The Evolutionary History of the Baobab Trees

By

Nisa Karimi

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
(Botany)

at the

UNIVERSITY OF WISCONSIN-MADISON
2018

Date of final oral examination: 12/11/2018

The dissertation is approved by the following members of the Final Oral Committee:

David A. Baum, Professor, Botany
Ken Sytsma, Professor, Botany
Cécile Ané, Professor, Botany & Statistics
Ken Cameron, Professor, Botany
Sean Schoville, Associate Professor, Entomology

Table of Contents

Co-Author Contributions

Chapter 1 was completed in collaboration with Corrinne E. Grover, Cécile Ané, Joseph P. Gallagher and Jonathan F. Wendel. Jonathan F. Wendel and Corrinne E. Grover provided guidance on the targeted sequence capture bait design and sequencing, Joseph P. Gallagher was responsible for the sequence captures, Corrinne E. Grover also assisted in the bioinformatics of obtaining haplotype sequences, and Cécile Ané performed the ancestral trait reconstructions and provided guidance on the network analyses. I was responsible for obtaining samples and performing DNA extractions, conducting read assemblies and dataset generation, phylogenetic and network analyses, and generating draft manuscripts.

Chapter 2 utilized the targeted sequence capture datasets developed for Chapter 1. I expanded on the initial dataset by collecting samples in Madagascar and refining the read assembly pipeline. I was responsible for all analyses conducted.

Chapter 3 was carried out in collaboration with Sarah Venter who provided guidance and field support in South Africa, and Ken Keefover-Ring who advised on the floral volatiles and nectar analyses. I designed the studies, carried out the field and the majority of laboratory work, supervised undergraduate students who generated some laboratory data, and wrote the manuscript.

Chapter 4 is adapted from Cron, G.V., Karimi, N., Glennon, K.L., Udeh, C.A., Witkowski, E.T., Venter, S.M., Assogbadjo, A.E., Mayne, D. and Baum, D.A., 2016. "One African baobab species or two? Synonymy of *Adansonia kilima* and *A. digitata*." *Taxon*, *65*(5), pp.1037-1049. Glynis V. Cron and I were co-first authors. Cron, Glennon, Chukwudi, Witkowski and Venter were responsible for the ordination of floral features from herbarium specimens. Assogbadjo and Diana Mayne sent samples for the molecular phylogenetics. I was

responsible for fieldwork in Tanzania, collecting measurements from these samples used for ordination, conducting chromosome counts, and molecular phylogenetic analysis.

Dissertation Summary

Baobabs have long fascinated botanists and travelers alike. Anyone who has had the pleasure of witnessing these majestic trees, their peculiar growth form, remarkable flowers, and colorful folklore is likely to want to know more about them. In my research I have sought to address some of the many scientific questions remaining unanswered to this day: What are the origins of this genus given their interesting geographic distribution? How and when did they come to occupy two separate continents, Africa and Australia? Why is it one species widespread across all of continental Africa yet 6 species are found endemic to Madagascar? What drivers have led to their remarkable diversity on Madagascar? What is the evolutionary history of their pollination systems and floral diversity? Do patterns of morphological variation in Africa correspond to cryptic species?

My dissertation research addressed these questions via a diversity of approaches that today make up the field of plant systematics, including molecular phylogenetics, pollination biology, and chemical ecology. Briefly, my chapters addressed the following:

1) We conducted phylogenomic analyses of hundreds of nuclear loci, finding evidence that the three lineages (Africa, Australia, Madagascar) diverged from one another rapidly and that there was a second dispersal to Madagascar, which introduced some genes from the African species into the ancestor of two Malagasy species, and may have facilitated the parallel evolution of mammal pollination in Africa and Madagascar.

2) We sampled the six species of baobabs endemic to Madagascar across their ranges to assess the validity of traditional species designations. Our results show strong geographic signals within and between species, and we were able to identify several instances of introgression between morphological species. Our data suggests that northern populations of the most widespread Malagasy species *A. za* should probably be recognized as their own species (*A. bozy*), which corresponds to historical variety designations.

3) We compiled a comprehensive understanding of the pollination ecology of the African baobab, *A. digitata*. We confirmed that the African baobab is self-incompatible, mostly likely by late-acting self-incompatibility. At least in populations in South Africa, although this species has a classic bat-pollination syndrome, hawkmoths are serving as the primary pollinator. Assessing floral scent profiles we find evidence of geographic variation in floral scent between West, East, and Southern Africa.

4) We assessed the validity of the recently described diploid African species, *A. kilima*. Based on ordination of floral features, phylogenetics and chromosome counts, we conclude that *A. kilima* is not distinct from the tetraploid *A. digitata*.

Chapter One:

Ancient Introgression in the Baobabs, *Adansonia* (Bombacoideae; Malvaceae) by Long Distance Dispersal Aids Floral Evolution

Nisa Karimi[1*], Corrinne E. Grover[2], Cécile Ané[1,4], Joseph P. Gallagher[2,3,] Jonathan F. Wendel[2], David A. Baum[1,5*]

[1]*Department of Botany, University of Wisconsin Madison, WI, 53706, United States of America*

[2]*Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, IA, 50011, United States of America*

[3]*Biology Department, University of Massachusetts, Amherst, MA 01003, United States of America*

[4]*Department of Statistics, University of Wisconsin Madison, WI, 53706, United States of America*

[5]*Wisconsin Institute for Discovery, 330 N Orchard St, Madison, WI 53715, United States of America*

*Email: nkarimi@wisc.edu, dbaum@wisc.edu

ABSTRACT

Keywords*: Targeted sequence capture, paralogy, reticulate evolution, network inference, introgression, species tree inference*

The baobab genus *Adansonia* (Malvaceae) includes eight species (Fig. 1) that have been delimited based on morphological and molecular data (Baum 1995b; Baum et al. 1998; Cron et

al. 2016). The group possesses an unusual geographic distribution, with one tetraploid species widespread across continental Africa (*A. digitata*), one diploid species endemic to Northwestern Australia (*A. gregorii*), and six diploid species restricted to Madagascar. Previous phylogenetic analyses identified three distinct lineages corresponding to geography; however, the relationships between these remain unresolved (Baum et al. 1998; Pettigrew et al. 2015).

Within Madagascar, the six recognized species are divided into two sections, *Brevitubae* and Longitubae, based on morphology (Baum 1995b). Section *Brevitubae* includes a pair of allopatric species, *A. grandidieri* in the southwest and *A. suarezensis* in the north, which form a clade that is well-supported by ITS sequences (Baum et al. 1998). Both species are characterized by flowers with short staminal tubes, white petals, cream-colored inner calyces, and mammal pollination (Baum 1995a; Andriafidison et al. 2006). All four species of Sect. *Longitubae* share elongated flowers, with long staminal columns, red or yellow petals, and a reddish inner calyx and style. This floral morphology is associated with pollination by long-tongued hawkmoths (Baum 1995a; Ryckewaert et al. 2011). Despite their floral similarities, prior molecular analyses have found limited support for the monophyly of the Malagasy *Longitubae*, though nuclear ribosomal ITS and, to some extent, plastid data support a clade composed of *A. madagascariensis*, *A. za*, and *A. perrieri* (hereafter referred to as the "core" *Longitubae*), to the exclusion of *A. rubrostipa* (Baum et al. 1998). Despite differences in flowering phenology among sympatric *Longitubae* species, prior work has suggested gene tree discordance in this group (Baum et al. 1998) and the possibility of introgression (Leong Pock Tsy et al. 2013).

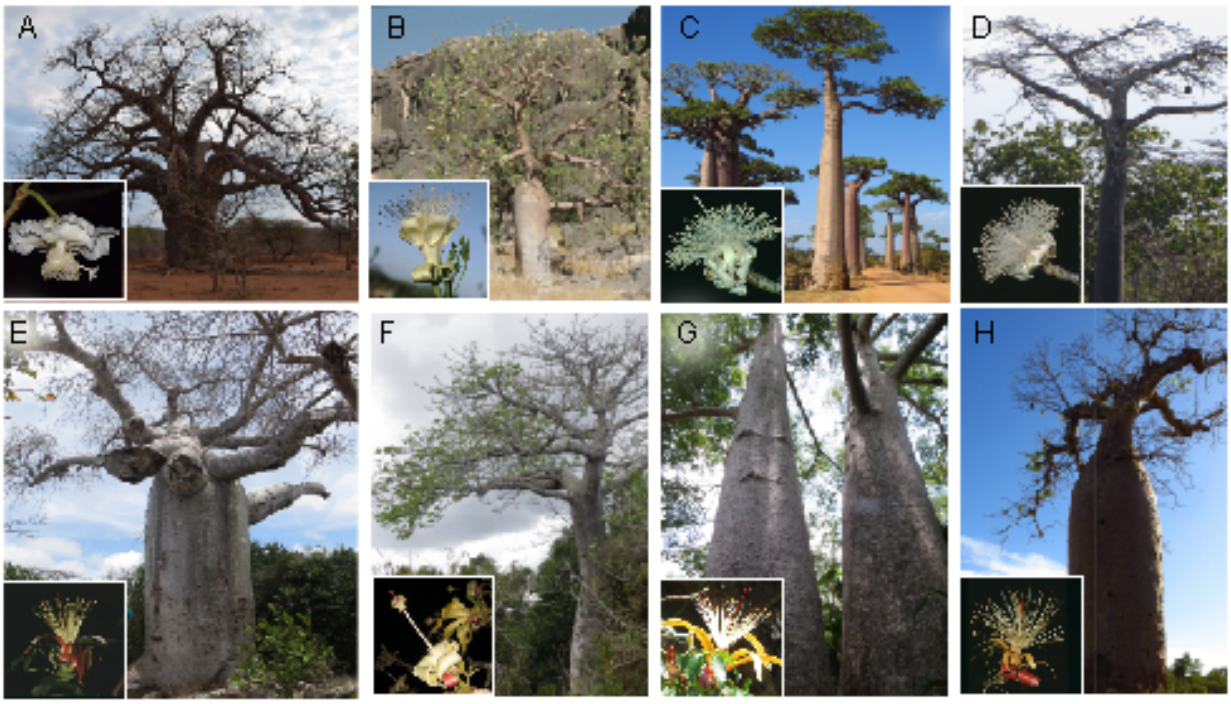


Figure 1. Eight species of *Adansonia.* A) *A. digitata*, continental Africa, B) *A. gregorii,* Australia, C) *A. grandidieri*, Madagascar, D) *A. suarezensis*, Madagascar, E) *A. madagascariensis*, Madagascar, F) *A. perrieri*, Madagascar, G) *A. za*, Madagascar, H) *A. rubrostipa,* Madagascar.

Inferring the genealogical history in many groups is challenging due to incomplete lineage sorting (ILS), reticulation (hybridization and introgression), and gene copy number variation, among other phenomena. Resolving the complex histories generated by these phenomena may be facilitated through comparison of many independent gene genealogies (Pamilo and Nei 1998; Raymond et al. 2002). Targeted sequence capture, or hyb-seq, has increased in popularity as a source of hundreds of low-copy nuclear genes for the purpose of multi-gene phylogenetic analyses in non-model systems (Ekblom and Galindo 2011; Zimmer and Wen 2015; Grover et al. 2015, Harvey et al. 2016, Chau et al. 2018; Wolf et al. 2018).

Genomic "baits" selectively enrich specific genes or regions in next-generation sequencing libraries. Although baits for non-model species are frequently derived from transcriptomic data, intron and/or intergenic regions are frequently recovered as part of the "splash zone" (Liston et al. 2014), both of which can be helpful for low-level phylogenetics (Grover et al. 2012; Weitemier et al. 2014). Furthermore, organellar genes can also be recovered from the off-target reads to obtain plastid and/or mitochondrial genomes (Weitemier et al. 2014). Targeted sequence capture is typically used with short-read sequencing, which is appropriate and useful for samples with degraded or poor-quality DNA, such as herbarium specimens (Hart et al. 2016).

Even when care is taken in bait design to target single-copy genes, the prevalence of tandem, segmental, and whole genome duplications in plants (Van de Peer et al. 2009; Jiao et al. 2011; Wendel 2015; Conover et al. 2018) often leads to the joint recovery of paralogs. Accordingly, caution needs to be exercised (Nicholls et al. 2015) in ortholog assignment and in screening for chimeric sequences (Philippe et al. 2011; Struck 2013). Several bioinformatic pipelines have been developed to facilitate assembly from hyb-seq data (i.e. Yang and Smith 2014; Johnson et al. 2016; Kamneva et al. 2017, Fér and Schmickl 2018).

Here we used custom-designed baits and targeted capture to obtain sequences of hundreds of independent low-copy nuclear loci for all species of baobab trees. We find support for a history of reticulation involving gene flow across the Mozambique channel from the ancestral African lineage into Sect. Brevitubae in Madagascar. Given that *A. digitata* is white-flowered and mammal pollinated, our result leads us to speculate that the evolution of mammal-pollination in the Brevitubae was facilitated by genes acquired from ancient introgression. This speculation is supported by ancestral trait reconstructions. We also find evidence of additional introgression events between species in Madagascar. Our analyses highlight the risk of cryptic

paralogy as a confounding factor in phylogenetic network analysis, and also point to the influence of read depth in paralog recovery.

METHODS

Bait Design

Total RNA was extracted from fresh leaf tissue of *Adansonia digitata* L. and *Bombax ceiba* L., as in Chang et al. (1993), followed by cleanup using the Qiagen RNeasy kit following manufacturer instructions (QIAGEN Inc., Valencia, California, USA). RNA quality assessment was performed by Agilent RNA PicoChip Analysis (Agilent Technologies, Inc., Santa Clara, California, USA) using 1µl of each sample diluted to 5 ng/µl. RNA library preparation was performed at the University of Wisconsin - Madison Biotechnology Center (Madison, WI) using an Illumina TruSeq RNA Sample Prep kit (Illumina Inc., San Diego, CA, USA) followed by purification with Agencourt AMPure XP beads (Beckman Coulter, USA). Library quantification was checked with a Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, USA) per manufacturer instructions. Samples were adjusted to a final concentration of 28-31ng/µl. Quality and quantity of the finished libraries were assessed using an Agilent DNA1000 chip (Agilent Technologies, USA) and Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, USA), respectively. Libraries were standardized to 2nM. Cluster generation was performed using Illumina TruSeq Cluster Kits and the Illumina Cluster Station (Illumina Inc., San Diego, CA, USA). Paired-end, 100bp sequencing was performed at the University of Wisconsin - Madison Biotechnology Center (Madison, WI) using SBS chemistry on an Illumina HiSeq2000 sequencer. Images were analyzed using the Illumina Pipeline, version 1.8.2.

Raw reads were assembled *de novo* with Trinity version 2.1.0 (Luo et al. 2012) and the resulting contigs were used as BLAST queries against each other, the *Arabidopsis* ultra-conserved sequence database (http://cgpdb.ucdavis.edu/cgpdb2/), and the *Gossypium* exome (Paterson et al. 2012). For each contig showing reciprocal best-BLAST matches between *Adansonia* and *Bombax, Adansonia-Bombax-Gossypium* alignments were identified that had >800bp of continuously aligned sequences and an average pairwise sequence similarity between the Bombacoids and cotton ≥ 93.5%. These candidates were screened for repetitive sequences by RepeatMasker (Smit et al., http://repeatmasker.org) and for potential paralog sequences (sequences that clustered with two or more genes from the cotton transcriptome). When base calls were ambiguous in *Adansonia*, the bait sequence was based on *Bombax*. When both species were polymorphic (i.e., heterozygous) at a position, the most common base among the two species' sequences was selected. The resulting targets were used as the basis for the design and synthesis of 120 bp, 2X-tiled MYbaits (Arbor Biosciences, formerly Mycroarray, Ann Arbor, MI, USA), available as Supplementary File S1.

Taxon Sampling and Targeted Sequence Capture

Sampling included one to three accessions per *Adansonia* species in addition to three outgroups (Table S1). DNA was extracted from silica-dried leaf tissue or seeds with the Qiagen DNeasy Plant Mini Plant Kit (Qiagen, USA), following manufacturer instructions, but with the following modifications: (1) increased lysis buffer to 650µl and included 10µl Proteinase K (25mg/mL), (2) tissue and lysis buffer was incubated at 65°C for 20 minutes rather than the recommended 10 minutes, (3) all centrifugation steps were performed at 4°C, and (4) final elution used heated buffer (approx. 80°C) and was then incubated at room temperature for 10 minutes before

centrifugation. DNA quality and quantity was estimated by 1% agarose gel electrophoresis. Qubit Fluorometric Quantitation (Life Technologies) was used for further quantification prior to DNA library preparation.

DNA library construction was performed at (1) the University of Wisconsin - Madison Biotechnology Center; or (2) the Genomics Core Facility at West Virginia University. Genomic DNA was sheared using a Covaris shredder to achieve a standard fragment range of 500-600 bp. Sequence capture was performed at Iowa State University using the MYbaits protocol version 2 (Arbor Biosciences, formerly Mycroarray, Ann Arbor, MI, USA). Briefly, libraries were denatured and hybridized with biotinylated RNA capture baits over 36 hours. Quantity and quality of the captured libraries were assessed via Quan-it PicoGreen dsDNA assay (ThermoFisher Scientific, USA) and Agilent Bioanalyzer 2100, respectively. Enrichment of post-capture capture libraries was verified via QPCR as described previously (Salmon 2012, Grover 2017).

Target-enriched, Illumina TruSeq libraries for an initial twelve accessions were sequenced on a single lane of Illumina MiSeq as 2x300bp by the UW-Madison Biotechnology Center (Madison, WI). Subsequent accessions were sequenced at the Beijing Genomics Institute (BGI, Hong Kong) on the Illumina HiSeq2500 as 2x250bp. Raw reads were quality trimmed using Trimmomatic v0.36 (Bolger et al. 2014) with the parameters *ILLUMINACLIP:Adapters.fa:2:30:15 LEADING:28 TRAILING:28 SLIDINGWINDOW:8:28 SLIDINGWINDOW:1:10 MINLEN:65 TOPHRED33* (scripts available online https://github.com/Wendellab/Adansonia).

Table S1. Taxon sampling including sample source (herbarium), sequencing platform and

number of read pairs obtained per sample. Wisconsin State Herbarium (WIS); George Brown Darwin Botanical Gardens, Australia (GBDBG); University of Wisconsin – Madison, Department of Botany Greenhouse (UWBG), Missouri Botanical Garden Herbarium (MO)

Taxon	**Sample ID**	**Sequencing Platform**	**Number of read pairs**	**Source**
Adansonia digitata L.	Adi001	MiSeq	10,698,792	Accession # UW11 (UWBG)
Adansonia digitata L.	Adi002	MiSeq	10,785,012	Accession # UW2291 (UWBG)
Adansonia digitata L.	Adi003	MiSeq	10,947,592	See Cron et al. 2016; GenBank ID: KU145771
Adansonia grandidieri Baill.	Aga001	MiSeq	9,040,756	Accession #97-B002010-1 (GBDBG)
Adansonia grandidieri Baill.	Aga002	MiSeq	8,093,124	Accession # 03-B000192-1 (GBDBG)
Adansonia gregorii F.Muell.	Age001	MiSeq	10,627,128	North Western Australia, D.A.Baum
Adansonia perrieri Capuron	Ape001	MiSeq	10,627,128	Accessions # 92-B000060-1 (GBDBG)

Adansonia perrieri Capuron	Ape009	HiSeq	18,871,820	Northern Madagascar, Karimi-2014-09 (WIS)
Adansonia madagascariensis Baill.	Ama006	HiSeq	18,105,976	Northern Madagascar, Karimi-2014-006 (WIS)
Adansonia madagascariensis Baill.	Ama018	HiSeq	17,421,708	Northern Madagascar, Karimi-2014-018 (WIS)
Adansonia rubrostipa Jum. and Perr.	Aru001	MiSeq	8,890,992	Southwestern Madagascar, D.A.Baum 313 (MO)
Adansonia rubrostipa Jum. and Perr.	Aru127	HiSeq	16,620,060	Western Madagascar, Karimi-2014-127 (WIS)
Adansonia suarezensis H.Perrier	Asu001	MiSeq	11,343,772	Accession #UW11 (GBDBG): Seed from Northern Madagascar Baum 320A (WIS)
Adansonia za Baill.	Aza037	HiSeq	18,017,204	Northern Madagascar, Karimi-2014-37 (WIS)
Adansonia za Baill.	Aza135	HiSeq	18,420,364	Southern Madagascar, Karimi-2014-135 (WIS)

Bombax ceiba L.	Bce020	MiSeq	7,398,256	Accession #UW10 (UWBG)
Pseudobombax croizatii A.Robyns.	Pcr070	MiSeq	7,398,256	Accession #UW1255 (UWBG): Seed from Puerto Ayacucho in Venezuela, Paul E. Berry (MO)
Scleronema micrantha Ducke	Smi165	MiSeq	7,398,256	See Alverson et al. (1999); GenBank: AF111735

Plastome Assemblies

Reference guided assembly of off-target plastid reads obtained after hyb-seq were performed using Burrows Wheeler Aligner (Li and Durban 2009) with the *bwa mem* algorithm and the *Gossypium raimondii* plastome sequence as reference (NCBI GenBank ID: HQ325744). Consensus sequences of the mapped reads were extracted from the BAM alignments and aligned with MAFFT (Katoh et al. 2012) using the FFT-NS-2 algorithm. We tested for recombination blocks within the plastomes using MDL (Ané 2011) with a minimum block length of one hundred parsimony informative sites. For each of the resulting partitions and the total concatenated alignment jModelTest 2 (Darriba et al. 2012) was used to select a model of evolution, after which Bayesian phylogenetic inference was performed with MrBayes v3.2.3 (Huelsenbeck and Ronquist 2001; Ronquist and Huelsenbeck 2003). Analyses were run for 2 million generations with 4 runs, 4 chains and a heat of 0.2 with 25% of generations discarded as

burn-in. We also used RAxML version 8.2.10 (Stamatakis 2006, 2014) to infer maximum likelihood trees on 100 bootstrap replicates using the GTR-Γ model and the rapid hill climbing algorithm.

Read Assembly and Dataset Selection

Reads were assembled for an initial dataset of 14 accessions (see Table S1), including 12 samples sequenced on the Illumina MiSeq platform and two accessions from the HiSeq platform (*A. za* and *A. madagascariensis*). Subsequent datasets (see below) added four additional accessions, all sequenced on HiSeq (one additional *A. za*, *A. madagascariensis*, *A. rubrostipa*, and *A. perrieri*), for a total of 18 accessions. Assembly was performed with the HybPiper package (Johnson et al. 2016) using our initial target set as a reference.

For the first phase of assembly (see Supplementary Fig. S1 for flow chart), all genes that mapped to the targets and did not yield additional sequence variants in the assembly were retained for an initial dataset. For targets yielding a "paralog warning" for any taxon in HybPiper, meaning that more than one contig was assembled that covered >85% of the target, we analyzed gene trees to identify paralogs. Initially, we extracted all contigs for all 14 accessions for that target and aligned them using MAFFT version 7.299 (Katoh et al. 2012). Maximum likelihood phylogenetic trees were generated in Geneious version 8.0.5 (http://www.geneious.com, Kearse et al. 2012) using the RaxML plugin and were inspected to determine if the multiple contigs most likely represented alleles (accessions from the same species form a clade) or paralogs (there were distinct clades each with multiple species). For gene trees without evidence of paralogs, we retained a single sequence for each accession (selected automatically by HybPiper in the case of alleles). For alignments that suggested

paralogs, due to having two sequences per accession that formed two distinct partitions, we retained each sequence but split the alignment into two putative paralogs. This yielded a dataset of 282 genes and 14 accessions, hereafter referred to as dataset "282-14". Dataset "282-18" used the same gene set as "282-14", but added four additional (HiSeq) accessions such that species in the Malagasy *Longitubae* clade were each represented by two accessions.

Further Diagnosis of Paralogous Targets

Rather than discarding targets with complicated paralogy (phase two of assembly, Supplementary Fig. S1), we sought to separate paralogs into paralog-specific targets using a gene tree-based orthology approach. For targets with putative paralogs, a separate consensus sequence was generated for each paralog for each accession. These paralog-specific consensus sequences were then used as operational targets for a new assembly in HybPiper. If the resulting tree suggested yet further paralogs, the process was repeated iteratively until each target yielded either one sequence or a set of putative alleles per accession. To corroborate paralog calling, all contigs from all putative paralogs derived from the same original target were aligned to each other and a maximum likelihood tree was examined to confirm that the contigs assembled from different targets formed symmetrical subtrees, as would be expected for true paralogs. Forty-five targets for which we could not resolve orthology-paralogy were excluded from further analysis.

Preliminary network analyses suggested that Malagasy *Longitubae* samples sequenced on the HiSeq platform might share erroneous signals of gene flow, perhaps due to the assembly of distant paralogs that were not found using the shallower read-depth of the MiSeq platform. Two morphologically and geographically distinct species, *A. rubrostipa* and *A. perrieri*, were both represented by one accession generated on the MiSeq platform (Aru001; Ape001) and one on

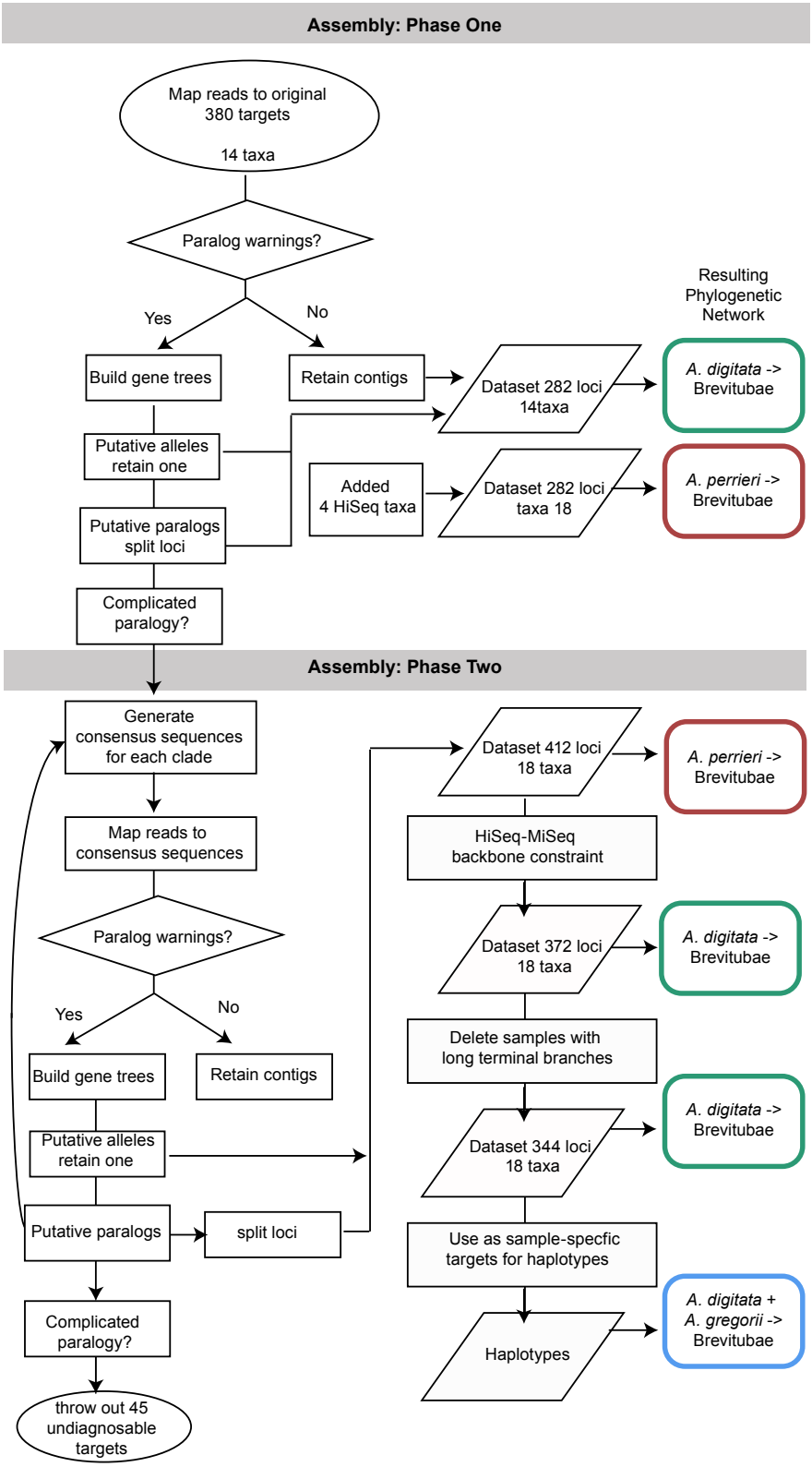
HiSeq (Aru127; Ape009). Therefore, we used PAUP* 4.0a (Swofford 2003) to identify all targets whose optimal gene trees (as estimated by maximum likelihood trees in RaxML; Stamatakis 2006) failed to satisfy the unrooted backbone constraint ((Aru001, Aru127),(Ape001, Ape009)). This resulted in a dataset of 372 genes, "372-18", which became our primary dataset for phylogenetic and network inference.

To continue to curate clean targets for haplotype assembly, the latter set of loci was pruned further, by removal of 28 loci whose optimal RaxML tree included single tips with very long branches (generally, at least five times longer than any other taxon, including outgroups), which might indicate misassembled sequences containing data properly belonging to a paralog, or alignments suggestive of bad read assembly. Genes with topologies that did not include a monophyletic *Adansonia* clade were also eliminated. This resulted in our most conservative dataset, "344-18".

In addition to analysis of the 344 gene trees from the HybPiper alignments, we also inferred haplotypes using this curated target set as reference. Trimmed reads were mapped to the accession-specific references using BWA v.0.7.15 (Li and Durban 2009) with the *bwa mem* algorithm. Haplotypes were inferred using HapHunt in BamBam (Page et al. 2014) under the following parameters: (1) 5 runs per accession; (2) a minimum of 20x coverage of each single nucleotide polymorphism (SNP); and (3) four haplotypes for the tetraploid *A. digitata* and two for all other taxa. We then generated a haplotype dataset which included 10 haplotype alignments for each gene, each alignment containing one sequence per accession, sampled at random from the haplotypes inferred for that accession.

To provide a baseline for comparison with curated datasets (addressed above), we also retained a dataset that was the result of direct read mapping to the original 380 targets for all 18

accessions (dataset “Original-baits”), ignoring any putative paralogs. This includes a single sequence for each sample, namely the assembly that had the greatest read depth or, if competing



Supplementary Figure S1. Read Assembly phases, datasets, and resulting phylogenetic networks.

contigs had similar depth, that with the highest percent identity to the original reference (Johnson et al. 2016).

Species-Tree Phylogenetic Analyses, Age estimation and Network Inference

Bayesian phylogenetic inference was performed on all datasets with MrBayes (Huelsenbeck and Ronquist 2001; Ronquist and Huelsenbeck 2003), as implemented in the TICR pipeline (Stenz et al. 2015), followed by Bayesian Concordance Analysis (Ané et al. 2007; Baum 2007), in BUCKy (Ané et al. 2007; Larget et al. 2011). BUCKy allows one to estimate the proportion of gene trees supporting a certain clade (the concordance factor, CF), while taking into account uncertainty in individual gene trees (Ané et al. 2007) and also estimates a population tree (="species tree") under the assumption that all discordance is due to incomplete lineage sorting (Larget et al. 2011). TICR complements this with a statistical test that assesses whether the concordance factors for all four-taxon sets are consistent with the multispecies coalescent on the optimal population tree, or a partial population tree with internal nodes collapsed to represent periods of panmixia in the past (Stenz et al. 2016). Rejection of the full or partial population tree model can be used to suggest a history of reticulation (Stenz et al. 2016).

MrBayes analyses used 4 linked chains with a heat of 0.2 and ran for two million generations, with 25% discarded as burn-in. The resulting posterior distributions were analyzed in BUCKy version 1.4.4 (Ané et al. 2006, Larget et al. 2010) using alpha=1 and 1,000,000 generations. For the nuclear haplotype dataset, we combined the set of 10 posterior distributions

generated with different allelic combinations for each locus. The quartet concordance factors output by BUCKy were used as input for the TICR test (Stenz et al. 2015).

To infer an explicit population network we used the maximum pseudolikelihood method implemented in SNaQ (Solís-Lemus and Ané 2016), which infers reticulate evolutionary histories while accounting for ILS. A starting population tree with branch lengths in coalescent units and the concordance factor table (generated by BUCKy in the TICR pipeline) were input into SNaQ to first optimize the starting tree and branch lengths for a network topology with zero hybridization edges (i.e., a tree). By default, TICR estimates a population tree using Quartet MaxCut. Although SNaQ optimizes input topology, to test for the possible effect of a different input tree for SNaQ, we also used the BUCKy population tree as the starting topology.

To avoid getting stuck on a local optimum, we routinely ran 50 independent runs of SNAQ for each dataset. The optimal network without hybridization edges (h0) was used as input for network searches with one hybridization edge (h1). Network searches were increased sequentially, up to 3 hybridization edges, starting from the previous (h1 or h2) optimal network. The preferred number of hybridizations was detected based on analysis of the slope of a plot of log-pseudolikelihood against hybridization number (Solís-Lemus and Ané 2016). The network with the best log-pseudolikelihood score for the optimal number of hybridizations (in each case one hybridization) was then selected as starting network for bootstrap analysis. We generated a total of 100 independent bootstrap replicates with 20 runs per replicate, where 10 runs/replicate started with the optimal hybridization network and the other 10 started with the h0 population tree.

Tests to Explain the Nuclear-Plastid Tree Discordance

To test for the possibility of two introgression events that would require SNaQ to move through intersecting cycles (Solís-Lemus and Ané 2017), which the program does not currently support, we pruned two outgroups, one *A. digitata* sample, and selected Malagasy taxa. Secondly, following the initial search for the optimal networks, we constructed a network that entails a *A. rubrostipa – A. za* hybridization edge and used that network as the start of a new search.

To test whether incomplete linage sorting (ILS), given the optimal network, could plausibly explain the recovered plastid phylogeny, we utilized a likelihood-based statistical significance test and applied it to the most-carefully curated dataset, "344-18". For the optimal network we multiplied branch lengths, which are in coalescent units, by 4 to account for the fact that plastid DNA experiences an effective population size ¼ of nuclear genes. External branches were set to an arbitrary length of 1. We then simulated 100,000 gene trees based on the network in the program hybrid-lambda (Zhu et al. 2015). Simulated tree topologies were input into PAUP* (version 4a; Swofford et al. 2003) along with the optimal plastid tree. We used tree filters to determine the frequency of targeted topologies among the 100,000 simulated trees. To see if all simulated trees could be rejected by the plastid data, tree likelihood scores of the optimal plastid tree and all topologies found by simulations were evaluated using one-tailed Approximately Unbiased (AU) and Shimodaira–Hasegawa (SH) tests with resampling of estimated log-likelihoods (RELL) and 10,000 bootstraps. These calculations assumed the GTR+ Γ model, with model parameters (transition:transversion bias and alpha) optimized individually for each tree.

Ancestral Trait Reconstruction

To infer the most likely flower color and pollinator of ancestral populations, we considered a Markov process for the evolution of these traits on the phylogenetic network. For pollinators, we considered a binary trait with 2 states (hawkmoth-pollinated and mammal-pollinated) evolving along each lineage according to a 2-state Markov process with 2 transition rates. For flower color, we considered a trait with 3 states (white, yellow, red) and all transitions possible with equal rates. We assumed an equal prior probability for all states at the root of the phylogenetic network. At a reticulation node in the network, we assumed that the trait of the hybrid population was inherited from either one of its parent populations, with probabilities equal to the proportion of genes contributed by each parent population. These inheritance probabilities are assumed to be those estimated from the genetic data with SNaQ. This model was already considered by Strimmer et al. (2001), who applied it to nucleotide data for the inference of ancestral recombination graphs. Like in Strimmer et al. (2001), we calculate the likelihood of the trait data as a linear combination of likelihoods from each tree displayed in the network. The transition rates between states were optimized with maximum likelihood, and ancestral reconstructions were obtained as the posterior probabilities of each state, given the trait data at the leaves of the phylogenetic network and the estimated rates. Our implementation is available in the open source Julia package PhyloNetworks (Solís-Lemus et al. 2017).

The trait evolution model used branch lengths in the network as a measure of evolutionary time. However, SNaQ infers branch lengths in coalescent units for internal branches, and does not infer any length for external branches (present-day populations). Therefore, we calibrated the branch lengths of the network using the approach developed by Bastide et al. (2018) and implemented in PhyloNetworks (Solís-Lemus et al. 2017). Briefly, we calculated the pairwise genetic distances between taxa from the individual gene trees, in which branch lengths measure the number of substitutions per site. To account for rate variation across

loci, the tree for each locus was re-scaled to a common median branch length (of 0.00162 substitutions / site / branch). The observed pairwise distance matrix was calculated by averaging the pairwise distances across all loci, for each pair of taxa. Ages of nodes in the network were optimized to yield phylogenetic distances that matched the observed pairwise distances as well as possible, according to the ordinary least-squares criterion. Ancestral trait reconstructions were then performed independently using the highest confidence nuclear dataset (344-18).

RESULTS

Targets and Dataset Selection

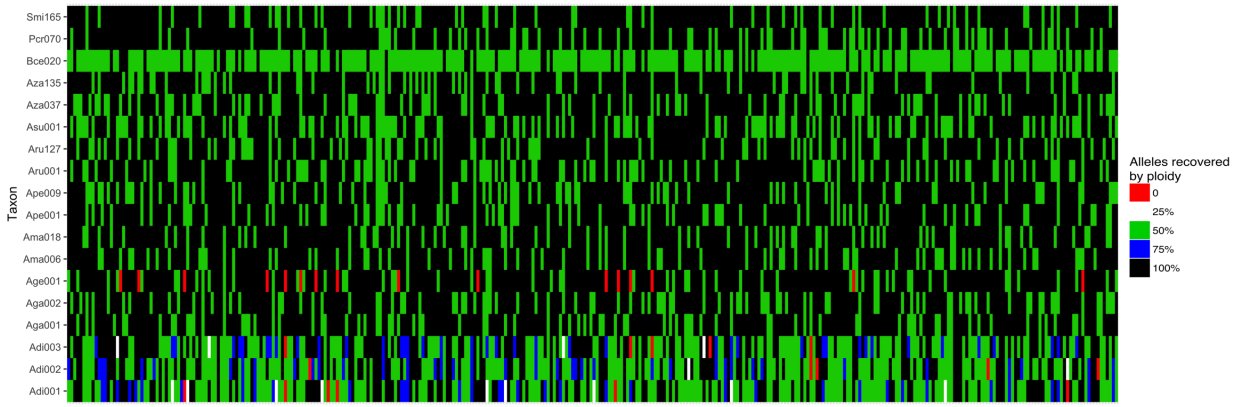
Analysis of transcriptomes recovered a total of 380 candidate sequences that met our criteria, for an initial total target space of 734,503 base pairs. The mean targeted locus was 1932.9 base pairs in length (minimum length of 763 bp and maximum of 7042 bp) with an average GC content of 42.8%. These targets are available as Supplementary File S1. An average of 9.3M and 17.9M read pairs were recovered for the accessions sequenced on the MiSeq and HiSeq, respectively. Although the MiSeq recovered an additional 100nt per read pair when sequenced as 2x300nt, the HiSeq still recovered an average of 3.4 Gb more sequence per sample when run as 2x250nt and multiplexed with 23 other samples on a lane of HiSeq2500 (Rapid Run). If all reads were on-target, this would represent an average of 5,000-12,000 fold coverage per nucleotide targeted for each sample. Allowing for the fact that about half of the sequences are off-target, we expected each targeted gene for each sample to have a minimum of 1000x coverage.

Following initial read mapping to the 380 gene bait set in HybPiper, we removed complicated genes and split paralogs as per the methods (see above), first leading to a total 282 loci. These include 244 inferred single copy loci, 32 paralogous genes (16 pairs), and 6 targets representing one of two inferred paralogs, where the other paralog was dropped from the dataset. Analyses were conducted with this set of genes for 14 taxa ("282-14"), including only the 12 accessions sequenced via MiSeq plus the two taxa that were only sequenced on HiSeq, and for all 18 taxa ("282-18").

Phase two of assembly sought to include targets with more complicated orthology/paralogy (Supplementary Fig. S1) resulting in our largest data set ("412-18"). We detected evidence that the increased coverage of sequencing in the HiSeq sequencing platform relative to the MiSeq platform resulted in paralogous sequences being assembled for subsets of HiSeq-based accessions, resulting in systematically misleading gene trees. Using the topological constraints described in the methods, we identified forty questionable loci, which were subsequently removed, generating our primary dataset ("372-18").

To develop a set of reference targets to be used as input for haplotype phasing, we dropped an additional 28 genes, as described in methods, and used the resulting 344 targets to call putative alleles in HapHunt. The "344-18" dataset used for haplotype assembly contained 218 single copy targets and 55 initial targets broken into 126 paralogous loci. Each paralog of this most conservative dataset had a median percent identity across all taxa of 89.75%. Haplotype recovery for the diploid *Adansonia* taxa resulted in an average of 73.5% loci were inferred as heterozygous, yielding two variant sequences. Across the three accessions of tetraploid *A. digitata*, 1.5% of loci were homozygous for a single allele, 50% yielded two variants, 12.5% yielded three, and 36% recovered four (Supplementary Figure S2). For each

gene we generated 10 alternative alignments, each comprising one randomly selected allele per accession, with downstream analyses integrating across these 10 subsampled datasets.



Supplementary Figure S2. Heatmap of recovery success for haplotype inference by HapHunt

Data Suggests Ancient Introgression

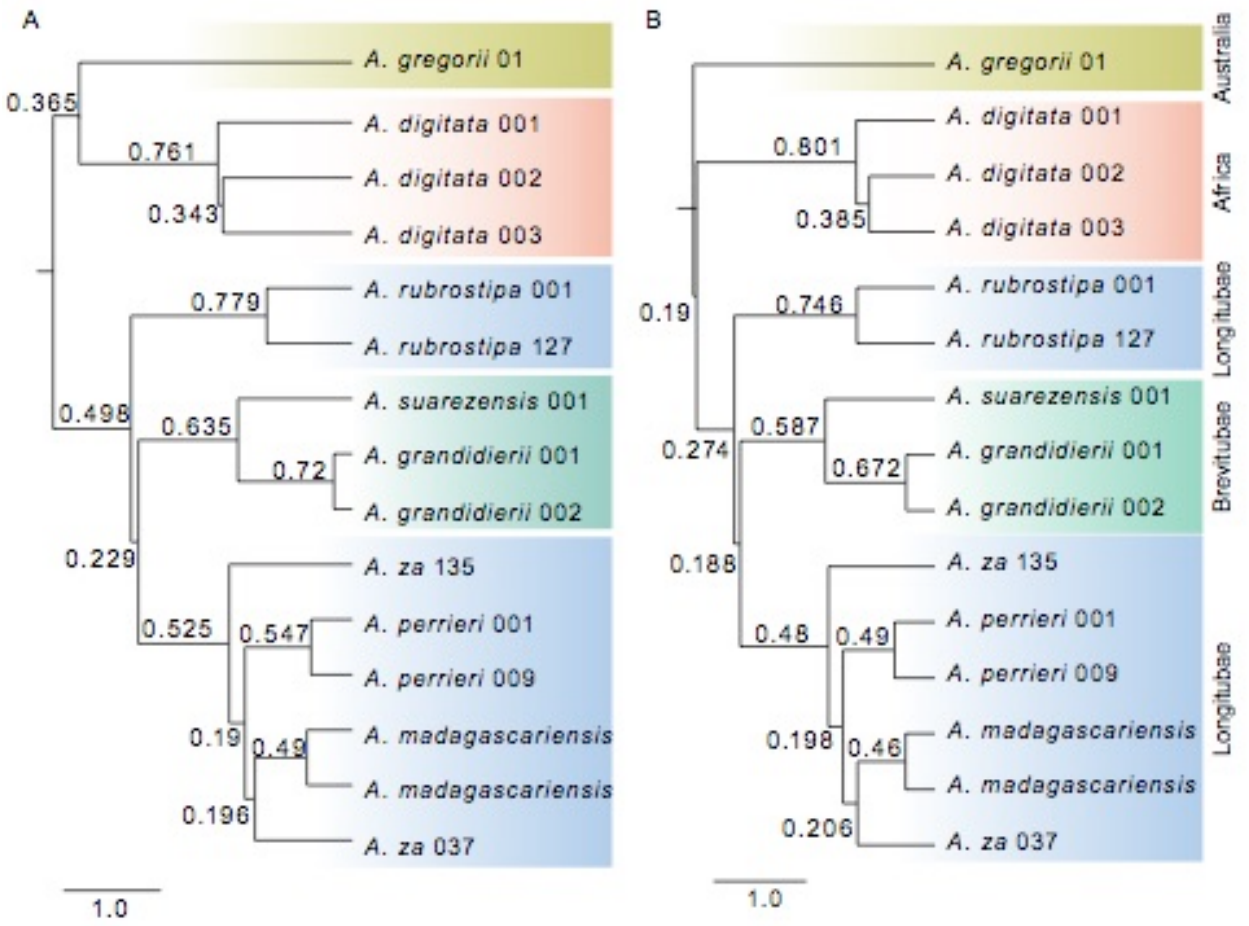


Figure 2. Population tree inferred with BUCKy from our (A) haplotype and (B) primary dataset with 372 genes. Trees scaled with branch lengths in coalescent units, and concordance factors added. Outgroups not shown.

Although multiple datasets were developed as described in the methods, here we report on the haplotype dataset and what we consider the most appropriately curated HybPiper dataset ("372-18"). In both cases, the primary population trees derived from BUCKy were identical to the inferred concordance trees (Fig. 2). Nonetheless, despite a large amount of data, relationships among the African, Malagasy, and Australian lineages were not resolved. Whereas the haplotype

dataset matched the plastid data (discussed further below) in supporting the Australia and African lineages as sister to each other (CF=0.365), the primary data set resolved the Australian species as sister to the rest of *Adansonia* (CF= 0.274). As expected, a clade of all three *A. digitata* accessions was strongly supported (CF= 0.761, CF=0.800), as was a clade of the six Malagasy species, though support is notably higher for this clade in the haplotype dataset (CF=0.498 vs. 0.274). For both nuclear data sets, monophyly of Sect. *Brevitubae* was supported (CF=0.635, CF=0.587), as expected based on morphology and previous molecular data. Somewhat unexpectedly, the data did not support a clade composed of all four Malagasy *Longitubae* taxa, but instead supported *A. rubrostipa* as sister to the rest of the Malagasy taxa (including Brevitubae). The concordance factor for a clade containing the four *Longitubae* species had a credibility interval that does not overlap that supporting a clade of *Brevitubae* plus "core" *Longitubae* (*A. madagascariensis*, *A. perrieri*, and *A. za*). This suggests, that despite many floral similarities, Sect. *Longitubae* is non-monophyletic or that some other process, such as reticulation, has influenced phylogenetic reconstruction.

The three species of core *Longitubae* (*A. madagascariensis*, *A. perrieri*, and *A. za)* are supported as a clade (CF= 0.525, 0.482). The two accessions of *A. za* are consistently resolved as non-monophyletic, with one *A. za* accession (from southern Madagascar) being sister to all sampled core *Longitubae* (all of which are from northern populations) including the other *A. za* accession. Similar non-monophyly on *A. za* was reported based on ITS analysis (Baum et al. 1998).

Network Inference

For the primary data set TICR rejected both fully and partially resolved tree-like structures due to an excess of outlier quartets, which suggests reticulated evolution (p=0.0053). The haplotype dataset, in contrast, rejected the tree model due to a *deficit* of outlier quartets (p=4.25e-44). The latter result we believe is an artifact of having manually curated the target set to remove anomalous topologies; as the dataset which was used references for HapHunt ("344-18") was which also was rejected (p=1.443886e-26) with this deficit.

Network inference in SNaQ found that a single hybridization event was optimal for all data sets. Our primary dataset "372-18" indicated introgression between the African lineage (the stem lineage of *A. digitata*) and the stem lineage of Sect. *Brevitubae*, with an estimated inheritance proportion of 12.4% (Fig. 3a). This hybridization edge was recovered in 73% of the bootstrap replicates on the major tree of the inferred network. The same network was inferred for our most conservative data set, "344-18", with the hybridization edge found in 95% of bootstrap replicates and having an inheritance proportion of 8% (Fig. 3a). While unexpected, introgression from the *A. digitata* lineage into *Brevitubae* is plausible given several homoplastic similarities between *A. digitata* and the *Brevitubae*, particularly in floral traits associated with mammal pollination.

The HapHunt-derived haplotype dataset also identified a network involving introgression between Sect. *Brevitubae* and a non-Malagasy taxon. However, the direction of introgression is reversed. Furthermore, perhaps because the population tree suggests a sister-group relationship between *A. digitata* and *A. gregorii*, the combined clade is identified (in 70% of bootstrap replicates) as potential recipient of genes (7.2%) from *Brevitubae* (Fig. 3a). Interpreted literally, this scenario is less likely for geographic and temporal reasons. However, given the difficulty of resolving the deepest splits in the genus and the fact that directionality sometimes cannot be determined reliably with SNaQ, (Solís-Lemus and Ané 2017), we interpret the haplotype data set

as being consistent with introgression from the *A. digitata* lineage into *Brevitubae* as obtained from the primary (372-18) and conservative (344-18) datasets.

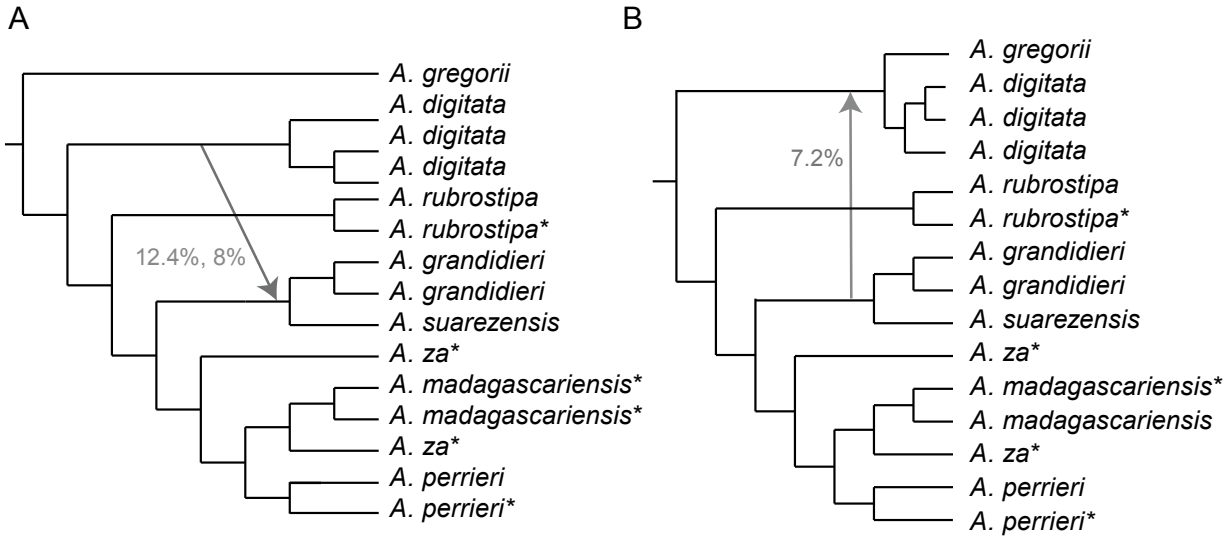


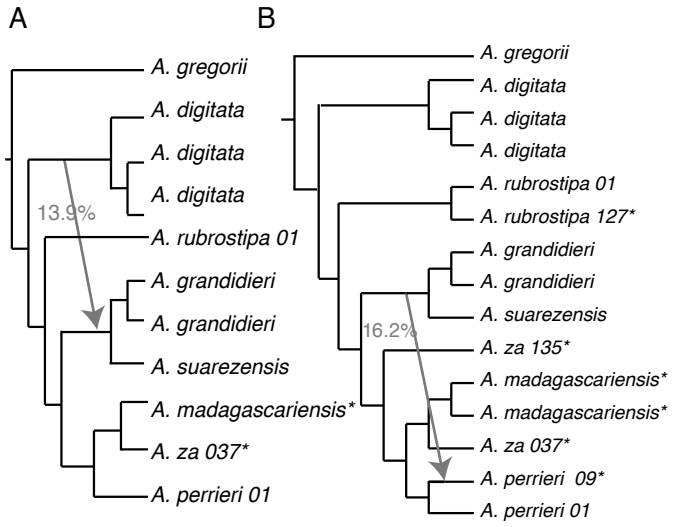
Figure 3. Phylogenetic networks inferred from SNaQ with A) primary datasets with 372 loci (12.4% inheritance) and 344 loci (8% inheritance), and B) inferred haplotypes.

Assembly Alters Phylogenetic Inference and Yields Artifactual Networks

Species network inference in SNaQ of datasets generated from phase one of assembly (see Supplementary Fig. S2) with 282 loci suggested the possibility of artifacts related to sequencing platform. Dataset "282-14", which was based only on MiSeq data, recovered an unexpected, but biologically plausible, hybridization edge from *A. digitata* into *Brevitubae*, with an inheritance probability of 13.9% (Fig. S2a). In contrast, dataset "282-18", which was identical but added four HiSeq-derived sequences, recovered an introgression event from the stem lineage of Brevitubae into one (MiSeq-derived) accession of *A. perrieri*, with an inheritance proportion 16.2% (Fig. S2b). *Adansonia perrieri* is the most geographically restricted and morphologically

cohesive taxon in the genus, making the latter result biologically implausible. A similar network was also seen in data set 412-18, indicating that such a network was an artifact of lingering paralog issues.

Since the *Brevitubae-perrieri* edge appeared when we added the additional HiSeq-derived accessions, we suspected that an artifact of gene assembly could be driving this result. Indeed, manually inspection of gene trees showed HiSeq-sequenced accessions showed an affinity as sister to each other rather than the same taxon counterpart. Targeted sequence capture baits can recover multiple paralogs for a single target, and it stands to reason that how many and which paralogs are assembled could be influenced by read depth. Initial read mapping of the original 380 reference targets resulted in 17 to 106 "paralog warning" flags from HybPiper per accession. Significantly, the average number of targets with paralog warnings was double for samples sequenced on the HiSeq platform (100 ±4 genes per accession) than for accessions sequenced on the MiSeq (44 ± 12 genes). This led us to suspect that even when paralog warnings are not issued, a sequence assembled for an accession is not the desired ortholog but a paralog.



To test the hypothesis that the *Brevitubae-perrieri* edge results from improper scoring of paralogs, we used a topology filter (see Methods) to remove targets that cluster accessions based on sequencing platform, to generate the primary dataset "372-18." We also removed entire targets or poorly aligned domains that contained signature of improper assemblies (i.e. long

terminal tips in HiSeq-derived accessions) to generate the data set "344-18". As described earlier, datasets 372-18 and 344-18 both strongly supported the origin *A. digitata-Brevitubae* hybridization and reject the *Brevitubae-perrieri* introgression, consistent with our hypothesis that assembly errors caused a false introgression edge to be inferred.

Plastid-Nuclear Incongruence Suggests Additional Reticulation Events Within Madagascar

Chloroplast sequences are relatively abundant in most whole genome sequencing libraries and are often recovered during sequence capture. Reference-guided assembly yielded nearly-complete plastid genomes for each taxon (number of reads assembled and read coverage is reported in Supplementary Table S2), with a shared alignment of 163,590 bp, containing 167-4,795 pairwise SNPs among taxa. Although it is generally assumed that the whole plastome has a single phylogenetic history, we first used MDL to identify possible recombination breakpoints. This recovered four partitions, representing 39.2, 21.7, 13.1, and 89.6 kilobases, respectively. All but the third partition (13.1 kb) support a combined Australian and African clade. Likewise, three partitions supported *A. suarezensis* as sister to the remaining Malagasy species, while partition two (21.7 kb) was unable to reject this arrangement, given bootstrap and posterior probabilities (Supplementary Fig. S2). As the conflicts among the plastid partitions are relatively minor and plastid recombination is unlikely, we used the concatenated plastid tree to represent the plastid history (Fig. 4). This tree strongly supports a clade composed of the African and Australian species and also implies non-monophyly of *Adansonia* Sect. *Brevitubae*, with *A. suarezensis* strongly supported as sister to the rest of the Malagasy clade. This is surprising given the morphological similarities shared between the two species of Sect. *Brevitubae* (*A. suarezensis* and *A. grandidieri*), including a distinctive crown architecture, white, upright

flowers with short-staminal tubes, winter flowering, large seeds, and cryptocotylar germination (Baum 1995a). Furthermore, both species are mammal-pollinated (Baum 1995a). It is worth noting that the plastid marker analyzed by Baum et al. (1998), *rpl16*, also supported a clade composed of all Malagasy species except *A. suarezensis,* which was sister to *A. digitata*. The plastid tree supports the monophyly of the four Malagasy Longitubae, which share elongated, yellow/red, primarily hawkmoth-pollinated flowers (*A. rubrostipa*, *A. za*, *A. perrieri*, and *A. madagascariensis*).

Supplementary Table S2. Plastid Genome Assembly Summary

Taxon (Sample ID)	**Number of reads**	**Length**	**Read coverage per position Mean,**
A. digitata (Adi001)	149,590	163,590	156, 160 (799)
A. digitata (Adi002)	187,363	165,284	200, 207 (1089)
A. digitata (Adi003)	64,640	163,583	58, 57 (764)
A. grandidieri (Aga001)	153,977	164,023	176, 179 (683)
A. grandidieri (Aga002)	128,584	164,895	144, 147 (858)
A. gregorii (Age001)	255,131	162,886	251, 206 (1237)
A. madagascariensis	723,721	174,559	1068, 1111 (1905)
A. madagascariensis	773,851	174,923	1139, 1218 (1753)
A. perrieri (Ape001)	112,439	165,008	115, 116 (886)
A. perrieri (Ape009)	469,746	171,757	689, 679 (2854)
A. rubrostipa (Aru001)	505,615	169,338	599, 625 (1245)
A. rubrostipa (Aru127)	880,514	175,580	1297, 1384 (2148)
A. suarezensis (Asu001)	228,382	166,192	256, 263 (1340)

A. za (Aza037)	230,053	169,304	332, 315 (2228)
A. za (Aza135)	412,706	171,535	604, 617 (1205)
Bombax ceiba (Bce020)	292,221	168,276	363, 379 (805)
Pseudobombax croizatii	43,136	162,438	35, 35 (516)
Scleronema micrantha	89,205	163,933	100, 102 (355)

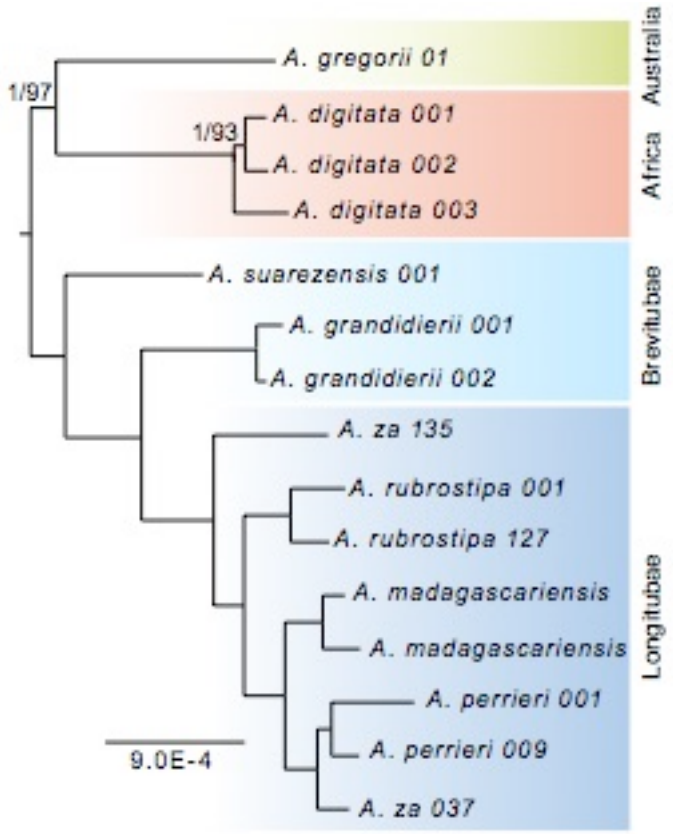


Figure 4: Concatenated plastid tree inferred by maximum likelihood and Bayesian phylogenetic inference. All branches have posterior probabilities (PP) of 1.0 and bootstrap support (BS) of 100%, unless otherwise indicated (PP/BS). Outgroups removed from figure.

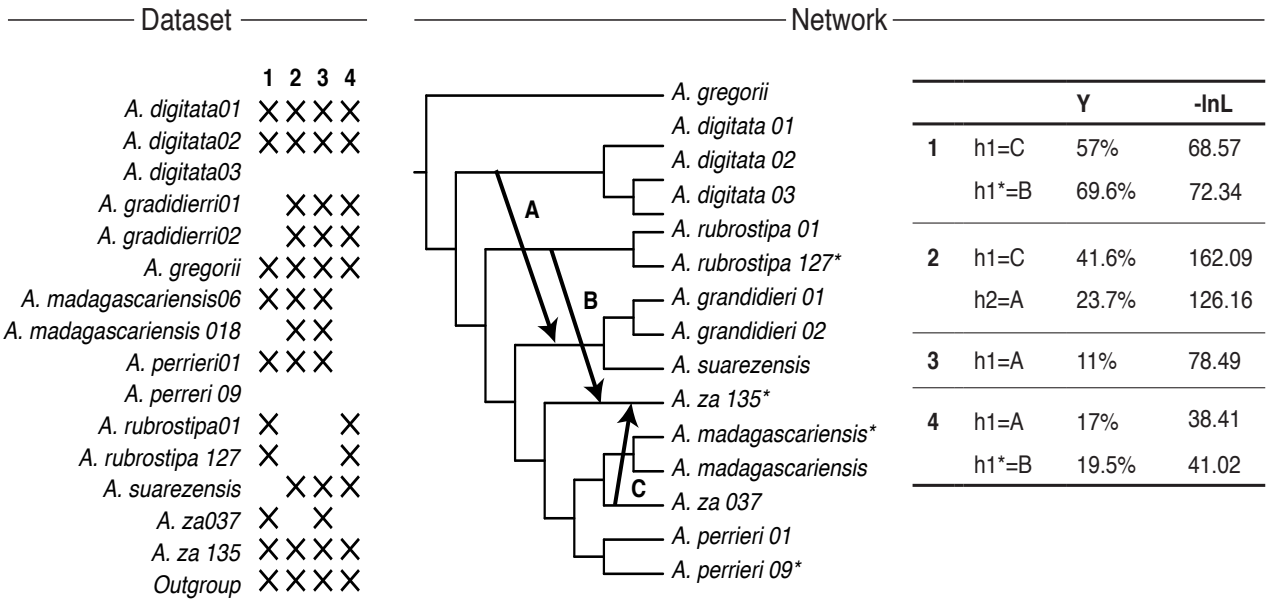
Testing Causes of Plastid-Nuclear Discordance

The plastid data were explained significantly better by the plastid maximum likelihood tree than any of the simulated gene trees derived from the optimal network from the primary data set (minimum difference of log-likelihood = 164.47). Of the 100,000 simulated trees, only one tree supported *A. suarezensis* as sister to the rest of the Malagasy clade. This suggests that the probability of *A. suarezensis* being sister to all other Malagasy species by incomplete lineage sorting alone is very low ($p=10^{-5}$).

The plastid tree also differs from the nuclear population tree in supporting southern *A. za* as sister to all Malagasy *Longitubae*, including *A. rubrostipa*. None of the simulated trees have this resolution, suggesting the possibility of additional reticulation events that affected the plastid genome. In this regard it is worth noting that after deleting both *A. suarezensis* and the two *A. za* accessions, only 22 simulated trees out of unique 2268 trees (2113 trees were discarded from the original 4381 due to redundancy after pruning taxa) cannot be rejected relative to the observed tree at the p=0.05 level. After deleting both *A. suarezensis* and the two *A. rubrostipa* accessions, 22 simulated tree topologies out of 2124 (2257 discarded due to redundancy) cannot be rejected relative to the observed tree at the p=0.05 level. Deleting *A. rubrostipa* and *A. za,* only two simulated trees out of 172 could not be rejected (3451 redundant trees discarded).

Evidence of A. rubrostipa - A. za Ancient Hybridization

Given the phylogenetic placement of *A. rubrostipa* as sister to the Malagasy clade on the nuclear and plastid phylogenies, we specifically tested for introgression involving Longitubae taxa. Because of SNaQ's inability to detect intersecting cycles, we eliminated selected accessions from the most curated dataset "344-18" and reran SNaQ (Fig 5). Deletion of *Brevitubae* taxa results in an introgression event from *A. rubrostipa* to *A. za* being recovered, with an inheritance proportion of 19.5% (Fig 5.). In datasets that included the *Brevitubae* and *A. digitata* clades, this event was sometimes recovered, though with lower likelihood than the *A. digitata* to *Brevitubae* event. When used the latter starting topology for the h1 search, the *A. rubrostipa* to *A. za* 135 hybridization event was again recovered (Fig. 5) Two pruned datasets yielded an optimal network with a *A. za* - *A. za* hybridization edge (with the *A. digitata-Brevitubae* event being added in the h2 search).



		Y	-lnL
1	h1=C	57%	68.57
	h1*=B	69.6%	72.34
2	h1=C	41.6%	162.09
	h2=A	23.7%	126.16
3	h1=A	11%	78.49
4	h1=A	17%	38.41
	h1*=B	19.5%	41.02

Figure 5. (Left) Four datasets and taxa selected for each subset using 344 loci. (Right) SNaQ results for each corresponding dataset. Resulting h1 networks with stars indicate results from altering the starting topology (see text).

Ancestral Trait Reconstruction: Introgression Explains the Shift in Pollination Syndrome in Sect. Brevitubae

Given the hypothesis of gene flow from the mammal-pollinated *A. digitata* lineage into an ancestor of the mammal-pollinated *Brevitubae* clade, we sought to explore whether reticulation may help explain variation in floral morphology and pollination system seen across *Adansonia* (Fig. 6). It has previously argued been that baobabs (Baum et al. 1998), are ancestrally hawkmoth-pollinated. Mapping current pollination onto the optimal tree topology suggests two independent origins of mammal pollination in the group, once occurring along the lineage of the African baobab and another in the lineage of Sect. *Brevitubae* (Fig. 6). This transition is consistent with and could have been facilitated by the gene flow event recovered between these two lineages. Using a time-calibrated network (see methods) the rate of transition from mammal to hawkmoth was estimated to be α=0.0 as contrasted to a rate of the hawkmoth to mammal transitions (β=35.7637). Furthermore, the optimal inference entails mammal pollination being acquired in *Brevitubae* through the hybridization edge. Similarly, it is plausible, if not so well supported, that there was a reversal from yellow to white flowers at around the same time that the ancestor of *Brevitubae* switched to mammal pollination, suggesting that the evolution of these floral traits was triggered by introgression.

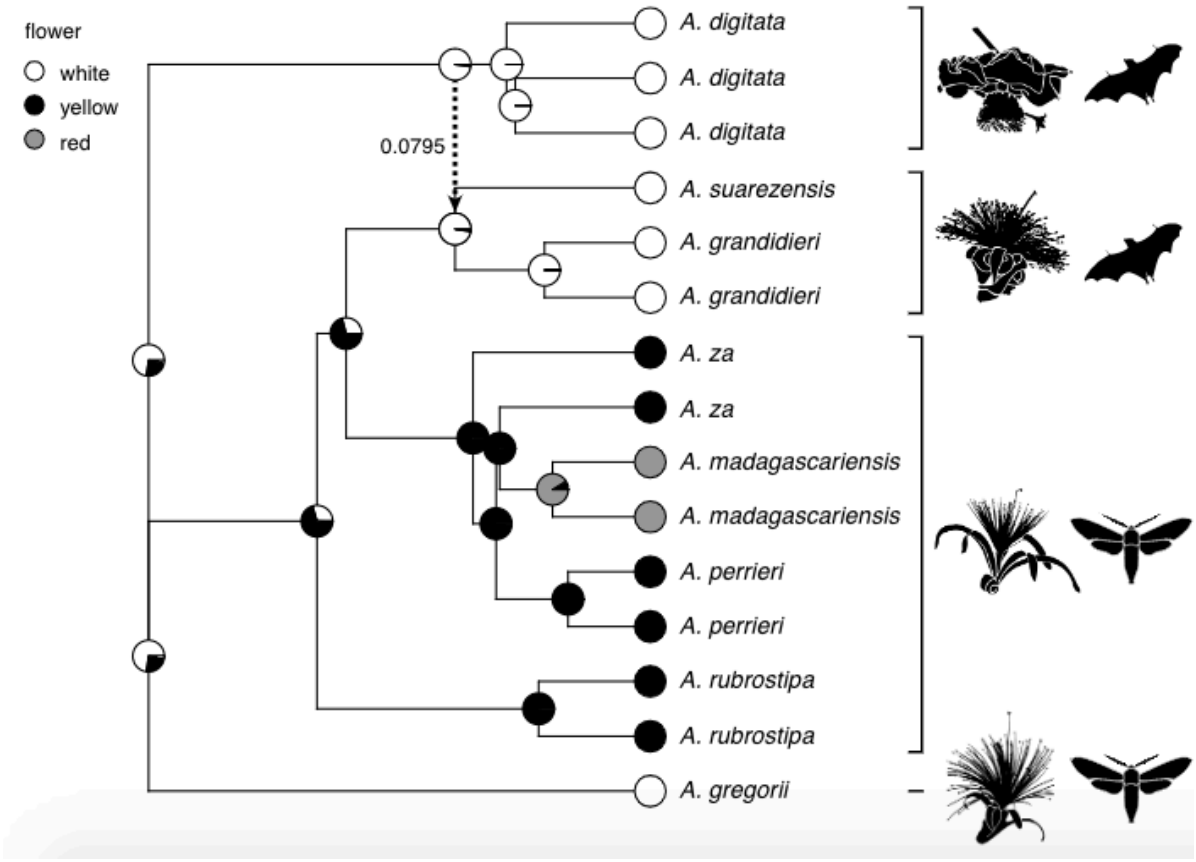


Figure 6. Ancestral state reconstruction of flower color inferred from the nuclear dataset. Current pollination systems added to each taxon on the right.

DISCUSSION

Targeted Sequence Capture offers new tools for inferring reticulate evolution

While caution is needed in assembling the alignments that form the basis of multilocus phylogenetic analysis, especially given the high likely of capturing paralogs at least in plants, it was still possible in our study to find strong evidence for two reticulation events. Such insights cannot easily be extracted from SNP-based methods, nor from analyses with few loci, justifying

the added analytical challenges of full locus sequencing. Furthermore, SNP-based methods do not allow for even the possibility of recovering annotated genes from the capture pool nor of obtaining organellar sequences. Analytic pipelines for dealing with putative paralogs, polyploid species, and identifying possible artifacts will hopefully improve with increasing utilization of such data. Also, we may hope that diagnostics will be improved for dealing with intersecting hybridization cycles and for adding an explicit temporal component to explicit networks. Although full Bayesian and Maximum Likelihood approaches are available (i.e. Phylonet, Than et al. 2008) these approaches are still limiting with this many loci and taxa.

Spurious Hybridization Events as Artifacts of Paralogy and Data Processing

Even though network inference was confounded by read assembly and paralogy errors, we were able to confidently understand the source of some obviously conflicting results. Although manual curation of gene trees and alignments may have resulted in overly clean data, to the point where variation was no longer statistically random (as determined by the TICR test results), it was clear that datasets generated by careful curation produced a consistent and biologically reasonable signal whereas raw datasets provided by HybPiper did not. Furthermore, given the *A. digitata-Brevitubae* hybridzation event did not disappear in other datasets, but was "swamped out" (such networks had higher log likelihood scores) by the erroneous reticulation event that was an artifact of paralog-calling issues supports our conclusions. SNaQ networks recovered from datasets with and without the four HiSeq samples (e.g., 282-14 vs. 282-18) suggests that our ability to decipher variants is masked (Sim et al. 2014) when paralog diagnosis is performed with samples of shallow sequencing depth. This is highlighted in the assembly statistics, in which HiSeq accessions had over twice as many variant sequences per target locus than the

MiSeq accessions. Most notably the dataset that ignored all paralog warnings, "Original-baits" contained loci of mixed paralogous sequences in the alignment. Discrepancies between sequencing platforms for identifying sequence variants (i.e. putative paralogs, alleles) during read assembly were amplified (average number of genes with alternate contigs found per sample: MiSeq=6, HiSeq=52), underscoring the need for deep sequencing for each taxon per locus versus more loci at shallower sequencing depths. This is not surprising given that additional taxa per clade and deeper sequencing has been shown to improve the ability to detect hidden paralogy (Philippe et al. 2011).

Causes of Plastid-Nuclear Tree Discordance in Malagasy Taxa

The primary population tree derived from nuclear genes was in conflict with the plastid tree, due to the placement of two species, *A. suarezensis* (member of Sect. *Brevitubae*) and *A. rubrostipa* (Sect. *Longitubae*). Such discordance might occur from various phenomena including chloroplast capture, hybridization, or ILS.

Although chloroplast capture via hybridization can be a source of nuclear-organellar discordance, this alone does not adequately explain the plastid topology we recovered with respect to the placement of *A. suarezensis* (the other member of strongly-supported monophyletic Sect. *Brevitubae*) as there is no hybridization scenario in which *A. suarezensis* would acquire a plastid that is sister to the rest of the Malagasy baobab clade. Given that plastomes are most often assumed to behave as a single non-recombining locus, they too are subject to ILS. However, results of our topological tests suggest that ILS alone could not result in the observed plastid tree, as demonstrated by the simulation of gene trees given the hybridization network. This suggests that other evolutionary forces such as selection may have contributed in

conjunction with chloroplast capture. If one considers the possibility of balancing selection in spatially heterogeneous environments, maintenance of a diversity in plastid haplotypes is feasible. Although poorly demonstrated, the maintenance of multiple plastid genotypes has been reported (Gurdon and Maliga 2014).

The plastid and nuclear discordance also suggests a history of gene flow between *A. rubrostipa* and *A. za* of Sect. *Longitubae*. This pattern of gene flow is in agreement with previous work that has shown introgression between these two taxa based on nuclear microsatellite loci (Leong Pock Tsy et al. 2013). Given this, we expected to recover evidence of hybridization between these species of Sect. *Longitubae* in the network analyses; however, no such networks were recovered in h2 searches. We attribute this to a number of factors. First, any potential paralogs that appeared exclusive to Sect. *Longitubae* were eliminated as we were concerned that such loci were artifacts of the MiSeq shallower sequencing not returning variants needed to recognize paralogous loci in all other taxa of the genus. However, the *A. rubrostipa* - *A. za* hybridization network was identified in the h1 search, though with less-optimal likelihood scores compared to the *A. digitata* - *Brevitubae* network. This would not be recovered in the search for a network with two hybridization edges by SNaQ, due to the method's assumption of a level-1 network. A network with both events would be of level 2, because the two event creates cycles that overlap (they share branches) (Solís-Lemus and Ané, 2017). This reticulation event was only recovered when used as the input network for the starting search.

Overwater dispersal facilitated ancient introgression

Our phylogenetic analyses of *Adansonia* support a rapid divergence into three geographic lineages involving multiple gene flow events across the Mozambique channel. This is plausible

given that many lineages endemic to Madagascar are of African origin (Haber et al. 2017; Yoder et al. 1996), thus multiple long-distance dispersal events to Madagascar within a lineage are not unlikely (Kainulainen et al. 2017). The inferred introgression event between the African lineage (*A. digitata)* into the Malagasy clade is presumed to be quite ancient given that it likely predated the origin of tetraploidy in *A. digitata* and also the divergence of the Malagasy *Brevitubae* clade into the extant *A. grandidieri* and *A. suarezensis* lineages (both diploids). This dispersal event was most likely due to fruit dispersal since pollen longevity is unlikely to be insufficient for long distance dispersal (even considering pollinator movement). Most species of baobab have fruit that are not only adapted to endozoochory via large bodied animals (Andriantsaralaza et al. 2010; Albert-Daviaud et al. 2018), but are also buoyant, potentially well-suited to hydrochory (Cornu et al. 2014) and have been shown to have viable seeds after extended seawater immersion (Leong Pock Tsy et al. 2009). This fact has been highlighted in helping to explain the dispersal of baobabs to Australia (Baum et al. 1998). Also, *A. digitata* fruit are known to be good dispersers, being found naturally on some islands (e.g., Cape Verde). They have also been reported to have washed up on Aldabra, which is just as far from the African coastline as Madagascar (Wickens 2008). It is worth noting that in Madagascar, specifically uninhabited areas of Analabe Peninsula (northwest Madagascar), there are a few long-term, naturalized populations of *A. digitata* (Leong Pock Tsy et al 2009) presumed to be introduced extant tetraploids, and some authors have suggested that some populations are derived from natural rather than anthropogenic colonization (Leong Pock Tsy et al 2009).

Our data suggest that perhaps 10% of the genome of extant *Brevitubae* descended from such an ancient *A. digitata* mariner. We may hypothesize that a single tree established, became reproductively mature, crossed with a local population and generated hybrid genotypes that were sufficiently favored or selected by a class of pollinators such that a significant fraction of the

recipient species' genome was replaced. Alternatively, perhaps a population of *A. digitata* established in Madagascar before hybridizing with resident species and has since been extirpated. Gene flow between the two lineages is plausible despite that the main pollinators of the two ancestral species likely being different, since nocturnal lemurs (primary pollinators of extant Brevitubae) can pollinate species that are primarily adapted for hawkmoth pollination, the pollinators of the Longitubae (Baum 1995a). This example illustrates that plants can undergo long-distance dispersal and then interbred with a genetically compatible recipient. While to our knowledge there is limited evidence of transoceanic dispersals and hybridization, this has famously been shown in cotton for which the allopolyploid cotton lineage of the Americas resulted from trans-oceanic dispersal of an A-genome taxon from Africa or Asia into the New World followed by hybridization with an indigenous American D-genome diploid (Wendel and Cronn 2003 and references therein)

Adaptive introgression of pollination traits

The ancestral trait reconstruction analysis suggests that the introgression of *A. digitata* genetic material into Sect. *Brevitubae* may have been a key factor in the latter lineage evolving mammal pollination and floral traits associated with such pollination syndrome. While some example of introgression facilitating pollinator shifts have been reported (i.e. Louisiana irises, Martin et al. 2006; Moneyflowers, Stankowski and Streisfeld 2015), this is the first case we are aware of that involves a hawkmoth to mammal transition. As a result of the introgression it seems there was a reversal to an all-white flower and changes in nectar volume, nectar chemistry, and scent profile (as summarized by Baum 1995a). Either the ancestral lineage leading to *A. digitata* had not yet evolved pendulous flowers, with wide, reflexed petals, or these traits failed to introgress.

Hybridization has been shown to facilitate floral isolation in incipient species (Chase et al., 2010).

Assuming that additional gene tree analyses to confirm the hybridization hypothesis, the Malagasy *Brevitubae* would be a prime example of a pollinator switch being facilitated by introgressive hybridization. We may hope, furthermore, that further genomic analyses would allow us to identify specific genes that were introgressed and see if they include candidate genes for various traits involved in differences in the observed pollination syndromes, such as flower color, nectar characteristics or floral scent chemistry.

FUNDING

This project is based on work supported by the National Science Foundation award DEB-1354268.

ACKNOWLEDGMENTS

We thank Steve Goldstein, Noah Stenz, and the Center for High Throughput Computing at the University of Wisconsin- Madison provided computational assistance and Claudia Solís-Lemus who advised on network analyses. Computational support at Iowa State University was provided through the ResearchIT Unit (https://researchit.las.iastate.edu/). Ben Wirth and George Brown Darwin Botanic Gardens kindly provided samples and Diana Mayne provided one of the *A. digitata* samples. Alison Dawn Scott conducted RNA extractions, and the University of Wisconsin Biotechnology Center DNA Sequencing Facility provided RNA-Seq library preparation and sequencing services, and assistance with bait design.

REFERENCES

Albert-Daviaud, Aurélie, Sarah Perillo, and Wolfgang Stuppy. "Seed dispersal syndromes in the Madagascan flora: the unusual importance of primates." *Oryx* (2018): 1-9.

Andriafidison, D., Andrianaivoarivelo, R.A., Ramilijaona, O.R., Razanahoera, M.R., MacKinnon, J., Jenkins, R.K. and Racey, P.A., 2006. Nectarivory by Endemic Malagasy Fruit Bats During the Dry Season 1. *Biotropica: The Journal of Biology and Conservation*, *38*(1), pp.85-90.

Andriantsaralaza, Seheno, Miguel Pedrono, Jacques Tassin, Roger Edmond, and Pascal Danthu. 2010. Baobabs de Madagascar: un anachronisme de la dispersion?. *Bois et Forêts des Tropiques*, 360:7-15.

Ané, C., 2011. Detecting phylogenetic breakpoints and discordance from genome-wide alignments for species tree reconstruction. *Genome biology and evolution*, 3:246-258.

Ané, C., Larget, B., Baum, D.A., Smith, S.D. and Rokas, A., 2006. Bayesian estimation of concordance among gene trees. *Molecular biology and evolution*, *24*(2), pp.412-426.

Bastide, P., Solis-Lemus, C., Kriebel, R., Sparks, K.W. and Ané, C., 2018. Phylogenetic comparative methods on phylogenetic networks with reticulations. *Systematic Biology*. *67*(5) pp. 800-820.

Baum, D.A., 1995a. The comparative pollination and floral biology of baobabs (Adansonia-Bombacaceae). *Annals of the Missouri Botanical Garden*, *82*(2), pp.322-348.

Baum, D.A., 1995b. A systematic revision of Adansonia (Bombacaceae). *Annals of the Missouri Botanical Garden*, *82*(3), pp.440-471.

Baum, D.A., DeWitt Smith, S., Yen, A., Alverson, W.S., Nyffeler, R., Whitlock, B.A. and Oldham, R.L., 2004. Phylogenetic relationships of Malvatheca (Bombacoideae and Malvoideae; Malvaceae sensu lato) as inferred from plastid DNA sequences. *American Journal of Botany*, *91*(11), pp.1863-1871.

Baum, D.A., Small, R.L. and Wendel, J.F., 1998. Biogeography and floral evolution of Baobabs Adansonia, Bombacaceae as inferred from multiple datasets. *Systematic Biology*, *47*(2), pp.181-207.

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics*, btu170.

Chang, S., Puryear, J. and Cairney, J., 1993. A simple and efficient method for isolating RNA from pine trees. *Plant molecular biology reporter*, *11*(2), pp.113-116.

Chase, M.W., Paun, O. and Fay, M.F., 2010. Hybridization and speciation in angiosperms: a role for pollinator shifts?. *BMC biology*, *8*(1), p.45.

Chau, J.H., Rahfeldt, W.A. and Olmstead, R.G., 2018. Comparison of taxon-specific versus general locus sets for targeted sequence capture in plant phylogenomics. *Applications in plant sciences*, *6*(3), p.e1032.

Cornu, Cyrille, Wilfried Ramahafaly, and Pascal Danthu. 2014. *Adansonia madagascariensis*, a marine hydrochory hypothesis. *Bois et Forets des Tropiques.* 320: 7-14.

Cron, G.V., Karimi, N., Glennon, K.L., Udeh, C.A., Witkowski, E.T., Venter, S.M., Assogbadjo, A.E. and Baum, D.A., 2016. One African baobab species or two? Synonymy of Adansonia kilima and A. digitata. *Taxon*, *65*(5), pp.1037-1049.

Darriba, D., Taboada, G.L., Doallo, R. and Posada, D., 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature methods*, 9:772-772.

Ekblom, R. and Galindo, J., 2011. Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity*, *107*(1), p.1.

Fér, T. and Schmickl, R.E., 2018. HybPhyloMaker: Target Enrichment Data Analysis From Raw Reads to Species Trees. *Evolutionary Bioinformatics*, *14*, p.1176934317742613.

Grover, C.E., Salmon, A. and Wendel, J.F., 2012. Targeted sequence capture as a powerful tool for evolutionary analysis1. *American journal of botany*, *99*(2), pp.312-319.

Grover, C.E., Gallagher, J.P., Jareczek, J.J., Page, J.T., Udall, J.A., Gore, M.A. and Wendel, J.F., 2015. Re-evaluating the phylogeny of allopolyploid Gossypium L. *Molecular phylogenetics and evolution*, *92*, pp.45-52.

Grover, C. E., J. P. Gallagher, E. P. Szadkowski, J. T. Page, M. A. Gore, J. A. Udall and J. F. Wendel. 2017. Nucleotide diversity in the two co-resident genomes of allopolyploid cotton. *Plant Systematics and Evolution* DOI 10.1007/s00606-017-1411-1.

Gurdon, C.S. and Maliga, P.A.L., 2014. Two distinct plastid genome configurations and unprecedented intraspecies length variation in the accD coding region in Medicago truncatula. *DNA research*, *21*(4), pp.417-427.

Haber, E.A., Kainulainen, K., Van Ee, B.W., Oyserman, B.O. and Berry, P.E., 2017. Phylogenetic relationships of a major diversification of Croton (Euphorbiaceae) in the western Indian Ocean region. *Botanical Journal of the Linnean Society*, *183*(4), pp.532-544.

Heled, J. and Drummond, A.J., 2009. Bayesian inference of species trees from multilocus data. *Molecular biology and evolution*, *27*(3), pp.570-580.

Hart, M.L., Forrest, L.L., Nicholls, J.A. and Kidner, C.A., 2016. Retrieval of hundreds of nuclear loci from herbarium specimens. *Taxon*, *65*(5), pp.1081-1092.

Harvey, M.G., Smith, B.T., Glenn, T.C., Faircloth, B.C. and Brumfield, R.T., 2016. Sequence capture versus restriction site associated DNA sequencing for shallow systematics. *Systematic biology*, *65*(5), pp.910-924.

Hellmuth, M., Wieseke, N., Lechner, M., Lenhof, H.P., Middendorf, M. and Stadler, P.F., 2015. Phylogenomics with paralogs. *Proceedings of the National Academy of Sciences*, *112*(7), pp.2058-2063.

Huelsenbeck, J.P. and Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, 17:.754-755.

Johnson, M.G., Gardner, E.M., Liu, Y., Medina, R., Goffinet, B., Shaw, A.J., Zerega, N.J. and Wickett, N.J., 2016. HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Applications in plant sciences*, 4:1600016.

Kainulainen, K., Razafimandimbison, S.G., Wikström, N. and Bremer, B., 2017. Island hopping, long-distance dispersal and species radiation in the Western Indian Ocean: historical biogeography of the Coffeeae alliance (Rubiaceae). *Journal of Biogeography*, *44*(9), pp.1966-1979.

Kamneva, O.K., Syring, J., Liston, A. and Rosenberg, N.A., 2017. Evaluating allopolyploid origins in strawberries (Fragaria) using haplotypes generated from target capture sequencing. *BMC evolutionary biology*, *17*(1), p.180.

Katoh, K., Misawa, K., Kuma, K.I. and Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic acids research*, 30:3059-3066.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Mentjies, P., and Drummond, A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data.*Bioinformatics*, 28:1647-1649.

Kubatko, L.S., Carstens, B.C. and Knowles, L.L., 2009. STEM: species tree estimation using maximum likelihood for gene trees under coalescence. *Bioinformatics*, *25*(7), pp.971-973.

Larget, B.R., Kotha, S.K., Dewey, C.N. and Ané, C., 2010. BUCKy: gene tree/species tree reconciliation with Bayesian concordance analysis. *Bioinformatics*, 26:2910-2911.

Leong Pock Tsy, J.M., Lumaret, R., Flaven-Noguier, E., Sauve, M., Dubois, M.P. and Danthu, P., 2013. Nuclear microsatellite variation in Malagasy baobabs (Adansonia, Bombacoideae, Malvaceae) reveals past hybridization and introgression. *Annals of botany*, *112*(9), pp.1759-1773.

Leong Pock Tsy, J.M., Lumaret R., Mayne D., Vall A.O., Abutaba Y.I., Sagna M., Raoseta S.N., Danthu P. Chloroplast DNA phylogeography suggests a West African centre of origin for the

baobab, *Adansonia digitata* L.(Bombacoideae, Malvaceae). Molecular Ecology. 2009 Apr;18(8):1707-15.

Li, H. and Durbin, R., 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, 25:1754-1760.

Liston, A., Weitemier, K., Fishbein, M., McDonnell, A., Straub, S.C., Schmickl, R. and Cronn, R.C., 2014. Hyb-Seq: Combining target enrichment and genome skimming for plant phylogenomics.

Liu, L., Yu, L. and Edwards, S.V., 2010. A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. *BMC evolutionary biology*, *10*(1), p.302.

Martin, N.H., Bouck, A.C. and Arnold, M.L., 2006. Detecting Adaptive Trait Introgression between Iris fulva and Iris brevicaulis in highly-selective field conditions. *Genetics*.

Mirarab, S., Reaz, R., Bayzid, M.S., Zimmermann, T., Swenson, M.S. and Warnow, T., 2014. ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics*, *30*(17), pp.i541-i548.

Nicholls, J.A., Pennington, R.T., Koenen, E.J., Hughes, C.E., Hearn, J., Bunnefeld, L., Dexter, K.G., Stone, G.N. and Kidner, C.A., 2015. Using targeted enrichment of nuclear genes to increase phylogenetic resolution in the neotropical rain forest genus Inga (Leguminosae: Mimosoideae). *Frontiers in Plant Science*, *6*, p.710.

Page, J.T., Liechty, Z.S., Huynh, M.D. and Udall, J.A., 2014. BamBam: genome sequence analysis tools for biologists. *BMC Research Notes*, *7*(1), p.829.

Pamilo, P. and Nei, M., 1988. Relationships between gene trees and species trees. *Molecular biology and evolution*, *5*(5), pp.568-583.

Park, H. J., and Nakhleh, L. (2012). Inference of reticulate evolutionary histories by maximum likelihood: the performance of information criteria. BMC bioinformatics, 13(Suppl 19), S12.

Pettigrew, F.R.S., Jack, D., Bell, K.L., Bhagwandin, A., Grinan, E., Jillani, N., Meyer, J., Wabuyele, E. and Vickers, C.E., 2012. Morphology, ploidy and molecular phylogenetics reveal a new diploid species from Africa in the baobab genus Adansonia (Malvaceae: Bombacoideae). *Taxon*, *61*(6), pp.1240-1250.

Philippe, H., Brinkmann, H., Lavrov, D.V., Littlewood, D.T.J., Manuel, M., Wörheide, G. and Baurain, D., 2011. Resolving difficult phylogenetic questions: why more sequences are not enough. *PLoS biology*, *9*(3), p.e1000602.

Pross, J., Contreras, L., Bijl, P.K., Greenwood, D.R., Bohaty, S.M., Schouten, S., Bendle, J.A., Röhl, U., Tauxe, L., Raine, J.I. and Huck, C.E., 2012. Persistent near-tropical warmth on the Antarctic continent during the early Eocene epoch. *Nature*, *488*(7409), p.73.

Raymond, O., Piola, F. and Sanlaville-Boisson, C., 2002. Inference of reticulation in outcrossing allopolyploid taxa: caveats, likelihood and perspectives. *Trends in Ecology and Evolution*, *17*(1), pp.3-6.

Ronquist, F. and Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 19:1572-1574.

Philippe Ryckewaert, Onja Razanamaro, Elysée Rasoamanana, Tantelinirina Rakotoarimihaja, Perle Ramavovololona, et al.. Les Sphingidae, probables pollinisateurs des baobabs malgaches. Cirad. Bois et forêts des tropiques, Bois et Forêts des Tropiques, pp.56-68, 2011, 307 (1).

Salmon, A., J. A. Joshua, J. A. Jeddeloh, and J. F. Wendel. 2012. Targeted capture of homoeologous coding and non-coding sequence in polyploid cotton. *G3: Genes, Genomes, Genetics* 2: 921-930

Stankowski, S. and Streisfeld, M.A., 2015. Introgressive hybridization facilitates adaptive divergence in a recent radiation of monkeyflowers. *Proc. R. Soc. B*, *282*(1814), p.20151666.

Than, C., Ruths, D. and Nakhleh, L., 2008. PhyloNet: a software package for analyzing and reconstructing reticulate evolutionary relationships. *BMC bioinformatics*, *9*(1), p.322.

Zhu, S., Degnan, J.H., Goldstien, S.J. and Eldon, B., 2015. Hybrid-Lambda: simulation of multiple merger and Kingman gene genealogies in species networks and species trees. *BMC bioinformatics*, *16*(1), p.292.

Smit, AFA, Hubley, R and Green, P. *RepeatMasker Open-4.0*. 2013-2015. <http://www.repeatmasker.org>

Snir, S. and Rao, S., 2012. Quartet MaxCut: a fast algorithm for amalgamating quartet trees. *Molecular phylogenetics and evolution*, *62*(1), pp.1-8.

Solís-Lemus, C. and Ané, C., 2016. Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting. *PLoS genetics*, 12:1005896.

Solís-Lemus, C., Bastide, P. and Ané, C., 2017. PhyloNetworks: a package for phylogenetic networks. *Molecular biology and evolution*, *34*(12), pp.3292-3298.

Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, *22*:2688-2690.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), pp.1312-1313.

Stenz, N.W., Larget, B., Baum, D.A. and Ané, C., 2015. Exploring tree-like and non-tree-like patterns using genome sequences: An example using the inbreeding plant species arabidopsis thaliana (l.) heynh. *Systematic biology*, 64:809-823.

Strimmer, K., Wiuf, C. and Moulton, V., 2001. Recombination analysis using directed graphical models. *Molecular Biology and Evolution*, *18*(1), pp.97-99.

Struck, T.H., 2013. The impact of paralogy on phylogenomic studies–a case study on annelid relationships. *PLoS One*, *8*(5), p.e62892.

Swofford, D.L., 2003. PAUP*: phylogenetic analysis using parsimony, version 4.0 b10.
Swofford, D.L.. (2003). PAUP*. Phylogenetic Analysis Using Parsimony (*and other Methods) Version 4. Sinauer, Sunderland, Massachusetts, USA. Nat. Biotechnol.. 18. 233-234.

Tiffney, B.H., 1985. Perspectives on the origin of the floristic similarity between eastern Asia and eastern North America. *Journal of the Arnold Arboretum*, *66*(1), pp.73-94.

Weitemier, K., Straub, S.C., Cronn, R.C., Fishbein, M., Schmickl, R., McDonnell, A. and Liston, A., 2014. Hyb-Seq: Combining target enrichment and genome skimming for plant phylogenomics.

Wendel, J.F. 2015. The wondrous cycles of polyploidy in plants. *American journal of botany* 102.11: pp.1753-1756.

Wendel, Jonathan F. and Cronn, Richard Clark, "Polyploidy and the Evolutionary History of Cotton" (2003). Botany Publication and Papers. 23.

Wesselingh, R.A. and Arnold, M.L., 2000. Pollinator behaviour and the evolution of Louisiana iris hybrid zones. *Journal of Evolutionary Biology*, *13*(2), pp.171-180.

Wickens, G.E., 2008. *The baobabs: pachycauls of Africa, Madagascar and Australia*. Springer Science & Business Media.

Wilde, V. and Frankenhäuser, H., 1998. The Middle Eocene plant taphocoenosis from Eckfeld (Eifel, Germany). *Review of Palaeobotany and Palynology*, *101*(1-4), pp.7-28.

Wolf, P.G., Robison, T.A., Johnson, M.G., Sundue, M.A., Testo, W.L. and Rothfels, C.J., 2018. Target sequence capture of nuclear-encoded genes for phylogenetic analysis in ferns. *Applications in Plant Sciences*, p.e01148.

Yang, Y. and S.A. Smith. 2014. Orthology inference in non-model organisms using transcriptomes and low-coverage genomes: improving accuracy and matrix occupancy for phylogenomics. Molecular Biology and Evolution. doi: 10.1093/molbev/msu245

Yoder, A.D., Cartmill, M., Ruvolo, M., Smith, K. and Vilgalys, R., 1996. Ancient single origin for Malagasy primates. *Proceedings of the National Academy of Sciences*, *93*(10), pp.5122-5126.

Yuan, Y.M., Wohlhauser, S., Möller, M., Klackenberg, J., Callmander, M.W. and Küpfer, P., 2005. Phylogeny and biogeography of Exacum (Gentianaceae): a disjunctive distribution in the Indian Ocean Basin resulted from long distance dispersal and extensive radiation. *Systematic Biology*, *54*(1), pp.21-34.

Zhang, N., Zeng, L., Shan, H. and Ma, H., 2012. Highly conserved low-copy nuclear genes as effective markers for phylogenetic analyses in angiosperms. *New Phytologist*, *195*(4), pp.923-937.

Zhu, S., Degnan, J.H., Goldstien, S.J. and Eldon, B., 2015. Hybrid-Lambda: simulation of multiple merger and Kingman gene genealogies in species networks and species trees. *BMC bioinformatics*, *16*(1), p.292.

Zimmer, E.A. and Wen, J., 2015. Using nuclear gene data for plant phylogenetics: Progress and prospects II. Next-gen approaches. *Journal of Systematics and Evolution*, *53*(5), pp.371-379.

Chapter Two:

Targeted sequence capture reveals gene flow among Madagascar's endemic baobabs

Key words: species delimitation, targeted sequence capture, reticulation, network inference

ABSTRACT

Using species tree and explicit network methods from targeted sequence capture data, we clarified the genealogical history of the six traditionally-recognized species of baobab (*Adansonia*; Malvaceae) that are endemic to Madagascar. The six species are divided into two sections based on floral morphology: *Brevitubae* comprises two species with short, white, primarily mammal-pollinated flowers, whereas the four species of *Longitubae* have elongated, hawkmoth-pollinated, yellow to red flowers. While Brevitubae is well supported as monophyletic, recent phylogenetic analyses have contradicted the monophyly of *Longitubae*. Species of *Longitubae* occur in regional or even local sympatry with one another and introgressive hybridization has been suggested. We use genome-scale data for accessions representing the full geographic ranges of each putative species, and included a few individuals that have been interpreted to be possible hybrids, to investigate species limits and patterns of gene flow within and between species. Explicit networks, which incorporate both incomplete lineage sorting and introgression, are estimated allowing us to test prior hypotheses and evaluate the role of geographic proximity in gene flow. Our data suggests patterns of gene flow based on geography and question the validity of one traditionally delimited species. Our study models new methods that can be used to explore discordance in non-model systems and illustrates the

capacity for phylogenomic methods to shed light on species delimitation and phylogeographic problems.

INTRODUCTION

Given the high degree of endemism on Madagascar, the island provides a model region for studying patterns and processes of diversification (Vences et al. 2009). Although Madagascar has been isolated from the other continents for 88-165 million years (Rabinowitz et al. 1983), many speciose lineages have been shown to originate more recently via long-distance dispersal from continental African (Meve and Liede 2002; Yoder and Nowak 2006; Janssens et al. 2016), or even the Indian Ocean Basin/Southeast Asia (Federman et al. 2015), as opposed to by Gondwanan vicariance. Numerous potential drivers of speciation have been proposed for groups radiating on Madagascar, including climatic stochasticity (Federman et al 2015; Dewar and Richard 2007), allopatric speciation based on watershed isolation (Pearson and Raxworthy 2009), and ecologically mediated range shifts (Losos and Glor 2003).

Studies of diversification are heavily dependent upon the inferred phylogeny and species designations, which can be challenging in cases of rapid radiations (Maddison and Knowles 2006). In such situations, integrating species-level phylogenetics with population-based approaches may be needed (Brito & Edwards 2009). Here we used targeted sequence capture and multilocus phylogenetics to explore patterns of diversification and gene flow and evaluate the validity of traditionally delimited species of baobabs in Madagascar (genus *Adansonia;* Malvaceae).

The Endemic Baobabs of Madagascar

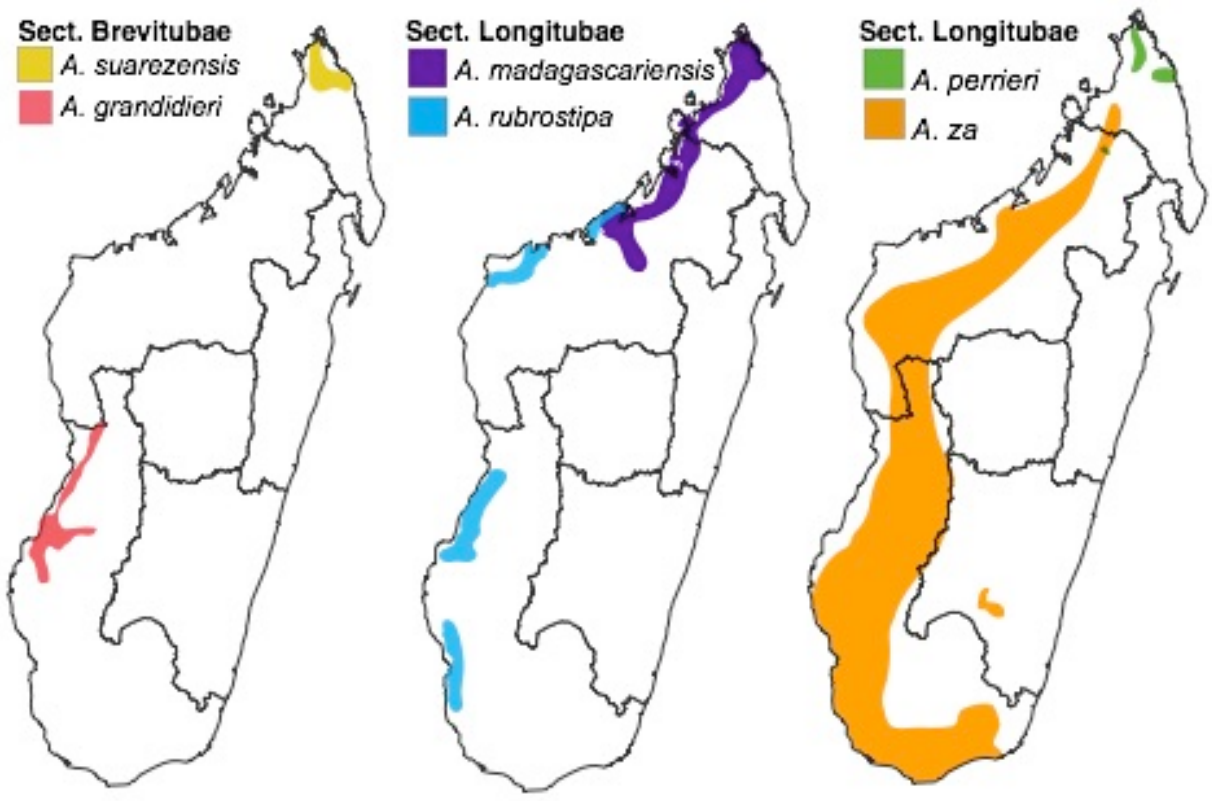


Figure 2. Distribution of the six baobab species endemic to Madagascar

Baobabs are iconic trees of western Madagascar, with six traditionally-recognized species endemic to the island (Baum 1995b). Systematic revision of the genus by Baum (1995b) affirmed two historically recognized sections. Sect. *Brevitubae* is composed of a pair of allopatric species (Fig 1.), *A. grandidieri* and *A. suarezensi*s, both massive trees with similar canopy structure and erect, white, mammal-pollinated flowers that open in the dry season when leafless. Based on nuclear data (Baum et al. 1998; Chapter 1) these two species are reciprocally monophyletic and show no evidence of gene flow with any of the remaining Malagasy species (Leong Pock Tsy et al. 2013), despite often occurring in sympatry with them.

The remaining four traditionally-recognized species (*A. madagascariensis*, *A. perrieri*, *A. rubrostipa*, and *A. za*) are classified within Sect. *Longitubae* based on similar elongated, primarily hawkmoth-pollinated flowers, which also occur in the Australian species *A. gregorii*.

The Malagasy species are further united by all having a red style, red inner-calyx, and either red or yellowish petals. Molecular phylogenetic data (Baum et al. 1998; Chapter 1) support a clade of six Malagasy species, but contradict the monophyly of the traditional *Longitubae*. The Malagasy *Longitubae* has since been shown to be non-monophyletic with *Brevitubae* sister to the "core" *Longitubae* (*A. madagascariensis*, *A. perrieri*, and *A. za*) and *A. rubrostipa* being sister to the rest of the Malagasy clade (Chapter 1).

A. rubrostipa is mostly known from the spiny forests of the southwest, but also occurs along the coast near Soalala in the northwest (Wickens 2008), and between Mahajanga and Nosy Hara (Cyrille Cornu, pers. comm.). This species has two distinct synapomorphies: leaflets with serrated leaf edges and flowers with a central cluster of stamens that are fused beyond the top of the staminal tube (Baum 1995b).

The “core *Longitubae*” are united by sharing a well-developed, annular nectar chamber (Baum, 1995a). Restricted to sub-humid forests in the north, *A. perrieri* is known from only a few populations, including Montagne D’Ambre, and the Ankarana limestone plateau. It grows in close sympatry with *A. madagascariensis* and *A. za*. *Adansonia madagascariensis* occurs in the north (Fig. 1) and is unique in having flowers with red petals, although yellow flowers also occur in some areas (Baum, pers. comm). Fruit are often distinctly globose in northernmost populations, though this is not consistent throughout the range. It is noteworthy, however, that this is the only taxon that flowers in fall (Mar.-Apr.).

A. za is the most widespread species occurring from Ambanja to almost Fort Dauphin in the south with a disjunct population inland near Ivohibe. There is variation in terms of flower color, from pale yellow to orange across its range. Two morphological traits have a clear pattern correlated with a latitudinal gradient (Fig. 2); leaflets are sessile in northern populations but are on long petiolules in southern populations. Similarly, fruit peduncles are narrow and elongated in the north but are distinctly swollen in the south. These traits are so pronounced that early systematists used them to distinguish three varieties: *A. za* var. *za* in the south, *A. za* var. *boinensis* in the northwest, and *A. za* var. *bozy* in the Sambirano region of the north. However, in his systematic revision of the group, Baum (1995b) deemed these characters too continuous to support formal taxonomic recognition of these varieties.

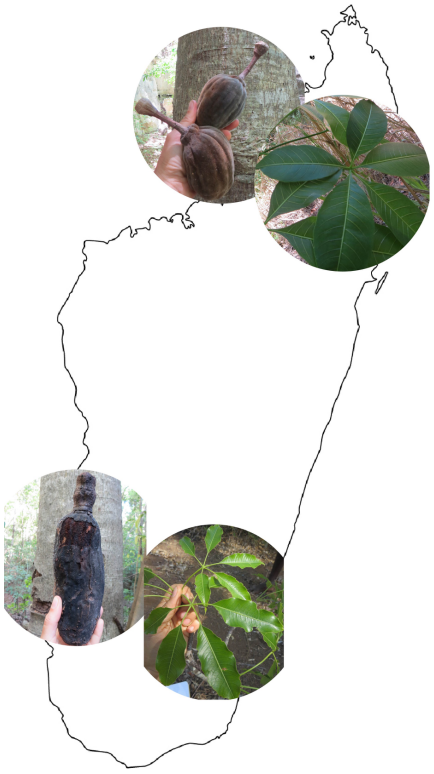


Figure 3. Morphological variations in *A. za.*

Previous Evidence of Gene Flow Between Traditional Species

Madagascar was most likely colonized from continental Africa, although the generic phylogeny is not conclusive (Baum et al. 1998; Chapter 1). Multilocus phylogenetic data suggest that, some time after Madagascar was initially colonized, there was a second migration across the Mozambique channel that resulted in introgression of genetic material from an African species (prior to the evolution of tetraploidy in that lineage) and the ancestor of Sect. *Brevitubae*

(Chapter 1). Nuclear and plastome discordance suggested additional gene flow events between *A. rubrostipa* and *A. za* (Chapter 1), but sampling in prior work was insufficient to determine the direction of gene flow or rule out additional hybridization events, as hypothesized by Leong Pock Tsy et al. (2014). Furthermore, samples representing *A. za* were consistently non-monophyletic in prior work (Baum et al 1998, Chapter 1), suggesting a more complicated genealogy.

Here we sampled broadly across the range of the Malagasy taxa in order to investigate the validity of species designations and reassess the status of *A. madagascariensis, A. perrieri,* and *A. za*, three species that may or may not be genealogically distinct entities. We also aimed to address whether geographic patterns of variation within Madagascar are best explained by clinal isolation-by-distance in a fragmented landscape or by vicariance followed by secondary contact. Finally, we clarified the role of introgression among all six putative baobab species in Madagascar.

METHODS

Taxon Sampling

Samples were collected in Madagascar in fall of 2014 under collection permit No. 195/14 granted on July, 29, 2014 by Madagascar agency La Conservation de la Biodiversite et du Systeme des Aires Protegees. Specimens are deposited at the Wisconsin State Herbarium and Parc Botanique et Zoologique de Tsimbazaza (PZBT) Herbarium in Antananarivo, Madagascar. See Appendix 1 for a list of specimens and locality information, and Figure 3 displays sampling. The full extent of most species' ranges are included except the extreme southeastern extent of *A.*

za and the northernmost *A. rubrostipa*. The biggest gap in our sampling is the inaccessible Ambongo-Boina region of the northwest, where Perrier de la Bâthie reported *A. madagascariensis, A. za*, and (along the coast) *A. rubrostipa* (from Wickens 2008). When possible, specimens from species found in sympatry were collected. Two samples correspond to morphological intermediates identified by Leong Pock Tsy et al. (2014) based on geographic coordinates and local knowledge. These include a putative *A. rubrostipa – A. za* hybrid from Kirindy in the southwest and a putative *A. za – A. perrieri* hybrid from Ambolobozo in the Sambirano, although the latter individual had flowers corresponding to *A. za* morphology. During our field sampling we located an individual tree that possesses an intermediate floral form between *A. perrieri* and *A. za* (see Fig. 4). We were kindly sent material of a putative new species intermediate between *A. rubrostipa* - *A. madagascariensis* that was located in coastal northwest Madagascar (Cyrille Cornu, pers. comm).

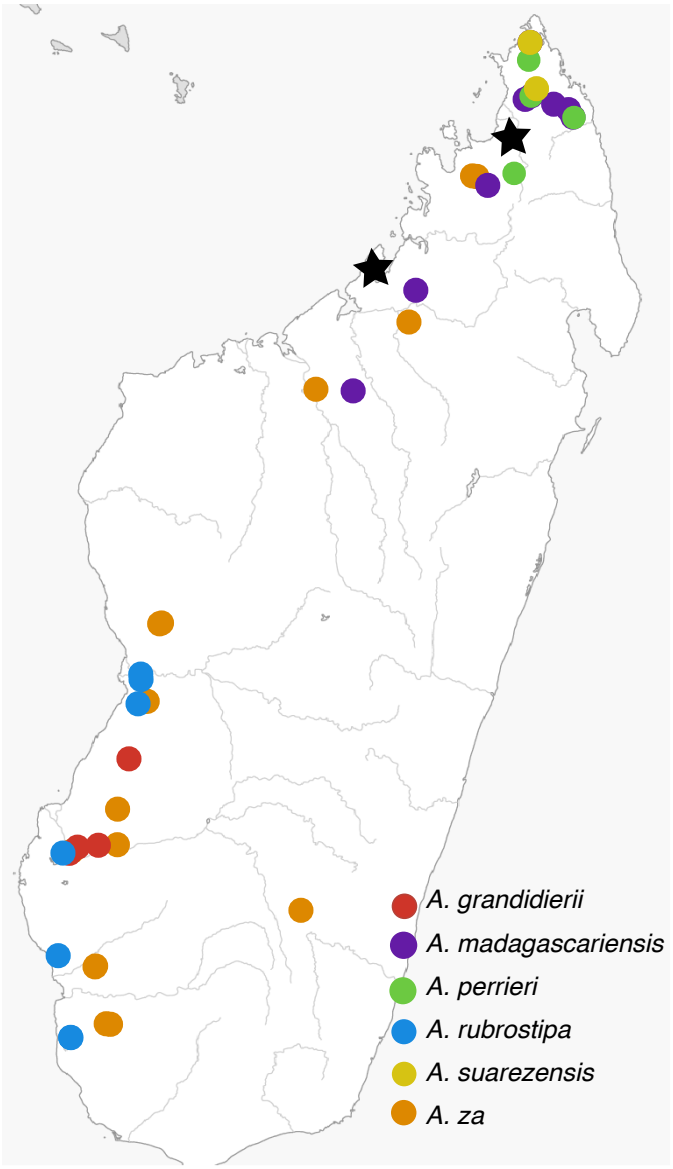


Figure 3. Sampling of the six species of baobab used in this study. Stars indicate individual morphologically-intermediate trees.

Our final dataset consisted of 56 samples, which also includes one sample each from the African and Australian baobabs (*A. digitata* and *A. gregorii*), and *Ceiba pentandra* and *Scleronema micrantha* as outgroups.

Figure 4. Flowers of *A. za (left)*, *A. perrieri* (right) and the morphological intermediate (center).

Sequencing

Plant tissue (silica-dried leaf or seed) was extracted with Qiagen DNeasy Plant Mini Kit (Qiagen, Netherlands) with modifications as in Cron et al. 2014. Bait design, DNA library construction, and captures were performed as described in Chapter 1. Target-enriched libraries were sequenced on a single lane of Illumina HiSeq2500 by the UW-Madison Biotechnology Center (Madison, WI) or at the Beijing Genomics Institute (BGI, Hong Kong) with paired-end reads of 250bp. Raw reads were trimmed using Trimmomatic v0.36 (Bolger et al., 2014) with the parameters *ILLUMINACLIP:Adapters.fa:2:30:15 LEADING:28 TRAILING:28 SLIDINGWINDOW:8:28 SLIDINGWINDOW:1:10 MINLEN:65 TOPHRED33.*

Dataset Assembly

Plastomes were assembled by mapping off-target quality-trimmed reads to *Gossypium raimondii* complete chloroplast (NCBI GenBank Accession HQ325744) using BWA *mem* algorithm of Burrows Wheeler Aligner (Li and Durban 2009). Plastome consensus sequences were extracted and aligned with MAFFT (Katoh et al. 2012) using the FFT-NS-2 algorithm. Maximum likelihood phylogenetic inference was performed with RAxML version 8.2.10 (Stamatakis 2006, 2014) using 100 bootstrap replicates with the GTR-Γ model.

Assembly of nuclear loci was performed using HybPiper (Johnson et al. 2016) and our original bait set targets as references (Chapter 1). Sequences without evidence of paralogs (i.e. no alternative contigs were recovered) were aligned with MAFFT (Katoh et al. 2012) and gene trees were inferred by RAxML version 8.2.10 (Stamatakis 2006, 2014). Gene trees were screened visually to remove targets that yielded putative paralogs (i.e. non-monophyletic ingroup taxa or symmetrical subtrees) or assembly errors (i.e. extremely long terminal branches). For all 160 target loci that met our filtering criteria, we also extracted intronic sequences from HybPiper using the intronerate.py pipeline, and used the combined intron+exon alignments for phylogenetic analysis. Gene trees were estimate using RAxML version 8.2.10 (Stamatakis 2006, 2014) with the GTR-Γ model and 100 bootstrap replicates.

Nuclear Admixture Visualization

To assist with visualizing geographic/clinal genetic variation, we wished to use Structure (Pritchard et al. 2000), but this method is only valid for independent SNPs, meaning that it is invalid to use multiple SNPs from a single gene alignment. We reasoned, however, that each well-supported branch on an unrooted gene tree could be coded as a "phyloSNP," where individual clade membership was assigned '1' for present in that clade or a '0' if outside. We

propose that, provided that only one phyloSNP is scored per branch, and that enough genes are scored to overcome topological errors in a single gene tree, admixture analysis can be usefully deployed to visualize genetic admixture.

Outgroups and *A. gregorii* were pruned from the gene trees and we used a custom R script to identify all partitions in all gene trees with a bootstrap score ≥50% and convert the tree into a binary phyloSNP matrix. The resulting matrices were concatenated over genes and were treated as haplotype genotypes input into Structure. Structure was run with 1 million generations, burn-in of 10,000 and *k=2-10*. The optimal number of clusters was selected with Clumpak (Kopelman et al. 2015).

Phylogenetic & Network Inference

A population (species) tree was inferred using ASTRAL (Mirarab et al. 2014) which implements the multi-species coalescent model (thus allowing for incomplete lineage sorting) given a set of gene trees. We used the maximum likelihood gene trees for both the coding sequences only, and the exons plus introns.

Phylogenetic networks were inferred with the pseudo-likelihood method used by SNaQ in the PhyloNetworks package of Julia (Solís-Lemus and Ané 2016). To test explicitly the hypothesis of hybridization given our samples, we produced subsets of the data in order to restrict the dataset size due to the computational challenges of large datasets on network inference. Sample selection was guided by the nuclear and plastid trees. Hybridization was explicitly tested for two individuals: the *A. perrieri* and *A. za* morphological intermediate (see Fig. 4), and the putative hybrid of *A. rubrostipa* - *A. madagascariensis.* Datasets were also compiled to test for hybridization between *A. rubrostipa* and *A. za* lineages, as suggested by the

plastid tree. Table 2 shows samples included in each dataset. For each analysis we generated a starting tree by pruning the excluded taxa from the ASTRAL population tree and used concordance factors from all quartets calculated from the maximum likelihood gene trees for all 160 nuclear genes. Network searches were performed starting at zero hybridization edges (h0) to first optimize topology and branch lengths. Network searches increased sequentially, up to three hybridization edges (h3), using the optimal network from the previous search. The optimal number of hybridizations per dataset was selected based slope heuristic of the log-pseudolikelihood scores (Solís-Lemus and Ané 2016).

Table 2. Samples selected (1) for each dataset used for SNaQ Analyses.

	vAzp29	vAzp_noSza	vAmad301	vRubZa
Adi004	0	0	0	0
Adi010	0	0	0	0
Aga095	0	0	1	1
Aga101	0	0	0	0
Aga108	1	1	0	0
Aga119	0	0	0	0
Agra111	0	0	0	0
Agra84	0	0	0	0
AgreM555	0	0	0	0
Ama006	0	1	1	1
Ama018	0	0	0	0
Ama023	0	0	0	0
Ama029	0	0	0	0
Ama034	1	1	1	0
Ama048	0	0	0	0
Ama054	0	0	0	0
Ama058	0	0	0	0
Ama201	0	0	0	0
Ama301	0	0	1	0
Amad1	0	0	0	0
Ape009	1	1	1	1
Ape013	0	0	0	0
Ape032	1	1	0	0
Ape201	1	1	0	0
Aper11	0	1	0	0
AperMad	0	0	0	0
Aru068	0	1	0	0

Aru076	0	0	0	0
Aru083	0	0	0	0
Aru085	0	0	1	1
Aru117	0	0	1	1
Aru124	0	0	1	1
Aru127	0	0	0	1
Aru128	1	1	0	1
Arub72	0	0	0	0
Asu012	1	1	1	1
Asua2	0	0	0	0
Asua3	0	0	0	0
Aza037	1	1	0	0
Aza038	1	1	1	0
Aza043	0	0	0	0
Aza055	1	0	0	1
Aza060	0	0	0	0
Aza081	1	0	0	0
Aza096	0	0	1	1
Aza122	0	0	0	1
Aza133	0	0	0	1
Aza135	0	0	1	1
Aza136	0	0	0	0
Aza143	0	0	0	1
Aza201	1	1	0	0
Aza92	0	0	0	0
AzaMad	0	0	0	0
Azp029	1	1	0	0
Cei070	1	1	1	1
Smi165	0	0	0	0

Clade Exclusivity

We calculated clade exclusivity factors as in Wright and Baum (2018). The exclusivity factor of a set of tips is defined as the maximum pairwise relatedness among members of the set minus the minimum pairwise relatedness between a member of the set and a tip outside the set (an outgroup). As a proxy for relatedness between a pair of tips, we used the patristic distance between them, averaged over nuclear genes. Thus, higher relatedness corresponds to a smaller distance. Thus, sets of accessions with positive exclusivity scores, exclusive taxa, are candidate taxa since they are composed of organisms that are more closely related to each other than to any

organism outside the group (Baum 2009). Such an approach is proposed as superior to a strict monophyly criterion, especially in cases with extensive gene tree incongruence and deep coalescence (Knowles and Carstens 2007, Wright and Baum 2018).

To calculate pairwise patristic distances, analyses used the maximum likelihood gene trees from the nuclear coding sequences that had no missing taxa (181 loci). One analysis used branch lengths for each gene, as estimated by RaxML. A second analysis first rendered these trees ultrametric by: rooting them with *C. pentandra* and then rescaling branches using Penalized Likelihood in the R package Ape ver. 5.2 (Paradis et al. 2004). In either case, patristic distances (pairwise) were calculated for each tip of the tree using the cophenetic function of APE and these were averaged over genes. The matrix of average pairwise distances was used to build an UPGMA tree, which should include as clades all exclusive taxa (Wright and Baum, 2018).

RESULTS

Plastid phylogeny suggests extensive gene flow, restricted by geography

Plastid data (Fig. 5) suggests non-monophyly of the two mammal-pollinated species of Sect. *Brevitubae*, *A. suarezensis* and *A. grandidieri*. This is consistent with previous phylogenetic analyses (Chapter 1) although here we find *A. suarezensis* to be sister to the African and Australia lineages (*A. digitata* and *A. gregorii*) rather than sister to the rest of the Malagasy clade, albeit with very low support (38% bootstrap). This result is likely a consequence of using a different (and more distant) outgroup in this case.

The plastid data strongly supports monophyly of the four species of Sect. *Longitubae* (BS=99). However, within this clade, traditionally named species do not consistently form

clades. One clade (BS=100), containing both southern populations of *A. za* and our southernmost samples of *A. madagascariensis*, is sister to the rest of Longitubae (Fig. 5). Two other southern accession of *A. za* are embedded within a paraphyletic grade containing all sampled *A. rubrostipa.* This *A. rubrostipa* – southern *A. za* clade (BS=100) is sister to a clade of northern populations of *A. madagascariensis, A. za,* and *A. perrieri*. With respect to three

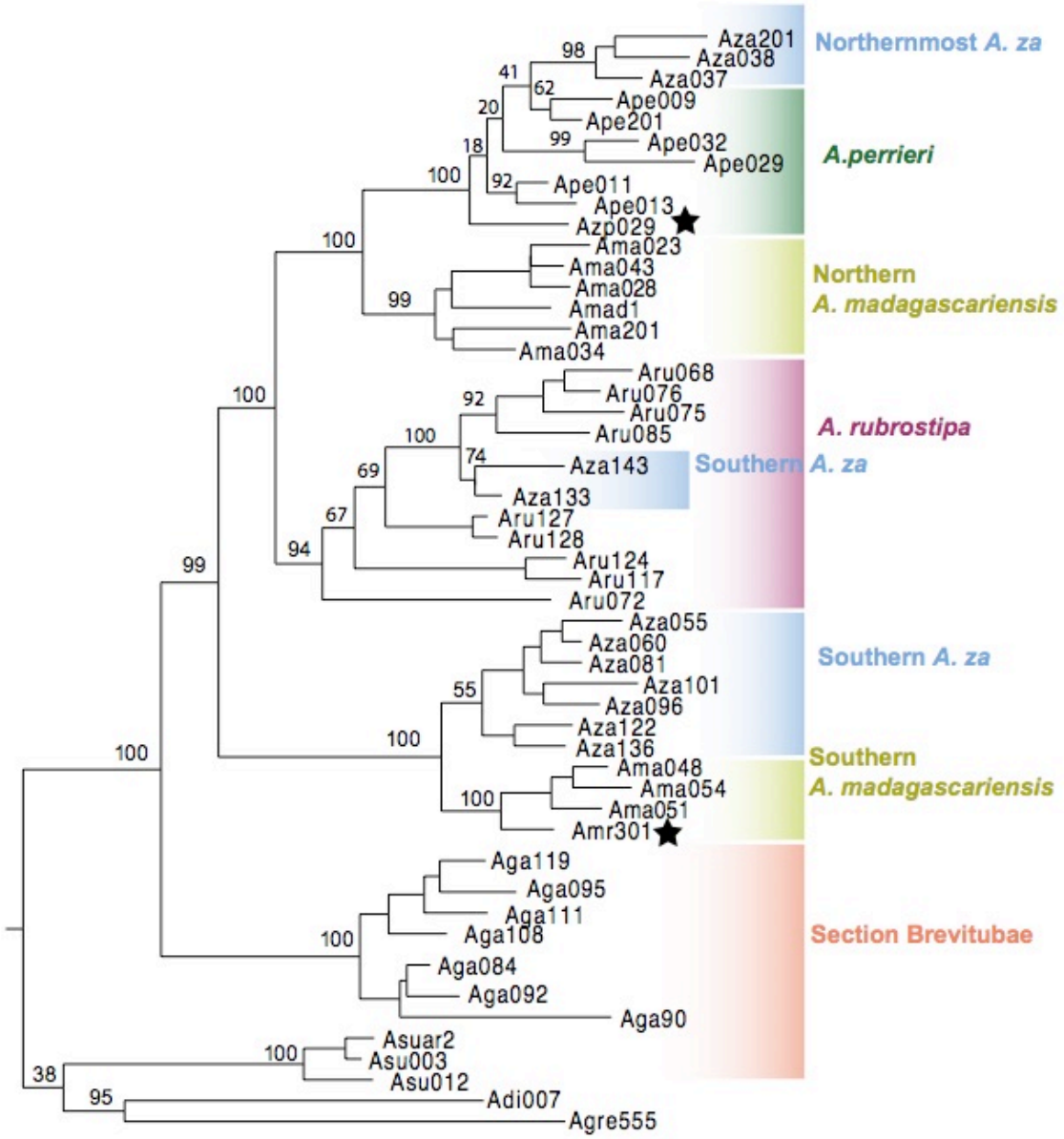


Figure 5. Maximum likelihood ingroup plastid tree. Bootstrap support for the primary clades are labeled.

northern *A. za* samples, *A. perrieri* (the highly localized and florally-distinct Longitubae taxon) is paraphyletic (see Fig. 5).

The accession that appeared intermediate between *A. za* and *A. perrieri* was found in a clade with *A. perrieri* (BS=100%). It is placed as sister to all sampled *A. perrieri* accessions but very weakly (BS=18%). The putative *A. rubrostipa – A. madagascariensis* hybrid is sister to a clade of southern *A. madagascariensis* accessions (BS=100%).

Nuclear Phylogeny and Introgression

The population tree topology inferred by ASTRAL from the nuclear coding sequences was identical to that inferred from the exons plus introns; therefore, Figure 6 shows the tree inferred with introns included. Each traditional species is supported as a clade with a posterior probability of 1.0 except for *A. za*, which is broken into a clear northern and southern clade. The northern *A. za* clade is sister to *A. madagascariensis* (PP = 0.87), and these together are sister to *A. perrieri* (PP = 1.0). As found previously (Chapter 1), *A. rubrostipa* in Sect. *Longitubae* is a monophyletic group sister to the rest of the Malagasy clade. This renders the Malagasy *Longitubae* paraphyletic due to the inclusion of Sect. *Brevitubae*. The nuclear population tree agrees with the plastid data in supporting the *A. za – A. perrieri* intermediate as a member of an *A. perrieri* clade (BS=100%). In contrast the putative *A. rubrostipa - A. madagascariensis* hybrid falls sister to the *A. rubrostipa* clade, which contrasts with its position as sister to *A. madagascariensis* with plastid data.

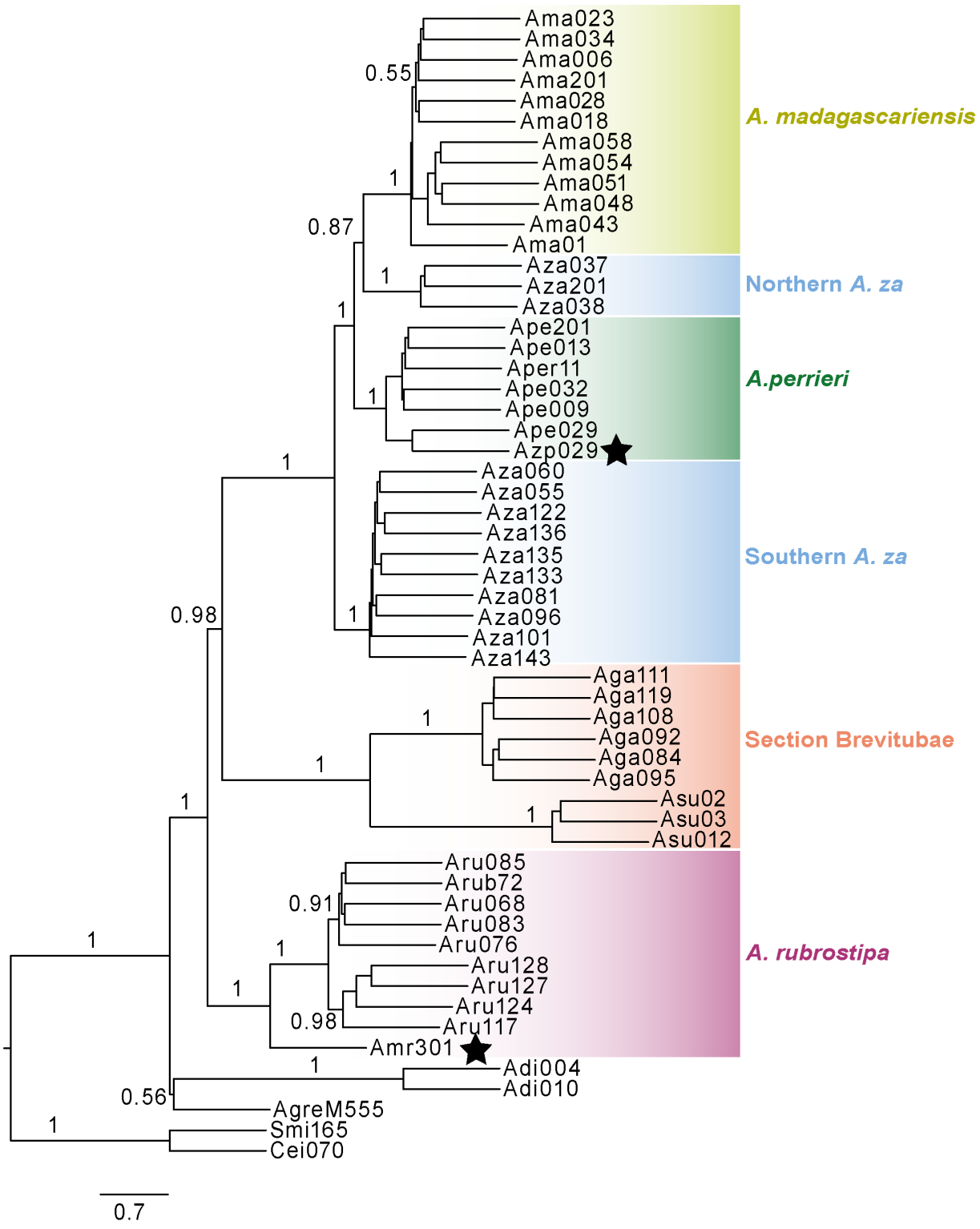


Figure 6. ASTRAL tree inferred from 160 loci (exons and introns). Stars indicate the two

putative hybrids. Only support values for the primary clades of interest are labeled for readability.

Network inference

Using SNaQ (Solís-Lemus and Ané 2016), we tested various hybridization hypotheses suggested by the plastid data including the individual samples (*A. za* - *A. perrieri* morphological intermediate and the *A. rubrostipa* - *A. madagascariensis* intermediate). SNaQ analysis of the *A. rubrostipa* - *A. madagascariensis* intermediate (Ama301) supported one hybridization edge and, indeed, recovered this individual tree as a hybrid based on nuclear data. SNaQ estimated that 68.5% of the genome of the hybrid had an *A. rubrostipa* origin, with 31.5% coming from *A. madagascariensis* as the donor lineage (Fig. 7).

Using a dataset of multiple *A. perrieri* and *A. za* clades (inclusion of both northern and southern *A. za*) did not suggest hybridization origin for this individual (Azp029) instead showing gene flow indicated between the geographic *A. za* clades. After removing the southern *A. za* samples, gene flow between the *A. perrieri* and *A. za* clades was still not detected (not shown).

Quantifying gene flow between the *A. rubrostipa* lineage and the *A. za* lineages was complicated. Based on the slope heuristic of log likelihood scores (as described in Solís-Lemus and Ané 2016) for h0-h3, the optimal number of reticulation edges is unclear as there seems to be a steady improvement with each additional hybridization edge, but no one significant change, suggesting that the optimal network has no edges. It is noteworthy, nonetheless, that the optimal h1 network implies gene flow from *A. za* into *A. rubrostipa*, contributing ~7% of the *A. rubrostipa* genome (Fig. 7). Similarly, using a different subset of *A. rubrostipa* and *A. za* samples also resulted in an h1 network between *A. za* into *A. rubrostipa* with 3% inheritance.

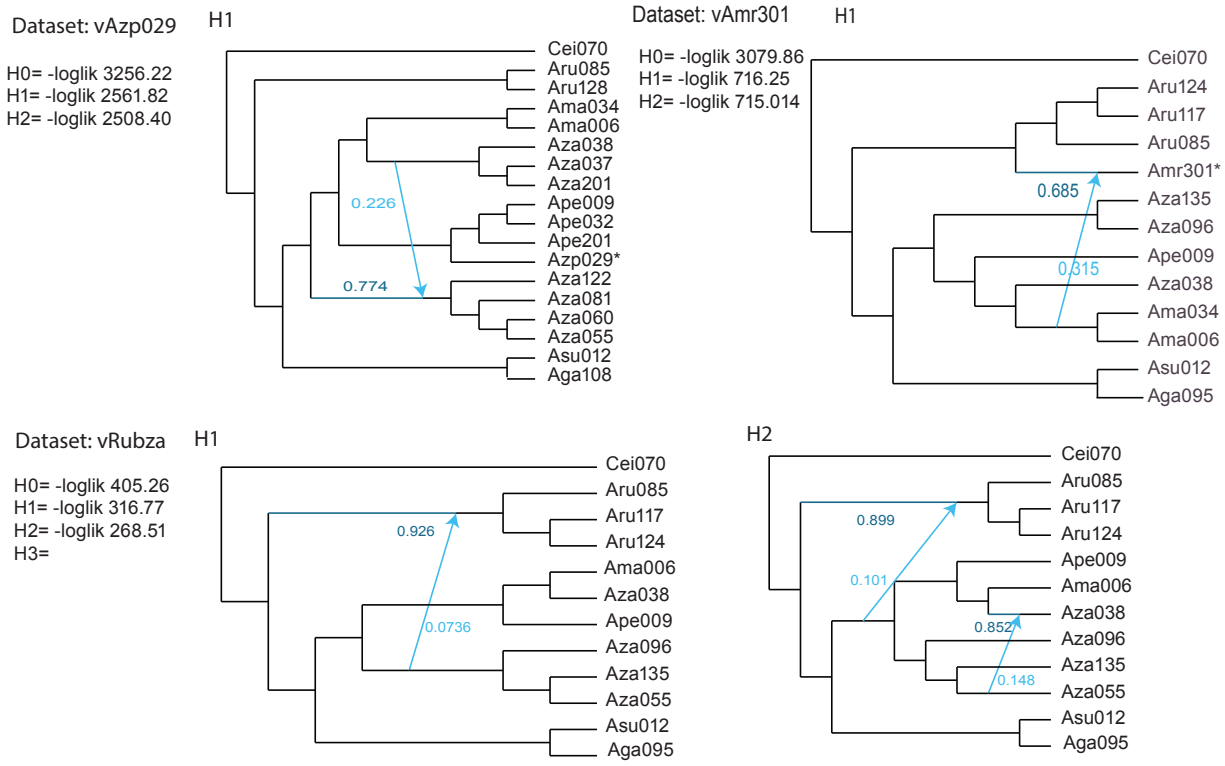


Figure 7. Phylogenetic networks inferred by SNaQ. Blue arrows indicate directionality of gene flow. Putative hybrid samples indicated with stars.

Using our nuclear maximum likelihood trees to generate "phyloSNPs" for nuclear admixture visualization (see methods), the optimal number of clusters was inferred at *k=9,* though with three inferred clusters having a membership proportion of zero. The resulting structure plot showed strong differentiation based on traditional species designations (Fig. 8).

The *A. perrieri* - *A. za* morphological intermediate is interpreted as a pure *A. perrieri* genotype. However, the individual sample labeled as Ape029 shows a predominately *A. perrieri* genotype with some *A. za* admixture, despite being consistently sister on both the plastid and nuclear trees to the intermediate sample. Geographically, this sample was collected in the vicinity of the intermediate.

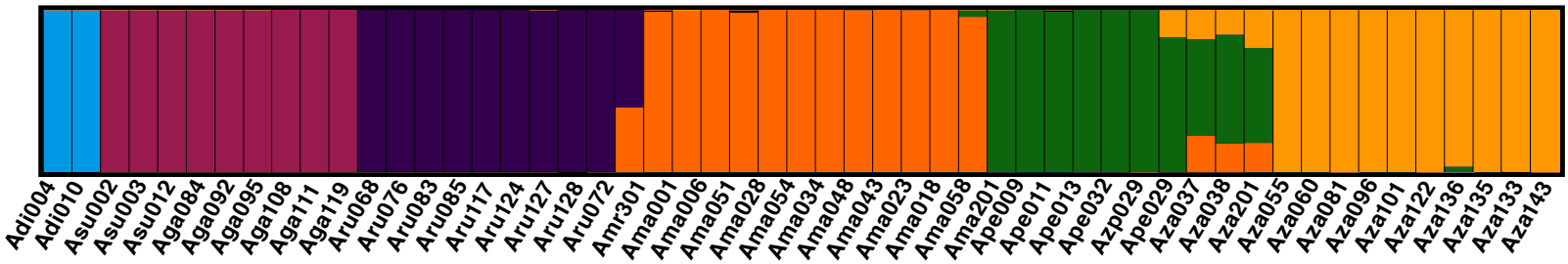


Figure 8. Structure plot (at *k=9*) using phyloSNPs. Blue represents *A. digitata* samples, pink corresponds to *Brevitubae* samples, purple to *A. rubrostipa,* orange is *A. madagascariensis,* green is *A. perrieri,* and yellow is *A. za*.

The three samples from the northernmost populations of *A. za* possess three admixed genotypes of otherwise found in southern *A. za, A. madagascariensis*, and *A. perrieri*. Two of these samples Aza037 and Aza038 were collected at a site studied by Leong Pock Tsy (2014) and identified as *A. perrieri* - *A. za* hybrids. These trees have the sessile petiolules and narrow peduncles consistent with previous descriptions of northern *A. za* populations (Baum 1995b and references within).

The *A. rubrostipa – A. madagascariensis* hybrid is clearly indicated as primarily *A. rubrostipa* genotype with a significant amount of admixed *A. madagascariensis* alleles, further supporting its hybrid status from SNaQ.

Clade Exclusivity

The analysis which first rendered the trees ultrametric resulted in a consensus UPGMA tree (Fig. 9) in which all traditionally designated species show positive exclusivity, with the exception of *A. perrieri* and *A. za*. Consistent with the inferred population tree, the latter is broken into two distinct exclusive clades, one of northern populations and one representing all samples from southern populations (Fig. 9). Despite apparent hybridization between *A. rubrostipa* with *A.*

madagascariensis, the apparent hybrid was sister to the *A. rubrostipa* clade, with both this combined clade and the *A. rubrostipa* clade showing positive exclusivity. Perhaps due to reticulation between *A. digitata* and Sect. *Brevitubae*, the Malagasy clade as a whole is not exclusive. The analysis using branch lengths from the RaxML maximum likelihood gene trees resulted in exclusive clades for only *A. rubrostipa*, and each of the *Brevitubae* (Fig. 10). We suspect that the lack of exclusivity for some clades when using raw branch length reflects some long terminal branch lengths (confirmed by visual inspection of gene trees) perhaps due to misassembled portions of alignments.

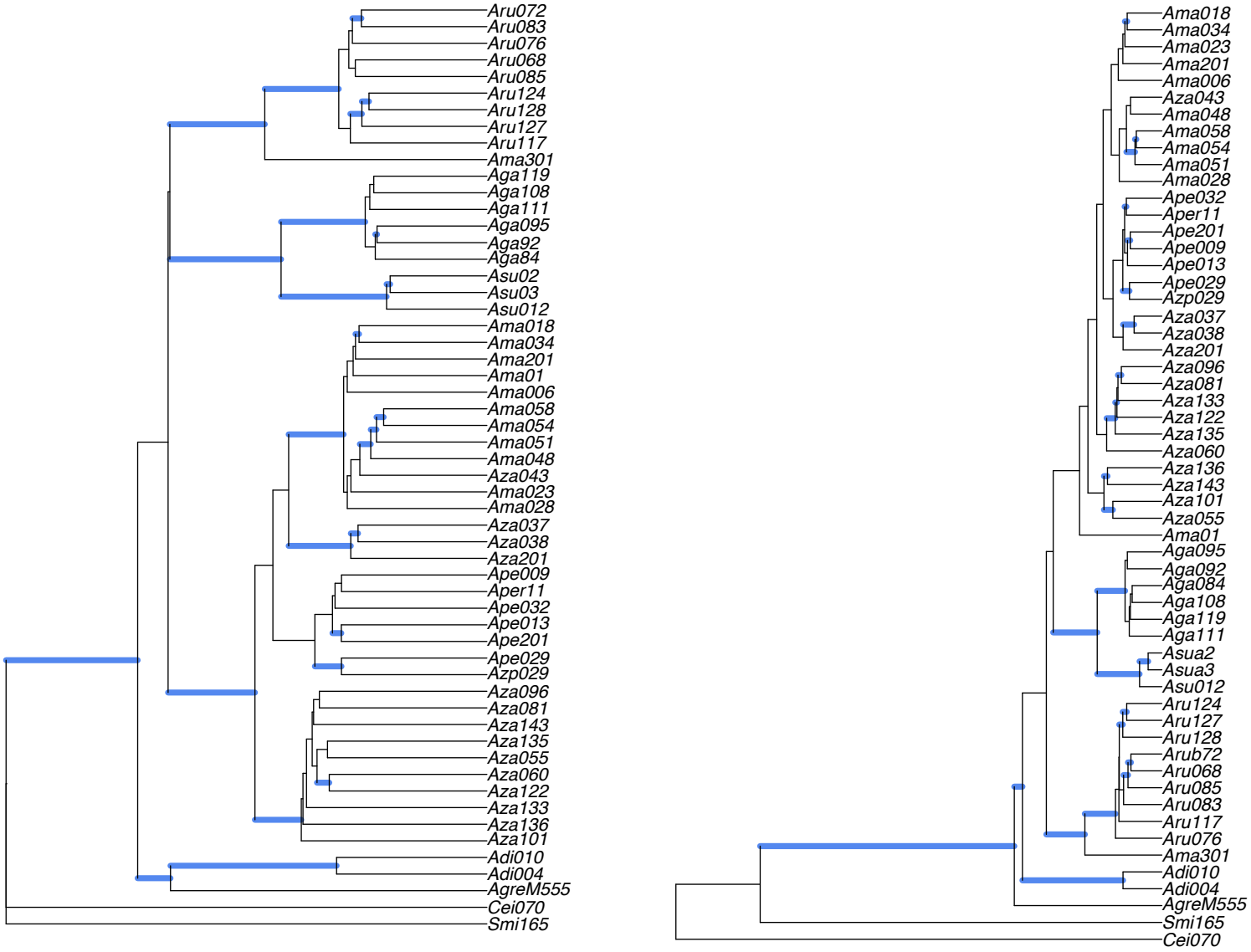


Figure 10. UPGMA tree of average patristic distances. Branch lengths made ultrametric by PL (left). Branch lengths from maximum likelihood gene trees (right). Branches in blue indicate clades with positive exclusivity (as in Wright and Baum 2018).

DISCUSSION

Phylogenomic tools for studying population-level patterns of gene flow in a spatial context

An important challenge in systematics is to understand the processes of diversification that occur on landscapes. This becomes even more challenging with hybridization and reticulate patterns of gene flow. Here we try to bridge gene tree and SNP based methods through the use of phylogenetically-derived admixture analysis coupled with explicit network methods. We have shown that despite several instances of reticulation during the diversification of the Malagasy baobabs, patterns of relationship are still significantly tree-like, with most instances of reticulate evolution involving populations that are likely to come into secondary contact.

Status and Implications of Individual Morphological Intermediates

The sample we suspected was a morphological intermediate between *A. perrieri* and *A. za* appears, instead, to be a member of a well-supported *A. perrieri* clade. The *A. perrieri* clade appears on the optimal plastid tree to be sister to a clade of the three northern *A. za*, while it appears as sister to a combined northern *A. za – A. madagascariensis* clade on the nuclear tree. One possibility is that northern *A. za* and *A. perrieri* form a clinal species complex, but this has been obscured by habitat fragmentation and limited sampling. Both species are hawkmoth pollinated and flower simultaneously (Baum 1995a; Ryckewaert et al. 2011). Based on geographic coordinates and local knowledge, we believe two individuals of the northernmost *A. za* clade (Aza037 and Aza038) are the same individuals assessed by Leong Pock Tsy et al. (2014), who deemed them to be introgressed hybrids of *A. za* - *A. perrieri*. It is unclear if these admixed genotypes represent localized hybrids of distinct taxa, or if northern *A. za* and *A. perrieri* represent a stable cline.

Adansonia rubrostipa represents is a distinct lineage from the rest of *Longitubae*, but has an apparent history of reticulation with both *A. madagascariensis and A. za.* Both the network analysis and nuclear admixture estimation showed that the putative *A. rubrostipa* - *A. madagascariensis* hybrid clearly possesses genetic material from both taxa. Geographically, this is reasonable since this individual occurs close to the ranges of both species; along the coast between Mahajanga and Nosy Ha (Cyrille Cornu, pers. comm.). Given that this individual was sister to *A. madagascariensis* in the plastid tree and sister to *A. rubrostipa* based on the nuclear population tree suggests that this hybridization was ancient. This raises the possibility that rather than being a rare individual of hybrid ancestry, all trees previously identified as "*A. rubrostipa*" in the northwest represent an introgressed/hybrid genotype. Unfortunately, we were unable to obtain additional samples from either species from this area, which leaves uncertain the extent and frequency of hybridization between these two species.

The plastid tree suggests gene flow between *A. rubrostipa* and *A. za.* Leong Pock Tsy et al. (2014) also proposed introgression between *A. rubrostipa* with *A. za* based on supposed morphological intermediates in areas of sympatry. Specifically, from one individual *A. rubrostipa* at Tsimanampetsotsa and two individuals from Kirindy. However, while we were able to sequence samples of both *A. rubrostipa* and *A. za* from the vicinity of Tsimanampetsotsa, those specific samples did not show evidence of introgressed genes. Furthermore, we believe we sampled an individual tree in Kirindy that Leong Pock Tsy et al. (2014) showed as admixted in their study, but our results suggest this tree is pure *A. rubrostipa*. Therefore, perhaps those exact individuals were not included in this analysis. Nonetheless, the plastid tree in combination with our nuclear network and SNaQ analyses does suggest gene flow between these species.

Hypothesis for Baobab Diversification in Madagascar

To develop a robust hypothesis of how a dynamic landscape and climate could have driven diversification in the Malagasy baobabs, we would need a dated phylogeny for the population tree. This is currently not possible due to the lack of fossils and secondary calibrations available given our sets of taxa. Nonetheless, assuming that extant distributions (currently impacted by deforestation) were stable in the recent past, we may speculate on how patterns of gene flow relate to the geographic radiation of the Malagasy baobabs.

Following an initial dispersal event from continental Africa, the ancestral Malagasy lineage most likely evolved a yellow-flowered, hawkmoth-pollinated population. From which, the ancestor of *A. rubrostipa* then separated from a lineage that later split into *Brevitubae* and the core *Longitubae*. Sometime thereafter, a subsequent dispersal event from Africa introduced a mammal pollinated taxon that interbred with the ancestral *Brevitubae* resulting in a shift in pollination system to mammals (i.e. bats and/or lemurs), as discussed in Chapter 1.

Perhaps due to the same geographic or climatic phenomena that resulted in allopatry of the two extant *Brevitubae* lineages, the core *Longitubae* split into the southern Longitubae (i.e. *A. za*), and a more complex northern *Longitubae* clade (representing extant *A. za var. bozy, A. madagascariensis* and *A. perrieri*). Apart from some gene exchange between southern *A. za* and *A. rubrostipa,* most of the subsequent diversification occurred in the north, reflective of the greater topographic and climatic heterogeneity in northern Madagascar.

These geographic and climatic complexities in the north of Madagascar may explain the adaptation of *A. perrieri* to higher elevation sub-humid forests. Similar patterns of diversification into the humid forests of the north have been described in lemurs (Goodman and Ganzhorn 2003). *Adansonia perrieri* could have evolved the elongated staminal tube as an adaptation to its primary pollinator, the famous Darwin hawkmoth *Xanthopan morganii* (Baum, 1995a). Perhaps

these habitats also served as refugia during range contractions of glacial maximum where it persisted while other populations contracted (Gasse and Van Campo, 1998).

Adansonia madagascariensis could have diverged from the ancestral population (the common ancestor to northern *A. za* and *A. madagascariensis*) perhaps due to evolution of fall flowering, resulting in a temporal reproductive isolation. Subsequently, populations of *A. madagascariensis* migrating southward could have intersected the range of *A. rubrostipa* and hybridized to generate an admixed northern *A. rubrostipa* form that acquired the plastid and about 20% of its nuclear genome from *A. madagascariensis,* supported by our sample with an admixed genotype. Although flowering times are staggered, the summer-flowering *A. rubrostipa* (February-March) has the potential to sometimes overlap in flowering with *A. madagascariensis* (March to April). This is consistent with the reported flowering time of March for our hybrid sample (Cyrill Cornu, pers. comm.).

While the plastid tree does not suggest gene flow between all sampled individuals representing these two lineages (*A. madagascariensis* and *A. rubrostipa*), it does suggest gene flow between southern *madagascariensis* and southern *A. za,* which form a clade sister to the rest of the *Longitubae*. The individuals representing the northern *A. za* samples and northern *A. madagascariensis,* given the plastid tree, are found north of the Sambirano River which could have acted as a barrier to gene flow, either as a barrier itself or as watershed climatic corridor serving as a refugia and/or recolonization pathway, both which have been shown as drivers of speciation in Madagascar (Wilmé, et al 2006; Pearson and Raxworthy 2009). Some individuals of ancestral *A. madagascariensis* could have crossed this barrier and interbred with the resident *Longitubae* to explain the sister southern *A. madagascariensis* and southern *A. za* clades on the plastid tree.

Species Delimitation and Taxonomic Implications Given Patterns of Gene Flow

Phylogenomic data consistently support the monophyly (on the population tree) and exclusivity of Sect. *Brevitubae*, *A. grandidieri* and *A. suarezensis,* despite having a history of reticulation with the African baobab lineage (Chapter 1). Species designations are also undeniable given morphology. *Adansonia rubrostipa* can be delimited as a valid, exclusive species, whether or not the admixed *A. rubrostipa – A. madagascariensis* sample is included in this taxon. It is possible however that additional sampling from the northwest will show that the admixed genotype applies to all "*A. rubrostipa*" in this area, in which case it might prove appropriate to split *A. rubrostipa* into two species, depending on morphology. It is noteworthy that, based on staminal tube length, Perrier de la Bâthie designated two *A. rubrostipa* varieties, with *A. rubrostipa* var. *rubrostipa* in the north and *A. rubrostipa* var. *fony* occurring in the south (as discussed in Baum 1995b and Wickens 2008). However, in the absence of more data (both molecular and morphological) and uncertainty regarding the genetic make-up of *A. rubrostipa* var. *rubrostipa*, it would be premature to elevate these varieties to species rank.

Patterns of gene flow observed between the three core *Longitubae* taxa (*A. madagascariensis*, *A. za*, and *A. perrieri)* challenge traditional species designations. A case can be made for retaining species status for *A. madagascariensis*, given the distinct flowering phenology, despite the optimal population tree grouping *A. madagascariensis* populations south of the Sambirano with northwestern populations of *A. za*, rather than with northern *A. madagascariensis,* While we do not have floral color assigned to each of our sampled specimens, the usually dark red-petaled flowers of *A. madagascariensis* are not diagnostic, given that yellow-petaled trees also exist, but it is possible that these variations correspond to the

geographic clades, with yellow-flowered occurring in the south, but we can only speculate this at this time.

Adansonia perrieri is the most restricted species in terms of overall range, habitat, and possesses a distinct flower form that seen in other *Longitubae*. While it does not have positive exclusivity, we believe this result is most likely an artifact due to some long terminal branches for a few gene trees, which would have the potential to distort estimates of relatedness in regions of the tree with short internal branches. Thus, we believe all evidence otherwise supports retaining its current species designation.

Adansonia za fails to emerge as a cohesive taxonomic unit. The exclusivity scores support distinct northern and southern *A. za* populations, and population tree suggests that these are not each others' closest relatives. It is unclear if the lack of exclusivity of *A. za* as a whole is due to a latitudinal gradient compounded by genetic admixture with *A. rubrostipa* in the south and *A. madagascariensis* (and maybe *A. perrieri*) in the north. Alternatively, the group may never have been a distinct taxonomic unit but was, instead, delimited by populations possessing presumed plesiomorphic traits (yellow flowers, long staminal column, spring flowering). There is considerable variation in flower petal color from pure yellow to yellow with orange striations to near-pure orange petals, in addition to a wide range in fruit shape.

Although limited sampling and extensive habitat fragmentation may contribute to the north-south distinction we detected, these populations likely deserve recognition as distinct entities. Our current data suggest that *A. za* populations from the Sambirano region, previously classified as *A. za* var. *bozy* (Perrier de la Bâthie), may warrant species status under the pre-existing name, *Adansonia bozy* Jum. & H. Perrier. However, more work would be needed to identify morphological traits that consistently separate *A. bozy* from other lineages formerly classified as *A. za*.

Species delimitation not only relies on a group's genealogical exclusivity but incorporates ranking and taxonomy based on practical considerations such as identifiability and historical precedent for the group in question. While more work is needed to determine the extent of reticulate evolution and especially to evaluate the impact of our limited sampling (in the remote region of Namoroka, and the Ambongo-Boina region), by integrating approaches and providing exclusivity factors for putative taxa among the Malagasy baobabs, we hope this study may serve as a model for species delimitation in the phylogenomic age.

ACKNOWLEDGEMENTS

We thank Diana Mayne and Cyrille Cornu for graciously sending samples. Botanists of Parc Botanique et Zoologique de Tsimbazaza (PBZT): Jacky Andriantiana Jacqueline Razanatosa, and Hanta Razafindraibe.The Missouri Botanical Gardens Madagascar for assistance with permitting and logistics. Dimby Raharinjanahary and land managers of Madagascar National Parks. Steve Goldstein for computational assistance. Chloe Pak Drummond for helpful discussions.

REFERENCES

Brito, P.H. and Edwards, S.V., 2009. Multilocus phylogeography and phylogenetics using sequence-based markers. *Genetica*, *135*(3), pp.439-455.

Cron, G.V., Karimi, N., Glennon, K.L., Udeh, C.A., Witkowski, E.T., Venter, S.M., Assogbadjo, A.E. and Baum, D.A., 2016. One African baobab species or two? Synonymy of Adansonia kilima and A. digitata. *Taxon*, *65*(5), pp.1037-1049.

Earl, D.A. and vonHoldt, B.M. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation genetics resources*, *4*(2), pp.359-361.

Johnson, M.G., Gardner, E.M., Liu, Y., Medina, R., Goffinet, B., Shaw, A.J., Zerega, N.J. and Wickett, N.J., 2016. HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. Applications in plant sciences, 4:1600016.

Knowles, L.L. and Carstens, B.C., 2007. Delimiting species without monophyletic gene trees. *Systematic biology*, *56*(6), pp.887-895.

Meve, U. and Liede, S., 2002. Floristic exchange between mainland Africa and Madagascar: case studies in Apocynaceae–Asclepiadoideae. *Journal of Biogeography*, *29*(7), pp.865-873.

Janssens, S.B., Groeninckx, I., Petra, J., Verstraete, B., Smets, E.F. and Dessein, S., 2016. Dispersing towards Madagascar: Biogeography and evolution of the Madagascan endemics of the Spermacoceae tribe (Rubiaceae). *Molecular phylogenetics and evolution*, *95*, pp.58-66.

Li, H. and Durbin, R., 2009. Fast and accurate short read alignment with Burrows–Wheeler

transform. Bioinformatics, 25:1754-1760.

Losos, J.B. and Glor, R.E., 2003. Phylogenetic comparative methods and the geography of speciation. *Trends in Ecology & Evolution*, *18*(5), pp.220-227.

Maddison, W.P. and Knowles, L.L., 2006. Inferring phylogeny despite incomplete lineage sorting. *Systematic biology*, *55*(1), pp.21-30.

Mirarab, S., Reaz, R., Bayzid, M.S., Zimmermann, T., Swenson, M.S. and Warnow, T., 2014. ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics*, *30*(17), pp.i541-i548.

Pearson, R.G. and Raxworthy, C.J., 2009. The evolution of local endemism in Madagascar: watershed versus climatic gradient hypotheses evaluated by null biogeographic models. *Evolution: International Journal of Organic Evolution*, *63*(4), pp.959-967.

Solís-Lemus, C. and Ané, C., 2016. Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting. PLoS genetics, 12:1005896.

Paradis, E., Claude, J. and Strimmer, K., 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, *20*(2), pp.289-290.

Pritchard, J.K., Stephens, M. and Donnelly, P., 2000. Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), pp.945-959.

Vences, M., Wollenberg, K.C., Vieites, D.R. and Lees, D.C., 2009. Madagascar as a model region of species diversification. *Trends in ecology & evolution*, *24*(8), pp.456-465.

Wilmé, L., S. M. Goodman, and J. U. Ganzhorn. 2006. Biogeographic evolution of Madagascar's microendemic biota. Science**312**:1063–1065.

APPENDIX 1.

Specimen ID	Genus	Species	Latitude	Longitude
1	*Adansonia*	*madagascariensis*	-14.400072	48.021352
2	*Adansonia*	*suarezensis*	-12.313285	49.336755
3	*Adansonia*	*suarezensis*	-12.313304	49.336788
9	*Adansonia*	*perrieri*	-12.488021	49.171227
11	*Adansonia*	*perrieri*	-12.681327	49.265419
12	*Adansonia*	*suarezensis*	-12.825883	49.263007
13	*Adansonia*	*perrieri*	-12.912677	49.199832
23	*Adansonia*	*madagascariensis*	-12.952348	49.128891
28	*Adansonia*	*madagascariensis*	-12.953985	49.127796
29	*Adansonia*	*za - perrieri intermediate*	-13.009238	49.466315
34	*Adansonia*	*madagascariensis*	-13.073107	49.641809
35	*Adansonia*	*madagascariensis*	-13.038739	49.536659
37	*Adansonia*	*za*	-13.755943	48.464887

38	*Adansonia*	*za*	-13.751909	48.464763
43	*Adansonia*	*madagascariensis*	-13.756581	48.360349
55	*Adansonia*	*za*	-16.380652	46.65465
60	*Adansonia*	*za*	-19.153416	44.807328
68	*Adansonia*	*rubrostipa*	-19.471019	44.51574
72	*Adansonia*	*rubrostipa*	-19.567427	44.527859
75	*Adansonia*	*rubrostipa*	-19.74947	44.583397
76	*Adansonia*	*rubrostipa*	-19.813435	44.586302
81	*Adansonia*	*za*	-20.072752	44.653399
85	*Adansonia*	*rubrostipa*	-20.11347	44.543008
92	*Adansonia*	*grandidieri*	-20.226403	44.430071
96	*Adansonia*	*za*	-21.346956	44.308989
108	*Adansonia*	*grandidieri*	-21.796134	43.835135
111	*Adansonia*	*grandidieri*	-21.79669	43.690202
117	*Adansonia*	*rubrostipa*	-21.866135	43.66286
122	*Adansonia*	*za*	-21.842691	43.771898
127	*Adansonia*	*rubrostipa*	-24.045826	43.753756
135	*Adansonia*	*za*	-23.212232	44.042275
136	*Adansonia*	*za*	-23.201938	44.049406
137	*Adansonia*	*za*	-23.196628	44.051668
143	*Adansonia*	*za*	-22.490774	46.293627

Chapter Three:

Evidence for Hawkmoth Pollination in a Classic Chiropterophilous Species, the African baobab (*Adansonia digitata*)

ABSTRACT

We investigated multiple aspects of the pollination biology and pollinator visitation dynamics of the African baobab, *Adansonia digitata*, to explain observed disparity in fruitset observed in populations across the species range. Given the delivery of pollen does not appear to be responsible, this points to receipt of pollen being a driving factor. Despite the African baobab having characteristics associated with bat-pollination, within population studies in South Africa we found an absence of bats and hawkmoth playing a significant role in pollination. To test for potential drivers of pollinator foraging differences, we investigated floral scent profiles, and nectar composition for primary sugars and amino acids. We found significant differences in nectar sugar ratios between trees previously published as having significant disparities in fruit set success. Floral fragrance profiles reveal geographic heterogeneity of scent compared to previously published work in Senegal. Our results suggest maintenance of variability in pollination characteristics could explain fruitset disparity as an evolutionary mechanism for fluctuating pollinator populations. Alternatively, such heterogeneity may be explained by evolutionary transition of pollination systems.

INTRODUCTION

Pollination is a dynamic process, with a number of factors that can influence pollen movement and reproductive success. Intraspecific variation in reproductive biology can include differences in floral sexual traits (Herrera 1993), selfing ability (Jarne & Charlesworth 1993) or resource allocation (Holland et al. 2004). In animal-pollinated systems, variations in floral traits (Ohashi & Yahara 1999), flowering phenology coupled with pollinator populations dynamics, and landscape patterns such as density and composition (Kunin 1997, Grindeland et al. 2005) can influence attraction of pollinators and foraging patterns. Efficacy of specific pollinators is dependent on characteristics of the plant population and modes of pollen transfer. Given that plant-pollinator interactions are not static nor is any one factor exclusive, many aspects of a pollination systems may lead to intraspecific disparities in reproductive success and fruit production.

The African baobab, *Adansonia digitata,* is widespread across continental Africa, with populations in West, East, and Southern Africa. Across their range, it has been noted that fruit set is highly variable within populations with individual trees ranging from ones that produce hundreds of fruit every year to ones that produce very little fruit, if any, year after year (Assogbadjo, 2008; Venter & Witkowski, 2011). This pattern is so marked that humans across Africa (i.e. Benin, South Africa, Tanzania) often classify individual baobabs as either “female (consistently produce fruit) or “male” (rarely if ever produce fruit), hereafter referred to as “producer” and “poor-producer” trees (Venter & Witkowski, 2011; Venter et al. 2015). Theoretically variation in fruit production among trees could reflect differences in (1) degrees of sterility given the overall pollen donor pool, (2) the ability to produce fruit without cross pollen (i.e., low rates of pollen flow between trees combined with variation in the strength of self-incompatibility or capacity for apomixis), (3) floral morphology, affecting pollen transfer and pollinator efficacy, (4) intrinsic differences in net attractiveness to pollinators, such as due to

floral chemistry, coloration, or display size (e.g., Gould 1978; Alder & Irwin 2012), or (4) geographical factors that might affect pollinator foraging behavior, such as location relative to other food resources or roost sites (e.g., Nattero et al. 2011).

Previous research has attempted to understand the pronounced differences in fruit production. After the proposal of a co-existing diploid African baobab (Pettigrew et al., 2012), the hypothesis of sterile triploids was proposed to explain the first hypothesis above, but it has since been shown that all baobabs are indeed tetraploid (Cron et al., 2016). Resource allocation trade-offs between number of flowers and fruit set has also been dismissed (Venter 2012). It has been found that poor-producer trees set fruit successfully when pollinated with cross pollen, which contradicts the first hypothesis we presented above (Venter et al., 2016). The second hypothesis described above is plausible given examples of species with variable levels of leaky self-incompatibility (Reinartz & Les, 1994; Zhang et al., 2014). While leaky self-incompatibility (Lobo et al., 2004) and apomixis (Oliveira et al., 1992; Mendes-Rodrigues et al., 2005) have been documented in another bombacoid species, *Ceiba pentandra*, it is not known if these occur in baobabs or if there is intraspecific variation in this trait. While Venter et al. (2016) concluded that baobabs are largely self-incompatible, one self-pollination flower did produce fruit suggesting possible “leakiness”. Therefore, this warrants further evaluation with a greater sample size. The third hypothesis, suggesting difference in floral traits, has not been adequately investigated, although variation in floral morphology has been shown to vary continuously without signs of a bimodal distribution (Cron et al. 2016). The fourth and fifth hypotheses also deserve additional consideration.

Baobabs have characteristics typically associated with bat pollinated species (Faegri and van der Pijl, 1979). Their flowers are borne on long, pendent peduncles and are nocturnal, usually opening at dusk and losing receptivity the following morning (Breitenbach &

Breitenbach, 1974; Baum, 1995a). Their morphology fits the classic chiropterophily syndrome with large white flowers, abundant nectar, and a "musty odor" (Faegri and van der Pijl, 1979). Indeed, previous research in West and East Africa has shown that baobabs are primarily bat pollinated (Jaeger 1945, 1950, 1954; Harris & Baker, 1959; Start, 1972; Ayensu, 1974; Baum, 1998; Djossa et al., 2015), although some pollination, at least in East Africa, may be due to bushbabies (Coe & Isaac, 1965). In contrast to the abundant work in East and West Africa, there has been no documentation of pollinator visits in Southern Africa.

Here we attempt to compile a comprehensive understanding of the pollination ecology of *Adansonia digitata* in South Africa, complemented by additional comparative studies in East Africa. We used pollinator observation and pollinator exclusion experiments to evaluate the role of different floral visitors and differences between different continental regions, crossing experiments to assess self-incompatibility/apomixis, analysis of floral scent and nectar profiles were conducted to compare high- and low producer trees and look for geographic differences, and microsatellite-based parentage analysis was used to estimate gene flow patterns. Our evidence suggests that hawkmoths, not bats, are currently the dominant pollinators in South Africa. Differences in floral scent at the continental scale may indicate some degree of local adaptation to hawkmoth pollination in South Africa, but cannot explain tree-to-tree fruit set differences. However, nectar analyses suggest that differences in sugar concentrations may drive selecting foraging. Together, our results suggest the possibility of different pollination modes existing across the African baobab's range.

METHODS

I. Pollinator Observations

Fieldwork was conducted at sites in the Limpopo Province of South Africa from November 01–30 and December 18-22 in 2015 and November 01-15, 2016. The dominant vegetation type is Mopane Bushveld; though trees are found within and around communal areas, with extensive farming and grazing occurring at most sites. A subset of the trees studied had previously been classified as "producer" and "poor-producer" based on multiple years of fruit-set data (Venter & Witkowski, 2011).

Nocturnal observations were conducted to document visitors and the frequency of their visits. Visual observations started at sunset, approximately 1900 hr local time, and continued typically until at least 2100 hr, corresponding to *crepuscular pollinators active periods*. Floral visitors were documented by direct observations conducted either from the ground or from within the tree canopy. Access into the tree canopy was by the single rope technique described by Smith & Padgett (1996). Flowers were illuminated by use of a red LED headlamp (Night Eyes LLC, PA, United States). Number of visits were recorded based on pollinator group classified as hawkmoths (Sphingideae), other large moths (i.e. Erebidae), and other (e.g., bushbabies). Visits were counted only if floral contact was made to the reproductive parts (anthers or stigma). The number of open flowers directly within a field of view was recorded to estimate the number of visits per flower per hour. To complement our observations, in the early morning we looked for evidence of claw scratch marks on the petals of flowers within the canopy or fallen flowers; as used by Djossa et al. (2015).

Wildlife cameras were also staged to document visitation outside of designated observation windows. Motion-triggered infrared wildlife cameras, a Reconyx XR6 UltraFire (Holmen, WI, USA) and a Moultrie M-1100i (EBSCO Industries, Inc., USA) were placed in

trees pointing at one to three flowers to collect photographs and record videos overnight. Photographs were downloaded and manually sorted to record visiting organisms and visitation frequency per individual flower.

II. Pollination Experiments

Pollination experiments were conducted on five trees. Open pollination/no treatment was also quantified for two additional trees, and three additional trees were used for pollinator exclusion experiments. Flowers for each treatment were selected indiscriminately.

a. *Pollinator Exclusion* - Flowers were enclosed in a chicken-wire cage (with 2.5cm-diameter hexagonal gaps) prior to anthesis. Different cages were constructed to meet two objectives: 1) Tubular, open at the bottom, that allow hawkmoths to access flowers from the bottom yet deter bat visits, and 2) Full cages that exclude bats, bushbabies, and hawkmoths, but allow bees and other small insect visitors.

b. *Self-incompatibility and apomixis* – Individually tagged flowers to be used for treatments were bagged in the late afternoon-early evening prior to anthesis, using either plastic pollination bags or brown paper bags.
    a. To test for apomixis, flowers were emasculated before anthesis and re-bagged.
    b. To test for passive self-pollination, flowers were left unemasculated and bagged.
    c. To evaluate self-compatibility, bagged flowers (unemasculated) were pollinated by manually rubbing stigma with the staminal bundle from another flower from the same tree.

d. To evaluate the efficiency of cross pollination, hand pollinations of flowers were conducted. Flowers were again left unemasculated to simulate natural conditions. Flowers from pollen donors were collected in late afternoon-early evening from either neighboring trees or elsewhere. Upon collection, the calyx was split open to confirm anthers had dehisced. A minimum of one hour after anthesis, when petals were in a fully reflexed position, bagged flowers were pollinated using one pollen donor, by manually rubbing anthers directly onto the receiving style. A subset of bagged flowers were cross pollinated in the morning (between 06:00 – 07:30 hr) to infer receptivity duration.
e. Open pollinated control flowers were tagged and left open (unaltered) to quantify natural fruit-set rates.

In all cases, except open pollination, flowers were bagged until floral organs abscised. Flowers were tracked to quantify fruit set except for a subset of the actively self- and cross-pollinated flowers (from "producer trees"), whose styles were collected to observe pollen tube growth and assess for late-acting self-incompatibility, as seen in *A. gregorii* species (Baum, 1995a). For these samples, twelve hours following self-pollination, ovaries and styles were collected and placed in 3:1 absolute ethanol: glacial acetic acid. After two to four hours, the styles were transferred to 70% ethanol for long-term storage until microscopy, described below.

III. Microscopy analyses of pollen-tube growth to assess late-acting self-incompatibility (LSI)

Pollen tube growth was observed by florescent imaging of callose using leuco-aniline blue (Baum 1995a). Stored self-pollinated and cross-pollinated pistils were rinsed in MilliQ water, cleared and softened in 8N NaOH at 60ºC for minimum of 20 minutes, and then left at room temperature for 24 to 72 hours. Pistils were again rinsed in MilliQ water and immersed in Tris-Buffer (pH 8.4) for 20 minutes. Pistils were dissected, mounted, and stained with one drop of 0.1% leuco-aniline blue with 20% glycerin (v/v/) and left for 24 hours in the dark at 4ºC. Visualization used an Olympus BX60 Epifluorescence Microscope at the Newcomb Imaging Center at the University of Wisconsin, Madison, with a WU filter cube under UV excitation at 370nm and emission at 509nm with images captured on an Olympus DP70 digital camera.

IV. Floral Fragrance Profiles

Scents were collected from ten trees in South Africa from multiple flowers (two to five) per individual tree for a total of 34 flowers. We also collected control samples from bags without flowers. Three trees were classified as "producer" trees and four as "poor producer" trees, designated based on Venter & Witkowski 2011, to test for significant differences between floral profiles. The remaining three South African trees were unclassified. Additionally, we collected from 16 flowers on five different trees in Tanzania, December 3-15, 2015.

Volatile Collection – Floral fragrance profiles of volatile organic compounds (VOCs) were determined by use of the active head-space absorption method. Volatiles were collected by bagging flowers just prior to, or soon after, the onset of anthesis with Reynolds Kitchens™ Oven Bags (Reynolds Consumer Products, IL, USA). Scent was collected by use of a battery-operated AirLite pump (SKC Inc., PA, USA) at an inflow rate of 200 ml/min for two hours and trapped in

glass tubes packed with 200mg of 50-80 mesh Porapak™ Q porous polymer adsorbent (Supelco Inc., PA, USA). Samples were immediately eluted from the absorbent with 200uL of GC-MS grade 95% n-hexane (Fisher Scientific Inc., NH, USA) into teflon-capped borosilicate glass vials and stored at -80°C upon returning from the field, until analysis.

Gas Chromatography-Mass Spectrometer (GC-MS) Analyses - Samples were prepared for GC-MS by combining 40µL of sample with 2µL of 1µL/ml m-Xylene (in n-hexane) as an internal standard. Samples were analyzed with a TRACE1300/1310 Gas Chromatograph (Thermo Fisher Scientific Inc., MA, USA) with a DB-5 column (30m long, 0.250mm internal diameter and 0.25µm film density). Samples were injected with a TriPlusRSH 57mm injection needle (Thermo Fisher Scientific Inc., MA, USA) with a split flow at 60.0:1 and a purge flow of 5.00. GC was programmed with an initial oven temperature of 45.0 ºC and an increase of 3.00 ºC/min., with a target value of 175.0ºC, followed by an increase of 25ºC per minute till 280.0ºC. Helium was used as the carrier gas. We scanned for masses in the range of 25-300 amu.

Compounds were identified by comparing Kovats retention indices derived from C8-C20 calibration standards and mass spectra to those in the NIST mass spectral library, the Chromeleon Chromatography Data System Software reference library and Pherobase (El-Sayed, 2017; http://www.pherobase.com/). Retention indices for compounds eluting before alkane C8 were identified based on mass spectra and relative elution times.

V. Nectar Composition

Sugar (i.e. sucrose and hexose), amino acids, or fatty acid composition of nectars are thought to influence the visitation rate of different pollinators (Baker & Baker 1975; Gottsberger et al.,

1984; Rodríguez-Peña et al., 2013). Therefore, we sought to compare these macronutrients on "producer" versus "poor-producer" trees.

Nectar samples were collected from 20 trees, 10 classified as "producer" and 10 as "poor-producer," (Venter & Witkowski 2011) using microcapillary tubes (Drummond Microcaps, Drummond Scientific Company, PA, USA). Nectar samples were extracted destructively after flowers were fully open, as determined by the reflexed position of the calyx and petals. Timing of collection varied somewhat from tree to tree, but typically occurred about one hour after sunset. Volume was recorded, and a drop of nectar was placed on a hand-held refractometer to estimate sugar concentration. The remaining sample was diluted at 1:4 in absolute ethanol and stored in glass vials at -80°C upon returning from fieldwork until analysis.

Derivatization of sugars for GC-MS analysis was adapted from Wang et al. (2017). Nectar samples were spun for 5 minutes at 20,000g and 100µL of supernatant was transferred to a glass GC vial and dried down in by centrifugation under vacuum at 32ºC for approximately 5 hours until a pellet formed. The pellets were resuspended in 40µL of anhydrous pyridine containing 15mg/mL methoxyamine hydrochloride. Vials were sonicated for 10 minutes and then incubated at 50ºC for 1.5 hours, interrupted by brief vortexing and sonication for 10 minutes after 1 hour. Following another round of vortexing and sonication, 40µL of N-methyl-N-trimethylsilytrifluoroacetamide with 1% Tertbutyldimetheylchlorosilane (MSTFA + 1% TMCS) was added to each vial. Samples were then vortexed, incubated at 50ºC for 1 hour and analyzed via GC-MS as described below.

An internal standard of 100mM L(+)norvaline (Arcos Organics, Thermo Fisher Scientific, MA, USA) was prepared (11.715mg/1ml DI water) and diluted with methanol to a final concentration of 100µM. Nectar aliquots to be derivatized for amino acid analysis were first treated to remove lipids. Nectar samples were dried under vacuum overnight, after which 400µL

of a 2:1 methanol:chloroform plus 100µM norvaline was added. Samples were sonicated for 5 minutes. 300µL of deionized water and125µL chloroform was added. Samples were vortexed for 1 minute and centrifuged at 10,000g for 5 minutes, separating the polar and nonpolar phases. Approximately 400µL of the aqueous phase was dried down in a speed-vac and the resulting pellet was dissolved in 100µL of methanol and sonicated for 5 minutes. Samples were centrifuged at 10,000g for 5 minutes and 100µL of supernatant was transferred into a glass GC vial insert.

To derivatize amino acids, nectar samples were dried down in a speed-vac at 32ºC for approximately 5 hours, until a pellet formed. 40µL of anhydrous pyridine were added to each sample along with 40µL of MTBSTFA (N-Methyl-N-[tert-butyldimethylsilyl] trifluoroacetamide with 1% Tertbutyldimetheylchlorosilane [t-BDMCS], catalog# M-108-5X1ML, Millipore Sigma, St. Louis, MO, USA). Samples were vortexed for 10 seconds, sonicated for 5 minutes, then incubated at 70ºC for 60 minutes.

Nectar samples were run as in Wang et al. (2017) for sugar analysis. Standards of glucose, fructose, sucrose, xylose, and arabinose, maltose and mannitol were also run at serial dilutions and standard sugar curves calculated. An amino acid standard was created with all 20 amino acids (except glutamine, asparagine and tryptophan which were run individually) with each at 2.5uM/mL, except cystine which had a concentration of 1.25uM/mL. Differences in concentrations between producer and low produce trees were compared by t-tests with a Bonferroni correction.

VI. Estimates of Gene Flow via Parentage Analysis

Scoring of microsatellite markers was used to infer gene flow estimates from parentage analysis. DNA was extracted from silica-dried leaf tissue or seeds using the DNeasy Plant Mini Kit (Qiagen Inc., Valencia, CA, USA) following the manufacturer's protocol with the following modifications: 1) Lysis Buffer AP1 was increased from 400µl to 650µl, and 10µl of Proteinase K (25mg/mL) was added, 2) tissue and lysis buffer were incubated at 65°C for 20 minutes instead of 10 minutes at 65°C, 3) neutralization buffer P3 was increased from 130µl to 195µl, 4) centrifugation steps were performed at 4°C instead of room temperature, with the exception of the last step, and 5) final elution was made with heated buffer (~85°C) and incubated at room temperature for ten minutes before centrifugation.

Eight *Adansonia digitata* microsatellite loci were selected from Larsen et al. (2009) based on successful polymerase chain reaction (PCR) amplification. Forward primers were synthesized and labeled with fluorescent dyes using Dye Set D (6FAM, HEX, TAMRA) by Eurofins Genomics (Eurofins Scientific Inc., Luxembourg). PCR reactions totaling 25µL in volume included 0.5-1µL of DNA template, 0.5µL of 10µM forward and reverse primers, 2µL of $MgCl_2$, 5µL of GoTaq Flexi buffer and 0.2µL of Flexi Taq polymerase (Promega, Madison, WI, USA), and 15.3µL of MilliQ water. PCR thermocycler parameters were as in Cron et al. (2016). Successful amplification was determined by bands in the expected size range on a 1.5% agarose gel stained with ethidium bromide. Successful PCR products were diluted based on flurophor dye type (1:40 dilution for 6FAM-dyed primers, 1:40 dilution for HEX-dyed, and 1:20 dilution for TAMRA-dyed) and 2µL of PCR product was added to 10µL of HiDi formamide (Fisher Scientific Company LLC., Pittsburg, USA) and 0.3µL of GENEFLO™ ROX-labeled 625 ladder (CHIMERx, Milwaukee, WI, USA). Fragment analysis was performed on an ABI 3170 at the University of Wisconsin- Madison Biotechnology Center (Madison, WI, USA). Chromatographs

of microsatellite alleles were scored manually in Geneious version 5.0 with the Microsatellite Plug-in.

Estimates of contemporary gene flow were based on parentage assignments for 10 seeds from 10 different fruit on a single "producer" focal tree. We genotyped the focal (maternal) tree and all potential pollen donor within a 350-meter radius of the focal tree, as well as 5 trees that were further away. Parentage assignment was performed manually by comparing shared alleles between the mother tree, offspring and potential fathers.

RESULTS

I. Pollinator Observations

After 62 hours of dedicated observations, as well as time conducting pollination treatments over the course of three weeks, no bats were ever seen visiting baobab flowers. Furthermore, we found no evidence of claw marks by bat visitations on either flowers in the tree canopy or on the ground. Instead of bats, hawkmoths (*Nephele comma*, *Agrius convolvuli*, and occasionally *Hippotion rosae*) were abundant visitors and often contacted the reproductive organs. Fifteen foraging hawkmoths were collected to confirm identification and specimens were deposited at the University of Venda, South Africa. Other moths seen visiting flowers included sundowner moths (*Sphingamorpha chlorea*), which were observed crawling on flowers. Additionally, both lesser bushbabies (*Galago moholi*) and greater thick-tailed bushbabies (*Otolemur crassicaudatus*) were observed visiting flowers. The contributions of bushbabies to cross-pollination are likely to be minimal due to their foraging on multiple flowers within an individual tree. Over 62 flower-hours of observation, we counted 266 hawkmoth visits (4.3 visits per

flower-hour), 81 non-sphingid moths (1.3 per flower-hour) and 4 bushbaby visits (<0.1 per flower-hour). Additional nocturnal floral visitors included other smaller moths, ants and unidentified beetles, considered most likely nectar-robbers and pollen-foragers.

Staged wildlife cameras totaled over one hundred hours flower-hours of observations, observing hawkmoths and one greater bushbaby visit (Fig. 1). Based on camera imaging of a single flower there was a hawkmoth visit approximately every 20 minutes between the hours of 20:00 and 22:00 and every hour thereafter until 4am.

Figure 1**.** Pictures from infrared camera capturing A) hawkmoth visitation of baobab flower, and B) greater bushbaby.

II. Pollination Experiments

A summary of the results for each pollination treatment are displayed in Table 1. Bat exclusion cages that allowed hawkmoths to visit (confirmed by observations) resulted in a total of 11/155 (7.1%) flowers yielding fruit. In contrast, flowers enclosed in cages that excluded both hawkmoths and bats (but allowed, bees and other small insects) yielded no fruit (0/49). This difference is nearly significant (chi-square $p = 0.055$). If we combine the 510 open-pollinated flowers, of which 48 (9.4%) set fruit, with the tubular (non-hawkmoth-excluding) cages, then the lack of fruit set in the fully caged flowers is judged significant (chi-square $p = 0.036$). Fruit-set

of control open pollinated flowers ranged from 2-16% depending on the individual tree, with a total fruit set of 48/510 (9.4%). The fact that no bat visits were seen and that hawkmoth-accessible/bat-excluded flowers had a similar fruit set to open-pollinated flowers whereas bat and moth-excluded flowers set no fruit suggests that hawkmoths are the main pollinators in these populations.

None of 32 flowers that had been emasculated and bagged set fruit, suggesting an absence of apomixis. Likewise, self-pollinated flowers yielded no fruit (0/79), whereas 71% (39/55) of nocturnally cross-pollinated flowers set fruit. This suggests quite effective self-incompatibility in *A. digitata*. Only three of 46 flowers cross-pollinated in the early morning (6.5%) set fruit, representing a significantly lower success rate compared to nocturnally-crossed flowers (chi-square p = 2E-05).

Table 1. Pollination treatment sample sizes and fruit set success rate.

Treatment	Fruit/flowers for each tree	Total fruit/flowers	%
Cages: bat-excluded, open to hawkmoths	0/5; 2/7; 0/4; 0/13; 0/4; 5/8; 3/80; 1/34	11/155	7.1
Cages: hawkmoth & bat excluded	0/6; 0/6; 0/6; 0/5; 0/14; 0/12	0/49	0
Self-pollinated	0/5; 0/5; 0/38; 0/21; 0/10	0/79	0
Emasculated & bagged (apomixis)	0/2; 0/6; 0/7; 0/12; 0/5	0/32	0
Cross-pollinated (nocturnal)	2/2; 4/5; 5/6; 4/6; 14/22; 10/14	39/55	71
Cross-pollinated (diurnal)	0/3; 0/27; 3/32; 0/3; 0/2	3/67	4.5
Open/no treatment	0/11; 7/17; 0/20; 0/43;	48/510	9.4

	4/95; 33/196; 11/128;		

III. Microscopy Analyses of pollen-tube growth

Thirteen self-pollinated styles were collected and their pollen tubes were compared with ten cross-pollinated styles (five from producer and five from poor-producer trees). Microscopic observations confirmed pollen tube formation and migration down the style in all flowers, with no difference based on pollen source (self vs. cross) or tree (producer vs. poor-producer). Combined with evidence of self-incompatibility from the pollination treatments described above, these data suggest that *A. digitata* shows late-acting self-incompatibility, as previously documented for *A. gregorii* (Baum 1995a).

IV. Floral Fragrance Profiles

The volatile organic compounds (VOCs) in *A. digitata* floral fragrances consisted of isoprenoids (monoterpenes and sesquiterpenes), sulfurous compounds, fatty acid ester derivatives, and benzenoids. Any compounds also detected in the controls were eliminated. Year 1 sampling included two South African trees (one “producer”, one “poor producer”) and five Tanzanian. Year two sampling was restricted to 10 South African trees, but perhaps due to sample instability the resulting mass spectra were of significantly poorer quality, with several compounds undetected, and not readily comparable to year 1. Therefore, we used year 1 data to consider geographic patterns and year 2 to assess the possibility of floral VOC difference between

"producer" and "poor-producer" trees. We compared our results to those previously reported for *A. digitata* in Senegal by Pettersson et al. (2004) in Table 2; Figure 3.

The compound 2-Hexenal was identified by us and by Pettersson et al. (2004), but since it also occurred in our controls we cannot be certain of its authenticity and therefore it was excluded from consideration. The most abundant compound in our samples was sesquiterpene *β*-caryophyllene, accounting for 60%-77% of the total profile (Table 2). This compound is associated with a sweet scent and is common in moth-pollinated flowers (Knudsen & Tollsten, 1993). Isomers iso-caryophyllene, *α*-humulene (syn. α-caryophyllene), and derivative *β*-caryophyllene oxide, were also found in varying amounts across samples. *C*aryophyllene compound was not detected in baobab scent by Pettersson et al. (2004), despite the fact that their method was suited to its detection and they found large amounts of this compound in other species studied.

A number of other isoprenoids were found including *β-elemene,* camphene, limonene, ß-Ocimene, a-Pinene, and germacrene D, all occurring in relative quantities of less than 1%- 2% of relative total. Isoprenoid ß-myrcene was detected in South African trees (from year 1), with an average relative abundance of 6.70%, but was completely absent from trees in Tanzania. In general, there was a greater abundance of monoterpenes, but a lower abundance of sesquiterpenes in South Africa than in Tanzania. In contrast, Pettersson et al. (2004) reported a complete lack of sesquiterpenes in Senegal, but a lot of "irregular terpenes, though some of which are very similar to, or possibly the same, compounds we classified as fatty-acid derivatives.

An abundant compound in the majority of samples were a pair of isomers, eluting with a retention index between 892-897 (Table 2), accounting for 2-6% of total. Based on the mass spectra, we identified these compounds as 9-Oxabicyclo-nonane stereo-isomers, which have

previously been reported in the essential oil of *Matricaria pubescens*, Asteraceae (Boutaghane, 2010). While this compound was not reported by Pettersson et al. (2004) by name, a pair of unidentified steroisomers detected in their study matches the most abundant ten ions of 9-Oxabicyclo-nonane.

As a chemical class, fatty acid derivatives were the numerous; We identified several of which are associated with "mushroom-like", pungent, fruity and "green-type" smells, including butanoic acid, butyl ester; 1-butanol 3-methyl- propanoate; 2-methyl-1-butyl acetate; 1-butanol, 3-methyl-, acetate, butyl acetate, and two enantiomeric forms of (*R*)- and (*S*)-2-methylbutanoic acid (see Table 2). These compounds appeared to be more abundant, on average in South Africa and Senegal than in Tanzania. A number of these or similar fatty acids have been reported in the bat-pollinated Bombacoid species *Ceiba trichastandra* (Knudsen & Tollsten 1995) and *Ceiba pentandra* (Pettersson et al. 2004).

Sulfur compounds are commonly associated with floral scent profiles of bat-pollinated plants (see discussion); we identified four sulfur compounds. Dimethyl-disulfide, which has an earthy or musky aroma even in very small quantities, was most nearly ten-fold more abundant in samples from South Africa than samples from Tanzania (Table 2). Additional sulfurous compounds detected included di-methyl tri-sulfide, S-methyl 3-methylbutanethioate, and methyl thiobutanoate. Based on mass spectra, we also recovered trace amounts of S-Methyl thiohexanoate. The only benzenoid compound found, the almond-smelling benzaldehyde, was occasionally detected in trace amounts.

We conducted a principal component analysis (PCA) of all VOCs detected and relative area for each sample collected. The first two principal components explained 83% of the variation in year 1 samples. The profiles from Tanzania were separated from the two South

African trees on PC 2 (Fig. 3). This separation seems driven by higher levels of dimethyl-disulfide and ß-myrcene in South African trees than Tanzanian ones.

Notwithstanding, the lower number of compounds detected in year 2 compared to year 1, we conducted PCA analysis to compare "producer" and "poor-producer" trees in South Africa (Fig. 3). 92% of the variance was explained by the first principal component given the two stero-isomers Oxabicyclo, but although the four lowest values on this axis came from "producer" trees, there was not significant difference between the two classes. The second principal component pulls out two flowers, from a single "poor-producer" tree, as having high values. However, since this axis explains just 4% of the variance, and the sample size is low, this result is not significant.

Table 2. Floral scent profiles for *A. digitata* of South Africa and Tanzania in year 1. Volatile organic compounds are organized by chemical class with Kovats Retention Index (RI) based on C8-20 alkane standards. Average relative abundance given as the percent of total as mean of area of GC peak and the standard deviation. Any compounds less than 0.1% of total are listed as trace (tr). Number of trees *(n)* and total number of flowers sampled per population is noted.

Chemical Class and Compounds	Retention Index (RI) log, linear	Mean relative abundance (% of total)	
		Tanzania *n=6 (16)*	South Africa *n= 2 (4)*

Isoprenoids			
Monoterpenes			
ß-myrcene	991, 989	-	6.70
D-Limonene	1031, 1028	0.17	0.84
ß-Ocimene	1040, 1036	0.27	0.28
1R-a-Pinene	935, 930	0.13	0.23
Camphene	949, 944	0.19	0.23
Sum		**0.76**	**8.28**
Sesquiterpenes			
α-copaene	1377, 1376	0.15	0.04
ß-Bourbonene	1385,1384	0.16	-
ß-Elemene	1393, 1392	0.12	0.05
Iso-caryophyellene	1406, 1406	2.47	0.51
β-caryophyellene	1421, 1420	77.61	60.1
α-humulene	1454,1543	4.73	1.08
Caryophyllene oxide	1583, 1583	2.16	0.99
Germacrene D	1482, 1481	0.49	0.09
Sum		**87.89**	**62.85**
Sulfurous			
di-methyl, di-sulfide	-	0.84	10.15
S-Methyl 3-methylbutanethioate	942, 937	0.15	1.68
Butanethioic acid, S-methyl ester	888,886	0.12	0.56
di-methyl tri-sulfide	971, 967	0.29	1.14
Sum		**1.40**	**13.53**
Fatty acid ester derivatives			
Butanoic acid, 2-methyl-, methyl ester	-	1.18	0.22
Butyl acetate	818, 815	-	0.65
1,7-Octadien-3-ol	877, 872	0.24	0.25
1-Butanol, 3-methyl-, acetate	879, 875	-	2.0
2-Methyl-1-butyl acetate	881, 877	-	0.54
2-Methyl-2-heptanethiol	913, 911	-	0.10
Butanoic acid, 3-methyl-, propyl ester	954, 949	0.11	0.35
1-Butanol, 3-methyl-, propanoate	974, 970	0.10	0.38
Butanoic acid, butyl ester	997, 997	-	0.56
Sum		**1.63**	**4.95**
OTHER			
Octane	800, 801	0.94	1.27
Acetyl valeryl	838, 833	0.28	0.28
1-Undecyne	1027,1024	0.28	0.05
Heptadecane, 2,6-dimethyl-	1203, 1202	0.08	-
9-Oxabicyclo[6.1.0]nonane-a	893, 892	3.92	4.87
9-Oxabicyclo[6.1.0]nonane-b	897, 896	2.51	3.00
Ions: 97 55 57 67 42 68 29 27 41 79			
Sum			

		11.27	**19.47**
Benzenoids			
Benzaldehyde	961, 956	tr	tr
Other			
DL-2-methylbutyric acid	-, -	tr	-
Hydroxyamine O-decyl-	900, 900	0.13	0.79
7-Tetradecene	1395,1394	0.098	tr

Comparing our floral scent profiles of *Adansonia digitata* to published profiles from populations in West Africa (Senegal) by Pettersson et al. 2004 suggests some degree of intraspecific variation in overall scent profile within this widespread species corresponding to geography. Interestingly, the most abundant compound we identified, *β*-caryophyellene nor its isomers were reported in the profiles recovered by Pettersson et al. (2004). Of the additional isoprenoids we found, only Ocimene was reported. The ions of stereo-isomers 9-Oxabicyclo matched those they reported as unidentified compounds, also comprising a large portion of the profile. Our identification of numerous fatty acid derivatives was also consistent with their findings though there are differences in specific compounds. Unfortunately, they did not report on calculated retention indices for comparison. Given the possibility of mis-identifying similar compounds of the same chemical class (most notably fatty acid derivatives), we also compared profiles based on chemical class compositions (Fig. 4). While we did also find 2-Hexenal, this compound was also identified in our controls, perhaps due to collection of ambient fragrance in the air; however, we cannot be certain of its authenticity and therefore was excluded. Given the higher quality from field season one, only those samples were used to compare profiles between South Africa (year 1), and Tanzania in Figure 3.

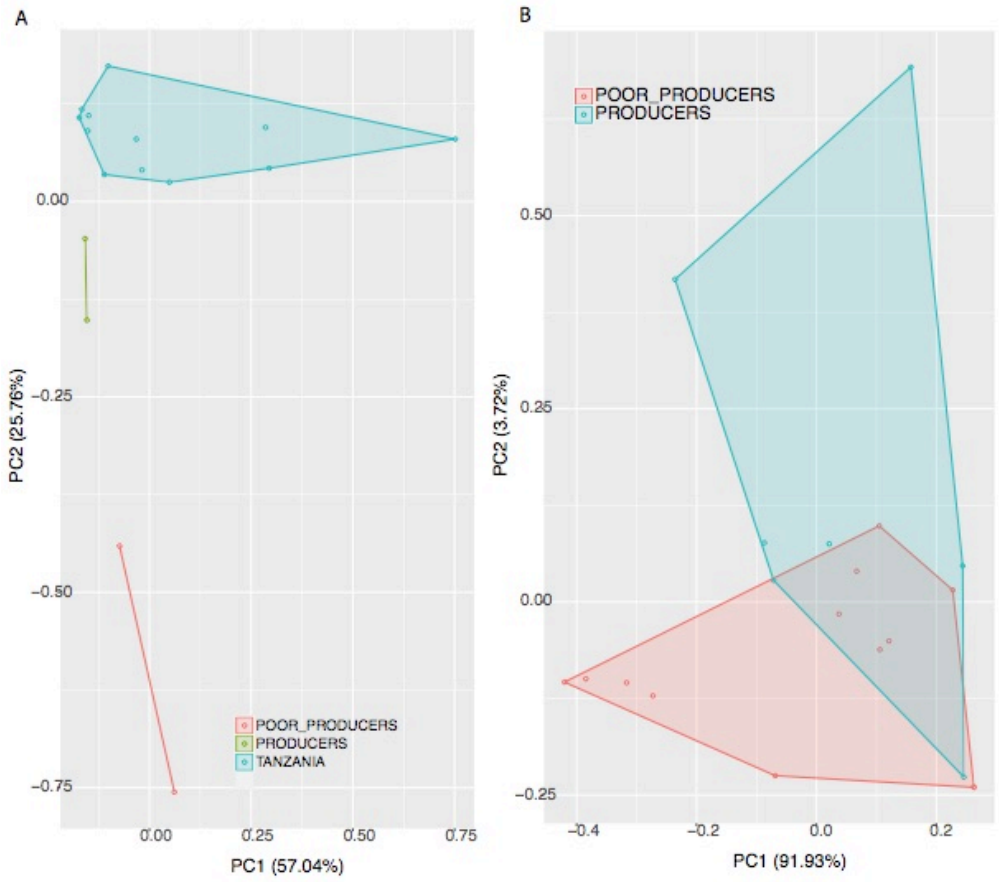


Figure 3. Plot of first two axes of a principal components analysis (PCA) showing the variations in floral scent profiles for A) South Africa and Tanzania from field season one, B) South Africa samples from year two.

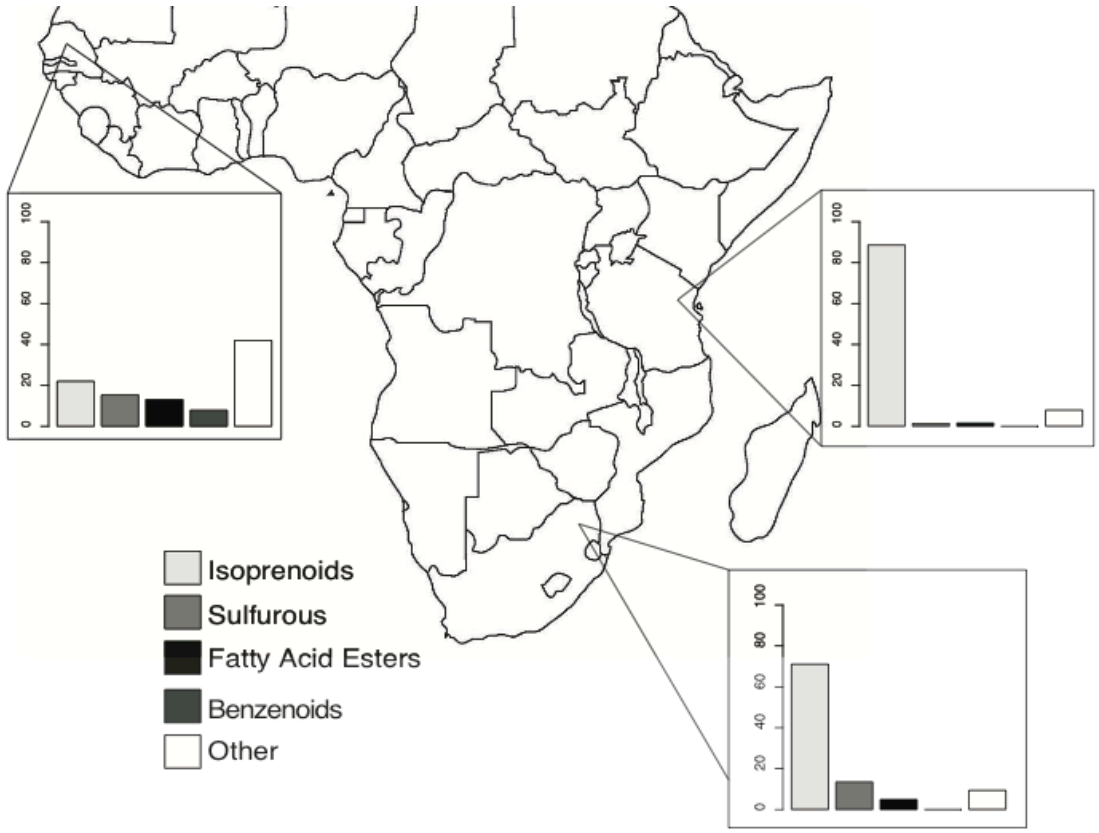


Figure 4. Variation in floral scent profiles based on average relative abundance of each chemical class per geographic population. Floral scent profiles of Senegal summarized from Pettersson et al., 2004.

V. Nectar Composition

Sugars and Amino Acids

Refractometer readings taken for 20 trees yielded nectar sugar concentrations ranged from 16%-23.4% between trees, with an average of 20.1% and a standard deviation of ±2.4%. Sugar analysis via GC-MS allowed for quantification of the primary sugars sucrose (S), fructose (F) and glucose (G), the latter being hexoses. As proposed by Baker (1975) a useful summary of

nectars is its sucrose to hexose ratio (SHR), S/(F+G), with ratios less than 1.0 being "hexose-rich."

We found *A. digitata* “poor-producer” trees to have hexose-rich nectar with a SHR of 0.57. This is consistent with ratios reported for other bat-pollinated species (Wolff 2006; Rodríguez-Peña et al. 2016). In contrast, the “producer” trees have an average SHR of 1.16, matching sucrose-dominant nectar profiles reported for sphingophilous flowers (Baker 1975, Wolff 2006). Interestingly, this is consistent with the presence of hawkmoths in these populations, but lack of bats, suggesting pollination preferences based on significant differences in the primary sugar nectar ($p= 0.003996$). This variation could also explain fruit set disparities in bat-pollinated populations; however, in those populations, we would expect the reverse, with producer trees having hexose-rich nectar corresponding to published bat pollination ratios.

In addition to glucose, fructose and sucrose, numerous other sugars and sugar alcohols were quantified (data not shown), although each of these other sugars and sugar products accounted for less than 1% of total nectar sugar. The disaccharide turanose was the most abundant atypical sugar, accounting for an average of 0.17±X % of nectar sugars in “poor-producers” trees and 0.37±X % in “producer” trees.

Table 3. Dominant sugars (glucose, fructose and sucrose) concentrations found in nectar samples (% total all sugars), with areas of stereo-isomers combined. Kovats Retention Indices (RI) were calculate based on alkane C8-20 standards. Pure standards of glucose, fructose and sucrose were used to verify compounds. Retention indices also matched published literature (Isidorov 2005). Sucrose eluted prior to alkanes C8-9 therefore RI was not calculated.

		Percentage of total (±) average deviation

	Retention Time, (RI log, linear)	Producer Trees *n=10*	Poor Producer Trees *n=10*
Sucrose	39.64 (-)	51.21% ± 6.29	34.31% ± 5.91
Fructose, anti-, syn-	26.26, 26.47 (1906,1923)	29.25% ± 5.97	32.91% ± 5.56
D- Glucose, E, Z	26.77, 27.05 (1932,1947)	14.93% ± 4.95	27.56% ± 6.05

We recovered all amino acids except arginine and selenocysteine (Table 4). Tryptophan was the most abundant amino acid (65% of total) followed by glutamine (14%) and the non-essential ornithine (6%). We found no significant differences (based on t-tests) between any amino acid concentrations in the "producer" compared to the "poor-producer" trees.

Table 4. Amino acids and concentrations found in nectar samples. Average retention times and calculated Kovats Retention Indices (RI) if applicable. Each amino acid had an authentic standard run.

		Percentage of total	
Amino Acids	Retention Times (RI)	Producer Trees	Poor-producer
aspartic acid	30.69	0.03	0.06
leucine	21.70 (1684)	0.11	0.05
glutamic acid	32.73	0	0.05
methionine_	27.22 (1957)	0.06	0.04
serine	27.69 (1980)	0.05	0.32

alanine	18.00 (1517)	0.46	0.34
phenylalanine	29.44	>.01	0.08
valine	20.73 (1640)	0.74	0.47
asparagine	33.18	0.04	0.57
proline	23.04 (1749)	0.55	0.27
tyramine	29.93	0.16	1.06
cystine	31.56	0	0.04
histidine	37.69	1.3	0.92
lysine	34.55	1.89	0.45
isoleucine	22.42 (1718)	1.63	3.63
tyrosine	38.52	0.05	1.21
glycine	18.64 (1546)	>.01	2.4
threonine	28.22	>.01	2.16
glutamine	35.24	6.05	14.15
tryptophan	39.09 (1649)	86.75	65.56

VI. Gene Flow and Parentage Analysis

In manually comparing the genotypes of the focal trees, offspring, and each potential pollen donor we could identify one potential father within a 350 meter radius for six out of ten seeds. Three seeds (from different fruit) were each potentially sired by the same pollen donor. Four of

the ten seeds could not be assigned a parent genotype suggesting relatively long-distance pollen transport; pollen receipt from trees greater than 350m away.

DISCUSSION

The experiments conducted support prior work (Baum 1995; Venter et al. 2016) in suggesting that *A. digitata* has flowers that are receptive primarily for a single night and have late-acting self-incompatibility. This means that fruit set in baobabs is dependent upon nocturnally-active floral visitors that can move pollen between trees. Furthermore, extrapolating from work in *Bombacopsis* (*Pachira*) *quinata* (Bombacoideae, Malvaceae), which concluded that 500 pollen grains are needed for successful fruit set (Quesada et al., 2010), successful pollination in the baobabs, whose fruit contain at least as many seed as *B. quinata*, likely requires either frequent pollinator visits or relatively large-bodied pollinators.

Of the diverse animals we observed visiting the flowers, many are likely to be unimportant for pollination. We observed bees and beetles visiting flowers in the morning, but there is often little available pollen in anthers at this time and our experimental crosses show limited stigma receptivity (wilting and browning of the stigma was often observed at this time). Limited receptivity is also supported by our nocturnal and diurnal cross-pollination data. Likewise, some small-bodied insects, such as settling moths and ants, visit baobab flowers at night, but these are unlikely to be effective pollinators because their small body size limits pollen loads and the rarely touch stigmas. Furthermore, smaller insects would have limited ability to move between trees, since in this study area reproductively mature baobab trees occur with mean densities of less than 2 trees/ha (Venter & Witkowski 2010). Occasional visits by bushbabies were detected, but their low visitation rate and tendency to forage for a long period of time in a

single tree, with few tree-to-tree movements, makes it unlikely that they are responsible for a significant amount of fruit set. Thus, we believe that bats and hawkmoths are the only floral visitors capable of effective pollination.

Although previous research has shown African baobabs are primarily bat pollinated (Jaeger 1945, 1950, 1954; Harris & Baker 1959; Start 1972; Ayensu 1974; Baum 1998; Djossa et al. 2015), our data show that only hawkmoths are significant pollinators in the Limpopo province of South Africa. *Rousettus aegyptiacus*, a species known to forage on baobabs in Kenya as much as 40 miles from their caves (Start 1972), has been recorded near one of our study sites (Bonaccorso et al. 2014, Markotter et al. 2016) but there are no reports of baobab pollination. A few times we observed fruit bats flying near the high canopy of flowering baobabs, but never observed visitation. This is corroborated by interviews with local residents, who do not report seeing bats in baobab trees. However, it is possible that competing food sources might explain their absence. Indeed, a study conducted in nearby Kruger National Park found that the foraging behavior of another potential pollinating bat, *Epomophorus wahlbergi*, varies significantly depending on the local abundance of figs, their preferred food source (Bonaccorso et al. 2014). Thus, it is possible, that the lack of flower-visiting bats could be a recent development, for example due to a temporal mismatch driven by climate-warming, as shown by Hegland et al. (2008). However, given the paucity of historical reports of bats visiting baobab flowers in South Africa, it is also possible that there are evolved differences between baobabs in different parts of Africa such that the trees in southern Africa are relatively unattractive to bats.

In the absence of bats, hawkmoths appear to be the major pollinators of *A. digitata* at our study sites. A high visitation rate of hawkmoths was seen during the night hours, and exclusion experiments showed that these visitors are sufficient to explain fruit set. While long-tongue hawkmoths have traditionally thought to be primarily nectar-robbers, our data suggests

otherwise. However, hawkmoths of the genus *Hippotion*, which we observed, have short proboscises, which might increase pollination efficiency (Peter et al. 2009).

However, despite the high visitation rate, *A. digitata* trees were found to be pollen-limited, as suggested by the 7.5-fold difference between the rate of fruit set in open-pollinated and manually cross-pollinate flowers. Hawkmoths were often seen moving from flower to flower within a single tree's canopy and between trees. Despite the fact that hawkmoths are known to fly long distances when foraging (Janzen 1984), and our evidence from paternity analysis that pollen sometimes moves more than 350m, cross-pollination by hawkmoths may be currently limiting fruit production.

The differences in nectar composition (sucrose to hexose ratios) found between the producer and poor-producer trees, and the current dependence on hawkmoths for pollination, may partly explain why some individual trees produce more fruit; being more favorable to hawkmoths. The sucrose-dominant nectar for "producer" trees is consistent with learned behavior of hawkmoths preferring higher sucrose nectars as opposed to fructose or glucose sugars (Kelber 2003). Additionally, atypical sugar turanose has also been found in the nectar of hawkmoth-pollinated species *Aquilegia* in significantly greater concentrations than that of hummingbird pollinated species (Noutsos et al. 2015). This is not contradicted by the lack of a difference between producer and poor-producer floral volatiles since analytical problems introduced noise into these data. Furthermore, even without long distance signals, hawkmoth are capable of exhibiting learned feeding behaviors (Kelber 2003), and could therefore visit "poor-producers" at a consistently lower rate than producer trees.

Bat-pollination flowers typically have a strong, musty odor, derived from sulfur-containing compounds, which are uncommon in most floral aromas but have been isolated

multiple times from phylogenetically-distinct chiropterophilous species in the Neotropics (Bestmann et al., 1997; Von Helversen, 2000). Pettersson et al. (2004) studied floral scents in eight different African bat-pollinated species, finding no sulfurous compounds in seven species, notably including *Ceiba pentandra* (Bombacoideae, Malvaceae), a species occurring in both South America and Africa. The African baobab, *A. digitata*, sampled from West Africa was the only bat-pollinated taxon that did produce sulfurous compounds. In further analysis, Pettersson & Knudsen (2014) found a total of seven sulfurous compounds, four of which were detected at such low concentrations that their absence in our samples is not definitive. The most surprising difference between the study by Pettersson et al. (2004) and ours is that they reported no sesquiterpenes, while we found them to be the most abundant compound. Given that sesquiterpenes are relatively easy compounds to detect and identify, there appears to be a real difference between the scents of baobabs in West Africa, where they collected, and East and Southern Africa, where we collected.

Sesquiterpenes are common chemical components of hawkmoth pollinated floral scents (Knudsen & Tollsten, 1993) but are usually absent in closely related bat-pollinated taxa (Riffell, 2008). Combined with our observation that hawkmoths but not bats are important pollinators of baobabs in South Africa, whereas bat-pollination of baobabs is well-documented in West Africa, it is worth considering the possibility that the proportional contributions of bats and hawkmoths to baobab pollination might vary across the continent. Phylogeographic work in *A. digitata* has suggested significant genetic differentiation of West African populations from those in Southern and Eastern Africa (Leong Pock Tsy et al., 2009; Cron et al., 2016). This raises the possibility that continental-scale differences in pollination ecology might have been in effect long enough to allow for the adaptive evolution of floral volatile synthesis, although the genetic basis of these differences is unknown.

Within Bombacoideae (Malvaceae), several taxa are thought to be specialized on either bats or hawkmoths for their pollination, and multiple evolutionary shifts between these pollinators have been inferred (von Balthazar et al. 2006; Fleming et al., 2009), including at least two transitions within *Adansonia* itself (Baum, 1995a; Baum et al., 1998). Evolutionary transitions in pollination are presumed to pass through intermediates in which both classes of pollinators visit flowers and contribute to fruit set, at least in some parts of the range and/or some times of year. Consequently, it is plausible that the different floral volatiles reported in different areas of Africa and the corresponding differences in the relative importance of bats and hawkmoths might indicate that *A. digitata* is currently in the midst of an evolutionary transition between two, alternative nocturnal pollination systems. For example, it could be that *A. digitata* was previously hawkmoth pollinated, but only the West African populations have yet shed some of the traits that resulted in high hawkmoth visitation. Conversely, it could be that baobabs in South Africa have been subjected to selection for increased attractiveness to hawkmoths, even at the expense of reduced bat attractiveness, perhaps because of a paucity of flower-visiting bats. However, the possibility of mixed modes of pollinator attractiveness within a population should also be considered and is consistent with this phenomenon across Africa. Consider balancing selection of these two pollination syndromes which would support mixed pollination (from hawkmoths and bats) allowing for reproductive success during fluctuating pollinator populations.

The results reported here suggest the need for further studies of geographic variation in baobab pollination ecology and reproductive biology, ideally including both secondary chemistry and studies of gene flow. Not only might these data help us evaluate the recent history and possible future evolution of pollination systems in these charismatic trees, but improved understanding of their reproductive biology might have economic significance, given the emerging role of baobab fruit and seed oil as a cash crop in some parts of Africa.

ACKNOWLEDGEMENTS

Work was partly funded by NSF grant DEB-1354793 to DAB and the NSF GROW with USAID program to NK. We thank Macy Madden for fieldwork assistance. Colin Schoeman and Hermann Staude for assistance in hawkmoth identification. The residents of Zwigodini, Muswodi Dipeni, Mbodi, Messina Nature Reserve, Pieter and Leonie van de Mercue, and Howard Knotts for their assistance and hospitality. Abby Schweiner assisted with pollen tube imaging. Hiroshi Maeda and Marcos Viana de Oliveira for nectar derivatization protocols and guidance. Peter Taylor provided helpful discussions about bats and Diana Mayne suggestions on site selections.

REFERENCES

Ayensu, E.S., 1974. Plant and bat interactions in West Africa. *Annals of the Missouri Botanical Garden*, pp.702-727.

Baker, H. G., and I. Baker. 1975. Studies of nectar-constitution and pollinator-plant coevolution. Pages 100-140 in L. E. Gilbert and P. H. Raven, eds. Coevolution of animals and plants. University of Texas Press, Austin.

Baum DA. 1995a. The comparative pollination and floral biology of baobabs (*Adansonia*–Bombacaceae). *Annals of the Missouri Botanical Garden* 82: 322–348.

Baum DA. 1995b. A systematic revision of Adansonia (Bombacaceae). *Annals of the Missouri Botanical Garden*. 82:440-471.

Baum, D. A., Small, R. L., & Wendel, J. F. 1998. Biogeography and floral evolution of Baobabs Adansonia, Bombacaceae as inferred from multiple data sets. *Systematic Biology*, *47*(2), 181-207.

Bestmann, H.J., Winkler, L. and von Helversen, O., 1997. Headspace analysis of volatile flower scent constituents of bat-pollinated plants. *Phytochemistry*, *46*(7), pp.1169-1172.

Bonaccorso, F.J., Winkelmann, J.R., Todd, C.M. and Miles, A.C., 2014. Foraging movements of epauletted fruit bats (Pteropodidae) in relation to the distribution of sycamore figs (Moraceae) in Kruger National Park, South Africa. *Acta Chiropterologica*, *16*(1), pp.41-52.

Coe, M.J. & Isaac, F.M. 1965. Pollination of the baobab (Adansonia digitata L.) by the lesser bush baby (Galago crassicaudatus E. Geoffroy). E. Afr. Wildl. J. 3, 123-124.

Cron, G. V. & Karimi, N., Glennon K. L., Chukwudi, U., Witkowski, E. T. F., Venter, S. M., Assogbadjo, A. E. and Baum, D. A. 2016. One African baobab species or two? Synonymy of *Adansonia kilima* and *A. digitata*. *Taxon*.

Cruden, R.W., 1976. Intraspecific variation in pollen-ovule ratios and nectar secretion--preliminary evidence of ecotypic adaptation. *Annals of the Missouri Botanical Garden*, pp.277-289.

Djossa, B.A., Toni, H.C., Adekanmbi, I.D., Tognon, F.K. and Sinsin, B.A., 2015. Do flying foxes limit flower abortion in African baobab (Adansonia digitata)? Case study in Benin, West Africa. *Fruits*, *70*(5), pp.281-287.

Erhardt, A., 1991. Pollination of Dianthus superbus L. *Flora*, *185*(2), pp.99-106.

El-Sayed AM 2017. The Pherobase: Database of Pheromones and Semiochemicals. http://www.pherobase.com.

Faegri, K. and Van Der Pijl, L., 2013. *Principles of pollination ecology*. Elsevier.

Gibbs, P.E., and Bianchi, M., 1993. Post-pollination events in species of Chorisia (Bombacaceae) and Tabebuia (Bigoniaceae) with late-acting self-incompatibility. *Botanica Acta*, 106(1), pp.64-71.

Gibbs, P., Bianchi, M.B. and Ranga, N.T., 2004. Effects of self-, chase and mixed self/cross-pollinations on pistil longevity and fruit set in Ceiba species (Bombacaceae) with late-acting self-incompatibility. Annals of Botany, 94(2), pp.305-310.

Gibbs, P.E. (2014) Late-acting self-incompatibility - the pariah breeding system in flowering plants. New Phytologist Vol. 203 (3) pp717-734

Gomez, J.M., Munoz-Pajares, A.J., Abdelaziz, M., Lorite, J. and Perfectti, F., 2013. Evolution of pollination niches and floral divergence in the generalist plant Erysimum mediohispanicum. *Annals of botany*, *113*(2), pp.237-249.

Gottsberger, G., Schrauwen, J. and Linskens, H.F., 1984. Amino acids and sugars in nectar, and their putative evolutionary significance. *Plant Systematics and Evolution*, *145*(1-2), pp.55-77.

Grindeland, J.M., Sletvold, N. and Ims, R.A., 2005. Effects of floral display size and plant density on pollinator visitation rate in a natural population of Digitalis purpurea. *Functional Ecology*, *19*(3), pp.383-390.

Haber, W.A. and Frankie, G.W., 1989. A tropical hawkmoth community: Costa Rican dry forest Sphingidae. *Biotropica*, pp.155-172.

Harris, B. J. & Baker, H. G. 1959. Pollination of flowers by bats in Ghana. Nigerian Field 24:151– 159.

Hegland, S.J., Nielsen, A., Lázaro, A., Bjerknes, A.L. and Totland, Ø., 2009. How does climate warming affect plant-pollinator interactions?. *Ecology letters*, *12*(2), pp.184-195.

Jaeger, P., 1945. Evanouissement et pollinisation de la fleur du baobab. *CR Acad. Sci.,*

Paris, *220*, pp.369-71.

Jaeger, P. 1950. La vie nocturne de la fleur de baobab (*Adansonia digitata L.,* bombacacées). La Nature (Paris) 78: p.28-29

Jaeger, P.,1954. Les aspects actuels du probleme de la Cheiropterogamie.*Bulletin Institut France Afrique Noire*, *16796*, p.821.

Jarne, P. and Charlesworth, D., 1993. The evolution of the selfing rate in functionally hermaphrodite plants and animals. *Annual Review of Ecology and Systematics*, *24*(1), pp.441-466.

Jones, O.R. and Wang, J., 2010. COLONY: a program for parentage and sibship inference from multilocus genotype data. *Molecular ecology resources*, *10*(3), pp.551-555.

Kelber, A., 2003. Sugar preferences and feeding strategies in the hawkmoth Macroglossum stellatarum. *Journal of Comparative Physiology A*, *189*(9), pp.661-666.

Knudsen, J.T., 2002. Variation in floral scent composition within and between populations of Geonoma macrostachys (Arecaceae) in the western Amazon.*American Journal of Botany*, *89*(11), pp.1772-1778.

Knudsen, J.T. and Tollesten, L. 1993. Trends in floral scent chemistry in pollination syndromes: floral scent composition in moth-pollinated taxa. Botanical Journal of the Linnean Society, 111(2), pp.263-284.

Knudsen, J.T. and Tollsten, L., 1995. Floral scent in bat-pollinated plants: a case of convergent evolution. *Botanical Journal of the Linnean Society*, *119*(1), pp.45-57.

Kunin, W.E., 1997. Population size and density effects in pollination: pollinator foraging and plant reproductive success in experimental arrays of Brassica kaber. *Journal of Ecology*, pp.225-234.

Lay, C.R., Linhart, Y.B. and Diggle, P.K., 2011. The good, the bad and the flexible: plant interactions with pollinators and herbivores over space and time are moderated by plant compensatory responses. *Annals of Botany*, *108*(4), pp.749-763.

Mendes-Rodrigues, C., Carmo-Oliveira, R., Talavera, S., Arista, M., Ortiz, P.L. and Oliveira, P.E., 2005. Polyembryony and apomixis in Eriotheca pubescens (Malvaceae-Bombacoideae). Plant Biology, 7(05), pp.533-540.

Nattero, J., Malerba, R., Medel, R. and Cocucci, A., 2011. Factors affecting pollinator movement and plant fitness in a specialized pollination system. *Plant Systematics and Evolution*, *296*(1-2), pp.77-85.

Noutsos, C., Perera, A.M., Nikolau, B.J., Seaver, S.M. and Ware, D.H., 2015. Metabolomic

Profiling of the Nectars of Aquilegia pubescens and A. Canadensis. *PloS one*, *10*(5), p.e0124501.

Pettigrew, F.R.S., Jack, D., Bell, K.L., Bhagwandin, A., Grinan, E., Jillani, N., Meyer, J., Wabuyele, E. and Vickers, C.E., 2012. Morphology, ploidy and molecular phylogenetics reveal a new diploid species from Africa in the baobab genus Adansonia (Malvaceae: Bombacoideae). *Taxon*, *61*(6), pp.1240-1250.

Pettersson, S., Ervik, F. and Knudsen, J.T., 2004. Floral scent of bat-pollinated species: West Africa vs. the New World. *Biological Journal of the Linnean Society*, *82*(2), pp.161-168.

Quesada, M., Fuchs, E.J. and Lobo, J.A., 2001. Pollen load size, reproductive success, and progeny kinship of naturally pollinated flowers of the tropical dry forest tree Pachira quinata (Bombacaceae). American Journal of Botany, 88(11), pp.2113-2118.

Rodríguez-Peña, N., Stoner, K.E., Flores-Ortiz, C.M., Ayala-Berdón, J., Munguía-Rosas, M.A., Sánchez-Cordero, V. and Schondube, J.E., 2016. Factors affecting nectar sugar composition in chiropterophilic plants. *Revista Mexicana de Biodiversidad*, *87*(2), pp.465-473.

Rosas-Guerrero, V., Aguilar, R., Martén-Rodríguez, S., Ashworth, L., Lopezaraiza-Mikel, M., Bastida, J.M. and Quesada, M., 2014. A quantitative review of pollination syndromes: do floral traits predict effective pollinators? *Ecology letters*, *17*(3), pp.388-400.

Rosevear, D.R. 1965. The Bats of West Africa. British Museum (Natural History), London.

Smith, B and Padgett, A. 1996. On Rope: North American vertical rope techniques. National Speleological Society, Huntsville, Alabama. Pp 266-272

Swynnerton, G.H. & Hayman, H.W. 1950. A checklist of the mammals of the Tanganyika Territory and the Zanzibar Protectorate. J. East Africa nat. Hist. SOC. 20, No. 6 and 7 (90), 274-392.

Tripp, E.A. and Manos, P.S., 2008. Is floral specialization an evolutionary dead-end? pollination system transitions in Ruellia (Acanthaceae). *Evolution*, *62*(7), pp.1712-1737.

Leong Pock Tsy, J.M.L., Lumaret, R., Mayne, D., Vall, A.O.M., Abutaba, Y.I., Sagna, M., Raoseta, S.N.O.R. and Danthu, P., 2009. Chloroplast DNA phylogeography suggests a West African centre of origin for the baobab, Adansonia digitata L. (Bombacoideae, Malvaceae). *Molecular Ecology*, *18*(8), pp.1707-1715.

Venter, S.M. 2012, The ecology of Baobabs (*Adansonia digitata*) in relation to sustainable utilization in northern Venda, South Africa. PhD thesis, University of the Witwatersrand.

Venter, S.M. and Witkowski, E.T., 2011. Baobab (Adansonia digitata L.) fruit production in communal and conservation land-use types in Southern Africa. *Forest Ecology and Management*, *261*(3), pp.630-639.

Venter, S.M. 2012, The ecology of Baobabs (Adansonia digitate) in relation to sustainable utilization in northern Venda, South Africa. PhD thesis, University of the Witwatersrand.

Venter, S. M., & Witkowski, E. T. 2013. Fruits of our labour: contribution of commercial baobab (Adansonia digitata L.) fruit harvesting to the livelihoods of marginalized people in northern Venda, South Africa. *Agroforestry systems*, 87(1), 159-172.

Von Balthazar, M., Schönenberger, J., Alverson, W.S., Janka, H., Bayer, C. and Baum, D.A., 2006. Structure and evolution of the androecium in the Malvatheca clade (Malvaceae sl) and implications for Malvaceae and Malvales. *Plant Systematics and Evolution*, *260*(2-4), pp.171-197.

von Helversen, O., Winkler, L. and Bestmann, H.J., 2000. Sulphur-containing "perfumes" attract flower-visiting bats. *Journal of Comparative Physiology A*, *186*(2), pp.143-153.

Wang, J., 2012. Computationally efficient sibship and parentage assignment from multilocus marker data. *Genetics*, *191*(1), pp.183-194.

Wang, M., Lopez-Nieves, S., Goldman, I.L. and Maeda, H.A., 2017. Limited tyrosine utilization explains lower betalain contents in yellow than in red table beet genotypes. *Journal of agricultural and food chemistry*, *65*(21), pp.4305-4313.

Wolff, D., 2006. Nectar sugar composition and volumes of 47 species of Gentianales from a southern Ecuadorian montane forest. *Annals of Botany*, *97*(5), pp.767-777.

Chapter Four:

One African baobab species or two? Synonymy of *Adansonia kilima* and *A. digitata*

This chapter is a reprint of the following publication: Cron, G.V., Karimi, N., Glennon, K.L., Udeh, C.A., Witkowski, E.T., Venter, S.M., Assogbadjo, A.E., Mayne, D. and Baum, D.A., 2016. One African baobab species or two? Synonymy of Adansonia kilima and A. digitata. *Taxon*, *65*(5), pp.1037-1049.

Abstract: We assessed the validity of a recently described baobab species *Adansonia kilima* that was suggested to be a diploid occurring in both eastern and southern Africa at high elevations within the range of the well-known tetraploid species *A. digitata*. We used a combination of phylogenetic analyses and statistical comparisons of various traits (e.g., flowers, stomata, pollen, chromosome counts) to test for the presence of two continental African baobab species. Ordination of the floral features of 133 herbarium specimens from across the natural range of *A. digitata*, including the putative type of *A. kilima* and other Tanzanian accessions as previously assigned *A. kilima*, revealed no distinct clusters of specimens. Likewise, stomatal size and density varied greatly across the specimens examined, with no clear bimodal pattern and no obvious association with altitude. The type specimen of *A. kilima* was found to have a chromosome number of 2n ≈ 166, showing it to be a tetraploid, like *A. digitata*. Phylogenetic analysis of the ITS region showed little resolution within the African baobab clade and a lack of distinction between the *A. kilima* type and *A. digitata* regional accessions. Among the 13 haplotypes detected, no distinct haplotype representing *A. kilima* was identified. Based on the data at hand we conclude that *A. kilima* is neither cytologically nor morphologically distinct and is here reduced to synonymy with *A. digitata*.

INTRODUCTION

The recent description of a second mainland African baobab species, *Adansonia kilima* Pettigrew, Bell, Bhagwandin, Grinan, Jillani, Meyer, Wabuyele & Vickers (Pettigrew *&* al., 2012), increased the number of known species of baobab (*Adansonia* L.) worldwide to nine. This count includes six species endemic to Madagascar, a single species in north-western Australia, and the African species *A. digitata* L. *Adansonia kilima* was reported to be diploid (2n = 88, Pettigrew & al., 2012), like the Malagasy and Australian species, but in contrast to *A. digitata,* which is tetraploid (2n ≈ 160, Baum & Oginuma, 1994). This led to the hypothesis that *A. kilima* might be a diploid progenitor of *A. digitata* (Pettigrew & al., 2012).

In addition to the ploidy difference, *Adansonia kilima* was distinguished from *A. digitata* by several morphological characters (Pettigrew & al., 2012), especially the size of its flowers, including the length of the pistil, the number of free staminal filaments, and the diameter of the "staminal corolla", which we understand to be the diameter of the free staminal ball (Fig. S1). Furthermore, petals of *A. digitata* were said to be longer than the androecium and 'strongly reflexed' in the open flower, whereas those of *A. kilima* are not reflexed and are shorter than the androecium. Pollen diameter and volume were also noted to be smaller in *A. kilima* than in *A. digitata*, with a greater spine density (Pettigrew & al., 2012) and, consistent with diploidy (e.g. Bingham, 1968; Stebbins, 1971; Mishra, 1997; Ramsey & Schemske, 2002), *A. kilima* was said to be characterized by having markedly smaller and more densely spaced stomata than *A. digitata*.

The comparative measurements of *A. kilima* and *A. digitata* reported in Pettigrew & al. (2012) were based on a relatively small sample – six of each species from East Africa. The six

specimens identified as *A. kilima* were trees occurring in the Central Highlands of Tanzania, whereas those assigned to *A. digitata* were from the coastal region near Mombasa, Kenya with one cultivated sample from the Brisbane Botanic Garden, Queensland, Australia. While detailed measurements were not reported from other sites, Pettigrew & al. (2012) claimed that trees matching those identified as *A. kilima* on the eastern slopes of Mt. Kilimanjaro, occurred in southern Tanzania, the Caprivi region of northern Namibia, north-eastern Botswana, and in the Venda region of northern South Africa, with the type specimen of *A. kilima* derived from the latter region. Pettigrew & al. (2012) suggested that the diploid species was limited to elevations between 650 and 1500 m and that the tetraploid species seldom occurred above 800 m. This meant that co-occurrence of the two species was rare, but did include the type locality in Venda, South Africa at an altitude of ~825 m above sea level (a.s.l.).

Long-term studies of populations of baobabs in South Africa (Venter & Witkowski, 2011) and West Africa (Assogbadjo & al., 2008) have shown that some trees are consistently poor producers of fruit, whereas others are prolific. Indeed, in both regions individual baobab trees are commonly characterized by locals as ‘male’ or ‘female’ trees based on fruit production. Despite the large differences in fruit production, Venter (2012) found no significant differences in flower production between ‘male’ and ‘female’ trees in either of two growing seasons, suggesting that the difference arises during or after pollination. Pettigrew & al. (2012) noted that *A. kilima* produced many flowers per tree, whereas *A. digitata* trees produced only a few flowers, suggesting the possibility that the male-female distinction might reflect different ploidy levels, either with one type being *A. kilima* and the other being *A. digitata* or with low-producing ‘male’ trees corresponding to sterile triploid hybrids.

In another study in southern Africa, Douie & al. (2015) found bimodal patterns in stomatal size and density in populations of baobabs in Zimbabwe, but no correlation with

altitude. They suspected that this reflected diploid *A. kilima* and tetraploid *A. digitata* occurring sympatrically in these populations, but did not confirm their assumptions of differences in ploidy using chromosome counts or flow cytometry.

The aim of this study was to evaluate whether *A. kilima* and *A. digitata* are distinct species using a morphological study of mainly herbarium specimens, phylogenetic analysis based on the nuclear ITS region, and ploidy assessment based on chromosome squashes and flow cytometry.

MATERIALS AND METHODS

Sampling.— Herbarium specimens of *A. digitata/A. kilima* with both leaves and flowers were acquired from across their natural range in Arabia, Africa and adjacent oceanic islands (including Zanzibar and Comores, as well as cultivated individuals from Madagascar, Hawaii, and the West Indies) via loans from the following herbaria: B, BM, BNRH, BR, EA, J, K, MO, MRSC, NY, PRE, UNIN, US and WIND (acronyms follow Holmgren & al., 1990). In addition, fresh specimens with leaves and flowers (with pollen) were collected from Venda, Limpopo Province, South Africa and the Caprivi region of Namibia and kept cool in damp tissue paper until examination.

To facilitate comparison with Pettigrew & al. (2012), we made collections of the trees that yielded the holotype and the paratype of *A. kilima*, from the Venda and Caprivi sites, respectively. The geographical coordinates for the holotype provided in Pettigrew & al. (2012) were found to be mistaken. However, based on the location description provided, as corroborated by correspondence with J. Pettigrew in which he provided the correct coordinates (22.90159 S 29.99656 E; J. Pettigrew pers. comm., 8 January 2013), the tree that yielded the

holotype was located. We also collected material from a tree in the Caprivi area of Namibia whose coordinates correspond to those provided by Pettigrew & al. (2012) for the paratype. While it is technically incorrect to consider these trees as nomenclatural types, we will for convenience refer to the Venda and Caprivi trees and new collections made from these trees as the *A. kilima* holotype and paratype, respectively. In addition, we collected nine specimens from the central highlands of Tanzania within the purported range of *A. kilima*. In each locality we collected samples from trees with small flowers, tentatively labeled '*A. kilima,*' as well as nearby trees with much larger flowers, labeled '*A. digitata*.' This sampling included collections at two localities listed in Pettigrew & al. (2012). We also collected from two further small-flowered '*A. kilima*' and five larger-flowered '*A. digitata*' at sites not specifically listed by Pettigrew & al. (2012).

Leaf material for DNA analysis was obtained from diverse sources and represents collections from Benin, Botswana, Burkina Faso, Côte d'Ivoire, Mozambique, Namibia, *São Tomé* Island, South Africa, Tanzania, Togo, and Zimbabwe (Appendix 1). Most were either collected as silica gel or air-dried leaf tissue or seeds from which DNA was extracted. In addition to DNA obtained from the holotype and paratype, we also analyzed DNA from 12 Tanzanian accessions, including four that correspond to *A. kilima* based on geographical coordinates provided in Pettigrew & al. (2012) and an additional three classified as '*A. kilima*' based on morphology (Appendix 1).

Seeds from mature fruits collected from the *A. kilima* holotype were germinated at the University of Wisconsin, Madison and the resulting seedlings were used for chromosome counts and flow cytometry. Leaf tissue and seeds from a Tanzanian *A. kilima* tree (matching Pettigrew & al. 2012; accession *J. Pettigrew 319*) were also collected for flow cytometry analysis.

Morphology: floral features.— One hundred and thirty three specimens (Appendix S1, see Supplemental Data with the online version of this article) were measured for eight floral characters suggested by Pettigrew & al. (2012) to distinguish *A. digitata* and *A. kilima*, viz. maximum calyx length (measured from base of ovary), calyx breadth at widest point, petal length, petal breadth at widest point, staminal tube length, staminal ball diameter (i.e. the "ball" of free stamens, Fig. S1), style length, and stigma branch length. Measurements were made on 1 to 3 flowers per specimen and were then averaged for further analysis.

Pollen diameter was measured for 43 of the herbarium specimens where pollen was of sufficient quality. Fresh pollen from eight recently collected specimens (seven for which floral measurements were not possible as the flowers had shriveled) was also measured, including the type specimens of *A. kilima* and three '*A. kilima*' Tanzanian specimens, of which two matched Pettigrew & al. (2012) coordinates. Pollen from all specimens (bar the three '*A. kilima*' Tanzanian ones) was viewed on microscope slides using a Zeiss Axio Imager M2 Compound Microscope and photographed with a Zeiss Axiocam 506 colour digital camera. Five pollen grains were measured from each specimen, except where it was only possible to measure three grains (one specimen), and the number of spines was counted. In addition, pollen grains from 23 specimens (including the three '*A. kilima*' Tanzanian specimens) were viewed using an FEI Quanta Scanning Electron Microscope to verify the diameter and volume measurements and to confirm pollen spine density counts. Pollen grains were transferred onto double-sided tape on a stub and were coated with gold/palladium for viewing.

Digital images with scale bars facilitated measurement of the diameter of the pollen grains and subsequent calculation of volume. Analyses were facilitated by the use of the image analysis program Axiovision (Zeiss corp.) or ImageJ (Rasband 1997–2015; Abramoff & al., 2004; Schneider & al. 2012; available at http://rsb.info.nih.gov/ij/). The surface area of the

pollen grain was calculated from the radius using the formula for a sphere. Pollen spine density (per 1000 μm^2) was calculated by dividing the number of spines on the visible surface of a pollen grain (one hemisphere) by half the surface area of the grain.

Principal component analysis (PCA) and non-metric multi-dimensional scaling (NMDS) were performed on floral features of the 133 specimens measured. Stigma branch length, which could only be measured on 27 of the specimens, was excluded from analysis. Analyses were performed using NTSYS-pc V.2.2 (Rohlf, 2008), with specimens coded as occurring at either low (< 800 m above sea level) or high (≥ 800 m a.s.l.) altitude. For NMDS, the ordination was repeated until the stress value reached an asymptote near zero, and a reasonable stress value (indicating goodness of fit) was obtained as per Kruskal (1964). Analyses were conducted both with and without pollen diameter, which had only been measured on 46 specimens.

Box and whisker plots were generated using R (R Core Team, 2014) for eight floral features (including pollen diameter) with the specimens grouped according to altitude (< 800 m a.s.l., ≥ 800 m a.s.l.). The relation between different traits and altitude was assessed using t-tests, conducted in Excel (Microsoft Office 7, 2006, Redmond, WA, USA).

Stomatal size and density.— Quick-dry clear nail polish (Revlon, USA) was applied to leaves to create epidermal peels from the abaxial and adaxial leaf surfaces of selected mature leaves from 52 herbarium specimen pockets (one leaf per specimen). Nine individuals for which freshly collected leaf material was available (three from Venda, one from Caprivi, and four from the central highlands of Tanzania) were also examined for comparison with the herbarium material. Epidermal impressions were examined on a ZEISS Axio Imager M2 or an Olympus BX60 with images captured using a ZEISS Axiocam 506 or an Olympus DP70 camera, respectively.

Measurements were made in three non-overlapping fields of view across each epidermal peel to ensure an accurate representation of each leaf (with the exception of three specimens for

which only one or two images were captured). Stomata were only measured and counted on the abaxial surface since stomata were consistently lacking on the adaxial surface. The claim in Pettigrew & al. (2012) that counts were made on the adaxial surface was likely a typographical error. Stomatal density was calculated for each field of view independently by dividing the number of stomata by the area of the field of view. The three field of view counts were averaged to estimate the mean stomatal density for a given accession. A total of 15 stomata per specimen (five stomata for each field of view) were measured, and mean stomatal length and width calculated per specimen. The area of a stoma was calculated using the area of the ellipse, which best represents the two guard cells together (area = (π(L/2)(W/2)), where L = length and W = width). It should also be noted that Pettigrew & al. (2012) reported stomatal density units as 'per 100 μm^2', but our measurements showed that no more than one stoma could be viewed within this area, therefore we calculated density based on 1000 μm^2.

In order to test whether our specimens followed an altitudinal pattern as suggested by Pettigrew & al. (2012), we conducted individual linear regressions of the four stomatal traits (length, width, area, and density) against altitude.

Chromosome counts.— Seeds of *A. kilima* holotype were soaked in water with a Kirstenbosch Instant Smoke Plus Seed Primer paper disc (Kirstenbosch National Botanical Gardens, South Africa) for 24 hours and germinated on a heat pad in 1:1 sterile potting mix:turf sand. The *A. digitata* and *A. rubrostipa* Jum. & H.Perrier specimens used as comparisons are established trees growing in the University of Wisconsin – Madison Department of Botany Greenhouse.

Root tips of germinated seedlings were treated in 2 mM 8-hydroxyquinoline at room temperature for two to four hours in the dark, rinsed in iced deionized water for 10 min and then transferred into freshly made 3:1 EtOH:Acetic Acid for a minimum of 24 hours. The root tips

were macerated on a glass slide, together with c. 100 µl of 45% acetic acid and smeared across the slide. They were then placed on a hot plate (50 ˚C) for 30 s, covered with 100 µl of 3:1 EtOH: Acetic Acid for 30 s, washed twice with this mix, and then air dried. Slides were further treated in 200 µL of 4% paraformaldehyde at room temperature for 15 min, with a cover slip, rinsed in 1x phosphate-buffered saline (10 mM, pH~7.4) for 5 s, washed in 2× saline sodium citrate (0.3 M NaCl, 0.03 M sodium citrate; pH 7.0) for five min, covered in 100 µL of 70% formamide at 75 °C for two minutes, cooled for five minutes in 70% ethanol chilled to −20 °C, and finally rinsed in 100% ethanol for 2 min. Slides were treated with 4', 6-diamidino-2-phenylindole 1:1000 in phosphate-buffered saline (pH ~7.2) for fluorescent staining. Prepared slides were viewed on an Olympus BX60 Epifluorescence Microscope and images were captured with a color Olympus DP70 digital camera. A minimum of six photographed cells were used for chromosome counting.

Flow cytometry.— Flow cytometric estimates of genome size were obtained for seed material or freshly collected leaves from greenhouse grown plants. Analyses were conducted by the Flow Cytometry Lab at the Benaroya Research Institute at Virginia Mason Hospital in Seattle, Washington. We used one biological replicate of leaves from *A. digitata*, *A. rubrostipa*, and *A. suarezensis* H.Perrier, two biological replicates of leaves for holotype *A. kilima*, and four biological replicates of holotype *A. kilima* seeds. One seed collected from an additional *A.kilima* matching coordinates for Pettigrew & al. (2012) accession *J. Pettigrew 319* was also analyzed [voucher *N. Karimi 2015-239* (WIS)].

DNA Sequencing and Phylogenetic analyses.— DNA sequences of the ITS region were obtained for a total of 78 specimens from across continental Africa for all taxa. This includes 70 samples assumed to be *A. digitata,* of which 15 trees are assigned as either producer or poor producer trees from Venda, Limpopo, South Africa, and eight putative *A. kilima* samples,

including the holotype and paratype (Appendix 1). Additional sequences (Appendix 1) downloaded from GenBank were included in the final aligned matrix, including the six *A. kilima* accessions from Pettigrew & al. (2012). *Cavanillesia platanifolia* (Bonpl.) Kunth. was used to root the tree due to its position as outgroup and close relative of *Adansonia* (Duarte & al., 2011).

The Qiagen Plant DNeasy Minikit (Venlo, Netherlands) was used for many DNA extractions, following the manufacturer's instructions, but doubling the volume of buffers AP1 and P3. In other cases DNA was extracted from leaves using a CTAB protocol, similar to Porebski & al. (1997), but modified as follows: (1) no proteinase K, (2) addition of an RNase A treatment for 10 minutes, (3) no addition of 5M NaCl to the recovered aqueous layer after chloroform extraction, and (4) precipitation using a half volume of 7.5M ammonium acetate in place of 1/10 vol. 2M Na acetate. Additionally, when extracting DNA from seed we used phenol:chloroform:isoamyl alcohol (25:24:1, v/v/v) instead of chloroform:octanol.

The nuclear internal transcribed spacer (ITS) region was amplified in Polymerase Chain Reactions (PCR) using the ITSLeu and ITS4 primers following Baum & al. (1998) or using Phusion Flash Master Mix (Thermo Scientific, Edenvale, South Africa) with an initial denaturation at 98 °C for 1 min, 33–37 cycles of 98 °C for 1 s, annealing at 55 °C for 5 s, extension at 72 °C for 1 s, with a final extension at 72 °C for 1 min. Purified PCR products were sequenced at the University of Wisconsin-Madison or at the Central Analytical Facility, Stellenbosch University, South Africa using the Applied Biosystems (Foster City, California, US) Big-Dye sequencing and the same primers as for PCR.

Forward and Reverse sequences were edited and aligned using Sequencher 5.2 (GeneCodes Corp., Ann Arbor, MI, USA) or Geneious v7.0 (Biomatters, Auckland, New Zealand). Double peaks were identified by visual inspection and were coded using the

appropriate IUPAC ambiguity codes. Consensus sequences were aligned using the MUSCLE Alignment plugin in Geneious with a gap open cost of 15.

Maximum likelihood analyses were performed using RAxML vers. 8.1.20 (Stamatakis, 2014) under the GTRGAMMA model with 100 rapid bootstrapping replicates. Bayesian analyses were performed using MrBayes vers. 3.2.2 (Huelsenbeck & Ronquist, 2001) with the GTR plus gamma model, nruns=4, nchains=4, and a heat of 0.2. Other settings were retained at default values. The full *Adansonia* dataset, including the outgroup, was run for 12,000,000 generations with a sample frequency of 2000. The first 25% of samples from the cold chain were discarded as burn-in.

A Maximum Likelihood midpoint rooted tree was inferred with RAxML vers. 8.1.20 (Stamatakis, 2014) under the GTRGAMMA model with 100 rapid bootstrapping replicates. Accessions of this phylogeny, with their associated flower size, was mapped based on geographic coordinates to visually display patterns of genetic relatedness and morphological variation. All Tanzanian *A. kilima* accessions from Pettigrew & al. (2012) were designated as small flowers with associated elevation above 800 m.

A haplotype network of *A. digitata* and *A. kilima* ITS sequences was estimated in TCSv1.21 (Clement & al., 2000) with a 95% maximum connection limit and gaps treated as missing data. Two *A. digitata* accessions obtained from GenBank (AF028525, JN400317) were used in the phylogenetic analysis, but were excluded from the haplotype network analysis because they are cultivated specimens with unknown origins. Populations were designated based on watershed and ecoregion resulting in six designated geographic populations: West Africa, Okavango (including east and central Namibia and Botswana), Limpopo Valley (which includes South Africa, Botswana and Zimbabwe), East Coast (Mozambique and low elevation coastal Tanzania) and Tanzanian Highlands.

RESULTS

Floral morphology.— For all the floral features measured, we found a broad range reflecting great variability in baobab flowers (Table 1). In most cases our observed variation includes values reported by Pettigrew & al. (2012) for *A. kilima* and *A. digitata*. Exceptions are the staminal ball diameter [which Pettigrew & al. (2012) call the 'staminal corolla'], where our largest measurement is smaller than they reported for *A. digitata*, and the pollen spine density, where our highest densities are still lower than they reported for *A. kilima* (Table 1).

Table 1. Comparison of morphological features used to distinguish *Adansonia kilima* and *A. digitata* by Pettigrew et al. (2012) and our results for all specimens across the range ('A. digitata') excluding the specimens matching Tanzanian trees designated as 'A. kilima'. [*Note: Stomatal density units given as 'per 100 μm^2' in Pettigrew et al. (2012) would not accommodate numbers of stomata reported; therefore we reported as per 1000 μm^2.] N=number of accessions analyzed, see methods for number of replicates.

	A. kilima (2x)	*A. digitata* (4x)	*A. digitata*	*'A. kilima'*
Characteristic	Range ± S.D. reported in Pettigrew et al. (2012), N = 6		Mean, ±SD, and range ± S.D. reported from our data	
Staminal ball diameter (mm)	38–42	95–106	47 ± 9.99, 28.5–75 N=123	41.75 ± 7.37, 33–51 N = 4
Number of free staminal filaments	270–640	700–1600	895–1259 N=3	833-916 N=2
Pollen spine density (per 1000 μm^3)	142–158	51–100	82.63 ± 25.63, 51.6–132.3 N=50	16.01 ± 3.78, 9.48–24.20 N=3
Pollen grain diameter (µm)	43.7 ± 3.4	63.4 ± 7.7	57.62 ± 8.8, 33.9–79	51.06 ± 5.59, 40.27–59.40

			N=49	N=3
Pollen volume ($\mu m^3 \times 10^3$)	42.8–44.6	118.8–113.0	106.99 ± 47.85, 20.4–258.2 N=49	72.14 ± 22.79, 33.65–110 N=3
Mean stomatal length (μm)	26.1 ± 5.7	38.3 ± 4.5	32.79 ± 4.45, 24.33–48.09 N=48	32.68 ± 9.26, 17.05–57.49 N=4
Stomatal density (per 1000 μm^{2*})	5.0 ± 2.1 per 100 μm^2	1.6 ±0.7 per 100 μm^2	196.13 ± 59.73, 123–432 N=48	159.86 ± 37.72, 114–224.79 N=9
Chromosome number (2*n*)	88	160		160–166

To facilitate comparisons with data in Pettigrew & al. (2012) we compared floral measurements made in the field with those obtained from dried herbarium vouchers collected from the same individual trees. These comparisons indicated 28–33% shrinkage in petal length, 16–30% in petal width, and 10–21% in staminal ball diameter. Given the magnitude of these differences, we conducted multivariate analyses only on herbarium specimens, for which we had the most data.

Including or excluding pollen diameter did not substantially affect the distribution of specimens, but excluding it increased the percentage variation represented by the first two components. Therefore results excluding pollen diameter are reported and presented. Of the variation in the data set (excluding pollen grain diameter), 59.3% was accounted for by the first principal component and 12.5% by the second (i.e. 71.5% by the first two principal components) in the PCA and 58.74% and 17.62% (total 76.4.%) in the NMDS. Eigenvectors that strongly influenced the distribution along the first principal component (in all cases positively) are style

length, calyx length and staminal ball diameter; whereas petal breadth and length (positively) and stamen tube length (negatively) influenced position along the second component (Table S1).

2D scatterplots of floral morphological variation do not indicate any separation between specimens from high and low altitude based on either PCA (not shown) or NMDS (Fig. 1; Fig. S2). Some specimens with very large flowers are located towards the right of the first axis, and these are all from altitudes lower than 800 m (e.g. *Texeira & al. 10115* (BR) from Angola; *Lebrun 2603* (BR, EA) from the DRC; *Roger s.n. sub K000452263* (K), Senegal. In Figure 1, the holotype (122) of *A. kilima* and Tanzanian trees identified by Pettigrew & al. (2012) as *A. kilima* are placed centrally while the paratype is to the mid-left. A similar lack of clustering was observed with 3D plots (not shown).

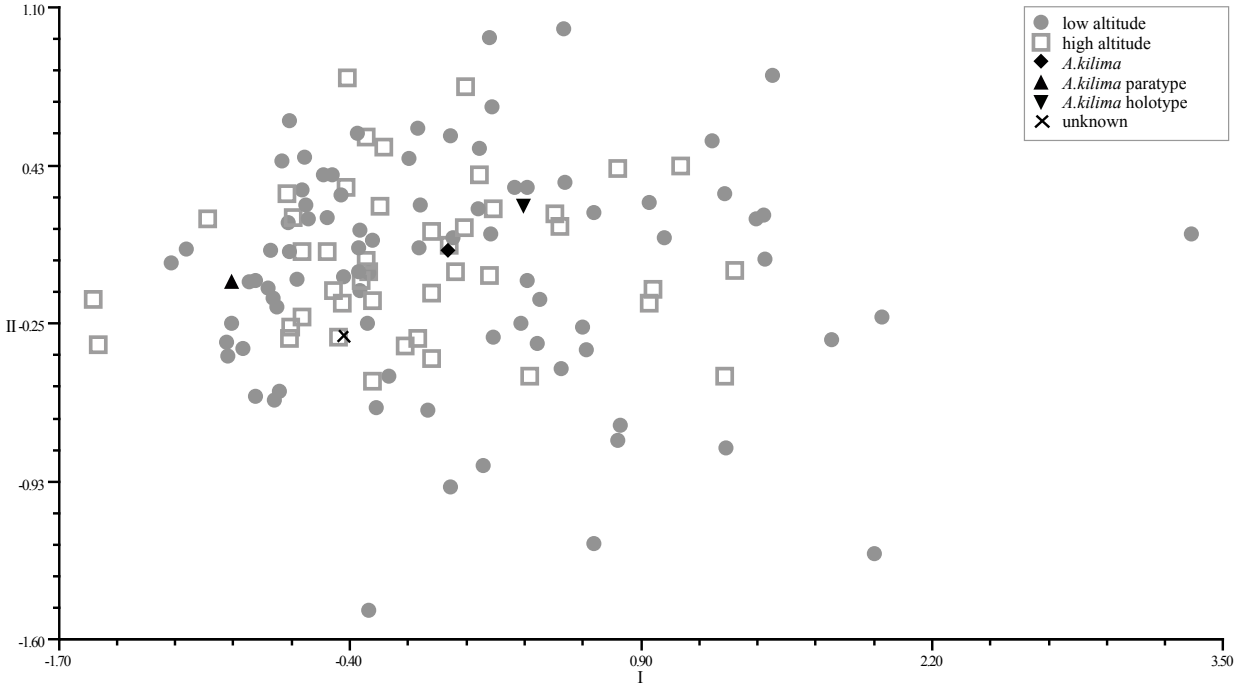


Figure 1. Two-dimensional plot of NMDS of 133 baobab specimens based on seven floral features (i.e., excluding pollen grain diameter). The two axes shown represent 76.4% of the

variation in the data. Circle < 800 m a.s.l. (low altitude), square = ≥ 800 m a.s.l. (high altitude), star = *Adansonia kilima* holotype, triangle = *A. kilima* paratype, diamond = Tanzanian specimens matching *A. kilima* as designated by Pettigrew et al. (2012), cross = altitude unknown.

Univariate analyses of floral and stomatal traits.— There is considerable overlap between high and low altitude accessions for each of the nine floral characters measured, with similar means for both altitude groups (Fig. 2). There were no significant differences in any of the floral features between the two altitudinal ranges.

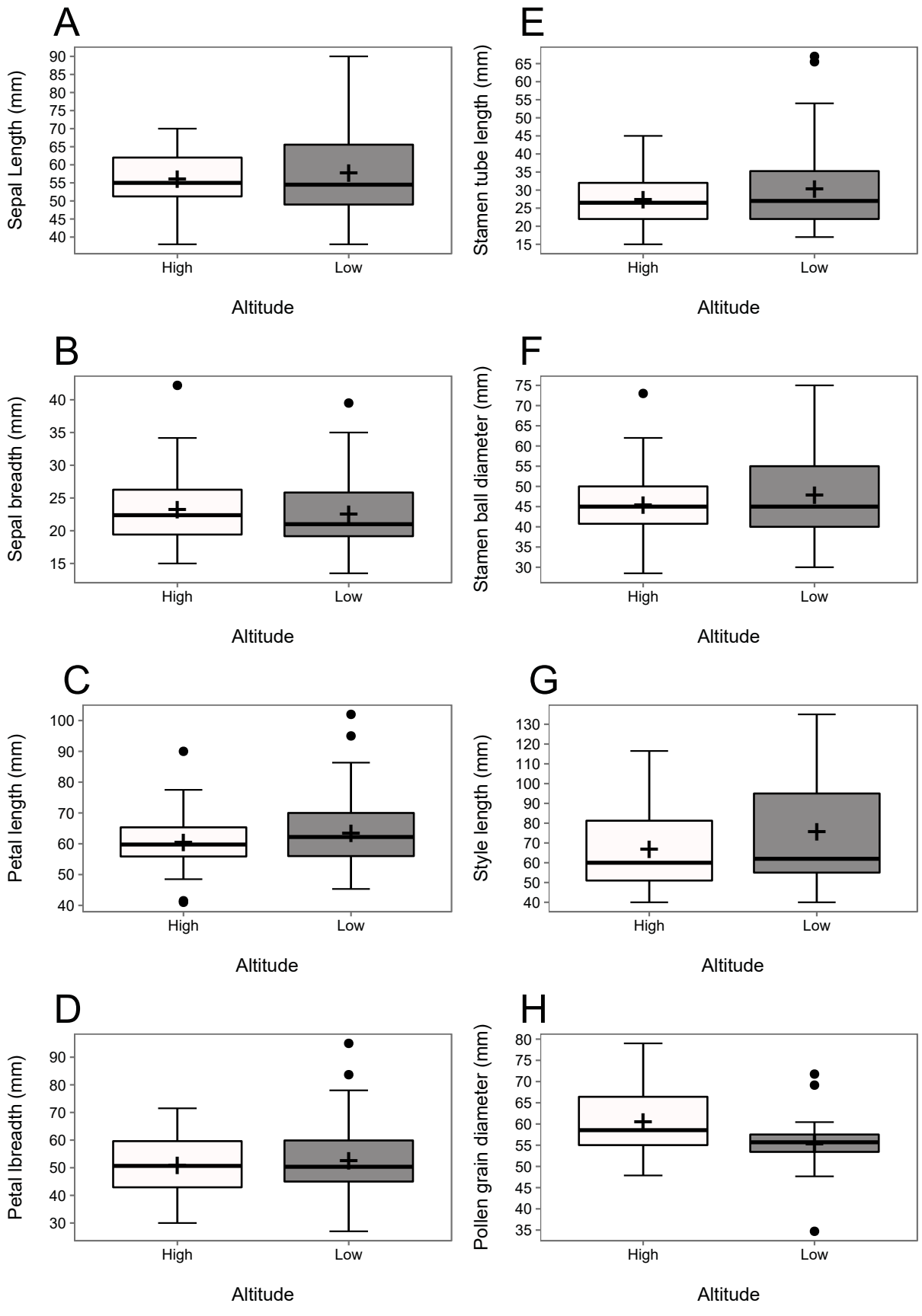

A
Sepal Length (mm)
90
85
80
75
70
65
60
55
50
45
40
High
Low
Altitude
E
Stamen tube length (mm)
65
60
55
50
45
40
35
30
25
20
15
High
Low
Altitude
B
Sepal breadth (mm)
40
35
30
25
20
15
High
Low
Altitude
F
Stamen ball diameter (mm)
75
70
65
60
55
50
45
40
35
30
High
Low
Altitude
C
Petal length (mm)
100
90
80
70
60
50
40
High
Low
Altitude
G
Style length (mm)
130
120
110
100
90
80
70
60
50
40
High
Low
Altitude
D
Petal lbreadth (mm)
90
80
70
60
50
40
30
High
Low
Altitude
H
Pollen grain diameter (mm)
80
75
70
65
60
55
50
45
40
35
High
Low
Altitude

Figure 2. Box and whisker plots of eight floral features of African baobab specimens [calyx length (A) and breadth (B), petal length (C) and breadth (D), stamen tube length (E), stamen mass diameter (F), style length (G), pollen grain diameter (H)] measured on baobab specimens from localities at low (< 800 m a.s.l.) and high (≥ 800 m a.s.l.) altitude. Line = median; whiskers are < and > 1.5 x IQR; Mean = + sign; outliers are plotted. No significant differences were detected for any of the features compared from low and high altitudes.

Comparison of freshly collected leaves and herbarium specimens found no significant difference between herbarium material and fresh material for any of the stomatal traits ($t = 2.009$, $df = 47$, $p = 0.536$; $t = 2.011$, $df = 47$, $p = 0.613$; $t = 2.012$, $df = 47$, $p = 0.352$, respectively). Linear regressions showed that none of the four stomatal traits were correlated with altitude (Fig. 3), although some tendency for decreasing stomatal size with altitude is indicated. Comparison of our stomatal density results with Pettigrew & al. (2012) is confounded by the fact that the stomatal density values in that paper are impossible, presumably because of an error in units:

each stoma is longer than 10 µm, so it is not possible to have 5 per 100 µm^2.

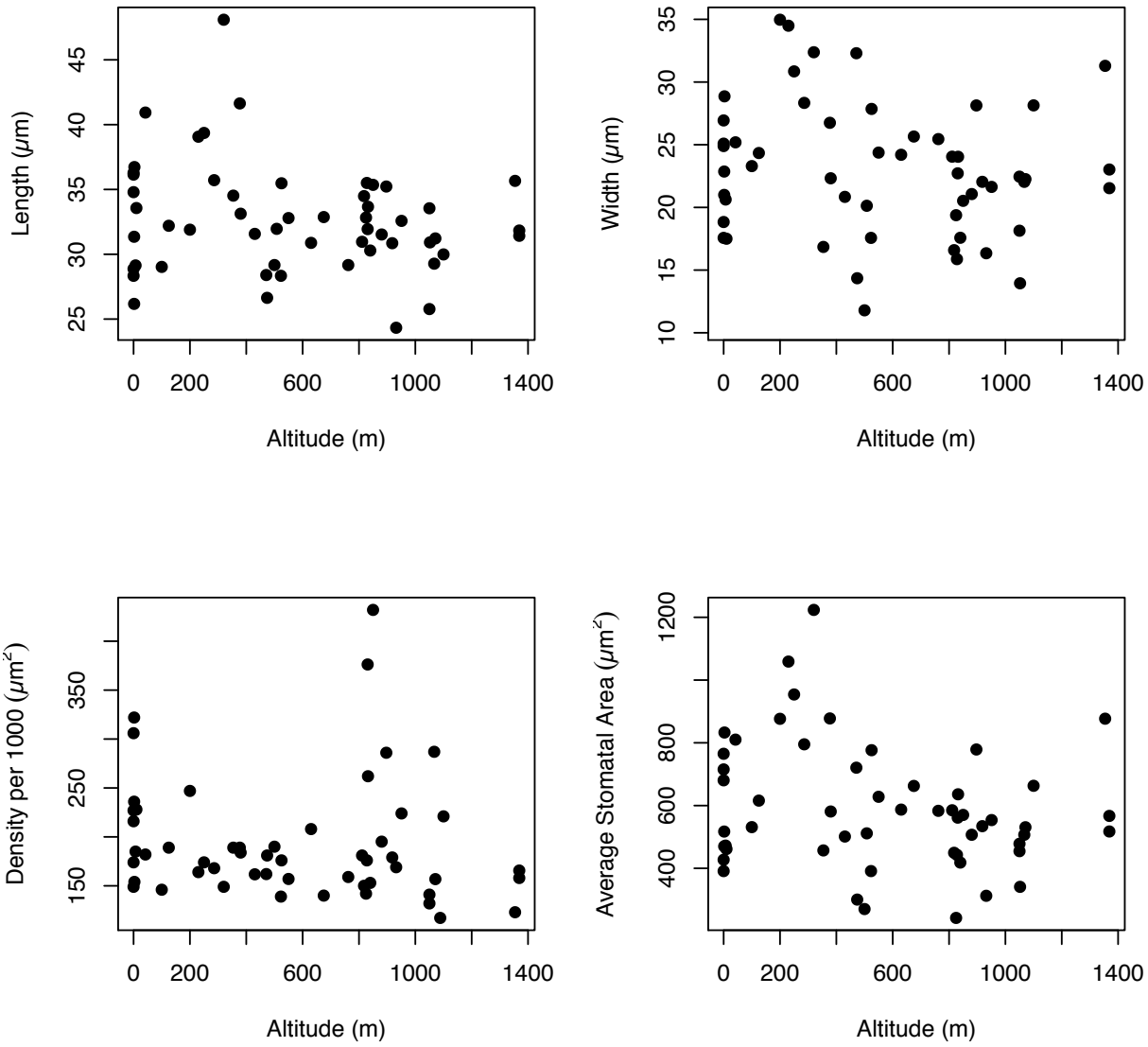


Figure 3. Linear regressions of four African *Adansonia* stomatal traits measured from 52 specimens against altitude. (A) length, (B) density, (C) area, (D) width.

Phylogenetic analysis.— Both Maximum Likelihood and Bayesian analyses went to completion. Bayesian analysis showed adequate mixing as indicated by standard deviation of split frequencies of 0.014. Both analyses both recovered three main *Adansonia* clades, organized by geography; continental Africa (*A. digitata*, *A. kilima*), Madagascar, and Australia (Fig. 4). The African clade was well supported (96% Maximum Likelihood bootstrap support, MLBS and 100% posterior probability, PP). Within the African clade, resolution was limited, although there appeared to be some geographic signal: with West African accessions (plus a few from southern Africa) forming a well-supported clade (MLBS=88, PP=1.0). There was no distinction between the producers and poor producers from Limpopo Province, South Africa. The *A. kilima* holotype and paratype were members of a large polytomy composed of *A. digitata* specimens from eastern and southern Africa, including three additional *A. kilima* sequences from Tanzania. Interestingly, there is some weak support (MLBS=48, PP=.61) for a clade comprised entirely by Tanzanian accessions, all but one designated as *A. kilima*. The one *A. digitata* sample in this clade was found in close geographic proximity to an *A. kilima* collection but had large flowers and resembled *A. digitata* by the criteria of Pettigrew & al. (2012).

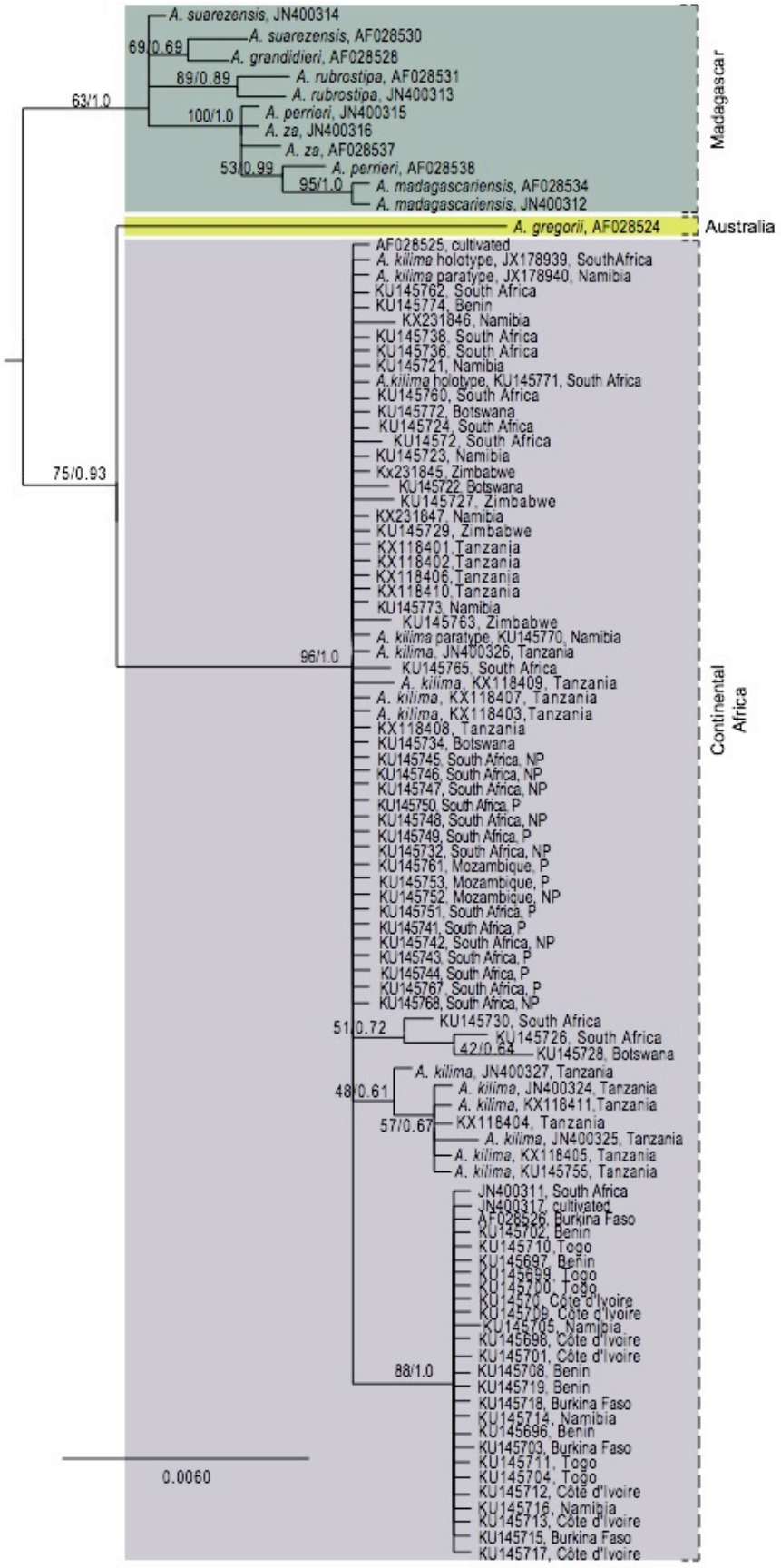

A. suarezensis, JN400314
A. suarezensis, AF028530
A. grandidieri, AF028528
A. rubrostipa, AF028531
A. rubrostipa, JN400313
A. perrieri, JN400315
A. za, JN400316
A. za, AF028537
A. perrieri, AF028536
A. madagascariensis, AF028534
A. madagascariensis, JN400312
Madagascar
A. gregorii, AF028524
Australia
AF028525, cultivated
A. kilima holotype, JX178939, SouthAfrica
A. kilima paratype, JX178940, Namibia
KU145762, South Africa
KU145774, Benin
KX231846, Namibia
KU145738, South Africa
KU145736, South Africa
KU145721, Namibia
A.kilima holotype, KU145771, South Africa
KU145760, South Africa
KU145772, Botswana
KU145724, South Africa
KU14572, South Africa
KU145723, Namibia
Kx231845, Zimbabwe
KU145722, Botswana
KU145727, Zimbabwe
KX231847, Namibia
KU145729, Zimbabwe
KX118401,Tanzania
KX118402,Tanzania
KX118406,Tanzania
KX118410,Tanzania
KU145773, Namibia
KU145763, Zimbabwe
A. kilima paratype, KU145770, Namibia
A. kilima, JN400326, Tanzania
KU145765, South Africa
A. kilima, KX118409, Tanzania
A. kilima, KX118407, Tanzania
A. kilima, KX118403,Tanzania
KX118408, Tanzania
KU145734, Botswana
KU145745, South Africa, NP
KU145746, South Africa, NP
KU145747, South Africa, NP
KU145750, South Africa, P
KU145748, South Africa, NP
KU145749, South Africa, P
KU145732, South Africa, NP
KU145761, Mozambique, P
KU145753, Mozambique, P
KU145752, Mozambique, NP
KU145751, South Africa, P
KU145741, South Africa, P
KU145742, South Africa, NP
KU145743, South Africa, P
KU145744, South Africa, P
KU145767, South Africa, P
KU145768, South Africa, NP
KU145730, South Africa
KU145726, South Africa
KU145728, Botswana
A. kilima, JN400327, Tanzania
A. kilima, JN400324, Tanzania
A. kilima, KX118411,Tanzania
KX118404, Tanzania
A. kilima, JN400325, Tanzania
A. kilima, KX118405, Tanzania
A. kilima, KU145755, Tanzania
JN400311, South Africa
JN400317, cultivated
AF028526, Burkina Faso
KU145702, Benin
KU145710,Togo
KU145697, Benin
KU145699, Togo
KU145700, Togo
KU145707, Côte d'Ivoire
KU145709, Côte d'Ivoire
KU145705, Namibia
KU145698, Côte d'Ivoire
KU145701, Côte d'Ivoire
KU145708, Benin
KU145719, Benin
KU145718, Burkina Faso
KU145714, Namibia
KU145696, Benin
KU145703, Burkina Faso
KU145711, Togo
KU145704, Togo
KU145712, Côte d'Ivoire
KU145716, Namibia
KU145713, Côte d'Ivoire
KU145715, Burkina Faso
KU145717, Côte d'Ivoire
Continental Africa
63/1.0
69/0.69
89/0.89
100/1.0
53/0.99
95/1.0
75/0.93
96/1.0
51/0.72
42/0.64
48/0.61
57/0.67
88/1.0
0.0060

Figure 4. Maximum likelihood (ML) tree comprising 86 *Adansonia* specimens from across Africa and Madagascar with Bayesian posterior probabilities added. Clades with < 60% ML bootstrap support were collapsed. The types of *A. kilima* and five specimens from the highlands of Tanzania tentatively referred to as *A. kilima* were included. *Cavanillesia platanifolia* was used to root the tree (not shown here to facilitate normalisation of branch lengths). Taxon names including '*' are producers and '**' are non-producers.

The midpoint rooted Maximum Likelihood phylogeny displayed with corresponding flower size shows the geographic distribution of variation in Tanzania (Fig. 5). One *A. kilima* accession (KX118409) matching published coordinates in Pettigrew & al. (2012) was found to occur at 430 m; therefore, we assigned this specimen to occur less than 800 m. The two largest phylogenetic clades show patterns of genetic relatedness based on geographic distribution (Eastern Highlands and Central Highlands), but show intermixing of both large and small flower types. Interestingly, two individuals, both with small flowers and occurring in one clade, are not found within the same geographic proximity.

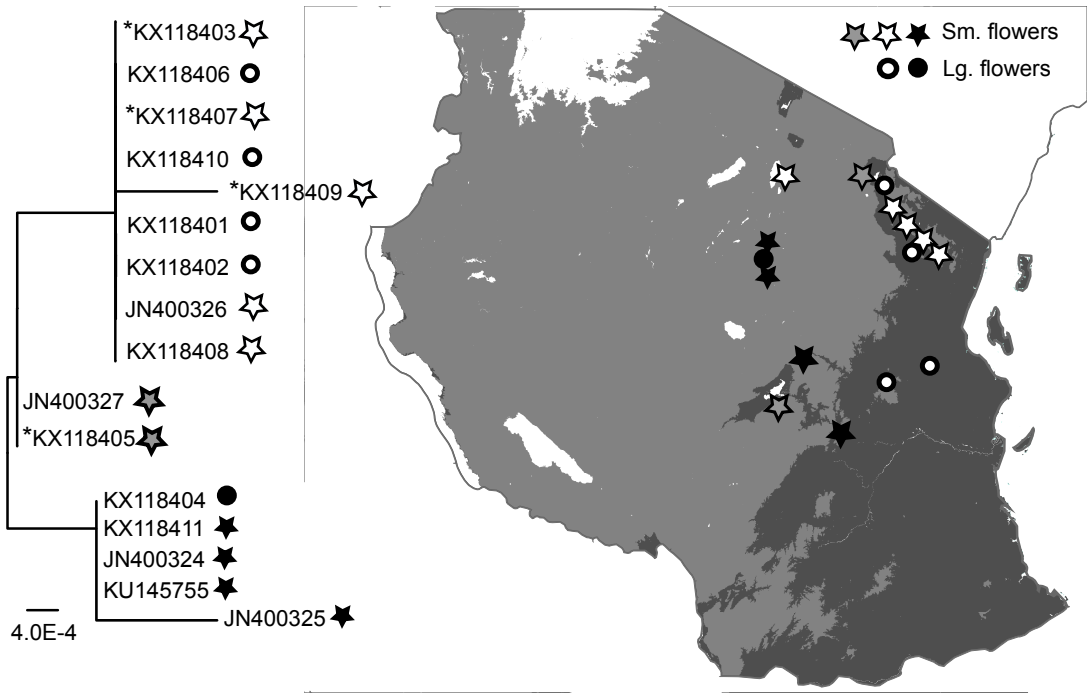


Figure 5. Midpoint rooted Maximum Likelihood tree of the ITS region and the geographic distribution of individuals sampled across Tanzania with associated flower size. Stars indicate duplicate samples matching coordinates to *A.kilima* trees in Pettigrew & al. (2012). Dark grey represents elevation above 800 m and light grey less than 800 m.

Haplotype analysis identified 13 haplotypes, seven represented by only one individual (Fig. 6). Haplotypes were roughly grouped by geography with populations from the Okavango ecoregion being the most diverse. The most common haplotype (H1) is represented by 39 individuals including six individuals from Tanzania of which five are '*A. kilima*', and the *A. kilima* holotype and paratype from southern Africa. Accessions from West Africa were mainly composed of haplotypes 2 and 3, which also occurred in one individual from Limpopo and three individuals from Okavango. One haplotype (H5) was represented by five individuals from Tanzania, four assigned to *A. kilima* and the other to *A. digitata*. Haplotypes H4–H8 were all

restricted to Tanzanian accessions and collectively included all East African specimens assigned to *A. kilima* based on morphology plus two assigned to *A. digitata*. One haplotype (H13) was composed only of three individuals from Mozambique.

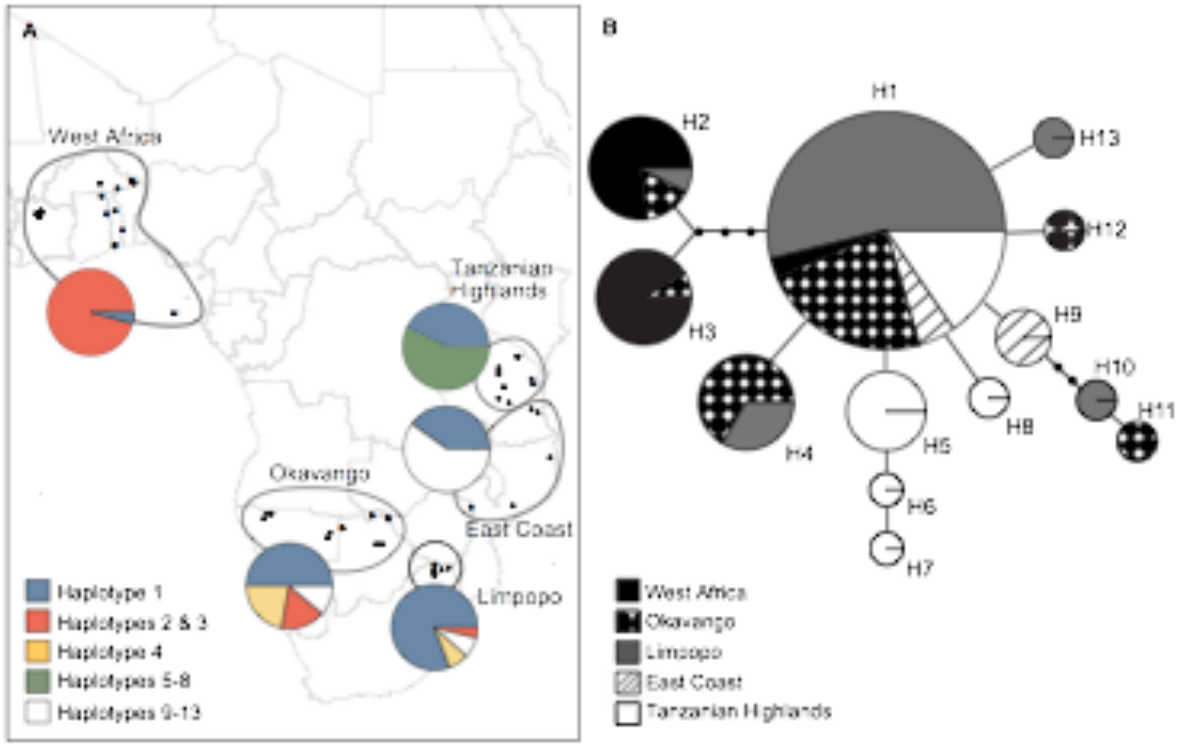


Figure 6. Thirteen haplotypes resulted from analysis of the ITS region of *Adansonia digitata* and *A. kilima* from across Africa. (A) Haplotype frequency for each geographic area; (B) Individuals by geography comprising each of the 13 haplotypes.

Chromosome counts.— Based on counts in seven metaphase cells, the chromosome count for the type specimen of *A. kilima* was estimated to be 2n = 160–166. Figure 7 shows a cell with 2n = 166. These counts are much higher than the 2n = 88 reported by Pettigrew & al. (2012). Chromosome squashes for *A. digitata* and *A. rubrostipa* were also examined as a point of reference. These revealed that the *A. kilima* type has a very similar or identical chromosome

number to *A. digitata*, which is about twice the diploid *A. rubrostipa*. For comparison, Figure 7B shows a diploid *A. rubrostipa* cell with about 80–90 chromosomes.

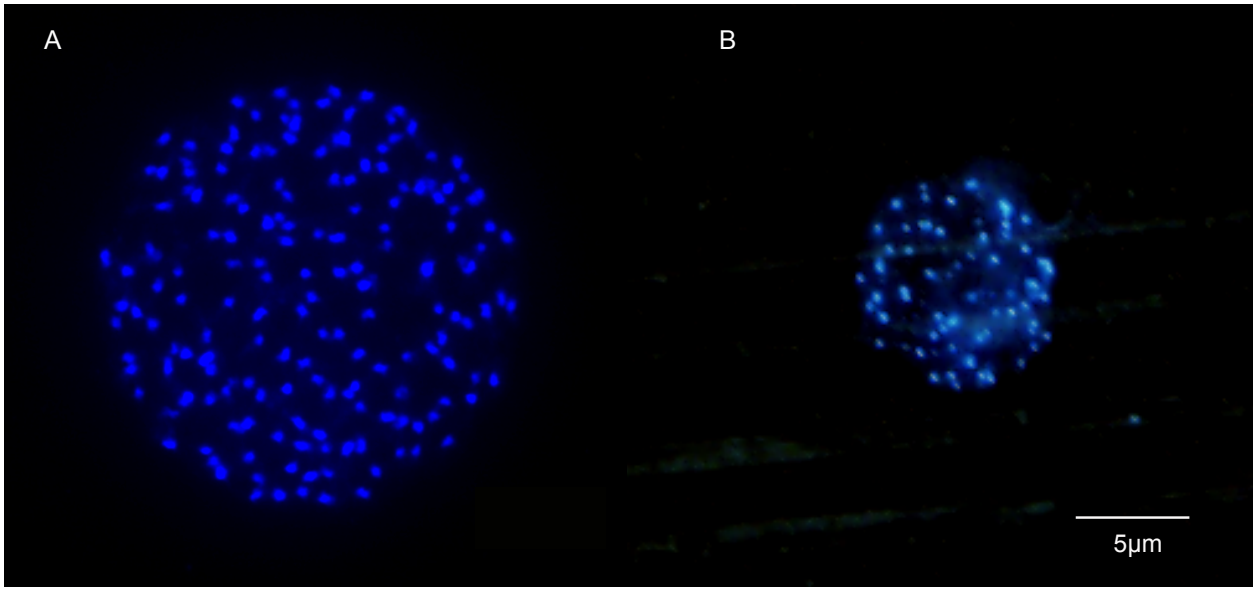


Figure 7. Chromosome squashes of root tips from (A) type specimen of *Adansonia kilima*, 2n = c. 166, and (B) *A. rubrostipa*, 2n = 80–90. (Photographs N. Karimi)

Flow cytometry.— Flow cytometry results are summarized in Table 2. DNA content for greenhouse grown *A. digitata* was 3.04 pg/2C, which is similar to greenhouse grown seedlings from the type specimen of *A. kilima* (2.81 and 3.12 pg/2C). Seeds from the holotype accession were also analyzed, yielding very similar C-values (2.86-3.24). The additional Tanzanian *A. kilima* analyzed from seed was found to have a C-value of 3.01. In contrast, the diploid species *A. suarezensis* and *A. rubrostipa* had DNA content measured at 1.60 pg/2C and 1.55 pg/2C, respectively. These data suggest that all African baobab samples have a genome size about twice that of the diploid Malagasy species.

Table 2. DNA content determined by flow cytometry for tetraploid *Adansonia digitata*, the type specimen of *A. kilima*, an additional *A.kilima* from Tanzania, and diploid species *A. suarezensis* and *A. rubrostipa.* DNA content reported as an average of four runs per sample.

Species, Sample name Tissue, # of samples	DNA content (pg/2C)	St. Dev±
A. digitata UW 291 Leaf, N=1	3.04	0.034
A. kilima type specimen Leaf, N=2	2.83	0.106
A. kilima type specimen Seed, N=4	3.05	0.145
A. kilima, KX118405 Seed, N=1	3.01	0.040
A. suarezensis UW10 Leaf, N =1	1.6	0.055
A. rubrostipa UW 1893 Leaf, N=1	1.55	0.036

DISCUSSION

This analysis of floral features from baobab specimens across Africa and its oceanic islands failed to reveal any bimodal trend, despite being based upon the same floral and pollen features that Pettigrew & al. (2012) used to distinguish *A. kilima* and *A. digitata* and including Tanzanian trees designated as *A. kilima* and *A. digitata* by Pettigrew & al. (2012). Furthermore, the variation that we observed did not appear to show an association with altitude. Consistent with this finding, field observations in southern Africa failed to identify any distinctions in gross morphology that would support the recognition of two species in this area. Although differences

in measurements could be due to using herbarium specimens versus fresh measurements, there is no reason to think that shrinkage would be different between *A. kilima* and *A. digitata*. Thus, while there might exist two morphotypes in the Tanzanian highlands, these both fall well within the range of morphological variation seen across the continent as a whole.

We scored stomatal size (guard cell length, width, and area) and density for 57 herbarium specimens from across Africa, but failed to find a bimodal distribution or an association with altitude. This partially contradicts Douie & al. (2015) who studied specimens from Zimbabwe and detected a bimodal pattern when plotting stomatal density against either length or width (e.g., our Fig. S3), but their pattern was intrapopulational and did not correlate with altitude. They suggested that the bimodal pattern might reflect ploidy variation, but did not directly measure ploidy. The difference between our results and theirs could be a consequence of us looking at a wider geographic area rather than individual populations, or a difficulty in accurately counting and measuring stomata on herbarium specimens. It is also noteworthy that abiotic factors can influence stomatal size and density along elevational, temperature, and water gradients (Körner & Cochrane, 1985; Kluge & Kessler, 2007; Wang & al., 2014). Such variation is likely in baobabs given evidence of higher stomatal density and smaller guard cells in high temperature and low rainfall areas in Benin, though this relationship was not also found in Malawi (Sanchez, 2010; Sanchez & al., 2010). Furthermore, stomatal density has been found correlated with the phenology of leaves, with younger leaves having higher densities (Chen et al., 2001); therefore, leaf maturity may also explain the discrepancies. As a result, the stomatal variation reported by Douie & al. (2015) could be a result of phenotypic plasticity or developmental changes rather than genetic differentiation.

Investigations of ploidy level using chromosome counts and flow cytometric measurement of genome size confirmed that the holotype of *A. kilima* is tetraploid, like *A.*

digitata. Likewise, flow cytometry shows that Tanzanian accessions from coordinates assigned to *A. kilima* by Pettigrew & al. (2012), and with a small-flowered morphology, are also tetraploid. This discrepancy implies that Pettigrew & al. (2012), who used a field microscope at 100× for the examination of chromosomes, which is difficult at the best of times due to the tiny and numerous chromosomes of *Adansonia*, were mistaken in their chromosome counts. There is, thus, no evidence of diploid baobabs in Africa.

Most studies of baobab population structure have focused on microsatellites (Larsen & al., 2009; Kyndt & al., 2009; Munthali & al., 2013; Wiehle & Prinz & al., 2014), plastid DNA (Leong Pock Tsy & al., 2009), or AFLPs (Assogbadjo & al., 2009). However, it was previously suggested that *A. kilima* might be distinct from *A. digitata* based partly on ITS sequences (Pettigrew & al., 2012), leading us to look for evidence of this in a broad sample of specimens.

Phylogenetic analysis of ITS sequences did not support the existence of subpopulations based on either altitude or productivity (viz. producer and poor-producer trees in Venda), as was also found with analysis of microsatellite data (Tivakudze, 2014; Venter & al., submitted.). This result mirrors analyses using AFLP markers, which failed to find a correlation between genetic variation and baobab morphotypes traditionally recognized by ethnic groups from Benin, Ghana and Senegal (Assogbadjo & al., 2009). Nonetheless, the ITS data support some degree of genetic differentiation, with West African accessions forming a clade (including a few other samples) relative to an unresolved grade composed of accessions from southern and eastern Africa. Our tree topology matches that reported by Pettigrew & al. (2012) whose sampled *A. digitata* were all derived from West Africa in contrast to the samples designated '*A. kilima*', which all came from southern Africa. With our much broader sampling in eastern and southern Africa, it is clear that there are no consistent differences in ITS sequence between *A. digitata* and *A. kilima*. Of particular note, the holotype and paratype specimens of *A. kilima* were not distinct from southern

African specimens assigned to *A. digitata* undermining the claim that *A. kilima* corresponds to a distinct species.

Even if *A. kilima* is considered a synonym of *A. digitata* based on the lack of a distinction between the type specimen and other southern African baobabs, this does not refute the hypothesis that two, tetraploid baobab species occur side-by-side in the highlands of East Africa. At first sight our ITS data might seem to support this inference. Several Tanzanian accessions, including all that have an '*A. kilima*' morphology have distinct haplotypes, not found elsewhere in Africa. Although a few accessions from Tanzania with an '*A. digitata*' morphology share these haplotypes, other '*A. digitata*' accessions carry more typical haplotypes. While the sample size remains small, this pattern suggests that two distinct genetic lineages might exist in Tanzania, and that genetic heritage is manifest to some degree in local morphological differentiation. We would speculate, therefore, that the occurrence of discrete morphotypes in sympatry in East Africa led Pettigrew & al. (2012) into the mistaken view that two sympatric baobab species exist when in fact the pattern instead may reflect recent migration of two genetic varieties into sympatry.

Looking at all of the data, there is no evidence of any phenotypic or genotypic distinction between the southern African types of *A. kilima* and *A. digitata*, whether we consider floral morphology, pollen, or stomatal traits, chromosome number, or ITS sequence. In short, the data all indicate that *A. kilima* is a synonym of *A. digitata*. Furthermore, even in the East African Highlands, where two morphologically distinct varieties occur in sympatry, there is evidence of neither a ploidy difference nor of the strong genetic differentiation needed to justify the recognition of more than one baobab species. At this time, the evidence supports the existence of just one African baobab species, *A. digitata*.

Formal taxonomy.—

Adansonia digitata L., Syst. Nat. (ed. 10), 2: 1144. 1759 – Lectotype (designated by Robyns, 1980): Type: s. loc., s. coll. (LINN 862.1).

= *Adansonia kilima* Pettigrew, K.L.Bell, Bhagw., Grinan, Jillani, Jean Mey., Wabuyele & C.E.Vickers, Taxon 61(6): 1248, f. 2A, C, f. 3"AK". 2012 – Type: South Africa, Limpopo, on south side of road between Tshkuwi and Tshituni, 16 km east of the N1 highway, 12 Nov 2009, *Pettigrew & Callistemon 1* (PRE barcode PRE0857831-0), syn. nov.

ACKNOWLEDGEMENTS

The authors thank X. Dong and A. Schweiner for assistance with DNA extractions and sequencing, Sarah Friedrich and Sue Medaris for assistance with artwork, and W. Hattingh for assistance with Box and Whisker plots in R. We thank collaborators who provided specimens (D. Mayne, A. Patrut, K. Bell, D. Murphy) as well as the Royal Botanic Garden, Victoria, Australia and J. Pettigrew for providing coordinates for the *A. kilima* type locality. We thank the herbaria listed herein for loan of specimens and permission to use leaf and pollen material for microscopic work. This work was partially funded by the National Research Foundation (NRF) of South Africa's Centre of Excellence in Tree Health Biotechnology (CTHB) and the NRF's Integrated Biodiversity Information Programme (grant 86959). KLG was supported by grant B8749.R01 from the Carnegie Corporation of New York to the Global Change and Sustainability Research Institute at the University of the Witwatersrand. DAB and NK were funded by the National Science Foundation (Grant DEB-1354793).

REFERENCES

Abramoff, M.D., Magalhaes, P.J. & Ram, S.J. 2004. Image Processing with ImageJ. *Biophotonics Int.*11: 36.

Assogbadjo, A.E., Kakai, R.G., Chadare, F.J., Thomson, L., Kyndt, T., Sinsin, B. & Van Damme, P. 2008. Folk classification, perception, and preferences of baobab products in West Africa: consequences for species conservation and improvement. *Econ. Bot.* 62: 74.

Assogbadjo, A.E., Kyndt, T., Chadare, F.J., Sinsin, B., Gheysen, G., Eyog-Matig, O. & Van Damme, P. 2009. Genetic fingerprinting using AFLP cannot distinguish traditionally classified baobab morphotypes. *Agrofor. Syst.* 75: 157.

Baum, D.A. & Oginuma, K.. 1994. A review of chromosome numbers in Bombacaceae with new counts for *Adansonia. Taxon* 43: 11.

Baum, D.A., Small, R.L. & Wendel, J.L. 1998. Biogeography and floral evolution of baobabs (*Adansonia*, Bombacaceae) as inferred from multiple data sets. *Syst. Biol.* 47: 181.

Bingham, E.T. 1968. Stomatal chloroplast alfafa at four ploidy levels. *Crop Sci.* 8: 509.

Chen, L.-Q., Li, C.-S., Chaloner, W.G., Beerling, D.J., Sun, Q.-G., Collinson, M.E. & P.L. Mitchell. 2001. Assessing the potential for the stomatal characters of extant and fossil Ginkgo leaves to signal atmospheric CO2 change. *Am. J. Bot.* 88: 1309.

Clement, M., Posada, D.C.K.A. & Crandall, K.A. 2000. TCS: a computer program to estimate gene genealogies. *Mol. Ecol.* 9: 1657.

Douie, C., Whitaker, J. & Grundy, I. 2015. Verifying the presence of the newly discovered African baobab, *Adansonia kilima*, in Zimbabwe through morphological analysis. *S. African J. Bot.* 100: 164.

Duarte, M.C., Esteves, G.L., Salatino, M.L.F., Walsh, K.C. & Baum D.A. 2011. Phylogenetic analyses of *Eriotheca* and related genera (Bombacoideae, Malvaceae). *Syst. Bot.* 36: 690.

Holmgren, P. K., Holmgren, N.H. & Barnett, L.C. (eds.), 1990. *Index herbariorum. Part 1: The herbaria of the world.* New York Botanical Garden, New York.

Huelsenbeck, J. P. & Ronquist, F. 2001. MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* 17: 754.

Körner, C.H. & Cochrane, P.M. 1985. Stomatal responses and water relations of *Eucalyptus pauciflora* in summer along an elevational gradient. *Oecologia* 66: 443.

Kluge J. & Kessler M. 2007. Morphological characteristics of fern assemblages along an elevational gradient: patterns and causes. *Ecotropica* 13: 27.

Kruskal, J. B. 1964. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* 29: 1.

Kyndt, T., Assogbadjo, A.E., Hardy, O.J., Kakai, R.G., Sinsin, B., Van Damme, P. & G. Gheysen. 2009. Spatial genetic structuring of baobab (*Adansonia digitata*, Malvaceae) in the traditional agroforestry systems of West Africa. *Am. J. Bot.* 96: 950.

Larsen, A.S., Vaillant, A., Verhaegen, D. & Kjær, E. D. 2009. Eighteen SSR-primers for tetraploid *Adansonia digitata* and its relatives. *Conserv. Genet. Res.* 1: 325.

Leong Pock Tsy, J. M., Lumaret, R., Mayne, D., Vall, A.O.M.,. Abutaba, Y.I.M, Sagna, M. Raoseta, S.O.R. & Danthu, P. 2009. Chloroplast DNA phylogeny suggests a West African centre of origin for the baobab, *Adansonia digitata* L. (Bombacoideae, Malvaceae). *Mol. Ecol.* 18:1707.

Mishra, M.K. 1997. Stomatal Characteristics at Different Ploidy Levels in *Coffea* L. *Ann. Bot-London* 80: 689.

Munthali, C.R.Y., Chirwa, P.W., Changadeya, W.J. & Akinnifesi, F.K. 2013. Genetic differentiation and diversity of *Adansonia digitata* L. (baobab) in Malawi using microsatellite markers. *Agrofor. Syst.* 82: 117.

Pettigrew, J.D., Bell, K.L., Bhagwandin, A., Grinan, E., Jillani, N., Meyer, J., Wabuyele, E., & C. Vickers. 2012. Morphology, ploidy and molecular phylogenetics reveal a new diploid species from Africa in the baobab genus *Adansonia* (Malvaceae: Bombacoideae). *Taxon* 61: 1240.

Porebski, S., Bailey, L.G. & Baum, B. R. 1997. Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Mol. Biol. Rep* 15: 8.

R Core Team. 2014. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Rasband, W.S. 1997-2015. ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA, http://imagej.nih.gov/ij/.

Ramsey, J. & Schemske, D.W. 2002. Neopolyploidy in flowering plants. *Ann. Rev. Ecol. and Syst.* 33: 589.

Robyns, A. 1980. Bombacaceae. In: M.D. Dassanayake & F.R. Fosberg (eds.) *A revised handbook of the flora of Ceylon, Vol. 1*. Amerind Publishing, New Delhi. Pp. 58–91.

Rohlf, F. J. 2008. *NTSYS-pc. Numerical taxonomy and multivariate analysis system.* Version 2.2. Applied Biostatistics, New York.

Sanchez, A.C. 2010. Predicting suitable areas for cultivation and conservation of the baobab tree and investigating superior sources of planting material. PhD Dissertation, University of Southampton, Faculty of engineering and the environment.

Sanchez, A.C., Haq, N. & Assogbadjo, A.E.. 2010. Variation in baobab (*Adansonia digitata* L.) leaf morphology and its relation to drought tolerance. *Genet. Resour. Crop Evol.* 57: 17.

Schneider, C.A., Rasband, W.S. & Eliceiri, K.W. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* 9: 671.

Stamatakis, A. 2014. RAxML Version 8. 2014. A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* (http://sco.h-its.org/exelixis/web/software/raxml/)

Stebbins, G.L. 1971. Chromosomal evolution in higher plants. London: Addison-Wesley.

Tivakudze, R. 2014. Assessing ploidy-level and gene flow between baobab (*Adansonia digitata*) fruit producers and poor producers in Limpopo. MSc Thesis, University of the Witwatersrand, Johannesburg, South Africa.

Venter, S.M. 2012. The ecology of baobabs (*Adansonia digitata* L.) in relation to sustainable utilization in northern Venda, South Africa. PhD Dissertation, University of the Witwatersrand, Johannesburg, South Africa.

Venter, S. M. & Witkowski, E.T. F. 2011. Baobab (*Adansonia digitata* L.) fruit production in communal and conservation land-use types in Southern Africa. *Forest Ecol. Manag.* 261: 630.

Wang, R., Yu, G., He, N., Wang, Q., Xia, F., Zhao, N. & Ge, J. 2014. Elevation-related variation in leaf stomatal traits as a function of plant functional type: Evidence from Changbai Mountain, China. *PLoS ONE*, *9*(12), e115395. doi:10.1371/journal.pone.0115395

Wiehle, M., Prinz, K., Kehlenbeck, K. & al. 2014. The African baobab (*Adansonia digitata*, Malvaceae): Genetic resources in neglected populations of the Nuba Mountains, Sudan. *Am. J. Bot.* 101: 1498.

Appendix 1. Accessions used for phylogenetic analyses of *Adansonia*. GenBank accession numbers for the ITS sequences and original publications and collected samples. * = producer; ** poor producer (per Venter & Witkowski, 2011). ***Taxon***: GenBank ITS accession; collection locale; geographic coordinates; reference (if not new accession) or voucher for new accession (if available).

***A. digitata* L.:** AF028625; -; Baum & al. 1998. AF028626.1; -; Baum & al. 1998. JN400317; Australia (cultivated); Pettigrew & al. 2012. JN400311; Limpopo, South Africa; Pettigrew & al. 2012. KU145696; Benin; 11.793881, 3.219264787; -. KU145697; Benin; 12.37063573, 2.833352107; -. KU145698; Côte d'Ivoire; 9.429778, −5.628944; -. KU145699; Togo; 6.591722, 1.384722; -. KU145700; Togo; 11.032433, 0.114733; -. KU145701; Côte d'Ivoire; 9.499194, −5.281194; -. KU145702; Benin; 12.017577, 3.2096878; -. KU145703; Burkina Faso; 12.1775, 0.0025; -. KU145704; Togo; 9.625433, 0.985089; -. KU145705; Namibia; −17.747733, 14.894883; -. KU145706; Benin; 12.14918819, 3.08182988; -. KU145707; Côte d'Ivoire; 9.605028, −5.303417; -. KU145708; Benin; 12.29708331, 2.875692609; -. KU145709; Côte d'Ivoire; 9.517361, −5.264806; -. KU145710; Togo; 9.326858, 0.584536; -. KU145711; Togo; 6.4385, 1.251861; -. KU145712; Côte d'Ivoire; 9.1400, −5.50775; -. KU145713; Côte d'Ivoire; 9.347389, −5.616944; -. KU145714; Namibia; −17.501833, 15.015567; -. KU145715; Burkina Faso; 11.22758, 0.768492; -. KU145716; Namibia; −17.502317, 14.983083; -. KU145717; Côte d'Ivoire; 9.378139, −5.632917; -. KU145718; Burkina Faso; 11.668694, 1.535361; -. KU145719; Benin; 7.9616, 1.97895; -. KU145720; Nxai Pan National Park, Botswana; −20.106339, 24.775291; -. KU145721; Namibia; −19.600267, 20.428867; -. KU145722; Botswana; 20.1896, 25.30539; -. KU145723; Namibia; −19.301017, 20.661433; -. KU145724; Limpopo, South Africa; −22.06583333, 30.00781667; -. KU145725; Limpopo, South Africa; −22.06583333,

30.00781667; -. KU145726; Limpopo, South Africa; −22.500871, 30.63325; -. KU145727; Zimbabwe; −17.902261, 25.842733; -. KU145728; Tsodilo Hills, Botswana; 18.75852, 21.73729; -. KU145729; Zimbabwe; −17.913817, 25.841783; -. KU145730; Skelmwater Tree 07, Limpopo, South Africa; −22.50004, 29.98339; -. KU145732; Limpopo, South Africa; −22.574250, 30.490371; -. KU145734; Tsodilo Hills, Botswana; 18.75368, 21.74015; -. KU145735; Mozambique; -16.771667, 37.031111; -. KU145736; Skelmwater Tree 13, Limpopo, South Africa; −22.50004, 29.98339; -. KU145738; Skelmwater Tree 10, Limpopo, South Africa; −22.50004, 29.98339; -. KU145739; Limpopo, South Africa; −22.900967, 29.99855; -. *KU145741; Limpopo, South Africa; −22.546845, 30.721447; -. **KU145742; Limpopo, South Africa; −22.546548, 30.721184; -. *KU145743; Limpopo, South Africa; −22.543778, 30.742028; -. *KU145744; Limpopo, South Africa; −22.403816, 30.930389; -. **KU145745; Limpopo, South Africa; −22.572381, 30.507167; -. **KU145746; Limpopo, South Africa; -22.550399, 30.716306; -. **KU145747; Limpopo, South Africa; −22.562639, 30.489816; -. **KU145748; Limpopo, South Africa; −22.562371, 30.490188; -. *KU145749; Limpopo, South Africa; −22.570139, 30.497389; -. *KU145750; Limpopo, South Africa; −22.538371, 30.732816; -. *KU145751; Limpopo, South Africa; -22.420990, 30.945622; -. **KU145752; Quelilea Island, Mozambique; −12 29 34.7, 40 36 13.7; -. *KU145753; Quelilea Island, Mozambique; -. −4.903983, 35.782267; -. KU145760; Limpopo, South Africa; −22.497717, 29.981817; -. *KU145761; Quelilea Island, Mozambique; −12 29 31.3, 40 36 03.5; -.KU145762; Skelmwater Tree 08, Limpopo, South Africa; −22.50004, 29.98339; -. KU145763; Zimbabwe; −17.9125, 25.841333; -. KU145765; Limpopo, South Africa; −22.423717, 31.284967; -. *KU145767; Limpopo, South Africa; −22.573028, 30.5070; -. **KU145768; Limpopo, South Africa; −22.471528, 30.673333; -. KU145769; Mozambique; −16.876944, 33.271944; -. KU145772; Tsodilo Hills, Botswana; −18.75638, 21.744507; -. KU145773; Zambezi (formerly

Caprivi), Namibia; −17.61711, 24.40019; -. KU145774; Benin; 12.350488, 2.849489; -. KX118401; Tanzania; −6.636493, 38.15217; -. KX118402; Tanzania; −6.874551, 37.269445; -. KX118404; Tanzania; −4.982818, 35.797188; *N.Karimi-2015-238* (WIS). KX118410; Tanzania; −4.791356, 38.195916; -. KX118406; Tanzania; −3.698297, 37.581407; -. KX231845; Victoria Falls, Zimbabwe; -17.913817, 25.841783; -. KX231846; Nyae Nyae Conservancy, Namibia, -19.65539 20.710662; -. KX231847; Nyae Nyae Conservancy, Namibia; -19.677910, 20.618027; -.

***A. grandidieri* Baill.:** AF028528.1; Baum & al. 1998.

***A. gregorii* F.Muell.:** AF028524; Baum & al. 1998.

***A. kilima* Pettigrew & al.:** JN00324; Tanzania; Pettigrew & al. 2012. JN400325; Tanzania; Pettigrew & al. 2012. JN400326; Tanzania; Pettigrew & al. 2012. JN400327; Tanzania; Pettigrew & al. 2012. JX178939; South Africa; Pettigrew & al. 2012. JX178940, Namibia; Pettigrew & al. 2012. KU145771; Holotype; Limpopo, South Africa; −22.901617, 29.996567; -. KU145766; Holotype tree; Limpopo, South Africa; -. KU145770; Paratype tree; Zambezi (formerly Caprivi), Namibia; −17.61751, 24.39892; -. KX118403; Tanzania; −3.775553, 35.967692; -. KX118407; Tanzania; -4.089764, 37.757617; *N.Karimi 2015-243* (WIS). KX8409; Tanzania; -4.791395, 38.19592; *N.Karimi 2015-248* (WIS). KX118411; Tanzania; −4.98298, 35.7982; N.Karimi-2015-251. KU145755; Kondoa, Tanzania; -4.903983, 35.782267; *N.Karimi 2015-250* (WIS). KX118405; Tanzania; −3.352094, 37.344533; *N.Karimi-2015-239* (WIS).

***A. madagascariensis* Baill.:** AF028534; -; Baum & al. 1998. JN400312; -; Pettigrew & al. 2012.

A. perrieri Capuron: AF028538; -; Baum & al. 1998. JN400315; -; Pettigrew & al. 2012.

***A. rubrostipa* Jum. & H.Perrier:** AF028530.1; -; Baum & al. 1998. JN400313.1; -; Pettigrew & al. 2012.

***A. suarezensis* H.Perrier:** AF028530.1; -; Baum & al. 1998. JN400314.1; -; Pettigrew & al. 2012.

***A. za* Baill.:** AF02837; -; Baum & al. 1998. JN400316; -; Pettigrew & al. 2012.

***Cavanillesia platanifolia* (Humb. & Bonpl.) Kunth.:** HQ658371; -; Duarte & al. 2011.