Contrasting genetic signal of recolonization after rainforest fragmentation in African trees with different dispersal abilities

Rosalía Piñeiro[a,b,c,1], Olivier J. Hardy[b], Carolina Tovar[d], Shyam Gopalakrishnan[a], Filipe Garrett Vieira[a], and M. Thomas P. Gilbert[a,e]

[a]The GLOBE Institute, University of Copenhagen, 1353 Copenhagen, Denmark; [b]Unit of Evolutionary Biology and Ecology, Faculté des Sciences, Université Libre de Bruxelles, B-1050 Brussels, Belgium; [c]Geography, College of Life and Environmental Sciences, CLES, University of Exeter, Exeter EX4 4RJ, United Kingdom; [d]Jodrell Laboratory, Royal Botanic Gardens, Kew, Richmond TW9 3AB, United Kingdom; and [e]University Museum, Norwegian University of Science and Technology, N-7491 Trondheim, Norway

Edited by Pamela S. Soltis, University of Florida, Gainesville, FL, and approved May 10, 2021 (received for review August 2, 2020)

Although today the forest cover is continuous in Central Africa, this may have not always been the case, as the scarce fossil record in this region suggests that arid conditions might have significantly reduced tree density during the ice ages. Our aim was to investigate whether the dry ice age periods left a genetic signature on tree species that can be used to infer the date of the past fragmentation of the rainforest. We sequenced reduced representation libraries of 182 samples representing five widespread legume trees and seven outgroups. Phylogenetic analyses identified an early divergent lineage for all species in West Africa (Upper Guinea) and two clades in Central Africa: Lower Guinea-North and Lower Guinea-South. As the structure separating the Northern and Southern clades—congruent across species—cannot be explained by geographic barriers, we tested other hypotheses with demographic model testing using δαδι. The best estimates indicate that the two clades split between the Upper Pliocene and the Pleistocene, a date compatible with forest fragmentation driven by ice age climatic oscillations. Furthermore, we found remarkably older split dates for the shade-tolerant tree species with nonassisted seed dispersal than for light-demanding species with long-distance wind-dispersed seeds. Different recolonization abilities after recurrent cycles of forest fragmentation seem to explain why species with long-distance dispersal show more recent genetic admixture between the two clades than species with limited seed dispersal. Despite their old history, our results depict the African rainforests as a dynamic biome where tree species have expanded relatively recently after the last glaciation.

phylogeography | Tropical Africa | genotyping by sequencing | spatial gradients of genetic diversity | glacial refugia

The rainforest cover in Tropical Africa has fluctuated widely over time. Today, the rainforests of West Africa (Upper Guinea) are disconnected from Central Africa (Lower Guinea) by the Dahomey Gap, a forest-savannah corridor along the coast of Benin, Togo, and eastern Ghana (Fig. 1). However, the fossil record shows that this region was forested under the humid conditions of the Holocene (from ∼8,400 y B.P.) while the current deforestation started only 4,500 y ago following the aridification of the climate (1). It is not clear whether the Dahomey Gap also became forested during the previous interglacials, as they do not seem to have been as humid as the last one (2). While the rainforests of Central Africa currently exhibit a continuous distribution, previous genetic studies indicate strong differentiation of tropical trees within the forest. Typically, such structure may be explained by geographic barriers. For instance, the river Sanaga, one of the main rivers of Cameroon, runs from inland Cameroon toward the coast, delimiting two different subspecies of chimpanzee (3) and also creating a deep genetic divergence of mandrill populations on both sides (4). In Gabon, the Ogooué river acts as an effective barrier for dispersal of mandrills (4) and gorillas (5). In contrast to primates, the genetic structure of tropical trees cannot be explained by effective barriers to dispersal by the main rivers and mountain chains in this area (6, 7).

Recent studies based on chloroplast DNA, nuclear microsatellites, and low-copy nuclear genes (6, 7) suggest that the historical isolation of the tree populations in Tropical Africa was caused by forest fragmentation during the cold and dry ice age periods, which occurred on several cycles, through the Pleistocene (8, 9). However, dating the fragmentation of the rainforest and determining where the ancestral populations of each species were has been challenging due to the low number of markers investigated. For example, with regards to dating, three prior studies have attempted to estimate the divergence of tree populations in the Pleistocene. These studies resulted in large amounts of uncertainty in age estimates due to the low numbers of molecular markers used (10–12). More recent studies based on the sequencing of whole plastomes or nuclear genes captured by baits place the divergence in the Pliocene/Pleistocene (13–15). As for the location of forest fragments that allowed tropical tree species to survive the ice ages, these have been postulated based on the fossil record and paleoclimatic reconstructions. Areas that harbor high species richness have been proposed as refugia

EVOLUTION

Significance

Although today the rainforest is continuous in Central Africa, the scarce fossil record suggests that arid conditions during the ice ages might have reduced tree density. However, the vast majority of the fossil pollen cores preserved in Tropical Africa are too young to inform about this period. Investigating whether the climate change left a genetic signature on trees can thus be useful to date past forest fragmentation. We use DNA technology to study five legume trees. Our results show significant differentiation of the populations of each species at a date compatible with forest fragmentation driven by ice age climatic oscillations. Contrasted timescales were obtained for each species, in agreement with different recolonization abilities in an expanding forest biome after fragmentation.

Author contributions: R.P., O.J.H., C.T., and M.T.P.G. designed research; R.P. performed research; R.P. contributed new reagents/analytic tools; R.P., C.T., S.G., and F.G.V. analyzed data; R.P., O.J.H., and M.T.P.G. wrote the paper; and M.T.P.G. hosted the lab work and analyses.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Published under the PNAS license.

[1]To whom correspondence may be addressed. Email: rosalia.pineiro@gmail.com.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2013979118/-/DCSupplemental.

Published July 1, 2021.

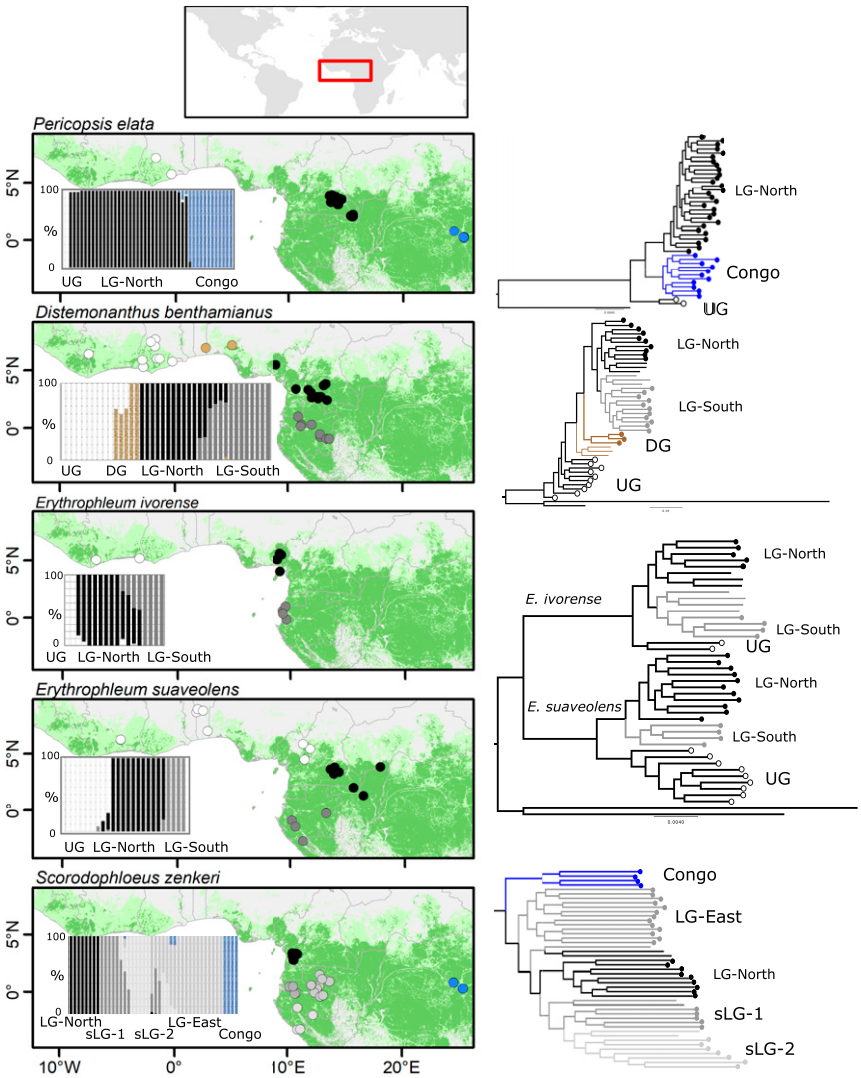

Fig. 1. Genetic ancestry and phylogenetic reconstructions for *P. elata*, *D. benthamianus*, *E. ivorense*, *E. suaveolens*, and *S. zenkeri* based on GBS. (*Left*) ADMIXTURE analyses. Bar plots show the probability of assignment of individuals to genetic clusters. (*Center*) Geographic distribution of individual trees according to the genetic cluster they belong to (admixed individuals with <70% assignment probability were removed). (*Right*) RAxML phylogenetic analyses with outgroups. Branch tips are colored by genetic cluster (branches without dots correspond to admixed individuals). Green areas in map represent present-day rainforest cover (69) (*SI Appendix*, Fig. S1). The genetic clusters were Upper Guinea (UG), Dahomey Gap (DG), Lower Guinea North (LG-North), Lower Guinea South (LG-South), Lower Guinea East (LG-East), and Congo.

(Fig. 2), assuming declines in the number of species outside the hypothesized refugia (16). Likewise, declines of genetic diversity with distance from refugia are expected to result from recolonization after forest fragmentation. Despite the potential of genetic diversity gradients to help locate the areas in which forest species survived, the available genetic data lack resolution to assess diversity gradients over space.

To overcome these challenges, we generated reduced representation genomic data using Illumina HiSeq2000 sequencing technology for five legume rainforest tree species in order to investigate the genetic signal of changes in the rainforest cover during the ice ages. All five species—*Pericopsis elata* (Harms) Meeuwen, *Distemonanthus benthamianus* Baill, *Erythrophleum ivorense* A. Chev., *Erythrophleum suaveolens* (Guill. & Perr.) Brenan, and *Scorodophloeus zenkeri* Harms—are widespread in the African rainforest (Fig. 1) and exhibit differences in light tolerance and dispersal capacity (*SI Appendix*, Table S1). Their ecology and dispersal biology is well known. *P. elata* (17) and *D. benthamianus* (18) are the most adapted to long-distance colonization. They are both light-demanding and have wind-dispersed seeds, with evidence of fat-tailed seed dispersal distributions, indicative of long-distance dispersal events. The two *Erythrophleum* spp. are light-demanding, and, besides their primary ballistic seed dispersal, they exhibit secondary dispersal of their seeds by animals, but no evidence of long-distance dispersal has been detected in direct measurements with molecular markers (18). *S. zenkeri,* a strict shade-tolerant species with limited seed dispersal nonassisted by wind or animals, exhibits the most limited colonizing capacity (19). In particular, we aimed to solve the following questions:

- Are the tree populations in West African rainforests (Upper Guinea) well-differentiated from Central African rainforests (Lower Guinea) or did the expansion of the rainforest during the humid Holocene favor gene flow and dispersal between the two forest blocks?

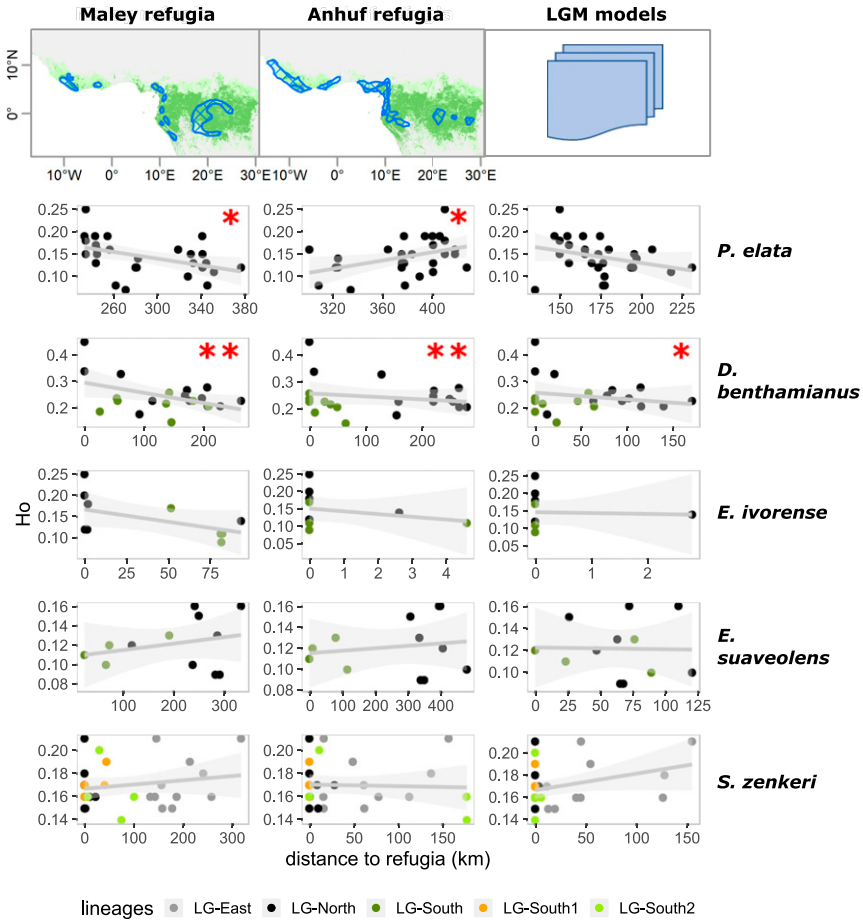

Fig. 2. Decline of genetic diversity, based on GBS, with distance from refugia in Central Africa (biogeographic region of Lower Guinea, LG) for *P. elata*, *D. benthamianus*, *E. ivorense*, *E. suaveolens*, and *S. zenkeri*. Linear regression of the genetic diversity—observed heterozygosity, H_o—of each genotyped individual tree against the geographical distance from the closest LGM refugia ~20,000 y B.P. and the genetic cluster as a random variable in order to account for genetic diversity differences across genetic lineages in Lower Guinea: North (LG-North), South (LG-South), and East (LG-East). Admixed individuals were excluded. Three different hypotheses of LGM forest refugia were considered. The forest refugia postulated by Maley and Anhuf based on paleoclimatic and palynological data: 1) LGM-Maley (16) and 2) LGM-Anhuf (41). In addition, specific forest refugia based on the LGM niche models of each species were tested (*SI Appendix*, Fig. S4): 3) LGM species niche models (see *Materials and Methods*). Significant correlations are indicated by $*P < 0.05$ and $**P < 0.01$.

- Are the divergence times estimated by sequencing large portions of the genome compatible with isolation and fragmentation of the Central African rainforest during the dry and cold ice ages?
- Can we trace the genetic signal of recolonization of tree populations from putative glacial refugia?
- What are the similarities and differences in the patterns of genetic diversity and differentiation in the light of different ecologies and seed and pollen dispersal capacities of the five tree species?

Results

A total of 182 samples were successfully sequenced (175 samples of the five study species and 7 outgroup taxa). Samples received an average number of reads between 3,387,236 in the *P. elata* and 5,518,179 in *D. benthaminanus* (*SI Appendix*, Table S2). Our protocol compensated the limitation of uneven coverage across loci typical of genotyping by sequencing (GBS) studies by building and sequencing two libraries per individual. For the GBS genotype calls without outgroups done with TASSEL, between 10,665 (*D. benthamianus*) and 27,838 single-nucleotide polymorphisms (SNPs) (*E. suaveolens*) were present in at least 50% of individuals. For the genotype likelihood framework, we retained between 568,240 reads in *P. elata* (16 depth coverage) and 1,561,119 reads in *D. benthamianus* (32 depth coverage). The number of SNPs ranged from 17,921 in *D. benthamianus* to 70,271 in *S. zenkeri* for the less stringent quality cutoff filtering 1 (Table 1 and *SI Appendix*, Table S2).

Genetic Ancestry and Phylogenetic Reconstructions.

Upper Guinea and Lower Guinea. We conducted RAxML phylogenetic analyses for each species using the GBS genotype calls with outgroups (Fig. 1, *Right* and *SI Appendix*, Fig. S1). For the species present in West Africa (*P. elata*, *D. benthamianus*, *E. ivorense*, and *E. suaveolens*) the Upper Guinean (UG) populations clustered in independent clades that were sister to the Lower Guinean (LG) clades. Similarly, ADMIXTURE bar plots of probability of assignment of individuals to populations (Fig. 1, *Left*) showed separate UG genetic clusters when the number of ancestral populations (K) was set to K = 3 (*E. ivorense*, *E. suaveolens*, and *P. elata*) or K = 4 (*D. benthamianus*). In the case of the species widespread in LG—*D. benthamianus*, *E. ivorense*, *E. suaveolens*, and *S. zenkeri*—two clades stand out for LG-North and LG-South (Fig. 1, *Right* and *SI Appendix*, Fig. S1). Similarly, in the ADMIXTURE analyses, a split between the LG-North

Table 1. Demographic history of *D. benthamianus*, *E. suaveolens*, and *S. zenkeri* using δaδi

	Dataset filtering 1			Dataset filtering 2		
	D. benthamianus	*E. suaveolens*	*S. zenkeri*	*D. benthamianus*	*E. suaveolens*	*S. zenkeri*
Divergence time (ky B.P.)						
LG-North – LG-South	28.98	359.28	3,436.99	7.70	101.78	3,085.94
UG – (LG-North, LG-South)	1,224.79	3,493.44		115.53	1,053.05	
LG-South1 – LG-South2			1,462.52			174.03
Model details						
Log-likelihood	−15,764.60	−25,466.29	−17,122.81	−11,098.82	−1,4168.11	−13,590.11
No. of SNPs	17,921.38	61,667.59	70,271.16	11,323.50	35,042.61	54,156.53

Divergence time in 1,000 y are shown for the best-fitting model under IM. Based on the admixture and phylogeny results, for the species widespread in Upper Guinea (UG) and Lower Guinea (LG)—*D. bentamianus*, *E. suaveolens*—we fitted a three-population model with the tree topology (UG, (LG-North, LG-South)). In addition to a model with symmetric migration among the three populations, we fitted a model with no migration between UG and LG-South after the split. Note that the best-fitting model for the three species does not include a migration component between UG and LG-South. For *S. zenkeri*, absent in UG, we estimated split times based on the tree topology (LG-North, (LG-South1, and LG-South2)). The demographic parameter estimates are provided for two datasets per species, depending on the stringency of the filtering criteria: filtering 1, the most lenient, and filtering 2, the intermediate filtering. Filtering 3 is not shown since it resulted in far fewer variable sites than filtering 1 and 2 (*SI Appendix*, Table S3).

and the LG-South genetic clusters was observed (Fig. 1) at K = 3 for *E. ivorense* and *E. suaveolens*, and at K = 4 for *D. benthamianus* and *S. zenkeri*. The two species sampled in Congo (*P. elata* and *S. zenkeri*) revealed independent genetic groups based on RAxML and ADMIXTURE analyses. However, the geographic coverage of our samples in this biogeographic region is incomplete.

D. benthamianus. The ADMIXTURE analysis at K = 2, where the minimum cross validation (CV) was found, revealed an UG and an LG cluster, with genetically intermediate samples in the Dahomey Gap. At K = 3, the Dahomey Gap is retrieved as an independent cluster, with admixture from the UG region in the west. At K = 4, in addition to the UG and Dahomey Gap clusters, LG splits into two clusters LG-North and LG-South, with ancestry shared between the two groups in the geographically intermediate areas. From K = 5, the samples with shared ancestry between LG-North and LG-South form an independent cluster. From K ≥ 6, further genetic subgroups are found within the UG and LG-North, with no geographical congruence. The rooted maximum likelihood phylogenetic tree agrees with K = 4. It consists of several basal clades for UG, two well-supported clades in LG-North and LG-South, and an intermediate clade for the Dahomey Gap. The admixed individuals fell in between the main clades.

E. ivorense. The ADMIXTURE analyses do not show a clear minimum in CV values. At K = 2, samples from UG and LG-North clustered together in one group and LG-South samples in a second cluster, with admixed individuals between the LG-North and the LG-South. At K = 3, the clustering corresponds to UG, LG-North, and LG-South, with admixed individuals in between the two latter. From K = 4 and higher, genetic subgroups were distinguished within the LG-North and the LG-South clusters, with no coherence across K. The rooted RAxML phylogeny revealed three main clades in agreement with K = 3: a basal clade in UG and two sister clades corresponding to LG-North and LG-South. The admixed individuals between LG-North and LG-South were placed together with the Southern cluster.

E. suaveolens. The ADMIXTURE analyses split the samples into an UG and an LG cluster at K = 2, where the minimum CV is reached. At K = 3, the genetic clusters retrieved were UG, LG-North, and LG-South. From K = 4 and higher, subgroups appear randomly within UG, and from K = 6 and higher within LG-North. The RAxML phylogenetic tree retrieves three clades corresponding to UG, LG-North, and LG-South. No individuals with shared ancestry between groups were detected. In this species, the UG cluster does not correspond to the West African Guineo-Congolian forests but to gallery forests embedded in West African savannahs, including the forest-savannah mosaic in the transition zone between the Guineo-Congolian and the Sudanian regions sensu White (20).

S. zenkeri. No clear minimum CV was shown in the ADMIXTURE analyses. At K = 2, LG-South splits from the rest of the samples. At K = 3, the following clusters are observed: LG-North, LG-South, and an LG-East/Congo, with significant admixture between LG-South and LG-East/Congo. At K = 4, the LG-East and the Congo clusters are retrieved as independent groups. At K = 5, the LG-South is split into a Northern (LG-South1) and a Southern cluster (LG-South2), with admixed individuals between them, and also with LG-East. From K = 6 and higher, random genetic groups arise within LG-South1, LG-South2, and LG-East.

P. elata. At K = 2, LG-North splits from UG/Congo (minimum CV). At K = 3, UG, LG-North, and Congo were revealed. From K = 4, random subgroups within LG-North and Congo are retrieved. The rooted RAxML phylogenetic tree revealed three well-supported clades in agreement with K = 3, where the UG clade is basal with respect to the two LG sister clades.

Demographic Inference. We inferred the demographic history of *E. suaveolens*, *D. benthamianus,* and *S. zenkeri* using δaδi (Table 1 and *SI Appendix*, Fig. S2 and Table S3). The fit substantially improved when considering a scenario of no migration between Upper Guinea and South Lower Guinea after divergence. For each scenario, different levels of data filtering gave different estimates but were congruent for each species. The scenarios with more SNPs (filtering level 1) gave older split dates while the scenarios with fewer SNPs (filtering level 2) fitted slightly more recent splits. The most restrictive filtering (filtering level 3) yielded fewer than 10,000 SNPs in all cases, so estimates are not reliable.

For the divergence between LG-North and LG-South, δaδi produced different estimates across species (Table 1). The most likely models estimated the most recent split ca. 29 ky B.P. for *D. benthamianus*, followed by *E. suaveolens* ∼359 ky B.P. The oldest was in *S. zenkeri* ca. 3.4 My B.P. The scenarios built using filtering level 2 generally inferred younger North–South splits, but the differences across species persisted (ca. 7.7 ky in *D. benthamianus*, ca. 102 ky B.P. in *E. suaveolens*, and ca. 3.1 My B.P. in *S. zenkeri*).

Gradients of Genetic Diversity over Space. We traced the spatial signal of recolonization after forest fragmentation by relating genetic diversity of individuals per species with distance from

hypothesized Last Glacial Maxima (LGM) forest refugia, accounting for differences across gene pools (Fig. 2). We found a significant negative relationship between observed heterozygosity, H_o, and distance to refugia only for *D. benthamianus* and *P. elata* (Fig. 2 and Table 2). Higher genetic diversity is found in individuals of *D. benthamianus* that are closer to the LGM forest refugia: LGM-Maley ($P < 0.01$), LGM-Ahnuf ($P < 0.01$), and LGM-species niche model ($P < 0.05$). The present-day distribution of *D. benthamianus* is mostly closer to the coast rather than in the core of the Congo forest (Fig. 1). All three postulated refugia suggest large areas of forest survival during the LGM along the coasts of Cameroon and Gabon; thus the observed gradient in H_o is dominated by the distance to the Coastal refugia rather than to the Congo refugia (Fig. 2). For *P. elata*, we detected a significant decline of genetic diversity as individuals are located farther away from LGM-Maley ($P < 0.05$). As the distribution of *P. elata* lies between the Coastal and the Congo refugia (Fig. 1 and *SI Appendix*, Fig. S3), we tested the relationship between H_o and the distances to both refugia separately. Here, we found a significant negative relationship with distance to the LGM-Maley-Congo ($P < 0.01$) and a positive relationship with scenarios LGM-Maley-Coastal. We also found a positive significant relationship with LGM-Anhuf (Fig. 2 and Table 2), where coastal refugia are closer to current *P. elata* populations. Finally, the LGM-species niche model, which shows restricted inland LGM refugia, does not exhibit any significant relationship (*SI Appendix*, Fig. S4).

In the remaining species, *E. ivorense*, *E. suaveolens*, and *S. zenkeri*, no significant relationships were found (Fig. 2, Table 2, and *SI Appendix*, Fig. S3). In the case of *E. ivorense*, the results need to be taken with caution as low sampling sizes may have reduced statistical power.

Discussion

For the five rainforest tree species investigated, at least three main intraspecific lineages were identified in the phylogenetic analyses. An early divergent lineage in West Africa (Upper Guinea) was detected in all species occurring in this forest block (*D. benthamianus*, *E. suaveolens*, *E. ivorense*, and *P. elata*). For all species widespread in Central Africa (Lower Guinea), two lineages were retrieved: LG-North and LG-South (*D. benthamianus*, *E. suaveolens*, *E. ivorense*, and *S. zenkeri*).

Early Divergence in West African Rainforests. The finding of early divergent lineages in West Africa for all the species is in tune with the hypothesis that Upper Guinea is an independent biogeographic region with numerous endemic species (20–23). Although the fossil record indicates that the Dahomey Gap was forested in the last interglacial, and thus Upper and Lower Guinea were connected between 8,400 and 4,500 y ago (1, 13), our data show that no genetic homogenization between the two forest blocks occurred. It is likely that previous interglacials were less humid than the last one, so there is no guarantee that the Dahomey Gap became forested during those periods (2). This seems to have favored the long-term differentiation of the two forest blocks. Previous genetic studies of tropical trees indicate genetic divergence between UG and LG populations (6, 7, 9, 24). However, very few studies estimate well-supported intraspecific phylogenies between UG and LG plant populations using outgroups in order to estimate the direction of evolution (10, 13). Based on relatively small numbers of markers, early divergent lineages, c.a. 500 ky. B.P., have also been found in West African populations of chimpanzees (3, 25, 26), woodpeckers (27), and frogs (28). Studies on other African lowland rainforest birds (29–31), forest-dwelling rodents (32), and African bushbucks suggest that haplotypes are rarely shared between populations sampled across the Dahomey Gap, although the relationships between clades are not well resolved. Finally, genome-wide data in lizards, snakes, and frogs revealed different levels of genetic divergence across taxa (33).

North–South Genetic Differentiation in Central Africa Suggests Rainforest Fragmentation during the Ice Ages. For each species, our phylogenetic analyses identified Northern and Southern lineages in Central Africa that do not correspond to any geographic barrier or current discontinuity in the distribution of the forest. These results are consistent with the genetic structuring of other tropical trees (6, 7, 14, 15, 34), frogs, and geckos (28, 35, 36) in Lower Guinea, as well as with the ADMIXTURE analyses, which provided additional insight into the levels of admixture among the central African lineages of each species. While *D. benthamianus* and *E. ivorense* showed admixture between the north and the south, *E. suaveolens* and *S. zenkeri* exhibited sharp differentiation and no admixed individuals between the two regions.

Using *δαδι* to reconstruct the population history of the four species present in Northern and Southern Lower Guinea, our best-fit comparison consistently supported models involving North–South divergence and subsequent gene flow. We also noticed that the fit substantially improved when considering no migration between Upper Guinea and Southern Lower Guinea after divergence. These clades diverged within the Upper Pliocene and the Pleistocene, with the oldest genetic split found in *S. zenkeri* ca. 3.4 to 3.1 My B.P., and the most recent in *D. benthamianus* ca. 29,000 to 7,700 y B.P. Our time estimates indicate North–South differentiation of the forest populations during the dry glacial climatic periods that took place from the Pliocene and especially during the Pleistocene (8). Altogether, our data are compatible with the differentiation of the genetic lineages in Northern and Southern Lower Guinea as a result of forest fragmentation during the dry glacial periods and subsequent admixture as a result of forest expansion during the humid

Table 2. Decline of genetic diversity with distance from refugia in Central Africa (biogeographic region of Lower Guinea) for *P. elata*, *D. benthamianus*, *E. ivorense*, *E. suaveolens*, and *S. zenkeri*

	LGM-Maley					LGM-Anhuf					LGM-spp.				
	Value	SE	DF	*t* value	*P* value	Value	SE	DF	*t* value	*P* value	Value	SE	DF	*t* value	*P* value
P. elata	**−3.80E−07**	1.50E−07	25	−2.44	**0.02**	4.70E−07	2.10E−07	25	2.27	0.03	−5.50E−07	3.40E−07	25	−1.61	0.12
D. benthamianus	**−4.40E−07**	1.50E−07	18	−2.90	**0.01**	**−5.90E−07**	1.20E−07	18	−4.79	**0.00**	**−5.90E−07**	2.70E−07	18	−2.21	**0.04**
E. ivorense	−5.80E−07	3.38E−07	8	−1.72	0.12	−6.92E−06	1.01E−05	8	−0.68	0.51	−5.95E−06	1.84E−05	8	−0.32	0.75
E. suaveolens	7.00E−08	7.40E−08	9	0.89	0.40	2.00E−08	4.70E−08	9	0.49	0.63	−2.00E−08	2.28E−07	9	−0.07	0.95
S. zenkeri	4.00E−08	3.80E−08	24	0.97	0.34	−1.00E−08	6.40E−08	24	−0.22	0.83	1.50E−07	8.20E−08	24	1.82	0.08

Linear regression of genetic diversity—observed heterozygosity, H_o—of each genotyped individual tree against the geographical distance from the closest LGM refugium. Negative, significant correlations are in bold. Three different hypotheses of LGM forest refugia were considered. The forest refugia postulated by Maley and Anhuf based on palynological and paleoclimatic data (Fig. 2) were tested: 1) LGM-Maley (16) and 2) LGM-Anhuf (41). In addition, specific forest refugia based on the LGM niche models of each species were tested (Fig. 2 and *SI Appendix*, Fig. S4): 3) LGM-spp.

interglacial periods. Similar windows of divergence times were reported in Annonaceae (14, 15, 34) and palms (14), but the phylogenetic methods used in these studies inferred divergence without subsequent gene flow. Several authors have pointed at a putative phenological barrier between the Northern and Southern tree populations that might contribute to the genetic differentiation (6, 7, 37). This would be driven by a sharp gradient in summer rainfall, known as Climatic Hinge. The boreal summer to the North is characterized by one dry and sunny season from June to August, while the austral summer to the South is characterized by one dry but cloudy season from December to January (38, 39). However, empirical data do not seem to show an inversion in flowering North and South of the Equator, including *E. suaveolens* (34, 40).

Decline of Genomic Diversity from Putative Glacial Refugia. We traced the spatial signal of recolonization during the humid periods by examining genetic diversity gradients over space. We found significant declines of genetic diversity from coastal refugia in *D. benthamianus* and from inland Congo refugia in *P. elata*. This finding suggests the survival of *D. benthamianus* in coastal refugia in Cameroon and Gabon, although the exact location cannot be determined since no differences were revealed among the different hypotheses of LGM forest refugia considered based on the paleoclimatic reconstructions (16, 41) or estimated from the potential climatic distribution of each species during the LGM. In the case of *P. elata* survival in inland forest refugia postulated by Maley in the Congo Basin is highly probable. For *S. zenkeri* and *E. suaveolens*, no significant declines in genetic diversity were found.

Genomic Diversity, Differentiation, and Dispersal Capacities. Our estimates of admixture and split dates are consistent with prior knowledge of the biology of the study species, in particular, with two traits that may be key for colonizing new areas after forest fragmentation: light tolerance and dispersal capacity. Long-distance dispersal of the seeds is especially relevant in the case of early successional communities and expanding populations, as it not only transports seeds over very long distances but also generates establishment opportunities. δαδι detected more recent splits between the North and South in *D. benthamianus* than for *Erythrophleum*. The oldest split was detected in the shade-tolerant, nonassisted dispersed species *S. zenkeri*. The fact that we detected more recent signals of North–South fragmentation in the species with long-distance dispersal capacity suggests that the old signals of fragmentation may be more easily erased in these species than in species with limited seed and pollen dispersal like *S. zenkeri*. The ability of long-distance dispersal may have made a difference over subsequent cycles of forest fragmentation and recolonization that took place during the Pleistocene. Our results also show a decline of genetic diversity from forest refugia only in species with long-distance dispersal. While *D. benthamianus* likely survived in coastal refugia, *P. elata* probably did so in inland refugia in the Congo Basin. The fact that we were able to trace significant declines of genetic diversity outside refugia in species with long-distance dispersal suggests that we may be detecting the signal of recent dispersal events for those species.

Conclusions

GBS data of five legume tree species widespread in African rainforests reveal 1) early divergence of the West African populations (Upper Guinea) from Central Africa (Lower Guinea) and 2) a clear North–South differentiation in Lower Guinea despite the absence of discontinuities in the rainforest cover or other geographic barriers, such as rivers and mountain chains. However, divergence times vary widely among species, from the Pliocene for shade-tolerant trees with nonassisted seed dispersal to late Pleistocene or Holocene for pioneer trees with long-distance wind-dispersed seeds. We conclude that different responses of tree species to recurrent forest fragmentation cycles driven by past climate fluctuations may explain why we observe congruent genetic spatial structures with contrasted timescales. Species with higher colonizing abilities seem to have been able to erase old signals of genetic differentiation compared with species with limited seed and pollen dispersal and light tolerance.

Materials and Methods

Study Species: Dispersal Capacity, Light Tolerance, Reproductive System. *P. elata* is a light-demanding, pioneer or nonpioneer tree with wind-dispersed seeds generally occurring in aggregates. Seeds disperse on average 314 m with a very fat tail. Most seeds disperse less than 100 m, but a significant amount of seeds experience long-distance dispersal >1 km (17). *D. benthamianus* is a pioneer light-demanding with wind-dispersed seeds (42). Individual trees are not aggregated in the field. It is an indicator of disturbed rainforests. Although most of the *D. benthamianus* seeds disperse over short distances (~70 m), the distribution of dispersal distances is very fat tailed, indicative of long-distance dispersal events. In addition, for a high proportion of seeds (30%), parents could not be detected within an exhaustively sampled plot larger than 3 km^2 (18). The two *Erythrophleum* spp. are light-demanding and nonpioneers (42). The fruits exhibit primary ballistic dispersal, and secondary dispersal by primates, *Cephalophus* spp., and rodents has been reported (43). The distribution is nongregarious, and they are indicators of secondary rainforests. Based on genetic markers, seed-dispersal distances of 294 m were detected in *E. suaveolens* (18), but long-distance dispersal events seem rare (18, 44). *S. zenkeri* is a shade-tolerant species that exhibits ballistic seed dispersal through the explosion of the pods. Genetic markers confirmed not only the limited dispersal suggested by the fruit morphology but also limited pollen dispersal (19). Trees are locally aggregated in the field. It requires high environmental humidity (45) and is an indicator of undisturbed rainforests.

DNA Extraction and Genotyping by Sequencing. Leaf and cambium materials were collected in the rainforests of West and Central Africa between 2005 and 2014 (*SI Appendix*, Table S4). The samples were immediately dried with silica gel in order to preserve the DNA quality. Between 18 and 46 samples of each species were selected in order to represent their distribution in Central Africa, with special emphasis in the biogeographic region of Lower Guinea. A few samples from the less accessible rainforest of the Congo Basin were included. Outgroup taxa (*SI Appendix*, Table S4) were selected based on available legume phylogenies (46, 47).

Overall, 362 GBS libraries from 182 individuals were initially sequenced on four Illumina lanes (HiSeq2000), using 100 base pair (bp) single-read chemistry. For each library, two DNA extractions were performed using the DNeasy Plant Minikit columns (Qiagen) and pooled in order to generate sufficient DNA for the GBS protocol. One blank per plate was included. DNA quality was checked on a 1.5% agarose gel, and DNA quantity was measured with Qbit HS (Life technologies). The DNA was purified with a ZR-96 DNA Clean up kit (Zymo Research Corp.). Subsequently, GBS was performed at the Genomic Diversity and Computational Biology Service Unit at Cornell University (Ithaca, NY) according to a published protocol (48). One microgram of DNA of each species was initially used in order to optimize the GBS protocol, in particular aiding the choice of the most appropriate restriction enzyme. Specifically, three libraries were built for each species using three different enzymes: ApeKI (4.5 base cutter), EcoT22I and PstI (both 6 base cutters) and checked for appropriate fragment sizes (<500 bp) and distribution on an Experion automatic electrophoresis system (Bio-Rad laboratories). Given these results, we elected to use the enzyme EcoT22I for subsequent data generation, as it yielded appropriate fragment sizes (<500 bp) and distributions for all study species.

Genotype Calling and Site Frequency Spectrum. We limited the impact of uneven coverage of samples typical for GBS data by building and sequencing two independent libraries for each individual. Two complementary bioinformatic pipelines were implemented for genotype calling and estimating summary statistics used for downstream analyses.

Genotype Calls without Outgroups. For those analyses that require SNP and genotype calls, we used the Universal Network-Enabled Analysis Kit (UNEAK) pipeline (49) within the software TASSEL 3.0 (50), suitable for analysis of GBS data at the intraspecific level in the absence of a reference genome. Reads were trimmed to 64 bp (to avoid sequencing errors at the ends of reads) and

identical reads collapsed into tags (i.e., alleles). Tag pairs having a single bp mismatch were identified as candidate loci and used for SNP calling. Tag pairs forming complicated networks, likely to result from repeats, paralogs, and sequencing error, were filtered out. In order to filter out false-positive SNPs, 1) the minimum number of reads per GBS tag (i.e., allele) was set to five, 2) an error tolerance rate of 0.05 was established, and 3) SNPs with a genotype missing rate >50% were removed.

Genotype Likelihood Framework for Analyses with Outgroups. Building a "reference" catalog: As the first step in estimating genotype likelihoods at variable sites across the sequenced loci, we constructed a reference catalog against which we could align the GBS reads. We used the tags built with TASSEL to obtain 64 bp long reads. Subsequently, the query read sequences from TASSEL were concatenated with a spacer of 200 ns to obtain the reference.

Genotype likelihood computation: For many of the downstream analyses, a genotype calling approach from next-generation sequencing (NGS) data may bias the population genetic estimates (51). Therefore, we computed the genotype likelihoods at all the variable sites using the software Analysis of Next-Generation Sequencing Data, ANGSD version 0.914 (52). First, we mapped the GBS reads from each sample to the constructed reference catalog using the PALEOMIX pipeline (53). As part of this pipeline, AdapterRemoval version 2 (54) was used to trim adapter sequences, merge the paired reads, and discard reads shorter than 30 bp. BWA version 0.7.15 was then used to map the processed reads to the reference (55). Minimum mapping quality was set at 15, while minimum base quality was set to 5. Subsequently, genotype likelihoods were calculated in ANGSD using the following filters: 1) the baq option was used to recompute base alignment qualities, allowing us to reduce false SNPs due to misalignment; 2) a Hardy–Weinberg Equilibrium $P > 0.001$; and 3) the total coverage at any site cannot exceed 55 times the number of samples. The depth distribution was plotted for each species, and a cutoff was chosen in order to exclude outlier sites at extremely high coverage. Given the diversity in our reads, the –C filter, downgrading mapping quality for reads containing excessive mismatches, was not used.

Site Frequency Spectrum: Multipopulation Site Frequency Spectrum (SFS) was estimated from the genotype likelihoods using the utility program realSFS provided with ANGSD. Three different quality cut-offs were used to compute three different SFS for each data set: (filtering 1) minimum mapping and base quality were both set to 10, and only sites where at least 50% of the samples were covered by one or more reads were retained; (filtering 2) minimum mapping and base quality were both set to 15, and only sites where at least 65% of the samples were covered by one or more reads were retained; (filtering 3) minimum mapping and base quality were both set to 30, and only sites where at least 75% of the samples were covered by one or more reads were retained.

Inference of Genetic Clusters. The software Admixture version 13 was used to calculate the probability of assignment of individuals to genetic clusters (56) from the genotype calls data sets without outgroups. Individuals exhibiting admixture between clusters were detected. For each species, K = 1 to 12 genetic clusters were tested with 20 randomly seeded replicate runs at each K and fivefold CV. Bar plots showing the probability of assignment of individuals to the genetic clusters were generated in R using the package ggplot2.

The value of K that best fits the data were estimated using the lowest fivefold CV. Since these values did not always work appropriately due to low sample sizes in some of the clusters (57), we used 1) the congruence of the genetic clusters with the geography and 2) the congruence of the genetic clusters with the phylogenies as additional criteria to select the optimal K.

Phylogenetic Relationships. Maximum likelihood phylogenies were built for each species using RAxML 8.0 (58) from the Genotype Calls data sets with outgroups. Rooting with outgroups allowed us to assess the direction of evolution. All SNPs were concatenated into a single alignment, with missing data and heterozygous positions entered as needed. Bootstrap support was calculated from 100 replicate searches with random starting trees using the GTR+ gamma nucleotide substitution model. Analyses with ascertainment bias correction using the Lewis method to avoid biases in the phylogeny when the number of nonvariable sites is not known revealed the same topology and support of the main phylogenetic clades (*SI Appendix*, Fig. S1).

Demographic Inference with δaδi. Population demographic models were estimated using δaδi (59), a method that uses the SFS of populations to infer their demographic history. Based on the admixture and phylogeny results, for the species widespread in Upper and Lower Guinea—*D. bentamianus*, *E. suaveolens*—we fitted a three-population model with the tree topology (Upper Guinea [UG], (Northern Lower Guinea [LG-North], Southern Lower Guinea [LG-South])) (*SI Appendix*, Fig. S2*A*). *E. ivorense* was excluded from this analysis due to poor sampling. In addition to a model with symmetric migration among the three clusters, we fitted a model with no migration allowed between UG and LG-South after the split. For *S. zenkeri*, absent in Upper Guinea, we estimated split times based on the tree topology (LG-North, (LG-South1, LG-South2)). The observed SFS was compared to the expected SFS under the Isolation with Migration (IM) model with symmetric gene flow (*SI Appendix*, Fig. S2 *B–D*). Using maximum likelihood, split time parameter values were estimated in generations. We assumed a mutation rate of $\mu = 2.5 \times 10 - 9$ ($1.7 \times 10 - 9 - 3.5 \times 10 - 9$) per site per year, estimated for *Populus* (60, 61) and a generation time of 100 y (12, 62). To avoid biasing the demographic inferences due to uneven depth of coverage, which is typical of GBS data, we excluded all singletons (i.e., alleles found only once among the populations), while estimating the demographic parameters. Furthermore, while estimating SFS, admixed individuals with less than 70% genetic ancestry to a single group were excluded.

For each such demographic model, we ran 100 replicates and selected the parameter estimates from the best fitting model (i.e., the model with the highest log-likelihood), with one important exception. Model fits in δaδi, which masked more than 5% of the SFS, were excluded. Finally, we performed 1,000 replicates for *D. benthamianus*, as this species yielded fewer SNPs and larger sample sizes, in order to get a stable estimate of the demography.

Gradients of Genetic Diversity over Space and Climatic Niche Modeling. In order to visualize the patterns of genetic diversity over space in Lower Guinea for each species, we calculated the genetic diversity of each of the genotyped individuals and plotted it against the geographical distance to the edge of the closest LGM refugium. A mixed effect model was performed using genetic diversity as the independent variable, the distance to LGM refugia as explanatory variable (minimal distance between the sample and the limit of a postulated refugia), and the genetic cluster as a random variable in order to account for genetic diversity differences across genetic lineages.

We calculated the observed heterozygosity (H_o) of each individual as a proxy for genetic diversity using GenAlEx (63). Admixed individuals with less than 70% genetic ancestry to a single group were excluded. Three different hypotheses of LGM forest refugia were considered. The classic forest refugia postulated by Maley (16) and Anhuf (41), based on species distributions as well as on paleoclimatic and palynological data (Fig. 2): 1) LGM-Maley and 2) LGM-Anhuf. In addition, specific forest refugia based on the LGM niche models of each species were built and tested (Fig. 2 and *SI Appendix*, Fig. S4): 3) LGM-species niche models (see *Species Distribution Models for the LGM*). For LGM-Maley and LGM-Anhuf, the shape files were digitized from the publications figures. For the LGM species, the polygon defined by the boundary of the ensemble model was used.

For *P. elata* and *E. suaveolens*, which exhibited inland populations that are equally likely to have survived either in the coastal refugia of Cameroon and Gabon or in the inland refugia in Congo, two additional mixed effects models were run: 4) LGM-Maley-Coast, including the coastal Maley refugia only and v5 LGM-Maley-Congo, including the Congo refugia only.

Species Distribution Models for the LGM. We collected occurrence records for each species from RAINBIO database (64) and our own records. We visually inspected records to eliminate outliers and deduplicated records per species to obtain one per pixel according to the climate layers resolution. After this, we had 1,029 occurrences distributed among species as follows: *D. benthamianus* ($n = 427$), *E. ivorense* ($n = 74$), *E. suaveolens* ($n = 180$), *P. elata* ($n = 118$), *S. zenkeri* ($n = 230$) (*SI Appendix*, Fig. S4).

We used the 19 bioclim layers from the Worldclim data set version 1.4 at 2.5 arc-min (65) representing the climate between 1960 and 1990 and calculated the total precipitation of the Austral Summer (December, January, and February) using the monthly layers. After performing correlation analysis and principal component analysis (PCA), we finally selected 5 variables that were minimally correlated to run the models: annual mean temperature (bio1), temperature annual range (bio7), total annual precipitation (bio12), precipitation of the driest month (bio14), and precipitation of December, January, and February (pp_djf). We also used the selected variables from simulations with the CCSM4 global climate model for the LGM from Worldclim to project species distributions for that time.

We modeled the distribution of the five species with the biomod2 package in R (66). We randomly selected 5,000 pseudoabsences. We used three different algorithms (generalized linear models, GLM; random forest, RF;

Maxent) and five repetitions, obtaining a total of 15 models for each species. Models were calibrated using 80% of the occurrences and evaluated using the remaining 20% of occurrences and the TSS and ROC statistics (67). We used a consensus approach to produce an ensemble model using the weighted mean of all models that had at least TSS values above 0.7 and ROC values above 0.8. Binary maps (presence/absence maps) were produced using the TSS threshold (*SI Appendix*, Fig. S4).

Data Availability. Genotyping by sequencing data generated for this paper were deposited in Dryad database https://doi.org/10.5061/dryad.tb2rbnzz1 (68). They include vcf files without outgroups and fasta files with outgroups of five Legume tree species widespread in the rainforests of West and Central Africa (*P. elata*, *D. benthamianus*, *E. ivorense*, *E. suaveolens* and *S. zenkeri*) and seven outgroup species (*Pericopsis laxiflora*, *Dicorynia guianensis*, *Dialium guianense*, *Erythrophleum africanum*, *Erythrophleum chlorostachys*, *Calpocalyx brevibracteatus*, and *Hymenostegia afzelii*.

ACKNOWLEDGMENTS. We thank Boris Demenou, Armel Donkpegan, Myriam Heuertz, Théophile Ayolle, Charlotte Hansen, Esra Kaymak, Jean-Louis Doucet, Jerôme Duminil, Jerôme Chave, and Hermann Daniel. This work received financial support from the Marie Curie FP7-PEOPLE-2012-IEF program (Project AGORA) awarded to R.P.

1. U. Salzmann, P. Hoelzmann, The Dahomey Gap: An abrupt climatically induced rain forest fragmentation in West Africa during the late Holocene. *Holocene* **15**, 190–199 (2005).
2. C. S. Miller, W. D. Gosling, Quaternary forest associations in lowland tropical West Africa. *Quat. Sci. Rev.* **84**, 7–25 (2014).
3. J. Hey, The divergence of chimpanzee species and subspecies as revealed in multipopulation isolation-with-migration analyses. *Mol. Biol. Evol.* **27**, 921–933 (2010).
4. P. T. Telfer *et al*., Molecular evidence for deep phylogenetic divergence in *Mandrillus sphinx*. *Mol. Ecol.* **12**, 2019–2024 (2003).
5. N. M. Anthony *et al*., The role of Pleistocene refugia and rivers in shaping gorilla genetic diversity in central Africa. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 20432–20436 (2007).
6. O. J. Hardy *et al*., 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 (2013).
7. M. Heuertz, J. Duminil, G. Dauby, V. Savolainen, O. J. Hardy, Comparative phylogeography in rainforest trees from Lower Guinea, Africa. *PLoS One* **9**, e84307 (2014).
8. P. Gibbard, J. Ehlers, P. Hughes, "Quaternary glaciations" in *International Encyclopedia of Geography: People, the Earth, Environment and Technology: People, the Earth*, D. Richardson et al., Eds. (Environmental Technology, 2016), pp. 1–10.
9. V. Plana, Mechanisms and tempo of evolution in the African Guineo-Congolian rainforest. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **359**, 1585–1594 (2004).
10. J. Duminil *et al*., 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 (2015).
11. A. Faye *et al*., Phylogenetics and diversification history of African rattans (Calamoideae, Ancistrophyllinae). *Bot. J. Linn. Soc.* **182**, 256–271 (2016).
12. R. Piñeiro, G. Dauby, E. Kaymak, O. J. Hardy, Pleistocene population expansions of shade-tolerant trees indicate fragmentation of the African rainforest during the Ice Ages. *Proc. Biol. Sci.* **284**, 20171800 (2017).
13. B. B. Demenou *et al*., Plastome phylogeography in two African rain forest legume trees reveals that Dahomey Gap populations originate from the Cameroon volcanic line. *Mol. Phylogenet. Evol.* **150**, 106854 (2020).
14. A. J. Helmstetter, K. Béthune, N. G. Kamdem, B. Sonké, T. L. P. Couvreur, Individualistic evolutionary responses of Central African rain forest plants to Pleistocene climatic fluctuations. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 32509–32518 (2020).
15. J. Migliore *et al*., Pre-Pleistocene origin of phylogeographical breaks in African rain forest trees: New insights from *Greenwayodendron* (Annonaceae) phylogenomics. *J. Biogeogr.* **46**, 212–223 (2019).
16. J. Maley, The African rain forest–main characteristics of changes in vegetation and climate from the Upper Cretaceous to the Quaternary. *Proc. R. Soc. Edinb. Biol. Sci.* **104**, 31–73 (1996).
17. D.-M. A. Angbonda, F. K. Monthe, N. Bourland, F. Boyemba, O. J. Hardy, Seed and pollen dispersal and fine-scale spatial genetic structure of a threatened tree species: *Pericopsis elata* (HARMS) Meeuwen (Fabaceae). *Tree Genet. Genomes* **17**, 27 (2021).
18. O. J. Hardy *et al*., Seed and pollen dispersal distances in two African legume timber trees and their reproductive potential under selective logging. *Mol. Ecol.* **28**, 3119–3134 (2019).
19. S. Vanden Abeele, S. B. Janssens, R. Piñeiro, O. J. Hardy, 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 Ecol. Evol.* **21**, 50 (2021).
20. F. White, *The Vegetation of Africa: A Descriptive Memoir to Accompany the UNESCO/AETFAT/UNSO Vegetation Map of Africa* (Natural Resources Research, UNESCO, Paris, France, 1983).
21. V. Droissart *et al*., Beyond trees: Biogeographical regionalization of tropical Africa. *J. Biogeogr.* **45**, 1153–1167 (2018).
22. A. Fayolle *et al*., Patterns of tree species composition across tropical African forests. *J. Biogeogr.* **41**, 2320–2331 (2014).
23. H. P. Linder *et al*., The partitioning of Africa: Statistically defined biogeographical regions in sub-Saharan Africa. *J. Biogeogr.* **39**, 1189–1205 (2012).
24. B. B. Demenou, R. Piñeiro, O. J. Hardy, 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 (2016).
25. P. Gagneux, M. K. Gonder, T. L. Goldberg, P. A. Morin, Gene flow in wild chimpanzee populations: What genetic data tell us about chimpanzee movement over space and time. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **356**, 889–897 (2001).
26. M. K. Gonder *et al*., A new west African chimpanzee subspecies? *Nature* **388**, 337 (1997).
27. J. Fuchs, R. C. Bowie, Concordant genetic structure in two species of woodpecker distributed across the primary West African biogeographic barriers. *Mol. Phylogenet. Evol.* **88**, 64–74 (2015).
28. A. D. Leaché, M. K. Fujita, Bayesian species delimitation in West African forest geckos (*Hemidactylus fasciatus*). *Proc. Biol. Sci.* **277**, 3071–3077 (2010).
29. P. Beresford, J. Cracraft, Speciation in African forest robins (*Stiphronis*): Species limits, phylogentic relationships, and molecular biology. *Am. Mus. Novit.* **3270**, 1–22 (1999).
30. B. K. Schmidt, J. T. Foster, G. R. Angehr, K. L. Durrant, R. C. Fleischer, A new species of African forest robin from Gabon (Passeriformes: Muscicapidae: *Stiphrornis*). *Zootaxa* **1850**, 27–42 (2008).
31. B. D. Marks, Are lowland rainforests really evolutionary museums? Phylogeography of the green hylia (*Hylia prasina*) in the Afrotropics. *Mol. Phylogenet. Evol.* **55**, 178–184 (2010).
32. V. Nicolas *et al*., The roles of rivers and Pleistocene refugia in shaping genetic diversity in *Praomys misonnei* in tropical Africa. *J. Biogeogr.* **38**, 191–207 (2011).
33. A. D. Leaché, J. R. Oaks, C. Ofori-Boateng, M. K. Fujita, Comparative phylogeography of West African amphibians and reptiles. *Evolution* **74**, 716–724 (2020).
34. A. J. Helmstetter *et al*., Phylogenomic approaches reveal how climate shapes patterns of genetic diversity in an African rain forest tree species. *Mol. Ecol.* **29**, 3560–3573 (2020).
35. R. C. Bell *et al*., Idiosyncratic responses to climate-driven forest fragmentation and marine incursions in reed frogs from Central Africa and the Gulf of Guinea Islands. *Mol. Ecol.* **26**, 5223–5244 (2017).
36. D. M. Portik *et al*., Evaluating mechanisms of diversification in a Guineo-Congolian tropical forest frog using demographic model selection. *Mol. Ecol.* **26**, 5245–5263 (2017).
37. C. Gonmadje, C. Doumenge, T. Sunderland, D. McKey, Environmental filtering determines patterns of tree species composition in small mountains of Atlantic Central African forests. *Acta Oecol.* **94**, 12–21 (2018).
38. M. Leroux, *Climate of Tropical Africa* (Champion, 1983).
39. J. P. Vande Weghe, *Forests of Central Africa: Nature and Man* (Protea Book House, Pretoria, 2004).
40. D.-Y. Ouédraogo *et al*., Latitudinal shift in the timing of flowering of tree species across tropical Africa: Insights from field observations and herbarium collections. *J. Trop. Ecol.* **36**, 159–173 (2020).
41. D. Anhuf *et al*., Paleo-environmental change in Amazonian and African rainforest during the LGM. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **239**, 510–527 (2006).
42. Q. Meunier, C. Moumbogou, J.-L. Doucet, *Les Arbres utiles du Gabon* (Presses agronomiques de Gembloux, 2015).
43. A.-P. Gorel, A. Fayolle, J.-L. Doucet, Ecology and management of the multipurpose *Erythrophleum* species (Fabaceae-Caesalpinioideae) in Africa. A review. *Biotechnol. Agron. Soc. Environ.* **19**, 415–429 (2015).
44. J. Duminil *et al*., CpDNA-based species identification and phylogeography: Application to African tropical tree species. *Mol. Ecol.* **19**, 5469–5483 (2010).
45. G. Abanda, *La surexploitation du Scorodophloeus zenkeri Harms (Arbre à ail): Atténuer l'impact de la gestion non durable de l'arbre à ail dans le Massif Forestier de Ngovayang du Sud Cameroun* (Editions Universitaires Europeennes, 2011).
46. A. Bruneau, M. Mercure, G. P. Lewis, P. S. Herendeen, Phylogenetic patterns and diversification in the caesalpinioid legumes. *Botany* **86**, 697–718 (2008).
47. N. Azani *et al*., A new subfamily classification of the Leguminosae based on a taxonomically comprehensive phylogeny the Legume Phylogeny Working Group (LPWG). *Taxon* **66**, 44–77 (2017).
48. R. J. Elshire *et al*., A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* **6**, e19379 (2011).
49. F. Lu *et al*., Switchgrass genomic diversity, ploidy, and evolution: Novel insights from a network-based SNP discovery protocol. *PLoS Genet.* **9**, e1003215 (2013).
50. P. J. Bradbury *et al*., TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633–2635 (2007).
51. R. Nielsen, J. S. Paul, A. Albrechtsen, Y. S. Song, Genotype and SNP calling from next-generation sequencing data. *Nat. Rev. Genet.* **12**, 443–451 (2011).
52. T. S. Korneliussen, A. Albrechtsen, R. Nielsen, ANGSD: Analysis of next generation sequencing data. *BMC Bioinformatics* **15**, 356 (2014).
53. M. Schubert *et al*., Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nat. Protoc.* **9**, 1056–1082 (2014).
54. S. Lindgreen, AdapterRemoval: Easy cleaning of next-generation sequencing reads. *BMC Res. Notes* **5**, 337 (2012).
55. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

56. D. H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
57. J. Wang, The computer program structure for assigning individuals to populations: Easy to use but easier to misuse. *Mol. Ecol. Resour.* **17**, 981–990 (2017).
58. A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
59. R. N. Gutenkunst, R. D. Hernandez, S. H. Williamson, C. D. Bustamante, Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet.* **5**, e1000695 (2009).
60. G. A. Tuskan *et al*., The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **313**, 1596–1604 (2006).
61. P. K. Ingvarsson, Multilocus patterns of nucleotide polymorphism and the demographic history of *Populus tremula*. *Genetics* **180**, 329–340 (2008).
62. T. R. Baker *et al*., Fast demographic traits promote high diversification rates of Amazonian trees. *Ecol. Lett.* **17**, 527–536 (2014).
63. R. Peakall, P. E. Smouse, GENALEX 6: Genetic analysis in excel. Population genetic software for teaching and research. *Mol. Ecol. Notes* **6**, 288–295 (2006).
64. G. Dauby *et al*., RAINBIO: A mega-database of tropical African vascular plants distributions. *PhytoKeys* **74**, 1–18 (2016).
65. R. J. Hijmans, S. E. Cameron, J. L. Parra, P. G. Jones, A. Jarvis, Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **25**, 1965–1978 (2005).
66. W. Thuiller, B. Lafourcade, R. Engler, M. B. Araújo, BIOMOD–a platform for ensemble forecasting of species distributions. *Ecography* **32**, 369–373 (2009).
67. K. H. Zou, A. Liu, A. I. Bandos, L. Ohno-Machado, H. E. Rockette, *Statistical Evaluation of Diagnostic Performance: Topics in ROC Analysis* (CRC Press, 2011).
68. R. Piñeiro *et al*., Genotyping by sequencing data of five Legume tree species widespread in the rainforests of West and Central Africa. *Dryad*, Dataset. https://doi.org/10.5061/dryad.tb2rbnzz1. Deposited 31 July 2020.
69. P. Mayaux, E. Bartholomé, S. Fritz, A. Belward, A new land-cover map of Africa for the year 2000. *J. Biogeogr.* **31**, 861–877 (2004).