

Nuclear ribosomal phylogeny of *Brachystegia*: new markers for new insights about rain forests and Miombo woodlands evolution

Arthur F. Boom^{1,2}, Jérémy Migliore^{1,3}, Esra Kaymak⁴, Pierre Meerts^{5,6,7}, Olivier J. Hardy¹

- 1 Université libre de Bruxelles, Faculté des Sciences, Service Evolution Biologique et Ecologie, Bruxelles, Belgium
- 2 Royal Museum for Central Africa, Biology Department, Section Vertebrates, Tervuren, Belgium
- 3 Muséum départemental du Var, Toulon, France
- 4 Okinawa Institute of Science and Technology, Okinawa, Japan
- 5 Université libre de Bruxelles, Faculté des Sciences, Service Ecologie végétale et Biogéochimie, Bruxelles, Belgium
- 6 Meise Botanic Garden, Meise, Belgium
- 7 Fédération Wallonie-Bruxelles, Service Général de l'Enseignement supérieur et de la Recherche scientifique, Brussels, Belgium

Corresponding author: Arthur F. Boom (boomarthur@gmail.com)

Academic editor: Isabel Larridon ♦ Received 17 November 2021 ♦ Accepted 09 May 2022 ♦ Published 12 August 2022

Abstract

Background and aims – *Brachystegia* is a species-rich tree genus found in tropical Africa and a typical element of Miombo woodlands, a widely distributed subtype of the Zambezi savanna. Plastid DNA was shown to be largely uninformative to assess species phylogenetic relationships due to widespread chloroplast capture among species. Here, we aim to assess the capacity of nuclear ribosomal DNA (rDNA) to clarify the phylogeny of *Brachystegia* species while accounting for intra-individual site polymorphisms (2ISPs), which are often present in rDNA and potentially phylogenetically informative.

Material and methods – Genome skimming sequencing on 47 samples representing 27 of the 29 currently recognized *Brachystegia* species, allowed us to retrieve complete nuclear ribosomal cistrons encoding for 18S, 5.8S, and 25S rRNA genes (35S rDNA). We reconstructed the *Brachystegia* phylogeny using Maximum Likelihood methods based on the standard substitution model or integrating 2ISPs (GENOTYPE implementation in RAXML-NG). We additionally tested the effect of partitioning the data (one partition for rDNA genes and one for the ITS1+ITS2). We also conducted network inferences (Neighbor-Net splits graph), as a strict bifurcative approach might not properly model topological uncertainty at shallow phylogenetic depth.

Key results – 2ISPs-aware and standard phylogenetic reconstructions are largely congruent. We identified several well-supported main clades clarifying the species relationships, including two clades of Miombo woodlands species. Miombo Group A includes species with ovoid to globose axillary dormant buds, while Miombo Group B species have flattened ones. Two morphologically close *Brachystegia* species (*B. kennedyi* and *B. leonensis*) found in Guineo-Congolian rain forests form also a robustly supported clade. 2ISPs coding allowed to identify an additional Guineo-Congolian clade (*B. eurycoma* and *B. nigerica*). Ribosomal DNA therefore proves more useful to explore the generic phylogeny than plastid DNA but the species relationships within and among the main clades remain poorly resolved, probably due to recent diversification and/or recurrent hybridization, so that the diversification of *Brachystegia* remains to be more properly characterised.

Conclusion – Nuclear and plastid phylogenetic reconstructions of *Brachystegia* species are discordant. Even if not well-resolved, rDNA phylograms and networks are characterized by taxonomic sorting, while we observe a strictly geographic sorting in the plastid dataset. Most of the species' relationships remain to be characterized using additional nuclear markers combined with in-depth morphological investigations.

Keywords

Brachystegia, genome skimming, phylogeny, RAxML-NG, ribosomal DNA, Zambezan woodlands, Miombo woodlands, 2ISPs

INTRODUCTION

Phylogenetic studies are needed to investigate the evolutionary processes that shaped the current African savannas (e.g. Maurin et al. 2014; Charles-Dominique et al. 2016; Davies et al. 2020). However, a critical evaluation of the phylogenetic history of typical Zambezan savanna flora elements is still lacking despite their wide distribution and key role in landscape dynamics through time (Linder 2014). The genus *Brachystegia* Benth. (Fabaceae, Detarioideae, Amherstieae in de la Estrella et al. 2018) constitutes one of the most iconic and dominant elements of the Zambezan Miombo woodlands (Frost 1996), which cover ca 2.7 million km² in southern, central, and eastern Africa (Campbell et al. 1996). Its 29 tree or shrub/suffrutex species (following Lebrun and Stork 2008) occur in Guineo-Congolian rain forests (eight species) or in Zambezan savannas and woodlands (21 species). *Brachystegia* is commonly known as one of the most taxonomically complex African tree genera (White 1962), with several species having blurred morphological boundaries (e.g. *B. boehmii*-*B. longifolia*, *B. bakeriana*-*B. spiciformis*, and the three species of the *B. tamarindoides* “complex” in Brummitt et al. 2007). Several species are additionally particularly variable in their morphology (e.g. *B. spiciformis*) and led to the description of varieties at a regional scale (Leonard et al. 1952, but see Brenan 1967; Brummitt et al. 2007). Hybridization is also suspected to occur, with up to 23 described putative hybrids (Palgrave 2002), but it was not demonstrated in a preliminary genetic investigation where putative hybrids were explained by morphological variation at the species scale (Palgrave 2002; Brummitt et al. 2007). The question remains open for Zambezan taxa, while there is no expected hybridization among Guineo-Congolian species (Palgrave 2002).

Plastid phylogeny

Boom et al. (2021) provided the first phylogeny of the genus *Brachystegia* by sequencing full plastomes but plastid haplotypes appeared shared among species at a local scale, suggesting widespread plastid capture (and thus, hybridization to some extent). Consequently, the plastid phylogeny reflects geographical rather than taxonomical affinities between samples, a pattern well documented for *Quercus*, *Fraxinus*, and *Macaranga* trees (e.g. Bänfer et al. 2006; Heuertz et al. 2006; Simeone et al. 2018). The plastid phylogeny delineates two main parapatric clades separating most of the Guineo-Congolian specimens from the Zambezan ones (Boom et al. 2021). The Zambezan clade is structured in three

parapatric clades, ranging from Tanzania to Angola (with eastern, central, and western subclades). The first cladogenesis event occurred during the late Miocene-Pliocene, while the different Zambezan plastid clades originated in the Pliocene-Pleistocene. Interestingly, a longitudinal gradient of time to the most recent common ancestor (TRMCA) is observed for the three Zambezan subclades and suggests a westward expansion of the Miombo woodlands, from an original range situated in East Africa. Although plastid clades have provided many insights into the biogeographic history of *Brachystegia*, the evolutionary relationships between the species remain to be characterised. To this end, phylogenetic information from the nuclear genome is required.

The 35S cistron

The nuclear-encoded 35S ribosomal DNA (rDNA) cistron, comprising the 18S, 5.8S, and 25S rRNA genes, is the most easily accessible nuclear DNA region using genome skimming, i.e. shallow genomic sequencing (Straub et al. 2012). However, interpreting variation in 35S rDNA sequences and their internal transcribed spacers (ITS1 and ITS2) can be challenging due to their mode of evolution (Feliner and Rosselló 2007). In plants, the nuclear-encoded 35S rDNA is located in one or several loci where hundreds of copies follow each other in arrays that tend to be homogeneous at the intra-individual scale due to concerted evolution (Eickbush and Eickbush 2007). Complete concerted evolution is however not universal (Bailey 2003). Hence, a certain degree of polymorphism can occur inside or between arrays, due to heterozygosity, homeology, or paralogy, including pseudogenes (array silencing, e.g. Volkov et al. 2017), causing Intra-Individual Sites Polymorphisms (2ISPs; Potts et al. 2014). When 2ISPs are not due to heterozygosity, they can sometime be fixed within a species, and therefore be phylogenetically informative. A potential trade-off between accuracy and simplicity in data analysis is to encode 2ISPs using IUPAC codes and to evaluate their putative impact on phylogenetic reconstructions. Potts et al. (2014) showed that treating 2ISPs as informative states rather than as ambiguous or missing characters could increase the resolution of phylogenetic reconstructions for highly polymorphic datasets (Potts et al. 2014, but see Fonseca and Lohmann 2020).

In addition to 2ISPs, other processes could blur the phylogenetic relationships between species at the nuclear genome, namely hybridization and incomplete lineage sorting (ILS) (Maddison 1997) or lack of divergence (e.g. Turner et al. 2016). Hybridization is suspected to occur among *Brachystegia* savanna species based on

morphological arguments (White 1962; Brenan 1967) but also given the evidence of recurrent plastid capture among species (Boom et al. 2021). ILS could also be likely, as dominant savanna tree species may have diversified recently and/or have large population sizes and long generation times, resulting in retention of ancestral polymorphism (Pennington and Lavin 2016). Finally, when DNA sequences are too short and/or evolve too slowly to accumulate enough mutations along the branches of a species tree, they offer limited resolution by lack of divergence.

Aims of this study

This work aims to investigate the diversification of *Brachystegia*, bridging the gap between genetic and morphological/taxonomic considerations through a multimarker phylogenetic approach. Namely, we use genome skimming on specimens from museum collections with degraded DNA (Zeng et al. 2018; Alsos et al. 2020) to assemble their nuclear ribosomal DNA, supplementing insights previously obtained from their plastomes (Boom et al. 2021). We will address the three following questions. Firstly, is the rDNA region phylogenetically informative for reconstructing the *Brachystegia* phylogeny? Secondly, does taking in account 2ISPs variability increase the resolution of our phylogenetic reconstructions? And thirdly, does rDNA provide complementary information to plastid DNA to infer the evolutionary history of the genus?

MATERIAL AND METHODS

Sampling and laboratory procedures

Brachystegia plant material (leaves) was collected on vouchers from the four following herbaria: BR, BRLU, FHO, and LISC (acronyms according to Index Herbariorum, Thiers continuously updated). In addition, we collected material ($n = 10$) from fieldwork and dried leaves using silica gel for DNA extraction purposes. In total, 47 individuals of 27 of the 29 described species of *Brachystegia* (following Lebrun and Stork 2008) were sequenced (Table 1). A sample of *Julbernardia paniculata* was added as outgroup to root the different reconstructed phylogenies. On both silica-dried and herbarium leaves, DNA extractions were performed using the DNeasy Plant Mini Kit (Qiagen, Netherlands) and the protocol detailed in Cappellini et al. (2010), except that the digestion step was performed overnight at 37°C rather than 55°C and we did not perform an initial wash step of the plant material with a bleach solution. DNA concentration and DNA size distribution were assessed with a Qubit® 2.0 Fluorometer (Thermo Fisher Scientific, USA) and with an electrophoresis on a 1% agarose gel. The different genomic libraries were prepared using the NEBNext Ultra II DNA Library Prep Kit (New England Biolabs, USA) and were

pooled equimolarly. After pooling, we sequenced the libraries on an Illumina NextSeq 500 instrument at the GIGA platform (Liège, Belgium), using the V2 mid-output reagent kit and targeting one million reads per library (2×150 paired-end reads).

Ribosomal DNA assembly and annotation

First, the quality of the genomic libraries was checked using FastQC v.0.11 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), and then they were trimmed to remove low-quality regions and Illumina adapters using Trim Galore! v.0.4.5 (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). De novo assemblies of the rDNA sequences were done for each sample using GetOrganelle v.1.6.2 (default parameters for ribosomal assembly; Jin et al. 2020). Resulting graphs were inspected using Bandage v.0.8.1 (Wick et al. 2015). The rDNA sequences were annotated using Infernal cmscan (Madeira et al. 2019) to delineate 18S rDNA, internal transcribed spacer 1 (ITS1), 5.8S rDNA, internal transcribed spacer 2 (ITS2), and 25S rDNA regions of the 35S cistron. Reads were finally mapped on their respective assembly using the Burrows-Wheeler Aligner BWA mem v.0.7.12 (Li and Durbin 2009), and depth coverage was computed using Samtools v.1.9 (Li et al. 2009). We polished the assembled rDNA sequences and identified 2ISPs positions by conducting a consensus calling for each individual with the following parameters: min-var-freq 0.01, --min-avg-qual 30, --min-freq-for-hom 0.9, and --min-coverage 3, using Varscan v.2.3.7 (Koboldt et al. 2012). Polymorphic sites were coded using IUPAC recommendations. Additionally, indels were coded as N. The polished sequences were deposited in GenBank (Table 1). The 18S–25S rDNA were aligned using MAFFT v.7 (Katoh et al. 2019) and the alignment was visually checked using MEGA7 (Kumar et al. 2016). The alignment is available in Supplementary file 1.

Phylogenetic reconstructions

In order to evaluate the impact of 2ISPs, we performed Maximum Likelihood (ML) based phylogenetic inferences with RAXML-NG software using three different implementations (Kozlov et al. 2019). (i) All 2ISPs were recoded as missing data (ML-N for missing data ‘N’). (ii) Binary 2ISPs were treated as ambiguous with both states of the coded ambiguity considered equiprobable, while 2ISPs involving three nucleotides were coded as missing data, using a GTR+I+G substitution model (ML-A for ‘Ambiguous’; Potts et al. 2014). (iii) Finally, we considered each 2ISP genotype as a transitional state, with the transitional probability included as additional parameter into the substitution model. This approach (ML-I for ‘Informative’) followed a comparable logic as in Potts et al. (2014), except that we preferred the new GENOTYPE implementation in RAXML-NG, using the GTGTR4+I+G substitution model (see Kozlov et al. 2022: supplementary

Table 1. Specimens used for the rDNA phylogeny. Most of them were also included in the plastid phylogeny of Boom et al. (2021) with two additional specimens of *Brachystegia utilis*, one of *B. mildbraedii*, and one of *B. wangermeanna*. Specimens are associated with herbarium vouchers, except the silica-dried sample ABo0065 (*B. utilis*), conserved in the silica-dried African plant leaves collection of Olivier Hardy (Université libre de Bruxelles, Belgium). The vouchers are hosted at BR, BRLU, FHO, and LISC. Two samples sequenced in Boom et al. (2021) were not included in this study due to taxonomic uncertainty (specimens *Boom* 38 attributed to *B. boehmii* and *Duvigneaud & Timperman 2317* attributed to *B. tamarindoides*).

No	Taxon	Voucher	Coordinates	Country	GenBank
1	<i>B. allenii</i> Hutch. & Burt Davy	<i>Milne-Redhead & Taylor 7663</i> (FHO)	-10.689, 38.945	Tanzania	OK335216
2	<i>B. allenii</i> Hutch. & Burt Davy	<i>White 2406A</i> (BR)	-14.960, 30.246	Zambia	OK335217
3	<i>B. angustistipulata</i> De Wild	<i>Jefford, Juniper & Newbould 2799</i> (BR)	-6.000, 30.000	Tanzania	OK335218
4	<i>B. bakeriana</i> Hutch. & Burt Davy	<i>Dechamps, Murta & da Silva 1327</i> (BR)	-14.817, 18.633	Angola	OK335219
5	<i>B. boehmii</i> Taub.	<i>Procter 262</i> (FHO)	-4.832, 29.962	Tanzania	OK335220
6	<i>B. boehmii</i> Taub.	<i>Duvigneaud 2833</i> (BRLU)		DR Congo	OK335221
7	<i>B. bussei</i> Harms	<i>White 2410</i> (BR)	-14.726, 30.762	Zambia	OK335222
8	<i>B. bussei</i> Harms	<i>Burt 4736</i> (BR)	-6.041, 37.519	Tanzania	OK335223
9	<i>B. cynometroides</i> Harms	<i>Forest Product Research Laboratory n/a</i> (FHO, collection date 6 Nov. 1969)		Cameroon	OK335224
10	<i>B. eurycoma</i> Harms	<i>Latilo & Daramola 28945</i> (BR)	7.710, 11.480	Nigeria	OK335225
11	<i>B. eurycoma</i> Harms	<i>Chapman 156</i> (FHO)	7.230, 10.628	Nigeria	OK335226
12	<i>B. floribunda</i> Benth.	<i>Barbosa 11037A</i> (LISC)	-12.179, 17.242	Angola	OK335227
13	<i>B. floribunda</i> Benth.	<i>Boom 41</i> (BRLU)	-11.530, 27.467	DR Congo	OK335228
14	<i>B. gossweileri</i> Hutch. & Burt Davy	<i>Barbosa 10988</i> (FHO)	-10.735, 14.981	Angola	OK335229
15	<i>B. gossweileri</i> Hutch. & Burt Davy	<i>Mendes dos Santos 1980</i> (FHO)	-12.148, 18.090	Angola	OK335230
16	<i>B. kennedyi</i> Hoyle	<i>Meikle & Keay 581</i> (BR)	6.105, 5.893	Nigeria	OK335231
17	<i>B. kennedyi</i> Hoyle	<i>Kennedy 2181</i> (FHO)	6.105, 5.893	Nigeria	OK335232
18	<i>B. laurentii</i> (De Wild.) Louis ex Hoyle	<i>Wieringa 4529</i> (BR)	-0.974, 10.925	Gabon	OK335233
19	<i>B. leonensis</i> Hutch. & Burt Davy	<i>Sesay 51</i> (BR)	8.913, -11.728	Sierra Leone	OK335234
20	<i>B. leonensis</i> Hutch. & Burt Davy	<i>Jongkind 9067</i> (BR)	5.646, -8.135	Liberia	OK335235
21	<i>B. longifolia</i> Benth.	<i>Boom 37</i> (BRLU)	-10.915, 28.517	DR Congo	OK335236
22	<i>B. longifolia</i> Benth.	<i>Dechamps, Murta & da Silva 1400</i> (BR)	-11.983, 18.283	Angola	OK335237
23	<i>B. longifolia</i> Benth.	<i>Boom 39</i> (BRLU)	-11.530, 27.466	DR Congo	OK335238
24	<i>B. manga</i> De Wild.	<i>Groome & Hoyle 1073</i> (FHO)	-7.623, 33.403	Tanzania	OK335239
25	<i>B. manga</i> De Wild.	<i>Duvigneaud 1214</i> (BR)	-11.187, 27.905	DR Congo	OK335240
26	<i>B. michelmorei</i> Hoyle	<i>Astle 797</i> (FHO)	-9.796, 29.295	Zambia	OK335241
27	<i>B. microphylla</i> Harms	<i>Leippert 6334</i> (BR)	-4.485, 35.758	Tanzania	OK335242
28	<i>B. mildbraedii</i> Harms	<i>Wieringa 5552</i> (BR)	-1.435, 10.478	Gabon	OK335243
29	<i>B. nigerica</i> Hoyle & A.P.D.Jones	<i>Chesters A124/30</i> (BR)	6.155, 6.770	Nigeria	OK335244
30	<i>B. nigerica</i> Hoyle & A.P.D.Jones	<i>Lapido 19061</i> (FHO)	7.134, 3.840	Nigeria	OK335245
31	<i>B. puberula</i> Hutch. & Burt Davy	<i>Bamps, Martins & Matos 4473</i> (BR)	-14.217, 14.033	Angola	OK335246
32	<i>B. russelliae</i> I.M.Johnst.	<i>Mendes 55</i> (BR)		Angola	OK335247
33	<i>B. russelliae</i> I.M.Johnst.	<i>Mendonça 4593</i> (FHO)	-12.476, 16.295	Angola	OK335248
34	<i>B. spiciformis</i> Benth.	<i>Liben 1742</i> (BR)	-5.863, 23.392	DR Congo	OK335249
35	<i>B. spiciformis</i> Benth.	<i>Boom 61</i> (BRLU)	-11.488, 27.600	DR Congo	OK335250
36	<i>B. spiciformis</i> Benth.	<i>Duvigneaud & Timperman 242 B2</i> (BRLU)	-10.598, 22.345	DR Congo	OK335251
37	<i>B. spiciformis</i> Benth.	<i>Barbosa, Henriques & Moreno 2164</i> (BRLU)	-14.831, 13.621	Angola	OK335252

Table 1. (continued) Specimens used for the rDNA phylogeny. Most of them were also included in the plastid phylogeny of Boom et al. (2021) with two additional specimens of *Brachystegia utilis*, one of *B. mildbraedii*, and one of *B. wangermeana*. Specimens are associated with herbarium vouchers, except the silica-dried sample ABo0065 (*B. utilis*), conserved in the silica-dried African plant leaves collection of Olivier Hardy (Université libre de Bruxelles, Belgium). The vouchers are hosted at BR, BRLU, FHO, and LISC. Two samples sequenced in Boom et al. (2021) were not included in this study due to taxonomic uncertainty (specimens Boom 38 attributed to *B. boehmii* and *Duvigneaud & Timperman 2317* attributed to *B. tamarindoides*).

No	Taxon	Voucher	Coordinates	Country	GenBank
38	<i>B. stipulata</i> De Wild.	Boom 7 (BRLU)	-11.510, 28.007	DR Congo	OK335253
39	<i>B. tamarindoides</i> Welw. ex Benth.	Torre 8678 (LISC)	-15.096, 13.565	Angola	OK335254
40	<i>B. taxifolia</i> Harms	Boom 24 (BRLU)	-11.477, 27.662	DR Congo	OK335255
41	<i>B. taxifolia</i> Harms	Boom 46 (BRLU)	-11.533, 27.463	DR Congo	OK335256
42	<i>B. taxifolia</i> Harms	Duvigneaud 3614br2 (BRLU)	-12.019, 27.784	DR Congo	OK335257
43	<i>B. torrei</i> Hoyle	Torre & Paiva 11521 (LISC)	-15.587, 39.614	Mozambique	OK335258
44	<i>B. utilis</i> Hutch. & Burt Davy	Boom silica collection ABo0065 (BRLU)	-11.515, 28.067	DR Congo	OK335260
45	<i>B. utilis</i> Hutch. & Burt Davy	Salubeni & Chikuni 6642 (FHO)	-12.069, 33.588	Malawi	OK335259
46	<i>B. wangermeana</i> De Wild.	Boom 10 (BRLU)	-11.516, 28.008	DR Congo	OK335261
47	<i>B. wangermeana</i> De Wild.	Plancke 154/2025 (BRLU)	-10.693, 23.182	DR Congo	OK335262
48	<i>J. paniculata</i> (Benth.) Troupin	Boom 51 (BRLU)	-11.432, 27.469	DR Congo	OK335215

note 2). For each method, branch supports were computed based on 1000 bootstrap replicates and are provided as nonparametric bootstrap support (BS).

Mutation patterns and rates differ between the rRNA genes and the internal transcribed spacers (Baldwin et al. 1995). To assess the robustness of our phylogenetic reconstructions, we also used RAxML-NG with the different implementations described above but using two partitions (ITS1+ITS2 vs all rRNA genes; GTR+I+G model for each partition in ML-N and ML-A; GTGTR4+I+G model for ML-I). We also evaluated how the different main subunits of the 35S cistron are phylogenetically informative and/or might provide conflicting signals by conducting separate RAxML-NG analyses on each region (ITS1, ITS2, 18S, and 25S) with the GTGTR4+I+G substitution model. Finally, at shallow phylogenetic levels, we can expect non-tree-like evolutionary patterns (i.e. not explained by strictly bifurcating trees) due to reticulation, low-level genetic divergence and ILS. We therefore also applied network approaches and computed a Neighbor-Net splits graph using p-distance (i.e. 2ISPs aware approach) computed with the R package phangorn v.2.5.5 (Schliep 2011; Schliep et al. 2017) and SplitsTree5 (Huson and Bryant 2006). We also computed a bootstrap consensus network (e.g. Schliep et al. 2017) using the ML-I bootstraps trees (ML-I 1 partition, 1000 BS trees; edge weights: tree size weighted mean; Threshold: 15).

Finally, to evaluate the ability of rDNA to retrieve clades with taxonomic significance, we assessed whether sequences divergence was correlated to taxonomy and/or geographical distance using Mantel tests computed with the R package vegan v.2.5-6 (Oksanen et al. 2019). For the different clades identified in the reconstructed phylogenies, we searched for correlation between matrices of genetic distances, computed with the R package ape v.5.4-1 (raw distance; Paradis and Schliep

2019), and taxonomic distances (0 and 1 indicating whether two individuals belong to the same species or not). Additionally, as geographic and genetic distance correlation can provide information regarding the geography of diversification (Abellán and Ribera 2017) and/or events of hybridization at a local scale (e.g. Boom et al. 2021), Mantel tests were also performed between geographic (i.e. shortest distances computed by the R package geosphere v.1.5-10; Hijmans 2019) and genetic distances (R Core Team 2019).

RESULTS

Ribosomal assemblies and alignment

After the trimming step, we obtained a mean of 1,227,134 reads per library (SD = 642,412). Raw sizes of the different rDNA assemblies ranged between 5,948 bp and 9,048 bp (mean = 7,526, SD = 757), and the length of the extracted 18S–ITS1–5.8S–ITS2–25S sequences ranged between 5,823 bp and 5,843 bp (mean = 5,828 bp). Mean depth coverage per sample ranged between 49X and 1,096X (mean = 260X, median = 192X). This wide range in mean depth coverage was also reflected in terms of relative rDNA reads content, i.e. between 0.19% and 3.35% of the total number of reads for each genomic library mapped on the corresponding assembly (mean = 0.81, SD = 0.66). The final alignment included a total of 5,973 sites and contained 166 variable and 78 parsimony-informative sites, when 2ISPs were not taken into account. When the outgroup *J. paniculata* was removed, these values drop to 108 and 75 variable and parsimony-informative sites, respectively. Taking into account the exclusive 2ISP sites, i.e. sites characterised by a single-defined nucleotide with ambiguity codes, the number of variable sites rose to

394 and 338, respectively with and without *J. paniculata*. Details regarding the exact base composition of each sequence are available in Supplementary file 2.

Phylogenetic analyses

ML-N and ML-A phylogenetic reconstructions supported similar groupings, with minor differences (more details in Supplementary file 3: Fig. S1). For clarity, we only present the ML-A and ML-I cladograms, i.e. without branch lengths, in the main article (Fig. 1). The ML-N, ML-A, and ML-I phylograms are available in Supplementary file 3 (Figs S2, S3, and S4). Both ML-A and ML-I inferences provided topologies with low bootstrap (BS) supports for most branches (Fig. 1). Additionally, among the 16 species represented by at least two individuals, only for four (ML-A) or six (ML-I) of those species, the different individuals were placed together. At the intra-generic level, we delineated four robustly supported clades that form coherent taxonomic, ecological, and/or spatial entities. One to two clades of rain forest species (*B. eurycoma*-*B. nigerica* and *B. kennedyi*-*B. leonensis* clades) and two clades of Miombo species are well supported (BS = 75–100

for ML-A, 84–99 for ML-I; Fig. 1). The positions of three species, *B. cynometroides*, *B. laurentii*, and *B. mildbraedii*, remain unresolved, even if the ML-I analysis suggested a *B. cynometroides*-*B. laurentii* grouping (BS = 64). The two identified Miombo clades are morphologically distinct according to floral, leaf, and axillary dormant bud features, corresponding roughly to the morphogroups A and B defined in the Flora of Tropical East Africa (Brenan 1967). The Miombo Group A clade includes *B. bakeriana*, *B. bussei*, *B. floribunda*, *B. manga*, *B. microphylla*, *B. puberula*, *B. spiciformis*, *B. tamarindoides*, *B. torrei*, and *B. utilis*, whereas the Miombo Group B clade includes *B. allenii*, *B. angustistipulata*, *B. boehmii*, *B. gossweileri*, *B. longifolia*, *B. russelliae*, *B. stipulata*, *B. taxifolia*, and *B. wangermeeana*. Except for *B. russelliae* in both the ML-A and ML-I method and for *B. boehmii* in ML-I, none of the Miombo species was sorted taxonomically. In addition, three species with similar leaf morphology (i.e. *B. bakeriana*, *B. floribunda*, and *B. spiciformis*) are placed together but with limited support in the three different implementations (BS = 40 for ML-N, 44 for ML-A, 39 for ML-I).

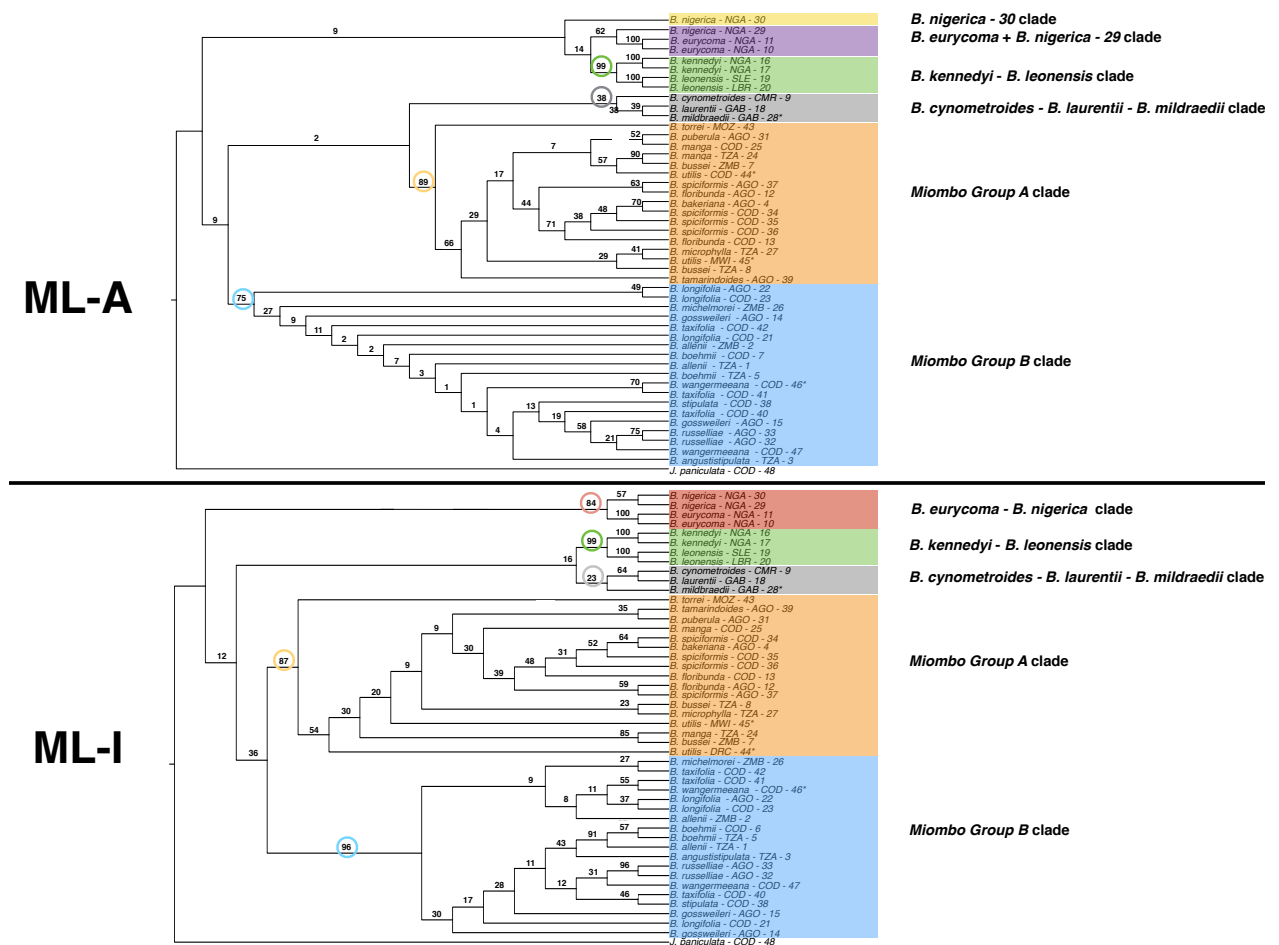


Figure 1. Maximum Likelihood phylogenetic inferences of *Brachystegia* species using rDNA sequences and two different coding schemes. Cladograms were produced using RAxML-NG software (Kozlov et al. 2019) and intra-individual site polymorphisms (2ISPs) were coded following the IUPAC nomenclature. 2ISPs are either considered as ambiguous (ML-A) or are coded as state into the substitution model (ML-I). Bootstrap supports (BS) are indicated on each branch.

Partitioning the data (ITS1+ITS2 and rDNA genes) did not produce major differences that are supported with high bootstrap values (Supplementary file 3: Figs S5, S6, and S7). We again observed three to four robustly supported clades, low BS values for most of the branches, and individuals from the same species were only placed together in four or six cases depending on the RAxML-NG implementation used (Supplementary file 3: Figs S5, S6, and S7).

The Neighbor-Net splits (Fig. 2) and the bootstrap consensus network graph (Supplementary file 3: Fig. S8) also identified four main clusters corresponding to the four main clades (i.e. Miombo Group A, Miombo Group B, *B. eurycoma*-*B. nigerica*, and *B. kennedyi*-*B. leonensis*). ITS2 was the most informative subunit as it identified the four main clades, contrary to the other subunits, although the bootstrap support values were moderate to low (BS = 19–65; Fig. 2). The least informative subunit was the 18S rDNA gene, where *B. eurycoma*-*B. nigerica* is the only main clade found, with moderate support (BS = 60; Fig. 2). The individual trees are available in Supplementary file 3 (Figs S9–S12).

The correlation between the genetic and taxonomic distances was significant for the Miombo species as a whole ($r = 0.19$, p value = 0.001, $n = 35$) but not within Miombo Group A ($r = 0.11$, p value = 0.138, $n = 17$) or Miombo Group B ($r = -0.069$, p value = 0.219, $n = 18$). Hence, rDNA discriminated between different Miombo taxonomic groups but did not provide fine-scale taxonomic information. The correlation between geographical and genetic distances was non-significant for Miombo species as a whole ($r = 0.12$, p value = 0.028, $n = 35$) and for Miombo Group B ($r = -0.02$, p value = 0.551, $n = 18$) but was significant within Miombo Group A ($r = 0.34$, p value = 0.001, $n = 17$). The mean genetic distance between individuals was higher in Miombo Group A than in Miombo Group B (i.e. differences in the nucleotide composition of 0.03% and 0.0055%, respectively) for roughly the same number of species ($n = 10$) and specimens ($n = 17$ – 18). Overall, the number of parsimony-informative sites within Miombo Group A and B was low ($n = 17$ in both clades).

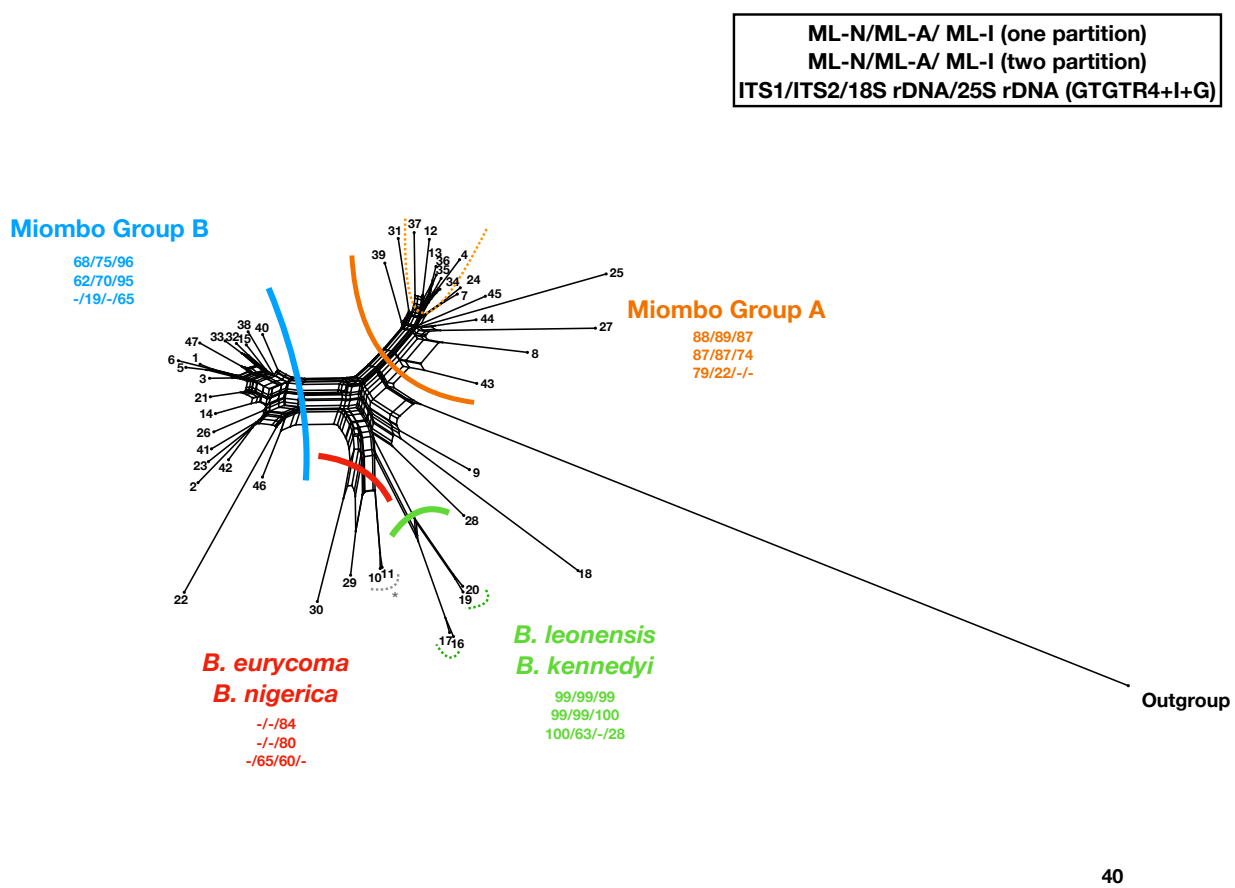


Figure 2. Network linking rDNA *Brachystegia* sequences (Neighbor-Net approach – p-distance). The four main clades identified in the different Maximum Likelihood (ML) phylogenetic inferences are delineated with thick coloured lines. The BS support values for these clades are given according to the different ML analyses (ML-N, ML-A, ML-I; one and two partitions; ML-I for ITS1, ITS2, 18S, and 25S). In Miombo Group A, the specimens of *B. bakeriana*, *B. floribunda*, and *B. spiciformis* are clustered together (delineated with the dotted orange line).

2ISPs coding and taxonomic consistency

Explicitly taking into account the retention of polymorphism allows for the detection of clades and species groups that align with morphology. For instance, regardless of the partitioning scheme (one vs two partitions), ML-N placed the specimen *B. nigerica* 30 (Lapido 19061, Nigeria) as poorly supported sister to the *B. kennedyi*-*B. leonensis* clade (BS = 22–28), while the specimen *B. nigerica* 29 (Chesters A124/30, Nigeria) grouped with *B. eurycoma* (BS = 61–63). In ML-A, the specimen *B. nigerica* 30 moved to the *B. eurycoma*-*B. nigerica* subtree but with decreased support (BS = 9). With two partitions, however, the specimen *B. nigerica* 30 remained sister, with low support, to the *B. kennedyi*-*B. leonensis* clade (BS = 25). In contrast, in ML-I, regardless of the partitioning scheme, both specimens of *B. nigerica* were resolved as sisters with moderate support (BS = 53–57) in a well-supported clade composed of *B. eurycoma* and *B. nigerica* (BS = 80–84).

Ribosomal and plastid topology

As described in Boom et al. (2021), the plastid phylogeny delineates clades that correspond to geographic regions (Fig. 3). Five main plastid clades and two additional singleton lineages delineated seven regions. Two clades are part of a larger rain forest (RF) clade (Fig. 3; pentagon = Upper Guinea and Southwest Nigeria region; diamond = Southeast Nigeria-Cameroon region). The RF sister clade encompasses two basal lineages (Fig. 3; triangle = the Eastern Arc Mountains and surroundings; reverse triangle = Lower Guinea) in addition to three parapatric Miombo woodlands (MW) clades (Fig. 3; circle = Eastern; star = Central; square = Western). *Brachystegia* specimens in Miombo regions share plastid sequences from different clades, regardless of the rDNA clades to which they belong (most specimens from rDNA Miombo A and B clades possess plastid sequences from one of the three parapatric Miombo clades). Nuclear ribosomal DNA is therefore sorted according to the taxonomy (to some extent), while

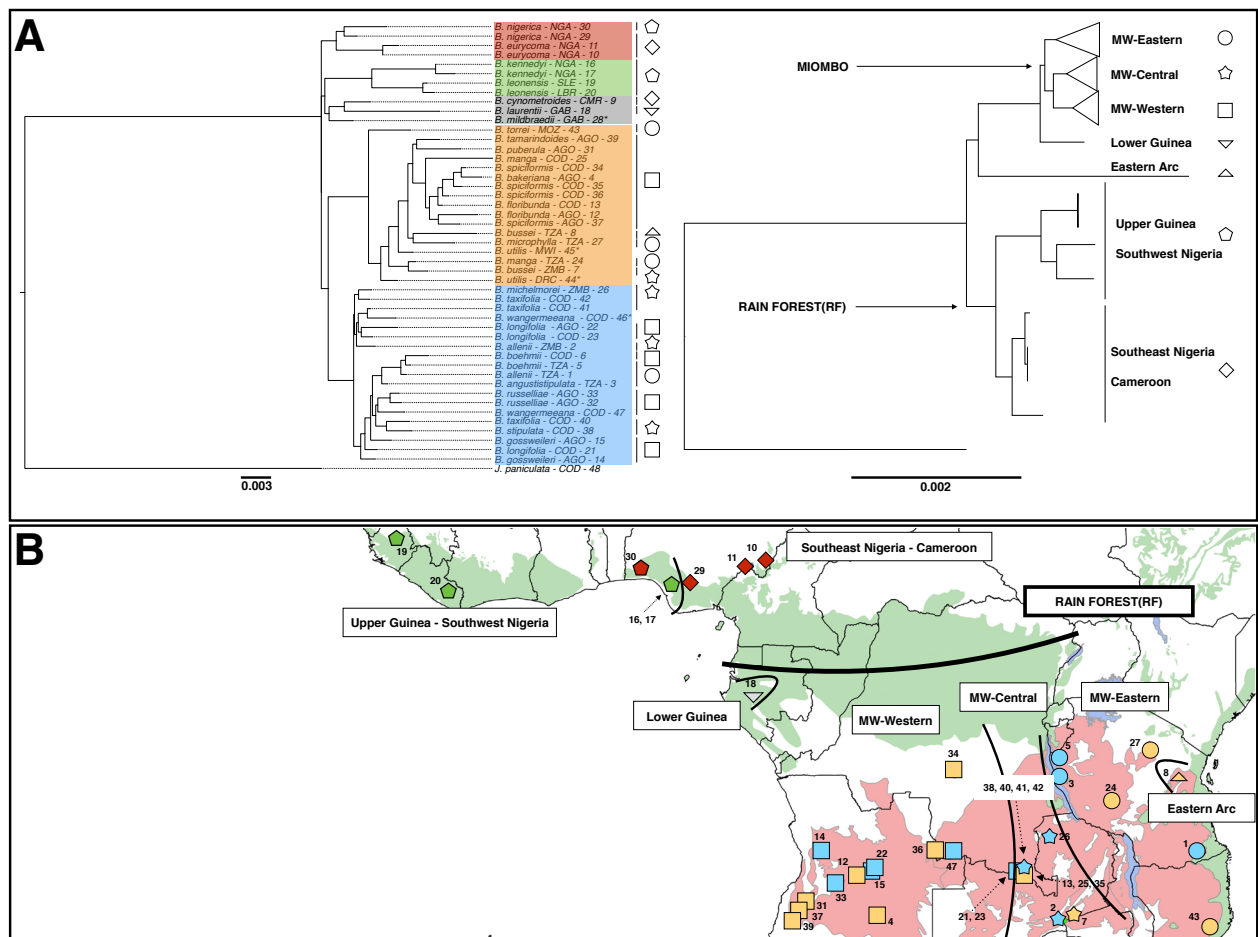


Figure 3. Comparison between the ribosomal (subfigure A, left; ML-I tree) and plastid (subfigure A, right) phylogenies of *Brachystegia* specimens, together with the geographic distribution of the specimens (subfigure B) in the miombo woodlands (MW, in red) and the African tropical rain forests (RF, in green). Four specimens labelled with * are present in the rDNA phylogram but absent in the plastid tree. The plastid phylogeny delineates five geographically coherent clades and two additional singletons (each represented by different shapes in each subfigure), independently of the clades delineated by rDNA. Bootstraps supports are given in Fig. 1 for the rDNA phylogram and are above 98 for all branches of the plastid phylogram.

plastid DNA is geographically structured. This supports cytoplasmic genome exchanges between all the different Miombo species, including between Miombo Groups, as well as between RF species. For example, the specimen *B. nigerica* 30 has a plastid sequence that is very similar to those in the geographically nearby specimens *B. kennedyi* 17 and 18 (Upper Guinea and Southwest Nigeria clade), while *B. nigerica* 29 plastome is part of the Southeast Nigeria-Cameroon region plastid clade including the two geographically nearby *B. eurycoma* specimens (see Table 1). In contrast, the ML-I rDNA phylogeny supports the specimen *B. nigerica* 30 as being part of the *B. nigerica*-*B. eurycoma* clade, suggesting at least one event of *B. kennedyi* plastid capture by *B. nigerica* in Southwest Nigeria.

DISCUSSION

The contribution of ribosomal DNA to decipher the evolutionary history of *Brachystegia*

The sorting of individuals according to their taxonomic species in the rDNA trees is, for the Miombo woodland species, the exception irrespective of the treatment of 2ISPs. However, reciprocal monophyly is supported for several Guineo-Congolian species and for the morphologically distinct Miombo Group A vs Miombo Group B. Correlation between geographic and genetic distances for Miombo Group A could suggest some inter-species gene flow at local scale but could not rule out past allopatric diversification. On the other hand, such correlation was not found in the Miombo Group B. Altogether, rDNA sequences provide insights regarding the *Brachystegia* evolutionary history, even if most of the relationships between species and clades remained unresolved due to insufficient phylogenetic information. This lack of discriminating signal between species could be due to the fairly recent origin of the different species (Boom et al. 2021), which could have evolved in parallel with the origin and expansion of C4 fire prone savannah during the late Miocene-Pliocene-Pleistocene (Maurin et al. 2014; Polissar et al. 2019). The relationships between the four main clades of *Brachystegia* and several of the Guineo-Congolian species (e.g. *B. cynometroides*, *B. mildbraedii*, *B. laurentii*) remain unresolved here, possibly because these lineages diverged in a short time (i.e. rapid radiation).

Nuclear ribosomal DNA also retrieved clades that are congruent with the taxonomy. The formerly recognized Miombo A and B groups in F.T.E.A. (Brenan 1967) (hereinafter “morphological Group A and B”) seem to represent two monophyletic groups, consistently supported as clades in our different reconstructions. Being established for East African species, the infrageneric system of Brenan (1967) does not include per se all Zambezian woodland species. Namely, four western Miombo species (*B. bakeriana*, *B. gossweileri*, *B. tamarindoides*, *B. russelliae*) and three narrowly

distributed species (*B. michelmorei*, *B. oblonga*, *B. torrei*) were not covered by Brenan (1967). However, these seven species were covered by the Flora Zambesiaca (Brummitt et al. 2007). They show clear morphological affinities with species included in Brenan (1967). *Brachystegia tamaridoides* and *B. torrei* share traits with *B. microphylla*, while *B. bakeriana* has many traits in common with *B. spiciformis*. On the other hand, *B. gossweileri*, *B. russelliae*, and *B. michelmorei* are, based on buds/flowers/stipules/auricles, part of the morphological Miombo Group B. Miombo Group species exhibit different vegetative and floral morphological trends (Brenan 1967). Morphological Group A species have globoid or ovoid buds, while species from morphological Group B have flattened buds enclosed in two large keeled scales. Morphological Group B species mostly have persistent stipules with basal reniform auricles, while in morphological Group A species stipules are generally caducous, mostly without auricles. Morphological differences are also reported for the bark, with relatively thick bark with vertical furrows in morphological Group B species. Other vegetative traits that have taxonomic value (e.g. the number and dimension of leaflets pairs) do not appear to show consistent differences between the two morphological groups. Regarding floral traits, tepals are absent, reduced, or shortly ciliate in morphological Group A species vs densely and long-ciliate in morphological Group B species. Most *Brachystegia* species have paniculate inflorescence, with the notable exception of three Miombo species: *B. stipulata* (morphological Group B) has raceme and/or paniculate inflorescences, while *B. bakeriana* and *B. spiciformis* species (morphological Group A) have racemose inflorescences (Brummitt et al. 2007). The latter two species are not always easy to discriminate in herbarium material (Brummitt et al. 2007), even if they differ in their habit, fruit, and bark traits (*B. bakeriana* is a shrub, while *B. spiciformis* is a tall tree) (White 1962; Brummitt et al. 2007). Sterile material of these species can be difficult to separate from *B. floribunda* (Brummitt et al. 2007). In the different implementations to reconstruct an rDNA tree (ML-A, ML-N, and ML-I), we systematically retrieved a moderate/low supported clade that includes individuals from the three aforementioned species. Such clade can reflect a close evolutionary relationship between these species. It is somehow unexpected, as *B. floribunda* is morphologically closer to other species from the Miombo Group A (i.e. several sepaloid/shortly ciliates tepals and panicles for *B. floribunda* and other Miombo group A species, while tepals are reduced or absent in other species). Alternatively, we cannot exclude that some of the specimens were not correctly identified and might contribute to this apparent and unexpected species cluster. Preliminary results using additional nuclear markers and a denser sampling suggests that the identification of *B. floribunda* specimens might be problematic (Boom 2021).

Apart from Zambezian species, the relationship between *B. leonensis* and *B. kennedyi* is congruent with the views of Hoyle (1955), who recognized close affinities

of leaf anatomy, with however substantial differences in floral traits.

Plastid and rDNA provide complementary insights on the evolution of *Brachystegia*

The rDNA phylogeny (Fig. 2) is in sharp contrast with the recently published plastid gene tree showing a strict geographical sorting (Boom et al. 2021). The large number of plastid introgression events observed in the genus resulting in large spatial clusters supports at least some interspecific gene flow within and even between the major groups. Hybrids are suspected to occur within each Miombo woodland group, while morphological intermediates between the two groups were rarely observed (Brenan 1967), in agreement with the rDNA data. Correlation between genetic and spatial distances was not found for Miombo Group B, and uncertainty on the exact reason for correlation in Miombo Group A prevents us to formally designate hybridization as the main driver of rDNA genetic diversity distribution among species. This may reflect allopatric speciation, hybridization, or even unbalanced sampling. However, low within-group resolution prevents us from pinpointing nuclear introgression between species.

Phylogenetic information provided by 2ISPs in rDNA sequences

Properly evaluating the effect of coding intra-individual polymorphisms is out of the scope of this paper. However, we note that considering the polymorphism as recommended in Potts et al. (2014) allowed to identify clades that are congruent with morphological observations and increase support of critical branches. Therefore, 2ISPs could contain valuable phylogenetic information, as in recognizing the *B. eurycoma* and *B. nigerica* grouping, which is significant given their frequency in rDNA sequences (we found about three times more variable sites among *Brachystegia* samples when including 2ISPs). The exact mechanisms producing 2ISPs remain to be characterized. Here, it is important to highlight that we mainly focused on tree topologies, and not on the other properties of phylogenetic trees, i.e. branch lengths. Works on *Scrophularia* and Bignoniaceae highlighted the strong impact of ambiguities coding schemes on branch lengths (Scheunert and Heubl 2017; Fonseca and Lohmann 2019) but they did not apply the here used RAXML-NG ML-I implementation. As seen in Supplementary file 3: Figs S2–S4, with our data, the treatment of 2ISPs did not substantially affect branch lengths despite a substantial impact on topology.

Perspectives: genomics, morphology, and species delineation

Overall, the rDNA gene trees in this study shed some light onto part of the evolutionary history of *Brachystegia*

species but they cannot resolve relationships between closely related species because the assembled 18S–25S rDNA data is not sufficiently divergent and thus not sufficiently informative. The lack of resolution between Miombo species could be explained mainly by the non-mutually exclusive following reasons: a recent evolutionary history of diversification and/or the occurrence of gene flow between the extant species.

If due to recent diversification, the proper characterisation of a species tree may need the use of several unlinked loci. Targeted enrichment could constitute an interesting strategy, as such methods allow the investigation of young and species rich genera (e.g. *Inga* in Nicholls et al. 2015). Moreover, specific baits already exist for the Detarioideae subfamily and already proved their potential to unravel species trees in recent genera, allowing a proper characterization of both evolutionary history and taxonomy (Ojeda et al. 2019; de la Estrella et al. 2020). Alternatively, Genotyping by Sequencing (GBS) or restriction-site associated DNA sequencing (RAD-seq) have been proven useful in order to evaluate species trees for plants, even in presence of polyploidy, hybridization, and incomplete lineage sorting (e.g. *Azalia* in Donkpegan et al. 2020; *Cycnoches* in Pérez-Escobar et al. 2020; *Quercus* in Hipp et al. 2020). Errors of identification in the diagnosed specimens (e.g. confusion between *B. bakeriana*, *B. floribunda*, and *B. spiciformis*) could be identified by such approaches, in addition by using a denser sampling for the problematic species (e.g. *Inga* in Dexter et al. 2010).

If the lack of resolution is due to reticulate history, the characterization of the nature, directionality, and extent of gene flow could be explored using targeted enrichment (e.g. *Brownea* in Schley et al. 2020). Moreover, the presence of interspecific gene flow could trigger a wider reflection on the species delineation within *Brachystegia*. The particularly weak nuclear genetic distances observed among the Miombo Group B individuals could potentially be explained by over-taxonomisation. Several species from the Miombo Group B form a morphological closely-related series with many intermediates (i.e. *B. allenii*, *B. angustistipulata*, *B. boehmii*, *B. longifolia*, and *B. wangermeeana* in Brenan 1967). Such a morphological continuum with different morphotypes might reflect shallow divergence, and the current taxa could be considered as having an infraspecific rank instead. Alternatively, the divergence between the current taxa could still be meaningful and all species would then be part of a wider interbreeding system (with limited gene flow between the different taxa), i.e. a syngameon (Grant 1981; e.g. oaks in Cannon and Petit 2020). The characterization of such a system could rely on combined genetic and morphological approaches (e.g. Tovar-Sánchez and Oyama 2004).

CONCLUSION

The analysis of the nuclear ribosomal DNA provides an overview of the evolutionary relationships between the different species of the African genus *Brachystegia*. As in some other tree genera, we found a general fit between nuclear phylogeny and morphology, and a near genus-wide decoupling of geographically sorted plastid signatures. A sole gene tree based on rDNA sequences associated with recent diversification did not allow to fully resolve the *Brachystegia* species tree. Other genomic approaches (i.e. target enrichment, GBS, RAD-seq) need to be tested towards this end. The data provided the opportunity to test different 2ISP scoring in phylogenetic inferences including the novel implementation in RAXML-NG. It proved here to be of some use, as it detects clades that would have been overlooked otherwise. The gain in topological accuracy however remained marginal in *Brachystegia* as the different coding schemes produced congruent and similar cladograms.

ACKNOWLEDGEMENTS

We thank all the curators and other persons who helped us with collecting samples for the genetic analyses. Namely, we thank the people from the following herbaria: Steven Janssens and Ann Bogaerts (BR), Tariq Stévant and Geoffrey Fadeur (BRLU), Stephen Harris and Serena Marner (FHO), Maria Cristina Duarte and Maria M. Romeiras (LISC). We extend our thanks to the people involved in the field collections: Michel Hasson, Annie and Richard De Cauwer, Eric Lowele, and Michel Anastassiou. We acknowledge Laurent Grumiau (ULB-EBE Molecular Biology platform, Belgium) and Latifa Karim (GIGA Liège, Belgium) for their technical support. We are indebted to the Faculté des sciences agronomiques de Lubumbashi (D.R. Congo) for the logistic support provided during the field collection (permit number 023/2016).

We are grateful to G.W. Grimm and an anonymous reviewer for their input during the reviewing process. The comments in addition to the analytical input improved both the content and the shape of this manuscript.

The study was funded by the Belgian “Fonds pour la Formation à la Recherche dans l’Industrie et l’Agriculture” – “Fonds National pour la Recherche Scientifique” (FRIA-FNRS PhD grant to A.F.B) and by the BRAIN-be BELSPO research program BR/132/A1/AFRIFORD (postdoctoral grant to J.M.).

REFERENCES

- Abellán P, Ribera I (2017) Using phylogenies to trace the geographical signal of diversification. *Journal of Biogeography* 44(10): 2236–2246. <http://doi.org/10.1111/jbi.13035>
- Alsos IG, Lavergne S, Merkel MKF, Boleda M, Lammers Y, Alberti A, Pouchon C, Denoed F, Pitelkova I, Puşcaş M, Roquet C, Hurdu B-I, Thuiller W, Zimmermann NE, Hollingsworth PM, Coissac E (2020) The treasure vault can be opened: large-scale genome skimming works well using herbarium and silica gel dried material. *Plants* 9: 432. <https://doi.org/10.3390/plants9040432>
- Bailey C (2003) Characterization of angiosperm nrDNA polymorphism, paralogy, and pseudogenes. *Molecular Phylogenetics and Evolution* 29(3): 435–455. <http://doi.org/10.1016/j.ympev.2003.08.021>
- Baldwin BG, Sanderson MJ, Porter JM, Wojciechowski MF, Campbell CS, Donoghue MJ (1995) The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. *Annals of the Missouri Botanical Garden* 82(2): 247–277. <https://doi.org/10.2307/2399880>
- Bänfer G, Moog U, Fiala B, Mohamed M, Weising K, Blattner FR (2006) A chloroplast genealogy of myrmecophytic *Macaranga* species (Euphorbiaceae) in Southeast Asia reveals hybridization, vicariance and long-distance dispersals. *Molecular Ecology* 15(14): 4409–4424. <http://doi.org/10.1111/j.1365-294X.2006.03064.x>
- Boom AF (2021) Diversification, evolution and population dynamics of the genus *Brachystegia*, a keystone tree of African miombo woodlands. PhD Thesis, Université libre de Bruxelles, Belgium.
- Boom AF, Migliore J, Kaymak E, Meerts P, Hardy OJ (2021) Plastid introgression and evolution of African miombo woodlands: new insights from the plastome-based phylogeny of *Brachystegia* trees. *Journal of Biogeography* 48(4): 933–946. <http://doi.org/10.1111/jbi.14051>
- Brenan JPM (1967) Leguminosae, subfamily Caesalpinioideae. In: Milne-Redhead E, Polhill RM (Eds) *Flora of tropical East Africa*. Crown Agents for Overseas Govts and Admin, 1–230.
- Brummitt R, Chikuni A, Lock J, Polhill R (2007) Leguminosae, subfamily Caesalpinioideae. *Flora Zambesiaca* 3(2): 1–218.
- Campbell B, Frost P, Byron N (1996) Miombo woodlands and their use: Overview and key issues. In: Campbell B (Ed.) *The Miombo in Transition: Woodlands and Welfare in Africa*: 1–10. Centre for International Forestry Research. <https://doi.org/10.17528/cifor/000465>
- Cannon CH, Petit RJ (2020) The oak syngameon: more than the sum of its part. *New Phytologist* 226(4): 978–983. <https://doi.org/10.1111/nph.16091>
- Cappellini E, Gilbert MTP, Geuna F, Fiorentino G, Hall A, Thomas-Oates J, Ashton PD, Ashford DA, Arthur P, Campos PF, Kool J, Willerslev E, Collins MJ (2010) A multidisciplinary study of archaeological grape seeds. *Naturwissenschaften* 97(2): 205–217. <http://doi.org/10.1007/s00114-009-0629-3>
- Charles-Dominique T, Davies TJ, Hempson GP, Bezeng BS, Daru BH, Kabongo RM, Maurin O, Muasya AM, van der Bank M, Bond WJ (2016) Spiny plants, mammal browsers, and the origin of African savannas. *Proceedings of the National Academy of Sciences* 113(38): E5572–E5579. <http://doi.org/10.1073/pnas.1607493113>

- Davies TJ, Daru BH, Bezeng BS, Charles-Dominique T, Hempson GP, Kabongo RM, Maurin O, Muasya AM, van der Bank M, Bond WJ (2020) Savanna tree evolutionary ages inform the reconstruction of the paleoenvironment of our hominin ancestors. *Scientific Reports* 10(1): 12430. <http://doi.org/10.1038/s41598-020-69378-0>
- de la Estrella M, Cervantes S, Janssens SB, Forest F, Hardy OJ, Ojeda DI (2020) The impact of rainforest area reduction in the Guineo-Congolian region on the tempo of diversification and habitat shifts in the *Berlinia* clade (Leguminosae). *Journal of Biogeography* 47(12): 2728–2740. <http://doi.org/10.1111/jbi.13971>
- de la Estrella M, Forest F, Klitgård B, Lewis GP, Mackinder BA, de Queiroz LP, Wieringa JJ, Bruneau A (2018) A new phylogeny-based tribal classification of subfamily Detarioideae, an early branching clade of florally diverse tropical arborescent legumes. *Scientific Reports* 8: 6884. <https://doi.org/10.1038/s41598-018-24687-3>
- Dexter KG, Pennington TD, Cunningham CW (2010) Using DNA to assess errors in tropical tree identifications: how often are ecologists wrong and when does it matter? *Ecological Monographs*, 80(2): 267–286. <https://doi.org/10.1890/09-0267.1>
- Donkpegan ASL, Doucet JL, Hardy OJ, Heuertz M, Piñeiro R (2020) Miocene diversification in the savannahs precedes tetraploid rainforest radiation in the African tree genus *Azelia* (Detarioideae, Fabaceae). *Frontiers in Plant Science* 11: 798. <http://doi.org/10.3389/fpls.2020.00798>
- Eickbush TH, Eickbush DG (2007) Finely orchestrated movements: evolution of the ribosomal RNA genes. *Genetics* 175(2): 477–485. <http://doi.org/10.1534/genetics.107.071399>
- Feliner GN, Rosselló JA (2007) Better the devil you know? Guidelines for insightful utilization of nrDNA ITS in species-level evolutionary studies in plants. *Molecular Phylogenetics and Evolution* 44(2): 911–919. <http://doi.org/10.1016/j.ympev.2007.01.013>
- Fonseca LHM, Lohmann LG (2019) An updated synopsis of *Adenocalymma* (Bignoniaceae, Bignoniaceae): new combinations, synonyms, and lectotypifications. *Systematic Botany* 44(4): 893–912. <https://doi.org/10.1600/036364419X15710776741341>
- Fonseca LHM, Lohmann LG (2020) Exploring the potential of nuclear and mitochondrial sequencing data generated through genome-skimming for plant phylogenetics: A case study from a clade of neotropical lianas. *Journal of Systematics and Evolution* 58(1): 18–32. <http://doi.org/10.1111/jse.12533>
- Frost P (1996) The ecology of miombo woodlands. In: Campbell B (Ed.) *The Miombo in Transition: Woodlands and Welfare in Africa*: 11–57. Centre for International Forestry Research. <https://doi.org/10.17528/cifor/000465>
- Grant V (1981) *Plant speciation*. Columbia University Press.
- Heuertz M, Carnevale S, Fineschi S, Sebastiani F, Hausman JF, Paule L, Vendramin GG (2006) Chloroplast DNA phylogeography of European ashes, *Fraxinus* sp. (Oleaceae): roles of hybridization and life history traits. *Molecular Ecology* 15(8): 2131–2140. <http://doi.org/10.1111/j.1365-294X.2006.02897.x>
- Hijmans RJ (2019) *Geosphere: spherical trigonometry*. R package version 1.5-10. Available from <https://CRAN.R-project.org/package=geosphere>
- Hipp AL, Manos PS, Hahn M, Avishai M, Bodénès C, Cavender-Bares J, Crowl AA, Deng M, Denk T, Fitz-Gibbon S, Gailing O, González-Elizondo MS, González-Rodríguez A, Grimm GW, Jiang X, Kremer A, Lesur I, McVay JD, Plomion C, Rodríguez-Correa H, Schulze E, Simeone MC, Sork VL, Valencia-Avalos S (2020) Genomic landscape of the global oak phylogeny. *New Phytologist* 226(4): 1198–1212. <https://doi.org/10.1111/nph.16162>
- Hoyle AC (1955) *Notulae Systematicae II. A new species of Brachystegia from Southern Nigeria (Caesalpinaceae)*. *Bulletin du Jardin botanique de l'État à Bruxelles* 25(2): 183–190. <http://doi.org/10.2307/3667064>
- Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* 23: 254–267. <https://doi.org/10.1093/molbev/msj030>
- Jin J-J, Yu WB, Yang J-B, dePamphilis CW, Yi T-S, Li D-Z (2020) GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biology* 21: 241. <https://doi.org/10.1186/s13059-020-02154-5>
- Katoh K, Rozewicki J, Yamada KD (2019) MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics* 20(4): 1160–1166. <http://doi.org/10.1093/bib/bbx108>
- Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK (2012) VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Research* 22(3): 568–576. <http://doi.org/10.1101/gr.129684.111>
- Kozlov AM, Alves JM., Stamakis A, Posada D (2022) CellPhy: accurate and fast probabilistic inference of single-cell phylogenies from scDNA-seq data. *Genome Biology* 23: 37. <https://doi.org/10.1186/s13059-021-02583-w>
- Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A (2019) RAXML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* 35(21): 4453–4455. <http://doi.org/10.1093/bioinformatics/btz305>
- Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* 33(7): 1870–1874. <http://doi.org/10.1093/molbev/msw054>
- Lebrun J-P, Stork AL (2008) *Tropical African flowering plants: ecology and distribution, Vol. 3: Mimosaceae - Fabaceae (incl. Derris)*. Conservatoire Botanique de Genève.
- Leonard J, Hauman L, Hoyle C (1952) *Caesalpinaceae IV. - Cynometrae et Amherstieae*. In: Boutique R (Ed.) *Flore du Congo Belge et du Ruanda-Urundi, Spermatophytes, vol. III*, 279–495.
- Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14): 1754–1760. <http://doi.org/10.1093/bioinformatics/btp324>
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup (2009). *The Sequence Alignment/Map*

- format and SAMtools. *Bioinformatics* 25(16): 2078–2079. <http://doi.org/10.1093/bioinformatics/btp352>
- Linder HP (2014) The evolution of African plant diversity. *Frontiers in Ecology and Evolution* 2: 38. <http://doi.org/10.3389/fevo.2014.00038>
- Maddison WP (1997) Gene trees in species trees. *Systematic Biology* 46(3): 523–536. <http://doi.org/10.1093/sysbio/46.3.523>
- Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD, Lopez R (2019) The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research* 47(W1): W636–W641. <http://doi.org/10.1093/nar/gkz268>
- Maurin O, Davies TJ, Burrows JE, Daru BH, Yessoufou K, Muasya AM, van der Bank M, Bond WJ (2014) Savanna fire and the origins of the ‘underground forests’ of Africa. *New Phytologist* 204(1): 201–214. <http://doi.org/10.1111/nph.12936>
- Nicholls J, Pennington R, Koenen E, Hughes C, Hearn J, Bunnefeld L, Dexter K, Stone G, Kidner C (2015) Using targeted enrichment of nuclear genes to increase phylogenetic resolution in the neotropical rain forest genus *Inga* (Leguminosae: Mimosoideae). *Frontiers in Plant Science* 6: 710. <http://doi.org/10.3389/fpls.2015.00710>
- Ojeda DI, Koenen E, Cervantes S, de la Estrella M, Banguera-Hinestroza E, Janssens SB, Migliore J, Demenou BB, Bruneau A, Forest F, Hardy OJ (2019). Phylogenomic analyses reveal an exceptionally high number of evolutionary shifts in a florally diverse clade of African legumes. *Molecular Phylogenetics and Evolution* 137: 156–167. <http://doi.org/10.1016/j.ympev.2019.05.002>
- Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGinn D, Minchin PR, O’Hara RB, Simpson GL, Solymos P, Stevens MHH, Szocs E, Wagner H (2019) Vegan: community ecology package. <http://CRAN.R-project.org/package=vegan>
- Paradis E, Schliep K (2019) ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35(3): 526–528. <http://doi.org/10.1093/bioinformatics/bty633>
- Palgrave CM (2002) Keith Coates Palgrave Trees of southern Africa, 3rd edition. House Struik.
- Pennington RT, Lavin M (2016) The contrasting nature of woody plant species in different neotropical forest biomes reflects differences in ecological stability. *New Phytologist* 210(1): 25–37. <http://doi.org/10.1111/nph.13724>
- Pérez-Escobar OA, Bogarín D, Schley R, Bateman RM, Gerlach G, Harpke D, Brassac J, Fernández-Mazuecos M, Dodsworth S, Hagsater E, Blanco MA, Gottschling M, Blattner FR (2020) Resolving relationships in an exceedingly young Neotropical orchid lineage using Genotyping-by-sequencing data. *Molecular Phylogenetics and Evolution* 144: 106672. <http://doi.org/10.1016/j.ympev.2019.106672>
- Polissar PJ, Rose C, Uno KT, Phelps SR, deMenocal P (2019) Synchronous rise of African C4 ecosystems 10 million years ago in the absence of aridification. *Nature Geoscience* 12: 657–660. <https://doi.org/10.1038/s41561-019-0399-2>
- Potts AJ, Hedderson TA, Grimm GW (2014) Constructing phylogenies in the presence of Intra-Individual Site Polymorphisms (2ISPs) with a focus on the nuclear ribosomal cistron. *Systematic Biology* 63(1): 1–16. <http://doi.org/10.1093/sysbio/syt052>
- R Development Core Team (2019). R: a language and environment for statistical computing. R Foundation for Statistical Computing. <https://www.R-project.org>
- Scheunert A, Heubl G (2017) Against all odds: reconstructing the evolutionary history of *Scrophularia* (Scrophulariaceae) despite high levels of incongruence and reticulate evolution. *Organisms Diversity & Evolution* 17: 323–239. <https://doi.org/10.1007/s13127-016-0316-0>
- Schliep KP (2011) phangorn: phylogenetic analysis in R. *Bioinformatics* 27(4): 592–593. <https://doi.org/10.1093/bioinformatics/btq706>
- Schliep KP, Potts AJ, Morrison DA, Grimm GW (2017) Intertwining phylogenetic trees and networks. *Methods in Ecology and Evolution* 8: 1212–1220. <https://doi.org/10.1111/2041-210X.12760>
- Schley RJ, Pennington RT, Pérez-Escobar OA, Helmstetter AJ, de la Estrella M, Larridon I, Sabino Kikuchi IAB, Barraclough, TG, Klitgård B (2020) Introgression across evolutionary scales suggests reticulation contributes to Amazonian tree diversity. *Molecular Ecology* 29(11): 4170–4185. <https://doi.org/10.1111/mec.15616>
- Simeone MC, Cardoni S, Piredda R, Imperatori F, Avishai M, Grimm GW, Denk T (2018) Comparative systematics and phylogeography of *Quercus* section *Cerris* in western Eurasia: inferences from plastid and nuclear DNA variation. *PeerJ* 6: e5793. <https://doi.org/10.7717/peerj.5793>
- Straub SCK, Parks M, Weitemier K, Fishbein M, Cronn RC, Liston A (2012) Navigating the tip of the genomic iceberg: next-generation sequencing for plant systematics. *American Journal of Botany* 99(2): 349–364. <http://doi.org/10.3732/ajb.1100335>
- Thiers B (continuously updated) Index Herbariorum: a global directory of public herbaria and associated staff, New York Botanical Garden’s Virtual Herbarium. <http://sweetgum.nybg.org/ih/> [04.05.2022]
- Tovar-Sánchez E, Oyama K (2004) Natural hybridization and hybrid zones between *Quercus crassifolia* and *Quercus crassipes* (Fagaceae) in Mexico: morphological and molecular evidence. *American Journal of Botany* 91(9): 1352–1363. <http://www.jstor.org/stable/4123932>
- Turner B, Paun O, Munzinger J, Chase MW, Samuel R (2016) Sequencing of whole plastid genomes and nuclear ribosomal DNA of *Diospyros* species (Ebenaceae) endemic to New Caledonia: many species, little divergence. *Annals of Botany* 117(7): 1175–1185. <http://doi.org/10.1093/aob/mcw060>
- Volkov RA, Panchuk II, Borisjuk NV, Hosiawa-Barabska M, Maluszynska J, Hemleben V (2017) Evolutional dynamics of 45S and 5S ribosomal DNA in ancient allohexaploid *Atropa belladonna*. *BMC Plant Biology* 17: 21. <https://doi.org/10.1186/s12870-017-0978-6>
- White F (1962) Forest Flora of Northern Rhodesia. Oxford University Press.

- Wick RR., Schultz MB, Zobel J, Holt KE (2015) Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics* 31(20): 3350–3352. <http://doi.org/10.1093/bioinformatics/btv383>
- Zeng C-X, Hollingsworth PM, Yang J, He Z-S, Zhang Z-R, Li D-Z, Yang J-B (2018) Genome skimming herbarium specimens for DNA barcoding and phylogenomics. *Plant Methods* 14: 43. <https://doi.org/10.1186/s13007-018-0300-0>
-

SUPPLEMENTARY FILES

Supplementary file 1

The 18S–25S rDNA alignment used in this study. Ambiguities between A and G, C and T, G and C, A and T, G and T, and A and C have been coded as R, Y, S, W, K.

Link: <https://doi.org/10.5091/plecevo.91373.suppl1>

Supplementary file 2

Nucleotide content of ribosomal DNA sequences for each specimen in the alignment used in this study. Nucleotide content is provided for the full 18S–ITS1–5.8S–ITS2–25S alignment, but also for the different main subunits (18S, ITS1, ITS2, and 25S). Apart the classic nucleotide (i.e. A, T, G, and C), intra-individual site polymorphisms have been coded using IUPAC recommendations. Namely ambiguities between A and G, C and T, G and C, A and T, G and T, and A and C have been coded as R, Y, S, W, K, and M. Some ambiguities appear frequent (e.g. Y, mean: 7, range: 1–31) when others are less frequent (e.g. W, mean = 0.5, range = 0–2). After mapping the reads on the reference, low quality bases, indels, and heterozygote positions with three or more possible nucleotides have been coded as N (mean = 1, range = 0–5). The number of gaps is also reported.

Link: <https://doi.org/10.5091/plecevo.91373.suppl2>

Supplementary file 3

The different supplementary figures (Figs S1–S12). Specific captions are provided for each figure.

Link: <https://doi.org/10.5091/plecevo.91373.suppl3>