

Genetic diversity of Malagasy baobabs: implications for conservation

Nisa KARIMI,
Corrinne E. GROVER,
Joseph P. GALLAGHER,
Justin L. CONOVER,
Emma R. MILLER,
Jonathan F. WENDEL &
David A. BAUM

DIRECTEUR DE LA PUBLICATION / *PUBLICATION DIRECTOR*: Bruno David
Président du Muséum national d'Histoire naturelle

RÉDACTEUR EN CHEF / *EDITOR-IN-CHIEF*: Thierry Deroin

RÉDACTEURS / *EDITORS*: Porter P. Lowry II; Zachary S. Rogers

ASSISTANT DE RÉDACTION / *ASSISTANT EDITOR*: Emmanuel Côté (adanson@mnhn.fr)

MISE EN PAGE / *PAGE LAYOUT*: Emmanuel Côté

COMITÉ SCIENTIFIQUE / *SCIENTIFIC BOARD*:

P. Baas (Nationaal Herbarium Nederland, Wageningen)
F. Blasco (CNRS, Toulouse)
M. W. Callmender (Conservatoire et Jardin botaniques de la Ville de Genève)
J. A. Doyle (University of California, Davis)
P. K. Endress (Institute of Systematic Botany, Zürich)
P. Feldmann (Cirad, Montpellier)
L. Gautier (Conservatoire et Jardins botaniques de la Ville de Genève)
F. Ghahremaninejad (Kharazmi University, Téhéran)
K. Iwatsuki (Museum of Nature and Human Activities, Hyogo)
A. A. Khapugin (Tyumen State University, Russia)
K. Kubitzki (Institut für Allgemeine Botanik, Hamburg)
J.-Y. Lesouef (Conservatoire botanique de Brest)
P. Morat (Muséum national d'Histoire naturelle, Paris)
J. Munzinger (Institut de Recherche pour le Développement, Montpellier)
S. E. Rakotoarisoa (Millenium Seed Bank, Royal Botanic Gardens Kew, Madagascar Conservation Centre, Antananarivo)
É. A. Rakotobe (Centre d'Applications des Recherches pharmaceutiques, Antananarivo)
P. H. Raven (Missouri Botanical Garden, St. Louis)
G. Tohmé (Conseil national de la Recherche scientifique Liban, Beyrouth)
J. G. West (Australian National Herbarium, Canberra)
J. R. Wood (Oxford)

COUVERTURE / *COVER*:

Réalisée à partir des Figures de l'article/*Made from the Figures of the article.*

Adansonia est indexé dans / *Adansonia is indexed in*:

- Science Citation Index Expanded (SciSearch®)
- ISI Alerting Services®
- Current Contents® / Agriculture, Biology, and Environmental Sciences®
- Scopus®

Adansonia est distribué en version électronique par / *Adansonia is distributed electronically by*:

- BioOne® (<http://www.bioone.org>)

Adansonia est une revue en flux continu publiée par les Publications scientifiques du Muséum, Paris
Adansonia is a fast track journal published by the Museum Science Press, Paris

Les Publications scientifiques du Muséum publient aussi / *The Museum Science Press also publish*: *Geodiversitas*, *Zoosystema*, *Anthropozoologica*,
European Journal of Taxonomy, *Naturae*, *Cryptogamie* sous-sections *Algologie*, *Bryologie*, *Mycologie*, *Comptes Rendus Palevol*

Diffusion – Publications scientifiques Muséum national d'Histoire naturelle
CP 41 – 57 rue Cuvier F-75231 Paris cedex 05 (France)
Tél.: 33 (0)1 40 79 48 05 / Fax: 33 (0)1 40 79 38 40
diff.pub@mnhn.fr / <http://sciencepress.mnhn.fr>

© Publications scientifiques du Muséum national d'Histoire naturelle, Paris, 2022
ISSN (imprimé / *print*): 1280-8571/ ISSN (électronique / *electronic*): 1639-4798

Genetic diversity of Malagasy baobabs: implications for conservation

Nisa KARIMI

Department of Botany, University of Wisconsin – Madison, WI, 53706 (United States)
nkarimi@wisc.edu

Corrinne E. GROVER

Department of Ecology, Evolution, and Organismal Biology,
Iowa State University, Ames, IA, 50011 (United States)

Joseph P. GALLAGHER

Department of Ecology, Evolution, and Organismal Biology,
Iowa State University, Ames, IA, 50011 (United States)
and Department of Biology, University of Massachusetts,
611 North Pleasant Street, Amherst, MA 01003 (United States)

Justin L. CONOVER

Emma R. MILLER

Jonathan F. WENDEL

Department of Ecology, Evolution, and Organismal Biology,
Iowa State University, Ames, IA, 50011 (United States)

David A. BAUM

Department of Botany, University of Wisconsin – Madison, WI, 53706 (United States)
and Wisconsin Institute for Discovery, 330 N Orchard St, Madison, WI 53715 (United States)

Submitted on 30 January 2021 | accepted on 11 May 2021 | published on 17 February 2022

Karimi N., Grover C. E., Gallagher J. P., Conover J. L., Miller E. R., Wendel J. F. & Baum D. A. 2022. — Genetic diversity of Malagasy baobabs: implications for conservation. *Adansonia*, sér. 3, 44 (6): 37-47. <https://doi.org/10.5252/adansonia2022v44a6>. <http://adansonia.com/44/6>

ABSTRACT

Assessing the genetic diversity of species and populations is critical for evaluating extinction vulnerability and provides important information for identifying populations of concern and/or those that should be targeted for breeding material. Baobabs (*Adansonia* L.) are botanical icons for conservation, with increasing attention regarding their threatened status from both scientists and non-scientists alike. Baobabs are of particular interest especially in Madagascar, where six of the eight species are endemic, and three are listed as Endangered or Critically Endangered by the International Union for Conservation of Nature (IUCN). Although *A. madagascariensis* Baill., *A. rubrostipa* Jum. & H. Perrier and *A. za* Baill. are more widespread and classified by IUCN as Least Concern, they show regional variation, which may reflect hidden genetic diversity or even the existence of cryptic species. Here we assess the genetic diversity of the Malagasy baobabs to serve as a basis for future conservation and management planning. Our study used a targeted sequence capture approach (hybrid enrichment) to obtain hundreds of low-copy nuclear loci with phased alleles to assess genetic diversity in the six species and their major regional subpopulations. We discuss the implications of proper delineation of species taxonomy for management issues associated with conservation. We hope such genetic information will guide more targeted population genetic assessments and inform conservation and management efforts, including identification of isolated or disjunct populations that may warrant targeted actions.

KEY WORDS

Genetic diversity,
Madagascar,
baobabs,
Adansonia,
conservation.

RÉSUMÉ

Diversité génétique des baobabs de Madagascar: implications pour la conservation.

L'évaluation de la diversité génétique des espèces et des populations est essentielle pour évaluer la vulnérabilité à l'extinction et fournit des informations importantes pour le tri des populations préoccupantes ou de celles qui devraient être ciblées pour le matériel de reproduction. Les baobabs (*Adansonia*) sont des icônes botaniques pour la conservation, avec une attention croissante de leur statut menacé de la part des scientifiques et des non-scientifiques, en particulier à Madagascar, où six des huit espèces de baobabs du monde sont endémiques. Trois de ces espèces endémiques sont classées en danger ou en danger critique d'extinction par International Union for Conservation of Nature (IUCN). Bien que *A. madagascariensis* Baill., *A. rubrostipa* Jum. & H. Perrier et *A. za* Baill. soient plus répandus et classés par IUCN comme Préoccupation mineure, ces deux derniers présentent une variation régionale, qui peut refléter une diversité génétique cachée ou même l'existence d'espèces cryptiques. Nous présentons ici la diversité génétique des baobabs malgaches pour servir de base à la planification future de la conservation et de la gestion. Notre étude a utilisé une approche de capture de séquence ciblée (enrichissement hybride) pour obtenir des centaines de loci nucléaires à faible copie avec des allèles phasés pour évaluer la diversité génétique des six espèces et de leurs principales sous-populations régionales. Nous discutons également des implications d'une bonne délimitation de la taxonomie des espèces pour les problèmes de gestion associés à la conservation. Nous espérons que ces informations génétiques pourront guider des évaluations génétiques des populations plus ciblées et éclairer les efforts de conservation et de gestion, y compris l'identification de populations isolées ou disjointes qui pourraient justifier des actions ciblées.

MOTS CLÉS
Diversité génétique,
Madagascar,
baobabs,
Adansonia,
préservation.

INTRODUCTION

Conservation efforts are informed by a multitude of factors, including demographic information (e.g., species distribution and abundance), dispersal ability and regeneration, extent of habitat threats or fragmentation, and socio-economic/cultural importance. Increasingly, these data are coupled with measures of genetic diversity, which can be used to influence policy and planning. Incorporating genetic diversity is an important criterion in assessing a species vulnerability to extinction (Cook & Sgrò 2017). Nonetheless, there is a need to further expand our understanding of genetic diversity to include many more global conservation targets (Laikre *et al.* 2010; Hoban *et al.* 2020).

From an evolutionary perspective, maintaining genetic diversity is critical for the long-term viability of populations, as high genetic diversity confers adaptive potential in the face environmental change (Schemske *et al.* 1994) and reduces fitness declines due to genetic drift of deleterious alleles (Agrawal & Whitlock 2012). Diversity assessments that couple population genetic approaches with phylogeographic methods are important tools for assessing species or population vulnerability (Millar & Libby 1991). Small population size and/or reduced gene flow through population fragmentation can result in low genetic diversity and/or inbreeding depression (Frankham 1996). For long-lived species with delayed reproductive maturity, the effects of population reduction and strong fragmentation may take longer to become apparent (Nunney 1993; Congdon *et al.* 1993), making it all the more important to evaluate the genetic health of their component populations.

Baobabs (*Adansonia* L.) are long-lived trees with hundreds of documented uses by humans (reviewed by Baum 1996; Wickens 2008). Revered across continental Africa, Australia, and Mada-

gascar, baobab trees are iconic, often being referred to as “mothers of the forest.” Their conservation status has drawn attention from both scientists and non-scientists alike, due to the observable population impacts of over-grazing and habitat destruction, as well as climate change (Wan *et al.* 2020). Six of the eight baobab species are endemic to Madagascar (Fig. 1), three of which are listed as either Endangered or Critically Endangered by the International Union for Conservation of Nature (IUCN), and all six face threats due to human activities, including excessive forest burning or clearance for agriculture. The Malagasy species *A. za* Baill. is currently classified as the least concerning (Letsara *et al.* 2019) based on its extensive distribution; however, recent studies have suggested that *A. za* is not monophyletic, with the southern *A. za* accessions being sister to a larger clade containing *A. madagascariensis* Baill., *A. perrieri* Capuron, and northern *A. za* accessions (Karimi *et al.* 2020). The discovery that northern populations of *A. za* are genetically distinct from southern populations is important from a conservation perspective because northern *A. za* populations are smaller and more isolated and, therefore, more vulnerable to genetic erosion.

To provide an initial assessment of genetic diversity of these charismatic trees and serve as a baseline for more targeted genetic studies, we used a SNP-based approach derived from targeted sequence capture to characterize a geographically representative sample of 4–14 trees per species. While microsatellites markers have historically been the most widely applied molecular markers for assessing genetic diversity across small spatial scales, next-generation sequencing approaches may yield a much larger number of SNPs and have the potential, therefore, to be more informative (Fischer *et al.* 2017). Here we report the first measures of genetic diversity in these iconic Malagasy trees and consider how these patterns of diversity may influence conservation efforts.



FIG. 1. — Flowers of the six species of baobabs endemic to Madagascar: **A**, *Adansonia grandidieri* Baill.; **B**, *A. madagascariensis* Baill.; **C**, *A. perrieri* Capuron; **D**, *A. rubrostipa* Jum. & H. Perrier; **E**, *A. suarezensis* H. Perrier; **F**, *A. za* Baill. Photographs by David A. Baum and Nisa Karimi.

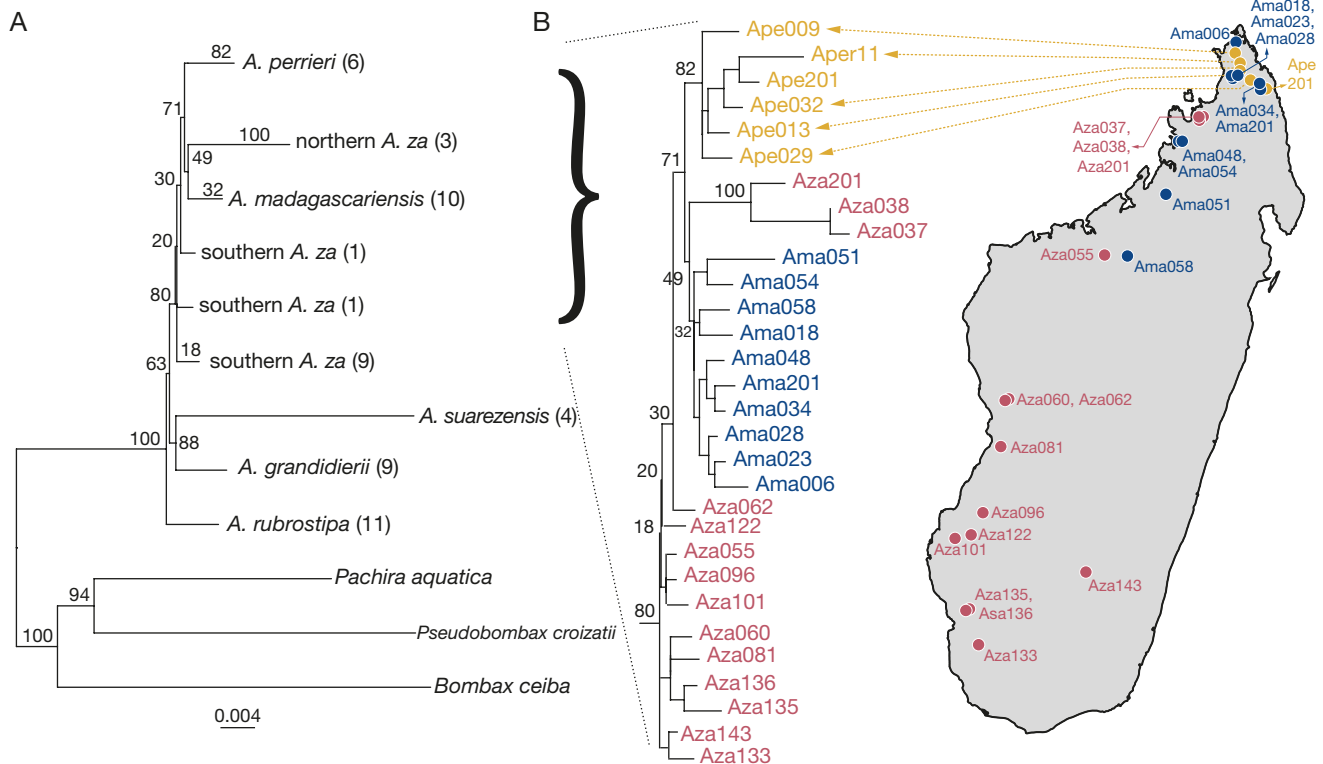


FIG. 2. — SNP-based tree from 2089 biallelic SNPs as inferred by IQ-TREE. Values on the branches represent bootstrap support for the primary clades: **A**, broad scale relationships with monophyletic taxa collapsed and number of samples per clade indicated in parentheses; **B**, detailed phylogeny for the core Longitubae clade (*A. perrieri* Capuron, *A. madagascariensis* Baill., and *A. za* Baill.), with individual accessions mapped to their respective geographic localities in Madagascar (**B**). *Adansonia za* (**Aza**) accessions are listed in red; *A. madagascariensis* (**Ama**), in blue; and *A. perrieri* (**Ape**) accessions in yellow.

METHODS

SAMPLING

We sampled broadly across each species range in Madagascar (Table 1) under permit No. 195/14 granted by Madagascar agency La Conservation de la Biodiversité et du Système des Aires Protégées. Specimens were deposited at the Wisconsin State Herbarium and Parc Botanique et Zoologique de Tsimbazaza (PZBT) Herbarium in Antananarivo, Madagascar. In total, we sampled ten accessions of *A. madagascariensis* Baill., six accessions of *A. perrieri* Capuron, six accessions of *A. grandidierii* Baill., 11 accessions of *A. rubrostipa* Jum. & H. Perrier, four of *A. suarezensis* H. Perrier, and 16 accessions of *A. za* Baill. We also included *Bombax ceiba* L., *Pachira aquatica* Aubl., and *Pseudobombax croizatii* A. Robyns as outgroups for phylogenetic analyses.

Allele phasing and SNP-calling

Bait design, library preparations, and sequence capture are detailed in Karimi *et al.* 2020. Previously generated bait reference sequences (Karimi *et al.* 2020) were mapped to the phylogenetically-related *Bombax ceiba* genome (Gao *et al.* 2018) using GMAP (Wu & Watanabe 2005) version gmap-snap/2019-05-12-zjqshxf from Spack (Gamblin *et al.* 2015) and returning only a single alignment. Output was parsed in Bash and transformed into *B. ceiba* genome coordinates for each target sequence on the *B. ceiba* genome (BED-format)

using a custom R script (prepare.bed.R). Any alignments greater than 10 kb were excluded (to remove artifactual assemblies), and then the BED coordinates were used to extract each target sequence from the *B. ceiba* genome using BEDTools ‘getfasta’ (Quinlan & Hall 2010).

Raw reads from each *Adansonia* and/or outgroup accession were trimmed using Trimmomatic (Bolger *et al.* 2014) version trimmomatic/0.36-lkkrba from Spack (Gamblin *et al.* 2015), and these were individually mapped to the bait-specific *B. ceiba* gene reference. A single alignment containing all alleles for each species was constructed from the mapped reads using bam2consensus (‘-m5 -p 4’) from the BamBam suite (Page *et al.* 2014). These alignments were subsequently processed with a custom script (filter_alignments), which uses PyCo-geant (Knight *et al.* 2007), to first remove alleles with >90% missing data and then remove aligned positions with >30% missing data. The resulting alignment was deduplicated using bioawk and parsed into species specific reference sequences (snps.workflow).

For each accession, trimmed reads were mapped to the newly generated species-specific reference sequences using bwa (Li & Durbin 2009) Spack version 0.7.17-rxh5dw. Single nucleotide polymorphisms (SNPs) were called on each accession individually using the Sentieon Genomics (Kendig *et al.* 2019) software suite (Spack version sentieon-genomics/201808.01-opfuvzr) following the DNaseq guidelines, which includes read deduplication, indel realignment, hap-

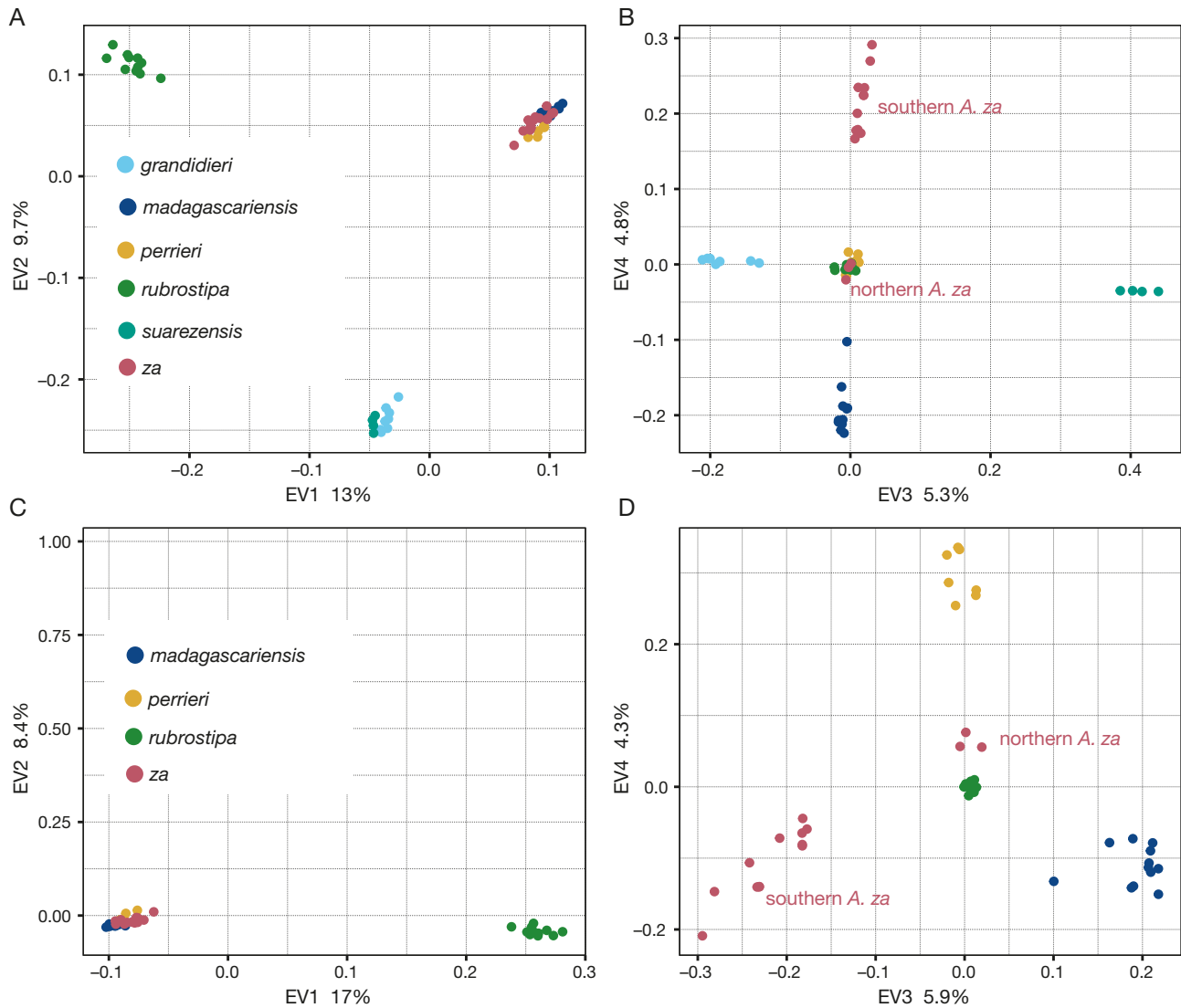


FIG. 3. — Principal components analysis (PCA) of Malagasy *Adansonia* L. colored by species. Panels represent different PCA axes from: **A**, 2089 SNPs for all Malagasy *Adansonia* for axis 1 vs axis 2; **B**, axis 3 vs axis 4; **C**, 1873 SNPs for Longitubae taxa only for axis 1 vs axis 2; and **D**, axis 3 vs axis 4. Northern and southern populations of *A. za* Baill. labeled.

lotyping, and joint genotyping. Sentieon represents an optimization of the GATK pipeline (McKenna *et al.* 2010).

Custom Bash (extract.info.bash) and R (make.variant.table.R) scripts were used to parse the VCF output and calculate the following metrics for each sample: 1) the number of reads that match the species-specific reference; 2) the number of reads that match an alternate SNP; 3) the total number of SNPs for that locus/allele; and 4) the relative number of SNPs per locus/allele length. These metrics were used as a basis to filter alleles that were likely misassembled sequences and/or paralogs: loci where greater than 20% of the reads mapped to either the reference or alternative, or having an overall percentage of variant sites for a given locus greater than 0.2, were dropped. Retained allelic sequences were aligned with MAFFT (Katoh *et al.* 2002) and consensus sequences were created using ‘cons’ from EMBOSS (Rice *et al.* 2000).

BWA was used to map the trimmed reads to the newly-generated, consensus reference sequence. SNPs were called against this new reference using the Sentieon pipeline to generate a single reference used for mapping and joint-genotyping of the samples. Finally, VCFtools (Danecek *et al.* 2011) was used to calculate the average SNP sequencing depth per sample, with only sites with an average depth per site >15 being retained for further analysis. All custom scripts referenced above are available on GitHub at <https://github.com/nkarimi/AdansoniaDiversity>.

Phylogenetic and Genetic Diversity Analyses

A SNP-based phylogeny was inferred using IQ-TREE (Nguyen *et al.* 2015) by estimating nucleotide substitution models with ascertainment bias correction and 100 bootstrap replicates.

Nucleotide diversity (π) was estimated based on a sliding window of 10kb in VCFtools (--window-pi), which, because all genes were limited to < 10kb, will result in a per gene diversity

TABLE 1. — Sampling. Wisconsin State Herbarium (WIS); George Brown Darwin Botanical Gardens, Australia (GBDBG); Missouri Botanical Gardens (MO); University of Wisconsin – Madison, Department of Botany Greenhouse (UWBG).

Taxon	Sample ID	Latitude	Longitude	Source
<i>Bombax ceiba</i> L.	Bce020			Accession #UW10 (UWBG)
<i>Pseudobombax croizatii</i> A.Robyns.	Pcr070			Accession #UW1255 (UWBG); Seed from Puerto Ayacucho in Venezuela, Paul E. Berry (MO)
<i>Pachira aquatica</i> Aubl.	Pacaqu06			B. Alverson (WIS), Finca La Selva, Costa Rica, 1988
<i>Adansonia grandidieri</i> Baill.	Aga001			Accession #97-B002010-1 (GBDBG)
	Aga002			Accession #03-B000192-1 (GBDBG)
	Aga095	–20.75214	44.443544 ●	
	Aga108	–21.79613	43.835135 ●	
	Aga111	–21.79669	43.690202 ●	
	Aga119	–21.86911	43.738777 ●	
<i>Adansonia madagascariensis</i> Baill.	Ama006	–12.27829	49.187668 ●	Northern Madagascar, Karimi-2014-006 (WIS)
	Ama018	–12.91038	49.20081 ●	Northern Madagascar, Karimi-2014-018 (WIS)
	Ama023	–12.95235	49.128891 ●	
	Ama029	–13.00924	49.466315 ●	
	Ama034	–13.07311	49.641809 ●	
	Ama048	–14.18814	48.079818 ●	
	Ama054	–15.20995	47.835857 ●	
	Ama058	–16.39997	47.094832 ●	
	Ama222	–13.16225	49.694467 ●	
<i>Adansonia perrieri</i> Capuron	Ape001			
	Ape009	–12.48802	49.171227 ●	Northern Madagascar, Karimi-2014-09 (WIS)
	Ape013	–12.91268	49.199832 ●	
	Ape029			
	Ape032	–13.16718	49.706729 ●	
	Ape222	–12.82702	49.265967 ●	
<i>Adansonia rubrostipa</i> Jum. & Perr.	Aru001			Southwestern Madagascar, D.A.Baum 313 (MO)
	Aru075	–19.74947	44.583397 ●	
	Aru076	–19.81344	44.586302 ●	
	Aru083	–20.10095	44.550986 ●	
	Aru085	–20.11347	44.543008 ●	
	Aru117	–21.86614	43.66286 ●	
	Aru124	–23.07974	43.606588 ●	
	Aru127	–24.04583	43.753756 ●	Western Madagascar, Karimi-2014-127 (WIS)
	Aru128	–24.04934	43.7571 ●	
<i>Adansonia suarezensis</i> H.Perrier	Asu001			Accession #UW11 (GBDBG); Seed from Northern Madagascar, Baum 320A (WIS)
	Asu003	–12.27837	49.187669 ●	
	Asu007	–12.82588	49.263007 ●	
	Asu012	–13.75594	48.464887 ●	
<i>Adansonia za</i> Baill.	Aza037	–13.75191	48.464763 ●	Northern Madagascar, Karimi-2014-37 (WIS)
	Aza038	–13.75658	48.360349 ●	
	Aza043	–16.38065	46.65465 ●	
	Aza055	–19.15508	44.806799 ●	Northern Madagascar, Karimi-2014-055 (WIS)
	Aza059	–19.15342	44.807328 ●	
	Aza060	–20.072	44.659204 ●	
	Aza079	–20.07275	44.653399 ●	
	Aza081	–21.34696	44.308989 ●	
	Aza096	–23.88507	44.174298 ●	
	Aza101	–21.77381	44.081807 ●	
	Aza132	–23.89124	44.225474 ●	
	Aza133	–23.21223	44.042275 ●	
	Aza135	–23.20194	44.049406 ●	Southern Madagascar, Karimi-2014-135 (WIS)
	Aza136	–22.54314	46.477414 ●	
	Aza144	–13.70648	48.478917 ●	
	Aza222	–21.84269	43.771898 ●	

estimate. The pairwise Weir & Cockerham (1984) F_{ST} statistic, which can be used to diagnose genetic differentiation among groups, was calculated between all species pairs using VCFtools. While this statistic is more appropriate for within species population differentiation, gene flow between multiple species has been shown previously (Leong Pock Tsy *et al.* 2013; Karimi

et al. 2020). Given the previously reported non-monophyly of *A. za* (Karimi *et al.* 2020), we also calculated genetic diversity metrics for the northern and southern populations of *A. za* separately. Principal components analysis (PCA) was used to visualize genetic diversity among accessions using SNPRelate (Zheng *et al.* 2012) in R v. 4.0.0 (R Core Team 2020).

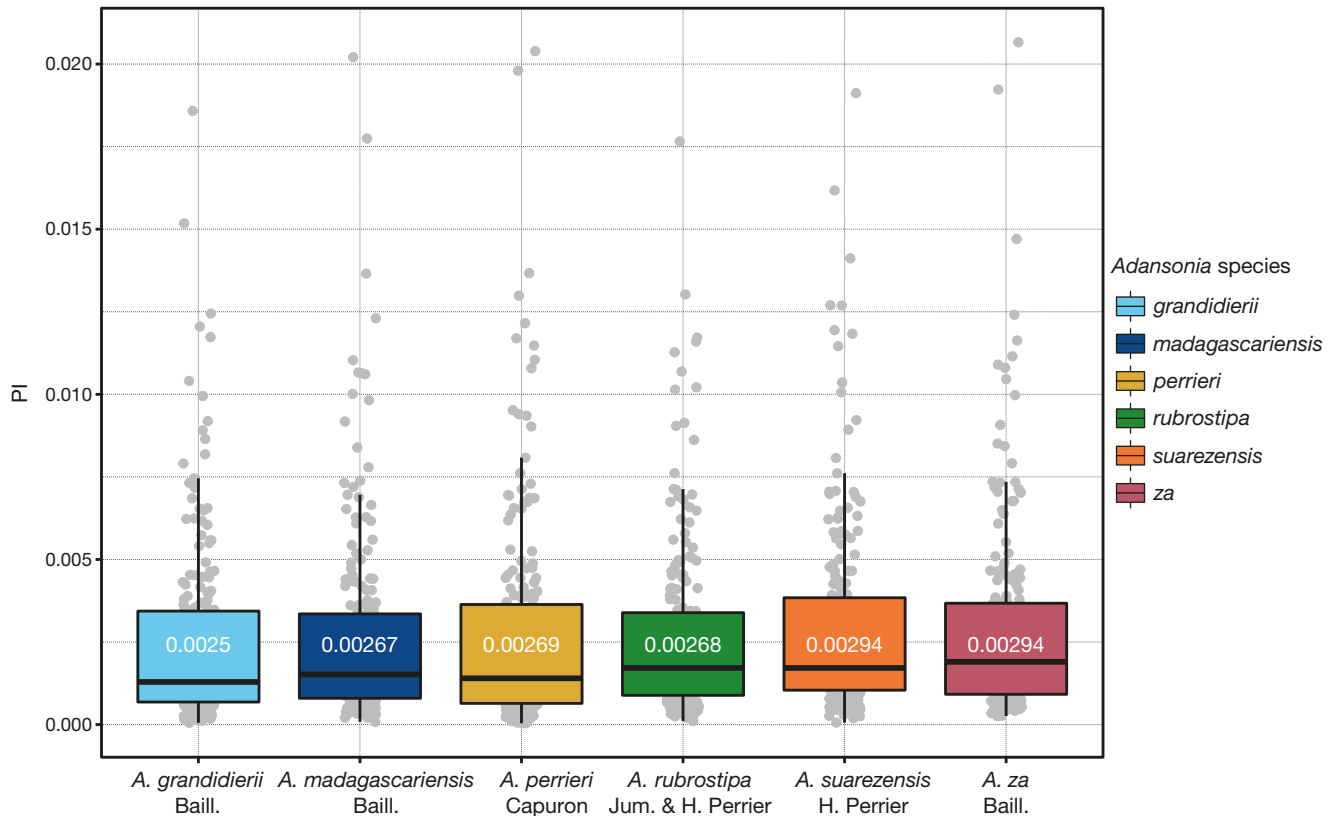


FIG. 4. — Observed nucleotide diversity π for the six *Adansonia* L. species endemic to Madagascar. Northern and southern populations of *A. za* Baill. are combined here. When distinguishing between geographic populations of *A. za*, nucleotide diversity was the same as combined for southern populations ($\pi = 0.0029$), but less for northern populations ($\pi = 0.0026$).

Signals of isolation by distance within the core Longitubae clade (*A. madagascariensis*, *A. perrieri*, and *A. za*; Karimi *et al.* 2020) were evaluated using a Mantel test with adegenet (Jombart 2008) in R with 9999 permutations of genetic diversity against geographic distance. As a measure of genetic distance, we used the pairwise patristic distance (on the maximum-likelihood trees obtained using IQ-TREE) averaged over alleles at each locus and, then, over all 206 loci. To correct for allelic rate-heterogeneity, gene trees were converted to ultrametric using the function chronoMPL in the R package ape v. 5.4-1 (Paradis & Schliep 2019) before calculating patristic distances. Geographic distances were calculated using the haversine method (distance between two geographic points on a sphere) implemented with the distm function in the R package geosphere v. 1.5-10 (Karney 2013).

RESULTS

Targeted sequence capture recovered 206 loci with a total alignment length of 501 792 bp and a mean locus length of 2424 bp $\pm$ 1235. From the 60 *Adansonia* accessions passing our quality filters (Table 1) we recovered a total of 2089 bi-allelic SNPs among the baobab accessions.

The SNP-based phylogeny constructed from these loci using IQ-TREE (Fig. 2) supports three primary lineages within a

monophyletic clade of the Malagasy taxa. Consistent with Karimi *et al.* (2020), we found *A. rubrostipa* as sister to the rest of the Malagasy taxa (bootstrap support; BS = 63%). The sister species *A. grandidierii* and *A. suarezensis* were supported as the Brevitubae clade (BS = 88%), sister to the remaining Malagasy taxa (BS = 80%). The remaining three species (i.e., *A. perrieri*, *A. madagascariensis*, and *A. za*) form the core Longitubae clade (BS = 80%). Although Karimi *et al.* (2020) found strong support for the deepest split in the core Longitubae clade being between southern *A. za* and the other accessions, we observe a lack of resolution at the base of the core Longitubae, with non-monophyly of the southern *A. za* samples. It is unclear if this difference is due to the larger number of accessions studied here or our use of SNP-based methods. As previously, however, the remaining core Longitubae lineages formed a clade (BS = 71%) within which monophyly of *A. perrieri* (BS = 82%) and the northern population of *A. za* (BS = 100%) is well supported. Monophyly of *A. madagascariensis* is uncertain (BS = 32%), as is its sister-group relationship to the northern *A. za* population (BS = 49%).

The SNP-based PCA separates species along multiple axes (Fig. 3), including separating of *A. rubrostipa*, *A. suarezensis* and *A. grandidierii* on axis 1 & axis 2, and axis 3, respectively. Within the core Longitubae, northern and southern *A. za* (Fig. 3C, D) are separated along axes 3 and 4, with the northern accessions clustering close to *A. madagascariensis* and

TABLE 2. — Mean pairwise Weir and Cockerham F_{ST} . Pairwise comparisons with northern and southern populations of *A. za* Baill. are presented separately.

	<i>A. grandidieri</i>	<i>A. suarezensis</i>	<i>A. rubrostipa</i>	<i>A. za</i> (south)	<i>A. za</i> (north)	<i>A. madagascariensis</i>
<i>A. grandidieri</i>	–					
<i>A. suarezensis</i>	0.124	–				
<i>A. rubrostipa</i>	0.193	0.185	–			
<i>A. za</i> (south)	0.159	0.122	0.149	–		
<i>A. za</i> (north)	0.226	0.171	0.197	0.052	–	
<i>A. madagascariensis</i>	0.175	0.151	0.164	0.067	0.058	–
<i>A. perrieri</i>	0.196	0.161	0.179	0.072	0.081	0.070

A. perrieri. Tests for isolation by distance within this clade recover a significant correlation between genetic and geographic distance, accounting for 27% of the genetic variance observed (Mantel: adjusted $r^2 = 0.27$, $p < 0.001$).

All species showed similar levels of nucleotide diversity (π) with outlier values for at least some loci (Fig. 4). Surprisingly, the geographically restricted *A. suarezensis* had one of the highest levels of diversity ($\pi = 0.0029$; Fig. 4) equivalent to that of widespread *A. za* ($\pi = 0.0029$). When distinguishing between populations of *A. za*, nucleotide diversity was somewhat higher in southern populations ($\pi = 0.0029$) than in as in northern populations ($\pi = 0.0026$), which is not unexpected given the much larger range and sample size of the southern *A. za* population.

Differentiation among the three primary clades (*A. rubrostipa*, Brevitubae, and the core Longitubae) was relatively high ($F_{ST} > 0.1$; Table 2) with mean pairwise values of Weir and Cockerham F_{ST} being highest between *A. grandidieri* and northern *A. za* ($F_{ST} = 0.226$). Measures of mean genetic differentiation among the three species of the core Longitubae clade were low ($F_{ST} = 0.053$ – 0.071). The pairwise F_{ST} for northern and southern populations of *A. za* was 0.052, the lowest observed.

DISCUSSION

Conservation of the iconic baobab trees in Madagascar is an important challenge and it is our hope that these data provide an initial framework for more comprehensive assessments and conservation action in the future. Although our results are preliminary, and subject to possible confounding factors, such as conflation of alleles and paralogs (Karimi *et al.* 2020), which might tend to inflate estimates of genetic diversity, they provide a useful baseline for further work on genetic diversity in Malagasy baobab populations.

Measures of genetic differentiation among species were generally low, consistent with a relatively recent radiation and low rates of molecular substitution in these long-lived trees. Differentiation among the core Longitubae species

was particularly low, with F_{ST} values even smaller than those reported for populations of the critically endangered Madagascar palm species, *Dypsis ambositrae* Beentje, which has *c.* 100 mature, extant individuals (Gardiner *et al.* 2017; $F_{ST} = 0.081$ versus 0.053–0.070 among *A. perrieri*, *A. madagascariensis*, and *A. za*). The significant correlation between geographic and genetic distance and the somewhat clinal pattern of genetic differentiation observed in the PCA of the core Longitubae clade (*A. perrieri*, *A. madagascariensis*, and *A. za*) is consistent with some degree of interspecific gene flow, as previously argued by Leong Pock Tsy *et al.* (2013). Because the northern populations of *A. za* occupy more humid gallery forests than the dry deciduous forests occupied by southern *A. za*, it is possible that introgression of genes from the wet-forest adapted *A. perrieri* had adaptive consequences (Vieilledent *et al.* 2013). However, more detailed studies correlating population structure with ecological tolerance within *A. za* may be warranted and may yield insight critical for conservation.

If future analyses confirm a distinction between the northern and southern *A. za*, splitting this species into two distinct species (*A. za* Baill. and *A. bozy* Jum. & H. Perrier) would have value in helping to highlight the critical need for conservation efforts in the northern populations. Although *A. za* is fairly widespread in southern Madagascar it has a restricted distribution in the Sambirano region of northern Madagascar. Climate niche modeling suggests loss of habitat for all species by 2050–2070 due to climate change alone (Wan *et al.* 2020), so the spatially restricted populations of northern *A. za* deserve more attention. This is a cause for concern since northern *A. za* has some ecological similarity to *A. perrieri*, which is considered to be endangered, with as few as 99 trees remaining (Vieilledent *et al.* 2013).

Genetic diversity is used as a measure of long-term population viability, as it is proxy for both adaptability and mutational load. Our analyses find that genetic diversity within the Malagasy baobabs was comparable to the low end of the range for long-lived conifers (e.g., $\pi = 0.0024$ – 0.0082 ; Mosca *et al.* 2012), but nearly half that found for long-lived Loblolly pine ($\pi = 0.00640$; Brown *et al.* 2004), and similar to that found within highly inbred *Gossypium* species (Malvaceae) where π ranges from *c.* 0.002–0.003 (Yuan *et al.* 2021). While outcrossed species typically have large effective population sizes (Jansson & Ingvarsson 2010), mammal-pollinated species, such as the Brevitubae, are more susceptible to reductions in genetic diversity compared to wind-pollinated trees (such as conifers; Loveless & Hamrick 1984). Interestingly, we found the highest genetic diversity in *A. suarezensis*, whose range is limited and whose populations are generally small (total population *c.* 15000 individuals; Vieilledent *et al.* 2013). Conversely, its sister species, the iconic *A. grandidieri*, exhibited the lowest genetic diversity, despite having extant populations estimated at over 1 million trees (Vieilledent *et al.* 2013). While counterintuitive, similar patterns have been shown for endangered Malagasy *Dypsis* Noronha ex Mart., where the more widespread species *D. decipiens* had

lower genetic diversity relative to *D. ambositrae*, which is composed of small, geographically isolated populations (Gardiner *et al.* 2017). Likewise, a genetic study of the leguminous genus *Delonix* Raf. (Rivers *et al.* 2014) revealed similar levels of genetic diversity between two species, i.e., the widespread *D. decaryi* and the Madagascar endemic *D. floribunda* (Baill.) Capuron. Notably, genetic diversity in *D. decaryi* was related to spatial distance, but not for the endemic *D. floribunda*. We see a similar pattern here, where estimates of habitat connectivity for the genetically diverse *A. suarezensis* (Vieilledent *et al.* 2013) suggest high fragmentation and low opportunity for gene flow among isolated populations/individuals.

Subject to poor recruitment and low seedling numbers for all species, natural regeneration is notably poor for many species, as exemplified by a demographic study of the widespread species *A. rubrostipa* (Metcalfe *et al.* 2007), which found very low densities and poor recruitment of younger trees in a previously logged area, despite the relatively high density of remaining mature trees (as previously suggested by Baum 1996). Furthermore, some species of baobab require scarification for germination (Razanameharizaka *et al.* 2006) and it is presumed that now-extinct species, such as *Archaeolemur* (Baum 1995) and giant tortoises *Aldabrachelys grandidieri* and *A. abrupta* (Andriantsaralaza *et al.* 2014) aided in their dispersal. A lack of ongoing recruitment would call into question the integrity and vulnerability of all populations, and suggest that the absolute number of existing mature individuals per species may not accurately predict long-term unassisted viability.

While the Madagascar Protected Area Network (Système des Aires Protégées de Madagascar 2016) is critical for increasing and delimiting conservation areas, nearly all baobab trees occur within fragmented forests, even those located in some type of reserve (e.g., Kirindy, La Réserve spéciale d'Ankarana, Avenue de Baobabs, Montagne d'Ambre). Such a lack of habitat connectivity likely restricts breeding pools and is expected to ultimately degrade genetic diversity. Therefore, population viability should be measured not only by genetic diversity and number of individuals within protected areas, but also includes other measures predictive of population success. For some species, such as *A. za* and *A. rubrostipa*, interventions may include introducing individuals from other populations with similar environmental conditions. For others, it may require ancillary efforts to restore pollinator (i.e., bat and hawkmoth), reduce seedling mortality, and restore habitat connectivity to increase gene flow. Input is needed from local residents and stakeholders for designing any conservation plan (Marie *et al.* 2009), but our data suggests that it may be advisable to immediately begin conserving genetic material of all species from populations outside of currently protected areas. Given the long-lived nature of baobab trees and their poor natural regeneration, prompt implementation of preservation and restoration actions is necessary to preserve the remaining genetic variation and to ensure the long-term persistence of these emblematic and ecologically important trees.

Acknowledgements

Funding for this project was supported by National Science Foundation award DEB-1354268 to DAB, CEG, and JFW. We are grateful to the botanists at Parc botanique et zoologique de Tsimbazaza (PZBT) for field support and staff at the Missouri Botanical Garden Madagascar Program for logistical support. We thank Diana Mayne for providing some samples, and the ReseachIT Unit at Iowa State University (<https://researchit.las.iastate.edu/>) for computational support. Lolona Ramamonjisoa and an anonymous reviewer are also thanked for their remarks on a previous version of the article. We would also like to thank Onja Razanamaro for inviting this work.

REFERENCES

- AGRAWAL A. F. & WHITLOCK M. C. 2012. — Mutation load: the fitness of individuals in populations where deleterious alleles are abundant. *Annual Review of Ecology, Evolution, and Systematics* 43: 115-135. <https://doi.org/10.1146/annurev-ecolsys-110411-160257>
- ANDRIANTSARALAZA S., PEDRONO M., TASSIN J., ROGER E., RAKOOUTH B. & DANTHU P. 2014. — The role of extinct giant tortoises in the germination of extant baobab *Adansonia rubrostipa* seeds in Madagascar. *African Journal of Ecology* 52: 246-249. <https://doi.org/10.1111/aje.12101>
- BAUM D. A. 1995. — A systematic revision of *Adansonia* (Bombacaceae). *Annals of the Missouri Botanical Garden* 82: 440-470. <https://doi.org/10.2307/2399893>
- BAUM D. A. 1996. — The ecology and conservation of the baobabs of Madagascar. *Primate Report* 46: 311-327.
- BOLGER A. M., LOHSE M. & USADEL B. 2014. — Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30 (15): 2114-2120. <https://doi.org/10.1093/bioinformatics/btu170>
- BROWN G. R., GILL G. P., KUNTZ R. J., LANGLEY C. H. & NEALE D. B. 2004. — Nucleotide diversity and linkage disequilibrium in loblolly pine. *Proceedings of the National Academy of Sciences* 101 (42): 15255-15260. <https://doi.org/10.1073/pnas.0404231101>
- CONGDON J. D., DUNHAM A. E. & VAN LOBEN SELS R. C. 1993. — Delayed sexual maturity and demographics of Blanding's turtles (*Emydoidea blandingii*): implications for conservation and management of long-lived organisms. *Conservation Biology* 7 (4): 826-833. <https://doi.org/10.1046/j.1523-1739.1993.740826.x>
- COOK C. N. & SGRÒ C. M. 2017. — Aligning science and policy to achieve evolutionarily enlightened conservation. *Conservation Biology* 31 (3): 501-512. <https://doi.org/10.1111/cobi.12863>
- DANECEK P., AUTON A., ABECASIS G., ALBERS C. A., BANKS E., DEPRISTO M. A., HANDSAKER R. E., LUNTER G., MARTH G. T., SHERRY S. T. & MCVEAN G. 2011. — The variant call format and Vcftools. *Bioinformatics* 27 (15): 2156-2158. <https://doi.org/10.1093/bioinformatics/btr330>
- FRANKHAM R. 1996. — Relationship of genetic variation to population size in wildlife. *Conservation Biology* 10: 1500-1508. <https://doi.org/10.1046/j.1523-1739.1996.10061500.x>
- KARNEY C. F. F. 2013. — Algorithms for geodesics. *Journal of Geodesy* 87: 43-55. <https://doi.org/10.1007/s00190-012-0578-z>
- HOBAN S., BRUFORD M., JACKSON J. D. U., LOPES-FERNANDES M., HEUERTZ M., HOHENLOHE P. A., PAZ-VINAS I., SJÖGREN-GULVE P., SEGELBACHER G., VERNESI C. & AITKEN S. 2020. — Genetic diversity targets and indicators in the Cdb post-2020 Global Biodiversity Framework must be improved. *Biological Conservation* 248: 108654. <https://doi.org/10.1016/j.biocon.2020.108654>
- GAO Y., WANG H., LIU C., CHU H., DAI D., SONG S., YU L., HAN L., FU Y., TIAN B. & TANG L. 2018. — De novo genome assembly of the red silk cotton tree (*Bombax ceiba*). *GigaScience* 7: giy051. <https://doi.org/10.1093/gigascience/gyi051>

- GAMBLIN T., LEGENDRE M., COLLETTE M. R., LEE G. L., MOODY A., DE SUPINSKI B. R. & FUTRAL S. 2015. — The Spack package manager: bringing order to Hpc software chaos, in *Sc'15: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis* 1-12. IEEE.
- GARDINER L. M., RAKOTOARINIVO M., RAJAOVELONA L. R. & CLUBBE C. 2017. — Population genetics data help to guide the conservation of palm species with small population sizes and fragmented habitats in Madagascar. *PeerJ* 5: 3248. <https://doi.org/10.7717/peerj.3248>
- QUINLAN A. R. & HALL I. M. 2010. — Bedtools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26: 841-842. <https://doi.org/10.1093/bioinformatics/btq033>
- FISCHER M. C., RELLSTAB C., LEUZINGER M., ROUMET M., GUGERLI F., SHIMIZU K. K., HOLDEREGGER R. & WIDMER A. 2017. — Estimating genomic diversity and population differentiation—an empirical comparison of microsatellite and Snp variation in *Arabidopsis halleri*. *BMC genomics* 18 (1): 1-15. <https://doi.org/10.1186/s12864-016-3459-7>
- HAMRICK J. L. & GODT M. W. 1996. — Effects of life history traits on genetic diversity in plant species. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 351 (1345): 1291-1298. <https://doi.org/10.1098/rstb.1996.0112>
- JANSSON S. & INGVARSSON P. K. 2010. — Cohort-structured tree populations. *Heredity* 105: 331-332. <https://doi.org/10.1038/hdy.2010.101>
- JOMBART T. 2008. — “adeigenet: a R package for the multivariate analysis of genetic markers”. *Bioinformatics* 24 (11): 1403-1405. <https://doi.org/10.1093/bioinformatics/btn129>
- KARIMI N., GROVER C. E., GALLAGHER J. P., WENDEL J. F., ANÉ C. & BAUM D. A. 2020. — Reticulate evolution helps explain apparent homoplasy in floral biology and pollination in baobabs (*Adansonia*; Bombacoideae; Malvaceae). *Systematic Biology* 69 (3): 462-478. <https://doi.org/10.1093/sysbio/syz073>
- KATO H., MISAWA K., KUMA K. I. & MIYATA T. 2002. — Mafft: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* 30 (14): 3059-3066. <https://doi.org/10.1093/nar/gkf436>
- KENDIG K. I., BAHETI S., BOCKOL M. A., DRUCKER T. M., HART S. N., HELDENBRAND J. R., HERNAEZ M., HUDSON M. E., KALMBACH M. T., KLEE E. W. & MATTSO N. R. 2019. — Sentieon Dnaseq variant calling workflow demonstrates strong computational performance and accuracy. *Frontiers in Genetics* 10: 736. <https://doi.org/10.3389/fgene.2019.00736>
- KNIGHT R., MAXWELL P., BIRMINGHAM A., CARNES J., CAPORASO J. G., EASTON B. C., EATON M., HAMADY M., LINDSAY H., LIU Z. & LOZUPONE C. 2007. — PyCogent: a toolkit for making sense from sequence. *Genome Biology* 8 (8): 171. <https://doi.org/10.1186/gb-2007-8-8-r171>
- LAIKRE L., ALLENDORF F. W., ARONER L. C., BAKER C. S., GREGOVICH D. P., HANSEN M. M., JACKSON J. A., KENDALL K. C., MCKELVEY K. E., NEEL M. C. & OLIVIERI I. 2010. — Neglect of genetic diversity in implementation of the convention on biological diversity. *Conservation Biology* 24 (1): 86-88. <https://doi.org/10.1111/j.1523-1739.2009.01425.x>
- LEACHÉ A. D., BANBURY B. L., FELSENSTEIN J., DE OCA A. N. M. & STAMATAKIS A. 2015. — Short tree, long tree, right tree, wrong tree: new acquisition bias corrections for inferring Snp phylogenies. *Systematic Biology* 64 (6): 1032-1047. <https://doi.org/10.1093/sysbio/syv053>
- LEONG POCK TSY J. M., LUMART R., FLAVEN-NOGUIER E., SAUVE M., DUBOIS M. P. & DANTHU P. 2013. — Nuclear microsatellite variation in Malagasy baobabs (*Adansonia*, Bombacoideae, Malvaceae) reveals past hybridization and introgression. *Annals of Botany* 112 (9): 1759-1773. <https://doi.org/10.1093/aob/mct230>
- LETSARA R., FARAMALALA M., RAZAFINDRAHAJA V. & FARANIRINA L. 2019. — *Adansonia za*. The IUCN Red List of Threatened Species 2019: e.T37680A64367122. <https://doi.org/10.2305/IUCN.UK.2019-3.RLTS.T37680A64367122.en>
- LI H. & DURBIN R. 2009. — Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25 (14): 1754-1760. <https://doi.org/10.1093/bioinformatics/btp324>
- LOVELESS M. D. & HAMRICK J. L. 1984. — Ecological determinants of genetic structure in plant populations. *Annual review of ecology and systematics* 15 (1): 65-95. <https://doi.org/10.1146/annurev.es.15.110184.000433>
- MARIE C. N., SIBELET N., DULCIRE M., RAFALIMARO M., DANTHU P. & CARRIÈRE S. M. 2009. — Taking into account local practices and indigenous knowledge in an emergency conservation context in Madagascar. *Biodiversity and Conservation* 18: 2759-2777. <https://doi.org/10.1007/s10531-009-9672-9>
- McKENNA A., HANNA M., BANKS E., SIVACHENKO A., CIBULSKIS K., KERNYTSKY A., GARIMELLA K., ALTSHULER D., GABRIEL S., DALY M. & DEPRISTO M. A. 2000. — The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research* 20: 1297-303. <https://doi.org/10.1101/gr.107524.110>
- METCALFE D. J., SANCHEZ A. C., CURRAN P. M., HAGA J. A., KJIA H. K., KLEYNHANS E. J., KOPP M., KOROGONE S. U., MADINDOU I. R., MINLEND A. & NDAGIJIMANA F. 2007. — Distribution and population structure of *Adansonia rubrostipa* in dry deciduous forest in western Madagascar. *African Journal of Ecology* 45 (4): 464. <https://doi.org/10.1111/j.1365-2028.2007.00755.x>
- MILLAR C. I. & LIBBY W. J. 1991. — Strategies for conserving clinal, ecotypic, and disjunct population diversity in widespread species, in FALK D. A. & HOLSINGER K. E. (eds), *Genetics and Conservation of Rare Plants*. Oxford University Press, New York: 149-170.
- MOSCA E., ECKERT A. J., LIECHTY J. D., WEGRZYŃ J. L., LA PORTA N., VENDRAMIN G. G. & NEALE D. B. 2012. — Contrasting patterns of nucleotide diversity for four conifers of Alpine European forests. *Evolutionary Applications* 5: 762-775. <https://doi.org/10.1111/j.1752-4571.2012.00256.x>
- NUNNEY L. 1993. — The influence of mating system and overlapping generations on effective population size. *Evolution* 47(5): 1329-1341. <https://doi.org/10.1111/j.1558-5646.1993.tb02158.x>
- NGUYEN L.-T., SCHMIDT H. A., VON HAESLER A. & MINH B. Q. 2015. — IQ-TREE: A fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Molecular biology and evolution* 32 (1): 268-274. <https://doi.org/10.1093/molbev/msu300>
- PAGE J. T., LIECHTY Z. S., HUYNH M. D. & UDALL J. A. 2014. — BamBam: genome sequence analysis tools for biologists. *Bmc Research Notes* 7: 1-5. <https://doi.org/10.1186/1756-0500-7-829>
- PARADIS E. & SCHLIEP K. 2019. — APE 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35: 526-528. <https://doi.org/10.1093/bioinformatics/bty633>
- R CORE TEAM. 2020. — *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.r-project.org/>
- RAZANAMEHARIZAKA J., GROUZIS M., RAVELOMANANA D. & DANTHU P. 2006. — Seed storage behaviour and seed germination in African and Malagasy baobabs (*Adansonia* species). *Seed Science Research* 16: 83-88. <https://doi.org/10.1079/Ssr2005231>
- RICE P., LONGDEN I. & BLEASBY A. 2000. — Emboss: the European molecular biology open software suite. *Trends in Genetics* 16: 276-277. [https://doi.org/10.1016/s0168-9525\(00\)02024-2](https://doi.org/10.1016/s0168-9525(00)02024-2)
- RIVERS M. C., BRUMMITT N. A., LUGHADHA E. N. & MEAGHER T. R. 2014. — Do species conservation assessments capture genetic diversity?. *Global Ecology and Conservation* 2: 81-87. <https://doi.org/10.1016/j.gecco.2014.08.005>
- SCHEMSKE D. W., HUSBAND B. C., RUCKELSHAUS M. H., GOODWILLIE C., PARKER I. M. & BISHOP J. G. 1994. — Evaluating approaches to the conservation of rare and endangered plants. *Ecology* 75 (3): 584-606. <https://doi.org/10.2307/1941718>
- VIEILLEDENT G., CORNU C., SANCHEZ A. C., POCK-TSY J. M. L. & DANTHU P. 2013. — Vulnerability of baobab species to climate

- change and effectiveness of the protected area network in Madagascar: Towards new conservation priorities. *Biological Conservation* 166: 11-22. <https://doi.org/10.1016/j.biocon.2013.06.007>
- WAN J. N., MBARI N. J., WANG S. W., LIU B., MWANGI B. N., RASOARAHONA J. R., XIN H. P., ZHOU Y. D. & WANG Q. F. 2020. — Modeling impacts of climate change on the potential distribution of six endemic baobab species in Madagascar. *Plant Diversity*. <https://doi.org/10.1016/j.pld.2020.07.001>
- WEIR B. S. & COCKERHAM C. C. 1984. — Estimating F-statistics for the analysis of population structure. *Evolution* 1: 1358-1370. <https://doi.org/10.2307/2408641>
- WICKENS G. E. 2008. — *The Baobabs: Pachycauls of Africa, Madagascar and Australia*. Springer Science, New York, 498 p.
- WU T. D. & WATANABE C. K. 2005. — GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* 21: 1859-1875. <https://doi.org/10.1093/bioinformatics/bti310>
- YUAN D., GROVER C. E., HU G., PAN M., MILLER E. R., CONOVER J. L., HUNT S. P., UDALL J. A. & WENDEL J. F. 2021. — Parallel and intertwining threads of domestication in allopolyploid cotton (*Gossypium*). *Advanced Science* 8 (10): 2003634. <https://doi.org/10.1002/advs.202003634>
- ZHENG X., LEVINE D., SHEN J., GOGARTEN S. M., LAURIE C. & WEIR B. S. 2012. — A high-performance computing toolset for relatedness and principal component analysis of Snp data. *Bioinformatics* 28: 3326-3328. <https://doi.org/10.1093/bioinformatics/bts606>

*Submitted on 30 January 2021;
accepted on 11 May 2021;
published on 17 February 2022.*