Botanical Journal of the Linnean Society, 2025, **XX**, 1–13
https://doi.org/10.1093/botlinnean/boaf033
Advance access publication 21 May 2025
Original Article

OXFORD

Original Article

Mid-Pleistocene origin and phylogeographical signatures of recurrent expansion-fragmentation of a highly inbred and endangered African timber legume

Surabhi Ranavat[1,*], Saskia Sergeant[1], Dieu-Merci Assumani[1,2,3], Bonaventure Sonké[4], Gaël U.D. Bouka[5], Nils Bourland[6], Jean-François Gillet[7], Jean-Louis Doucet[8], Olivier J. Hardy[1]

[1]Evolutionary Biology and Ecology, Faculté des Sciences, Université Libre de Bruxelles (ULB), Av. F.D. Roosevelt 50, 1050, Brussels, Belgium
[2]Faculté de Gestion des Ressources Naturelles Renouvelables, Université de Kisangani, B.P. 2012, Av. Kitima 3, Kisangani, Democratic Republic of the Congo
[3]Institut National pour l'Etude et la Recherche Agronomiques (INERA – Yangambi), Yangambi, Democratic Republic of the Congo
[4]University of Yaoundé I, Plant Systematic and Ecology Laboratory, Yaoundé, PO Box 047, Cameroon
[5]Laboratory of Biodiversity and Ecosystems and Environmental Management, Faculty of Science and Technology, Marien Ngouabi University, Brazzaville, PO Box 69, Republic of the Congo
[6]Royal Museum for Central Africa, Service of Wood Biology, 13 Leuvensesteenweg, Tervuren, 3080, Belgium
[7]Nature Forest Environment, Porcheresse/Daverdisse, 6929, Belgium
[8]TERRA Teaching and Research Centre, Gembloux Agro-Bio Tech, Université of Liège, 5030, Gembloux, Belgium

[*]Corresponding author. Evolutionary Biology and Ecology, Faculté des Sciences, Université Libre de Bruxelles (ULB), Av. F.D. Roosevelt 50, 1050, Brussels, Belgium.
E-mail: ranavatsurbs@gmail.com

ABSTRACT

Past climatic oscillations have influenced the genetic diversity and distribution patterns of tropical African tree species, and possibly their mating system. To explore these effects, we investigated the phylogeography of *Pericopsis elata* (Fabaceae), an endangered timber species with a high selfing rate and a fragmented Guineo-Congolian distribution with three gene pools: Upper Guinea (UG, west Africa), the Sangha River Interval (SRI, western Central Africa), and Congolia (C, Congo Basin). Our dated phylogeny of 51 plastomes indicates that *P. elata* diverged from *P. angolensis*, a dry woodland species, during the Mid-Pleistocene and had spread to UG by 210 000 years ago. Central African plastomes diverged 99 000 years ago but those from the SRI show more recent divergence. Nuclear microsatellites confirm the strong differentiation between the C and SRI clusters, and reveal contrasting evolutionary histories. While the C cluster exhibited moderate inbreeding and secondary selfing rate, the SRI cluster displayed a westward decay of diversity with high secondary selfing rate and strong fine-scale spatial genetic structure. Our findings suggest recurrent range expansion–fragmentation of *P. elata* since the Mid-Pleistocene, with a probable refugium in northern Republic of Congo, followed by a recent westward expansion into Cameroon facilitated by selfing, providing clues about the vegetation history of the SRI. The genetic peculiarities of the different gene pools must be considered in efforts to conserve and exploit this species sustainably.

Keywords: African rain forest; genetic diversity; inbreeding; *Pericopsis*; Pleistocene; range expansion

INTRODUCTION

The Guineo-Congolian rain forest in Africa is one of the largest rain forests in the world, second only to the Amazon rain forest (Malhi *et al.* 2013). This region is characterized by three floristic sub-centres of endemism, Upper Guinea (UG) in the west, Lower Guinea (LG) in the Eastern-Central, and Congolia (C) in the Western-Central African tropics (White 1979, 1983). While UG and LG are separated by the Dahomey Gap, a region currently dominated by savannas but forested during the Holocene African Humid Period that ended about 5500 years ago, LG and C are separated by the Sangha River Interval (SRI), a region nowadays covered by a mosaic of different forest formations (Gillet and Doucet 2012), which was highly perturbed during the late Holocene climatic crisis 2500–2000 years ago (Maley *et al.* 2018). The SRI was probably involved in the expansion of Bantu agriculturists (Bostoen *et al.* 2015) but few studies provide insights into the history of its vegetation (Elenga *et al.* 2004, Schmitt *et al.* 2023).

Received 25 November 2024; revised 19 March 2025; accepted 7 April 2025

Figure 1. Experimental plantation of *Pericopsis elata* established in 1945 in Yangambi (INERA), Democratic Republic of the Congo. A, *Pericopsis elata* plantation. B, Close-up of the trunk. Photographs by Olivier J. Hardy.

Past climatic changes have played a key role in shaping patterns of biodiversity and distribution of tree species across the Guineo-Congolian rain forest. In particular, the repeated cycles of glaciation during the Pleistocene presumably led to forest contraction and possibly fragmentation during the colder and drier glacial periods and expansion during the warmer and humid interglacial periods (Dupont *et al.* 2001, Hardy *et al.* 2013, Miller and Gosling 2014). While phylogeographical patterns are partially consistent across several species, they may have arisen at different timescales due to different responses of these species to climatic oscillations based on their ecological and environmental traits (Hardy *et al.* 2013, Demenou *et al.* 2020, Helmstetter *et al.* 2020, Piñeiro *et al.* 2021). Hence, studying the demographic history of species showing particular distribution ranges or ecological traits is relevant to complete our current understanding.

The signatures of past climate changes can be traced by examining past demographic events using genetic data and fossil records (e.g. Hardy *et al.* 2013, Helmstetter *et al.* 2020, Matvijev *et al.* 2022). Increasing genetic data for various tree species reveal distinct biogeographical regions (Vanden Abeele *et al.* 2021), and the extent of genetic differentiation between and within these regions provides insight into historical patterns. As several tree species of commercial importance in tropical Africa are highly exploited, examining their genetic diversity and demographic history is also necessary for their conservation and sustainable exploitation (Bourland *et al.* 2012).

Pericopsis elata (Harms) Meeuwen is a large, light-demanding, gregarious tree species (Fig. 1) with a remarkably disjoint distribution in the Western and Central African tropics (Fig. 2), and an original mixed-mating system (Bourland *et al.* 2012, Angbonda *et al.* 2021). It is a wind-dispersed species found in moist semi-deciduous forests in three main areas: in UG (between Ghana and Ivory Coast), in the SRI (from north-west Republic of Congo to eastern Cameroon) and in C [north-east of the Democratic Republic of the Congo (DRC)] (Fig. 3). A few small populations are also reported between these main areas (Vivien and Faure 1985), suggesting an ongoing fragmentation of a once more continuous range, or a species expansion marked by occasional long-distance colonization events. Logged for its excellent timber, it has been over-exploited in Western Africa where its range has been considerably reduced by agriculture. Currently, this species is only exploited in Central Africa. It is listed as 'Endangered A2cd' in the IUCN Red List (Hills 2020), and included in the CITES Appendix II due to a deficit of natural regeneration in its native distribution range (Bourland *et al.* 2012, 2015).

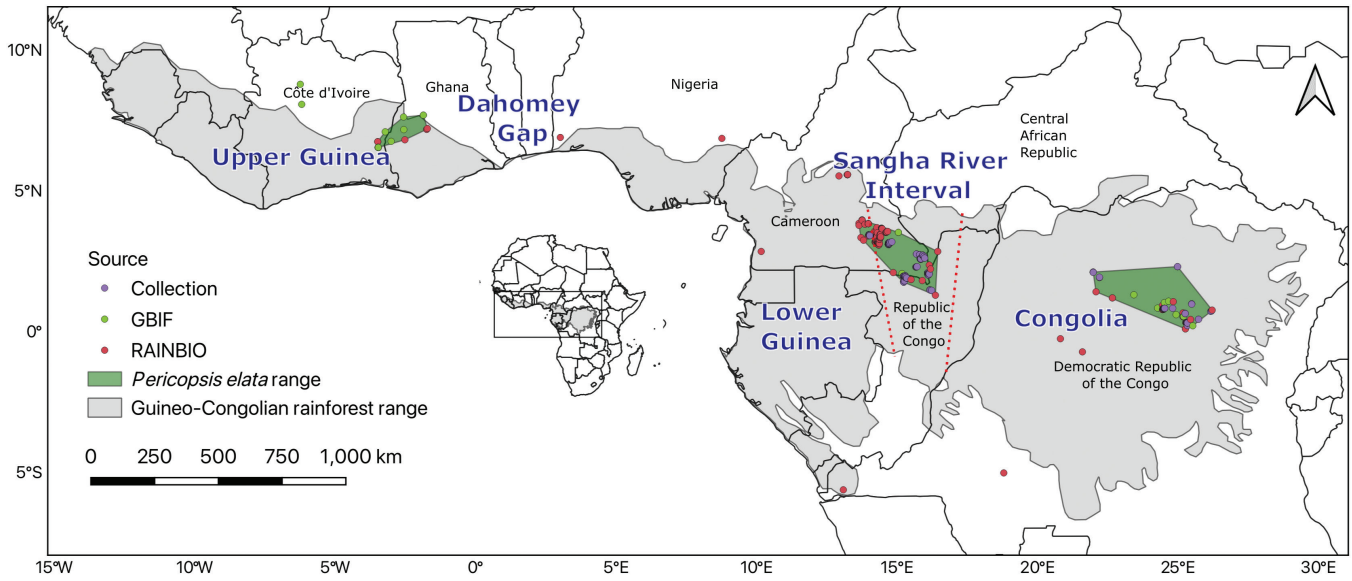

Figure 2. Distribution range of *Pericopsis elata* in tropical Africa based on the GBIF and RAINBIO databases, and collected samples. The biogeographical centres of endemism defined by White (1979) are labelled in blue and, within these, the Sangha River Interval is delineated approximately within the red dotted lines, following Elenga *et al.* (2004) and Schmitt *et al.* (2023).

Knowledge of the population structure of *P. elata* across its distribution range is central to its conservation management, as it will help inform the appropriate geographical units for consideration. Genotyping-by-sequencing data revealed the presence of three well-differentiated clades associated with the fragmented distribution of *P. elata*, where the UG clade was found to be early diverging compared to the SRI and C clades (Piñeiro *et al.* 2021; note that what we refer to as the SRI clade was called the LG-north clade by these authors who compared several species). Microsatellite markers showed that individuals from Cameroon had a much lower diversity than those from DRC, indicating a dearth of heterozygotes in the Cameroon gene pool (Micheneau *et al.* 2011). Studying the population structure within each of these groups is important, as it will help uncover where and when the populations expanded during favourable periods. The presence of a gene pool centred on the SRI could also provide insights on the vegetation history of this little known region.

P. elata has a mixed-mating system, producing 54% of selfed seeds or seedlings in DRC, although the selfing rate drops to 26% in adults, indicating inbreeding depression (Angbonda *et al.* 2021), which was confirmed in experimental plantations (Angbonda *et al.* 2024). Seeds have mainly a short dispersal range, <75 m, although occasional long-distance dispersal probably occurs in the presence of strong winds. This short-distance dispersal enables biparental inbreeding, as 80% outcrossing events happened between individuals <500 m (Angbonda *et al.* 2021). Therefore, investigating the occurrence of inbreeding and inbreeding depression is crucial for the sustainable management of remnant populations.

In this study, we hypothesize that the high selfing rate of *P. elata* led to increased genetic drift, which in turn caused strong spatial genetic structuring. We expect that this structure could help reveal the history of its populations. We utilized the nuclear microsatellite markers developed by Micheneau *et al.* (2011) to precisely delimit the gene pools occurring in Cameroon and Republic of Congo, within the SRI, and in DRC (C). The spatial patterns of inbreeding and of observed heterozygosity were also investigated to verify whether the high inbreeding rate resulting from selfing reported in a population of DRC (Angbonda *et al.* 2021) is also observed throughout Central Africa. We also aimed to detect potential diversity gradients that could highlight refuge areas as well as recently colonized areas. Finally, we utilized whole plastome data to estimate the divergence date of *P. elata* with its congeneric species, to estimate divergence dates between its gene pools, and to infer phylogeography and past colonization events within gene pools. More specifically, we aim to address the following questions: (i) How do the two gene pools in Central Africa differ in terms of genetic diversity and the extent of inbreeding? (ii) What is the crown age of *P. elata* and when did the plastid lineages from the two gene pools diverge? (iii) What is the demographic history of *P. elata* in Central Africa, and in particular within the SRI?

METHODS

Sampling and DNA extraction

Cambium or leaflets of 608 *P. elata* trees (in general adult trees) were collected in natural forests in Cameroon and Republic of the Congo (SRI), and DRC (C) between 2006 and 2014, mostly in forest concessions, and a small area (a few cm^2) was dried directly with silica-gel to preserve DNA (Fig. 3; Supporting Information Table S1). While the sampling was well spread over the distribution area in the SRI, where six subpopulations were delineated (W0–W5, 398 individuals), in C, most samples originated from two forests extending over *c.* 3 × 6 km (186 individuals), in Yangambi (Eyg) and Yoko (Eyk), the 24 remaining well-spread samples constituting another subpopulation (E0). In addition, one *P. elata* and two *Pericopsis laxiflora* (Benth. ex Baker) Meeuwen were collected in Ghana, one *P. laxiflora* in Burkina Faso, and one *Pericopsis angolensis* (Baker) Meeuwen in southern DRC.

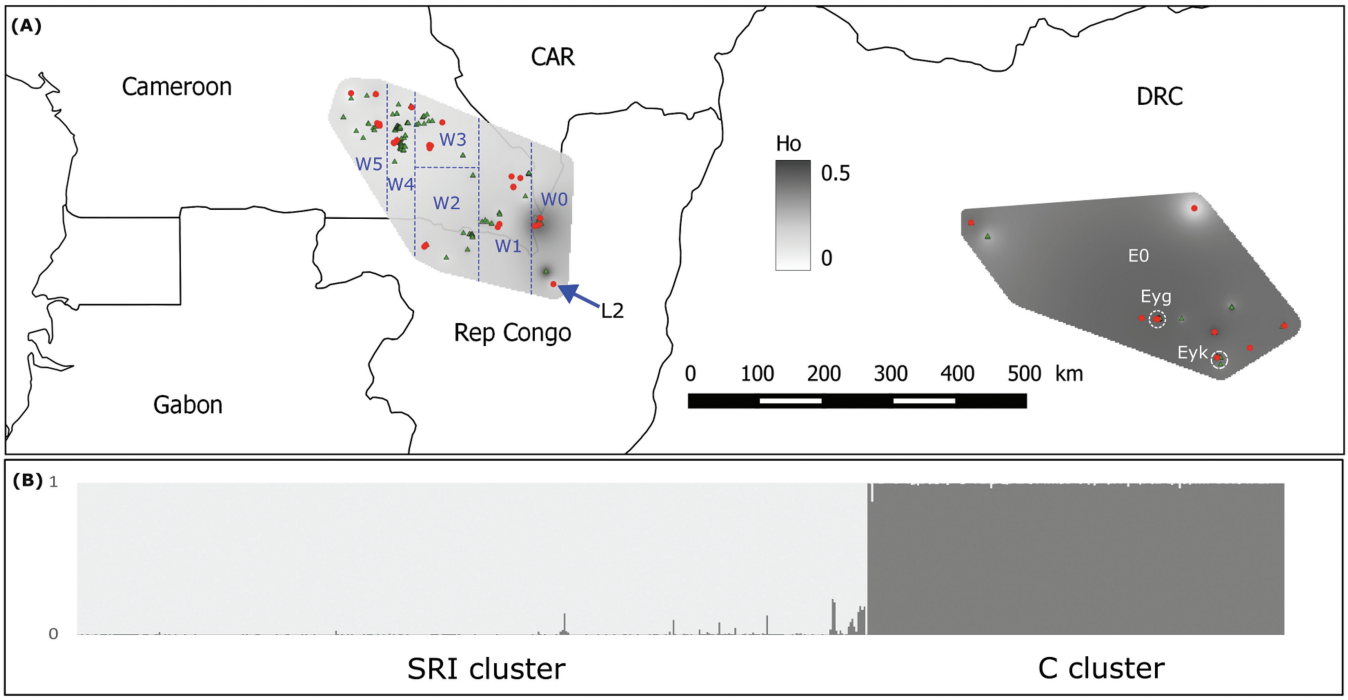

Figure 3. A, Distribution of 608 *Pericopsis elata* trees in Central Africa genotyped at 11 microsatellite loci (green triangles) among which 42 had their full plastome sequenced (red circles). The grey shaded areas delimit the estimated distribution of the species in Central Africa [Sangha River Interval (SRI) to the west and Congolia (C) to the east], considering two convex hulls surrounding extreme observations with a border of 20 km. The intensity of the grey shading is a spatial interpolation of the observed heterozygosity (H_o) measured at the individual level. The arrow shows the sole sample belonging to lineage L2, the most divergent plastome found in Central Africa. The stippled lines or circles delimit nine subpopulations with genetic diversity parameters described in Table 2. B, Bar plot of membership probabilities for each individual, ranked from west to east, according to the clustering results using the STRUCTURE algorithm, delimiting two clusters.

For these samples, DNA was extracted using the Nucleospin 96 Plant II kit (Macherey-Nagel) following the manufacturer's instructions. To complete our sampling, we also extracted DNA from two herbarium samples (one *P. angolensis* from Tanzania, one *P. laxiflora* from Ivory Coast), using a modified cetyltrimethylammonium bromide (CTAB) protocol (Doyle and Doyle 1990) that included an additional sorbitol washing step.

Library preparation for genome skimming and microsatellite genotyping

A subset of 50 samples (43 *P. elata*, three *P. angolensis*, and four *P. laxiflora*) were used to sequence their plastomes (Fig. 3; Supporting Information Table S2). Of the *P. elata* samples, 32 were from SRI (26 from Cameroon, six from Republic of the Congo), 10 from C, and one C individual from GenBank. The extracted DNA was quantified using a Qubit 2.0 Fluorimeter (Life Technologies, Invitrogen), sheared to a size of *c.* 400 bp using a Biorupter Pico Sonicator for eight cycles with 15 s on and 90 s off (Diagenode, SA), except in the case of the two herbarium samples. Library preparation was done with the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs) as per the protocol in Migliore *et al.* (2020). The libraries were pooled in equimolar concentrations and sequenced on an Illumina NovaSeq 6000 at GIGA (Liège, Belgium) at 150-bp paired-end mode.

Eleven microsatellite markers were genotyped following the protocol of Micheneau *et al.* (2011) on the 608 samples from Central Africa (Fig. 3).

Whole plastome phylogeny and divergence dating

DNA sequence quality control was done using FastQC v.0.11.9 (Andrews 2010). Reads were trimmed using fastp v.0.23.2 (Chen *et al.* 2018). Ten million reads were sampled per fastq file using seqtk (https://github.com/lh3/seqtk). The reads were then aligned to the chloroplast genome of *P. elata* available on GenBank (accession number: MZ274113.1) using the bwa mem option in BWA v.0.7.17 (Li 2013), and the aligned files were converted from SAM to BAM format, sorted, and indexed using SAMtools v.1.17 (Danecek *et al.* 2021). All these steps were carried out on the Consortium des Équipements de Calcul Intensif (CÉCI) 'Hercules2' cluster.

We performed divergence dating in two steps: (i) A family-level phylogeny allowed us to estimate the divergence between the plastomes of different *Pericopsis* species and the crown age of *P. elata* plastomes using 77 protein coding sequences (CDS) that were aligned with those of other legume tree species (Choi *et al.* 2022) using calibration points from Koenen *et al.* (2021), (ii) A genus-level phylogeny based on the complete plastid sequences of 51 available *Pericopsis* samples was time calibrated with the crown age of *P. elata* determined in the previous step to infer the phylogeographical history of *P. elata* maternal lineages.

For the family-level phylogeny, after identifying the main plastid lineages occurring within *P. elata*, we kept nine *P. elata* samples representative of each lineage, the two *P. angolensis*, and three *P. laxiflora* (Supporting Information Table S2), aligned their plastid sequences together with the annotated *P. elata*

plastome (MZ274113.1) using MAFFT, and transferred the annotations of CDS to all sequences using Geneious v.7.1.3 (https://www.geneious.com). We manually corrected the alignment and checked the presence of stop codons within CDS, eventually correcting gene boundaries. We then concatenated 77 CDS as in Choi *et al.* (2022) for other legumes and aligned the *Pericopsis* dataset with the nucleotide file from Choi *et al.* (2022) using DECIPHER v.2.28.0 (Wright 2015, 2020) (File S1). Because the substitution rate tends to be higher for non-woody taxa (Smith and Donoghue 2008), we removed all such taxa present in Choi *et al.* (2022), ending with a final alignment containing 95 taxa. We used RaxML-NG (Kozlov *et al.* 2019) to produce a maximum likelihood (ML) tree under the GTR+Gamma model and default parameters, and converted the ML tree to an ultrametric tree using ape v.5.7-1 (Paradis and Schliep 2019), phytools v.1.9-16 (Revell 2012), and geiger v.2.0.11 (Pennell *et al.* 2014) in RStudio v.4.3.1 (Rstudio Team 2020) by setting *Cercis canadensis* as the outgroup, minimum age as 66 Mya, and lambda as 1 using the 'chronopl' function in ape. Figtree v.1.4.4 (https://github.com/rambaut/figtree/) was used to convert the tree from NEXUS into newick format. Using the calibrations for the crown ages of Fabaceae as 66 Mya (SD ± 2) and of the Robinioid clade as 33.9 Mya (SD ± 2) from Koenen *et al.* (2021), we used BEAST 2.7.5 (Bouckaert *et al.* 2019) to date the divergence time of *P. elata*, carried out on the 'ebe1' (Evolutionary Biology and Ecology) server at Université Libre de Bruxelles. The XML input was prepared using BEAUTi with Gamma Site Model, HKY substitution model, relaxed clock log normal, coalescent constant population, and partitioning the codon positions as {1,2}+3. The ultrametric tree generated previously was used as the starting tree, and the operators Bactrian Subtree Slide, Narrow Exchange, Wide Exchange, and Wilson-Balding were set to zero to prevent local and global rearrangement of the tree. The Markov chain Monte Carlo (MCMC) chain length was 150 000 000 with the logs stored every 10 000 generations. Tracer v.1.7.2 was used to visualize the progress of the run. Once the Effective Sample Size (ESS) values reached ≥200 for the statistics of interest, we used LogCombiner to combine the trees file, followed by TreeAnnotator to produce a Maximum Clade Credibility (MCC) tree with median heights. To visualize the tree, we used FigTree v.1.4.4.

For the genus-level phylogeny, the whole plastomes of 42 *P. elata*, two *P. angolensis*, four *P. laxiflora* individuals, and two samples of *P. elata* and *P. angolensis* from GenBank (MZ274112, MZ274113) (Supporting Information Table S2) were aligned using MAFFT. We excluded one inverted repeat and checked the alignment manually, removing short regions that seemed misaligned or contained many missing bases or ambiguous nucleotides (usually in AT-rich regions with short repeats) (File S2).

The crown age of *P. elata* was determined as 0.217 Mya (SD ± 0.02) based on the median node age [95% highest posterior density (HPD) interval 0.194–0.238]. Using this calibration, we used BEAST 2.7.5 to date the plastid phylogeny of *P. elata* (including *P. angolensis* and *P. laxiflora*) on the 'ebe1' server at ULB. The XML input was prepared using BEAUTi with Gamma site model and HKY substitution model, strict clock, and coalescent constant population. The MCMC chain length was 50 000 000 with the logs stored every 5000 generations. Once the ESS values reached ≥200 for all the statistics, an MCC tree with median heights was produced using TreeAnnotator and visualized using Figtree v.1.4.4.

The whole chloroplast alignment of *P. elata* individuals was used to construct a median joining haplotype network using PopART v.1.7 (Leigh and Bryant 2015), where epsilon was set to zero.

Genetic diversity analyses of nuclear microsatellite data

We used the Bayesian clustering algorithm implemented in STRUCTURE v.2.3.4 (Pritchard *et al.* 2000) to identify genetic clusters among the 608 Central African *P. elata* samples, using default parameters under the admixture model, correlated allele frequencies, assuming a number of clusters ranging from $K = 1$ to 6 with 10 iterations for each, with burn-in period and number of MCMC replicates set at 100 000 each. We declared the presence of recessive alleles for each locus to consider the possible presence of null alleles. We estimated the fixation index and genetic differentiation between the two inferred clusters by computing F_{ST} and R_{ST} using SPAGeDi v.1.5d (Hardy and Vekemans 2002). We applied the allele size permutation test to assess whether R_{ST} is significantly larger F_{ST}, which would mean that stepwise mutations of microsatellites contributed to genetic differentiation and that the number of generations of isolation is not negligible compared to the reciprocal of the mutation rate (Hardy *et al.* 2003).

For each genetic cluster, we characterized the genetic diversity (rarefied allelic richness *AR*, expected heterozygosity H_e, observed heterozygosity H_o, inbreeding coefficient F_{IS}) and estimated the secondary selfing rate (i.e. proportion of selfed adult trees) based on identity disequilibrium (David *et al.* 2007), as it is a robust method in the presence of null alleles and biparental inbreeding (Hardy 2016, Bürkli *et al.* 2017). These parameters were also estimated at the level of the nine subpopulations to assess how they vary geographically, using SPAGeDi v.1.5d. We also modelled the individual heterozygosity within each cluster by spatial interpolation using the Inverse Distance Weighting GDAL method implemented in QGIS v.2.18 (QGIS.org 2023).

Finally, we characterized the spatial genetic structure (kinship–distance curve) within each cluster by computing pairwise kinship coefficients between individuals [estimator of Loiselle *et al.* (1995), implemented in SPAGeDi v.1.5d], averaging values over a set of distance intervals (delimited by 0.15, 0.3, 0.5, 1, 2, 5, 10, 100, 200, and 500 km), and regressing values on log(distance) for pairs separated by less than 2 km to quantify the level of fine-scale spatial genetic structuring using the *Sp* statistic (Vekemans and Hardy 2004). Standard errors of mean kinship and *Sp* estimates were obtained by jackknifing over loci.

RESULTS

Plastid phylogeny and divergence dating

After subsampling 10 million read pairs per sample, we obtained an average of 294 249 reads (SD ± 181 386) mapped on to the reference plastome of *P. elata* (156 426 bp), with a mean coverage of 265× (SD ± 176×). The aligned plastomes of *Pericopsis* were 126 723 nucleotides long after removing one of the inverted repeats and filtering ambiguous bases. After extracting the 77 protein coding genes used in Choi *et al.* (2022), their concatenated sequences were 67 572–67 593 bp long.

According to the family-level plastid phylogeny based on 77 protein coding genes (Supporting Information Fig. S1), the crown age of *Pericopsis* was estimated to be 3.43 Mya (95% HPD interval: 1.25–7.24) and that of *P. elata* was 0.217 Mya (95% HPD interval: 0.091–0.423). This value was used for calibrating the genus-level phylogeny (Fig. 4; Table 1).

Both the family- and genus-level phylogenies indicate that *P. laxiflora* and *P. elata* are monophyletic, while *P. angolensis* appears paraphyletic because *P. elata* plastomes are more closely related to the *P. angolensis* plastomes from southern DRC, from which they diverged 0.805 Mya (95% HPD interval: 0.582–1.057), than to a *P. angolensis* plastome from Zambia, from which they diverged 0.916 Mya (95% HPD interval: 0.670–1.209).

Among *P. elata* plastomes, five lineages (L1–L5) were identified and were well supported with a posterior probability of 1, except for the split between L3 and L4 + L5 which was 0.5 (Fig. 4B). The first two diverging lineages, L1 and L2, are each represented by a single sample: L1 was found in Ghana and diverged 0.213 Mya (95% HPD interval: 0.174–0.254), and L2 was found in Republic of Congo where it is our westernmost and southernmost sample (Fig. 3) and diverged 0.099 Mya (95% HPD interval: 0.066–0.133). Based on the haplotype network (Fig. 4A), 32 and 16 substitutions separated L1 and L2 from their most closely related haplotypes. Two lineages, L3 and L4, represented respectively by seven and four samples, were present only in DRC (Congolia), diverged from each other 0.089 Mya (95% HPD interval: 0.061–0.123), and respectively included three and two haplotypes that differed by up to four substitutions. Finally, lineage L5, represented by 31 samples, occurred only in Lower Guinea (five in Republic of Congo, 26 in Cameroon), with a crown age of 0.036 Mya (95% HPD interval: 0.021–0.054), and consisted in a well-represented haplotype (27 samples) and four singleton haplotypes differing by one or two substitutions from the most common haplotype.

Hence, from a geographical point of view, UG was represented by a single sample bearing the earliest diverging lineage L1, LG was represented by two lineages, an early diverging one (L2) with one sample in Republic of Congo and one (L5) widespread from Republic of Congo to Cameroon, and C (DRC) was represented by two related lineages (L3 and L4).

Genetic diversity and extent of inbreeding in the two clusters

STRUCTURE analysis of 608 microsatellite genotypes of *P. elata* from Central Africa showed that the log-likelihood of the data increased sharply from $K = 1$ to $K = 2$ clusters and then slightly for higher K values (Supporting Information Fig. S2). However, above $K = 2$, most individuals remained admixed between clusters that did not delimit well-defined areas. Hence, we considered two genetic clusters, a western cluster in the SRI and an eastern cluster in C (Fig. 3). There were no admixed genotypes, and no additional subclusters were identified when running STRUCTURE separately for each cluster. The marked differentiation between the two clusters was confirmed by high fixation index ($F_{ST} = 0.52 \pm 0.10$) and allele size differentiation measure ($R_{ST} = 0.60 \pm 0.15$). However, the allele size permutation test indicated that R_{ST} was not significantly higher than F_{ST}, suggesting that drift, rather than stepwise mutations, was responsible for the high level of differentiation.

Genetic diversity was higher in the C cluster than SRI, with expected and observed heterozygosities being respectively two and three times higher in the former (Table 2). Conversely, the mean inbreeding coefficient was two times higher in the SRI cluster (Table 2), and levels of identity disequilibrium in adult trees led to selfing rate estimates four times higher in the SRI cluster (53%) than in the C cluster (13%).

Genetic diversity within subpopulations showed moderate variation between eastern subpopulations but substantial variation among the western ones, characterized by a steep westward decay of genetic diversity (Table 2). Specifically, while *AR*, H_e and H_o were relatively high near the Sangha river (subpopulation W0), approaching values found in subpopulations of the C cluster (E0, Eyg, Eyk; Table 2), they were lower in the subpopulations situated up to 180 km westward (W1, W2; H_o twofold lower), and still lower in the subpopulations situated up to 300 km to the north-west (W3, W4, W5; H_o threefold lower). This spatial gradient is also illustrated by the spatial interpolation of the proportion of heterozygous loci per individual (Fig. 3).

Finally, the spatial genetic structure (kinship–distance curve) was more pronounced at a fine scale (distance range <1 km) in the SRI cluster than the C cluster (Fig. 5). The resulting *Sp* statistic at a fine scale, based on the regression slope of kinship coefficients against log(distance) for pairs separated by <2 km, was four times larger in the SRI cluster ($Sp = 0.144 \pm 0.029$) than the C cluster ($Sp = 0.032 \pm 0.005$).

DISCUSSION

Our study combining whole plastome sequencing and microsatellite markers provides an overview of the phylogenetic origin, demographic history, and degree of inbreeding of *P. elata*. In summary, this rain forest species seems to have emerged relatively recently, at most 1 Mya (Mid-Pleistocene), from a population of *P. angolensis*, which is characteristic of dry woodland and savannah from austral Africa. It must have spread across a large portion of the Guineo-Congolian forest, up to West Africa, by at least 200 000 years ago. Its current highly fragmented distribution suggests subsequent population decline and fragmentation. In Central Africa, two disjoint populations form distinct genetic clusters that would have differentiated less than 100 000 years ago, despite their strong genetic differentiation ($F_{ST} = 0.52$). The western Central African cluster, located in the SRI, shows converging evidence of a recent range expansion from the plastome phylogeny and the microsatellite diversity gradient. We hypothesize that the expansion started from a refuge area located east of the Sangha River, in the north of the Republic of Congo, and has progressed westwards into Cameroon for about 300 km. This expansion, which seems to have occurred at the end of the Pleistocene or the early Holocene, was accompanied by high inbreeding, presumably due to a high selfing rate, which might have facilitated the range expansion through reproductive assurance, and recurrent founder events.

Hereafter, we discuss the evidence supporting this overall scenario in depth, showing how *P. elata* appears original compared to other African trees, and what are the consequences for the sustainable management of this threatened species.

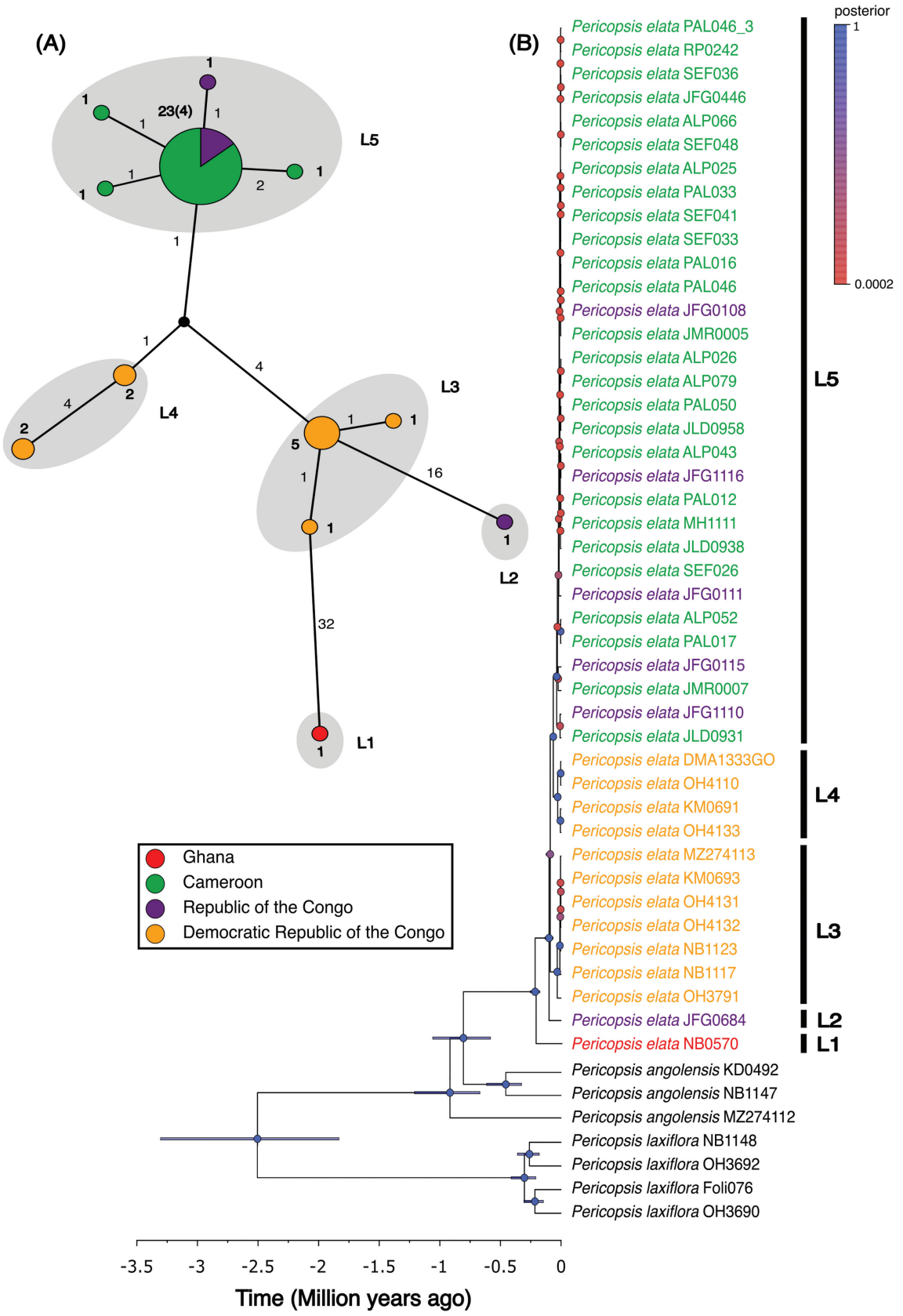

Figure 4. A, Haplotype network based on the whole plastome alignment for *Pericopsis elata* using PopART v.1.7. Numbers on the branches are the substitutions and numbers in bold represent the number of individuals per haplotype (majority number, with minority number in parentheses). We identified five lineages that are highlighted in grey (L1–L5). B, Maximum Clade Credibility (MCC) tree of the genus *Pericopsis* using BEAST v.2.7.5. Tip labels indicate the species name and collection or GenBank accession number. Clade labels and the colours of tip labels correspond to the lineages and location colours respectively from A for *P. elata*. Circles on the nodes represent the posterior probability corresponding to the gradient values on the top right corner. Purple bars indicate the 95% highest posterior density (HPD) interval.

Table 1. Stem and crown age (million years ago) of plastid lineages of *Pericopsis elata* along with the 95% highest posterior density (HPD) intervals estimated using BEAST v.2.7.5.

Lineage	Stem age [95% HPD]	Crown age [95% HPD]
Pericopsis laxiflora	2.505 [1.833, 3.304]	0.302 [0.208, 0.411]
Pericopsis angolensis	2.505 [1.833, 3.304]	0.916 [0.670, 1.209]
Pericopsis elata	0.805 [0.582, 1.058]	0.213 [0.174, 0.254]
L1[a]	0.213 [0.174, 0.254]	-
L2[a]	0.099 [0.066, 0.133]	-
L3	0.089 [0.061, 0.123]	0.031 [0.014 0.053]
L4	0.064 [0.040, 0.090]	0.028 [0.012, 0.045]
L5	0.064 [0.040, 0.090]	0.036 [0.021, 0.054]

[a]Represented only by one individual.

Table 2. Genetic diversity parameters of the Sangha River Interval (SRI) and Congolian (C) genetic clusters of *Pericopsis elata* in Central Africa based on 11 microsatellite markers; parameters for nine subpopulations arranged from west to east are also given.

Population	*N*	*AR*	H_e	H_o	*F*	*S*
SRI	398	4.25	0.24	0.14	0.41	0.53 ± 0.09
C	210	5.96	0.51	0.41	0.20	0.13 ± 0.04
Subpopulation						
W5	41	1.69	0.21	0.13	0.38	0.49 ± 0.16
W4	228	1.89	0.21	0.11	0.45	0.48 ± 0.12
W3	18	1.80	0.16	0.12	0.28	0.00 ± 0.32
W2	53	2.53	0.24	0.19	0.23	0.52 ± 0.20
W1	39	2.54	0.29	0.18	0.37	0.09 ± 0.23
W0	19	3.42	0.44	0.38	0.13	0.00 ± 0.05
E0	24	4.28	0.56	0.38	0.34	0.27 ± 0.16
Eyg	142	3.77	0.47	0.42	0.11	0.09 ± 0.05
Eyk	44	4.16	0.52	0.39	0.25	0.23 ± 0.11

N, sample size; *AR*, allelic richness (among $k = 394$ allele copies for SRI and C, or among 34 allele copies for subpopulations); H_e, expected heterozygosity; H_o, observed heterozygosity; *F*, inbreeding coefficient; *S*, selfing rate estimate (± SE).

The recent origin of *P. elata* following a biome shift

The genus *Pericopsis* Thwaites comprises four species, with three in Africa and one in the Asian tropics [*P. mooniana* (Thwaites) Thwaites, from Sri Lanka to Caroline Islands]. *Pericopsis elata and P. mooniana* are tall rainforest tree species whereas *P. angolensis* and *P. laxiflora* are small tree or shrub species found in savannahs. The time to the most recent common ancestor (TMRCA) of the three African *Pericopsis* species possibly lies in the Pliocene/Pleistocene boundary. The plastid phylogeny suggests that *P. elata* and *P. angolensis* from southern DRC are nested within *P. angolensis* from Zambia (Fig. 4B), diverging as recently as 0.9 Mya, while they are found in nonoverlapping habitats (*P. elata* in moist semi-deciduous forests and *P. angolensis* in Miombo woodlands, a wooded savannah from Austral Africa). Hence, *P. elata* being the sole African *Pericopsis* species living in rain forest and deriving from the dry woodland-dwelling species *P. angolensis*, we hypothesize that it has undergone a biome shift from savannah to forest. This is rather unusual, as Gorel *et al.* (2022) showed that the majority of biome shifts inferred among African trees occurred in the reverse direction, which is consistent with the general trend of aridification of tropical Africa since the Miocene (Couvreur *et al.* 2021), pushing rain forest species to adapt to drier environments. From its dry woodland origin, *P. elata* seems to have retained a rather good resistance to water stress, because its seedlings are able to grow under full sunlight with low mortality after being raised for a few months in a nursery (Pieters 1994, Kouadio 2009). The recent, Mid-Pleistocene, origin of *P. elata* according their plastomes (stem age *c.* 0.9 Mya, crown age = 0.25 Mya) is also unusual, as most tree species have an origin pre-dating the Pleistocene [but see De La Estrella *et al.* (2020) also for an example of recent speciation in trees]. Similarly, *P. laxiflora*, found in the savannah of Western Africa, has a crown age of 0.3 Mya and is also a recently diverged species.

Range expansion and fragmentation cycles in *P. elata*

Another peculiar feature of *P. elata* is its wide but highly fragmented distribution. At least three main geographical groups are known: one in Upper Guinea ranging from Ivory Coast to Ghana (but the species might have been wiped out of the former due to overexploitation and deforestation for cocoa), one in the Sangha River Interval ranging from eastern Cameroon to north-western Republic of Congo (our 'SRI' cluster) and one in Congolia, limited to a part of northern DRC (our 'C' cluster). Small relict populations may also occur in between these regions according to some reports (Vivien and Faure 1985) but should be confirmed. Within each region, *P. elata* is usually known as a gregarious species, so that its distribution is also aggregated at a local scale. West and Central African plastomes share an ancestor about 213 000 years ago, indicating that the species was once able to expand quickly over a vast area. Its diaspores (pods) are dispersed by wind, mostly within 100 m from the mother tree (Angbonda *et al.* 2021), but occasional long-distance dispersal probably occurs during storms. Furthermore, indehiscent pods are probably dispersed by water since the species occurs along rivers in the SRI. These factors could explain its ability to rapidly expand its range. However, the subsequent range fragmentation seems to indicate that *P. elata* alternates between phases of range expansion and phases of range fragmentation and contraction. To explain this pattern, we hypothesize that *P. elata* might flourish only when the forest cover is recurrently disturbed (e.g. periods of climatic instability, itinerant slash-and-burn agriculture), colonizing the large forest openings, while it would tend to disappear when the closed-canopy forest is stable over several centuries because its seedlings do not survive under low-light conditions (Bourland *et al.* 2015).

Whether geographical barriers could also partly explain the fragmented distribution of *P. elata* is unclear. The vast area of flooded forests that span between north-eastern Congo and western DRC is probably unsuitable for *P. elata* and could explain its disjoint distribution in Central Africa. However, we still lack reliable palaeovegetation data to reconstruct the extent of this potential barrier over the last 200 000 years.

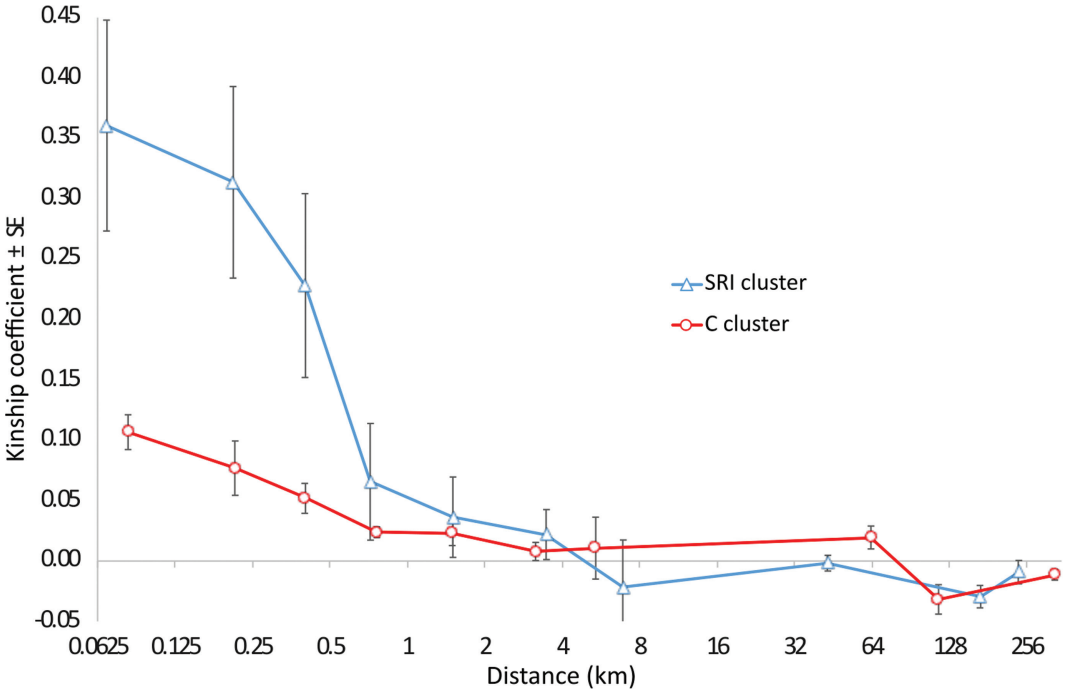

Figure 5. Spatial genetic structure (kinship–distance curve) within the Congolia (C) and Sangha River Interval (SRI) genetic clusters of *P. elata* in Central Africa (mean pairwise kinship coefficients between pairs of trees for each of 10 distance intervals and their standard errors, SE).

Currently, most natural populations of *P. elata* show a strong deficit of natural regeneration because seedlings need more light to grow and survive than what they usually find under a closed canopy, as observed in other light-demanding African canopy trees (Doucet 2003, Biwolé *et al.* 2015, Morin-Rivat *et al.* 2017, Zébazé *et al.* 2023). The fact that a significant portion of African dominant canopy tree species do not reproduce well nowadays is usually explained by changing human impacts over recent centuries (Morin-Rivat *et al.* 2017). About two centuries ago, much of the Central African forest was possibly perturbed by slash-and-burn agriculture. This practice led to the creation of patches with large gaps, favouring light-demanding tree species to thrive in temporarily abandoned fields (Bourland *et al.* 2015). European colonization pushed agricultural populations towards the main communication roads or towns, and reduced forest elephant populations, a keystone species able to maintain large openings in the forest, so that large expanses of Central African forests probably became less perturbed, reducing establishment opportunities for light-demanding species (Morin-Rivat *et al.* 2017). Although industrial logging started during this period, selective logging of a few prized species causes limited perturbation (one to three logged trees per hectare), which is not sufficient for the regeneration of the most light-demanding species (Guidosse *et al.* 2022). Hence, the history of *P. elata* might be characterized by a succession of phases of population expansion when forests were perturbed, and phases of population decline and fragmentation when forests were more stable.

It is worth noting that some other African tree species with very distinct ecologies also show evidence of range expansion dated about 200 000–300 000 years ago: *Staudtia kamerunensis*, another rain forest species, expanded throughout Congolia (Matvijev *et al.* 2022); *Podocarpus latifolius*, a montane species, spread across western Central Africa before declining recently (Migliore *et al.* 2020); and *Brachystegia* spp. trees spread from East Africa towards the western savannah of austral Africa (Boom *et al.* 2021). These seemingly concomitant events may reflect a period of abrupt environmental changes across tropical Africa, but additional evidence is needed to confirm this hypothesis because the limited precision of molecular dating does not allow us to demonstrate that these signatures of range expansion were really synchronous.

Post-glacial range expansion in Central Africa from a refuge area in the Sangha River Interval

Our plastid phylogeny suggests that the C and SRI genetic clusters of Central Africa differentiated after the Penultimate Glacial Maximum (~130 kya) at ~99 kya, followed by a second split at ~65 kya. Such recent differentiation is supported by the absence of phylogeographical signal in microsatellite allele sizes (R_{ST} not higher than F_{ST}), which is expected when the number of generations since a population split is lower than the reciprocal of the mutation rate (Hardy *et al.* 2003). Assuming a mutation rate of the order of $\mu = 10^{-3}$ for microsatellites and a generation time of the order of 100 years for canopy trees, we would expect $R_{ST} > F_{ST}$ after about 1000 generations, or 100 000 years, of isolation. Despite their recent divergence, the very high fixation index ($F_{ST} = 0.53$) between these clusters compared to other African tree species also highlights the intensity of genetic drift that mediated this differentiation. Genetic drift must have affected the SRI cluster in particular, which hosts a much lower genetic diversity than the C cluster. In fact, the SRI group appears to have a recent range expansion as 27 individuals share the same plastid haplotype, separated by one to two substitutions with four singleton haplotypes from this region, leading to a star-like haplotype network in lineage L5 (Fig. 4A) and extremely short branch lengths in the plastid phylogeny (Fig.

4B). The westward decline in microsatellite genetic diversity (Table 2) and individual homozygosity (Fig. 3) is also indicative of recent divergence and expansion into eastern Cameroon. This is in line with the relatively low genetic diversity found at plastid markers for other tree species in this region (Dauby *et al.* 2014), which could be explained by a strong decline of forest cover in the SRI during the Last Glacial Maximum followed by a postglacial recolonization. We hypothesize therefore that *P. elata* survived in a small refuge population in the eastern part of the SRI during the last glacial period. From this refuge, *P. elata* expanded westwards after the last glaciation, or possibly following forest perturbations recorded during Holocene climatic crises and characterized by extensive forest fires that occurred 8000–6500 and 2500–1500 years BP (Hubau *et al.* 2015, Maley *et al.* 2018). The later perturbation period also coincides with the development of slash-and-burn agricultural practices by the Bantu populations that started in Central Africa *c.* 3000 years ago (Bostoen *et al.* 2015). This timing is consistent with our results, as some individuals from the SRI diverged between *c.* 3300 and 1100 years ago. Being a hardwood, *P. elata* trees might have been left by Bantu farmers in newly opened fields, as they are difficult to cut down with traditional tools, favouring their regeneration after fields were abandoned (Bourland *et al.* 2012, 2015). Nevertheless, the exact timing of this range expansion requires further investigations with methods able to detect demographic changes over tens to hundreds of generations.

The second most divergent *P. elata* lineage (L2) was found in north-western Republic of Congo, along the Sangha river (Fig. 3), where microsatellite diversity was also higher (subpopulation W0) than in the rest of the SRI cluster, suggesting that this area served as a refuge for *P. elata*. In fact, the SRI seems to have been particularly perturbed by climatic and anthropogenic events about 3000–2500 years ago (Bostoen *et al.* 2015), and may therefore have been a region favourable for the persistence of forest species requiring recurrent perturbations such as *P. elata* (Bourland *et al.* 2015). Whether this perturbation occurred due to anthropogenic causes or solely due to climate crises *c.* 2500 years ago in the SRI, or a combination of the two, is contentious, as various studies point to differing histories. Some indicate that this region experienced a savannah expansion during the Late Holocene Forest Crisis (Grollemund *et al.* 2015, Maley *et al.* 2018), followed by establishment of pioneer tree species (Brncic *et al.* 2009, Morin-Rivat *et al.* 2017, Giresse *et al.* 2020). Others argue that this region did not undergo a significant savannah expansion, and was composed of patchy forests with pioneer tree species (Seidensticker 2024). Therefore, studying the demographic history of various species within the SRI could help clarify our understanding of past events in the SRI, and provide a deeper insight into its vegetation history.

High selfing rate, reproductive assurance, and colonization ability in *P. elata*

One of the peculiar features of *P. elata* compared to most tropical trees is its high selfing rate (Angbonda *et al.* 2021), detectable even among adults (Table 2). While numerous plant species have a mixed-mating system (primary selfing rate between 20% and 80%; Winn *et al.* 2011), this is much less common among trees, which are predominantly cross-pollinated (Dick *et al.* 2008). However, in a *P. elata* population from DRC (C cluster), about 55% of seeds or seedlings and about 26% of adults resulted from self-pollination (Angbonda *et al.* 2021). The lower selfing rate among adults is indicative of inbreeding depression, as confirmed by the lower growth rate of selfed than outcrossed seedlings (Angbonda *et al.* 2024).

As our genetic sampling essentially included adult trees, our selfing rate estimates (Table 2) integrate the impact of inbreeding depression on seedling/sapling survival (secondary selfing rates). In the C cluster, we inferred an overall secondary selfing rate of 13%, ranging from 9% to 27% according to subpopulations, in line with the estimate of Angbonda *et al.* (2021) for adults (26%). In the SRI cluster, secondary selfing rate estimates were even much higher (48–52% in subpopulations represented by at least 40 individuals), although low values (0–9%) were also recorded but usually in subpopulations with low sample sizes leading to high standard errors (Table 2). Whether the higher secondary selfing rate detected at the adult stage in the SRI cluster compared to the C cluster results from lower inbreeding depression or a higher primary selfing rate cannot be assessed with current data, but preliminary results suggest a higher primary selfing rate in the SRI cluster (75–85%; our unpublished data).

The higher selfing rate in the SRI cluster probably impacts the spatial genetic structure of its population, which appears stronger at a fine scale than in the eastern cluster (Fig. 5). The steeper slope of the kinship–log(distance) curve in the SRI cluster manifests only at short distances (less than a few kilometres, Fig. 5). This is confirmed by the *Sp* statistic measured at the local scale. The origin of this contrast probably lies in the higher secondary selfing rate found in the SRI cluster, reducing pollen-mediated gene dispersal, given that outcrossing pollen disperses further than seeds in *P. elata* (Angbonda *et al.* 2021).

The maintenance of a high selfing rate in *P. elata*, despite substantial inbreeding depression, may be related to the reproductive assurance it confers during range expansion. Most *P. elata* pods disperse over short distances (Angbonda *et al.* 2021) but range expansion may occur when pods are dispersed over several kilometres by strong winds, for example during thunderstorms, or along rivers. The uncommon ability of *P. elata* to generate selfed seedlings surviving until adulthood may confer it a higher potential to expand its range faster than most other wind-dispersed tree species for which trees isolated following a long-distance seed dispersal event cannot reproduce due to a lack of cross-pollination. Hence, *P. elata* founders can exhibit high levels of selfing, facilitating their establishment further into other regions (Koski *et al.* 2019). Selfing should dramatically increase genetic drift at the colonization front, so that the steep decay of genetic diversity observed in the SRI cluster probably reflects the accumulation of founder events.

Therefore, investigating whether selfing evolved along with the ability to colonize rain forests or whether *P. elata* had high selfing as its ancestral state remains to be tested. The persistence of selfing or a mixed-mating system in a long-lived tropical tree is a rare phenomenon, and hence understanding why selfing is so common in this species and whether it is contributing to a high genetic load remains to be studied. The extent of inbreeding across its geographical range can serve as a view to its evolutionary history, which remains to be investigated in the Upper

Guinean and SRI populations. An inventory of existing *P. elata* populations needs to be carried out, as it is under-sampled across its distribution range.

Implications for the conservation and sustainable management of *P. elata*

Our results provide useful information for the conservation and management of *P. elata*, which is threatened by a deficit of natural regeneration in current African forests (Bourland *et al.* 2012). Although its trade is regulated by the CITES convention, which limits the volume legally exploited in each country, *P. elata* suffers also from illegal logging in part of its range, notably in DRC (UNEP-WCMC 2018). Sylviculture (i.e. tree plantation) can help compensate for the deficit of natural regeneration and has started in some forest concessions under FSC certification (Forest Stewardship Council, a label of sustainable forest management; http://www.fsc.org/), leading to encouraging preliminary results (Doucet *et al.* 2016).

The origin of plants can be important for the success of plantations and to guarantee the conservation of the species genetic diversity. In this respect, the finding of differentiated gene pools (Piñeiro et al. 2020; Fig. 3) can orient the choice of seed sources to keep enough diversity. In the SRI, in particular, natural populations along the Sangha river host an original and relatively high genetic diversity that merits protection against overexploitation. Seed orchards should also be created to assemble sources representative of the genetic diversity at the level of the species or local cluster. Seed orchards can furthermore be used to evaluate whether some traits of interest (growth rate, good conformation, disease resistance, etc.) are inheritable and could be selected to augment the value of plantations in the future.

Another important observation is the high inbreeding of natural populations, particularly in the SRI cluster, presumably due to a high primary selfing rate. As inbreeding depression occurs in *P. elata* (Angbonda *et al.* 2021, 2024), using outcrossed seeds rather than selfed seeds could considerably improve the success of plantations (higher growth and/or lower mortality rates). Currently, only genetic markers allow assessing the selfed or outcrossed status of seedlings but if selfing rate varies among seed trees, it could be of interest to identify trees producing more outcrossed seeds as seed sources.

Finally, the very low genetic diversity and high inbreeding observed on the west side of the SRI cluster (Fig. 3), which we interpret as a colonization front, may have fostered a higher genetic load, decreasing the fitness of trees. It is therefore useful to compare the characteristics and growth of provenances situated along the genetic diversity gradient to test if populations on the colonization front suffer from lower performance.

CONCLUSION

Our study aimed to examine the demographic history of *P. elata* revealing distinct evolutionary histories between two Central African gene pools. The Congolia gene pool exhibited higher heterozygosity compared to the Sangha River Interval pool across the sampling range. Our divergence estimates suggest that *P. elata* is a relatively young species, with a recent range expansion from the Republic of Congo into Cameroon, possibly involving multiple founder events facilitated by selfing. As a tree species exhibiting high levels of inbreeding, *P. elata* shows varied population histories in Central Africa, making it a compelling study system to investigate the impact of past climatic changes. Our findings provide baseline insights for devising strategies for conservation and sustainable exploitation of this timber tree species.

SUPPLEMENTARY DATA

Supplementary data are available at *Botanical Journal of the Linnean Society* online.

ACKNOWLEDGEMENTS

We thank Claire Micheneau and Esra Kaymak for their contribution towards developing and genotyping the microsatellite markers for *P. elata*. Samples in Cameroon were collected under research permit no. 000102/MINRESI/B00/C00/C10/C13. We thank the companies Pallisco, AlpiCAM, and SEFAC (Cameroon) as well as SIFCO and CIB (Republic of Congo) for access to management units of forest concessions. Computational resources were provided by the Consortium des Équipements de Calcul Intensif (CÉCI), supported by the F.R.S-FNRS under grant no. 2.5020.11 and by the Walloon Region. We thank two anonymous referees who provided useful comments that helped to improve the manuscript.

CREDIT STATEMENT

O.J.H. conceived the study. D.M.A., G.U.D.B., B.S., N.B., J.-F.G., J.-L.D., and O.J.H. collected the samples. S.R. and S.S. performed the lab work. S.R. and O.J.H. analysed the data and wrote the manuscript. All authors commented on the final manuscript.

CONFLICT OF INTEREST

None declared.

FUNDING

This work was supported by F.R.S-FNRS (Fonds de la Recherche Scientifique) grant no. PDR T.0119.20 to O.J.H.

DATA AVAILABILITY

The sample information, multiple sequence alignments for the family-level and genus-level plastome phylogeny, the family-level tree, and the log-likelihood plot for the STRUCTURE analysis can be found in the supplementary files. The scripts used for this study can be found on the GitHub repository https://github.com/surabhiranavat/pericopsis_phylogeography

REFERENCES

Andrews S. *FastQC: A Quality Control Tool for High Throughput Sequence Data*. 2010. http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Angbonda D-MA, Monthe FK, Bourland N *et al.* Seed and pollen dispersal and fine-scale spatial genetic structure of a threatened tree species: *Pericopsis elata* (Harms) Meeuwen (Fabaceae). *Tree Genetics & Genomes* 2021;**17**:27.

Angbonda D-MA, Ilunga-Mulala CM, Bourland N *et al.* Inbreeding depression affects the growth of seedlings of an African timber species with a mixed mating reproductive system, *Pericopsis elata* (Harms) Meeuwen. *Heredity* 2024;**133**:238–48. https://doi.org/10.1038/s41437-024-00709-x

Biwolé AB, Dainou K, Fayolle A *et al.* Light response of seedlings of a central African timber tree species, *Lophira alata* (Ochnaceae), and the definition of light requirements. *Biotropica* 2015;**47**:681–8. https://doi.org/10.1111/btp.12258

Boom AF, Migliore J, Kaymak E *et al.* Plastid introgression and evolution of African miombo woodlands: new insights from the plastome-based phylogeny of *Brachystegia* trees. *Journal of Biogeography* 2021;**48**:933–46. https://doi.org/10.1111/jbi.14051

Bostoen K, Clist B, Doumenge C *et al.* Middle to late Holocene paleoclimatic change and the early Bantu expansion in the rain forests of Western Central Africa. *Current Anthropology* 2015;**56**:354–84. https://doi.org/10.1086/681436

Bouckaert R, Vaughan TG, Barido-Sottani J *et al.* BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* 2019;**15**:e1006650. https://doi.org/10.1371/journal.pcbi.1006650

Bourland N, Kouadio YL, Fétéké F *et al.* Ecology and management of *Pericopsis elata* (Harms) Meeuwen (Fabaceae) populations: a review. *Biotechnology, Agronomy, Society and Environment* 2012;**16**:486–98.

Bourland N, Cerisier F, Daïnou K *et al.* How tightly linked are *Pericopsis elata* (Fabaceae) patches to anthropogenic disturbances in Southeastern Cameroon? *Forests* 2015;**6**:293–310. https://doi.org/10.3390/f6020293

Brncic TM, Willis KJ, Harris DJ *et al.* Fire and climate change impacts on lowland forest composition in Northern Congo during the last 2580 years from palaeoecological analyses of a seasonally flooded swamp. *The Holocene* 2009;**19**:79–89. https://doi.org/10.1177/0959683608098954

Bürkli A, Sieber N, Seppälä K *et al.* Comparing direct and indirect selfing rate estimates: when are population-structure estimates reliable? *Heredity* 2017;**118**:525–33. https://doi.org/10.1038/hdy.2017.1

Chen S, Zhou Y, Chen Y *et al.* fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 2018;**34**:i884–90. https://doi.org/10.1093/bioinformatics/bty560

Choi I-S, Cardoso D, De Queiroz LP *et al.* Highly resolved Papilionoid legume phylogeny based on plastid phylogenomics. *Frontiers in Plant Science* 2022;**13**:823190. https://doi.org/10.3389/fpls.2022.823190

Couvreur TLP, Dauby G, Blach-Overgaard A *et al.* Tectonics, climate and the diversification of the tropical African terrestrial flora and fauna. *Biological Reviews* 2021;**96**:16–51.

Danecek P, Bonfield JK, Liddle J *et al.* Twelve years of SAMtools and BCFtools. *GigaScience* 2021;**10**:giab008. https://doi.org/10.1093/gigascience/giab008

Dauby G, Duminil J, Heuertz M *et al.* Congruent phylogeographical patterns of eight tree species in Atlantic Central Africa provide insights into the past dynamics of forest cover. *Molecular Ecology* 2014;**23**:2299–312. https://doi.org/10.1111/mec.12724

David P, Pujol B, Viard F *et al.* Reliable selfing rate estimates from imperfect population genetic data. *Molecular Ecology* 2007;**16**:2474–87. https://doi.org/10.1111/j.1365-294X.2007.03330.x

De La Estrella M, Cervantes S, Janssens SB *et al.* 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* 2020;**47**:2728–40. https://doi.org/10.1111/jbi.13971

Demenou BB, Migliore J, Heuertz M *et al.* Plastome phylogeography in two African rain forest legume trees reveals that Dahomey Gap populations originate from the Cameroon Volcanic Line. *Molecular Phylogenetics and Evolution* 2020;**150**:106854. https://doi.org/10.1016/j.ympev.2020.106854

Dick CW, Hardy OJ, Jones FA *et al.* Spatial scales of pollen and seed-mediated gene flow in tropical rain forest trees. *Tropical Plant Biology* 2008;**1**:20–33. https://doi.org/10.1007/s12042-007-9006-6

Doucet J-L. *L'alliance délicate De la Gestion Forestière et de la Biodiversité dans les Forêts du Centre du Gabon*. Ph.D. Thesis, University of Liége, 2003.

Doucet J-L, Daïnou K, Ligot G *et al.* Enrichment of Central African logged forests with high-value tree species: testing a new approach to regenerating degraded forests. *International Journal of Biodiversity Science, Ecosystem Services & Management* 2016;**12**:83–95. https://doi.org/10.1080/21513732.2016.1168868

Doyle JJ, Doyle JL. Isolation of plant DNA from fresh tissue. *Focus* 1990;**12**:13–5.

Dupont LM, Donner B, Schneider R *et al.* Mid-Pleistocene environmental change in tropical Africa began as early as 1.05 Ma. *Geology* 2001;**29**:195–8. https://doi.org/10.1130/0091-7613(2001)029<0195:mpecit>2.0.co;2

Elenga H, Maley J, Vincens A *et al. Palaeoenvironments, Palaeoclimates and Landscape Development in Atlantic Equatorial Africa: A Review of Key Sites Covering the last 25 Kyrs. Past Climate Variability through Europe and Africa*. Dordrecht: Springer, 2004, 181–98.

Gillet J-F, Doucet J-L. A commented checklist of woody plants in the Northern Republic of Congo. *Plant Ecology and Evolution* 2012;**145**:258–71. https://doi.org/10.5091/plecevo.2012.648

Giresse P, Maley J, Chepstow-Lusty A. Understanding the 2500 yr BP rainforest crisis in West and Central Africa in the framework of the Late Holocene: Pluridisciplinary analysis and multi-archive reconstruction. *Global and Planetary Change* 2020;**192**:103257.

Gorel A, Hardy OJ, Dauby G *et al.* Climatic niche lability but growth form conservatism in the African woody flora. *Ecology Letters* 2022;**25**:1164–76. https://doi.org/10.1111/ele.13985

Grollemund R, Branford S, Bostoen K *et al.* Bantu expansion shows that habitat alters the route and pace of human dispersals. *Proceedings of the National Academy of Sciences of the United States of America* 2015;**112**:13296–301. https://doi.org/10.1073/pnas.1503793112

Guidosse Q, du Jardin P, White L *et al.* Gabon's green gold: a bibliographical review of thirty years of research on okoumé (*Aucoumea klaineana* Pierre). *Biotechnology, Agronomy, Society and Environment* 2022;**26**:30–42.

Hardy OJ. Population genetics of autopolyploids under a mixed mating model and the estimation of selfing rate. *Molecular Ecology Resources* 2016;**16**:103–17. https://doi.org/10.1111/1755-0998.12431

Hardy OJ, Vekemans X. SPAGeDi. a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes* 2002;**2**:618–20.

Hardy OJ, Charbonnel N, Fréville H *et al.* Microsatellite allele sizes: a simple test to assess their significance on genetic differentiation. *Genetics* 2003;**163**:1467–82. https://doi.org/10.1093/genetics/163.4.1467

Hardy OJ, Born C, Budde K *et al.* Comparative phylogeography of African rain forest trees: a review of genetic signatures of vegetation history in the Guineo-Congolian region. *Comptes Rendus. Géoscience* 2013;**345**:284–96. https://doi.org/10.1016/j.crte.2013.05.001

Helmstetter AJ, Béthune K, Kamdem NG *et al.* Individualistic evolutionary responses of Central African rain forest plants to Pleistocene climatic fluctuations. *Proceedings of the National Academy of Sciences of the United States of America* 2020;**117**:32509–18. https://doi.org/10.1073/pnas.2001018117

Hills R. *Pericopsis elata*. The IUCN Red List of Threatened Species. 2020:e.T33191A67802601. https://dx.doi.org/10.2305/IUCN.UK.2020-3.RLTS.T33191A67802601.en

Hubau W, Van Den Bulcke J, Van Acker J *et al.* Charcoal-inferred Holocene fire and vegetation history linked to drought periods in the Democratic Republic of Congo. *Global Change Biology* 2015;**21**:2296–308. https://doi.org/10.1111/gcb.12844

Koenen EJM, Ojeda DI, Bakker FT *et al.* The origin of the legumes is a complex paleopolyploid hylogenomic tangle closely associated with the Cretaceous–Paleogene (K–Pg) mass extinction event. *Systematic Biology* 2021;**70**:508–26. https://doi.org/10.1093/sysbio/syaa041

Koski MH, Layman NC, Prior CJ *et al.* Selfing ability and drift load evolve with range expansion. *Evolution Letters* 2019;**3**:500–12. https://doi.org/10.1002/evl3.136

Kouadio YL. *Mesures Sylvicoles en vue d'améliorer la Gestion des Populations d'essences Forestières Commerciales de l'Est du Cameroun*. Ph.D. Thesis, University of Liége, 2009.

Kozlov AM, Darriba D, Flouri T *et al*. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* 2019;**35**:4453–5. https://doi.org/10.1093/bioinformatics/btz305

Leigh JW, Bryant D. Full-feature software for haplotype network construction. *Methods in Ecology and Evolution* 2015;**6**:1110–6.

Li H. *Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM*. 2013, **arXiv**:1303.3997v2 [q-bio.GN].

Loiselle BA, Sork VL, Nason J *et al*. Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). *American Journal of Botany* 1995;**82**:1420–5. https://doi.org/10.1002/j.1537-2197.1995.tb12679.x

Maley J, Doumenge C, Giresse P *et al*. Late Holocene forest contraction and fragmentation in central Africa. *Quaternary Research* 2018;**89**:43–59. https://doi.org/10.1017/qua.2017.97

Malhi Y, Adu-Bredu S, Asare RA *et al*. African rainforests: past, present and future. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences* 2013;**368**:20120312. https://doi.org/10.1098/rstb.2012.0312

Matvijev K, Dellicour S, Kaymak E *et al*. Spatially explicit phylogeographical reconstruction sheds light on the history of the forest cover in the Congo Basin. *Journal of Biogeography* 2022;**49**:2256–68. https://doi.org/10.1111/jbi.14507

Micheneau C, Dauby G, Bourland N *et al*. Development and characterization of microsatellite loci in *Pericopsis elata* (Fabaceae) using a cost-efficient approach. *American Journal of Botany* 2011;**98**:e268–70. https://doi.org/10.3732/ajb.1100070

Migliore J, Lézine A-M, Hardy OJ. The recent colonization history of the most widespread Podocarpus tree species in Afromontane forests. *Annals of Botany* 2020;**126**:73–83. https://doi.org/10.1093/aob/mcaa049

Miller CS, Gosling WD. Quaternary forest associations in lowland tropical West Africa. *Quaternary Science Reviews* 2014;**84**:7–25. https://doi.org/10.1016/j.quascirev.2013.10.027

Morin-Rivat J, Fayolle A, Favier C *et al*. Present-day central African forest is a legacy of the 19th century human history. *eLife* 2017;**6**:e20343. https://doi.org/10.7554/eLife.20343

Paradis E, Schliep K. Ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 2019;**35**:526–8. https://doi.org/10.1093/bioinformatics/bty633

Pennell MW, Eastman JM, Slater GJ *et al*. geiger v2.0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. *Bioinformatics* 2014;**30**:2216–8. https://doi.org/10.1093/bioinformatics/btu181

Pieters A. *Natural Regeneration in the Equatorial Forest of the Yangambi Region, Applied to Afrormosia elata Harms*. Leuven, 1994.

Piñeiro R, Hardy OJ, Tovar C *et al*. Contrasting genetic signal of recolonization after rainforest fragmentation in African trees with different dispersal abilities. *Proceedings of the National Academy of Sciences of the United States of America* 2021;**118**:e2013979118. https://doi.org/10.1073/pnas.2013979118

Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics* 2000;**155**:945–59. https://doi.org/10.1093/genetics/155.2.945

QGIS.org, 2023, *QGIS Geographic Information System*. Open Source Geospatial Foundation Project. http://qgis.org

Revell LJ. phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* 2012;**3**:217–23. https://doi.org/10.1111/j.2041-210x.2011.00169.x

Rstudio Team. *RStudio: Integrated Development for R*. 2020. http://www.rstudio.com/

Schmitt DN, Lupo KD, Amaye GT *et al*. New Holocene radiocarbon dates from the heart of the Sangha River Interval, Southwestern Central African Republic. *Journal of African Archaeology* 2023;**21**:81–91. https://doi.org/10.1163/21915784-bja10025

Seidensticker D. Pikunda-Munda and Batalimo-Maluba. *African Archaeological Review* 2024;**41**:317–44. https://doi.org/10.1007/s10437-024-09576-7

Smith SA, Donoghue MJ. Rates of molecular evolution are linked to life history in flowering plants. *Science* 2008;**322**:86–9. https://doi.org/10.1126/science.1163197

UNEP-WCMC. *Report on Species/Country Combinations Selected for Review by the Plants Committee Following CoP17*. Cambridge: UNEP World Conservation Monitoring Centre, 2018.

Vanden Abeele S, Janssens SB, Piñeiro R *et al*. Evidence of past forest fragmentation in the Congo Basin from the phylogeography of a shade-tolerant tree with limited seed dispersal: *Scorodophloeus zenkeri* (Fabaceae, Detarioideae). *BMC Ecology and Evolution* 2021;**21**:50. https://doi.org/10.1186/s12862-021-01781-1

Vekemans X, Hardy OJ. New insights from fine-scale spatial genetic structure analyses in plant populations. *Molecular Ecology* 2004;**13**:921–35. https://doi.org/10.1046/j.1365-294x.2004.02076.x

Vivien J, Faure J-J. *Arbres Des Forêts Denses d'Afrique Centrale*. Paris: Ministére des relations extérieures, cooperation et développement, Agence de cooperation culturelle et technique, 1985, 414–5.

White F. The Guineo-Congolian region and its relationships to other Phytochoria. *Bulletin du Jardin botanique national de Belgique/Bulletin van de National Plantentuin van België* 1979;**49**:11–55.

White F. *The Vegetation of Africa; a Descriptive Memoir to Accompany the UNESCO/AETFAT/UNSO Vegetation Map of Africa*. Paris: UNESCO, 1983.

Winn AA, Elle E, Kalisz S *et al*. Analysis of inbreeding depression in mixed-mating plants provides evidence for selective interference and stable mixed mating. *Evolution* 2011;**65**:3339–59. https://doi.org/10.1111/j.1558-5646.2011.01462.x

Wright ES. DECIPHER: harnessing local sequence context to improve protein multiple sequence alignment. *BMC Bioinformatics* 2015;**16**:1–14.

Wright ES. RNAconTest: comparing tools for noncoding RNA multiple sequence alignment based on structural consistency. *RNA* 2020;**26**:531–40. https://doi.org/10.1261/rna.073015.119

Zébazé D, Gorel A, Gillet J-F *et al*. Natural regeneration in tropical forests along a disturbance gradient in South-East Cameroon. *Forest Ecology and Management* 2023;**547**:121402. https://doi.org/10.1016/j.foreco.2023.121402