlap from occurrence and spatial environmental data. *Glob. Ecol. Biogeogr.* **21**, 481–497 (2012).
107. Wang, Y. et al. Plants maintain climate fidelity in the face of dynamic climate change. *Proc. Natl Acad. Sci. USA* **120**, e2201946119 (2023).
108. Di Cola, V. et al. ecospat: an R package to support spatial analyses and modeling of species niches and distributions. *Ecography* **40**, 774–787 (2017).
109. Chen, T. et al. The Genome Sequence Archive family: toward explosive data growth and diverse data types. *Genom. Proteom. Bioinform.* **19**, 578–583 (2021).
110. CNCR-NGDC Members and Partners. Database resources of the National Genomics Data Center, China National Center for Bioinformation in 2023. *Nucleic Acids Res.* **51**, D18–D28 (2023).
111. Wan, J. N. Genome sequencing of baobabs. *figshare* <https://doi.org/10.6084/m9.figshare.25422502.v2> (2024).

Acknowledgements The genome sequencing, assembly and annotation were conducted by BGI Genomics, Shenzhen, China. The baobab genome project was funded by the Scientific Research Program of Sino-Africa Joint Research Center (SAJC202101, SAJC202105), additionally supported by National Natural Science Foundation of China (32100166). We thank X. Yan from SAJOREC for the assistance on application of scientific permissions. We acknowledge J. K. Muchuku from Jomo Kenyatta University of Agriculture and Technology for the field guide, as well as assistance on plant sampling and DNA extraction in Madagascar. We also thank Y.-X. Zhu from Wuhan University, X.-Q. Wang from Institute of Botany, CAS, and J. Yang from Fudan University for their useful comments on the conception of this study. We thank S.-L. Liu from China Agriculture University for his kind suggestions on revising the manuscript. We also thank M. Borreggine from the Department of Earth and Planetary Sciences at Harvard University for providing data and the analysis of RSL change along the west coast of Madagascar.

Author contributions Q.F.W. and T.W. conceived and initiated the study, managing the project of baobab genomes. T.W. and Q.F.W. designed the main scientific objectives and led the manuscript preparation together with J.N.W., S.W.W., A.R.L., I.J.L., J.B.J., H.P.X. and C.D. The collaboration between groups was close in all aspects of the study. J.N.W., S.W.W., A.R.L., I.J.L.

Article

and J.B.J. are joint first authors. J.N.W. and J.B.J. mainly performed the bioinformatic analyses and data processing. Z.Y.W., H.Q.W., C.X.Z. and M.Z.B. worked on the initial sequence assembly and construction of the Hi-C libraries. S.W.W. mostly contributed to the ecological analyses with assistance from G.E.O., R.W.G., G.M., S.L.D. and C.D. The field sampling and DNA extractions were mainly performed by G.E.O. guided by M.R., G.W.H. and H.P.X. Additional valuable comments were addressed by N.W., G.W.H., S.C.C. and X.Y.C. with input from all authors.

Competing interests The authors declare no competing interests.

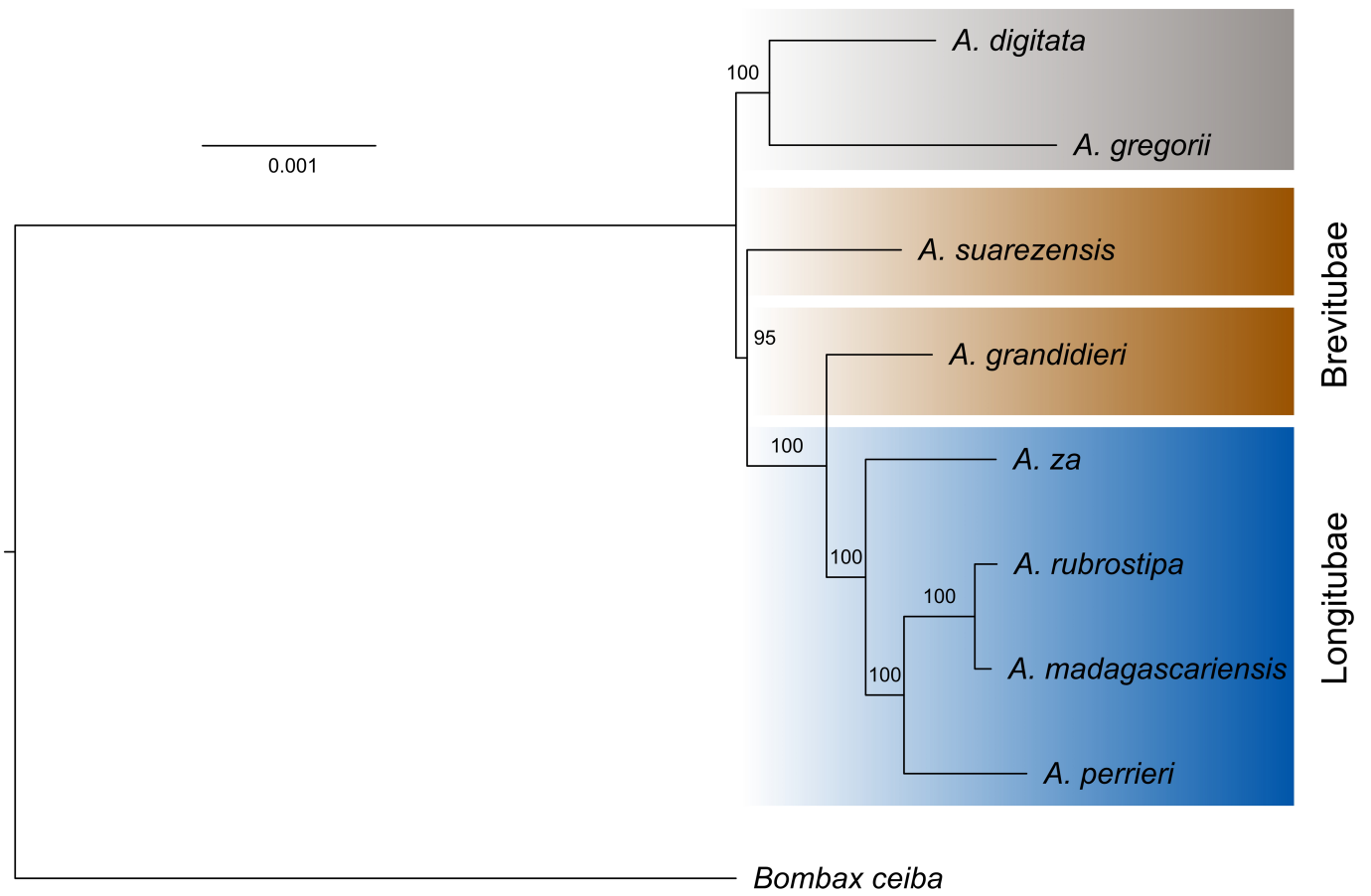
Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41586-024-07447-4>.

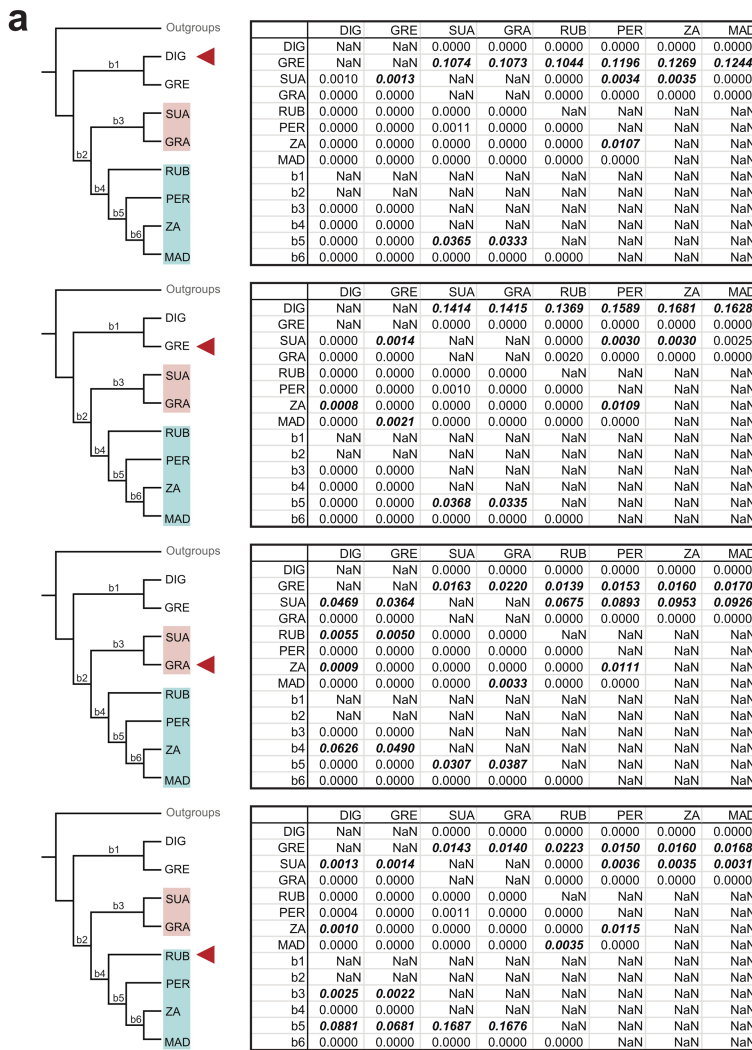
Correspondence and requests for materials should be addressed to Tao Wan or Qing-Feng Wang.

Peer review information *Nature* thanks David Baum, Jerry Mitrovica, Eshchar Mizrachi, Aaron Ragsdale and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

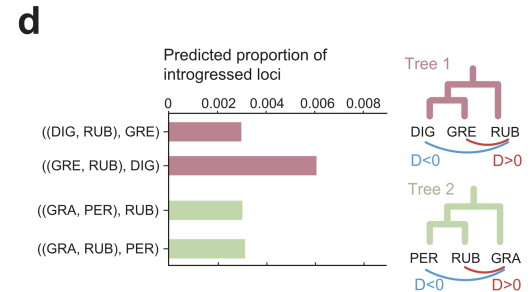
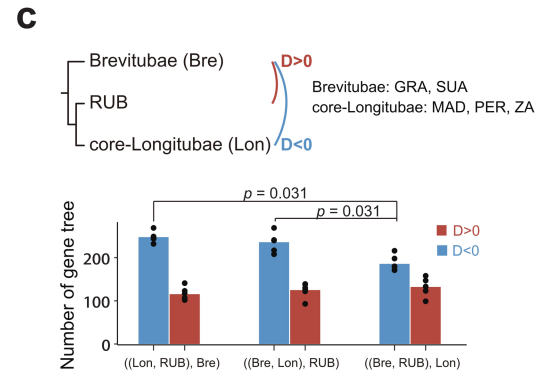
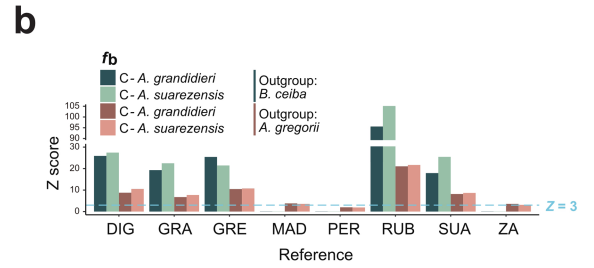
Reprints and permissions information is available at <http://www.nature.com/reprints>.



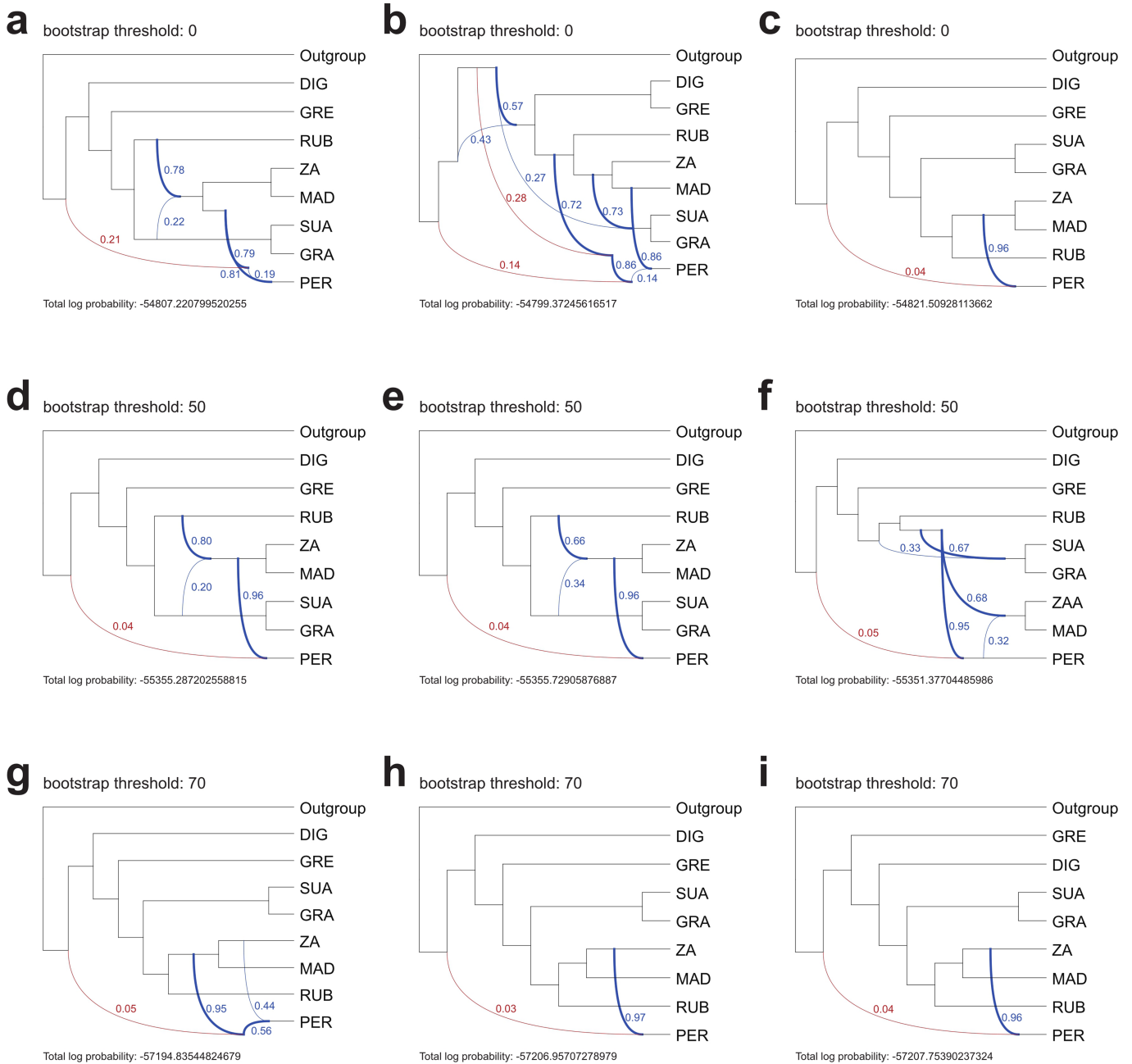
Extended Data Fig. 1 | Plastid genome tree inferred from maximum-likelihood phylogenetic analysis. There was some discordance between nuclear and plastid phylogenies in the placement of *A. suarezensis*, perhaps because it has retained a different ancestral plastome than the other species.



Extended Data Fig. 2 | Test for introgression using different reference genomes. **a**, The inferred D -statistics were calculated and f -branch scores are shown between the species branches or internal tree branches (branch b , ordered b1-6) and baobab species (P3, horizontal row) using *Bombax ceiba* as the outgroup. The reference genome in each case is indicated by a red arrow. '0' indicates a zero score returned in the analysis while significant cells with a Z-score > 3 are highlighted in italics. 'NaN' indicates an unperformable test.



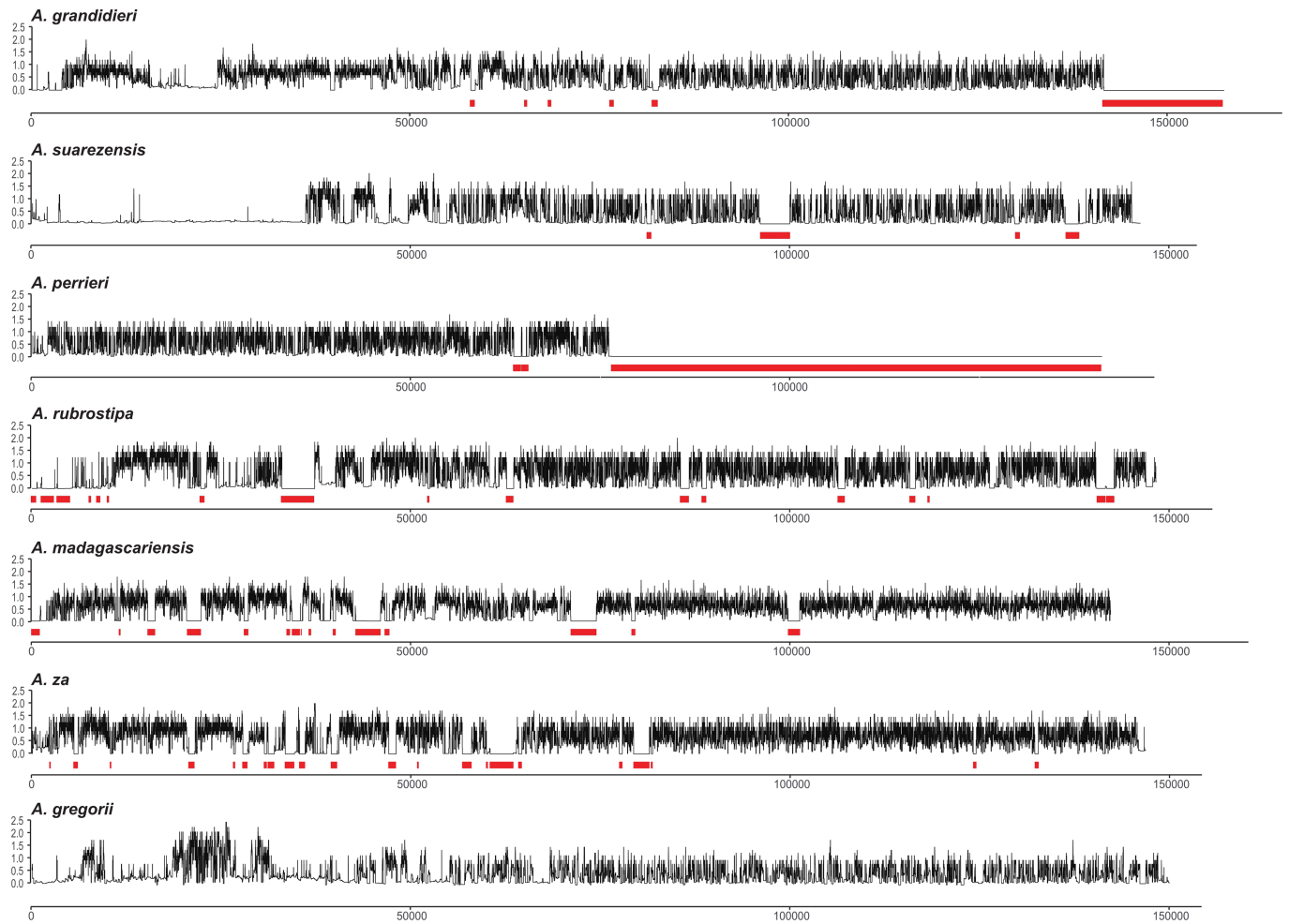
b, The Z-score statistics of f_b using different reference assembly genomes and outgroups (Z-scores > 3 indicate f -branch scores that were significantly larger than zero). **c**, Characterization of different SCN gene topologies from the genomic regions showing introgression signals. Two-sided Wilcoxon-test for paired samples ($n = 6$). **d**, The predicted proportion of introgressed loci that generate the triplet species topologies.



Extended Data Fig. 3 | The reticulation history of baobabs inferred from PhyloNet. The best-supported networks (from left to right) were derived from analyses of 999 concatenated SCN gene trees using PhyloNet in three replicated runs with either (a-c) no bootstrap threshold, or with 50 (d-f) or 70

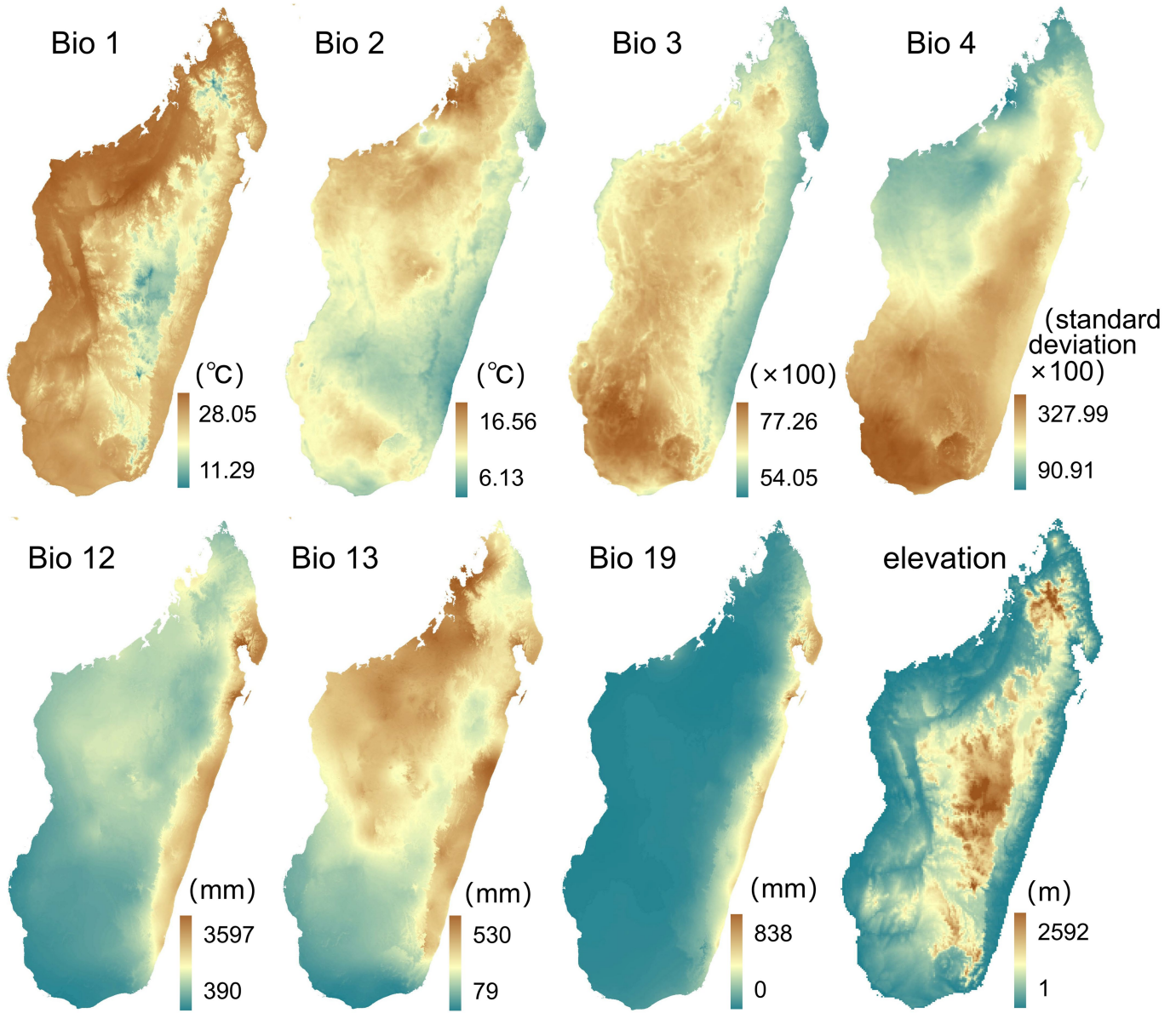
(g-i) bootstrap thresholds. Reticulation edges are coloured in blue, except for the reticulation edge involving *A. perrieri* which is in red. The inheritance probabilities are indicated next to the reticulation edges.

Article



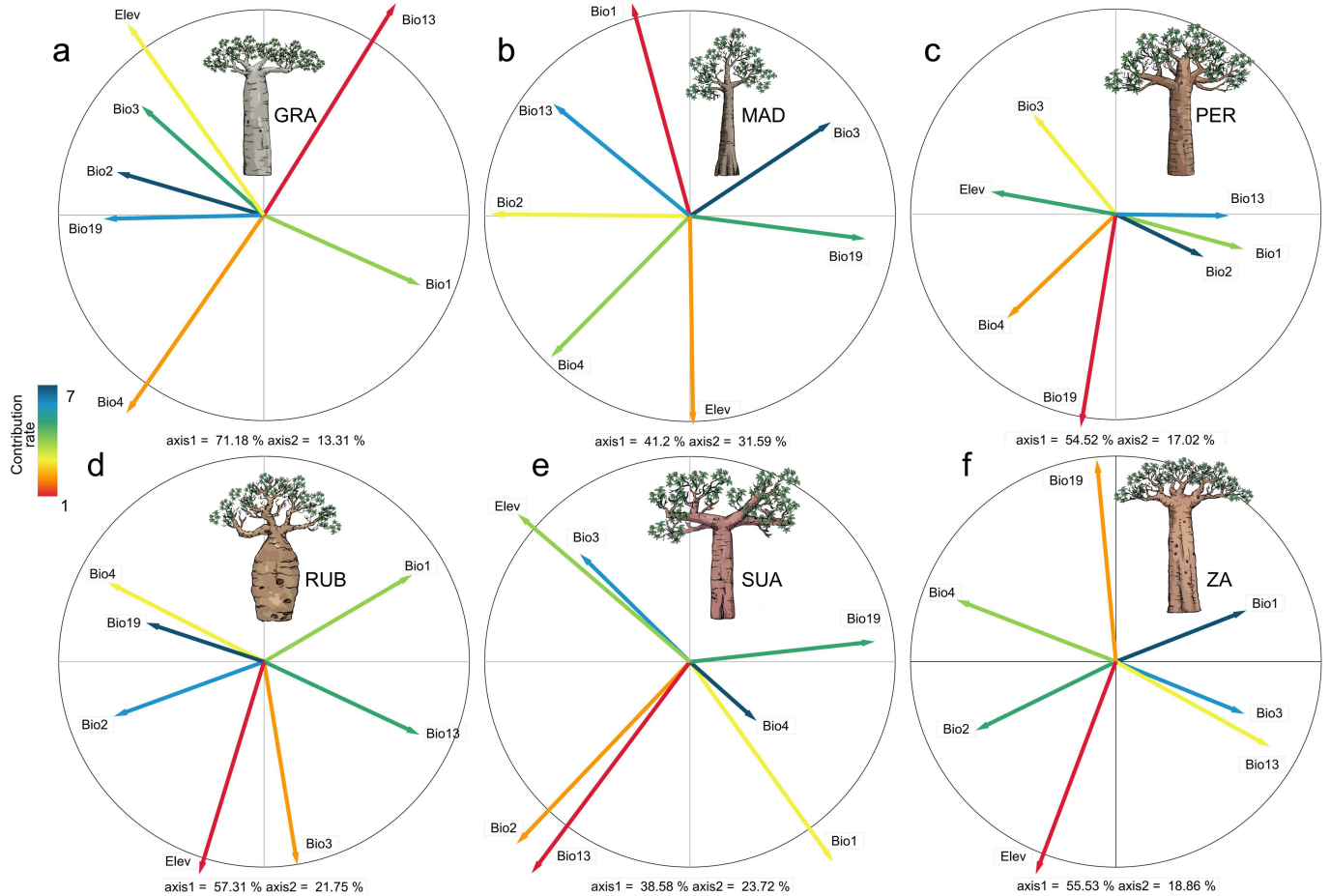
Extended Data Fig. 4 | Visualized runs of homozygosity (ROH) along orthologous chromosome 15 in each baobab species. The y-axis shows the time to the most recent common ancestor (TMRC) and the x-axis shows

positions along the chromosome. Red bars below the black signal indicate regions that are inferred by the pairwise sequentially Markovian coalescent (PSMC) model to have coalesced recently and be ROH.



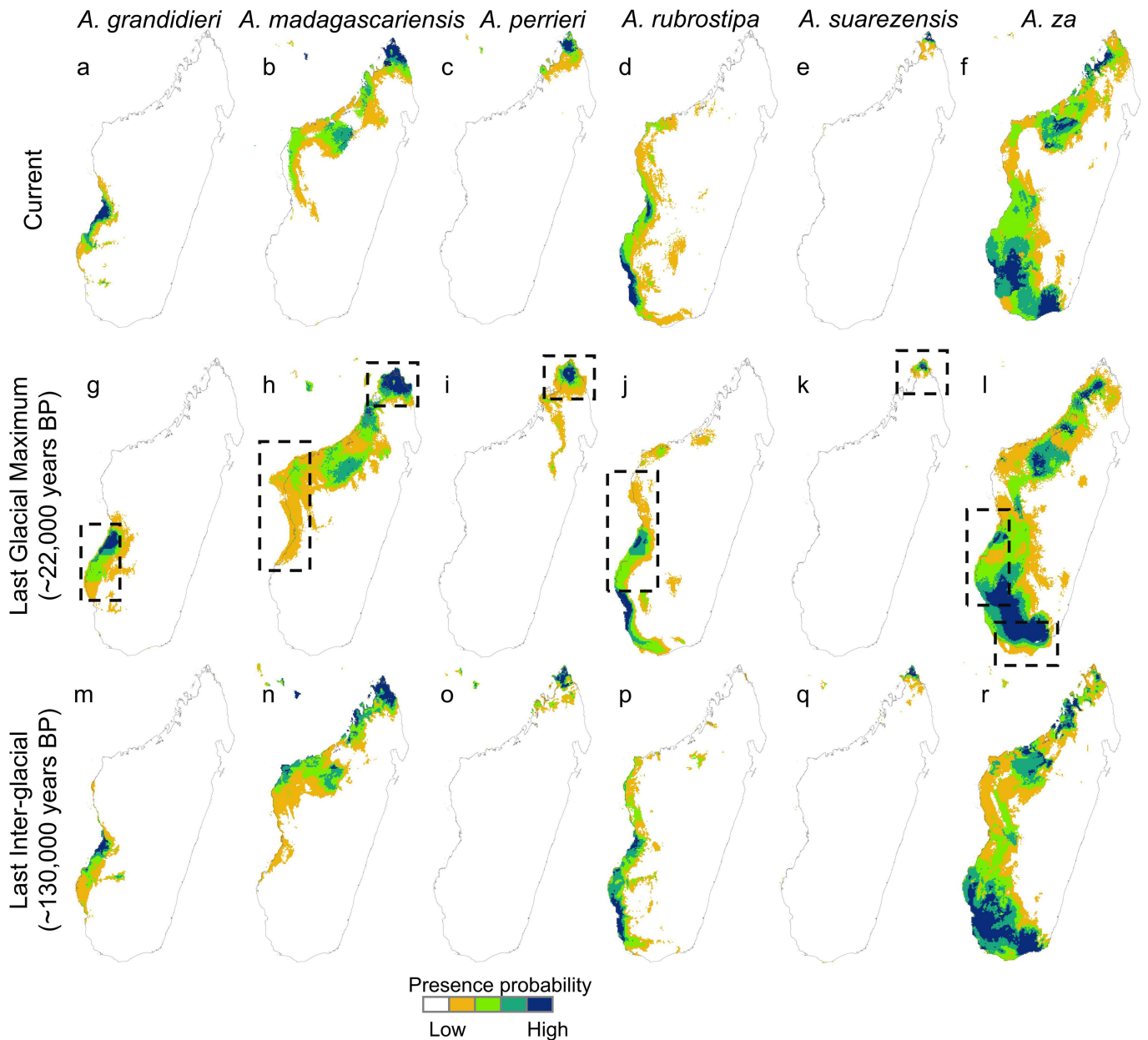
Extended Data Fig. 5 | Patterns of spatial heterogeneity of environmental factors in Madagascar. Eight of 19 bioclimatic variables were filtered using Pearson correlation analysis (Pearson $r < 0.8$). The differences in climates were visualized in ArcGIS (version 10.2). Bio1 = Annual Mean Temperature (°C), Bio2 = Mean Diurnal Range (mean of monthly temperature ranges (monthly

max temp – monthly min temp) (°C), Bio3 = Isothermality (Bio2/Bio7) (×100), Bio4 = Temperature Seasonality (standard deviation ×100), Bio12 = Annual Precipitation (mm), Bio13 = Precipitation of Wettest Month (mm), Bio19 = Precipitation of Coldest Quarter (mm).



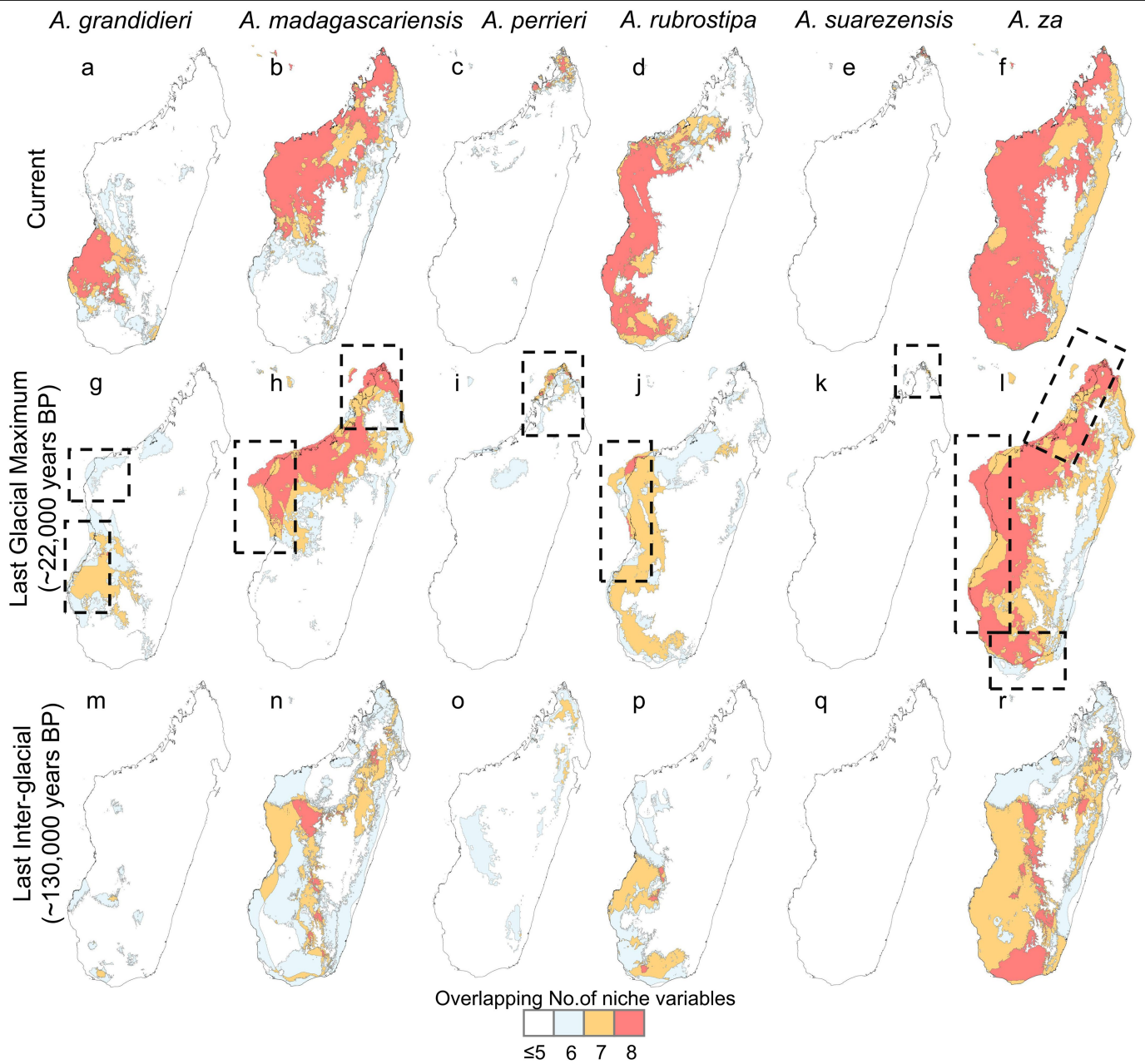
Extended Data Fig. 6 | Principal component analysis of the main factors influencing the distribution of baobabs. For each species (species acronyms are given in Fig. 1a), the figures (a-f) show in the first two principal components (PC1 vs PC2) the relative contribution (line colour) of the bioclim variables. Bio1 = Annual Mean Temperature (°C); Bio2 = Mean Diurnal Range (mean of monthly temperature ranges (monthly max temp – monthly min temp) (°C); Bio3 = Isothermality (Bio2/Bio7) (×100); Bio4 = Temperature Seasonality (standard deviation ×100); Bio5 = Max Temperature of Warmest Month (°C); Bio6 = Min Temperature of Coldest Month (°C); Bio7 = Temperature Annual Range (BIO5-BIO6) (°C); Bio8 = Mean Temperature of Wettest Quarter (°C); Bio9 = Mean Temperature of Driest Quarter (°C);

of Warmest Quarter (°C); Bio11 = Mean Temperature of Coldest Quarter (°C); Bio12 = Annual Precipitation (mm); Bio13 = Precipitation of Wettest Month (mm); Bio14 = Precipitation of Driest Month (mm); Bio15 = Precipitation Seasonality (Coefficient of Variation) (mm); Bio16 = Precipitation of Wettest Quarter (mm); Bio17 = Precipitation of Driest Quarter (mm); Bio18 = Precipitation of Warmest Quarter (mm); Bio19 = Precipitation of Coldest Quarter (mm). The results demonstrated that the spatial heterogeneity of altitude, temperature and precipitation act as the main environmental factors influencing the current distribution of Malagasy species, with *A. madagascariensis* and *A. rubrostipa* mainly influenced by temperature and altitude, while the remaining species were mainly affected by precipitation and altitude.



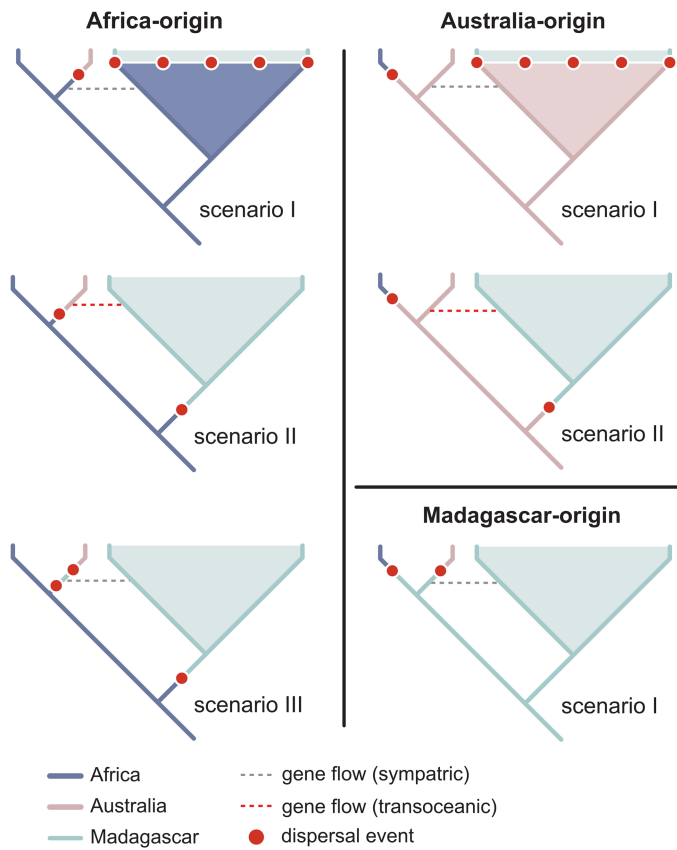
Extended Data Fig. 7 | The potential distribution of six Malagasy baobab species in two past epochs modeled using MaxEnt software based on their current occurrence. Coloured patches reveal predicted current (a-f) and past (g-r) occurrences of suitable habitats for each species. The dashed boxes in g-l show probable range expansions of all species about 22,000 years ago, during the last glacial maximum (LGM). During that period, *A. grandidieri*,

A. madagascariensis and *A. rubrostipa* probably had more extensive suitable habitat areas on the western coast of Madagascar, a habitat that has now largely disappeared (g, h, j). It is also likely that *A. suarezensis*, *A. perrieri* and *A. madagascariensis* had more suitable habitats in the north of Madagascar during the LGM period (h, j, k). In contrast, *A. za* is likely to have had more suitable habitats on the south of the island at that time (l).



Extended Data Fig. 8 | Potential suitable habitats for Malagasy baobab species and their possible recent dispersal trends. Predicted current and past distributions of suitable habitats for six species, derived from the spatial overlay of ecological amplitude layers of each species under each environmental factor (niche variables, spatial superimposition modeling). A higher number of

parameter overlaps represents a higher probability of suitable habitat. The results show similar patterns to the paleo-distributions shown in Extended Data Fig. 7, except that this approach predicts more current suitable habitats for *A. za* on the northwest side of the island.



Extended Data Fig. 9 | Different scenarios explaining three hypotheses of geographic origin for baobabs. The branch colors denote the geographic lineages of the species under each scenario with the occurrence of predicted, independent dispersal events (red dots, multiple dispersals are represented by five red dots, regardless of the exact number of times).

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

No software used for data collection.

Data analysis

K-mer analysis on genome size estimation: SOAPnuke (v1.5.6); Jellyfish (v2.3.0); Genomescope1.0; Genomescope 2.0; Smudgeplot (v0.2.5).
 Genome assembly and contigs anchoring: Falcon (v0.5); Pilon (v1.23); Hifisam (v0.15.1); bwa (v0.7.12); Juicer pipeline (v1.5); 3d-dna (version 180922); JuiceBox (v1.9.0); RepBase (v21.12).
 Genome annotation and assessment: RepBase (v21.12); RepeatModeler (v2.0); LTR_Finder (v1.07); REPEATMASKER (v3.3.0); Tandem Repeats Finder (v4.09); hista2 (v2.1.0); stringtie (v2.1.6); GeMoMa (v1.6.4); AUGUSTUS (v3.2.3); SNAP (version 2006-07-28); Maker-P (v2.31.11).
 Historical activity of LTR retrotransposons: LTR_Finder (v1.0.7); Genome Tools (v1.5.8); LTR_retriever (v1.9); TESorter (v1.3.0).
 Syntenic analysis and whole genome duplication detection: JCVI (v1.1.11); MCScanX-170908; PAML (v4.5).
 Orthologous gene set construction: BLASTP (2.9.0+); OrthoMCL (version 1.4).
 Phylogenetic inference and molecular dating: MUSCLE (v3.8.31); MrBayes (v3.1.2); PAML (v4.5); RAxML-8.2.11; ASTRAL-5.6.1; LASTZ 1.04; MULTIZ (v11.2); CONSEL (v1.20); NOVOPlasty (v4.3.1).
 Variant calling and filtering: bwa (v0.7.17); samtools (v1.9); Picard (v2.21.6); GATK (v4.1.2.0).
 Identification of copy number variations: CNVpytor (v1.2.1); R package "ape" (version 5.6-2).
 Gene flow: Dsuite (version 0.5 r47); genomics_general-master2 (v0.4); QuIBL (<https://github.com/miriammiyagi/QuIBL>); PhyloNet (v3.8.2).
 Genome wide heterozygosity estimation: MIRho (v2.9).
 Demographic inference and ROH estimation: PSMC (version 0.6.5-r67).
 Coupled logistic modeling: R package "deSolve" (version 1.35).
 Species distribution data: Global Biodiversity Information Facility (<https://www.gbif.org/>; November 28, 2022).
 Environmental variables: WorldClim database version 2.0 (www.worldclim.org); ArcGIS 10.2.
 Modeling procedure: MaxEnt (v3.4.1).

Paleo-distribution prediction by ecological niche consistency: ArcGIS 10.2.
 The analysis of niche overlap: R package "ecospat" (version 3.4).
 The geological history of Madagascar: Global Elevation Layer (<https://www.gmrt.org/>); ArcGIS 10.2.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The raw and whole genome sequence data reported in this paper have been deposited in the Genome Sequence Archive¹²³ in National Genomics Data Center¹²⁴, China National Center for Bioinformation / Beijing Institute of Genomics, Chinese Academy of Sciences (GSA: CRA015211) that are publicly accessible at <https://ngdc.cncb.ac.cn/gsa>. The distribution data of baobabs used in our analyses are freely available in GBIF (www.gbif.org) and the doi representing the download is: <https://doi.org/10.15468/dl.ajuz23>.

The data sources of outgroup species: Arabidopsis thaliana (https://phytozome-next.jgi.doe.gov/info/Athaliana_TAIR10); Bombax ceiba (<https://gigadb.org/dataset/100445>); Carica papaya (https://phytozome-next.jgi.doe.gov/info/Cpapaya_ASGBPv0_4); Durio zibethinus (https://www.ncbi.nlm.nih.gov/assembly/GCF_002303985.1); Gossypium arboreum (https://www.cottongen.org/species/Gossypium_arboreum/A2_WHU); Gossypium raimondii (https://www.cottongen.org/species/Gossypium_raimondii/NSF-D5); Theobroma cacao (<https://cocoa-genome-hub.southgreen.fr/download>); Populus trichocarpa (https://phytozome-next.jgi.doe.gov/info/Ptrichocarpa_v3_1).

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical method was used for determining sample size. One individual of each species was sequenced.
Data exclusions	The tetraploid species <i>A. digitata</i> was only excluded from the genome-wide estimation of theta and ROH, and the analysis of PSMC, due to the incommensurability of these estimations for species have a higher polyploid level.
Replication	No such experiments.
Randomization	The sampling of individual for sequencing was randomly conducted without any specific selection. No group allocation was needed.
Blinding	Blinding was not relevant to our study as comparison were performed by computer software not influenced by investigator.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

- | | |
|-------------------------------------|--|
| n/a | Involvement in the study |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Antibodies |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology and archaeology |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Animals and other organisms |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Dual use research of concern |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Plants |

Methods

- | | |
|-------------------------------------|---|
| n/a | Involvement in the study |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |

Dual use research of concern

Policy information about [dual use research of concern](#)

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

- | | |
|-------------------------------------|---|
| No | Yes |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Public health |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> National security |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Crops and/or livestock |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Ecosystems |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Any other significant area |

Experiments of concern

Does the work involve any of these experiments of concern:

- | | |
|-------------------------------------|--|
| No | Yes |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Demonstrate how to render a vaccine ineffective |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Confer resistance to therapeutically useful antibiotics or antiviral agents |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Enhance the virulence of a pathogen or render a nonpathogen virulent |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Increase transmissibility of a pathogen |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Alter the host range of a pathogen |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Enable evasion of diagnostic/detection modalities |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Enable the weaponization of a biological agent or toxin |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Any other potentially harmful combination of experiments and agents |

Plants

Seed stocks	Young leaves of <i>A. digitata</i> and <i>A. gregorii</i> were collected from trees in wild populations. The ripe fruits of Malagasy baobabs were collected from natural populations and seeds were further germinated for fresh young leaves at a local nursery garden in Antananarivo, Madagascar. <i>A. digitata</i> Lon:-17.19111, Lat: 14.74497, Senegal; <i>A. gregorii</i> Lon: 132.26266, Lat: -14.45994, Australia; <i>A. grandis</i> Lon: 41.19654, Lat: -20.255134, Madagascar, 2021-11-05; <i>A. madagascariensis</i> Lon: 49.11928, Lat: -12.96816, Madagascar, 2021-11-05; <i>A. perrieri</i> Lon: 49.171495, Lat: -12.490713, Madagascar, 2021-11-05; <i>A. rubrostipa</i> Lon: 43.6290503, Lat: -23.1193637, Madagascar, 2021-11-16; <i>A. suarezensis</i> Lon: 49.16806, Lat: -12.26667, Madagascar, 2021-11-16; <i>A. za</i> Lon: 44.534665, Lat: -20.335426, 2021-11-05.
Novel plant genotypes	None
Authentication	No such experiments.