



HAL
open science

Miocene diversification in the Savannahs precedes tetraploid rainforest radiation in the african tree genus *Afzelia* (Detarioideae, Fabaceae)

Armel Donkpegan, Jean-Louis Doucet, Olivier Hardy, Myriam Heuertz,
Rosalía Piñeiro

► To cite this version:

Armel Donkpegan, Jean-Louis Doucet, Olivier Hardy, Myriam Heuertz, Rosalía Piñeiro. Miocene diversification in the Savannahs precedes tetraploid rainforest radiation in the african tree genus *Afzelia* (Detarioideae, Fabaceae). *Frontiers in Plant Science*, 2020, 11, pp.1-14. 10.3389/fpls.2020.00798 . hal-02906073

HAL Id: hal-02906073

<https://hal.inrae.fr/hal-02906073>

Submitted on 24 Jul 2020

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

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



Distributed under a Creative Commons Attribution 4.0 International License



Miocene Diversification in the Savannahs Precedes Tetraploid Rainforest Radiation in the African Tree Genus *Azelia* (Detarioideae, Fabaceae)

Armel S. L. Donkpegan^{1,2,3*}, Jean-Louis Doucet¹, Olivier J. Hardy², Myriam Heuertz^{4†} and Rosalía Piñeiro^{5,6†}

OPEN ACCESS

Edited by:

Juan Viruel,
Royal Botanic Gardens, Kew,
United Kingdom

Reviewed by:

Carolina Carrizo García,
Instituto Multidisciplinario de Biología
Vegetal (IMBIV), Argentina
Jun Ying Lim,
Nanyang Technological University,
Singapore

*Correspondence:

Armel S. L. Donkpegan
armel.donkpegan@gmail.com

† These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Plant Systematics and Evolution,
a section of the journal
Frontiers in Plant Science

Received: 13 December 2019

Accepted: 19 May 2020

Published: 17 June 2020

Citation:

Donkpegan ASL, Doucet J-L,
Hardy OJ, Heuertz M and Piñeiro R
(2020) Miocene Diversification
in the Savannahs Precedes Tetraploid
Rainforest Radiation in the African
Tree Genus *Azelia* (Detarioideae,
Fabaceae). *Front. Plant Sci.* 11:798.
doi: 10.3389/fpls.2020.00798

¹ Forest is Life, TERRA Teaching and Research Centre, Gembloux Agro-Bio Tech, University of Liège, Gembloux, Belgium, ² Evolutionary Biology and Ecology Unit, Faculté des Sciences, Université Libre de Bruxelles, Brussels, Belgium, ³ INRAE, BFP, University of Bordeaux, Villenave d'Omon, France, ⁴ INRAE, BIOGECO, University of Bordeaux, Cestas, France, ⁵ Department of Geography, College of Life and Environmental Sciences, University of Exeter, Exeter, United Kingdom, ⁶ Evolutionary Genomics, Centre for Geogenetics – Natural History Museum of Denmark, Copenhagen, Denmark

The dating of diversification events, including transitions between biomes, is key to elucidate the processes that underlie the assembly and evolution of tropical biodiversity. *Azelia* is a widespread genus of tropical trees, threatened by exploitation for its valuable timber, that presents an interesting system to investigate diversification events in Africa. Africa hosts diploid *Azelia* species in the savannahs north and south of the Guineo-Congolian rainforest and autotetraploid species confined to the rainforest. Species delimitation and phylogenetic relationships among the diploid and tetraploid species remained unresolved in previous studies using small amounts of DNA sequence data. We used genotyping-by-sequencing in the five widespread *Azelia* species in Africa, the savannah species *A. africana* and *A. quanzensis* and the rainforest species *A. bipindensis*, *A. pachyloba*, and *A. bella*. Maximum likelihood and coalescent approaches resolved all species as monophyletic and placed the savannah and rainforest taxa into two separate clades corresponding to contrasted ploidy levels. Our data are thus compatible with a single biome shift in *Azelia* in Africa, although we were unable to conclude on its direction. SNAPP calibrated species trees show that the savannah diploids started to diversify early, at 12 (9.09–14.89) Ma, which contrasts with a recent and rapid diversification of the rainforest tetraploid clade, starting at 4.22 (3.12 – 5.36) Ma. This finding of older diversification in a tropical savannah clade vs. its sister rainforest clade is exceptional; it stands in opposition to the predominant observation of young ages for savannahs lineages in tropical regions during the relatively recent expansion of the savannah biome.

Keywords: *Azelia*, Leguminosae (Detarioideae), high-throughput sequencing, phylogenomics, coalescent approaches, biome shift, molecular dating, species trees

INTRODUCTION

The biogeographic history of the African flora has been marked by an overall trend toward continental aridification since the wet and warm conditions of the Paleocene (66 – 56 Ma, Senut et al., 2009; Jacobs et al., 2010). Paleobotanical evidence from the north of Africa suggests that rainforest was the most common biome during the Paleocene and the beginning of the Eocene (56 – 33.9 Ma, Jacobs, 2004; Jacobs et al., 2010). More open vegetation appeared in central Africa in the middle Eocene (47.8 – 38 Ma) concomitant with increased temperatures and aridification (Jacobs et al., 2010). A global cooling at the Eocene-Oligocene boundary (33 Ma) led to large-scale extinctions (Zachos et al., 2008; Jacobs et al., 2010) and the grass-dominated savannah biome began to expand in the middle Miocene (16 Ma, Jacobs, 2004), becoming a well established component of tropical vegetation from the late Miocene (ca. 8 Ma, Cerling et al., 1997). The alternation of cold/dry and hot/humid climates of the Miocene (23 – 5.3 Ma), Pliocene (5.3 – 2.6 Ma) and Pleistocene (2.6 – 0.01 Ma) has affected the distribution the major tropical biomes - rainforest, woodland and savannah - with repetitive phases of major expansion or contraction, resulting in the modern distribution of tropical African biomes (Sarnthein and Fenner, 1988; Morley, 2000; Plana, 2004; Salzmann and Hoelzmann, 2005; Anhuf et al., 2006; Miller and Gosling, 2014).

These historical contractions and expansions of the major African biomes have probably triggered biome shifts and diversification in the evolution of tropical plant lineages. Understanding and dating biome shifts is key to understanding the processes that underlie the assembly and evolution of African tropical biodiversity (Wiens and Donoghue, 2004), however, biome shifts have been little studied in the African floras. Multiple transitions from rainforests to dry forests/savannahs have been inferred in the diversification of the tree genus *Guibourtia* in Africa (Tosso et al., 2018). Similarly, three biome transitions from humid forest to dry or montane forests have been documented in the tree genus *Entandrophragma* (Meliaceae) along with ecological adaptations to drier habitat (Monthe et al., 2019). The literature suggests that most biome shifts in tropical Africa support the transition from closed habitats to open habitats (Holstein and Renner, 2011; Veranso-Libalah et al., 2018), which is congruent with paleobotanical evidence for rainforest to be ancient and savannahs to be a more recent biome (Jacobs et al., 2010).

In African forest trees, phylogenetics or population genetics studies have led to the discovery of many new species that could not *a priori* be distinguished based on morphological features (Koffi et al., 2010; Duminil et al., 2012; Heuertz et al., 2014; Dainou et al., 2016; Ikabanga et al., 2017; Lissambou et al., 2018). High-throughput sequencing can facilitate the study of taxonomically difficult groups that contain closely related, weakly differentiated species. Sequencing large portions of the genome of non-model organisms can help generate resolved phylogenies of these complex groups. For non-model taxa, reduced representation sequencing methods such as genotyping-by-sequencing (GBS, – Elshire et al., 2011) can provide thousands of single nucleotide polymorphisms (SNPs) for phylogenetic

analysis without prior knowledge of the genome (Eaton and Ree, 2013; Escudero et al., 2014; Hipp et al., 2014; Ariani et al., 2016; Nicotra et al., 2016; Fernández-Mazuecos et al., 2018).

Afzelia Smith (Detarioideae – Caesalpinioideae) is a widespread and taxonomically complex genus of valuable timber trees that provides an excellent opportunity to apply genomic tools for species delimitation and investigate the role played by biome shifts in species diversification in tropical Africa. *Afzelia* is a Paleotropical genus distributed in Sub-Saharan Africa, where it is known as “doussié,” and Southeast Asia (Donkpegan et al., 2014). The genus exhibits large morphological variability within and between species and can be considered a species complex (Donkpegan et al., 2014). At present, most taxonomists agree that it contains 11 species (Chevalier, 1940; Léonard, 1950; Institut National pour l’Etude Agronomique du Congo-belge [INEAC], 1952; Aubréville, 1959, 1968, 1970; Satabié, 1994). Seven species occur in sub-Saharan Africa: five of them are widely distributed, the savannah species *A. africana* Sm. ex Pers., and *A. quanzensis* Welw., and the rainforest species *A. bipindensis* Harms, *A. bella* Harms, *A. pachyloba* Harms (**Figure 1**); and two are local endemics, *A. parviflora* (Vahl) Hepper occurring in rainforest habitat in West Africa, and *A. peturei* De Wild, probably the least documented species of the *Afzelia* clade in Africa, being found in the transition zone between the rainforest and the Zambesian savannah. The remaining four species, *A. xylocarpa* (Kurz) Craib, *A. rhomboidea* (Blanco) S. Vidal, *A. javanica* (Miq) J. Léonard and *A. palembanica* Baker, occur in Southeast Asia in scattered locations in dry, mixed deciduous or evergreen dipterocarp forest. Based on a fossil attributed to *Afzelia* discovered in the Guang River flora in north-western Ethiopia and dating from the Late Oligocene (27.23 Ma, Pan et al., 2010), it is likely that the genus originated in Africa and that it dispersed subsequently into tropical Asia. Most *Afzelia* species are categorized as vulnerable according to the International Union for the Conservation of Nature (IUCN) Red List because they are substantially exploited for the international timber market (International Union for Conservation of Nature and Natural Resources [IUCN], 2012).

In the evolution of *Afzelia* species in Africa, biome shifts seem to have taken place in association with ploidy levels. The rainforest species *A. bipindensis*, *A. bella*, *A. pachyloba*, and *A. parviflora* – sympatrically distributed across the Guineo-Congolian rainforest – have recently been shown to be autotetraploids using nuclear microsatellites and flow cytometry, whereas the savannah species *A. africana* and *A. quanzensis* – situated north and south of the Guineo-Congolian rainforest, respectively – are diploids (Donkpegan et al., 2015). However, based on Sanger sequencing of two nuclear (nDNA) and three plastid (pDNA) regions, and on full plastome sequences for each of the species, the phylogenetic relationships in the polyploid complex remain uncertain. First, the phylogenetic relationships between the forest and the savannah species were not resolved. Therefore, it remains uncertain whether the forest-savannah transitions have happened once or multiple times in the evolution of the genus. There was cyto-nuclear incongruence with respect to the placement of *A. quanzensis*: pDNA placed the savannah species *A. quanzensis* as sister to the tetraploid forest

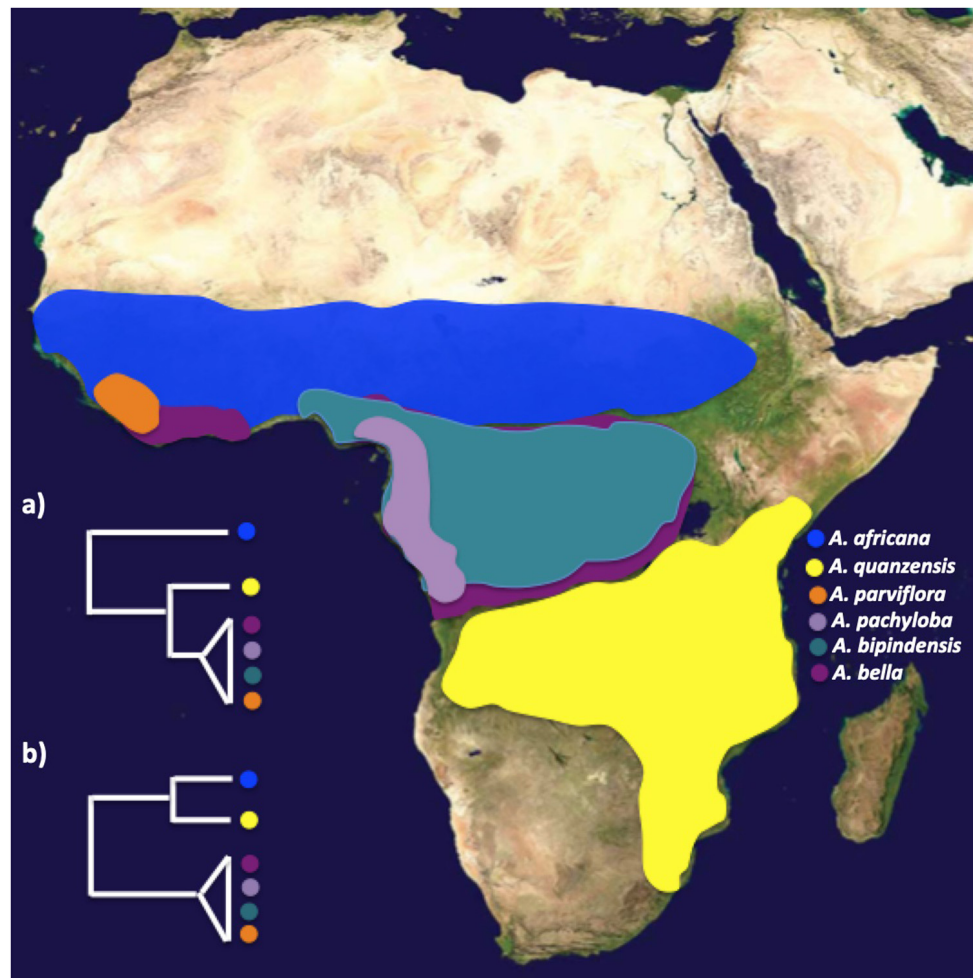


FIGURE 1 | Biogeographic ranges of *Afzelia* sequenced in this study, and alternative phylogenetic relationships **(a,b)** recovered from a previous study (Donkpegan et al., 2017). Plastid data (*psbA*, *trnL*, *ndhF* and genome-wide SNPs via plastomes) suggest that savannah species are paraphyletic with respect to forest taxa **(a)**, whereas nuclear markers (ribosomal *ITS* and the single-copy *PEPC E7* gene) recovered distinct savannah and tropical forest clades **(b)**. Map image: public domain from www.simplemappr.net.

clade, whereas nuclear markers identified the savannah and the forest species as two monophyletic sister clades (**Figures 1a,b**; Donkpegan et al., 2017). Second, the forest taxa showed little genetic differentiation and displayed extensive plastid and nDNA haplotype sharing across species based on few genetic markers, which could be due to incorrect taxonomy, recent speciation with large effective population sizes or ongoing hybridization (Pennington and Lavin, 2016), whereas the savannah species were genetically well differentiated (Donkpegan et al., 2017, 2020). A phylogeny based on genome-wide genetic markers has the potential to improve the taxonomic classification and our understanding of the evolutionary history of this genus of economically important African trees, including the history of biome shifts between forests and savannahs, and shed light on the speciation process in the rainforest taxa.

In this study, we used GBS to sequence the five most abundant species of the genus *Afzelia* in Africa (*A. africana*, *A. quanzensis*, *A. bipindensis*, *A. bella*, and *A. pachyloba*) in order to assess the

phylogenetic relationships among them using multiple methods and datasets. We addressed the following questions.

1. Given the previously unresolved phylogeny, can genome-wide genetic markers provide additional insights into the phylogenetic relationships between diploid savannah and tetraploid rainforest species in *Afzelia*?
2. If so, can molecular dating of the phylogeny inform on which biome shifts occurred during the diversification of *Afzelia* in Africa?
3. Given the extensive haplotype sharing previously observed, can multiple genomic markers delimit species and provide insights into the timing of diversification and/or hybridization in the polyploid complex of rainforest *Afzelia* taxa?

We find strong support for the delimitation of the investigated *Afzelia* species and for phylogenetic relationships between

species. This study represents the most comprehensive phylogenomic evaluation of *Afzelia* to date.

MATERIALS AND METHODS

Sampling, DNA Extraction, Genomic Libraries and Sequencing

We used 41 accessions of *Afzelia* and six accessions (Supplementary Material S1) of other Leguminosae species as outgroups. Our sampling represents the five widely distributed species of the genus *Afzelia* in Africa (Donkpegan, 2017): the diploid savannah species *A. africana* (12 accessions) and *A. quanzensis* (7 accessions) and the tetraploid rainforest species *A. bipindensis* (14 accessions), *A. bella* (4 accessions) *A. pachyloba* Harms (4 accessions). Two species of *Afzelia* (*A. parviflora* and *A. peturei*), which have very restricted ranges, were not included in this study due to lack of recently collected plant material required for GBS. The outgroups were chosen according to the latest available phylogenies in the legume family (Bruneau et al., 2008; LPWG, 2017): namely *Scorodophloeus zenkeri* Harms (one accession), *Prioria balsamifera* (Vermeesen) Breteler (2 accessions), *Prioria oxyphylla* (Harms) Breteler (2 accessions), *Peltogyne* sp. (one accession) and the putative sister species to the *Afzelia* clade, *Intsia bijuga* (Colebr.) Kuntze (one accession). Metadata on all accessions are given in Supplementary Material S1.

DNA was extracted from silica-dried leaves collected in the field and four recent herbarium specimens (National Herbarium of the Netherlands Wageningen, WAG; African Botanical Library of Université Libre de Bruxelles, BRLU; and the Botanic Garden Meise, BR). For each accession, total genomic DNA was extracted using a CTAB protocol (Doyle and Doyle, 1987) and further purified using the QIAquick method (Qiagen, Venlo, Netherlands). We then quantified and controlled the quality of DNA using a QIAxcel (Qiagen). Before library construction, DNA extracts were further purified using a ZR-96 DNA Clean up kit (Zymo Research, Orange, CA, United States) to remove secondary metabolites. DNA quality was checked on a 1.5% agarose gel and DNA quantity was measured with Qbit HS (Thermo Fisher Scientific, Karlsruhe, Germany). To reach the high concentrations required for the genotyping-by-sequencing (GBS) protocol, two extractions per individual sample were pooled at this second purification step whenever possible.

Overall, 180 GBS libraries were built and sequenced on two Illumina lanes (HiSeq2000, San Diego, CA, United States), using 100-bp Single Read chemistry. Given the large genome sizes of our study species (4.9 – 5.0 pg in the diploids and 8.5 – 9.9 in the tetraploids, Donkpegan et al., 2015), two or three independent GBS libraries per individual were built and sequenced for the diploid and tetraploid individuals, respectively. GBS was performed at the Institute for Genomic Diversity and Computational Biology Service Unit at Cornell University (Ithaca, NY, United States) according to a published protocol (Elshire et al., 2011). To select the best enzyme for the GBS protocol, one microgram of DNA of *Afzelia bipindensis* was

used to build test libraries using three different enzymes: *ApeKI* (4.5-base cutter), *EcoT22I* and *PstI* (both 6-base cutters). Libraries were checked for appropriate fragment sizes (<500 bp) and distribution on an Experion automated electrophoresis system (Bio-Rad Laboratories, Hercules, CA, United States). The enzyme *EcoT22I* gave appropriate fragment sizes (<500 bp) and was selected.

Bioinformatics Analyses

De novo Assembly of Reference Sequence

Single-end reads were checked for quality using FastQC 0.11.5 software (Andrews, 2010). Fastq-formatted GBS data was demultiplexed with Saber software¹. Low quality bases and adapter contamination were removed with TRIMMOMATIC version 0.33 (Bolger et al., 2014) with the following options: ILLUMINACLIP 2:30:10, LEADING 3, TRAILING 3, SLIDINGWINDOW 4:15, MINLEN 36. The trimmed reads of all *Afzelia* accessions were *de novo* assembled using PyRAD v.3.0.2 software (Eaton and Ree, 2013, see parameters file in Supplementary Material S2): sequences were clustered within individuals using VSEARCH allowing for indels and nucleotide polymorphisms and assuming a minimum similarity rate of 85% (Rognes et al., 2016). Consensus allele sequences of each cluster (GBS locus) were generated for each individual based on the jointly estimated heterozygosity (H) and the error rate (E). A two-step approach was used to enrich the final dataset in *Afzelia* loci, and genotype all accessions including outgroups for the same loci. In a first step, clustering was performed using all *Afzelia* accessions in order to build a reference catalog with a minimum similarity rate of 85%. Preliminary genotyping was conducted with PyRAD assuming a minimum depth of 8 and a maximum of four shared polymorphic sites across individuals, to minimize the inclusion of paralogs. In a second step, the best *Afzelia* accession was selected, based on the lowest amount of missing data, as the reference catalog of GBS loci in PyRAD for read mapping and final SNP calling (next section) of all accessions, including outgroups.

SNP Discovery and Genotyping

The trimmed reads of all accessions, including outgroups, were then aligned to the reference sequence using the Burrows-Wheeler Aligner BWA mem 0.7.5a-r405 (Li and Durbin, 2009) with -M and -B 4 options, to generate SAM files. SAM files were processed using SAMtools 0.1.17 (Li et al., 2009) and Picard Tools v1.96² to convert from SAM to BAM (Binary Alignment Map) (Sam Format Converter module), sorting the BAM files by position (Sort Sam module) and adding read groups (AddOrReplaceReadGroups module). The resulting BAM files were used as input for Genome Analysis Toolkit (GATK) v3.7 (Depristo et al., 2011). HaplotypeCaller variant discovery was run using emitRefConfidence GVCF mode separately for each sample (Depristo et al., 2011) and GenotypeGVCFs was run on the combined GVCF files to call

¹<https://github.com/najoshi/sabre>

²<http://sourceforge.net/projects/picard/files/picard-tools/1.96/>

and genotype SNPs. Although both diploids and tetraploids were present in the dataset, we used a diploid genotyping model in GATK to facilitate analyses involving both ploidy levels and because this practice is known to provide robust results, with only minor loss in numbers of SNPs in the tetraploids (Anderson et al., 2017). VCFtoolsv0.1.15 (Danecek et al., 2011) was used to filter out indels and non-biallelic variants. Different filter criteria were tested for missing data per sample and per SNP (see section “Results”). To evaluate the utility of GBS to infer phylogenetic relationships in *Afzelia*, we considered two separate datasets: the first containing all *Afzelia* samples with the outgroup taxa (hereafter *Afzelia with outgroups* dataset) and the second containing *Afzelia* samples only (hereafter *Afzelia dataset*). The filtered VCF files were converted to a fasta files containing a single consensus sequence per individual using PGDSpider version 2.1.1.5 (Lischer and Excoffier, 2012).

Phylogenetic Analyses and Estimation of Divergence Times

For the *Afzelia with outgroups* dataset we applied Maximum Likelihood (ML) methods to perform phylogenetic analyses. ML analyses were conducted with the GTR + GAMMA substitution model and 100 bootstrap replicates running RAxML 7.2.6 (Stamatakis, 2014) with default parameters through the CIPRES Portal 2.1 (Miller, 2009)³. Phylogenetic trees were visualized in FigTree 1.4.3 (Rambaut, 2007). The analyses were repeated with ascertainment bias correction using the Lewis method to avoid overestimation of branch lengths and biases in the phylogeny when the number of non-variable sites is not known (Lewis, 2001). We then estimated divergence times within *Afzelia* based on the resulting phylogeny using Bayesian MCMC analysis implemented in BEAST 1.7.4 (Drummond et al., 2012). To facilitate comparison with previously estimated divergence times, we used an *Afzelia* fossil dated to 27.23 Ma (Pan et al., 2010) as in Donkpegan et al. (2017). Since our RAxML phylogeny was unable to resolve the positioning of *Intsia* with respect to *Afzelia* and because of the similarity of species from both genera for morphological characters used to describe the fossil (Kadiri and Olowokudejo, 2008; Pan et al., 2010), we considered the fossil to represent the minimum age for diversification of the *Afzelia-Intsia* clade. Bayesian analyses were done using the following priors: an uncorrelated lognormal relaxed clock model, a Yule process of speciation which is adequate to analyze data at the interspecific level (Yule, 1925; Heled and Drummond, 2015), and the selected nucleotide substitution model. The MCMC analyses were run for 50,000,000 generations, sampling trees every 1,000 generations. To evaluate convergence and ensure sufficient effective sample sizes (ESS values) for all BEAST parameters, we used Tracer 1.6 (Rambaut and Drummond, 2016). Runs were combined with LogCombiner after removing the first 10,000 samples as burn-in. Maximum Clade Credibility trees were produced in

TreeAnnotator 1.8 (Drummond and Rambaut, 2007) and plotted in FigTree 1.4.4⁴.

Species Tree Inference: Species Delimitation and Estimation of Divergence Times

To provide additional support for species delimitation in *Afzelia*, we used several models to estimate species trees directly from the *Afzelia dataset*. We used the multi-species coalescent approach in SNAPP v.1.3.0 (Bryant et al., 2012) implemented in BEAST2 v.2.5.2 (Bouckaert et al., 2014) to estimate species trees from a multilocus SNP matrix. SNAPP is based on a Bayes Factor Delimitation method (BFD*) (Leaché et al., 2015) which allows for the comparison of alternative species delimitation models in an explicit multispecies coalescent framework. The corresponding VCF file was converted to PHYLIP format using the Python script “vcf2phylip”⁵ and the XML input file for SNAPP analyses was prepared using the Ruby script “snapp_prep.rb”⁶. Two independent Markov-Chain Monte Carlo (MCMC) simulations were run for one million generations each, sampling trees at 1000 step intervals. Stationarity and convergence of chains were visually checked in TRACER v1.6 (ESS > 1000; Rambaut and Drummond, 2016)⁷. The program DensiTree v.2.2.6 (Bouckaert and Heled, 2014) was used to visualize the SNAPP trees after discarding the first 10% of each MCMC chain as burn-in. The resulting tree and log files were combined with Logcombiner 1.8.2⁸ with a burn-in of 10 000 for each run. A maximum-clade-credibility summary tree was generated with TreeAnnotator (Drummond and Rambaut, 2007) and visualized in FigTree v.1.4.3⁹. Divergence dating based on the multispecies coalescent has been shown to provide more accurate results than dating of concatenation-based species trees for the ages of younger nodes, which are commonly overestimated when concatenation is used (Stange et al., 2018). We dated the resulting species tree in SNAPP using the same fossil as above. To generate a SNAPP input XML file for divergence dating, we followed the protocol of Stange et al. (2018), using their provided Ruby script (“snapp_prep.rb”). According to their approach, a molecular clock and effective population sizes are shared between all species. We assumed the age of the root to be within a normal distribution of mean = 27.23 Ma and standard deviation being 10% of that variation, with an offset of 20 Ma.

We then used the Generalized Mixed Yule Coalescent (GMYC) model (Pons et al., 2006; Fujisawa and Barraclough, 2013) based on a likelihood method for species delimitation in *Afzelia*. To account for uncertainty in species delimitation, we used the single (*sGMYC*) and multiple (*mGMYC*) threshold species delimitation models using packages APE (Paradis et al., 2004) and SPLITS (Ezard et al., 2013) in R (R Core Team, 2017). The GMYC method requires an ultrametric

⁴<http://tree.bio.ed.ac.uk/software/figtree/>

⁵<https://github.com/edgardomortiz/vcf2phylip>

⁶https://github.com/mmatschiner/snapp_prep

⁷<http://beast.bio.ed.ac.uk/tracer/>

⁸<http://beast.bio.ed.ac.uk/LogCombiner>

⁹<http://tree.bio.ed.ac.uk/software/figtree/>

³www.phylo.org

tree (i.e., calibrated with a molecular clock), which was constructed using BEAST v.1.8 (Bouckaert et al., 2014). We used a relaxed log-normal clock with a coalescent tree prior as these have been identified as best prior parameters for GMYC analyses (Esselstyn et al., 2012). Monte Carlo Markov chains (MCMC) were run for one million iterations, sampling every 1000 iterations. Convergence of chains was assessed using Tracer v.1.6 (Rambaut and Drummond, 2016). The consensus tree (maximum clade credibility tree; 10% burn in; tree not presented) was constructed with TreeAnnotator v.1.7 (Drummond and Rambaut, 2007).

Finally, we used Bayesian (*bPTP*) and ML (*mlPTP*) implementations of the Poisson tree processes model (PTP) (available at <http://species.h-its.org/ptp/>) to estimate the number of speciation events in the *Afzelia* rooted phylogenetic tree based on nucleotide substitutions (Zhang et al., 2013). Because this approach does not require ultrametrization of trees, it constitutes a reasonable alternative to other species delimitation models such as the General mixed Yule coalescent model (Pons et al., 2006). In PTP models, the numbers of substitutions (branch lengths) represent speciation or branching events and, therefore, they only require a phylogenetic tree as input. PTP analyses were conducted on the web server for PTP (available at <http://species.h-its.org/ptp/>) using the best ML tree resulting from the RA × ML analysis (see below).

Inference of Interspecific Hybridization History

We used TreeMix v1.12 (Pickrell and Pritchard, 2012) to infer historical relationships among *Afzelia* species. This method builds a maximum likelihood graph that connects species with their common ancestor, using the covariance structure of allele frequencies between species and a Gaussian approximation for genetic drift. Migration events, i.e., hybridization events, can be modeled to improve the fit of the inferred graph. To meet TreeMix requirements, the *Afzelia* dataset was reduced to SNPs without missing data, a single SNP was selected per GBS locus using VCFtools, and species-level allele frequencies were computed from the VCF to generate the TreeMix infile. We modeled the interspecific evolutionary history in *Afzelia* using $m = 0$ to $m = 4$ migration events.

RESULTS

GBS Data Production and Reference Sequence Construction

Unambiguous barcodes were found in a total of 295 million sequencing reads. After trimming, cleaning, and quality filtering an average of 4.7 million reads per accession were retained. Genotyping *Afzelia* accessions with PyRAD yielded accession-level heterozygosity estimates between 0.0288 and 0.0794 and error rates between 0.0043 and 0.0169 (Supplementary Material S3). The accession with the lowest amount of missing data in the PyRAD genotyping – AD657 of *A. bipindensis* – was used as a GBS reference sequence. It comprised a total of 221,334 loci

representing 3,489,577 bp, including 52,314 polymorphic sites, and 3749 scaffolds (Supplementary Material S3). This reference is available in FASTA format in DRYAD¹⁰.

Mapping and SNP Calling

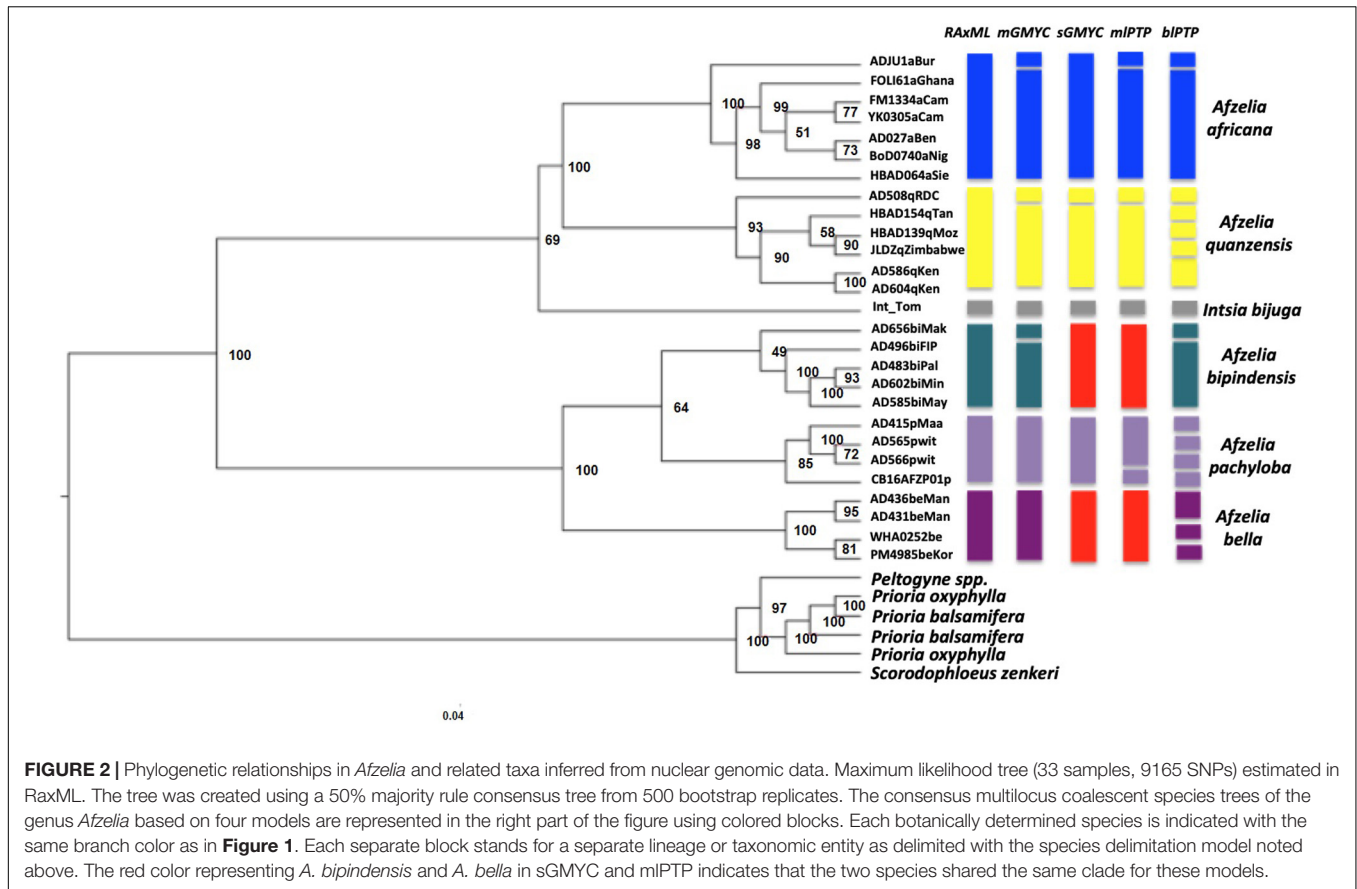
Mapping the reads from all accessions against the reference and genotyping with GATK allowed us to obtain VCF files for two datasets. For the *Afzelia with outgroups* dataset 21,150 SNPs were discovered and 9,165 SNPs were retained in 26 *Afzelia* accessions and all seven outgroups, after filtering INDELs, non-biallelic sites, and sites with more than 40% of missing data. For the *Afzelia* dataset 23,694 SNPs were discovered and 4,823 SNPs were retained (26 accessions) after filtering INDELs, non-biallelic sites, and sites with more than 20% of missing data. The *Afzelia with outgroups* and *Afzelia* datasets were used to generate phylogenetic trees in RAxML. For species delimitation based on the multispecies coalescent in SNAPP, a subset of 2,370 bi-allelic SNPs without missing data in at least one species was retained.

Concatenation-Based Tree and Timing of *Afzelia* Diversification

For the *Afzelia with outgroups* dataset, different datasets (on the percentage of missing data) were tested for phylogenomics of the genus. The dataset (9,165 SNPs) of at most 40% missing data yielded the phylogenetic relationships that were most congruent with the known topology in the legumes (Bruneau et al., 2008; LPWG, 2017). The *Afzelia* clade (including *Intsia bijuga*) formed a monophyletic group in the RAxML tree (Figure 2). We also obtained two strongly supported monophyletic clades that correspond to the diploid and tetraploid lineages in the *Afzelia-Intsia* clade. All *Afzelia* species – represented by multiple accessions each – appeared monophyletic with strong support for four species and lower support for *A. bipindensis*. Such well-supported species delimitation had never been found in previous studies of phylogeny in the genus, especially in the rainforest clade of tetraploid species. The differentiation of the savannah species was well supported. In contrast, the topology of the diversification of the three rainforest species was not well resolved, as revealed by low bootstrap supports. The analyses with ascertainment bias correction to avoid overestimation of branch lengths and biases in the phylogeny using SNP datasets, resulted in identical topology and did not improve the resolution of the rainforest clades (results not shown).

Based on molecular dating of the concatenation-based phylogeny, diversification appeared to occur earliest in the diploid lineage of *Afzelia*, while the position of the sister species *Intsia bijuga* remained ambiguous on the topology of the tree (Figure 3). The diversification of the *Afzelia-Intsia* clade started in the Oligocene, the posterior mean age of the common ancestor (MRCA, node A) of the clade being estimated at 33.31 Ma [95% highest posterior density (HPD) 28.64–41.04 Ma] (Figure 3). The divergence of each monophyletic species on the basis of the well-resolved phylogenetic tree at the genus level suggests that savannah species diversified in the Middle Miocene, *A. africana* (node F, 12.33 Ma) and *A. quanzensis* (node E, 14.68 Ma),

¹⁰<https://doi.org/10.5061/dryad.95~x~69p8gf>



whereas the rainforest species would have appeared in the Upper Miocene: *A. bipindensis* (node G, 10.06 Ma), *A. pachyloba* (node H, 10.08 Ma) and *A. bella* (node I, 08.39 Ma).

Coalescent-Based Species Trees and Timing of *Afzelia* Diversification

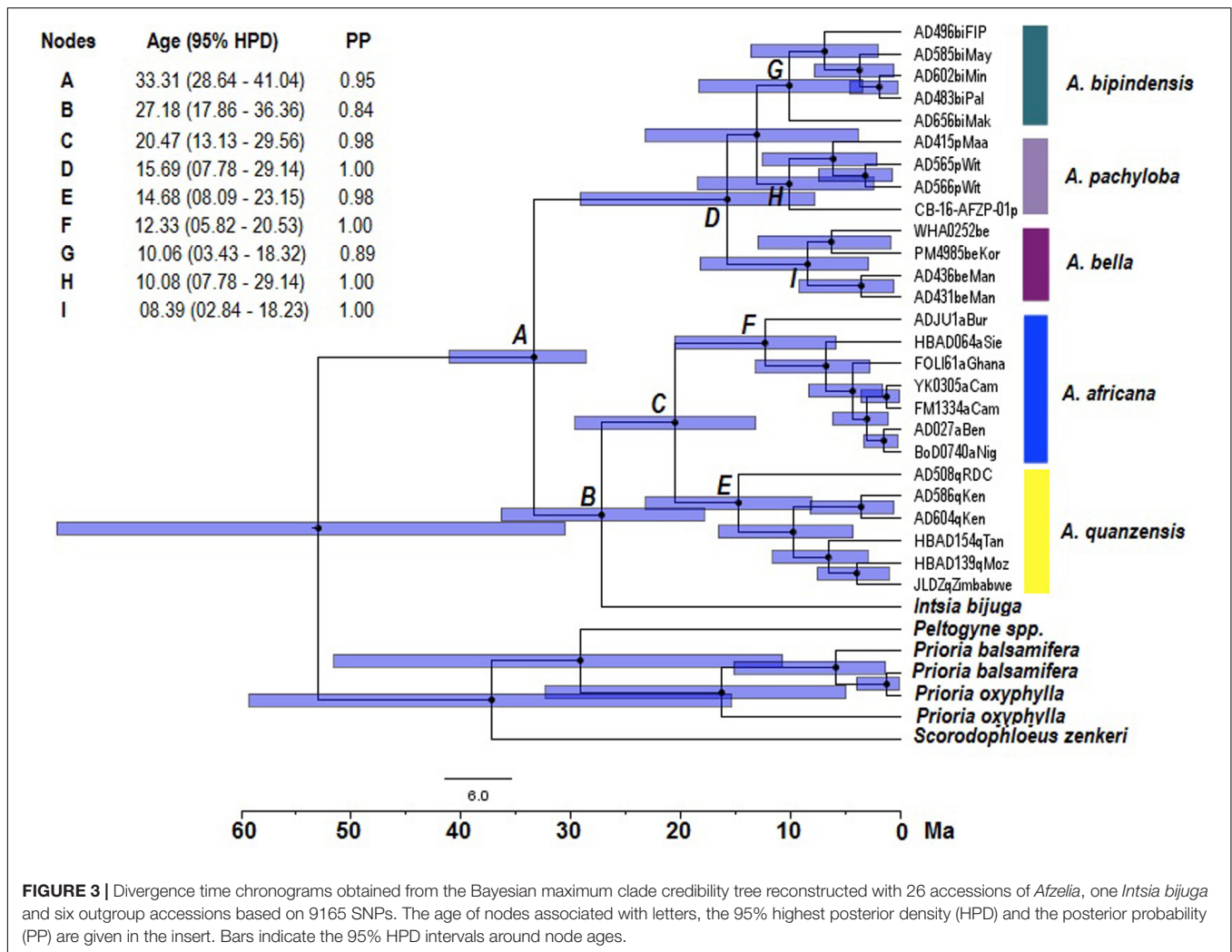
For the *Afzelia* dataset the SNAPP analysis resolved five well-differentiated clades that support the monophyly of all species. The visualization of the species trees superimposed in the DensiTree plot (**Figure 4**) resolved the evolutionary relationships among species with no signs of conflict among trees, even within the rainforest clade: *A. pachyloba* is the sister species of the clade containing *A. bipindensis* and *A. bella*. The results of the GMYC and PTP models are plotted against the RAxML phylogeny in **Figure 2**. They resolved between 6 and 15 sub-lineages within *Afzelia*. The sGMYC and mlPTP models placed *A. bella* and *A. bipindensis* into the same species cluster, in line with the close relationship revealed by the SNAPP species tree. The SNAPP-calibrated tree revealed, as expected, later diversification dates than the concatenated gene tree. The diversification of *Afzelia* has a the posterior mean age of the MRCA in the late Oligocene at 26.93 Ma (95% HPD 21.06 – 32.92 Ma; **Figure 4**). The rainforest lineage diversified rapidly between the Pliocene and the early Pleistocene (mean *A. pachyloba* split at 4.22 Ma and mean *A. bipindensis/A. bella* split at 2.78 Ma) and the savannah lineage earlier in the Miocene (mean at 12 Ma).

Interspecific Hybridization History

The genetic relationships among species revealed by TreeMix distinguished the diploid and tetraploid clades, in agreement with the phylogenetic relationships evidenced with concatenation or coalescent-based methods, and confirmed the placement of *A. pachyloba* as sister of the clade containing *A. bipindensis* and *A. bella*, as revealed by species delimitation methods (**Figure 5A**). The proportion of variance in the data explained by the model was high, PVE = 0.975, for a model without migration. The addition of migration events improved the proportion of variance explained to PVE = 0.998 for $m = 1$ and PVE = 0.9999 for $m = 2$ migration events. The first migration event links diploid *A. africana* with tetraploid *A. bella* whereas the second links an ancestor of diploid *A. quanzensis* with tetraploid *A. bipindensis* (**Figure 5**).

DISCUSSION

Our phylogenetic reconstructions provided the most robust phylogenetic framework of the tropical tree genus *Afzelia* in Africa produced to date. Both the SNP concatenated gene tree (RAxML tree; **Figure 2**) and the coalescent-based species tree (SNAPP tree; **Figure 4**) highly supported two major monophyletic clades associated with habitat and ploidy levels: a diploid savannah clade and a tetraploid



rainforest clade. The calibrated phylogeny and the species tree (Figures 3, 4) show an earlier diversification of the savannah clade followed by a later speciation within the rainforest clade. Species delimitation within these two major clades, with all species resolved as monophyletic, was also the most robust to date.

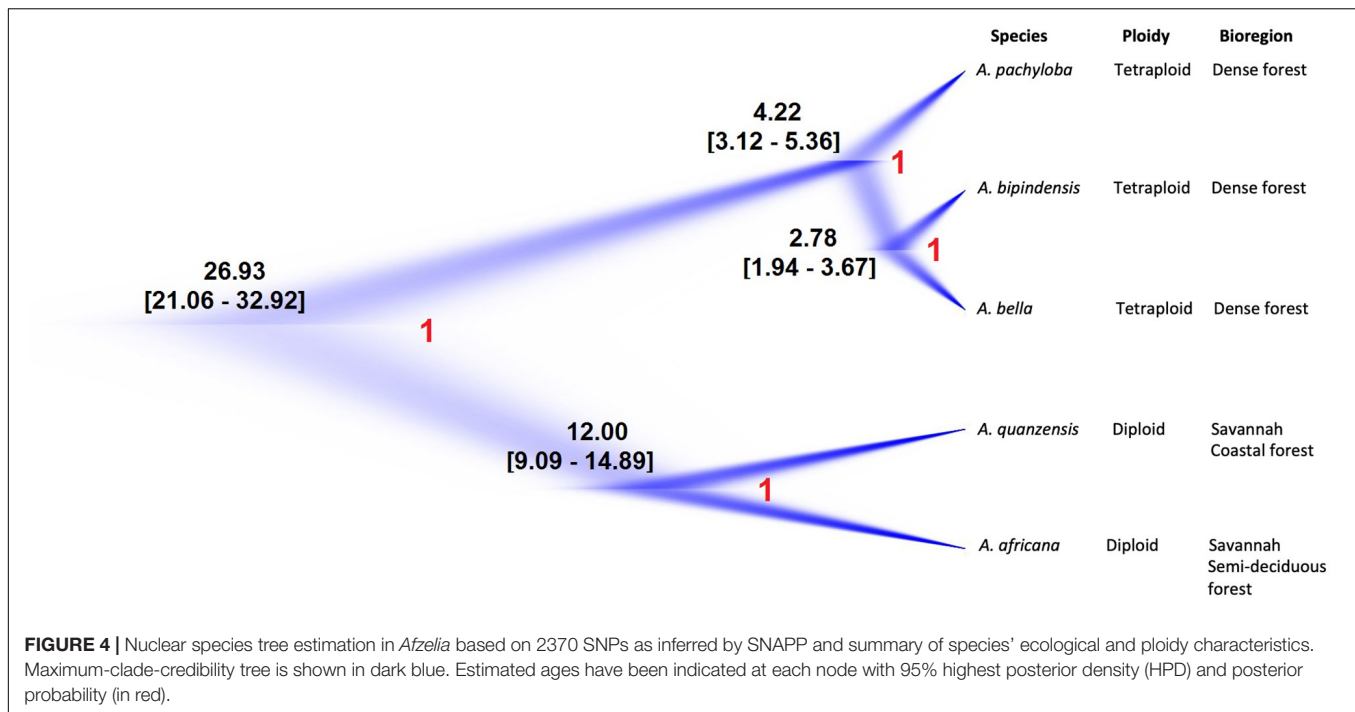
A Single Biome Shift in African *Azelia*

Genotyping-by-sequencing data strongly supported the monophyly of two major habitat-specific clades in *Azelia*. This suggests a single transition between the savannah and the rainforest biomes in Africa. Previous plastid DNA sequence data – Sanger sequences of a few loci and full plastome sequences – placed the savannah species *A. quanzensis* as sister to the rainforest clade (Donkpegan et al., 2017). In contrast, the phylogenetic analyses of the multilocus genomic GBS data strongly supported the monophyly of the savannah clade and suggested a unique forest-savannah shift in the diversification of *Azelia* in Africa.

The calibrated phylogeny and the dated species tree indicate a diversification of the savannah clade in the Miocene,

earlier than the rainforest clade. This early diversification of the savannah species is exceptional in the context of the global plant evolutionary patterns reported in tropical Africa and South America that point to a relatively young age for savannah lineages. In the South African tree flora, the majority of divergence times between sister taxa of savannah trees were dated within the Pleistocene (the last 2 Ma), which is more recent than those between sister taxa of forest trees (Maurin et al., 2014). In the Brazilian savannah (Cerrado) the origin of woody plants restricted to the Cerrado is also estimated to be recent (<10 Ma), most of them in the Pliocene (<4 Ma; Simon et al., 2009). This seems to reflect the relative ages of the biomes, with rainforests dating from at least the early Paleocene (Burnham and Johnson, 2004; Jacobs et al., 2010), whereas savannahs are much younger, arising only in the Miocene (Jacobs, 2004; Senut et al., 2009; Simon et al., 2009; Pennington and Hughes, 2014).

Evolutionary shifts of plant lineages from the rainforest to savannah or to other dry biomes seem to have been significantly more frequent than switches of plant lineages



into the rainforest (Simon et al., 2009; Simon and Pennington, 2012, Donoghue and Edwards, 2014, Maurin et al., 2014, Tosso et al., 2018; Freitas et al., 2019; Monthe et al., 2019). It seems logical that older biomes, i.e., the rainforests, will act as sources of lineages for younger biomes, i.e., the savannahs. Of course, there is no *a priori* reason to expect biome switches to be unidirectional, and given enough time, savannah lineages may re-enter rainforests. This might have been the case of the African *Afzelia*, where molecular phylogenies indicate recent splits within the rainforest lineage in the Upper Miocene and Pleistocene. If we assume that the splits of extant species occurring in the same biome indicate a shared ancestor within that biome, it is possible that the African *Afzelia* originated as a diploid savannah lineage followed by whole genome duplication and subsequent diversification in the rainforest. However, we should note that biome conservatism is more common than biome shifts in phylogenies of Southern Hemisphere plants (Crisp et al., 2009) and that the topology of the *Afzelia* phylogeny alone did not allow us to conclude on the direction of the biome shift, as both monophyletic clades diversified into their respective rainforest or savannah biomes.

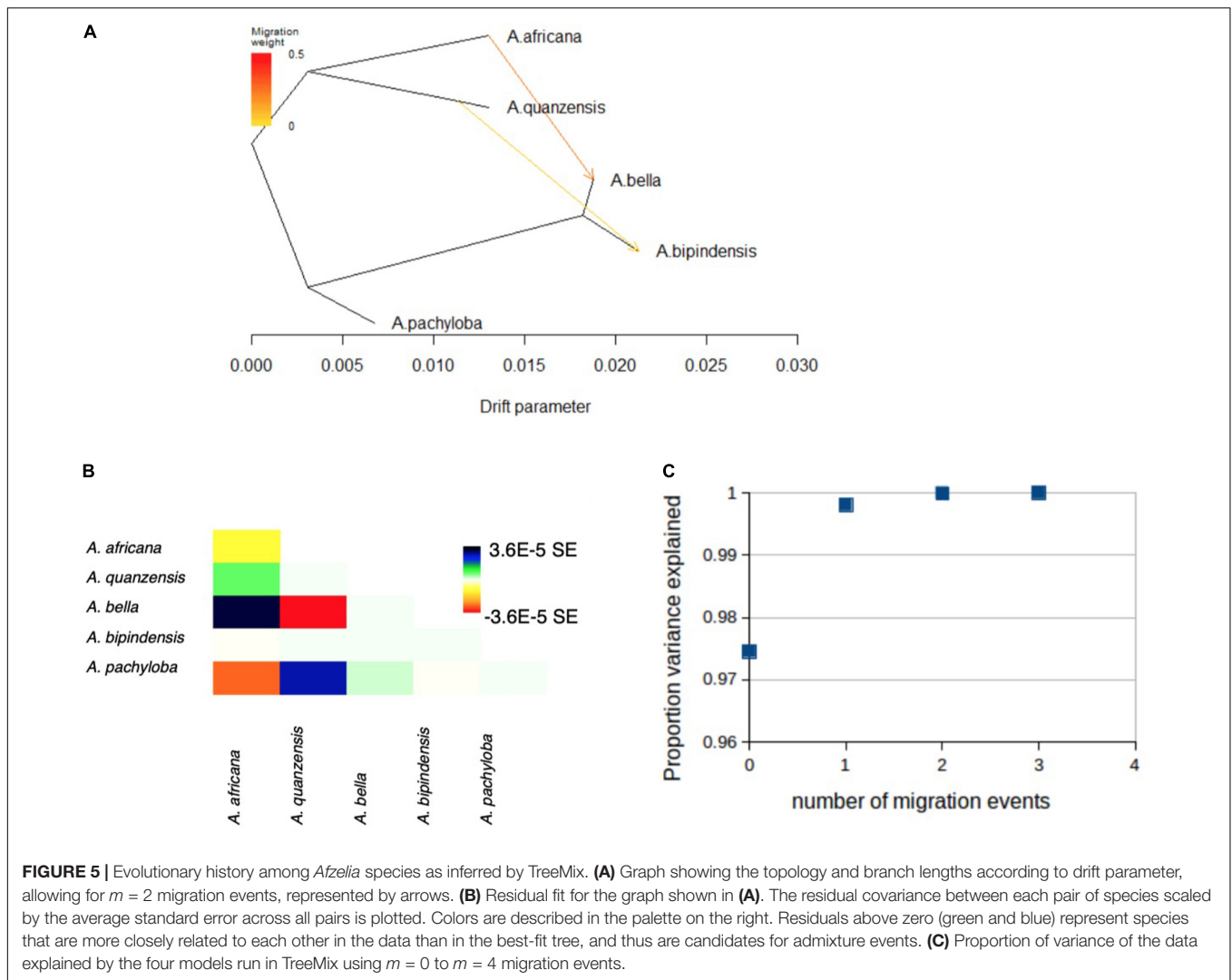
In terms of morphological trait variation in *Afzelia*, dry forest species (*A. africana* and *A. quanzensis*) are clearly differentiated from rainforests species (*A. bella*, *A. bipindensis*, and *A. parviflora*) based on vegetative and floral variables (Donkpegan et al., 2017). The main morphological discriminant variables which separate dry forest versus rainforest species are (i) the rounded versus truncated-attenuated basal leaflet shape; (ii) at the distal end of the leaflets, the scalloped versus acuminate and

mucronate shapes (sometimes with black spots). Leaves are longer and have a smaller number of secondary veins in dry forest vs. rainforest *Afzelia* species (Donkpegan et al., 2017). It is, however, difficult to associate these leaf and vein trait differences with a clear habitat-adaptive role in the light of the literature (Sack and Scoffoni, 2013; Tng et al., 2013) and to our knowledge there are no studies available that treat functional or physiological trait variation in *Afzelia*.

Evolutionary Radiation of the African Rainforest Species

Previous Sanger sequencing of two nuclear and three plastid regions revealed extensive allele sharing across the rainforest clade of *Afzelia*. Accessions of the same species were scattered across the trees (Donkpegan et al., 2017), which could be due to incorrect taxonomy, recent speciation or ongoing hybridization. The GBS data effectively resolved species delimitation and diversification times of *Afzelia* in the rainforest. All species – which are morphologically similar and broadly sympatric – were resolved as monophyletic, validating the latest taxonomic revisions. The data also revealed a recent diversification process, and no trace of hybridization between species.

In *Afzelia*, autopolyploidization occurred prior to rapid speciation in the rainforests, suggesting the role of whole-genome duplications in the onset of adaptive radiations. Polyploidy represents an immediate source of genetic novelty that may promote evolutionary changes and divergence (Wood et al., 2007). Rapid speciation events immediately after polyploidization are well known in the evolution of



plant groups and point at the success of whole-genome duplications as triggers of speciation (Soltis et al., 2007). The rainforest *Azelia* species correspond to monophyletic clades, although the clade support varied depending on the filtering parameters used to generate the datasets as well as on the phylogenetic method chosen. This weaker phylogenetic support suggests incomplete lineage sorting in the tetraploid species, which is consistent with their more recent diversification and the larger effective population sizes of tetraploid organisms (Arnold et al., 2012). In addition, the stronger support for monophyly in the savannah than rainforest species is in agreement with similar observations in South America (Pennington and Lavin, 2016). The pattern of strong support for monophyletic species in seasonally dry tropical forests in South America has been attributed to maintenance of resident lineages adapted to a stable, seasonally dry ecology; conversely lower support for monophyly in rainforest trees was attributed to lesser habitat stability creating opportunities for immigration and speciation from taxa with large effective population sizes

extending over large areas (Pennington and Lavin, 2016; Heuertz et al., 2020).

Coalescent-Based Phylogenetics and Population Genetics of Multiple Nuclear Loci for Species Delimitation in Recent Radiations

In the rainforest clade of *Azelia*, phylogenomic analysis revealed short branches among species, in line with a scenario of recent radiation, and revealed the superior performance of the coalescent-based over the concatenation-based methods. Using concatenation-based trees, the topology changed depending on the filtering parameters, which might be explained by the poor performance of this approach under highly incomplete lineage sorting, typically found in cases of rapid speciation (Fernández-Mazuecos et al., 2018). Coalescent-based species trees are based on tests of alternative hypotheses of species delimitation. In this context, lineages do not need to be resolved as monophyletic in gene trees, which leads to more reliable estimates of the

evolutionary relationships and divergence times. Overall, an approach combining multiple nuclear loci with coalescent-based phylogenetic analyses was revealed as the optimal approach to resolve and date the radiation of the rainforest clade.

Forcing the rainforest radiation into bifurcating phylogenetic trees may be problematic because of the age of the radiation and potential for interspecific gene flow. Therefore, we explored the possibility of reticulate evolution using a population genetics approach implemented in TreeMix. Using genome-wide SNPs we found no evidence for gene flow between rainforest species despite their rapid genetic differentiation and sympatric distribution. Given the limited sample sizes within species it is possible that we have missed introgression events limited to sympatric populations of both species. Nevertheless, our analyses identified possible hybridization events that led to incorporation of genetic variation from savannah to rainforest species after the split of the main lineages. Such hybridization events could potentially explain the previously observed cyto-nuclear incongruence, where *A. quanzensis* grouped with rainforest species for plastid DNA but with *A. africana* for two nuclear regions (Donkpegan et al., 2017). *Afzelia quanzensis* could thus have captured plastid DNA from rainforest species during an ancient hybridization event, or else, incomplete lineage sorting may explain the pattern. Altogether, phylogenetic evidence in *Afzelia* points to the existence of hybridization and introgression between species despite differences in ploidy, as has been observed in other plant lineages, such as orchids of the genus *Epidendrum* (Pinheiro et al., 2010) or grasses of the genus *Spartina* (Ainouche et al., 2009).

Similar to Donkpegan et al. (2017) our study did not permit to solve the phylogenetic position of *Intsia bijuga*, the putative sister clade to *Afzelia*. *Intsia bijuga* was placed within the *Afzelia* clade, confirming the close relatedness of both genera. Multilocus data of Asian *Afzelia* species and increased taxon sampling in *Intsia* with multiple accessions per species may help improve the phylogenetic positioning of *Intsia*.

CONCLUSION

We have elucidated the evolutionary history of the widespread emblematic and threatened African tree species of the genus *Afzelia* using genome-wide multilocus data. While the genus was previously recognized as a species complex (Donkpegan et al., 2015), we showed that based on our set of accessions and markers, all species were resolved as monophyletic, and that diversification in the savannah clade preceded that of the rainforest clade. Phylogenomic studies thus represent promising approaches to clarify evolutionary relationships in taxonomic groups that show a high level of variation.

DATA AVAILABILITY STATEMENT

Fasta reference sequence is available from the Dryad data repository: <https://datadryad.org/stash/share/QK6Ay8vq7gprepB6>

6aGWp42I8PCwjrpSR2IXUqX4rRk. Raw fastq sequences of GBS data are being submitted to GenBank's Sequence Read Archive.

AUTHOR CONTRIBUTIONS

AD, MH, and RP conceived the study. All authors collected the data, performed the analyses, interpreted the results, and contributed to drafting and writing the manuscript.

FUNDING

The authors thank the “Fonds pour la Formation à la Recherche dans l'Industrie et l'Agriculture (FRIA-FNRS, Belgium),” the Marie Curie FP7-PEOPLE-2012-IEF program (project AGORA awarded to RP), the Fonds de la Recherche Scientifique (F.R.S.-FNRS through grant J.0292.17F), the Belgian Science Policy (project AFRIFORD), and the DynAffor project (funded by FFEM-AFD) for funding this research. This work has also benefited from an “Investissements d'Avenir” grant managed by Agence Nationale de la Recherche (CEBA: ANR-10-LABX-25-01).

ACKNOWLEDGMENTS

The authors acknowledge a “Patrimoine de l'Université de Liège” and Labex COTE mobility grant provided to ASLD at INRAE; Nature+, Esra Kaymak and Barbara Leal for technical assistance, M. Thomas P. Gilbert for hosting the GBS labwork performed by RP at the University of Copenhagen, Jérôme Chave et Bernadette Grosso for providing *Peltogyne* specimens (BRIDGE collection) and R. Toby Pennington (Geography, College of Life and Environmental Sciences, University of Exeter – United Kingdom) for contributing to the discussion on tropical biome-shifts.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2020.00798/full#supplementary-material>

MATERIAL S1 | Sample origins (fresh and herbarium material) of plant tissue samples used for the phylogenetic analyses of *Afzelia*. “na” means that no collector or voucher names are mentioned on the specimen consulted and studied.

MATERIAL S2 | The parameters used in PyRAD for *de novo* assembly.

MATERIAL S3 | Results from *de novo* assembling of 30 *Afzelia* and *Intsia* accessions. (nlloc) is the number of loci identified in the intra-accession clustering; (f1loc) is the number of loci with >N depth coverage; (f2loc) number of loci with >N depth and passed paralog filter; (nsites) number of sites across f loci; (npoly) number of polymorphic sites in nsites; (poly) frequency of polymorphic sites. H and E are heterozygosity and error rate, respectively.

REFERENCES

- Ainouche, M. L., Fortune, P. M., Salmon, A., Parisod, C., Grandbastien, M. A., Fukunaga, K., et al. (2009). Hybridization, polyploidy and invasion: lessons from *Spartina* (Poaceae). *Biol. Invasions* 11:1159. doi: 10.1007/s10530-008-9383-2
- Anderson, B. M., Thiele, K. R., Krauss, S. L., and Barrett, M. D. (2017). Genotyping-by-sequencing in a species complex of Australian hummock grasses (Triodia): methodological insights and phylogenetic resolution. *PLoS One* 12:e0171053. doi: 10.1371/journal.pone.0171053
- Andrews, S. (2010). *FastQC: A Quality Control Tool for High Throughput Sequence Data*. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc> (accessed March, 2017).
- Anhuf, D., Ledru, M. P., Behling, H., Da Cruz, F. W. Jr., Cordeiro, R. C., Van der Hammen, T., et al. (2006). Paleo-environmental change in Amazonian and African rainforest during the LGM. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 239, 510–527. doi: 10.1016/j.palaeo.2006.01.017
- Ariani, A., Berny, M., Teran, J. C., and Gepts, P. (2016). Genome-wide identification of SNPs and copy number variation in common bean (*Phaseolus vulgaris* L.) using genotyping-by-sequencing (GBS). *Mol. Breed.* 36:87.
- Arnold, B., Bombliès, K., and Wakeley, J. (2012). Extending coalescent theory to autotetraploids. *Genetics* 192, 195–204. doi: 10.1534/genetics.112.140582
- Aubréville, A. (1959). *La flore forestière De La Côte d'Ivoire*, Vol. 1, 2 Edn. Nogent-sur-Marne: Centre Technique Forestier Tropical.
- Aubréville, A. (1968). Légumineuses – Césalpinioïdées. Flore du Gabon. *Muséum Natl. d'Histoire Nat.* 15, 111–118.
- Aubréville, A. (1970). Légumineuses - Césalpinioïdées (Leguminosae - Caesalpinioideae). *Flore Cameroun* 9:339.
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Bouckaert, R., and Heled, J. (2014). DensiTree 2: seeing trees through the forest. *bioRxiv* [Preprint]. doi: 10.1101/012401
- Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., et al. (2014). BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* 10:e1003537. doi: 10.1371/journal.pcbi.1003537
- Bruneau, A., Mercure, M., Lewis, G. P., and Herendeen, P. S. (2008). Phylogenetic patterns and diversification in the caesalpinoid legumes. *Botany* 86, 697–718. doi: 10.1139/b08-058
- Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N. A., and RoyChoudhury, A. (2012). Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Mol. Biol. Evol.* 29, 1917–1932. doi: 10.1093/molbev/mss086
- Burnham, R. J., and Johnson, K. (2004). South American palaeobotany and the origins of neotropical rainforests. *Philos. T. Roy. Soc. B* 359, 1595–1610. doi: 10.1098/rstb.2004.1531
- Cerling, T. E., Harris, J. M., McFadden, B. J., Leakey, M. G., Quade, J., Eisenmann, V., et al. (1997). Global vegetation change through the Miocene/Pliocene boundary. *Nature* 389, 153–158. doi: 10.1038/38229
- Chevalier, A. (1940). Sur un arbre du Cameroun et du Gabon à bois utilisable (*Afzelia pachyloba* Harms). *Bot. Appl. Agricult. Trop.* 19, 484–488. doi: 10.3406/jatba.1939.6006
- Crisp, M. D., Arroyo, M. T., Cook, L. G., Gandolfo, M. A., Jordan, G. J., McGlone, M. S., et al. (2009). Phylogenetic biome conservatism on a global scale. *Nature* 458, 754–756. doi: 10.1038/nature07764
- Dainou, K., Blanc-Jolivet, C., Degen, B., Kimani, P., Ndiade-Bourobou, D., Donkpegan, A. S. L., et al. (2016). Revealing hidden species diversity in sister species using SNPs and SSRs on a systematically collected sample – A case study in the African tree genus *Milicia*. *BMC Evol.* 16:259. doi: 10.1186/s12862-016-0831-9
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. doi: 10.1093/bioinformatics/btr330
- DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43, 491–498.
- Donkpegan, A. S. L. (2017). *Evolutionary History of Afzelia Smith (Leguminosae - Caesalpinioideae) Complex in Forest and Savannah Ecosystems of Tropical Africa*. 175. PhD thesis, University of Liège – Gembloux Agro-Bio Tech, Belgium.
- Donkpegan, A. S. L., Doucet, J.-L., Dainou, K., and Hardy, O. J. (2015). Microsatellite development and flow cytometry in the African tree genus *Afzelia* (Fabaceae, Caesalpinioideae) reveal a polyploid complex. *Appl. Plant Sci.* 3:1400097. doi: 10.3732/apps.1400097
- Donkpegan, A. S. L., Doucet, J.-L., Migliore, J., Duminiel, J., Dainou, K., Rosalia, P., et al. (2017). Evolution in African tropical trees displaying ploidy-habitat association: the genus *Afzelia* (Leguminosae). *Mol. Phyl. Evol.* 107, 270–281. doi: 10.1016/j.ympev.2016.11.004
- Donkpegan, A. S. L., Hardy, O. J., Lejeune, P., Oumorou, M., Dainou, K., and Doucet, J.-L. (2014). Un complexe d'espèces d'*Afzelia* des forêts africaines d'intérêt économique et écologique (synthèse bibliographique). *Biotechnol. Agron. Soc. Environ.* 18, 233–246.
- Donkpegan, A. S. L., Piñeiro, R., Heuert, M., Duminiel, J., Dainou, K., Doucet, J.-L., et al. (2020). Population genomics of the widespread African savannah trees *Afzelia africana* and *Afzelia quanzensis* reveals no significant past fragmentation of their distribution ranges. *Am. J. Bot.* 107, 498–509. doi: 10.1002/ajb2.1449
- Donoghue, M. J., and Edwards, E. J. (2014). Biome shifts and niche evolution in plants. *Ann. Rev. Ecol. Evol. Syst.* 45, 547–572. doi: 10.1146/annurev-ecolsys-120213-091905
- Doyle, J. J., and Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 19, 11–15.
- Drummond, A. J., and Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7:214. doi: 10.1186/1471-2148-7-214
- Drummond, A. J., Suchard, M. A., Xie, D., and Rambaut, A. (2012). Bayesian P hylogenetics with BEAUti and the BEAST 1.7 Research article. *Soc. Mol. Biol. Evol.* 29, 1969–1973. doi: 10.1093/molbev/mss075
- Duminiel, J., Kenfack, D., Viscosi, V., Grumiau, L., and Hardy, O. J. (2012). Testing species delimitation in sympatric species complexes: the case of an African tropical tree, *Carapa* spp. (Meliaceae). *Mol. Phylogenet. Evol.* 62, 275–285. doi: 10.1016/j.ympev.2011.09.020
- Eaton, D. A., and Ree, R. H. (2013). Inferring phylogeny and introgression using RADseq data: an example from flowering plants (Pedicularis: Orobanchaceae). *Syst. Biol.* 62, 689–706. doi: 10.1093/sysbio/syt032
- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., et al. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6:e19379. doi: 10.1371/journal.pone.0019379
- Escudero, M., Eaton, D. A. R., Hahn, M., and Hipp, A. L. (2014). Genotyping-by-sequencing as a tool to infer phylogeny and ancestral hybridization: a case study in *Carex* (Cyperaceae). *Mol. Phylogenet. Evol.* 79, 359–367. doi: 10.1016/j.ympev.2014.06.026
- Esselstyn, J. A., Evans, B. J., Sedlock, J. L., Anwarali Khan, F. A., and Heaney, L. R. (2012). Single-locus species delimitation: a test of the mixed Yule-coalescent model, with an empirical application to Philippine round-leaf bats. *Proc. R. Soc. B Biol. Sci.* 279, 3678–3686. doi: 10.1098/rspb.2012.0705
- Ezard, T., Fujisawa, T., and Barraclough, T. (2013). *R package splits: SPecies' Limits by Threshold Statistics, version 1.0-18/r45*. Available online at: <http://r-forge.r-project.org/projects/splits/> (accessed September, 2019).
- Fernández-Mazuecos, M., Mellers, G., Vigalondo, B., Sáez, L., Vargas, P., and Glover, B. J. (2018). Resolving recent plant radiations: power and robustness of genotyping-by-sequencing. *Syst. Biol.* 67, 250–268. doi: 10.1093/sysbio/syx062
- Freitas, C. G., Bacon, C. D., Souza-Neto, A. C., and Collevatti, R. G. (2019). Adjacency and area explain species bioregional shifts in neotropical palms. *Front. Plant Sci.* 10:55. doi: 10.3389/fpls.2019.00055
- Fujisawa, T., and Barraclough, T. G. (2013). Delimiting species using single-locus data and the Generalized Mixed Yule Coalescent approach: a revised method and evaluation on simulated data sets. *Syst. Biol.* 62, 707–724. doi: 10.1093/sysbio/syt033
- Heled, J., and Drummond, A. J. (2015). Calibrated birth–death phylogenetic time-tree priors for bayesian inference. *Syst. Biol.* 64, 369–383. doi: 10.1093/sysbio/syu089
- Heuert, M., Caron, H., Scotti-Saintagne, C., Pétronelli, P., Engel, J., Tysklind, N., et al. (2020). The hyperdominant tropical tree *Eschweilera coriacea*

- (Lecythidaceae) shows higher genetic heterogeneity than sympatric *Eschweilera* species in French Guiana. *Plant Ecol. Evol.* 153, 67–81. doi: 10.5091/plecevo.2020.1565
- Heuertz, M., Duminil, J., Dauby, G., Savolainen, V., and Hardy, O. J. (2014). Comparative phylogeography in rainforest trees from Lower Guinea. *Africa. PLoS One* 9:e84307. doi: 10.1371/journal.pone.0084307
- Hipp, A. L., Eaton, D. A. R., Cavender-Bares, J., Fitzek, E., Nipper, R., and Manos, P. S. (2014). A framework phylogeny of the American oak clade based on sequenced RAD data. *PLoS One* 9:e93975. doi: 10.1371/journal.pone.0093975
- Holstein, N., and Renner, S. S. (2011). A dated phylogeny and collection records reveal repeated biome shifts in the African genus *Coccinia* (Cucurbitaceae). *BMC Evol. Biol.* 11:28. doi: 10.1186/1471-2148-11-28
- Ikabanga, D. U., Stévant, T., Koffi, G. K., Monthe, F. K., Doubindou, N. E. C., Dauby, G., et al. (2017). Combining morphology and population genetic analysis uncover species delimitation in the widespread African tree genus *Santiria* (Burseraceae). *Phytotaxa* 321, 166–180.
- Institut National pour l'Etude Agronomique du Congo-belge [INEAC] (1952). *Spermatophytes. Flore du Congo Belge et du Ruanda-Urundi*. Bruxelles: INEAC, 579.
- International Union for Conservation of Nature and Natural Resources [IUCN] (2012). *IUCN Red List of Threatened Species*. Available online at: www.iucnredlist.org (accessed January 08, 2012).
- Jacobs, B. F. (2004). Palaeobotanical studies from tropical Africa: relevance to the evolution of forest, woodland and savannah biomes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 359, 1573–1583. doi: 10.1098/rstb.2004.1533
- Jacobs, B. F., Pan, A. D., and Scotese, C. R. (2010). *A review of the Cenozoic vegetation history of Africa. Cenozoic Mammals of Africa*. Berkeley: University of California Press, 57–72.
- Kadiri, A. B., and Olowokudejo, J. D. (2008). Comparative foliar epidermal morphology of the West African species of the genus *Afzelia* Smith (Leguminosae: Caesalpinioideae). *Gayana Bot.* 65, 84–92.
- Koffi, K. G., Heuertz, M., Doumenge, C., Onana, J. M., Gavory, F., and Hardy, O. J. (2010). A combined analysis of morphological traits, chloroplast and nuclear DNA sequences within *Santiria trimeria* (Burseraceae) suggests several species following the Biological Species Concept. *Plant Ecol. Evol.* 143, 160–169. doi: 10.5091/plecevo.2010.433
- Leaché, A. D., Banbury, B. L., Felsenstein, J., de Oca, A. N. -M., and Stamatakis, A. (2015). Short tree, long tree, wrong tree: new acquisition bias corrections for inferring SNP phylogenies. *Syst. Biol.* 64, 1032–1047. doi: 10.1093/sysbio/syv053
- Léonard, J. J. G. (1950). Notes sur les genres paleotropicaux *Afzelia*, *Intsia* et *Pahudia* (Legum. Caesalp.). *Reinwardtia* 1, 61–66.
- Lewis, P. O. (2001). A likelihood approach to estimating phylogeny from discrete morphological character data. *Syst. Biol.* 50, 913–925. doi: 10.1080/106351501753462876
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- Lischer, H. E. L., and Excoffier, L. (2012). PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics* 28, 298–299. doi: 10.1093/bioinformatics/btr642
- Lissambou, B.-J., Hardy, O. J., Atteke, C., Stevant, T., Dauby, G., Mbatshi, B., et al. (2018). Taxonomic revision of the African genus *Greenwayodendron* (Annonaceae). *Phytokeys* 114, 55–93. doi: 10.3897/phytokeys.114.27395
- LPWG (2017). A new subfamily classification of the Leguminosae based on a taxonomically comprehensive phylogeny. *Taxon* 66, 44–77.
- Maurin, O., Davies, T. J., Burrows, J. E., Daru, B. H., Yessoufou, K., Muasya, A. M., et al. (2014). Savanna fire and the origins of the ‘underground forests’ of Africa. *New Phytol.* 204, 201–214. doi: 10.1111/nph.12936
- Miller, M. A. (2009). *The CIPRES Portals. CIPRES. Arch. By WebCite(r)*. Available online at: https://www.webcitation.org/5imQJJeQa (accessed December, 2019).
- Miller, C. S., and Gosling, W. D. (2014). Quaternary forest associations in lowland tropical West Africa. *Quat. Sci. Rev.* 84, 7–25. doi: 10.1016/j.quascirev.2013.10.027
- Monthe, F. K., Migliore, J., Duminil, J., Bouka, G., Demenou, B. B., Doumenge, C., et al. (2019). Phylogenetic relationships in two African Cedreloideae tree genera (Meliaceae) reveal multiple rain/dry forest transitions. *Perspect. Plant Ecol. Evol. Syst.* 37, 1–10. doi: 10.1016/j.ppees.2019.01.002
- Morley, R. J. (2000). *Origin and Evolution of Tropical Rain Forests*. New York, NY: John Wiley & Sons.
- Nicotra, A. B., Chong, C., Bragg, J. G., Ong, C. R., Aitken, N. C., Chuah, A., et al. (2016). Population and phylogenomic decomposition via genotyping-by-sequencing in Australian *Pelargonium*. *Mol. Ecol.* 25, 2000–2014. doi: 10.1111/mec.13584
- Pan, A. D., Jacobs, B. F., and Herendeen, P. S. (2010). Detarieae *sensu lato* (Fabaceae) from the Late Oligocene (27.23 Ma) Guang River flora of north-western Ethiopia. *Bot. J. Linn. Soc.* 163, 44–54. doi: 10.1111/j.1095-8339.2010.01044.x
- Paradis, E., Claude, J., and Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20, 289–290. doi: 10.1093/bioinformatics/btg412
- Pennington, R. T., and Hughes, C. E. (2014). The remarkable congruence of New and Old World savannah origins. *New Phytol.* 204, 4–6. doi: 10.1111/nph.12996
- Pennington, R. T., and Lavin, M. (2016). The contrasting nature of woody plant species in different neotropical forest biomes reflects differences in ecological stability. *New Phytol.* 210, 25–37. doi: 10.1111/nph.13724
- Pickrell, J. K., and Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* e1002967. doi: 10.1371/journal.pgen.1002967
- Pinheiro, F., De Barros, F., Palma-Silva, C., Meyer, D., Fay, M. F., Suzuki, R. M., et al. (2010). Hybridization and introgression across different ploidy levels in the Neotropical orchids *Epidendrum fulgens* and *E. puniceoluteum* (Orchidaceae). *Mol. Ecol.* 19, 3981–3994. doi: 10.1111/j.1365-294x.2010.04780.x
- Plana, V. (2004). Mechanisms and tempo of evolution in the African Guineo-Congolian rainforest. *Philos. Trans. R. Soc. B Biol. Sci.* 359, 1585–1594. doi: 10.1098/rstb.2004.1535
- Pons, J., Barraclough, T. G., Gomez-Zurita, J., Cardoso, A., Duran, D. P., Hazell, S., et al. (2006). Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Syst. Biol.* 55, 595–609. doi: 10.1080/10635150600852011
- R Core Team (2017). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available online at: http://www.R-project.org/
- Rambaut, A. (2007). *FigTree, a Graphical Viewer of Phylogenetic Trees*. Available online at: http://tree.bio.ed.ac.uk/software/figtree/ (accessed March, 2020).
- Rambaut, A., and Drummond, A. J. (2016). *Tracer v 1.6*. Scotland: University of Edinburgh, 2007.
- Rognes, T., Flouri, T., Nichols, B., Quince, C., Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 4:e2584. doi: 10.7717/peerj.2584
- Sack, L., and Scoffoni, C. (2013). Leaf venation: structure, function, development, evolution, ecology and applications in the past, present and future. *New Phytol.* 198, 983–1000. doi: 10.1111/nph.12253
- Salzmann, U., and Hoelzmann, P. (2005). The Dahomey Gap: an abrupt climatically induced rain forest fragmentation in West Africa during the late Holocene. *Holocene* 15, 190–199. doi: 10.1191/0959683605hl799rp
- Sarnthein, M., and Fenner, J. (1988). Global wind-induced change of deep-sea sediment budgets, new ocean production and CO₂ reservoirs ca. 3.3–2.35 Ma BP. *Philos. Trans. R. Soc. London, Ser. A* 318, 487–504. doi: 10.1098/rstb.1988.0020
- Satabié, B. (1994). Biosystème et Vicariance dans la flore Camerounaise. *Bull. Jard. Bot. Nat. Belg.* 63, 125–170.
- Senut, B., Pickford, M., and Ségalen, L. (2009). Neogene desertification of Africa. *Comptes Rendus. Geosci.* 341, 591–602. doi: 10.1016/j.crte.2009.03.008
- Simon, M. F., Grether, R., de Queiroz, L. P., Skema, C., Pennington, R. T., and Hughes, C. E. (2009). Recent assembly of the Cerrado, a neotropical plant diversity hotspot, by in situ evolution of adaptations to fire. *Proc. Natl. Acad. Sci. U.S.A.* 106, 20359–20364. doi: 10.1073/pnas.0903410106
- Simon, M. F., and Pennington, R. T. (2012). The evolution of adaptations of woody plants in the savannas of the Brazilian cerrado. *Int. J. Plant Sci.* 173, 711–723.

- Soltis, D. E., Soltis, P. S., Schemske, D. W., Hancock, J. F., Thompson, J. N., Husband, B. C., et al. (2007). Autopolyploidy in angiosperms: have we grossly underestimated the number of species? *Taxon* 56, 13–30.
- Stamatakis A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033
- Stange, M., Sanchez-Villagra, M. R., Salzburger, W., and Matschiner, M. (2018). Bayesian divergence-time estimation with genome-wide single-nucleotide polymorphism data of sea catfishes (Ariidae) supports Miocene closure of the Panamanian Isthmus. *Syst. Biol.* 6, 681–699. doi: 10.1093/sysbio/syy006
- Tng, D. Y. P., Jordan, G. J., and Bowman, D. M. J. S. (2013). Plant traits demonstrate that temperate and tropical giant eucalypt forests are ecologically convergent with rainforest not savanna. *PLoS One* 8:e0084378. doi: 10.1371/journal.pone.0084378
- Tosso, F., Hardy, O. J., Doucet, J. L., Dainou, K., Kaymak, E., and Migliore, J. (2018). Evolution in the Amphi- Atlantic tropical genus *Guibourtia* (Fabaceae, Detarioideae), combining NGS phylogeny and morphology. *Mol. Phylogenet. Evol.* 120, 83–93. doi: 10.1016/j.ympev.2017.11.026
- Veranso-Libalah, M. C., Kadereit, G., Stone, R. D., and Couvreur, T. L. P. (2018). Multiple shifts to open habitats in Melastomateae (Melastomataceae) congruent with the increase of African Neogene climatic aridity. *J. Biogeogr.* 45, 1420–1431. doi: 10.1111/jbi.13210
- Wiens, J. J., and Donoghue, M. J. (2004). Historical biogeography, ecology, and species richness. *Trends Ecol. Evol.* 19, 639–644. doi: 10.1016/j.tree.2004.09.011
- Wood, T. E., Takebayashi, N., Barker, M. S., Mayrose, I., Greenspoon, P. B., and Rieseberg, L. H. (2007). The frequency of polyploid speciation in vascular plants. *Proc. Natl. Acad. Sci. U.S.A.* 106, 13875–13879. doi: 10.1073/pnas.0811575106
- Zachos, J. C., Dickens, G. R., and Zeebe, R. E. (2008). An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. *Nature* 451, 279–283. doi: 10.1038/nature06588
- Zhang, J., Kapli, P., Pavlidis, P., and Stamatakis, A. (2013). A general species delimitation method with applications to phylogenetic placements. *Bioinformatics* 29, 2869–2876. doi: 10.1093/bioinformatics/btt499
- Yule, G. U. (1925). A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis. F.R.S. *Philos. Trans. R. Soc. B Biol. Sci.* 213, 21–87. doi: 10.1098/rstb.1925.0002

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Donkpegan, Doucet, Hardy, Heuertz and Piñero. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.