OXFORD UNIVERSITY PRESS | Systematic Biology

Reticulate evolution helps explain apparent homoplasy in floral biology and pollination in baobabs (Adansonia; Bombacoideae; Malvaceae)

Journal:	*Systematic Biology*
Manuscript ID	USYB-2019-138
Manuscript Type:	Regular Manuscript
Date Submitted by the Author:	25-Jul-2019
Complete List of Authors:	Karimi, Nisa; University of Wisconsin Madison, Botany; Morton Arboretum, Grover, Corrinne; Iowa State University, Department of Ecology, Evolution, and Organismal Biology Wendel, Jonathan; Iowa State University, Ecology, Evolution, and Organismal Biology Department Gallagher, Joseph ; Iowa State University, Department of Ecology, Evolution, and Organismal Biology Ane, Cecile; University of Wisconsin-Madison, Dpt of Statistics Baum, David; University of Wisconsin, Botany;
Keywords:	Targeted sequence capture, Hyb-Seq, network inference, introgression, species tree inference, comparative methods

TITLE: Reticulate evolution helps explain apparent homoplasy in floral biology and pollination in baobabs (*Adansonia*; Bombacoideae; Malvaceae)

RUNNING HEAD: Reticulation in the Baobabs

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

Keywords*: Targeted sequence capture, Hyb-Seq, reticulate evolution, network inference, introgression, species tree inference, comparative methods*

ABSTRACT

Baobabs (*Adansonia)* are a cohesive group of tropical trees with a disjunct distribution in Australia, Madagascar, and continental Africa, and diverse flowers associated with two

pollination modes. We used custom targeted sequence capture in conjunction with new and existing phylogenetic comparative methods to explore the evolution of floral traits and pollination systems while allowing for reticulate evolution. Our analyses suggest that relationships in *Adansonia* are confounded by reticulation, with network inference methods supporting at least one reticulation event. The best supported hypothesis involves introgression between *A. rubrostipa* and core Longitubae, both of which are hawkmoth pollinated with yellow/red flowers, but there is also some support for introgression between the African lineage and Malagasy Brevitubae, which are both mammal-pollinated with white flowers. New comparative methods for phylogenetic networks were developed that allow maximum-likelihood inference of ancestral states and applied to study the apparent homoplasy in floral biology and pollination mode seen in *Adansonia*. This analysis supports a role for introgressive hybridization in morphological evolution even in a clade with highly divergent and geographically widespread species. Our new comparative methods for discrete traits on species networks are implemented in the software PhyloNetworks.

It is widely accepted that hybridization and its genetic consequence, introgression, are widespread phenomena in plants (Arnold, 1992; Reiseberg and Wendel, 1993; Soltis & Soltis, 2009; Payseur & Rieseberg, 2016), and equally in other clades, including animals (Mallett, 2008; Schwenk et al., 2008; Payseur & Rieseberg, 2016). Furthermore, it has also long been appreciated that introgression can be an agent of adaptive evolution in cases where beneficial traits are transferred from a donor to a recipient taxon (Arnold 2004; Reiseberg and Wendel, 1993; Arnold and Kunte, 2017; Suarez-Gonzalez et al. 2018). These facts raise the possibility that adaptive introgression could sometimes carry ecologically important traits between species, and that such events might explain apparent homoplasy when these traits are mapped onto binary trees that do not incorporate the true hybridization history. However, we know relatively little about this possibility. Macroevolutionary methods for systematically detecting reticulation using phylogenomic data are relatively new (Yu et al. 2011, 2014; Park and Nakhleh, 2012; Solís-Lemus et al. 2016; Zhang et al. 2017), and still not widely deployed due to the large amounts of data they require and their computational intensity. Furthermore, comparative methods for ancestral state reconstruction on phylogenetic networks are needed to assess the likelihood that a trait was acquired via a minor hybridization edge, but such methods are also in their infancy (Bastide at al. 2018). Here, we used phylogenomic data to infer the history of reticulation in baobabs and then developed and deployed network-aware ancestral state reconstruction methods to evaluate the possible role of adaptive introgression in explaining apparent floral homoplasy.

The baobab genus *Adansonia* (Malvaceae) includes eight morphologically distinct species (Fig. 1) (Baum 1995b; Cron et al. 2016). The group possesses an unusual geographic distribution, with one tetraploid species (*A. digitata*) that is widespread across continental Africa, one diploid species (*A. gregorii*) endemic to Northwestern Australia, and six diploid species that

are restricted to Madagascar. Previous phylogenetic analyses identified three distinct lineages corresponding to geography; however, relationships among these remain unresolved (Baum et al. 1998; Pettigrew et al. 2015).

Within Madagascar, the six recognized species are assigned to two sections based on morphology (Baum 1995b). Brevitubae includes a pair of allopatric species, *A. grandidieri* in the southwest and *A. suarezensis* in the north, which form a clade well-supported by ITS sequences (Baum et al. 1998). Both species are dry-season flowering whose flowers are characterized by short staminal tubes, white petals, cream-colored inner calyces, and are inferred to be mammal pollinated (Baum 1995a; Andriafidison et al. 2006). The four species of Longitubae are wet-season flowering and 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, although they do support a clade composed of *A. madagascariensis*, *A. za*, and *A. perrieri* (hereafter referred to as "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).

In order to infer explicit phylogenetic networks, that is genealogical histories that take account of both incomplete lineage sorting (ILS) and reticulation (hybridization and introgression), information from many independent gene genealogies are needed (Raymond et al. 2002; Yu et al. 2011, 2014; Solís-Lemus et al. 2016; Zhang et al. 2017). Targeted sequence capture, or hyb-seq, has increased in popularity as a source of hundreds of low-copy nuclear

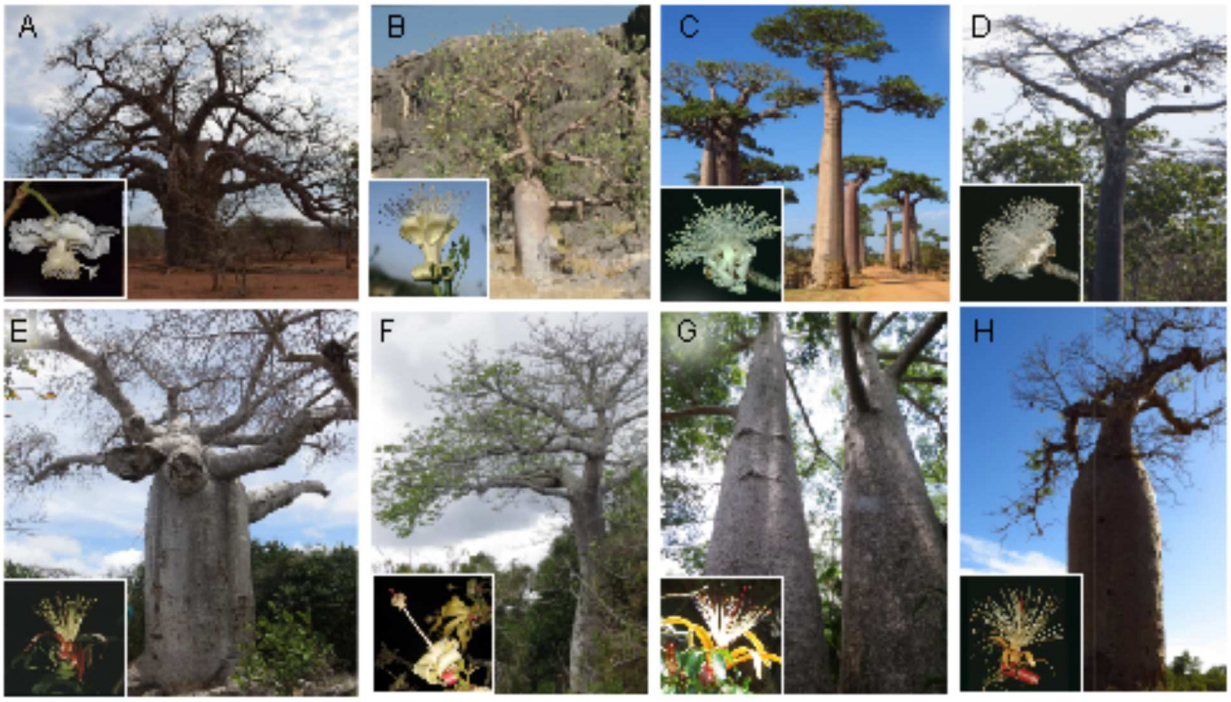


Figure 1. Eight species of *Adansonia.* A) *A. digitata*, continental Africa, B) *A. gregorii,* Australia, C) *A. grandidieri*, Madagascar, D) *A. suarezensis*, Madagascar, E) *A. madagascariensis*, Madagascar, F) *A. perrieri*, Madagascar, G) *A. za*, Madagascar, H) *A. rubrostipa,* Madagascar.

genes for the purpose of multigene phylogenetics 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). Biotinylated RNA "baits" selectively enrich targeted genomic loci in next-generation sequencing libraries. Furthermore, organellar and ribosomal sequences can usually be recovered from the off-target reads (Weitemier et al. 2014). Targeted sequence capture is typically used with short-read sequencing, which makes it appropriate and useful for samples with degraded or poor-quality DNA, such as herbarium specimens (Hart et al. 2016, Villaverde et al. 2018).

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 (Yang and Smith 2014; Johnson et al. 2016; Kamneva et al. 2017, Fér and Schmickl 2018).

Here we used custom-designed baits to obtain sequences of hundreds of independent low-copy nuclear loci for all species of baobab. Despite the large size of our data set, relationships among the three geographic lineages could not be resolved. Within Madagascar the optimal population trees support a sister group relationship between Brevitubae and core Longitubae to the exclusion of *A. rubrostipa*. This implies floral homoplasy since core Longitubae and *A. rubrostipa* share elongated, yellow and red, hawkmoth-pollinated flowers, whereas Brevitubae and *A. digitata* have short, white, mammal-pollinated flowers (*A. gregorii* has a relatively short, white, mainly hawkmoth-pollinated flowers; Baum 1995a). These data show that relationships in *Adansonia* are confounded by reticulation, with network inference methods supporting at least one major reticulation event, and possibly a second, though this inference is confounded by methodological limitations and read assembly challenges, possibly due to paralogy. We developed new phylogenetic comparative methods for species networks, in which discrete morphological traits may be inherited via reticulation. Using this method we show that reticulation edges help explain evolution of floral biology and apparent homoplasy in baobabs.

METHODS

Bait Design

124

125 Total RNA was extracted from fresh leaf tissue of *Adansonia digitata* L. and *Bombax*
126 *ceiba* L., as in Chang et al. (1993), followed by cleanup using the Qiagen RNeasy kit following
127 manufacturer instructions (QIAGEN Inc., Valencia, California, USA). RNA quality assessment
128 was performed by Agilent RNA PicoChip Analysis (Agilent Technologies, Inc., Santa Clara,
129 California, USA) using 1µl of each sample diluted to 5 ng/µl. RNA library preparation was
130 performed at the University of Wisconsin - Madison Biotechnology Center (Madison, WI) using
131 an Illumina TruSeq RNA Sample Prep kit (Illumina Inc., San Diego, CA, USA) followed by
132 purification with Agencourt AMPure XP beads (Beckman Coulter, USA). Library quantification
133 was checked with a Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, USA) per
134 manufacturer instructions. Samples were adjusted to a final concentration of 28-31ng/µl. Quality
135 and quantity of the finished libraries were assessed using an Agilent DNA1000 chip (Agilent
136 Technologies, USA) and Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, USA),
137 respectively. Libraries were standardized to 2nM. Cluster generation was performed using
138 Illumina TruSeq Cluster Kits and the Illumina Cluster Station (Illumina Inc., San Diego, CA,
139 USA). Paired-end, 100bp sequencing was performed at the University of Wisconsin - Madison
140 Biotechnology Center (Madison, WI) using SBS chemistry on an Illumina HiSeq2000
141 sequencer. Images were analyzed using the Illumina Pipeline, version 1.8.2, and raw reads were
142 assembled *de novo* with Trinity version 2.1.0 (Luo et al. 2012). The resulting contigs were used
143 as BLAST queries against each other, the *Arabidopsis* ultra-conserved sequence database
144 (http://cgpdb.ucdavis.edu/cgpdb2/), and the *Gossypium* exome (Paterson et al. 2012). For each
145 contig showing reciprocal best-BLAST matches between *Adansonia* and *Bombax, Adansonia-*
146 *Bombax-Gossypium* alignments were identified that had >800bp of continuously aligned

147 sequences and an average pairwise sequence similarity between the Bombacoids and cotton ≥
148 93.5%. These candidates were screened to remove candidate targets with repetitive sequences by
149 RepeatMasker (Smit et al. 2015) and for potential gene families (sequences that clustered with
150 two or more genes from the cotton transcriptome). When base calls were ambiguous in
151 *Adansonia*, the bait sequence was based on *Bombax*. When both species were polymorphic at a
152 position, the most common base call was selected. The resulting targets were used as the basis
153 for the design and synthesis of 120 bp, 2X-tiled MYbaits (Arbor Biosciences, formerly
154 Mycroarray, Ann Arbor, MI, USA), available as Supplementary File S1. Transcriptomes are
155 available on NCBI Sequence Read Archive under accession PRJNA493960 (Conover et al.
156 2018).
157
158 *Taxon Sampling and Targeted Sequence Capture*
159
160 Sampling included one to three accessions per *Adansonia* species in addition to three
161 outgroups (Table S1). DNA was extracted from silica-dried leaf tissue or seeds with the Qiagen
162 DNeasy Plant Mini Plant Kit (Qiagen, USA), following manufacturer instructions, but with the
163 following modifications: (1) increased lysis buffer to 650µl and included 10µl Proteinase K
164 (25mg/mL), (2) tissue and lysis buffer was incubated at 65°C for 20 minutes rather than the
165 recommended 10 minutes, (3) all centrifugation steps were performed at 4°C, and (4) final
166 elution used heated buffer (approx. 80°C) and was then incubated at room temperature for 10
167 minutes before centrifugation. DNA quality and quantity were estimated by 1% agarose gel
168 electrophoresis. Qubit Fluorometric Quantitation (Life Technologies) was used for further
169 quantification prior to DNA library preparation.

DNA library construction was performed at the University of Wisconsin - Madison Biotechnology Center or 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 (see Table S1) were sequenced on a single lane of Illumina MiSeq as 2x300bp by the UW-Madison Biotechnology Center (Madison, WI). An additional four accessions (see Table S1) 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* (all scripts available at https://github.com/nkarimi/Adansonia_HybSeq)

Read Assembly and Dataset Selection

The first phase of assembly (see Supplementary Fig. S1 for flow chart) used the HybPiper package (Johnson et al. 2016) with our initial target set as a reference. HybPiper,

yields a "paralog warning" whenever more than one contig is assembled that covered >85% of the target. For putative single-copy targets, which yielded no paralog warnings for any of the initial 12 (MiSeq-derived) accessions, we retained the resulting sequences and aligned them using MAFFT version 7.299 (Katoh et al. 2012). Whenever HybPiper gave a paralog warning for any of these taxa we retained all sequence variants for all accessions, aligned them using MAFFT, and generated maximum likelihood trees in Geneious version 8.0.5 (http://www.geneious.com, Kearse et al. 2012) using the RAxML plugin. The resulting gene trees were inspected to determine if the multiple contigs most likely represented alleles (accessions from the same species form a clade) or paralogs (distinct clades each with multiple species). For gene trees without evidence of paralogs, we retained a single allelic sequence for each accession (the longest one, as selected by HybPiper).

Rather than discarding targets with paralogs, we sought to generate paralog-specific references using gene tree-guided orthology identification. As summarized in Supplementary Fig. S1, a separate consensus sequence was generated from all sequences of each putative paralog and these paralog-specific consensus sequences were used as new, operational targets for assembly in HybPiper. If the resulting assemblies lacked paralog warnings, and also appeared as distinct clades on gene trees generated after alignment of all paralogs for a given original target, each paralog assembly was retained. If sequences from the operational targets did not separate on the resulting gene trees, suggesting orthology-paralogy mixing, the target was dropped. If, alternatively, gene trees suggested yet further paralogs, the process was repeated iteratively until each target was either dropped due to orthology-paralogy mixing or yielded one or multiple alignments of putatively orthologous sequences.

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 version 8.2.10 (Stamatakis 2006) failed to satisfy the unrooted backbone constraint ((Aru001, Aru127),(Ape001, Ape009)). Removal of these genes resulted in our primary HybPiper data set of 372 putative orthologs.

Although prior work has shown that allele phasing may have minimal impact on phylogenetic inference (Kates et al. 2017), we sought to obtain allelic information as an additional set of alignments. To generate clean targets for haplotype assembly, we used consensuses of all targets (for all accessions) in the primary HybPiper data set after dropping a further 28 targets whose optimal RAxML trees that suggested the possibility of mis-assembly. This included trees with single terminal branches at least three times longer than others on that tree, or topologies that could not be rooted to support a monophyletic *Adansonia* clade. The 344 "cleaned" targets meeting these criteria were used to infer phased haplotypes. Trimmed reads were mapped to the 344 accession-specific references using BWA v.0.7.15 (Li and Durban 2009) with the *bwa mem* algorithm. Haplotypes were then 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 allowed for the tetraploid *A. digitata* and two for all other taxa. We then generated 10 alternative HapHunt

238 alignments (HH1-10), each containing one haplotype sequence per target per accession, sampled
239 at random from the haplotypes inferred for that accession.
240
241 *Species-Tree Phylogenetic Analyses and Network Inference*
242
243 Bayesian phylogenetic inference was performed on all datasets with MrBayes
244 (Huelsenbeck and Ronquist 2001; Ronquist and Huelsenbeck 2003), either as a free-standing
245 application or as implemented in the TICR pipeline (Stenz et al. 2015). MrBayes analyses used 4
246 linked chains with a heat of 0.2 and ran for two million generations, with 25% discarded as burn-
247 in. The resulting posterior distributions were analyzed in BUCKy version 1.4.4 (Ané et al. 2006,
248 Larget et al. 2010) using alpha=1 and 1,000,000 generations. The nuclear HapHunt datasets,
249 HH1-10, were analyzed individually. Additionally, to shed light on the effect of allele sampling,
250 we combined the 10 posterior tree distributions for each gene into a composite posterior and used
251 the resulting combined HapHunt data set for several downstream analyses.
252 Bayesian Concordance Analysis (Ané et al. 2007; Baum 2007) was implemented with
253 BUCKy (Ané et al. 2007; Larget et al. 2011). BUCKy allows one to estimate the proportion of
254 gene trees supporting a certain clade (the concordance factor, CF), while taking into account
255 uncertainty in individual gene trees and also estimates a population tree (="species tree") under
256 the assumption that all discordance is due to ILS.
257 We also generated maximum likelihood gene trees with RAxML version 8.2.10
258 (Stamatakis 2006, 2014) and then used these to infer a population tree in ASTRAL-III (Mirarab
259 et al. 2014; Zhang et al. 2018). Concatenated nuclear gene alignments were also used to infer a

population tree using SVD quartets (Chifman and Kubatko 2015) in PAUP* (version 4a; Swofford et al. 2003).

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. The concordance factor table generated by BUCKy in the TICR pipeline was used as input into SNaQ. To test for the effect of input tree, starting population trees with branch lengths in coalescent units were obtained from both BUCKy and from the TICR pipeline, the latter of which uses Quartet MaxCut (Snir and Rao 2012).

We ran 50 independent runs of SNaQ for each dataset to infer the optimal network without hybridization edges (h0). This was then used as input for network searches with one hybridization edge (h1). Network searches were increased sequentially, up to 3 hybridization edges, in each case starting from the previous (h1 or h2) optimal network. The preferred number of hybridizations was determined 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 (sampling from the confidence intervals for each possible quartet CF) 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.

Four-Taxon D-statistic (ABBA-BABA)

Guided by sensitivity of SNaQ to taxon sampling (see Results), we also used the D-statistic "ABBA-BABA" test (Kulathinal et al. 2009) on the concatenated gene alignments from the HybPiper dataset using a modification of the "CalcD" function from the R package evobiR (Blackmon et al. 2015) with ambiguous sites dropped. To test for a hybridization between *A. digitata* and Brevitubae we analyzed all 162 BLdo quartets, namely those that contain one member each of Brevitubae (*A. grandidieri*; *A. suarezensis;*), core Longitubae (*A. madagascariensis*; *A. perrieri*; *A. za*), *A. digitata*, and the outgroup. Likewise, to test for a hybridization between *A. rubrostipa* and the core Longitubae we tested all 108 BLro quartets, namely those that contain one member each of Brevitubae, core Longitubae, A. *rubrostipa* and the outgroup. To test for sensitivity to taxon selection, we also used *A. gregorii* in place of the outgroups, yielding datasets BLdg and BLrg. For each quartet we calculated both the total number of ABBA and BABA sites in the concatenated alignment and the number of genes having more ABBA sites than BABA sites, or vice versa. Positive D-statistics show an excess of ABBA sites, while negative values show an excess of BABA sites. To evaluate significance, we used non-parametric bootstrap resampling of genes to obtain (for each 4-taxon data set) a Z-score, which is the calculated D-statistic divided by the bootstrap-estimated standard deviation of the D-statistic.

Plastome Assembly and Analysis

Reference guided assemblies 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. On 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 using the GTR-Γ model with 100 bootstrap replicates.

To test whether ILS, given the optimal network, could plausibly explain the recovered plastid phylogeny, we simulated expected plastid trees given two alternative population networks as inferred from the nuclear data. For each, we simulated 100,000 gene trees in the program hybrid-lambda (Zhu et al. 2015), after first multiplying all branch lengths in the network by 4 to account for the fact that plastid DNA experiences an effective population size one quarter that of nuclear genes. External branches were set to an arbitrary length of 1. Simulated tree topologies were input into PAUP* (version 4a; Swofford et al. 2003). We used tree filters to determine the frequency of targeted topologies among the 100,000 simulated gene trees.

Phylogenetic Comparative Method for Discrete Traits on a Species Network

To analyze flower color and pollinator mode in baobabs, we implemented maximum likelihood estimation of evolutionary parameters for discrete traits. We consider a trait with k

possible states as evolving under a Markov process along each edge of a known species network. At a reticulation node in the network, we assume that the trait of the hybrid population, Y_h, was inherited from either one of its parent populations, Y_1 or Y_2, with probabilities equal to the proportion of genes contributed by each parent population γ_1 and γ_2:

$$Y_h = \begin{cases} Y_1 & \text{with probability } \gamma_1 \\ Y_2 & \text{with probability } \gamma_2 = 1 - \gamma_1 \end{cases}.$$

This model may be interpreted in various ways. For instance, the trait might be controlled by a single gene, but this gene is unknown and probabilities γ_1 and γ_2 serve as prior probabilities that this unknown gene came from one parent or the other. Alternatively, the trait might be controlled by many genes of small effects, of which proportions γ_1 and γ_2 are expected to come from each parental population a priori. The γ inheritance probabilities are assumed to be known, along with the network topology and branch lengths. At the root of the phylogenetic network, the prior probability of each state may be assumed to be uniform across the k states, or may be taken as the stationary distribution of the process given the transition rates. 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 are estimated with maximum likelihood.

Conditional on the estimated rates, ancestral state estimations are obtained as the posterior probabilities of each state given the trait data at the leaves of the phylogenetic network. At each reticulation node, the posterior probability p_{gf} that the trait was inherited via gene flow is calculated as the posterior probability that the trait state was inherited from the minor parent population (with inheritance $\gamma<0.5$) given the trait data at the leaves of the network and given the estimated rates. This posterior probability p_{gf} is compared to the prior probability γ, to get a

Bayes factor comparing the hypotheses of inheritance via gene flow versus vertical inheritance: $\frac{p_{gf}}{\gamma}\frac{1-\gamma}{1-p_{gf}}$.

Our implementation is available in the open source Julia package PhyloNetworks (Solís-Lemus et al. 2017), version 0.9.1.

Ancestral Trait Reconstruction and Network Calibration

To infer the most likely flower color and pollinator of ancestral populations, we considered a Markov process for the evolution of these traits on inferred phylogenetic networks, as described above. For pollinators, we considered a binary trait with states hawkmoth-pollinated and mammal-pollinated and for flower color we considered a binary trait with states white and pigmented (=yellow or red). We scored the closest outgroup, *Scleronema micranthaum* as white flowered. Given that the methods cannot handle multiple states, *Scleronema* was dropped from the pollination analysis due to being interpreted as having mixed bat and moth pollination (van Dulmen 1998). In each case traits evolved along each lineage according to a 2-state Markov process with 2 transition rates and we assumed an equal prior probability for all states at the root of the phylogenetic network.

The trait evolution model uses 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 patristic distance between outgroup taxa and ingroup taxa (all loci have at least one outgroup taxon and one ingroup taxon). Note that the total tree length was not used to normalize gene trees, because genes with missing taxa are expected to have a lower tree length due to missing branches. 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. Networks inferred from the primary data set of 372 genes were calibrated using the pairwise distances obtained from the RAxML trees of these 372 genes. Networks inferred from the 344-gene HapHunt data set were calibrated using the pairwise distances obtained from the RAxML trees of one of the haplotype sets (HH6), whose network estimated with SNaQ has an estimated γ closest to that of the combined HH1-10 network. Ancestral trait reconstructions were then performed independently using each calibrated network.

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 each accession sequenced on the MiSeq and HiSeq, respectively. MiSeq was run at 2x300nt and, thus, recovered an additional 100nt per read pair than HiSeq, which was run at 2x250nt. Nonetheless, even multiplexing 24 samples per HiSeq2500 (RapidRun) lane as compared to 12 samples per MiSeq lane, HiSeq still recovered an average of 3.4 Gb more sequence per sample. If all reads were on-target the average MiSeq and HiSeq coverage per nucleotide in the target would 5,000-fold to 12,000-fold, respectively. Even allowing for the fact that about half of the sequences are off-target, we expected all targeted genes to exceed 1000x coverage.

Following initial read mapping to the 380 gene bait set in HybPiper, we removed 45 targets and split some of the other targets into more than one discrete paralog. The resulting data set included 412 genes, some of which seemed to cluster sequences by platform (MiSeq vs. HiSeq) rather than species. We infer that the increased coverage of sequencing in using HiSeq resulted in additional paralogous sequences being assembled for some HiSeq-based accessions, resulting in erroneous clustering of these accessions. Using the topological constraints described in the methods, we identified forty loci whose gene trees might have been distorted by a HiSeq-MiSeq artifact, which were subsequently removed, generating our primary HybPiper dataset of 372 genes. These 372 genes were from 241 original targets which yielded single copy assemblies and 55 original targets yielded multiple paralogs. Of the latter, 53 were split into two paralogs (in five cases we did not retain both copies), five original targets were split into three paralogs, and one target was split into each of four, five, and six paralogs. In total 131 paralogous alignments were included in the 372-gene HybPiper data set.

After dropping a further 28 targets, as described in Methods, 344 modified targets were used to call haplotypes with HapHunt. Consistent with baobabs being outcrossing (Baum, 1995a;

Venter et al. 2017), haplotype recovery for the diploid *Adansonia* taxa resulted in an average of 73.5% loci inferred to be heterozygous (i.e., with two variant sequences per accession). 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 (Fig. S2).

Population Tree Inference

We generated population (=species) trees using three approaches: BUCKy and ASTRAL, which use gene trees, and SVDQuartets, which uses the concatenated data. Results from all approaches were in agreement with the exception of the SVDQuartets tree from the HybPiper dataset (Fig. S3a), which placed *A. rubrostipa* sister to the Brevitubae taxa (PP=59) and placed *A. perrieri* as sister to the rest of the core Longitubae (PP=93). Using the other methods, the optimal population trees inferred for the HybPiper and HapHunt datasets were identical except in their resolution of the basal node within *Adansonia* (Fig. 2; Fig. S3). Whereas the HybPiper data placed *A. gregorii* sister to the rest of the genus (Posterior Probability=PP=0.50, CF=0.19), the HapHunt data placed it sister to *A. digitata* (PP=1, CF=0.37). In other regards, the optimal population tree is robust to assembly method. Furthermore, the topology matches the concordance tree constructed by BUCKy. The concordance factors are similar for the two data sets, but generally higher for the HapHunt data. The most dramatic difference relates to the monophyly of the Malagasy clade whose CF is 0.50 for the HapHunt data, as contrasted with CF=0.27 for the HybPiper data set.

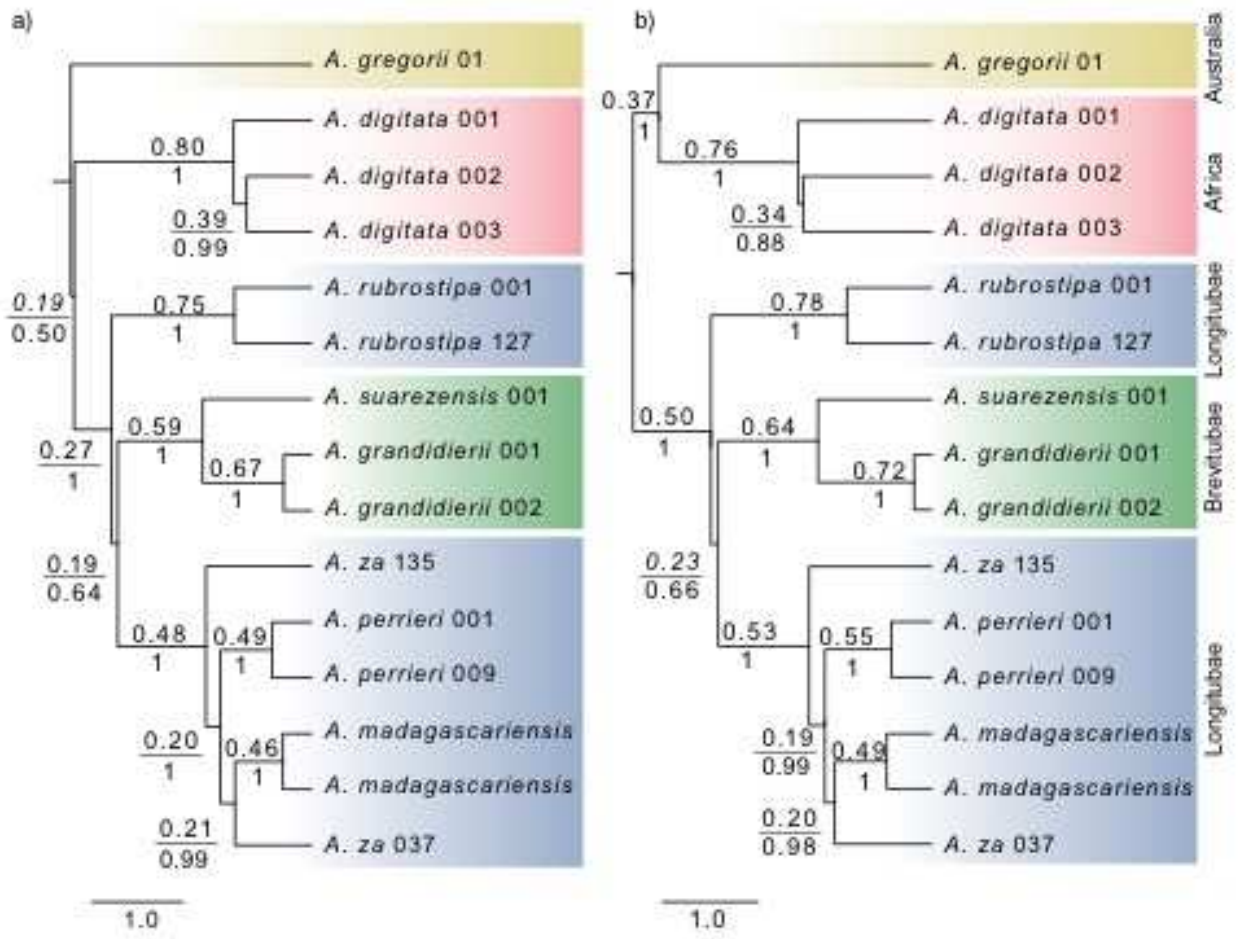


Figure 2. Population trees from (a) the primary HybPiper dataset with 372 genes and (b) the combined HapHunt dataset with 344 genes. Outgroups not shown. BUCKy and ASTRAL yielded the same tree topologies. BUCKy concordance factors added above branches (numbers in italics are *not* significantly higher than at least one conflicting clade); ASTRAL posterior probabilities added below. Trees scaled with branch lengths in coalescent units (from BUCKy).

On the optimal tree for both datasets, the Malagasy Longitubae do not form a clade. Instead, *A. rubrostipa* is sister to a clade including Brevitubae plus core Longitubae (*A. madagascariensis*, *A. perrieri*, and *A. za*). Although the Brevitubae-core Longitubae clade has a concordance factor of only 0.19 based on the HybPiper data, the corresponding credibility

interval (0.169 - 0.210) does not include the CF of a four-species Longitubae clade (0.116 - 0.151). In contrast, the Brevitubae-core Longitubae clade in the summed HapHunt dataset has a CF=0.23 whose credibility interval (0.198 - 0.253) does overlap that of 4-taxon Longitubae clade (0.151 - 0.206). The Brevitubae plus core Longitubae clade is supported in ASTRAL analyses (posterior probability of 0.64 and 0.66, respectively).

The consistent signal contradicting the monophyly of Malagasy Longitubae is surprising, given that the four species have very similar, elongated red-yellow flowers, which are quite different from the shorter, white flowers of Brevitubae and the outgroups. As a result, these data suggest either homoplasy of floral traits or that genes have discordant histories, whether due to ILS or introgression or both.

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

Phylogenetic Network Inference

Based on the slope heuristic, SNaQ analyses favored h1 networks with a single reticulation edge for all data sets, with γ, the proportion of genes inferred to have followed this edge, ranging from 7-23%. It is not surprising that the h1 network is supported since, when we allowed 2 or more reticulation edges, higher order reticulation events had estimates of $\gamma < 2\%$. It

475 should be noted, however, that SNaQ is constrained to ignore intersecting hybridization
476 scenarios, meaning that once a preferred introgression edge is added, many other potential edges
477 are not able to be recovered in the search (Solís-Lemus and Ané 2017).

478 With all taxa included, the HybPiper data supported introgression from the African
479 lineage (the stem lineage of *A. digitata*) into the stem lineage of Brevitubae, with $\gamma = 12.4\%$ (Fig.
480 3a). This hybridization edge was recovered in 73% of the bootstrap replicates. Reducing this data
481 set to those genes that were used for haplotype inference, yielded the same network (as in Fig.
482 3a), with the hybridization edge being found in 95% of bootstrap replicates and $\gamma = 8\%$.
483

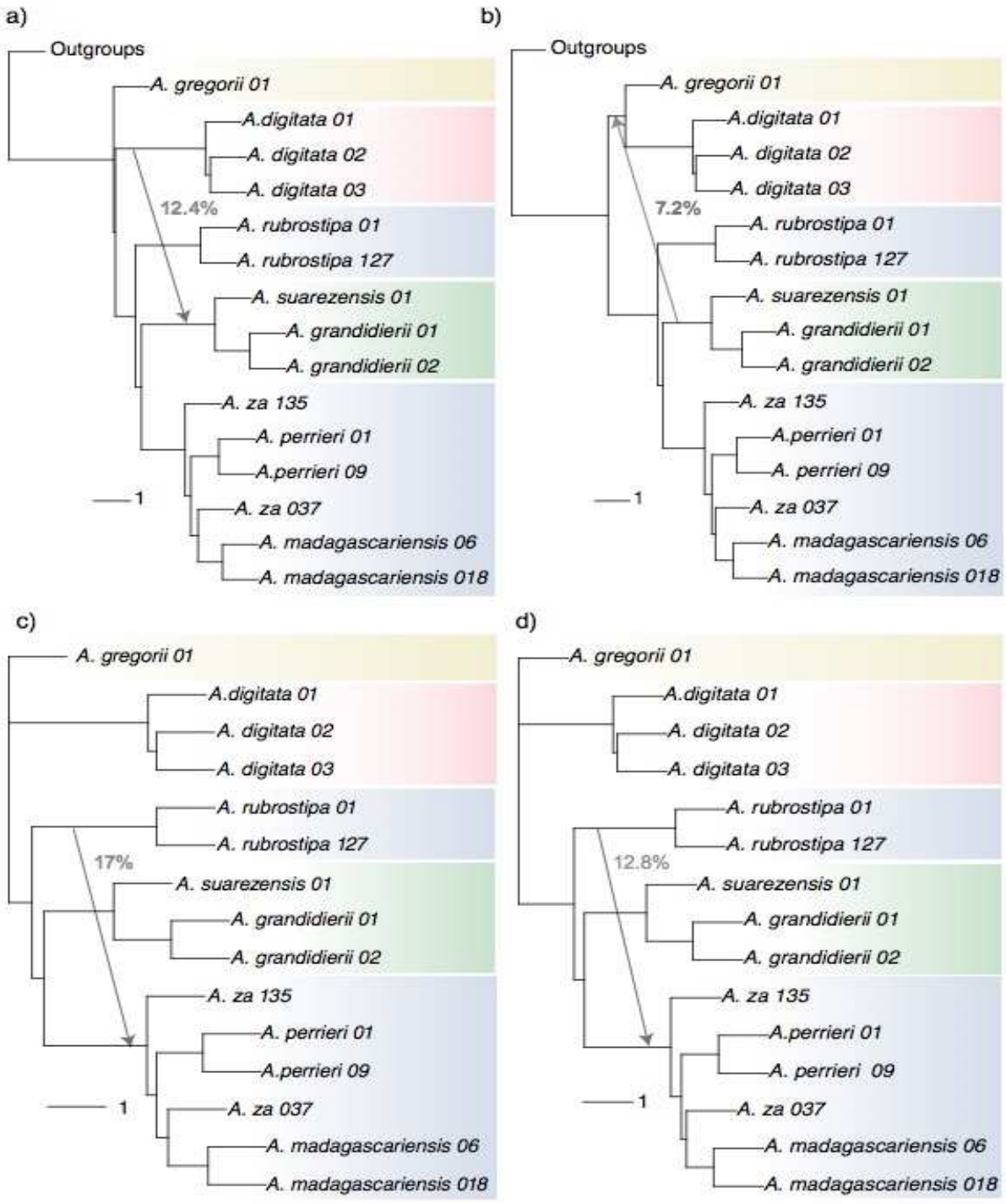


Figure 3. Phylogenetic networks as inferred by SNaQ from the HybPiper dataset (left) and HapHunt dataset (right), with all taxa included (a & b) or after deleting outgroups (c & d). Branch lengths scaled in coalescent units and introgression fractions, γ, shown in grey.

488

489 The HapHunt dataset with all taxa included also identified (in 70% of bootstrap
490 replicates) an introgression between Brevitubae and a non-Malagasy lineage (Fig. 3b). However,
491 the direction of introgression is reversed, suggesting introgression of 7.2% of genes from
492 Brevitubae into a common ancestor of *A. digitata* and *A. gregorii*. Interpreted literally, this
493 scenario is extremely unlikely for geographic and temporal reasons.
494 Given the difficulty of resolving the deep splits and the fact that directionality sometimes
495 cannot be determined reliably with SNaQ, (Solís-Lemus and Ané 2017), we explored the
496 sensitivity of these results to taxon sampling. In the process we discovered that deletion of the
497 outgroups suggested a different reticulation history. For both the HybPiper and HapHunt data
498 sets, the optimal networks after pruning outgroups implies introgression (γ = 13-17%) between
499 *A. rubrostipa* and the stem lineage of the core Malagasy Longitubae clade (Fig. 3c,d).
500 Given the impossibility of inferring intersecting hybridization cycles with SNaQ, we
501 sought to test for additional introgression events after deleting selected taxa (Fig. S5). Including
502 an outgroup but deleting Brevitubae taxa from the HybPiper data set supported gene flow from
503 *A. rubrostipa* to the southern *A. za* sample, with γ = 19.5% (Fig. S5d). In contrast, deleting *A.*
504 *rubrostipa* supported reticulation between *A. digitata* or *A. digitata*+*A.gregorii* and Brevitubae (γ
505 = 11-15%; Fig. S5a,c), as well as additional gene flow between northern and southern *A. za*
506 accessions (γ = 42%, Fig. S5b). Taken together the nuclear data provides evidence of
507 introgression between *A. rubrostipa* and core Longitubae or from an African lineage into
508 Brevitubae, plus the possibility of additional gene flow between accessions of *A. za*.
509 When analyzed each of the ten individual HapHunt datasets, six yielded results consistent
510 with the combined HapHunt dataset (Fig. S4) only differing in the hybridization fraction, γ,

511 which ranged from 7.2 - 8.5%. Of the remaining four, two indicated a hybrid edge between *A.*
512 *digitata* and one member of Brevitubae, *A. suarezensis,* similar to the combined HapHunt dataset
513 when *A. gregorii* was pruned (no shown). One out of the ten HapHunt datasets yielded a
514 network similar to the one obtained without outgroups, which involves gene flow from *A.*
515 *rubrostipa* to the core Longitubae clade (γ=13.2%; Fig. S4, I).
516
517 *D-statistic (ABBA-BABA)*
518
519 Depending on taxon-sampling, SNaQ yielded one of two reticulation edges: between *A.*
520 *digitata* and Brevitubae, or between *A. rubrostipa* and Longitubae. On an unrooted network (Fig.
521 4), these two scenarios are quite close and form intersecting loops. This means that SNaQ, can
522 never yield support for both occurring on the same network, leaving us unsure if one edge is true,
523 with the other being an artifact of some kind, or if both occurred.

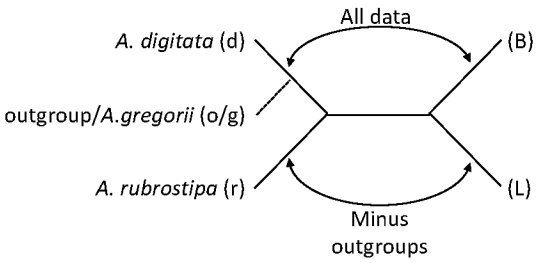


Figure 4. Scenarios for intersecting cycles given an unrooted phylogenetic network for all accessions (top) or all accessions minus outgroups (bottom).

To evaluate these two hypotheses separately we conducted ABBA-BABA tests targeted at either of the two hybridizations individually. To test the *A. digitata*-Brevitubae edge, we dropped *A. rubrostipa* and looked for polymorphisms supporting each partition of the four remaining groups, B, L, d, and o/g (Fig. 5). Likewise, we analyzed the BLro and BLrg datasets to test for gene flow between *A. rubrostipa* and core Longitubae. Support for the reticulation events would

534 come from an excess of BABA over ABBA sites in BLdo/BLdg and the reverse under

535 BLro/BLrg.

536

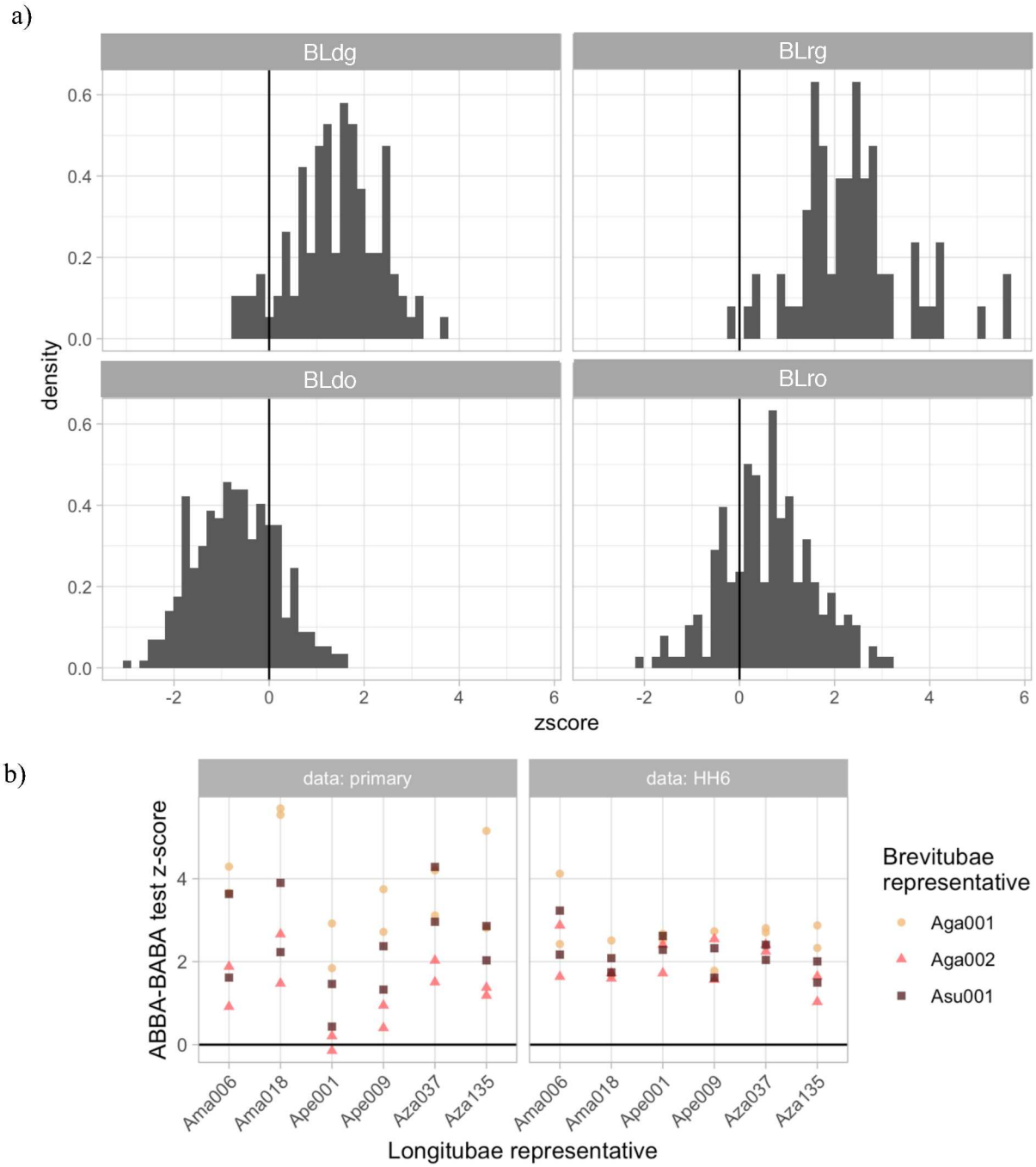


Figure 5. Summary of ABBA-BABA tests. Effect of quartet on the distribution of Z-scores (a). Taxon selection and resulting Z-scores for BLrg quartet (b). See Supplemental Data for details.

537

538 As summarized in Figure 5a, summing over both the HybPiper and HapHunt data sets,
539 only BLrg quartets yielded statistically consistent deviations from null expectations, supporting
540 positive values of D and, consequently indicating introgression between core Longitubae and *A.*
541 *rubrostipa*. The support for this conclusion is especially consistent for the HapHunt data set,
542 which yields a positive Z-score, regardless of which set of 4 taxa are included in the test (Fig.
543 5b).

544

545 *Plastid-Nuclear Incongruence Suggests Additional Reticulation Events Within Madagascar*

546

547 Reference-guided assembly yielded nearly-complete plastid genomes for each accession
548 (number of reads assembled and read coverage is reported in Supplementary Table S2), with a
549 shared alignment of 163,590 bp, containing 167- 4,795 pairwise SNPs among taxa. Although it is
550 generally assumed that the whole plastome has a single phylogenetic history, we first used MDL
551 (Ané 2011) to identify possible recombination breakpoints. This recovered four partitions,
552 representing 39.2, 21.7, 13.1, and 89.6 kilobases, respectively. All but the third partition (13.1
553 kb) supported a combined Australian and African clade (as seen in the HapHunt data set).
554 Likewise, three partitions supported *A. suarezensis* as sister to the remaining Malagasy species,
555 while partition two (21.7 kb) was unable or reject this arrangement based on bootstrap and
556 posterior probabilities (Fig. S6). As the conflicts among the plastid partitions are relatively minor
557 and plastid recombination is unlikely, we used the concatenated plastid tree to represent the
558 plastid history (Fig. 6).

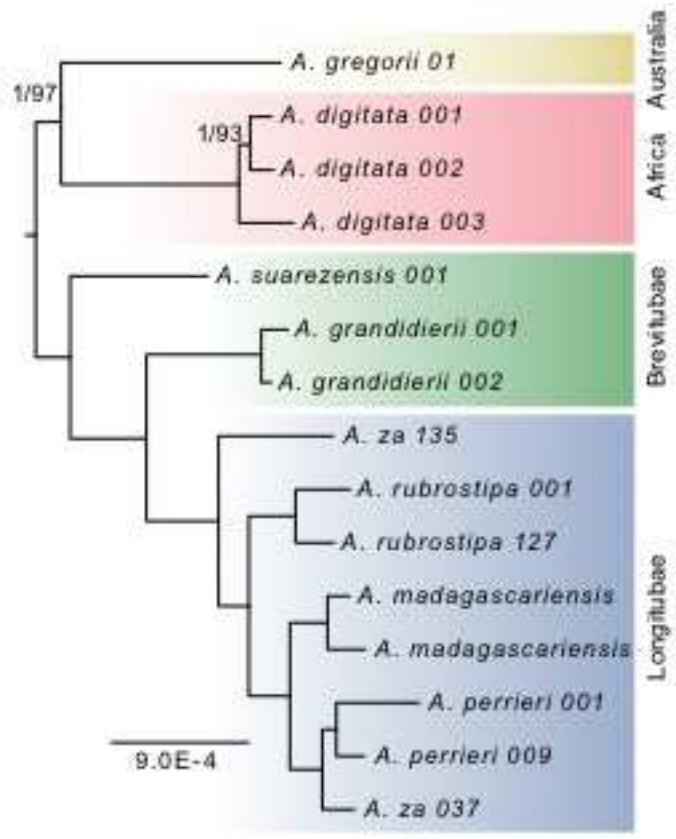


Figure 6: Concatenated plastid tree inferred by maximum likelihood and Bayesian phylogenetic inference. All branches have posterior probabilities (PP) of 1.0 and bootstrap support (BS) of 100%, unless otherwise indicated (PP/BS). Outgroups removed from figure.

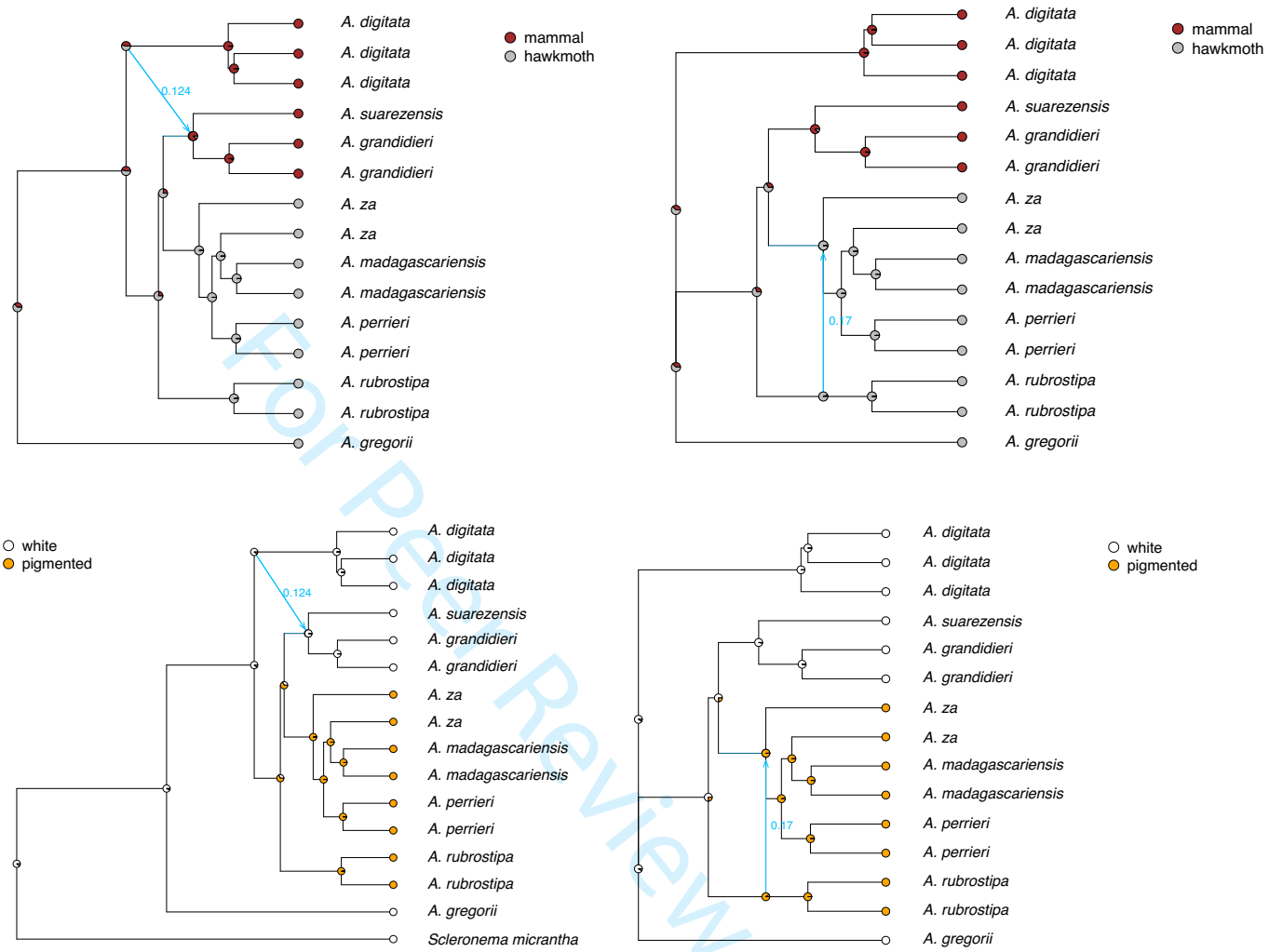
The plastid data strongly support a clade composed of the African and Australian species. The tree also implies non-monophyly of Brevitubae, with *A. suarezensis* strongly supported as sister to the rest of the Malagasy clade. It is worth noting that the plastid marker analyzed by Baum et al. (1998), *rpl16*, also supported a clade composed of all Malagasy species except *A. suarezensis*. This is surprising given the many morphological similarities shared between *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). The plastid tree is also discordant with the nuclear population tree in supporting the monophyly of the four Malagasy Longitubae (*A. rubrostipa*, *A. za*, *A. perrieri*, and *A. madagascariensis*).

To evaluate whether nuclear-plastid discordance can be explained by ILS on the nuclear-derived network, we simulated 100,000 plastid trees on the optimal networks for the 372-gene

582 HybPiper dataset, either with all taxa included (network with *A. digitata*-Brevitubae
583 introgression) or with outgroups excluded (network with *A. rubrostipa*-core Longitubae
584 introgression). Only one simulated tree (out of 100,000) supported *A. suarezensis* as sister to the
585 rest of the Malagasy clade when using the *A. digitata*-Brevitubae introgression network, and four
586 when using the *A. rubrostipa*-core Longitubae introgression network. Despite suggesting rarity
587 of this topology, it is noteworthy that a clade of all Malagasy species except *A. suarezensis* was
588 found in nine maximum likelihood gene trees, out of the 372 genes (2.4%) in the primary data
589 set, with three of these nine trees placing *A. suarezensis* sister to the rest of the Malagasy taxa.
590 The plastid tree also differs from the nuclear population tree in supporting southern *A. za*
591 as sister to all Malagasy Longitubae, including *A. rubrostipa*. None of the simulated trees under
592 the *A. digitata*-Brevitubae introgression have this resolution, but this topology is found in four
593 trees simulated under the *A. rubrostipa*-core Longitubae introgression network. Similarly, among
594 the 372 individual maximum likelihood gene trees inferred from the primary dataset, four (1%)
595 had all Malagasy Longitubae (including *A. rubrostipa*) monophyletic sister to southern *A. za*.
596 While there is discordance under both networks, the plastid data is easier to reconcile with *A.*
597 *rubrostipa* - core Longitubae introgression than *A. digitata*-Brevitubae introgression.
598
599 *Ancestral Trait Reconstruction: Introgression Explains the Shift in Pollination Syndrome in*
600 Brevitubae
601
602 Given that the estimated population tree shows non-monophyly of Malagasy Longitubae,
603 which includes all the colored, hawkmoth pollinated species, we sought to explore the possibility
604 that reticulation edges could help explain the evolution of floral morphology and pollination

systems in *Adansonia*. We conducted ancestral state reconstruction of flower color and inferred pollination mode on time-calibrated networks. Time-calibrated networks from the HapHunt data featured many collapsed internal branches, reflecting temporally improbable introgression edges, especially that representing gene flow between the common ancestor of *A. gregorii* and *A. digitata* and Brevitubae.

For each model and trait, AIC favored equal rates of gain vs. loss and, with these parameters, there was consistent support for introgressed genetic material having played a role in floral evolution (Fig. 7; Table S3). A Bayes factor (Bf) over 1 indicates positive evidence that trait was acquired from the minor rather than from the major parent at a reticulation node. By this criterion, focusing on the HybPiper networks, there is strong evidence of flower-color introgression, whether it be of pigment from *A. rubrostipa* to Longitubae on the network without outgroups (Bf = 7.4) or of non-pigment from *A. digitata* to Brevitubae on the network with outgroups (Bf = 8.0). There is also some support, with a Bayes factors of 2.4-3.0, for introgression introducing hawkmoth pollination into core Longitubae (no outgroups) or mammal pollination into Brevitubae (with outgroups). In either case, these data illustrate the potential for introgressive hybridization to transfer ecologically important traits between lineages.



621

622 Figure 7. Ancestral state reconstruction of flower color (top) and pollination system (bottom) on
623 time-calibrated networks inferred from the HybPiper nuclear dataset. Analyses on the left were
624 based on networks with outgroups included, although these outgroups were excluded from
625 analysis of pollination mode because *Scleronema* has a mixed pollination system (see Methods).

626

627 **DISCUSSION**

628

629 *Inference of Reticulation with Hyb-Seq data*

630

631 Our data support a history of reticulation in *Adansonia* but, despite using a custom set of
632 baits for 380 nuclear loci, the detailed reticulation history proved sensitive to assembly method
633 and taxon inclusion. Recently, Lambert et al. (2019) also found that detection of introgression
634 was sensitive to sequencing batch. Deeper sequencing, as achieved in our data set with the HiSeq
635 platform used on some accessions, has been shown to improve detection of hidden paralogy
636 (Philippe et al. 2011). Given this, we suspect that using a mixture of accessions sequenced with
637 different platforms resulted in different paralogs being assembled for some accessions. To
638 overcome these difficulties, we developed a thorough but laborious procedure of iterative
639 assembly and tree examination to arrive at a conservative, manually curated data set.

640 Even after careful curation, conflicting reticulation histories were suggested depending
641 on whether we did or did not include outgroups: analysis of all taxa supports a Brevitubae - *A.*
642 *digitata* hybridization network, suggesting gene flow from Africa to Madagascar, whereas
643 exclusion of the outgroups supports an *A. rubrostipa* – core Longitubae network, and hence
644 introgression within Madagascar. Based on ABBA-BABA tests, concordance with the plastid
645 data, and geographical proximity, we believe that exclusion of outgroups yields a more plausible
646 result. This highlights the need to carefully examine alternative assemblies and taxon inclusion
647 sets when using Hyb-Seq data for network inference.

648

649 *Biogeography of Adansonia*

650

651

In this study, despite obtaining hundreds of low-copy nuclear loci, relationships among the three primary geographic lineages of *Adansonia* could not be resolved. This begs the question as to the geographic origins of the genus. Given that the sister lineage to *Adansonia* is Neotropical (Carvalho et al. 2016), it was previously suggested that the stem lineage of *Adansonia* migrated across the Atlantic to Africa (Baum, 1998), perhaps as part of the Boreotropical migration route that was open during periods of warmer climate in the Eocene (Tiffney 1985). This hypothesis is supported by the presence of Bombacoid pollen fossils in North America (Wolfe 1975) and Europe (Krutzsch 1989). However, the alternative of trans-Pacific dispersal into Australia and subsequently along the Indian Ocean rim needs also to be considered given the presence of bombacoid pollen fossils in Antarctica during the Eocene (Pross et al., 2012). Whereas the trans-Atlantic hypothesis is compatible with any resolution of the three geographic lineages of *Adansonia*, the trans-Pacific hypothesis predicts, by parsimony, that the African and Malagasy taxa would be sister to one another.

The nuclear data provide do not resolve the basal relationships of *Adansonia*: branch lengths are short and different assembly methods favor (albeit very weakly) different resolutions. The plastid data give strong support for an Africa+Australia clade, thus contradicting trans-Pacific dispersal. While one might attribute this to incomplete lineage sorting on the short internal branches of the inferred population trees, such a scenario would imply that multiple lineages carrying divergent plastid haplotypes dispersed across the Indian Ocean to Africa/Madagascar, which seems unlikely. Instead it seems more probable that there was an almost simultaneous divergence into the three extant lineages somewhere in northwest Africa or the Middle East. Nonetheless, pending datasets with additional *A. gregorii* accessions and fewer

paralogy-calling uncertainties, we currently consider the biogeographic history of baobabs to be unresolved.

Introgression between the African lineage and Brevitubae by Overwater Dispersal

While less certain than the *A. rubrostipa* – core Longitubae introgression, our analyses suggests the possibility of gene flow from an ancestor of *A. digitata*, presumably living in continental Africa, into an ancestor of Malagasy Brevitubae. While this conclusion could be an artifact driven by inclusion of relatively distant outgroups, and is not strongly supported by ABBA-BABA tests, it is certainly plausible that there was more than one dispersal event of baobabs to Madagascar. Many lineages endemic to Madagascar are of African origin (Haber et al. 2017; Yoder et al. 1996) and multiple long-distance dispersal events to Madagascar within a single lineage are known (Kainulainen et al. 2017).

Networks that support *A. digitata* – Brevitubae introgression suggest that perhaps 10% of the genome of extant Brevitubae descended from *A. digitata*. This suggests that a single tree established in Madagascar, became reproductively mature, crossed with a local population and generated hybrid genotypes that were sufficiently favored (perhaps by mammal pollinators) that a significant fraction of the recipient species' genome was replaced. Alternatively, a population of *A. digitata* could have established in Madagascar before hybridizing with resident species and then was extirpated. While there are few clear cases of transoceanic dispersal 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

696 the New World followed by hybridization with an indigenous American D-genome diploid
697 (Wendel and Grover 2015, and references therein).

698 If introgression occurred from the African lineage (*A. digitata)* into the Malagasy clade, it
699 must have been quite ancient. It presumably predated the origin of tetraploidy in *A. digitata* and
700 the transition of the Malagasy Brevitubae to dry season flowering or its divergence into its two
701 extant species (*A. grandidieri* and *A. suarezensis*). If the two lineages occurred in sympatry, gene
702 exchange is plausible despite the main pollinators being different; based on extant species,
703 nocturnal primates and hawkmoths would occasionally visit both flower types (Baum 1995a). To
704 date, no artificial crossing studies have been done testing interfertility.

705

706 *Ancient Introgression between A. rubrostipa and Longitubae*

707

708 Plastid and nuclear discordance suggests a history of gene flow between *A. rubrostipa*
709 and core Longitubae. This conclusion is supported by SNaQ analyses that exclude outgroups and
710 gains strong statistical support from ABBA-BABA tests. Given that *A. rubrostipa* and the other
711 extant Longitubae share similar flower morphologies and pollination systems, and also have
712 widely overlapping ranges and some potential for occasional flowering season overlap, such
713 hybridization is plausible. Indeed, there are reasons to infer that there was not just one ancient
714 hybridization event but several instances of gene flow between *A. rubrostipa* and core
715 Longitubae both before and after the divergence of the latter into its three extant species. Such
716 gene flow is in agreement with analyses based on nuclear microsatellite data (Leong Pock Tsy et
717 al. 2013). Furthermore, a specific *A. rubrostipa* - *A. za* hybridization was identified in a h1
718 search after pruning Brevitubae taxa, though with lower likelihood than the corresponding h1 *A.*

digitata - Brevitubae network. Due to intersecting cycles (shared branches) an *A. rubrostipa* - *A. za* reticulation edge cannot be found by SNaQ on an h2 network with either an *A. digitata* – Brevitubae or an *A. rubrostipa* – core Longitubae edge. Reticulation between *A. rubrostipa* and *A. za* was recovered, however, when used as the input network for the starting search. This also agrees with our unpublished data documenting additional reticulation events among the Malagasy Longitubae, all of which are wet-season flowering and primarily hawkmoth pollinated.

Plastid-Nuclear Tree and the Plastid Non-Monophyly of Brevitubae

The primary population tree derived from nuclear genes conflicts markedly with the plastid tree in the placement of two species, *A. suarezensis* and *A. rubrostipa*. Cases of plastid-nuclear tree discordance are commonly attributed to "chloroplast capture" (Rieseberg and Soltis 1991; Tsitrone et al. 2003; Feliner et al. 2017), which is to say introgression affecting the plastid but not (much of) the nuclear genome. Since we reconstructed an explicit network from the nuclear data, rather than just a single tree, we could assess whether the same hybridization history, combined with incomplete lineage sorting, could explain plastid-nuclear discordance.

The plastid data provides strong support (100% bootstrap and 1.0 PP) for the placement of *A. suarezensis* as sister to the rest of the Malagasy baobab clade. Simulations suggest that this resolution is unlikely, with a probability of $<4 \times 10^{-5}$, given neutral evolution and ILS along the nuclear-inferred networks. Nonetheless, the nuclear genes themselves suggest that such a signal also exists in the nuclear genome, with 2% of nuclear gene trees placing *A. suarezensis* as sister to the rest of the Malagasy clade. This hints at a possible undetected minor introgression event, an "introgressive kiss," either involving the *A. grandidieri* lineage and the stem lineage of the

742 Malagasy Longitubae or between *A. suarezensis* and a now extinct sister lineage to the entire
743 Malagasy clade. While such additional hybridizations seem relatively unlikely given the long
744 stem lineage of Brevitubae (i.e., *A. suarezensis* and *A. grandidieri*) and their synapomorphy of
745 winter-flowering, it cannot be ruled out.

746 The plastid tree is also distinct in embedding *A. rubrostipa* within the Malagasy
747 Longitubae and placing southern accessions of *A. za* sister to the remainder of the Longitubae
748 clade. This topology has a low probability of arising simply by ILS, especially on networks with
749 a Brevitubae-*A. digitata* reticulation. Consequently, this result further supports the possibility of
750 additional undetected reticulation events within the Longitubae clade.

751

752 *Adaptive Introgression of Pollination Traits*

753

754 We developed methods for time-calibrating networks, inferring the probabilities of
755 alternative character states at ancestral nodes, and then estimating the posterior probability that a
756 trait was acquired from the minor or major parent during hybridization. These methods should
757 have broad applicability for studying character evolution in the context of phylogenetic
758 networks. As an example of their application we used ancestral trait reconstruction analysis to
759 explore whether adaptive introgression of floral pigment and pollination mode might explain the
760 implied homoplasy associated with these traits when mapped onto the dominant population tree.

761 Considering, first, the best supported network, namely that entailing introgression from
762 *A. rubrostipa* into a common ancestor of the three species of core Longitubae, we found strong
763 support for the hypothesis that introgression transferred the shared floral traits of a red style, red
764 interior calyx and colored petals from the *A. rubrostipa* lineage into Longitubae. Furthermore, as

765 shown by analysis of pollination mode evolution on this network, a consequence of this
766 introgression was likely to include a shift towards hawkmoth pollination. Indeed, it is plausible
767 that adaptive introgression of a suite of floral traits, including not just pigmentation but staminal
768 tube length, nectar composition, and scent, could have been driven by them promoting the
769 frequency and efficiency of pollinator visits by hawkmoths.

770 Although less well-supported by our data, the *A. digitata* – Brevitubae gene flow
771 hypothesis also supports a role for introgression in floral homoplasy. In this case, the highest
772 likelihood reconstructions imply an introgression-mediated reversal in Brevitubae to an all-white
773 flower and a concomitant switch to mammal pollination. Although not formally analyzed here,
774 we presume that introgression would also explain similarities between the donor and recipient
775 taxa in shared traits such as nectar volume, nectar chemistry, and scent profile. On the other
776 hand, to explain some striking differences in flower form between Brevitubae and *A. digitata*,
777 including the latter's long-pendulous flowers and extremely wide, reflexed petals, one could
778 either suppose that these traits evolved in *A. digitata* after the introgression event or that these
779 traits we present but failed to introgress.

780 While examples of introgression facilitating pollinator shifts have been reported (i.e.
781 Louisiana irises, Wesselingh 2006; Monkeyflowers, Stankowski and Streisfeld 2015), this is the
782 first case we are aware of that involves transitions between hawkmoth and mammal pollination.
783 In the future it would be exciting to look more broadly at the genomes of all *Adansonia* species
784 in the hopes of identifying candidate genes for various traits involved in the observed pollination
785 syndromes, such as flower color, nectar characteristics or floral scent chemistry. Such genome-
786 scale work would not only solidify the history of ancient introgression in the baobabs, but could

provide a model for understanding mechanisms involved in introgressive pollination system evolution.

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 for 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, George Brown, and Diana Mayne, and the Darwin Botanic Gardens kindly provided 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

Andriafidison D., Andrianaivoarivelo R.A., Ramilijaona O.R., Razanahoera M.R., MacKinnon J., Jenkins R.K., Racey P.A. 2006. Nectarivory by endemic Malagasy fruit bats during the dry season 1. *Biotropica.* 38:85-90.

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

Ané C., Larget B., Baum D.A., Smith S.D., Rokas A. 2006. Bayesian estimation of concordance among gene trees. *Mol. Biol. Evo.* 24:412-426.

Arnold M.L., Kunte K. 2017. Adaptive genetic exchange: a tangled history of admixture and evolutionary innovation. *Trends Ecol. Evol.* 32:601–611.

Arnold M.L. 2004. Transfer and origin of adaptations through natural hybridization: were Anderson and Stebbins right? *Plant Cell.* 16:562–570.

Arnold M.L., 1992. Natural hybridization as an evolutionary process. *Annu. Rev. Ecol. Syst.* 23:237-261.

Bastide P., Solis-Lemus C., Kriebel R., Sparks W.K., Ané C. 2018. Phylogenetic comparative methods on phylogenetic networks with reticulations. *Sys. Bio.* 67:800-820.

Baum D.A. 1995a. The comparative pollination and floral biology of baobabs (*Adansonia*-Bombacaceae). *Ann. Mo. Bot. Gard.* 82:322-348.

833 Baum D.A. 1995b. A systematic revision of *Adansonia* (Bombacaceae). *Ann. Mo. Bot. Gard.*
834 82:440-471.
835
836 Baum D.A., DeWitt Smith S., Yen A., Alverson W.S., Nyffeler R., Whitlock B.A., Oldham R.L.
837 2004. Phylogenetic relationships of Malvatheca (Bombacoideae and Malvoideae; Malvaceae
838 sensu lato) as inferred from plastid DNA sequences. *Am. J. Bot.* 91:1863-1871.
839
840 Baum D.A., Small R.L., Wendel J.F. 1998. Biogeography and floral evolution of Baobabs
841 Adansonia, Bombacaceae as inferred from multiple datasets. *Sys. Bio.* 47:181-207.
842
843 Blackmon H., Richard A.A. 2015. EvobiR: Tools for comparative analyses and teaching
844 evolutionary biology. doi:10.5281/zenodo.30938.
845
846 Bolger A. M., Lohse M., Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina
847 Sequence Data. *Bioinformatics*. 170.
848
849 Carvalho-Sobrinho J. G., Alverson W. S., Alcantara S., Queiroz L. P., Mota A. C., Baum D. A.
850 2016. Revisiting the phylogeny of Bombacoideae (Malvaceae): Novel relationships,
851 morphologically cohesive clades, and a new tribal classification based on multilocus
852 phylogenetic analyses. *Mol. Phylogenet. Evol.* 101:56-74.
853
854 Chang S., Puryear J., Cairney J. 1993. A simple and efficient method for isolating RNA from
855 pine trees. *Plant Mol. Bio. Rep.* 11:113-116.

856

857 Chau J.H., Rahfeldt W.A., Olmstead, R.G. 2018. Comparison of taxon-specific versus general
858 locus sets for targeted sequence capture in plant phylogenomics. *Appl. Plant Sci*. 6:1032.

859

860 Chifman J., Kubatko L. 2015. Identifiability of the unrooted species tree topology under the
861 coalescent model with time-reversible substitution processes, site-specific rate variation, and
862 invariable sites. *J. Theor. Biol.* 374:35-47.

863

864 Conover J.L., Karimi N., Stenz N., Ane C., Grover C.E., Skema C., Tate J.A., Wolff K., Logan
865 S.A., Wendel J.F., Baum D.A. 2019. A Malvaceae mystery: A mallow maelstrom of genome
866 multiplications and maybe misleading methods? *J. Integr. Plant Biol.* 61:12–31

867

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

871

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

874

875 Ekblom R., Galindo J. 2011. Applications of next generation sequencing in molecular ecology of
876 non-model organisms. *Heredity*. 107:1.

877

878 Feliner G. N., Álvarez I., Fuertes-Aguilar J., Heuertz M., Marques I., Moharrek F., Piñeiro R.,
879 Riina R., Rosselló J.A., Soltis P.S., Villa-Machío I. 2017. Is homoploid hybrid speciation that
880 rare? An empiricist's view. *Heredity*. 118:513–516.
881
882 Fér T., Schmickl R.E. 2018. HybPhyloMaker: target enrichment data analysis from raw reads to
883 species trees. *Evol. Bioinform.* 14:1176934317742613.
884
885 Grover C.E., Salmon A., Wendel J.F. 2012. Targeted sequence capture as a powerful tool for
886 evolutionary analysis1. *Am. J Bot.* 99:312-319.
887
888 Grover C.E., Gallagher J.P., Jareczek J.J., Page J.T., Udall J.A., Gore M.A., Wendel J.F. 2015.
889 Re-evaluating the phylogeny of allopolyploid Gossypium L. *Mol. Phylogenet. Evol.* 92:45-52.
890
891 Grover C.E., Gallagher J.P., Szadkowski E.P., Page J.T., Gore M.A., Udall J.A., Wendel J.F.
892 2017. Nucleotide diversity in the two co-resident genomes of allopolyploid cotton. *Plant Sys.*
893 *Evol.* 303: 1021
894
895 Haber E.A., Kainulainen K., Van Ee B.W., Oyserman B.O., Berry P.E. 2017. Phylogenetic
896 relationships of a major diversification of *Croton* (Euphorbiaceae) in the western Indian Ocean
897 region. *Bot. J. Linn. Soc.* 183:532-544.
898
899 Hart M.L., Forrest L.L., Nicholls J.A., Kidner C.A. 2016. Retrieval of hundreds of nuclear loci
900 from herbarium specimens. *Taxon*. 65:1081-1092.

Harvey M.G., Smith B.T., Glenn T.C., Faircloth B.C., Brumfield R.T. 2016. Sequence capture versus restriction site associated DNA sequencing for shallow systematics. *Sys. Bio.* 65:910-924.

Huelsenbeck J.P., 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., Wickett N.J. 2016. HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Appl. Plant Sci*. 4:1600016.

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

Kamneva O.K., Syring J., Liston A., Rosenberg N.A. 2017. Evaluating allopolyploid origins in strawberries (*Fragaria*) using haplotypes generated from target capture sequencing. *BMC Evol. Bio*. 17:180.

Kates H.R., Johnson M.G., Gardner E.M., Zerega N.J., Wickett, N.J. 2018. Allele phasing has minimal impact on phylogenetic reconstruction from targeted nuclear gene sequences in a case study of Artocarpus. *Am J Bot*. 105:404-416.

924 Katoh K., Misawa K., Kuma K.I., Miyata T. 2002. MAFFT: a novel method for rapid multiple
925 sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30:3059-3066.
926
927 Kearse M., Moir R., Wilson A., Stones-Havas S., Cheung M., Sturrock S., Buxton S., Cooper A.,
928 Markowitz S., Duran C., Thierer T., Ashton B., Mentjies P., Drummond A. 2012. Geneious
929 Basic: an integrated and extendable desktop software platform for the organization and analysis
930 of sequence data. *Bioinformatics.* 28:1647-1649.
931
932 Krutzsch W. 1989. Paleogeography and historical phytogeography (paleochorology) in the
933 Neophyticum. *Plant Syst. Evol*. 162:5–61.
934
935 Kubatko L.S., Carstens B.C., Knowles L.L. 2009. STEM: species tree estimation using
936 maximum likelihood for gene trees under coalescence. *Bioinformatics*. 25:971-973.
937
938 Kulathinal R.J., Stevison L.S., Noor M.A. 2009. The genomics of speciation in Drosophila:
939 diversity, divergence, and introgression estimated using low-coverage genome sequencing. *PLoS*
940 *Genet*. 5:1000550.
941
942 Lambert S.M., Streicher J.W., Fisher-Reid M.C., Méndez de la Cruz F.R., Martínez-Méndez N.,
943 García-Vázquez U.O., Montes de Oca A.N., Wiens J.J. 2019. Inferring introgression using
944 RADseq and D_{FOIL}: Power and pitfalls revealed in a case study of spiny lizards (*Sceloporus*).
945 *Mol. Ecol. Resour.* 19: 18–837.
946

947 Larget B.R., Kotha S.K., Dewey C.N., Ané C. 2010. BUCKy: gene tree/species tree
948 reconciliation with Bayesian concordance analysis. *Bioinformatics*. 26:2910-2911.
949
950 Leong Pock Tsy J.M., Lumaret R., Flaven-Noguier E., Sauve M., Dubois M.P., Danthu P. 2013.
951 Nuclear microsatellite variation in Malagasy baobabs (*Adansonia*, Bombacoideae, Malvaceae)
952 reveals past hybridization and introgression. *Ann. Bot*. 112:1759-1773.
953
954 Li H.,Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform.
955 *Bioinformatics*. 25:1754-1760.
956
957 Mallet J. 2008. Hybridization, ecological races and the nature of species: empirical evidence for
958 the ease of speciation. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 363:2971-2986.
959
960 Martin N.H., Bouck A.C., Arnold M.L. 2006. Detecting adaptive trait introgression between *Iris*
961 *fulva* and *Iris brevicaulis* in highly-selective field conditions. *Genetics*. 172:2481–2489
962
963 Mirarab S., Reaz R., Bayzid M.S., Zimmermann T., Swenson M.S., Warnow T. 2014. ASTRAL:
964 genome-scale coalescent-based species tree estimation. *Bioinformatics*. 30:541-548.
965
966 Nicholls J.A., Pennington R.T., Koenen E.J., Hughes C.E., Hearn J., Bunnefeld L., Dexter K.G.,
967 Stone G.N., Kidner C.A. 2015. Using targeted enrichment of nuclear genes to increase
968 phylogenetic resolution in the neotropical rain forest genus *Inga* (Leguminosae:
969 Mimosoideae). *Front. Plant Sci.* 6:710.

Page J.T., Liechty Z.S., Huynh M.D., Udall J.A. 2014. BamBam: genome sequence analysis tools for biologists. *BMC Res. Notes*. 7:829.

Park H. J., Nakhleh L. 2012. Inference of reticulate evolutionary histories by maximum likelihood: the performance of information criteria. *BMC Bioinformatics*. 13:S12.

Payseur B.A., Rieseberg L.H. 2016. A genomic perspective on hybridization and speciation. *Mol. Eco.* 25: 2337–2360.

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

Philippe H., Brinkmann H., Lavrov D.V., Littlewood D.T.J., Manuel M., Wörheide G., Baurain D., 2011. Resolving difficult phylogenetic questions: why more sequences are not enough. *PLoS Biol.* 9:1000602.

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., Huck C.E. 2012. Persistent near-tropical warmth on the Antarctic continent during the early Eocene epoch. *Nature*. 488:73.

992 Raymond O., Piola F., Sanlaville-Boisson C. 2002. Inference of reticulation in outcrossing
993 allopolyploid taxa: caveats, likelihood and perspectives. *Trends in Ecol. Evol.* 17:3-6.
994
995 Rieseberg L.H., Soltis D.E. 1991. Phylogenetic consequences of cytoplasmic gene flow in plants.
996 *Evolutionary Trends in Plants*. 5.
997
998 Rieseberg L.H., Wendel J.F. 1993. Introgression and its consequences in plants. Hybrid zones
999 and the evolutionary process, 70–109. New York, NY: Oxford University Press.
1000
1001 Ronquist F., Huelsenbeck J.P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed
1002 models. *Bioinformatics*. 19:1572-1574.
1003
1004 Ryckewaert, P., Razanamaro, O., Rasoamanana, E., Rakotoarimihaja, T., Ramavovololona, P.
1005 2011. Les Sphingidae, probables pollinisateurs des baobabs malgaches. Cirad. *Bois et Forêts des*
1006 *Tropiques*. 307:56-68.
1007
1008 Salmon A., Joshua J. A., Jeddeloh J.A., Wendel J.F. 2012. Targeted capture of homoeologous
1009 coding and non-coding sequence in polyploid cotton. *G3: Genes, Genomes, Genetics* 2:921-930
1010
1011 Schwenk K., Brede N., Streit B. 2008. Introduction. Extent, processes and evolutionary impact
1012 of interspecific hybridization in animals. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 363:2805–
1013 2811.
1014

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

Soltis P.S., Soltis D.E. 2009. The role of hybridization in plant speciation. *Annu. Rev. Plant Biol.* 60:561–588.

Suarez-Gonzalez A., Lexer C., Cronk Q.C. 2018. Adaptive introgression: a plant perspective. *Biology Letters.* 14:20170688.

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

Snir S., Rao S. 2012. Quartet MaxCut: a fast algorithm for amalgamating quartet trees. *Mol. Phylogenet. Evol*. 62:1-8.

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

Solís-Lemus C., Bastide P., Ané, C. 2017. PhyloNetworks: a package for phylogenetic networks. *Mol. Bio. Evol.* 34:3292-3298.

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*. 30:1312-1313.

Strimmer K., Wiuf C., Moulton V. 2001. Recombination analysis using directed graphical models. *Mol. Bio. Evol.* 18:97-99.

Struck T.H. 2013. The impact of paralogy on phylogenomic studies–a case study on annelid relationships. *PLoS One*. 8:62892.

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. *J. Arnold Arbor*. 66:73-94.

Tsitrone A., Kirkpatrick M., Levin D.A. 2003. A model for chloroplast capture. *Evolution.* 57:776-1782.

van Dulmen A., 2001. Pollination and phenology of flowers in the canopy of two contrasting rain forest types in Amazonia, Colombia. *Tropical forest canopies: ecology and management*. 3:85.

Villaverde, T., Pokorny L., Olsson S., Rincón-Barrado M., Johnson M.G., Gardner E.M., Wickett N.J., Molero J., Riina R., Sanmartín I. 2018. Bridging the micro- and macroevolutionary levels in phylogenomics: Hyb-Seq solves relationships from populations to species and above. *New Phytol.* 220:636–650.

Weitemier K., Straub S.C., Cronn R.C., Fishbein,M., Schmickl R., McDonnell A., Liston A. 2014. Hyb-Seq: Combining target enrichment and genome skimming for plant phylogenomics. *Appl. Plant Sci*. 9:1400042

Wendel J.F. 2015. The wondrous cycles of polyploidy in plants. *Am. J. Bot.* 102:1753-1756.

Wendel J. F., Cronn R.C. 2003. Polyploidy and the evolutionary history of cotton. *Adv. Agron.* 78:*139–186.*

Wendel J. F., Grover C.E. 2015. Taxonomy and evolution of the cotton genus, Gossypium. Cotton *Agron. Monogr*. 57:25-44.

Wesselingh R.A. 2000. Pollinator behaviour and the evolution of Louisiana iris hybrid zones. *J. Evol. Bio.* 13:171-180.

Wolfe J. A. 1975. Some aspects of plant geography of the northern hemisphere during the Late Cretaceous and Tertiary. *Ann Mo Bot Gard*. 62:264–279.

1083 Wolf P.G., Robison T.A., Johnson M.G., Sundue M.A., Testo W.L., Rothfels C.J. 2018. Target
1084 sequence capture of nuclear-encoded genes for phylogenetic analysis in ferns. *Appl. Plant Sci.*
1085 6:01148.
1086
1087 Yang Y., Smith S.A. 2014. Orthology inference in non-model organisms using transcriptomes
1088 and low-coverage genomes: improving accuracy and matrix occupancy for phylogenomics. *Mol.*
1089 *Bio. Evol.* 11:3081-92.
1090
1091 Yoder A.D., Cartmill M., Ruvolo M., Smith K., Vilgalys R. 1996. Ancient single origin for
1092 Malagasy primates. *Proc. Natl. Acad. Sci.* 93:5122-5126.
1093
1094 Yu Y., Dong J., Liu K.J., Nakhleh L. 2014. Maximum likelihood inference of reticulate
1095 evolutionary histories. *Proc. Natl. Acad. Sci*. 111:16448-16453.
1096
1097 Yu Y., Than C., Degnan J.H., Nakhleh L. 2011. Coalescent histories on phylogenetic networks
1098 and detection of hybridization despite incomplete lineage sorting. *Sys. Bio.* 60:138-149.
1099
1100 Zhang C., Ogilvie H.A., Drummond A.J., Stadler T. 2017. Bayesian inference of species
1101 networks from multilocus sequence data. *Mol. Bio. Evol*. 35:504-517.
1102
1103 Zhang C., Rabiee M., Sayyari E., Mirarab S. 2018. ASTRAL-III: Polynomial Time Species Tree
1104 Reconstruction from Partially Resolved Gene Trees. *BMC Bioinformatics.* 19:153.
1105

1106 Zhu S., Degnan J.H., Goldstien S.J., Eldon B. 2015. Hybrid-Lambda: simulation of multiple
1107 merger and Kingman gene genealogies in species networks and species trees. *BMC*
1108 *Bioinformatics*. 16:292.
1109
1110 Zimmer E.A., Wen J. 2015. Using nuclear gene data for plant phylogenetics: Progress and
1111 prospects II. Next-gen approaches. *J. Sys. Evol.* 53:371-379.
1112
1113 FIGURE CAPTIONS:
1114

1115 Figure 2. Eight species of *Adansonia.* A) *A. digitata*, continental Africa, B) *A. gregorii,*
1116 Australia, C) *A. grandidieri*, Madagascar, D) *A. suarezensis*, Madagascar, E) *A.*
1117 *madagascariensis*, Madagascar, F) *A. perrieri*, Madagascar, G) *A. za*, Madagascar, H) *A.*
1118 *rubrostipa,* Madagascar.

1119
1120 Figure 2. Population trees from (a) the primary HybPiper dataset with 372 genes and (b) the
1121 combined HapHunt dataset with 344 genes. Outgroups not shown. BUCKy and ASTRAL
1122 yielded the same tree topologies. BUCKy concordance factors added above branches (numbers
1123 in italics are *not* significantly higher than at least one conflicting clade); ASTRAL posterior
1124 probabilities added below. Trees scaled with branch lengths in coalescent units (from BUCKy).

1125
1126 Figure 3. Phylogenetic networks as inferred by SNaQ from the HybPiper dataset (left) and
1127 HapHunt dataset (right), with all taxa included (a & b) or after deleting outgroups (c & d).
1128 Branch lengths scaled in coalescent units and introgression fractions, γ, shown in grey.

Figure 4. Scenarios for intersecting cycles given an unrooted phylogenetic network for all accessions (top) or all accessions minus outgroups (bottom).

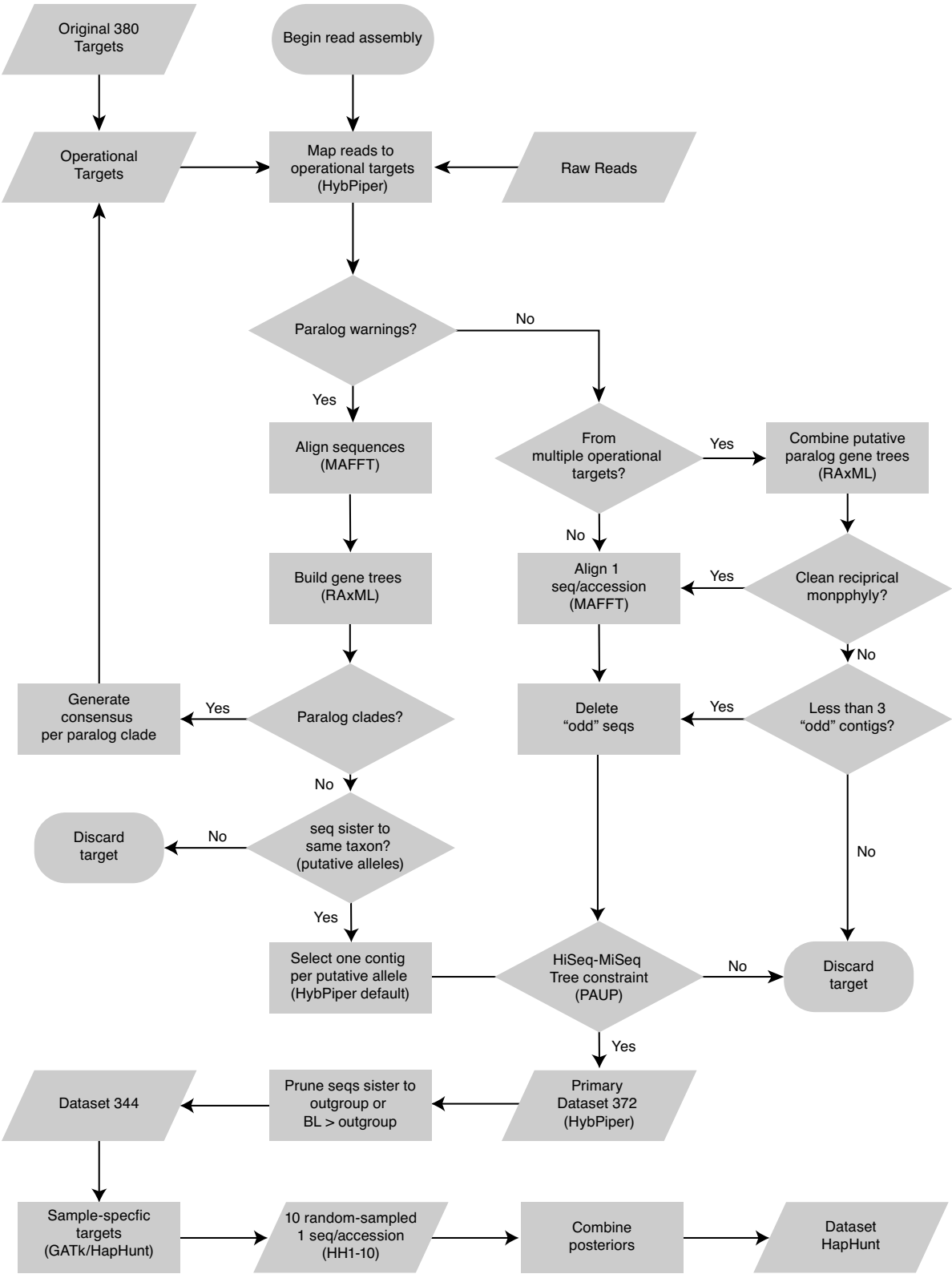
Figure 5. Summary of ABBA-BABA tests. Effect of quartet on the distribution of Z-scores (a). Taxon selection and resulting Z-scores for BLrg quartet (b). See Supplemental Data for details.

Figure 6: 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.

Figure 7. Ancestral state reconstruction of flower color (top) and pollination system (bottom) on time-calibrated networks inferred from the HybPiper nuclear dataset. Analyses on the left were based on networks with outgroups included, although these outgroups were excluded from analysis of pollination mode because *Scleronema* has a mixed pollination system (see Methods).

SUPPLEMENTAL MATERIAL

Figure S1. Read assembly pipeline.

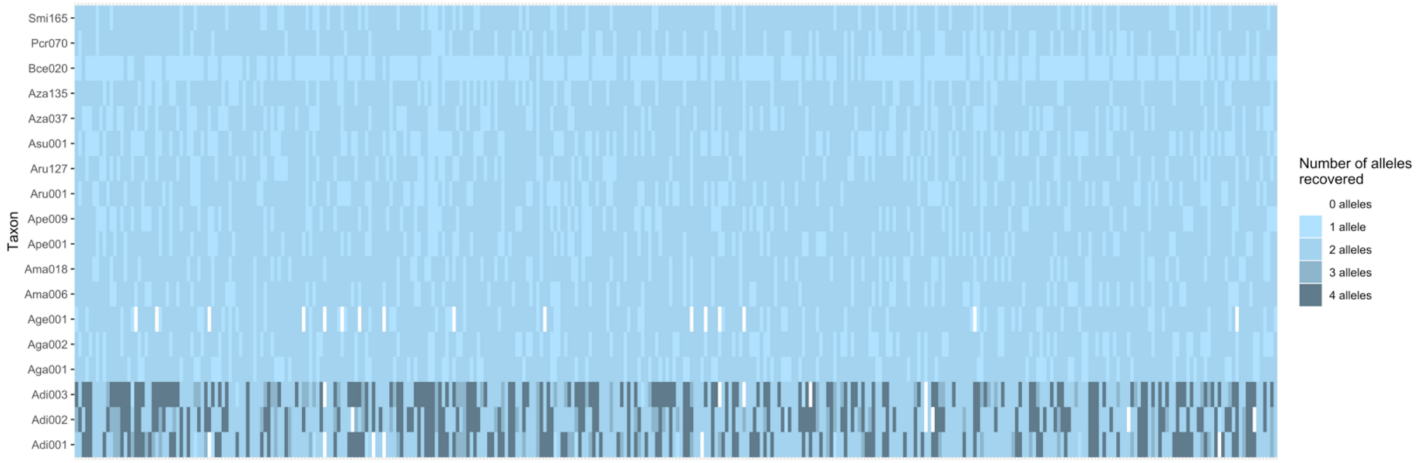

Original 380 Targets
Begin read assembly
Operational Targets
Map reads to operational targets (HybPiper)
Raw Reads
Paralog warnings?
No
Yes
Align sequences (MAFFT)
From multiple operational targets?
Yes
Combine putative paralog gene trees (RAxML)
No
Build gene trees (RAxML)
Align 1 seq/accession (MAFFT)
Yes
Clean reciprical monpphyly?
No
Generate consensus per paralog clade
Yes
Paralog clades?
Delete "odd" seqs
Yes
Less than 3 "odd" contigs?
No
No
Discard target
No
seq sister to same taxon? (putative alleles)
Yes
No
Select one contig per putative allele (HybPiper default)
HiSeq-MiSeq Tree constraint (PAUP)
No
Discard target
Yes
Dataset 344
Prune seqs sister to outgroup or BL > outgroup
Primary Dataset 372 (HybPiper)
Sample-specfic targets (GATk/HapHunt)
10 random-sampled 1 seq/accession (HH1-10)
Combine posteriors
Dataset HapHunt

1147

1148

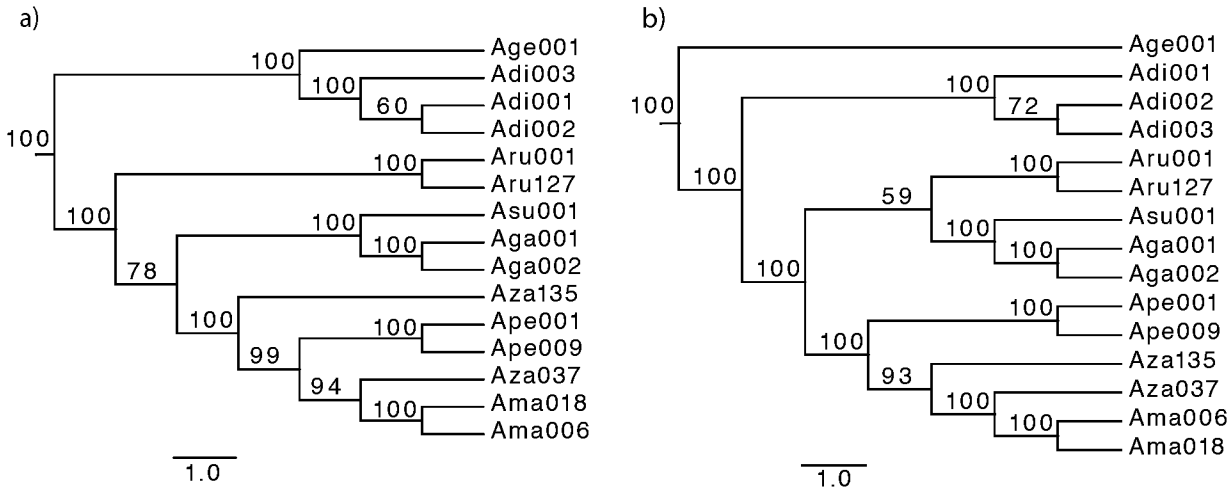
1149

1150 Figure S2. Haplotype recovery heatmap.



1151

1152 Figure S3. SVDQuartets tree from the (a) HybPiper and (b) HapHunt datasets.



1153

1154

1155 Table S1. Taxon sampling.

1156

TABLE S1. Sampling. 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. & Perr.	Aru001	MiSeq	8,890,992	Southwestern Madagascar, D.A.Baum 313 (MO)
Adansonia rubrostipa Jum. & 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

Table S2. Plastid genomes assembly statistics.

Taxon (Sample ID)	**Number of reads assembled**	**Length**	**Read coverage per position Mean, Median (Max)**
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 (Ama006)	723,721	174,559	1068, 1111 (1905)
A. madagascariensis (Ama018)	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)

Table S3: Transition rates were obtained after scaling each network to a unit height from the crown node of the *Adansonia* clade to the tips. In other words, transition rates represent the expected number of transitions along one lineage from the *Adansonia* crown node to the present.

trait	network_data	network_sampling	rates	likelihood	aic	bf	rate1	rate2
pollinator	primary	with_outgroups	unconstrained	-6.8396016	17.6792031	4.62653448	0.84619726	0
pollinator	primary	with_outgroups	equal	-7.1081293	16.2162586	3.00777996	0.48039017	
pollinator	primary	only_ingroup	unconstrained	-6.6229669	17.2459338	1.31188554	0	0.41610137
pollinator	primary	only_ingroup	equal	-6.8878524	15.7757049	2.42869936	0.34390221	
pollinator	haplotype	with_outgroups	unconstrained	-7.3578779	18.7157557	0.48726114	0.8037555	0.69949582
pollinator	haplotype	with_outgroups	equal	-7.3644757	16.7289514	0.46824808	0.75649948	
pollinator	haplotype	only_ingroup	unconstrained	-6.622447	17.2448939	1.35134808	0	0.55952334
pollinator	haplotype	only_ingroup	equal	-6.7584202	15.5168404	2.63642415	0.49816399	
flower_color	primary	with_outgroups	unconstrained	-7.3516439	18.7032878	15.451702	0.2250505	0
flower_color	primary	with_outgroups	equal	-7.5244653	17.0489307	8.00153837	0.25784836	
flower_color	primary	only_ingroup	unconstrained	-6.1020516	16.2041032	9.86561932	0.2911427	0
flower_color	primary	only_ingroup	equal	-6.4977478	14.9954955	7.35164968	0.21914787	
flower_color	haplotype	with_outgroups	unconstrained	-6.2030723	16.4061447	0.07258259	0.1765216	0.75842377
flower_color	haplotype	with_outgroups	equal	-6.6398703	15.2797405	0.29044977	0.41915283	
flower_color	haplotype	only_ingroup	unconstrained	-6.2711902	16.5423803	2.78646403	0.21776477	0.79936157
flower_color	haplotype	only_ingroup	equal	-6.4754053	14.9508107	4.66515625	0.36115074	

Data available from the Dryad Digital Repository:

File S1. Hyb-Seq Targets; http://dx.doi.org/10.5061/dryad.[NNNN]