Molecular Phylogenetics and Evolution 150 (2020) 106854

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

# Plastome phylogeography in two African rain forest legume trees reveals that Dahomey Gap populations originate from the Cameroon volcanic line

Boris B. Demenou[a,*], Jérémy Migliore[a], Myriam Heuertz[b], Franck K. Monthe[a], Dario I. Ojeda[a,c], Jan J. Wieringa[d,e], Gilles Dauby[a,f], Laura Albreht[a], Arthur Boom[a], Olivier J. Hardy[a]

[a] Evolution Biologique et Ecologie, Université Libre de Bruxelles, Faculté des Sciences, CP160/12, Av. F. D. Roosevelt 50, BE-1050 Brussels, Belgium
[b] BIOGECO, INRA, Univ. Bordeaux, F-33610 Cestas, France
[c] Norwegian Institute of Bioeconomy Research, Høgskoleveien 8, 1433 Ås, Norway
[d] Naturalis Biodiversity Center, National Herbarium of the Netherlands, Darwinweg 2, 2300 RA Leiden, the Netherlands
[e] Wageningen University & Research, Biosystematics Group, Droevendaalsesteeg 1, 6708 PB Wageningen, the Netherlands
[f] AMAP Lab, IRD, CNRS, INRA, Univ Montpellier, Montpellier, France

ARTICLE INFO

*Keywords:*
African rain forest
Cameroon volcanic line
Colonization origin
Dahomey gap
Phylogeography
Plastid genome sequencing

ABSTRACT

Paleo-environmental data show that the distribution of African rain forests was affected by Quaternary climate changes. In particular, the Dahomey Gap (DG) – a 200 km wide savanna corridor currently separating the West African and Central African rain forest blocks and containing relict rain forest fragments – was forested during the mid-Holocene and possibly during previous interglacial periods, whereas it was dominated by open vegetation (savanna) during glacial periods. Genetic signatures of past population fragmentation and demographic changes have been found in some African forest plant species using nuclear markers, but such events appear not to have been synchronous or shared across species. To better understand the colonization history of the DG by rain forest trees through seed dispersal, the plastid genomes of two widespread African forest legume trees, *Anthonotha macrophylla* and *Distemonanthus benthamianus*, were sequenced in 47 individuals for each species, providing unprecedented phylogenetic resolution of their maternal lineages (857 and 115 SNPs, respectively). Both species exhibit distinct lineages separating three regions: 1. Upper Guinea (UG, i.e. the West African forest block), 2. the area ranging from the DG to the Cameroon volcanic line (CVL), and 3. Lower Guinea (LG, the western part of the Central African forest block) where three lineages co-occur. In both species, the DG populations (including southern Nigeria west of Cross River) exhibit much lower genetic diversity than UG and LG populations, and their plastid lineages originate from the CVL, confirming the role of the CVL as an ancient forest refuge. Despite the similar phylogeographic structures displayed by *A. macrophylla* and *D. benthamianus*, molecular dating indicates very contrasting ages of lineage divergence (UG diverged from LG since c. 7 Ma and 0.7 Ma, respectively) and DG colonization (probably following the Mid Pleistocene Transition and the Last Glacial Maximum, respectively). The stability of forest refuge areas and repeated similar forest shrinking/expanding events during successive glacial periods might explain why similar phylogeographic patterns can be generated over contrasting timescales.

## 1. Introduction

Major changes in tropical forest cover during the Pleistocene are thought to have greatly influenced large-scale biogeographic patterns and ecological processes. Past climatic fluctuations probably caused major fragmentation-expansion cycles of the African Guineo-Congolian rain forest (Dupont et al., 2001; Miller and Gossling, 2014), affecting population genetic structure, distribution and biodiversity patterns (Demenou et al., 2017; Demenou et al., 2016; Hardy et al., 2013; Piñeiro et al., 2017). Environmental barriers (e.g. dry corridors separating wet ecosystems) can impede the homogenization of species pools while favoring genetic divergence between conspecific populations, as recently shown for the Dahomey Gap (Demenou et al., 2017; Demenou et al., 2016; Iloh et al., 2017). The Dahomey Gap (hereafter DG) is a *c.* 200 km wide corridor (Demenou, 2018) dominated by open vegetation in eastern Ghana, Togo, Benin, and western Nigeria, which separates two major Guineo-Congolian forest blocks: Upper Guinea (hereafter UG), which extends from Sierra Leone to Ghana, and Lower Guinea

* Corresponding author.
*E-mail addresses:* bdemenou@ulb.ac.be, ohardy@ulb.ac.be (B.B. Demenou).

https://doi.org/10.1016/j.ympev.2020.106854
Received 2 October 2019; Received in revised form 8 May 2020; Accepted 13 May 2020
Available online 19 May 2020

(hereafter LG), which extends from southern Nigeria to the western part of the Republic of Congo (White 1986, 1979).

The establishment of an open vegetation in the DG is linked to past climate changes since at least 1.05 Ma ago (Demenou et al., 2017; Demenou et al., 2016; Dupont et al., 2001; Dupont and Weinelt, 1996; Miller and Gossling, 2014; Salzmann and Hoelzmann, 2005; Tossou, 2002). However, paleo-vegetation reconstructions using sediment cores collected in the DG, notably in Lake Sélé, reveal that the region was occupied by a closed forest during the early and mid-Holocene (from *c.* 8400 to 4500 yr BP). Then, savannas expanded rapidly from 4500 to 3400 cal. yr BP following the Climatic Pejoration of the Holocene (HCP), a dry period characterized by less rainfall and more seasonality (Marchant and Hooghiemstra, 2004). Although a return to wetter conditions led to a forest-savanna mosaic from 3300 to 1100 yr BP, savannas re-expanded and have dominated the DG over the last 1100 yr (Salzmann and Hoelzmann, 2005). Adjacent to the DG, lake Bosumtwi in Ghana (now in a moist semi-deciduous forest area) yielded half a million years of paleo-ecological records, revealing six main periods of forest vegetation (Miller and Gossling, 2014), congruent with those reported in a marine core of the Niger delta (Dupont, 2011). These forested periods are interpreted as synchronous to the most recent interglacials, whereas open vegetation dominated during the glacial periods. A marine core facing the coast of DG also contained evidence of extended rain forests during the early and mid-Holocene and the Eemian interglacial (130–115 ka), and extended savannas and open woodlands during the last glacial period, particularly between 70 and 15 ka (Dupont and Weinelt, 1996).

Hence, open vegetation seems to have dominated the DG during most of the late Quaternary. However, few forested phases have occurred during the interglacial periods, as was the case during the mid-Holocene and probably during the peaks of other interglacial periods when the African rain forest reached its maximal extension to form a single block. Consequently, it is expected that populations of forest species underwent recurrent range shifts, demographic changes, and fragmentation or population admixture events during the Quaternary, leaving signatures in their genetic constitution. Based on nuclear microsatellites, the DG populations of two Guineo-Congolian tree species, *Distemonanthus benthamianus* Baill. (Demenou et al., 2016) and *Terminalia superba* Engl. and Diels (Demenou et al., 2017), have been inferred to originate from the admixture of populations from UG and LG, with a higher contribution of UG (71–80%). However, these admixture events do not seem synchronous, as admixture was inferred to have occurred, respectively, after and before the Last Glacial Maximum (LGM) in these species. In the absence of precise estimation of mutation rates we cannot rule out effects of estimation bias on this result. Moreover, pollen-mediated gene flow might result in admixture at the nuclear genome even if the DG was colonized through seed dispersal from one forest block only. New insights into the history of the DG populations could thus be derived from data from maternally inherited genomes, like the plastome in most angiosperms (Petit and Vendramin, 2007), in particular regarding the colonization process through seed dispersal. A stronger phylogeographical pattern should be expected for plastid DNA in comparison with nuclear DNA because of its haploid nature, which enhances genetic drift, and because it is probably not affected by gene flow through pollen dispersal. DNA sequence data also allows more robust molecular dating using fossil calibration than microsatellites for which mutation rates are difficult to estimate.

The present study aims to use whole plastid genome sequencing to better trace the origin and history of the DG populations of two legume trees, *Anthonotha macrophylla* P.Beauv. (Detarioideae) and *Distemonanthus benthamianus* (Dialioideae). These two species are mainly distributed in secondary, semi-deciduous to evergreen forest throughout the Guineo-Congolian region, but they also occur in isolated forest fragments and gallery forest of the DG. We infer the history of population fragmentation and colonization based on the phylogeographic structure of dated plastome genealogies. More specifically, we address the following questions: (i) Was the DG colonized from both UG and LG during the humid Holocene period, as supported by microsatellite data in *D. benthamianus*? (ii) Are phylogeographic patterns similar and synchronous between the two species studied, suggesting shared histories driven by the same processes? (iii) Is there evidence that some regions were colonized recently (rapid spread of recent lineages), while other ones served as museums of diversity (coexistence of old lineages)? Finally, we compare insights from plastomes with those obtained with nuclear genetic data.

## 2. Material and methods

### 2.1. Plant material and DNA extraction

*Distemonanthus benthamianus* is a large, deciduous, wind-dispersed, light-demanding pioneer tree distributed from UG to LG. *Anthonotha macrophylla* is an evergreen shrub to medium-sized tree, rarely lianescent, non-pioneer, occurring in lower strata of tropical rain forest and in secondary forest, distributed from UG to Congolia (the Eastern part of Central African forest), whose seeds are consumed but possibly also dispersed by monkeys. We sequenced the plastome of 47 georeferenced individuals of each species sampled across their distribution ranges. The samples were composed of field-collected leaves dried in silica gel (18 individuals of *A. macrophylla* and 47 of *D. benthamianus*) and from herbarium specimens (29 individuals of *A. macrophylla*) collected at the National Herbarium of the Netherlands, Naturalis (WAG) and the Herbarium of the Université Libre de Bruxelles (BRLU) (Table S1 in Appendix S1). Total genomic DNA was extracted for each sample (*c.* 10 μg) from 20 to 30 mg of dried leaves, using the cetyltrimethylammonium bromide (CTAB) method (Doyle and Doyle, 1987). Extracted DNA was purified with QIAquick purification kits (Qiagen, Venlo, Netherlands) followed by Qubit ® 2.0 Fluorometer DNA quantification (Life Technologies, Invitrogen, Foster City, USA) and by QIAxcel (Qiagen) DNA quality control.

### 2.2. Whole plastome sequencing

Extracted genomic DNA was first sheared using a Bioruptor® Pico sonicator (Diagenode SA., Liège, Belgium) to target DNA fragments around *c.* 400 bp (8 cycles of 15′ sonication, 90′ wait, brief centrifugation); this step was not conducted on extracts from herbarium specimens because their DNA was already fragmented. The sizing was based on AMPure XP Solid Phase Reversible Immobilisation beads (Agencourt Bioscience, Beckman Coulter, Brea, USA), to target fragments from 200 to 600 bp. The preparation of genomic libraries followed the approach of Rohland and Reich (2012) and Mariac et al. (2014), including a fast DNA end-repair step (NEB End Repair enzyme), a blunt-end ligation (NEB T4 DNA ligase), and a nick fill-in amplification of adapters (NEB Bstl DNA polymerase), including 6-bp barcodes for samples multiplexing. Each step was followed by AMPure XP bead-based clean-up, QIAxcel DNA quality control (Qiagen, Venlo, Netherlands), and Qubit® 2.0 Fluorometer quantification (Life Technologies, Invitrogen, Foster City, USA). For *A. macrophylla*, the libraries were enriched in plastome sequences using a protocol detailed hereafter. We then performed real-time quantitative PCRs (qPCRs) using the KAPA SYBR® FAST qPCR Kit (Kapa Biosystems) with Illumina i5/i7 adapters on a StepOnePlus (Applied Biosystems, Foster City, USA) to generate the ready-to-load libraries. After qPCR, multiplexed libraries were paired-end sequenced (2 × 150 bp) on an Illumina NextSeq sequencer with reagent kit V2 mid-output at the GIGA platform (Liège, Belgium).

Two strategies of massively parallel sequencing were applied. First, whole genome skimming data was used to reconstruct reference plastid genomes, targeting > 500,000 reads per sample to guarantee that all parts of the plastome can be assembled. This approach was applied for all the samples of *D. benthamianus* and for one sample of *A. xanderi* Breteler. Second, to maximize the sequencing of plastomes, we

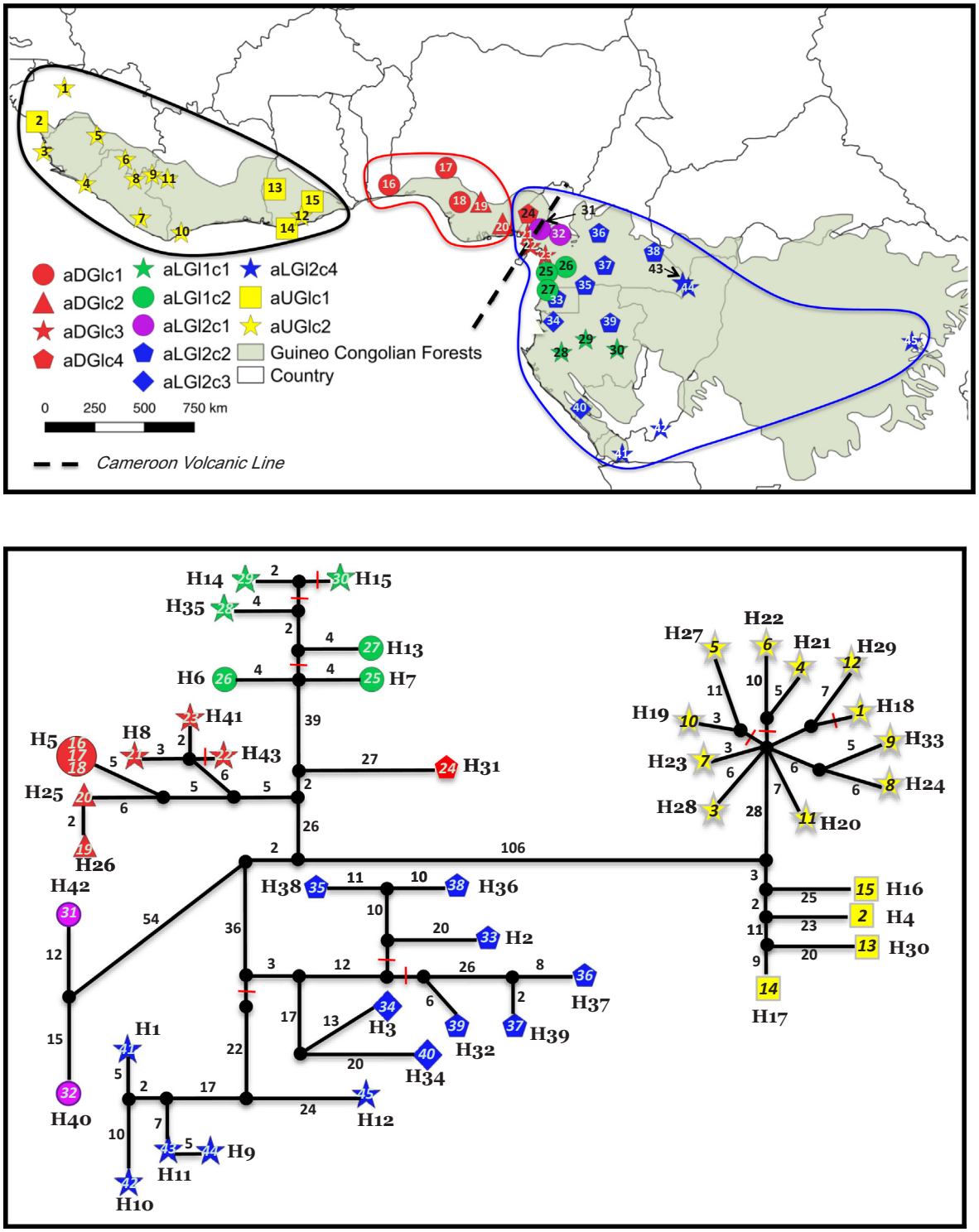


**Fig. 1.** Phylogeographical structure and plastid haplotype diversity in the legume tree *Anthonotha macrophylla*. The map shows the distribution of each individual used in this study (numbers represent individual IDs) while the network shows the phylogenetic relationships between haplotypes (haplotype IDs are preceded by the letter H, their symbol size increases with the number of individuals bearing that haplotype, and individual IDs are given inside each symbol or under parentheses). The colours identify 5 main lineages while different symbols of the same colour identify subclades or isolated haplotypes. The numbers along the network segments indicate the number of mutations (if $>$ 1, segment length is not proportional to number of mutations) and red bars across the segments indicate when there is a single mutation. Three populations were delineated: Upper-Guinea (UG) circled in black, Dahomey Gap (DG) in red and Lower-Guinea (LG) in blue. The dotted line indicates the Cameroon Volcanic Line (CVL) axis. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

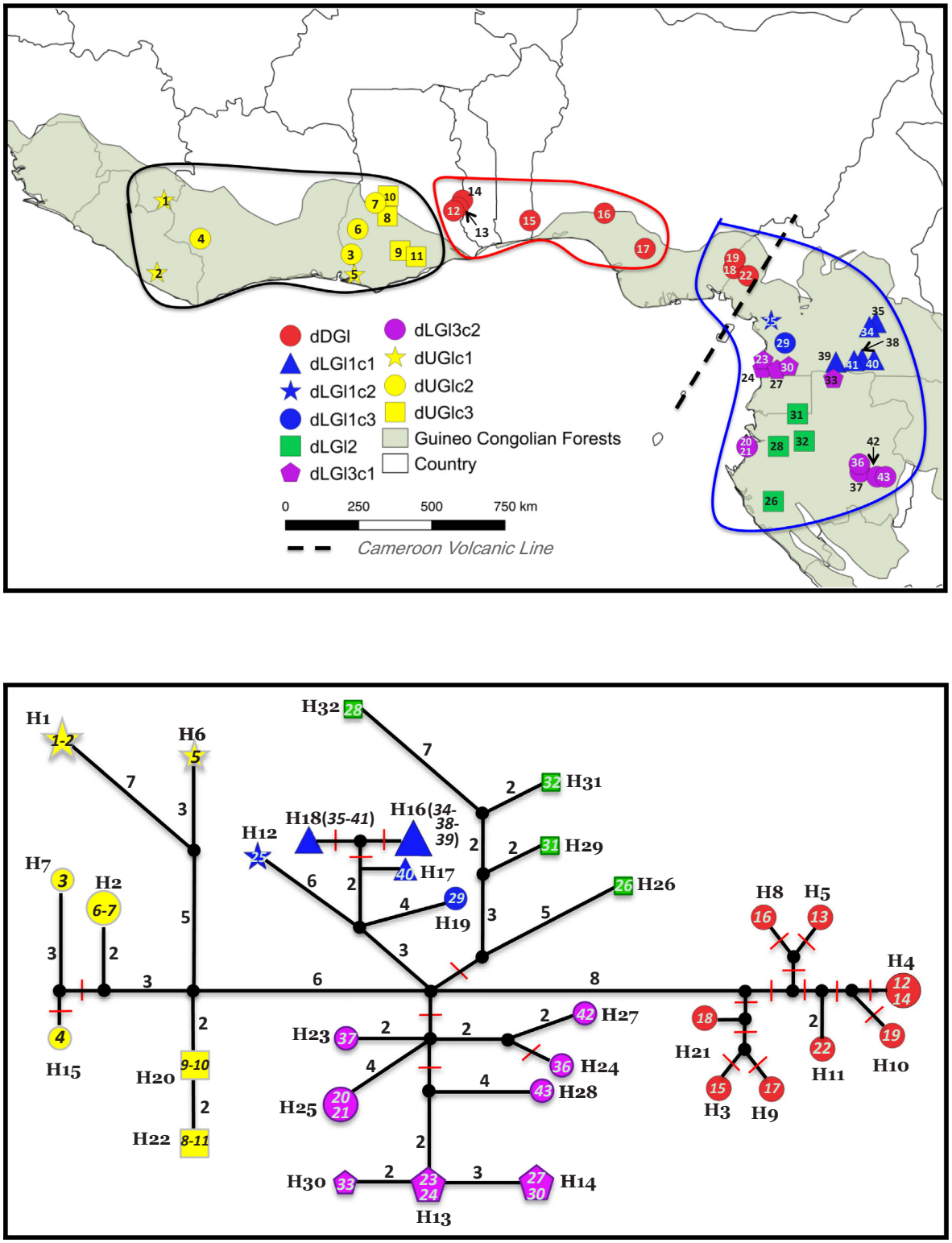


**Fig. 2.** Phylogeographical structure and plastid haplotype diversity in the legume tree *Distemonanthus benthamianus*. See legend of Fig. 1.

performed plastome enrichment for *Anthonotha*, reducing the number of reads needed for reconstructing plastomes (150,000 reads per sample).

Plastome enrichment was based on biotinylated probes previously produced in our laboratory at Université Libre de Bruxelles for another Leguminosae-Detarioideae tree, *Guibourtia tessmannii* (Tosso et al., 2018). As attempts to produce specific probes for the two species were not very conclusive (results not shown) and *Anthonotha* is closely related to *Guibourtia*, we opted for an enrichment using *Guibourtia* plastome probes. As detailed in Tosso et al. (2018), long-range PCR reactions were performed with 16 universal PCR primer combinations (Uribe-Convers et al., 2014) using the LongAmp® Taq PCR Kit (New England Biolabs, Beverly, USA) on one *Guibourtia* sample. After shearing, sizing, blunt-end ligation of L1L2 adapters, and biotinylation with L2Biotine primers (Phusion® High-Fidelity PCR Master Mix, NEB), these probes were used to capture plastid DNA directly by hybridization

with the genomic DNA libraries of *Anthonotha*, during 48 h at 65 °C (see Mariac et al., 2014). The plastid DNA-enriched libraries were then submitted to qPCR and sequenced on a NextSeq Illumina sequencer.

### 2.3. Bioinformatic analyses: data cleaning, SNP calling, missing data filtering

The quality of the Illumina data generated was first checked with FastQC. The demultiplexing of samples based on the 6-bp barcodes was performed using Sabre (https://github.com/najoshi/sabre), using a pe command without the m option that allows for mismatches in the barcodes. Adapters and low-quality bases were removed using Trimmomatic 0.36 (http://www.usadellab.org/cms/?page=trimmomatic; Bolger et al., 2014) with the following options: LEADING and TRAILING = 20 (quality cutoff the start and the end of a read); MINLEN (minimum length) = 40; seed mismatches = 2; palindrome clip threshold = 30 and simple clip threshold = 10.

We used the software Mitobim 1.5 (Hahn et al., 2013) to reconstruct the plastid genome of *D. benthamianus*, assembling the reads from the non-enriched genomic library. Mitobim mapped these reads using MIRA 3.4.1.1 (Chevreux et al., 1999) to highly conserved regions of a closely related reference genome available in GenBank: *Haematoxylum brasiletto* (Leguminosae Detarieae; NC_026678.1).

The *Anthonotha xanderi* plastid genome was assembled using the software GetOrganelle v1.6.2e (Jin et al., 2018). The assembly was visually presented and edited using the software Bandage 0.8.1 (Wick, Schultz, Zobel & Holt, 2015) and the remaining contigs were joint using GetOrganelle utilities.

The plastid DNA-enriched libraries for *A. macrophylla* and the non-enriched ones for *D. benthamianus* were mapped on the plastid reference genomes previously constructed, using Bwa Mem 0.7.15-r1142-dirty (Li and Durbin, 2009) with -M and -B 4 options.

To call SNPs, we then used the combination of Samtools 1.3 (Li and Durbin, 2009) with –B option to generate a mpileup file and Varscan 2.3.7 mpileup2snp and mpileup2cns (Koboldt et al., 2012) with a –min-var-freq of 0.5, a –min-freq-for-hom of 0.5, a minimum base quality of 30 (as recommended by Scarcelli et al., 2016), and a min coverage of 3 for *D. benthamianus* (lower depth of sequencing) and 8 for *A. macrophylla*. We then filtered and cleaned the vcf files. To do so, we used vcftools (https://github.com/vcftools/vcftools, Danecek et al., 2011) with the following options: (i) –remove-indels –recode –recode-INFO-to remove all indels; (ii) –min-alleles 2 –max-alleles 2 to keep only biallelic sites; (iii) –exclude-positions to exclude SNPs position with observed heterozygosity > 0.1 and SNPs with missing data (N) in the reference; (iv) –minDP 5 –minGQ 10 –max-missing 0.8 to respectively keep only SNPs with a minimum depth of 5, a minimum genotype quality of 10 and with at most 20% of missing data. We finally used Beagle (https://faculty.washington.edu/browning/beagle/beagle.html; Browning and Browning, 2007) to impute missing data. The final vcf files generated were then converted to a fasta multi-alignment file.

### 2.4. Analysis of plastome variation: haplotype diversity, distribution and differentiation between populations

We used DnaSP v.5.10.01 (Librado and Rozas, 2009) to infer haplotypes for each individual (excluding sites with gaps/missing data). The number of haplotypes (Nh) was computed and the genetic relatedness between them was visualized via a median–joining haplotype network (Bandelt et al., 1999) using the software NETWORK. The geographic distribution of haplotypes was visualized by plotting them on a map using QGIS (QGIS Development Team, 2009).

Based on phylogenetic networks and the geographic distribution of haplotypes, we identified within each species five groups of closely related haplotypes showing a restricted distribution, hereafter referred to as lineages as they were well supported in phylogenetic analyses (see results). These lineages represent the major groups of haplotypes within which phylogeographic structure at a finer scale was observed (Figs. 1 and 2). In addition, we delimited in both species three populations whose limits are based on the geographical distribution of lineages within the three major biogeographic regions investigated (UG, DG and LG).

The SNPs data for both species were used to estimate the nucleotide diversity ($\pi$) and the gene diversity (h: gene diversity corrected for sample size; Nei, 1978) for each population and over all populations using DnaSp 5.10.01 (Librado and Rozas, 2009).

Using SPAGeDi (Hardy and Vekemans, 2002), we estimated the following genetic differentiation indices between populations (Pons and Petit, 1996): $G_{ST}$, based on the identity of haplotypes, and $N_{ST}$, based on the pairwise nucleotide diversity matrices between haplotypes, which were generated using MEGA 7.0.21 (Kumar et al., 2016) with pairwise deletion for gaps when treating missing data. We then tested for the presence of a phylogeographic signal (i.e. when $N_{ST} > G_{ST}$) indicating that distinct haplotypes within population are more related on average than haplotypes sampled between populations (Pons and Petit, 1996).

### 2.5. Time-calibration of plastome phylogenies

We time-calibrated the plastome phylogenetic trees in two steps: (i) we aligned our plastid sequences to published coding sequences from other legume species to estimate the TMRCA (time to the most recent common ancestor) of all samples within *A. macrophylla* and *D. benthamianus*, (ii) we used these secondary calibration points to build dated phylogenetic trees within each target species using all plastome data available. For the first step, we used the alignment of 72 protein-coding plastid genes in Koenen et al. (2020), retaining the four Dialioideae and 13 Detarioideae species available, as these subfamilies diverged at the start of the Tertiary (Koenen et al., 2020), plus *Quillaja saponaria* (Quillajaceae) as outgroup within the order Fabales. We aligned these genes on our plastome sequences using MAFFT 7 (Katoh and Standley 2013), keeping eventually 45 genes (some plastome regions were lost from our *Anthonotha* plastomes due to the enrichment step), totaling 44 109 nucleotides. We used PARTITIONFINDER2 (Lanfear et al., 2017) to define partitions for subsequent phylogenetic analyses on nucleotide sequences, considering the different codon positions per gene. We used RAxML-NG (Kozlov et al. 2019) to generate a maximum likelihood phylogenetic tree with 100 bootstraps, and BEAST 1.10.4 (Drummond et al., 2012) to time-calibrate it by setting the divergence date between Dialioideae and Detarioideae subfamilies as a normally distributed random variable (mean = 64 Ma, sd = 1). We considered a strict clock model with a coalescent prior tree of constant effective size, and tested further the impact of assuming a Yule process of speciation, an uncorrelated relaxed clock model assuming a log-normal distribution of substitution rates among branches, or a single partition under the GTR + Gamma + Invariant sites model. Using the relaxed clock model, we also constrained the crown age of Detarioideae to be > 22.5 Ma (based fossil evidence discussed in Koenen et al., 2020) and the crown age of the order Fabales between 80 and 85 Ma. The MCMC analyses were run for up to 200,000,000 generations, sampling trees every 10,000 generations. BEAST log output files were examined using Tracer 1.7.1 (Rambaut and Drummond, 2016) to evaluate convergence and ensure sufficient effective sample sizes (ESS values) for all parameters (ESS > 200). A maximum clade credibility (MCC) tree with posterior node age distributions was obtained using TreeAnnotator 1.10.4 and represented using FigTree 4.3 (Rambaut, 2007).

The TMRCA between major conspecific lineages estimated above was then used to time-calibrate the phylogenetic tree of all available plastid sequences within each species. At this level, all available nucleotide sequences were aligned using MAFFT 7 over their full length, reaching 69982 bp for *A. macrophylla* and 151,666 bp for *D. benthamianus*, after removing gaps. RAxML-NG and BEAST were again used, setting the TMRCA of *A. macrophylla* and *D. benthamianus* as log-normally distributed random variables of mean ± SD = 7.05 ± 1.66 Ma

and 0.72 ± 0.25 Ma, respectively, following estimates obtained in the first step for the uncorrelated relaxed clock model with additional age constrains. We used a strict clock model, a reasonable assumption within species, GTR + G and HKY + G nucleotide substitution models as identified with jModeltest 2.1.10 (Maleŝ and Sarić, 2009) for *A. macrophylla* and *D. benthamianus* respectively, without partitioning because this had nearly no impact on the divergence times estimated in the first step. We tested coalescent models of constant population size, exponential growth, logistic growth, Bayesian skyline and extended Bayesian skyline plot, and obtained similar results (results not shown) so that the constant size model was eventually kept.

### 2.6. Relationship between lineage age and geographic range

For each node of the phylogenetic trees, the maximal geographical distance between the samples descending from this node was used as a measure of distribution range, which was plotted against the node height. We then compared whether the age-range relationship differed between regions (UG, DG, LG), expecting that recently colonized regions contain wide ranging recent lineages, as we expect for the DG.

## 3. Results

### 3.1. Genomic libraries and plastome reconstruction

The reference plastid genome of *D. benthamianus* (GiD0728, Gabon) was assembled after 7 iterations of Mitobim using 1,055,681 reads (R1-R2 paired), leading to a 158,673 bp long plastome (Table 1, Genbank accession n° MN604403). The plastid genome of *A. xanderi* was assembled by GetOrganelle using 1,260,568 paired reads and had a total length of 159,874 bp (Table 1, Genbank accession n° MT442044).

An average of 303,300 reads per individual were generated for *A. macrophylla* (enriched libraries) and 1,037,465 for *D. benthamianus* (non-enriched libraries). We obtained an average depth of 189x (sd = 212) for *A. macrophylla* and 29x (sd = 10) for *D. benthamianus*. The number (proportion) of mapped reads was on average 146,272 (48.2%) for *A. macrophylla* and 35,903 (3.5%) for *D. benthamianus*, demonstrating the efficiency of the enrichment step in *A. macrophylla*. A minimum of 5x coverage was achieved on average for 52.4% of the *A. macrophylla* plastome and for 95.9% of the *D. benthamianus* plastome, showing that nearly half of the *A. macrophylla* plastome was not well enriched using the baits developed from the *Guibourtia* plastome (Table 1).

We obtained for all individuals a plastome sequence length of 69,982 bp for *A. macrophylla* and 151,666 bp for *D. benthamianus*, representing respectively 43.9% and 95.6% of the total length of the reference for each species. After filtering, cleaning and imputation of missing data (4.3% in *A. macrophylla* and 4.8% in *D. benthamianus*), we identified a total of 857 SNPs in 45 individuals of *A. macrophylla* and 115 SNPs in 43 individuals of *D. benthamianus*. This set of markers and individuals were used in subsequent analyses.

### 3.2. Plastid haplotypes and distribution

A total of 44 haplotypes were identified for *A. macrophylla* and 32 for *D. benthamianus* (Figs. 1 and 2; Table S1 in Appendix S1). Hence, almost all haplotypes were unique in *A. macrophylla*, except H5, which was shared by 3 individuals (one from Benin, two from Nigeria; Fig. 1 and Table S1 in Appendix S1). For *D. benthamianus*, 22 haplotypes were unique while 10 haplotypes were shared by 2 or 3 individuals, generally situated in the same country (Fig. 2 and Table S1 in Appendix S1). The haplotype networks allowed the identification of five main lineages in each species (monophyletic in the MCC trees, see later, and represented by the same colour in Figs. 1 and 2). These lineages tend to occur in parapatry, allowing to distinguish three populations with very similar biogeographic limits in both species: (i) Upper Guinea (UG), (ii) a region spanning from the Dahomey Gap to the Cameroonian Volcanic Line (DG-CVL), and (iii) Lower Guinea (LG), including Congolia in the case of *A. macrophylla* (Figs. 1 and 2). In each species, the UG lineage was the most divergent one, as confirmed by the MCC trees (Fig. 3), and the lineage including samples from the DG was always phylogenetically closer to LG lineages than to the UG one (Figs. 1 and 2). In most of the main lineages, subclades also occurred and displayed phylogeographic substructures (Figs. 1 and 2, where colours identify the main lineages and symbols the subclades).

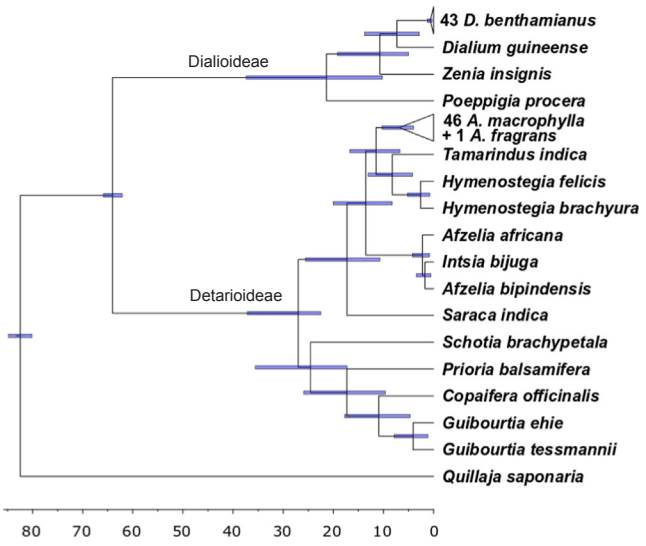


**Fig. 3.** Dated phylogeny of legume plastomes from the Dialioideae and Detarioideae subfamilies, rooted with the non-legume *Quillaja saponaria*, as inferred using 45 protein coding genes. The tree includes 43 *Distemonanthus benthamianus* and 45 *Anthonotha macrophylla* samples (forming two collapsed subtrees) whose respective TMRCA were estimated with BEAST under the uncorrelated relaxed clock model, a coalescent tree, 48 partitions of nucleotidic data (44109 sites), and age constrains on the divergence time between Dialioideae and Detarioideae (64 Ma, sd = 1), the TMRCA of Detarioideae (> 22.5 Ma), and tree height (80–85 Ma). The blue bars represent the 95% highest posterior density intervals for the age of each node (scale is in million years). All represented nodes were highly supported. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

**Table 1**
Sample size for each species and summary values of raw and mapped reads. Raw reads correspond to the number of reads obtained with the NextSeq Illumina sequencing platform. Plastid DNA enrichment was applied on *A. macrophylla* but not on *D. benthamianus*.

| Species | Number of individuals | Enrichment percentage | Mean raw number of reads | Mean number of mapped reads (%) | Mean coverage per sample | Coverage of at least 5X (%) |
|---|---|---|---|---|---|---|
| *Anthonotha macrophylla* | 47 (45)* | 60 | 303,300 | 146,272 (48.2%) | 189 ± 212 | 52.4 |
| *Distemonanthus benthamianus* | 47 (43)* | – | 1,037,465 | 35,903 (3.5%) | 29 ± 10 | 95.9 |

* Final number of individuals used for different analyses between parentheses.

**Table 2**
Haplotype diversity and differentiation parameters of the three populations recognized in *A. macrophylla* and *D. benthamianus* species.

| | Region | Differentiation parameters: $G_{ST}/N_{ST}$ | | | Diversity parameters | | | |
|---|---|---|---|---|---|---|---|---|
| | | UG | DG | LG | n | Nh | h | π |
| *A. macrophylla* | UG | – | 0.884 * | 0.666 * | 15 | 15 | 1 | 0.00051 |
| | DG | 0.150 | – | 0.408 * | 5 | 3 | 0.7 | 0.00011 |
| | LG | 0 | 0.150 | – | 25 | 25 | 1 | 0.00148 |
| | All samples $G_{ST}$ = 0.100 $N_{ST}$ = 0.700 | | | | 45 | 43 | 0.99 | 0.00187 |
| *D. benthamianus* | UG | – | 0.738 * | 0.433 * | 11 | 8 | 0.95 | 0.00007 |
| | DG | 0.227 | – | 0.528 * | 6 | 5 | 0.60 | 0.00001 |
| | LG | 0.047 | 0.220 | – | 26 | 19 | 0.96 | 0.00009 |
| | All samples $G_{ST}$ = 0.165 $N_{ST}$ = 0.577 | | | | 43 | 32 | 0.98 | 0.00010 |

UG: Upper Guinea; DG: Dahomey Gap (including southern Nigeria west of Cross River); LG: Lower Guinea.
Differentiation: $G_{ST}$ below and $N_{ST}$ above the diagonal; tests of phylogeographical signal (i.e. $N_{ST} > G_{ST}$): * for $P < 0.05$.
Diversity: n: sample size, Nh: Number of haplotypes; h: gene diversity corrected for sample size, Nei, 1978, π: nucleotide diversity.

### 3.3. Genetic diversity and differentiation between regions

Despite the much higher nucleotide diversity found in *A. macrophylla* ($\pi = 19\ 10^{-4}$) compared to *D. benthamianus* ($\pi = 1.0\ 10^{-4}$), both species showed the same regional gradient of nucleotide diversity with LG > UG > DG, and a nearly ten-fold difference between DG and LG (Table 2).

Genetic differentiation measures between regions based on haplotype frequencies ($G_{ST}$) were low due to the high haplotype diversity found within regions (except in DG; Table 2). However, differentiation measures accounting for the phylogenetic information ($N_{ST}$) were relatively high, and tests of phylogeographic signals were significant ($N_{ST} > G_{ST}$) indicating, as expected, that distinct haplotypes within each region are more related on average than haplotypes sampled from different regions (Table 2).

### 3.4. Timing of divergence between lineages and origin of DG populations

The first analyses using coding genes inferred much older TMRCA of plastomes for *A. macrophylla* (5.0–7.05 Ma according to models) than *D. benthamianus* (0.32–0.72 Ma), whatever the models (Table 3; Fig. 3). TMRCA estimates were little impacted when assuming prior trees following a Yule rather than a coalescent process, or using a single partition rather than all 48 partitions identified by PARTITION FINDER (Table 3). However, compared to the strict clock models, the uncorrelated relaxed clock models gave older estimates and wider confidence intervals (95% HPD) in *D. benthamianus* and to a lower extent in *A. macrophylla*, a trend increased when adding age constrains on the TMRCA of Detarioideae (Table 3). The relaxed clock models required to run 10 chains of 200 million generations each to reach satisfactory ESS values (> 200) because the TMRCA estimates displayed autocorrelation times of 6 to 10 million MCMC steps. As the posterior distribution of the standard deviation in substitution rates among branches did not encompass zero (95% HPD 1.4–4.3 $10^{-4}$), we used the results of the relaxed clock model, including the additional age constrains, to derive the secondary calibration points for the within species plastid phylogenies. In both species, sequences from UG diverged first from the other ones, so that the TMRCA correspond to the divergence between the UG lineage and the other ones. Interestingly, a plastome of *A. fragrans* is situated among the *A. macrophylla* subtree, in a sister position to sample 24.

Lineages from the LG in *A. macrophylla* (Fig. 4), started diverging around 4.4 Ma [95% HPD: 2.4–6.4], while those in UG started around 1.9 Ma [95% HPD: 1.1–2.9]. The lineage including samples from the CVL and the DG region (samples 16–24) diverged from its closest LG lineage 2.1 Ma [95% HPD: 1.1–3.1]. It contains a clade endemic to the DG region and southern Nigeria (samples 16–20) that diverged 0.78 Ma [95% HPD: 0.41–1.27] ago from the most related CVL samples (21–23). A subclade including the three samples (16 to 18) from Benin and south-central Nigeria bear the same haplotype H5 and diverged from the south-eastern Nigerian samples (19 and 20) around 0.35 Ma [95% HPD: 0.16–0.62] (Fig. 4).

In *D. benthamianus* (Fig. 4) lineage divergence started 0.43 Ma [95% HPD: 0.18–0.81] in LG, and 0.39 Ma [95% HPD: 0.17–0.77] in UG. The clade including samples from the CVL and the DG region (samples 12–19 and 22) diverged from its closest LG lineage 0.43 Ma [95% HPD: 0.18–0.81] and the TMRCA of its sequences is 0.12 Ma [95% HPD: 0.035–0.25]. Contrary to the analogous *A. macrophylla* clade, this one does not form geographically segregated subclades as at least two subclades range from the CVL to the DG, and the divergence time between DG samples (12–17) and closest CVL samples (18–19 and 22) range from 0.026 Ma (between samples 14 or 17 and 19) to 0.081 Ma (between samples 13 or 16 and 22; Fig. 3).

### 3.5. Age-range relationships in different regions

As expected, the geographical range of a lineage tends to increase

**Table 3**
Estimation of the TMRCA (in Ma), and 95% HPD intervals (under square brackets), of plastid sequences for each focal species (*Anthonotha macrophylla* and *Distemonanthus benthamianus*) and their corresponding subfamilies (Detarioideae and Dialioideae, respectively) under different models implemented in BEAST. To this end, 45 plastid protein coding genes were aligned for 43 *D. benthamianus* samples, 45 *A. macrophylla* samples, and published sequences from three additional Dialioidea species, 13 Detarioidea species and *Quillaja saponaria* used as outgroup (Koenen et al., 2020). We considered the strict clock model or the uncorrelated relaxed clock model with Lognormal distribution of substitution rates, the Yule or Coalescent tree models, 48 partitions and their molecular evolution models inferred by PARTITIONFINDER2 or a single partition under the GTR + G + I model, and a single fossil-based calibration (divergence between subfamilies at N(64,1) Ma), plus additional age constrains in one case (TMRCA Detarioidea > 22.5 Ma, TMRCA of whole tree = Fabales between 80 and 85 Ma).

| Taxon | Strict Clock Coalescent 48 partitions | Strict Clock Yule 48 partitions | Strict Clock Coalescent 1 partition | Relaxed Clock Coalescent 48 partitions | Relaxed Clock Coalescent 48 partitions + age constrains |
|---|---|---|---|---|---|
| *Anthonotha* | 5.36 [4.66–6.11] | 5.58 [4.86–7.14] | 5.00 [4.31–5.67] | 6.48 [3.42–9.93] | 7.05 [4.12–10.42] |
| *Distemonanthus* | 0.32 [0.21–0.45] | 0.37 [0.26–0.50] | 0.32 [0.22–0.44] | 0.66 [0.28–1.12] | 0.72 [0.32–1.21] |
| Detarioideae | 20.05 [18.82–21.37] | 20.22 [18.99–21.45] | 20.38 [19.19–21.70] | 24.45 [10.13–43.75] | 28.1 [22.5–37.11] |
| Dialioideae | 24.28 [22.46–26.09] | 24.38 [22.45–26.09] | 24.05 [22.31–25.83] | 21.62 [9.43–35.66] | 22.23 [9.88–37.00] |

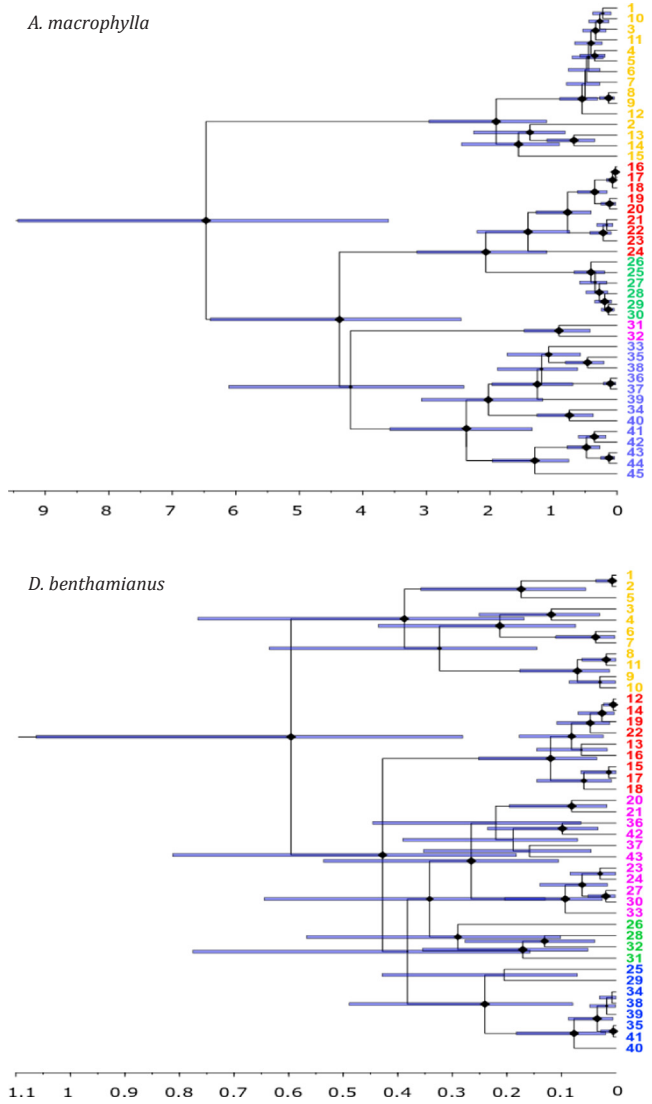


**Fig. 4.** Dated genealogy of *A. macrophylla* (above) and *D. benthamianus* (below) plastomes inferred with BEAST (MCC trees). The blue bars represent the 95% highest posterior density intervals for the age of each node. The numbers indicated next to terminal taxa correspond to the IDs of individuals and their colour indicate the main lineages (orange for the UG lineage, red for the DG + CVL lineage, and green, blue and violet for LG lineages). Nodes with a diamond were well supported (> 0.9). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

with its age (Fig. 5). However, the age-range relationship differs strikingly between species and regions. In *A. macrophylla*, most lineages younger than 0.5 Ma span < 500 km, except in UG where even lineages of 0.2 Ma age span more than *c.* 1000 km (Fig. 5). In *D. benthamianus*, the timescale is much more recent because the root of the phylogenetic tree is only 0.6 Ma and spans > 2500 km (Fig. 5). Here, most lineages younger than 0.2 Ma are distributed over < 500 km, except in the DG-CVL region where lineages span > 1000 km.

## 4. Discussion

This study is among the first relying on whole plastome sequencing to elucidate the demographic history of African plants (Faye et al., 2016, Migliore et al., 2019) across the entire tropical African rain forest. The gain in information is considerable compared to classical studies relying on Sanger sequencing of one or a few plastid regions giving low numbers of polymorphic sites and haplotypes. For example, in the Guineo-Congolian tree *Santiria*, Koffi et al. (2011) found only 14 SNPs defining 12 haplotypes among 377 individuals sequenced at three plastid regions, with a total 1598 bp. Previous attempts using Sanger sequencing of plastid genes performed on *D. benthamianus* failed to detect any polymorphism, while our whole plastome sequencing revealed 115 SNPs, which means that < 0.1% of the sites are polymorphic, explaining in turn the previous lack of variation. Polymorphism in *A. macrophylla* was much higher with 857 SNPs, thus sequencing a few genes would have been sufficient to detect the main lineages but the phylogenetic resolution here obtained allowed us to investigate the genetic structure at a finer scale within each lineage.

### 4.1. Congruent phylogeographic structures in the two legume tree species

Both species show some congruent phylogeographical patterns: (i) Upper Guinea (UG) is composed of one distant lineage, (ii) Central Africa is composed of several lineages occurring in parapatry or sympatry, (iii) the Dahomey Gap (DG) is occupied by one lineage spanning from the Cameroon Volcanic Line (CVL) to the DG, (iv) the CVL contains endemic lineages divergent from the most widespread Central African lineages. These patterns support the hypothesis of past forest fragmentation isolating rain forest populations from West (UG) and Central (LG) Africa for long periods of time. This supports the view that past climatic events have created genetic structures within many African tropical forest plant species, as obtained with several types of markers and species (e.g. Budde et al., 2013; Daïnou et al., 2014; Demenou et al., 2017; Demenou et al., 2016; Duminil et al., 2015; Faye et al., 2016; Hardy et al., 2013; Iloh et al., 2017; Migliore et al., 2018 and others). Our results also highlight the key role of the area surrounding the Cameroonian Volcanic Line for hosting ancient lineages, consistently with previous studies (Budde et al., 2013; Dauby et al., 2014; Migliore et al., 2019). Within regions (UG, DG and LG), strong spatial phylogeographic structures also appear in both species (Figs. 1 and 2), indicating that dispersal (a priori through seeds for plastomes) remains limited compared to the mutation rate at the whole plastome level.

Strong phylogeographic structure was observed in *A. macrophylla* for the LG and DG, while in the UG the structure is less evident. The absence of spatial phylogenetic structure may be indicative of relatively recent population expansion (or recent colonization) that had not enough time to accumulate mutations in different areas, although such interpretation must be taken with caution given our limited sampling size. However, for both species, the DG was occupied by phylogenetically very close haplotypes (Figs. 1 and 2), that are also closely related to haplotypes distributed in Nigeria and around the CVL. The absence of a clear phylogeographic structure in the DG for *D. benthamianus*, in contrast to *A. macrophylla*, suggests a faster and/or a more recent colonization of the DG by this species, as confirmed by phylogenetic dating (Fig. 4).

If we consider only the lineages occurring in Central Africa (LG) and excluding the CVL, there is much less congruence in the phylogenetic structures of the two species. While *A. macrophylla* displays two main lineages with largely overlapping distribution ranges in Cameroon and Gabon (Fig. 1), *D. benthamianus* displays three lineages that are mainly parapatric, two occurring in south-western Cameroon and/or Gabon and one restricted to central Cameroon (Fig. 2) with a distribution corresponding well to the northern LG genetic cluster previously identified using nuclear microsatellite data (Demenou et al. 2016). Hence, within Central African forests, the two species seem to have undergone very distinct biogeographic histories.

### 4.2. Haplotype diversity and centres of diversity

The plastid genome showed similar gradients of diversity across regions: for both species, nucleotide diversity ($\pi$) was much higher in

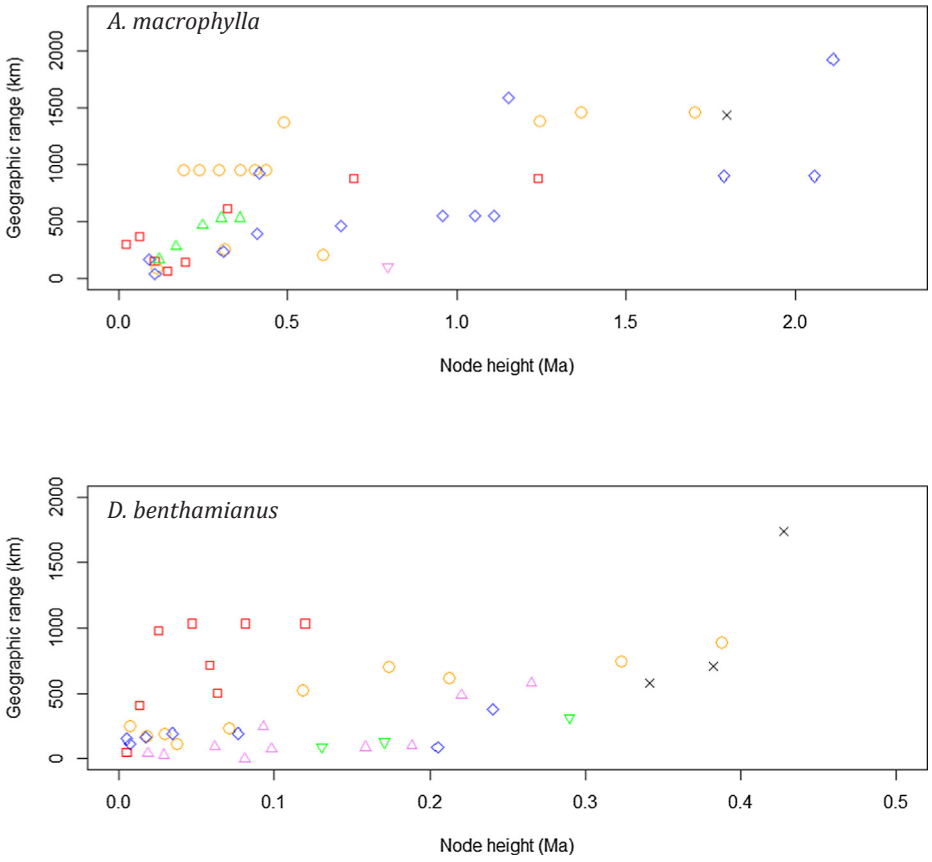


**Fig. 5.** Age-range relationship: the geographical range of the samples descending from each node is plotted against node height for the plastid phylogenetic trees of *A. macrophylla* (above) and *D. benthamianus* (below). Nodes are distinguished according to the main lineage they belong to: UG lineage (orange circles), DG-CVL lineage (red squares), LG lineages (green and violet triangles, blue diamonds), other nodes (black crosses). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

both west and central African forest compared to the DG. It was also higher in central African forest than in the west African forest, suggesting that the former is likely the centre of origin for the two species studied here, as confirmed by the dated phylogenies (Fig. 4). These results are congruent with nuclear microsatellite data on both species (Demenou, 2018; Demenou et al., 2017; Demenou et al., 2016), showing higher genetic diversity for the forested areas than in the DG. The low genetic diversity in DG may reflect small effective population sizes, as nuclear microsatellites data suggest a recent decline or bottleneck in DG populations for both species (Demenou et al., 2017; Demenou et al., 2016), similarly observed in *Terminalia superba* (Demenou, 2018). High genetic diversity has been observed in tropical rain forest organisms when associated to stable species-specific habitats, e.g., animals in the Coastal Atlantic Forests of Brazil (Carnaval et al., 2009) and the Wet Tropics in Australia (Moussalli et al., 2009). Similar results were found in other regions such as in the Iberian Peninsula (Abellán and Svenning, 2014), in the Himalayas (Qu et al., 2014) and more recently in the African palm genus *Podococcus* (Faye et al., 2016). Our study demonstrated thus that central African forests could be considered as an ancient and relatively stable habitat for our studied species.

### 4.3. Contrasted divergence time in the two legume tree species

Despite their congruent phylogeographic subdivision in three main regions (UG, DG, LG), the two species studied show a contrasting timing of population divergence. In *A. macrophylla*, plastid lineages started diverging since the Miocene (7 Ma), lineages within the UG and LG diverged since the Pliocene (4.4 Ma), and the clade found in the DG diverged from its sister clade situated in the CVL since the mid-Pleistocene (0.78 Ma). Such timescale is comparable to that obtained recently for another Guineo-Congolian tree (*Greenwayodendron suaveolens*, Annonaceae). In the latter species, Migliore et al. (2019) concluded that the phylogeographic structure of plastomes were established during the Pliocene or early Pleistocene, and while it might have been reinforced during subsequent glacial–interglacial cycles, the forest expansion expected during inter-glacial periods did not allow a remixing of the different lineages. In contrast, plastid lineages in *D. benthamianus* diverged much more recently during the late Pleistocene, *c.* 0.6 Ma between UG and LG lineages. Moreover, the haplotypes found in the DG diverged only 0.026 to 0.081 Ma ago from their closest relatives situated in the CVL. These estimates have low precision (95% HPD 0.003–0.069 and 0.023–0.177, respectively) but are not inconsistent with microsatellite data suggesting that the DG population originates from the early Holocene (0.013–0.007 Ma; Demenou et al., 2016).

The fact that two tree species display similar phylogeographic breaks across the entire African rain forest despite contrasted divergence times indicates that the same barriers and forest refuges appeared repeatedly at different geological periods. It highlights also the role played by the CVL as a museum of diversity, which can fuel the recolonization of newly available habitats, in accordance with some previous works (e.g. Budde et al., 2013; Dauby et al., 2014; Migliore et al., 2019).

### 4.4. Plastome data definitively clarify the origin of DG populations

One of the main objectives of this study was to compare nuclear and plastome data to clarify the origin of the current DG populations of the studied species. Under the assumption that plastid DNA is transmitted maternally, the phylogeography of the plastomes should trace the history of population colonization. This study supports the CVL as a source for the colonization of the Dahomey Gap region because, for *A. macrophylla*, haplotypes found in the DG form a derived subclade of a larger clade where all other samples came from the CVL. For *D. benthamianus*, evidence of a CVL origin is less strong if we consider only plastid

phylogenetic information but our previous work based on microsatellites showed that the nuclear genome of the DG samples result from recent (post last glacial maximum) admixture between Upper Guinean and Lower Guinean gene pools (Demenou et al. 2016). Hence, in *D. benthamianus* also, the most parsimonious explanation for the origin of the DG plastomes given their phylogenetic relationships with other regions is that they originate from Lower Guinea, and more specifically from the CVL.

Based on our results, we suggest that the current DG population of *A. macrophylla* may be the result of a progressive colonization that could have started as early as 780 ka (divergence between CVL and eastern Nigerian samples) and continued 350 ka ago (divergence between DG and eastern Nigerian samples), long before the penultimate glacial period (PGM, 140–190 ka). Nuclear microsatellite data confirm that the DG gene pool derives from a LG gene pool without contribution from UG (Demenou, 2018). In *D. benthamianus*, the DG population does not form a unique plastid clade but contains at least two clades ranging from the CVL to the DG, and the divergence between DG samples and their closest CVL relatives dates back from 26 to 81 ka. Even more recent dates are supported by nuclear microsatellite data (Demenou et al., 2016), but the latter indicates that the nuclear gene pool in the DG results from admixture between UG and LG gene pools. Hence, the colonization of the DG by *D. benthamianus* seems to have occurred around or after the last glacial maximum (LGM) through a westward range expansion of a CVL population, followed by substantial admixture due to pollen flow from the UG population. It may well be possible that the DG was reached in earlier stages as well by populations from LG and the CVL, but this is no longer evident in our dataset. Such populations either became extinct again during glacial cycles, or the maternal lineages ended by coincidence or got lost in an overflow of incoming immigrants. In addition, we cannot totally exclude that seed-mediated colonisation from UG also occurred and that these lineages were not sampled due to our limited sample size.

The contrasted colonization timescales between *A. macrophylla* and *D. benthamianus* might result from their contrasted dispersal abilities and ecological strategies. Pods of *D. benthamianus* are wind-dispersed and parentage analyses revealed that while the majority of seeds disperse over $< 100$ m, still 20 to 30% of them disperse over $> 500$ m, possibly aided by storm winds carrying pods of trees dominating the surrounding canopy (Hardy et al., 2019). Pollen dispersal by insects is also extensive, 70% dispersing over 1000 m (Hardy et al., 2019). Being a pioneer species, *D. benthamianus* could thus rapidly extend its range when new habitats become available, and long-distance pollen dispersal could favour admixture between populations, explaining its recent phylogeographic history. In contrast, seeds and pods of *A. macrophylla* are much heavier, fall from a lower height (subcanopy tree $\leq 20$ m height) and probably disperse over much shorter distances, although direct measurements are lacking and dispersal by water and/or monkeys might play a role. Being non-pioneer, this species probably extended its range more progressively during periods of forest expansion, explaining its ancient phylogeographic structure.

Finally, it must also be kept in mind that what we infer here is the history of plastid lineages, not of species per se. In the case of *Anthonotha*, we observed that the plastome of an *A. fragrans* sample was sister to one of our *A. macrophylla* samples, so that plastid captures seem to occur among *Anthonotha* species. This hypothesis deserves further investigation but, if confirmed, the inferred history of *Anthonotha* plastomes might reflect the colonization history of a group of *Anthonotha* species subject to cytoplasmic exchanges rather than the history of *A. macrophylla* only.

## 5. Conclusion

Whole genome sequencing of the plastome demonstrated its power for plant evolutionary study, providing high levels of polymorphism and a high phylogeographic resolution (Faye et al., 2016; Monthe et al., 2019). Our comparative study on the phylogeography and the demographic history of two tropical African rain forest plants, *A. macrophylla* and *D. benthamianus*, showed congruent origins of the Dahomey Gap (DG) populations but different timing of colonization. Both species exhibit a strong phylogeographic structure separating forest blocks longitudinally, and similar gradients of genetic diversity (Central African forest > West African forest > DG). The DG was colonized from the Cameroon volcanic line (CVL), much before the Last Glacial Maximum (LGM) in the case of *A. macrophylla*, and around or after the LGM, possibly during the humid Holocene, for *D. benthamianus*. The CVL origin of the DG population was congruent with nuclear data for *A. macrophylla* but not for *D. benthamianus*, where pollen and/or seed influx from UG seems to have occurred. Overall, our results clarify the origin of DG populations, highlight the role of the CVL as a reservoir of genetic diversity for the DG, and illustrate that similar phylogeographic patterns can result from processes that occurred at contrasting timescales.

## Data accessibility

Dataset used for this study was deposited in Dryad Digital Repository (doi:https://doi.org/10.5061/dryad.zpc866t68).

## CRediT authorship contribution statement

**Boris B. Demenou:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing - original draft, Writing - review & editing. **Jérémy Migliore:** Methodology, Formal analysis, Validation, Writing - review & editing. **Myriam Heuertz:** Methodology, Formal analysis, Validation, Writing - review & editing. **Franck K. Monthe:** Methodology, Resources, Formal analysis, Validation, Writing - review & editing. **Dario I. Ojeda:** Methodology, Data curation, Writing - review & editing. **Jan J. Wieringa:** Resources, Writing - review & editing. **Gilles Dauby:** Methodology, Writing - review & editing. **Laura Albreht:** Formal analysis, Writing - review & editing. **Arthur Boom:** Formal analysis, Writing - review & editing. **Olivier J. Hardy:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing - original draft, Writing - review & editing.

## Acknowledgements

We thank the “Fonds pour la Formation à la Recherche dans l’Industrie et l’Agriculture” (FRIA, Belgium) for providing PhD grants to BBD and FKM, and the “Fonds de la Recherche Scientifique” (F.R.S.-FNRS, grant T.0163.13) and the Belgian Science Policy (project AFRIFORD under BRAIN program) for financial support. We also thank the Labex COTE for funding a three-months mobility to INRA-Pierroton for BBD. We thank Esra Kaymak for her help in the laboratory, Hugo Darras and Chedly Kastally for their help with bioinformatic analyses, and Erik Koenen for his advices for calibrating the phylogenetic trees. Finally, we would like to thank the Herbarium of the Muséum National d'Histoire Naturelle (MNHN) at Paris and the Naturalis Biodiversity Center at Leiden for allowing us to sample leaf tissue from their herbarium collections. Computational resources have been provided by the Consortium des Équipements de Calcul Intensif (CÉCI), funded by the Fonds de la Recherche Scientifique de Belgique (F.R.S.-FNRS) under Grant No. 2.5020.11 and by the Walloon Region. We are also grateful to the Genotoul bioinformatics platform Toulouse Midi-Pyrenees (Bioinfo Genotoul) for providing help and/or computing and/or storage resources.

## Author contribution

BBD and OJH designed the study. BBD, JJW, GD and OJH contributed to collection of plant material. BBD, JM, FKM performed the laboratory analyses. BBD, FKM, OJH and MH conducted the data analyses. LA and AB reconstructed the reference plastome for *A. xanderi*. BBD and OJH wrote the first draft and all authors contributed to the final version of the manuscript.

## Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2020.106854.

## References


Abellán, P., Svenning, J.C., 2014. Refugia within refugia–patterns in endemism and genetic divergence are linked to Late Quaternary climate stability in the Iberian Peninsula. Biol. J. Linn. Soc. 113, 13–28.

Bandelt, H.-J., Forster, P., Röhl, A., 1999. Median-joining networks for inferring intraspecific phylogenies. Mol. Biol. Evol. 16, 37–48.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina Sequence Data. Bioinformatics 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Browning, S.R., Browning, B.L., 2007. Rapid and accurate haplotype phasing and missing data inference for whole genome association studies by use of localized haplotype clustering. Am. J. Hum. Genet. 81, 1084–1097.

Budde, K.B., Gonzalez-Martinez, S.C., Hardy, O.J., Heuertz, M., 2013. The ancient tropical rainforest tree Symphonia globulifera L. f. (Clusiaceae) was not restricted to postulated Pleistocene refugia in Atlantic Equatorial Africa. Heredity 111, 66–76.

Carnaval, A.C., Hickerson, M.J., Haddad, C.F.B., Rodrigues, M.T., Moritz, C., 2009. Stability predicts genetic diversity in the Brazilian Atlantic forest hotspot. Science 323, 785–789.

Chevreux, B., Wetter, T., Suhai, S., 1999. Genome sequence assembly using trace signals and additional sequence information computer science and biology. Proc. German Conf. Bioinform. 99, 45–56.

Daïnou, K., Mahy, G., Duminil, J., DicK, C.W., Doucet, J.-L., Donkpegan, A.S.L., et al., 2014. Speciation slowing down in widespread and long-living tree taxa: insights from the tropical timber tree genus Milicia (Moraceae). Heredity 113, 74–85.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R., 1000 Genomes Project Analysis Group, 2011. The variant call format and VCFtools. Bioinformatics 27 (15), 2156–2158.

Dauby, G., Duminil, J., Heuertz, M., Koffi, G.K., Stévart, T., Hardy, O.J., 2014. Congruent phylogeographical patterns of eight tree species in Atlantic Central Africa provide insights into the past dynamics of forest cover. Mol. Ecol. 23, 2299–2312. https://doi.org/10.1111/mec.12724.

Demenou, B.B., 2018. Origin, Diversity and Phylogeography of the Guineo-Congolian flora of the Dahomey Gap. PhD Thesis. Université Libre de Bruxelles, pp. 254.

Demenou, B.B., Doucet, J.-L., Hardy, O.J., 2017. History of the fragmentation of the African rain forest in the Dahomey Gap: insight from the demographic history of *Terminalia superba*. Heredity 120, 547–561.

Demenou, B.B., Piñeiro, R., Hardy, O.J., 2016. Origin and history of the Dahomey Gap separating West and Central African rain forests: insights from the phylogeography of the legume tree *Distemonanthus benthamianus*. J. Biogeogr. 43, 1020–1031.

Doyle, J.J., Doyle, J.L., 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem. Bullet. 19, 11–15.

Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1. 7 research article. Soc. Mol. Biol. Evol. 29, 1969–1973. https://doi.org/10.1093/molbev/mss075.

Duminil, J., Mona, S., Mardulyn, P., Doumenge, C., Walmacq, F., Doucet, J.-L., et al., 2015. Late Pleistocene molecular dating of past population fragmentation and demographic changes in African rain forest tree species supports the forest refuge hypothesis. J. Biogeogr. 42, 1443–1454.

Dupont, L.M., 2011. Orbital scale vegetation change in Africa. Quat. Sci. Rev. 30, 3589–3602.

Dupont, L.M., Donner, B., Schneider, R., Wefer, G., 2001. Mid-Pleistocene environmental change in tropical Africa began as early as 1.05 Ma. Geology 29, 195–198.

Dupont, L.M., Weinelt, M., 1996. Vegetation history of the savannah corridor between the Guinean and the Congolian rain forest during the last 150,000 years. Veg. Hist. Archaeobot. 5, 273–292.

Faye, A., Delauwe, V., Mariac, C., Richard, D., Sonké, B., Vigouroux, Y., Couvreur, T.L.P., 2016. Phylogeography of the genus Podococcus (Palmae/Arecaceae) in Central African rain forests: climate stability predicts unique genetic diversity. MPE 105, 126–138.

Hahn, C., Bachmann, L., Chevreux, B., 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. Nucl. Acid Res. 41 (13) e129-e129.

Hardy, O.J., Born, C., Budde, K., Daïnou, K., Dauby, G., Duminil, J., et al., 2013. Comparative phylogeography of African rain forest trees: a review of genetic signatures of vegetation history in the Guineo-Congolian region. C R Geosci. 345, 284–296.

Hardy, O.J., Vekemans, X., 2002. SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. Mol. Ecol. Notes 2, 618–620.

Hardy, O.J., Delaide, B., Hainaut, H., Gillet, J.-F., Gillet, P., Kaymak, E., Vankerckhove, N., Duminil, J., Doucet, J.-L., 2019. Seed and pollen dispersal distances in two African legume timber trees and their reproductive potential under selective logging. Mol. Ecol. 28 (12). https://doi.org/10.1111/mec.15138.

Iloh, A.C., Schmidt, M., Muellner-Riehl, A.N., Ogundipe, O.T., Paule, J., 2017. Pleistocene refugia and genetic diversity patterns in West Africa: insights from the liana Chasmanthera dependens (Menispermaceae). PLoS ONE 12 (3), e0170511. https://doi.org/10.1371/journal.pone.0170511.

Jin, J.-J., Yu, W.-B., Yang, J.-B., Song, Y., Yi, T.-S., Li, D.-Z., 2018. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. bioRxiv 256479. https://doi.org/10.1101/256479.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30 (4), 772–780. https://doi.org/10.1093/molbev/mst010.

Koboldt, D., Zhang, Q., Larson, D., et al., 2012. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 22, 568–576.

Koenen, E.J.M., Ojeda, D.I., Steeves, R., Migliore, J., Bakker, F.T., Wieringa, J.J., Kidner, C., Hardy, O.J., Pennington, R.T., Bruneau, A., Hughes, C.E., 2020. Large-scale genomic sequence data resolve the deepest divergences in the legume phylogeny and support a near-simultaneous evolutionary origin of all six subfamilies. New Phytol. 225 (3), 1355–1369. https://doi.org/10.1111/nph.16290.

Koffi, K.G., Hardy, O.J., Doumenge, C., Cruaud, C., Heuertz, M., 2011. Diversity gradients and phylogeographic patterns in Santiria trimera (Burseraceae), a widespread African tree typical of mature rainforests. Am. J. Bot. 98, 254–264.

Kozlov, A.M., Darriba, D., Flouri, T., Morel, B., Stamatakis, A., 2019. RAxML-NG: a fast, scalable, and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics btz305. https://doi.org/10.1093/bioinformatics/btz305.

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870–1874.

Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., Calcott, B., 2017. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Mol. Biol. Evol. 34, 772–773. https://doi.org/10.1093/molbev/msw260.

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

Librado, P., Rozas, J., 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25, 1451–1452.

Maleŝ, Željan, Sarić, F., 2009. Kvantitativna analiza fenolnih spojeva ljepljivog omana - Inula viscosa (L.) Ait. Farmaceutski Glasnik 65, 143–148. https://doi.org/10.1038/nmeth.2109.

Marchant, R., Hooghiemstra, H., 2004. Rapid environmental change in Africa and South American tropics around 4000 years before present: a review. Earth-Sci. Rev. 66, 217–260.

Mariac, C., Scarcelli, N., Pouzadou, J., Barnaud, A., Billot, C., Faye, A., Kougbeadjo, A., Maillol, V., Martin, G., Sabot, F., 2014. Cost-effective enrichment hybridization capture of chloroplast genomes at deep multiplexing levels for population genetics and phylogeography studies. Molecul. Ecol. Resour. 14, 1103–1113.

Migliore, J., Kaymak, E., Mariac, C., Couvreur, T.L.P., Lissambou, B.-J., Pineiro, R., Hardy, O., 2019. Pre-pleistocene origin of phylogeographical breaks in African rain forest trees: new insights from Greenwayodendron (Annonaceae) phylogenomics. J. Biogeogr. https://doi.org/10.1111/jbi.13476.

Miller, C.S., Gossling, W.D., 2014. Quaternary forest associations in lowland tropical West Africa. Quat. Sci. Rev. 84, 7–25.

Monthe, F.K., Migliore, J., Duminil, J., Bouka, G., Demenou, B.B., Doumenge, C., Blanc-Jolivet, C., Ekué, M.R.M., Hardy, O.J., 2019. Phylogenetic relationships in two African Cedreloideae tree genera (Meliaceae) reveal multiple rain/dry forest transitions. Perspect. Plant. Ecol. Evolut. Syst. 37, 1–10.

Moussalli, A., Moritz, C., Williams, S.E., Carnaval, A.C., 2009. Variable responses of skinks to a common history of rainforest fluctuation: concordance between phylogeography and palaeo-distribution models. Mol. Ecol. 18, 483–499.

Nei, M., 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89, 583–590.

Petit, R.J., Vendramin, G.G., 2007. Phylogeography of organelle DNA in plants: an introduction. In: Weiss, S., Ferrand, N. (Eds.), Phylogeography of Southern European Refugia. Kluwer, Amsterdam.

Piñeiro, R., Dauby, G., Kaymak, E., Hardy, O.J., 2017. Pleistocene population expansions of shade-tolerant trees indicate fragmentation of the African rainforest during the Ice Ages. Proc. R Soc. B 284, 20171800.

Pons, O., Petit, R., 1996. Measuring and testing genetic differentiation with ordered versus unordered alleles. Genetics 144, 1237–1245.

QGIS Development Team, 2009. QGIS Geographic Information System. Open Source Geospatial Foundation Project. http://qgis.osgeo.org (accessed 12 July 2018).

Qu, Y., Ericson, P.G., Quan, Q., Song, G., Zhang, R., Gao, B., Lei, F., 2014. Long-term isolation and stability explain high genetic diversity in the Eastern Himalaya. Mol. Ecol. 23, 705–720.

Rambaut, A., 2007. FigTree, a graphical viewer of phylogenetic trees. See: http://tree.bio.ed.ac.uk/software/figtree.

Rambaut, A., Drummond, A.J., 2016. Tracer v 1.4. 8. Institute of Evolutionary Biology, University of Edinburg, 2007.

Rohland, N., Reich, D., 2012. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. Genome Res. 22, 939–946.

Salzmann, U., Hoelzmann, P., 2005. The Dahomey Gap: an abrupt climatically induced rain forest fragmentation in West Africa during the late Holocene. The Holocene 15, 1–10.
Scarcelli, N., Mariac, C., Couvreur, T.L.P., Faye, A., Richard, D., Sabot, F., Berthouly-Salazar, C., Vigouroux, Y., 2015. Intra-individual polymorphism in chloroplasts from NGS data: where does it come from and how to handle it? Molecul. Ecol. Resour. 16, 434–445.
Tosso, F., Hardy, O.J., Doucet, J.-L., Daïnou, K., Kaymak, E., Migliore, J., 2018. Evolution in the amphi-Atlantic tropical genus Guibourtia (Fabaceae, Detarioidaea), combining NGS phylogeny and morphology. Mol. Phylogenet. Evol. 120, 83–93.
Tossou, G.M., 2002. Recherche palynologique sur la vegetation Holocene du Sud-Bénin (Afrique de l'Ouest). PhD Thesis. Université de Lomé, Togo.
Uribe-Convers, S., Duke, J.R., Moore, M.J., Tank, D.C., 2014. A long PCR–based approach for DNA enrichment prior to next-generation sequencing for systematic studies. Appl. Plant Sci. 2, 1300063.
White, F., 1979. The Guineo-Congolian region and its relationships to other phytochoria. Bulletin du Jardin Botanique National de Belgique/Bulletin van de National Plantentuin van België 49, 11–55.
White, F., 1986. La végétation de l'Afrique. ORSTOM-UNESCO, 384p.
Wick, R.R., Schultz, M.B., Zobel, J., Holt, K.E., 2015. Bandage: interactive visualisation of de novogenome assemblies. Bioinformatics 31, 3350–3352.