

1

2

Classification: Biological Science, Evolution

3

Title: Contrasting dates of rainforest fragmentation in Africa inferred from trees with  
different dispersal abilities

4

5 Rosalía Piñeiro<sup>a,b,c,\*</sup>, Olivier J. Hardy<sup>b</sup>, Carolina Tovar<sup>c</sup>, Shyam Gopalakrishnan<sup>a</sup>, Filipe Garrett  
6 Vieira<sup>a</sup>, and M Thomas P Gilbert<sup>a,d</sup>

7

8 <sup>a</sup> Evolutionary Genomics, Natural History Museum of Denmark, University of Copenhagen,  
9 Copenhagen, Denmark, Øster Voldgade 5-7, 1350 København K

10

11 <sup>b</sup> Unit of Evolutionary Biology & Ecology, Faculté des Sciences, Université Libre de Bruxelles,  
12 Brussels, Belgium, Av. F.D. Roosevelt, 50, CP 160/12, B-1050 Brussels, Belgium

13

14 <sup>c</sup> Biodiversity Informatics & Spatial Analysis, Royal Botanic Gardens, Kew, Richmond,  
15 Surrey TW9 3AB

16 <sup>d</sup> University Museum, Norwegian University of Science and Technology, N-7491 Trondheim,  
17 Norway

18

19

20 Corresponding author: Rosalía Piñeiro, Geography, Laver Building, University of Exeter,  
21 North Park Road, Exeter, EX4 4QE, UK, +441392 725297, rosalia.pineiro@gmail.com

22

23 Keywords: rainforests, Tropical Africa, Genotyping by Sequencing, Gradients of Genetics  
24 Diversity, Glacial Refugia

25 ABSTRACT

26 The rainforests of Tropical Africa have fluctuated over time. Although today the forest cover  
27 is continuous in Central Africa this may have not always been the case, as the scarce fossil  
28 record in this region suggests that more arid conditions might have significantly reduced the  
29 density of trees during the Ice Ages. Our aim was to investigate whether the dry ice-age  
30 periods left a genetic signature on tree species that can be used to date the past  
31 fragmentation of the rainforest. We sequenced reduced representation libraries of 182  
32 samples representing five Legume tree species that are widespread in African rainforests and  
33 seven outgroups. Phylogenetic analyses identified an early divergent lineage for all species in  
34 West Africa (Upper Guinea), and two clades in Central Africa: Lower Guinea-North and  
35 Lower Guinea-South. As the structure separating the Northern and Southern clades cannot  
36 be explained by geographic barriers, we tested other hypotheses using demographic model  
37 testing. The best estimates recovered using  $\partial a\partial I$  indicate that the two clades split between  
38 the Upper Pliocene and the Pleistocene, a date compatible with forest fragmentation driven  
39 by ice-age climatic oscillations. Furthermore, we found remarkably older split dates for the  
40 shade-tolerant tree species with non-assisted seed dispersal than for light-demanding long-  
41 distance wind-dispersed trees. We also show that the genetic diversity significantly declines  
42 with the distance from ice-age refugia in the two long-distance dispersed species only.  
43 Different recolonisation abilities after recurrent cycles of forest fragmentation seem to  
44 explain why we observe congruent genetic spatial structures across species with contrasted  
45 timescales.

46  
47  
48 SIGNIFICANCE STATEMENT

49 *Although today the rainforest cover is continuous in Central Africa, the scarce fossil record*  
50 *suggests that arid conditions during the Ice Ages might have reduced the density of trees*  
51 *during the Ice Ages. However, the vast majority of the fossil pollen records preserved in*  
52 *Tropical Africa is too young to inform about this period. Investigating whether the past*  
53 *climate change left a genetic signature on trees can thus be useful to date past forest*  
54 *fragmentation. However, most genetic studies available to date lack resolution as they use*  
55 *limited numbers of loci. In this study we use modern DNA technology to study five Legume*  
56 *trees. Our results show significant differentiation of the populations of each species at a*  
57 *date compatible with forest fragmentation driven by ice-age climatic oscillations.*  
58 *Contrasted timescales were obtained for each species, which probably reflects their*  
59 *different recolonisation abilities after forest fragmentation.*

60  
61 INTRODUCTION

62 The rainforest cover in Tropical Africa has fluctuated widely over time. Today the rainforests  
63 of West Africa (Upper Guinea) are disconnected from Central Africa (Lower Guinea) by the  
64 Dahomey Gap, a forest-savannah corridor along the coast of Benin, Togo and eastern Ghana  
65 (Figure 1). However, the fossil record shows that this region was forested under the humid  
66 conditions of the last interglacial (from ca. 8,400 years BP) while the current deforestation  
67 started only 4,500 years ago following the aridification of the climate (1). It is not clear  
68 whether the Dahomey gap also became forested during the previous interglacials, as they do  
69 not seem to have been as humid as the last one (2). While the rainforests of Central Africa  
70 (Lower Guinea) currently exhibit a continuous distribution, previous genetic studies indicate  
71 strong differentiation of tropical trees within the forest. Typically, such structure may be  
72 explained by geographic barriers. For instance the river Sanaga, one of the main rivers of  
73 Cameroon, runs from inland Cameroon towards the coast delimiting two different  
74 subspecies of chimpanzee (3) and also creating a deep genetic divergence of mandrill

75 populations on both sides (4). In Gabon, the Oougué river acts as an effective barrier for  
76 dispersal of mandrills (4) and gorillas(5). In contrast to primates the genetic structure of  
77 tropical trees cannot be explained by effective barriers to dispersal such as the main rivers  
78 and mountain chains in this area (6, 7).

79

80 Recent studies based on chloroplast DNA, nuclear microsatellites, and low-copy nuclear  
81 genes (6-8) suggest that the observed historical isolation of the tree populations in Tropical  
82 Africa was caused by forest fragmentation during the cold and dry Ice-Age periods, which  
83 occurred on several cycles, along the Pleistocene (9). However, dating the fragmentation of  
84 the rainforest and determining where the ancestral populations of each species was, has been  
85 challenging due to the low number of markers investigated. For example with regards to  
86 dating, three prior studies have attempted to estimate the divergence of tree populations in  
87 the Pleistocene, although with large uncertainty due to the low numbers of molecular  
88 markers used (10-12). A fourth study on the genus *Greenwayodendron* places the divergence  
89 estimates in the Pliocene/Pleistocene(13). As for the location of forest fragments that allowed  
90 tropical tree species to survive the Ice Ages, these have been postulated based on the fossil  
91 record and palaeoclimatic reconstructions. Areas that harbour high species richness have  
92 been proposed as refugia (Figure 2), assuming declines in the number of species outside the  
93 hypothesized refugia (14). Likewise, declines of genetic diversity with distance from refugia  
94 are expected to result from recolonisation after forest fragmentation. Despite the potential of  
95 genetic diversity gradients to help locate the areas in which forest species survived, the  
96 available genetic data lack resolution to assess diversity gradients over space.

97

98 To overcome these challenges, we generated reduced representation genomic data using  
99 Illumina sequencing technology, for five Legume rainforest tree species, in order to  
100 investigate the genetic signal of changes in the rainforest cover during the Ice Ages. All five  
101 species -*Pericopsis elata* (Harms) Meeuwen, *Distemonanthus benthamianus* Baill.  
102 *Erythrophleum ivorense* A. Chev., *Erythrophleum suaveolens* (Guill. & Perr.) Brenan, and  
103 *Scorodophloeus zenkeri* Harms- are widespread in the African rainforest (Figure 1) and  
104 exhibit differences in light tolerance and dispersal capacity (S1). Based on their ecology and  
105 dispersal biology, *P. elata* and *D. benthamianus* are the most adapted to long-distance  
106 colonization. They are both light-demanding and wind-dispersed, with significant long-  
107 distance dispersal events (15). The two *Erythrophleum* species are light-demanding and  
108 besides its primary ballistic dispersal they exhibit secondary dispersal of their seeds by  
109 animals, but no evidence of long-distance dispersal has been detected in direct  
110 measurements with molecular markers (15). *S. zenkeri*, a strict shade-tolerant species with  
111 seed dispersal non-assisted by wind or animals, exhibits the most limited colonising  
112 capacity.

113

114 In particular we aimed to solve the following questions:

- 115
- 116 • Are the tree populations in West African rainforests (Upper Guinea) well-  
117 differentiated from Central African rainforests (Lower-Guinea) or did the expansion  
118 of the rainforest during the humid Holocene favoured dispersal and geneflow  
between the two forest blocks?
  - 119 • Are the divergence times estimated by sequencing large portions of the genome  
120 compatible with isolation and fragmentation of the Central African rainforest during  
121 the dry and cold Ice Ages?

- 122 • Can we trace the genetic signal of recolonisation of tree populations from putative  
123 glacial refugia?  
124 • What are the similarities and differences in the patterns of genetic diversity and  
125 differentiation in the light of different ecologies and dispersal capacities of the five  
126 tree species?  
127

## 128 RESULTS

129 A total of 182 samples were successfully sequenced (175 samples of the five study species and  
130 7 outgroup taxa). Samples received an average number of reads between 3,387,236 in the *P.*  
131 *elata* and 5,518,179 in *D. benthamianus* (S2). Our protocol compensated the limitation of  
132 uneven coverage across loci typical of GBS studies by sequencing two libraries per individual.  
133

134 For the GBS genotype calls without outgroups done with TASSEL between 10,665 (*D.*  
135 *benthamianus*) and 27,838 SNPs (*E. suaveolens*) were present in at least 50% of individuals.  
136 For the genotype likelihood framework we retained between 568,240 reads in *P. elata* (16  
137 depth coverage) and 1,561,119 reads in *D. benthamianus* (32X coverage). The number of  
138 SNPs ranged from 17,921 in *D. benthamianus* to 70,295 in *S. zenkeri* for the less stringent  
139 quality cut-off filtering 1 (Table 1, S2).  
140

### 141 Genetic Ancestry and Phylogenetic reconstructions

#### 142 *Upper Guinea and Lower Guinea*

143 We conducted RAxML phylogenetic analyses for each species using the GBS genotype calls  
144 with outgroups (Figure 1, Right; S3). For the species present in West Africa (*P. elata*, *D.*  
145 *benthamianus*, *E. ivorensis*, and *E. suaveolens*) the Upper Guinean populations clustered in  
146 independent clades that were sister to the Lower Guinean clades. Similarly, ADMIXTURE  
147 barplots of probability of assignment of individuals to populations (Figure 1, Left) showed  
148 separate Upper Guinean (UG) genetic clusters when the number of ancestral populations (K)  
149 was set to K=3 (*E. ivorensis*, *E. suaveolens*, and *P. elata*) or K=4 (*D. benthamianus*). In the  
150 case of the species widespread in Lower Guinea (LG) -*D. benthamianus*, *E. ivorensis*, *E.*  
151 *suaveolens*, and *S. zenkeri*- two reciprocally monophyletic clades stand out for LG-North and  
152 LG-South (Figure 1, Right; S3). Similarly, in the ADMIXTURE analyses a split between the  
153 LG-North and the LG-South genetic clusters was observed (Figure 1) at K=3 for *E. ivorensis*  
154 and *E. suaveolens*, and at K=4 for *D. benthamianus* and *S. zenkeri*. The two species sampled  
155 in Congo (*P. elata*, and *S. zenkeri*) revealed independent genetic groups based on RAxML  
156 and ADMIXTURE analyses. However, the geographic coverage of our samples in this  
157 biogeographic region is incomplete.

158

#### 159 *D. benthamianus*

160 The ADMIXTURE analysis at K=2, where the minimum cross validation (CV) was found,  
161 revealed an UG and a LG cluster, with genetically intermediate samples in the Dahomey Gap.  
162 At K=3 the DG is retrieved as an independent cluster, with admixture from the UG region in  
163 the West. At K=4 in addition to the UG, and DG clusters, LG splits into two clusters LG-  
164 North and LG-South, with ancestry shared between the two groups in the geographically  
165 intermediate areas. From K=5 the samples with shared ancestry between LG-North and LG-  
166 South form an independent cluster. From K=>6 further genetic subgroups are found within  
167 the UG and LG-North, with no geographical congruence. The rooted ML phylogenetic tree  
168 without admixed individuals agrees with K=4. It consists of several basal clades for UG, two

169 reciprocally monophyletic well-supported clades in LG–North and LG–South, and an  
170 intermediate clade for the DG. The admixed individuals fell in between the main clades.

171 *E. ivorensis*

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

181 *E. suaveolens*

182 The ADMIXTURE analyses split an UG and a LG clusters at K=2, where the  
183 minimum CV is reached. At K=3 the genetic clusters retrieved were UG, LG-North,  
184 and LG-South. From K=4 and higher, subgroups appear randomly within UG, and  
185 from K=6 and higher within LG-North. The RAxML phylogenetic tree retrieves three  
186 clades corresponding to UG, LG-North and LG-South. No individuals with shared  
187 ancestry between groups were detected. In this species the UG cluster does not  
188 correspond to the West African Guineo-Congolian forests but to gallery forests  
189 embedded in West African savannahs, including the forest-savanna mosaic of  
190 Cameroon, XI. Guinea-Congolia/ Sudania regional transition zone (16).

191 *S. zenkeri*

192 No clear minimum CV was shown in the ADMIXTURE analyses. At K=2 LG-South splits  
193 from the rest of the samples. At K=3 the following clusters are observed: LG-North, LG-  
194 South, and a LG-east/Congo, with significant admixture between LG-South and LG-  
195 east/Congo. At K=4 the LG-east and the Congo clusters are retrieved as independent groups.  
196 At K=5 the LG-South is split into a Northern (LG-South<sub>1</sub>) and a Southern cluster (LG-  
197 South<sub>2</sub>), with admixed individuals between them, and also with LG-East. From K=6 and  
198 higher random genetic groups arise within LG-South<sub>1</sub>, LG-South<sub>2</sub> and LG-East.

199 *P. elata*

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

204

## 205 **Demographic inference**

206 We inferred the demographic history of each species using  $\partial a \partial I$  (Table 1, S4, S5). The fit  
207 substantially improved when considering a scenario of no migration between Upper Guinea  
208 and South Lower Guinea after divergence, in all cases except for one of the *E. ivorensis*  
209 scenarios. For each scenario, the two different levels of data filtering gave different estimates  
210 but congruent for each species. The scenarios with more SNPs (filtering level 1) gave older  
211 split dates while the scenarios with fewer SNPs (filtering level 2) fitted slightly more recent  
212 splits. The most restrictive filtering (filtering level 3) yielded fewer than 10,000 SNPs in all  
213 cases so estimates are not reliable.

214 For the divergence between LG-North and LG-South,  $\partial a \partial i$  produced different estimates  
215 across species (Table 1). The most likely models estimated the most recent split ca. 29 Kyr.  
216 BP for *D. benthamianus*, followed by *E. ivorensis* ca. 103 Kyr BP and *E. suaveolens* ca. 359  
217 Kyr. BP. The oldest was in *S. zenkeri* ca. 3.5 Myr. BP. The scenarios built using filtering level  
218 2 generally inferred younger North-South splits but the differences across species persisted  
219 (ca. 7.6 Kyr in *D. benthamianus*, ca. 77 Kyr. BP in *E. ivorensis*, ca. 102 Kyr. BP in *E.*  
220 *suaveolens*, and ca. 3.1 Myr BP in *S. zenkeri*).

221

## 222 **Gradients of genetic diversity over space**

223 We traced the spatial signal of recolonisation after forest fragmentation, by relating genetic  
224 diversity of individuals per species with distance from the three hypothesised LGM forest  
225 refugia, accounting for differences across gene pools (Figure 2). We found a significant  
226 negative relationship between observed heterozygosity,  $H_o$ , and distance to refugia only for  
227 *D. benthamianus* and *P. elata* (Figure 2, Table 2). Higher genetic diversity is found in  
228 individuals of *D. benthamianus* that are closer to the LGM forest refugia: LGM-Maley  
229 ( $p < 0.01$ ), LGM-Ahnuf ( $p < 0.01$ ), and LGM-species niche model ( $p < 0.05$ ). The present-day  
230 distribution of *D. benthamianus* is mostly closer to the coast rather than in the core of the  
231 Congo forest (Figure 1), All three postulated refugia suggest large areas of forest survival  
232 during the LGM along the coasts of Cameroon and Gabon thus, the observed gradient in  $H_o$   
233 is dominated by the distance to the Coastal refugia rather than to the Congo refugia (Figure  
234 2). For *P. elata* we detected a significant decline of the genetic diversity as individuals are  
235 located further away from LGM-Maley ( $p < 0.05$ ). As the distribution of *P. elata* lays between  
236 the Coastal and the Congo refugia (Figure 1, S6), we tested the relationship between  $H_o$  and  
237 the distances to both refugia separately. Here we found a significant negative relationship  
238 with distance to the LGM-Maley-Congo ( $p < 0.01$ ) and a positive relationship with scenarios  
239 LGM-Maley-Coastal. We also found a positive significant relationship with LGM-Anhuf  
240 (Figure 2, Table 2), where coastal refugia is closer to current *P. elata* populations. Finally,  
241 the LGM- species niche model, which shows only inland LGM refugia restricted to NE  
242 Cameroon, does not exhibit any significant relationship.

243 In the remaining species *E. ivorensis*, *E. suaveolens*, and *S. zenkeri* no significant  
244 relationships were found (Figure 2, Table 2, S6). In the case of *E. ivorensis* the results need to  
245 be taken with caution as low sampling sizes may have reduced statistical power.

246

## 247 **DISCUSSION**

248 For the five rainforest tree species investigated, at least three main intraspecific lineages  
249 were identified in the phylogenetic analyses. An early divergent lineage in West Africa  
250 (Upper Guinea, UG) was detected in all species occurring in this forest block (*D.*  
251 *benthamianus*, *E. suaveolens*, *E. ivorensis*, and *P. elata*). For all species widespread in  
252 Central Africa (Lower Guinea, LG) two lineages were retrieved: Northern Lower Guinea, LG-  
253 north) and a Southern Lower Guinea, LG-south (*D. benthamianus*, *E. suaveolens*, *E.*  
254 *ivorensis*, and *S. zenkeri*).

255

## 256 **Early divergence in West African rainforests**

257 The finding of early divergent lineages in West Africa for all the species is in tune with the  
258 hypothesis that Upper Guinea is an independent biogeographic region with numerous  
259 endemic species (8, 16-18). Although the fossil record indicates that the Dahomey gap was

260 forested in the last interglacial, and thus Upper and Lower Guinea were connected between  
261 8,400 and 4,500 years ago (1, 19), our data show that no genetic homogenisation between  
262 the two forest blocks occurred. It is likely that previous interglacials were less humid than  
263 the last one so there is no guarantee that the Dahomey Gap became forested during those  
264 periods (2). This seems to have favoured the long-term differentiation of the two forest  
265 blocks. To our knowledge this is the first study that estimates well-supported intraspecific  
266 phylogenies between Upper and Lower Guinean plant populations using outgroups in order  
267 to estimate the direction of evolution. Based on relatively small numbers of markers, early  
268 divergent lineages, c.a. 500 Kyr. BP, have also been found in west African populations of  
269 chimpanzees (3, 20, 21) and woodpeckers (22). Studies on other African lowland rainforest  
270 birds (23-25), forest-dwelling rodents (26) and African bushbucks suggest that haplotypes  
271 are rarely shared between populations sampled across the Dahomey Gap although the  
272 relationships between clades is not well resolved.  
273

#### 274 ***North-South Genetic differentiation in Central Africa suggests Rainforest*** 275 ***Fragmentation during the Ice-Ages***

276 For each species, our phylogenetic analyses identified northern and southern lineages in  
277 Central Africa that do not correspond to any geographic barrier or current discontinuity in  
278 the distribution of the forest. These results are consistent with the genetic structuring of  
279 other tropical trees in Lower Guinea (6, 7), as well as with the ADMIXTURE analyses, which  
280 provided additional insight into the levels of admixture among the central African lineages of  
281 each species. While *D. benthamianus* and *E. ivorensis* showed admixture between the North  
282 and the South, *E. suaveolens* and *S. zenkeri* exhibited sharp differentiation and no admixed  
283 individuals between the two regions.

284

285 Using  $\delta\text{a}\delta\text{i}$  to reconstruct the population history of the four species present in North and  
286 South Lower Guinea, our best-fit comparison consistently supported models involving  
287 North-South divergence and subsequent gene flow. We also noticed that the fit substantially  
288 improved when considering no migration between Upper Guinea and South Lower Guinea  
289 after divergence. These clades diverged within the Upper Pliocene and the Pleistocene, with  
290 the oldest genetic split found in *S. zenkeri* ca. 3.4-3.1 Myr. BP, and the most recent in *D.*  
291 *benthamianus* ca. 29,000-7,700 yr. BP. Our time estimates indicate North-South  
292 differentiation of the forest populations during the dry glacial climatic periods that took  
293 place from the Pliocene and especially during the Pleistocene (9). Altogether, our data are  
294 compatible with the differentiation of the genetic lineages in Northern and Southern Lower  
295 Guinea as a result of forest fragmentation during the dry glacial periods and subsequent  
296 admixture as a result of forest expansion during the humid interglacial periods.

297

#### 298 ***Decline of genomic diversity from putative glacial refugia***

299 We traced the spatial signal of recolonisation during the humid periods, by examining  
300 genetic diversity gradients over space. We found significant declines of genetic diversity from  
301 coastal refugia in *D. benthamianus* and from inland Congo refugia in *Pericopsis elata*. This  
302 finding suggests the survival of *D. benthamianus* in coastal refugia in Cameroon and Gabon,  
303 although the exact location cannot be determined since no differences were revealed among  
304 the different hypotheses of LGM forest refugia considered based on the palaeoclimatic  
305 reconstructions (14, 27) or estimated from the potential climatic distribution of each species  
306 during the LGM. In the case of *P. elata* survival in inland forest refugia postulated by Maley

307 in the Congo Basin is highly probable. For *S. zenkeri* and *E. suaveolens* no significant  
308 declines of the genetic diversity were found.

309

### 310 *Genomic diversity and differentiation and dispersal capacities*

311 Our estimates of admixture and split dates are consistent with prior knowledge of the biology  
312 of the study species. In particular, with two traits that may be key for colonising new areas  
313 after forest fragmentation: light tolerance and dispersal capacity. Long-distance dispersal of  
314 the seeds is especially relevant in the case of early successional communities and expanding  
315 populations, as it not only transports seeds over very long distances but also generates  
316 establishment opportunities. *ada* detected more recent splits between the North and South  
317 in *D. benthamianus*, than for the two *Erythrophleum* species. The oldest split was detected  
318 in the shade-tolerant, non-assisted dispersed species *S. zenkeri*. The fact that we detected  
319 more recent signals of North-South fragmentation in the species with long-distance dispersal  
320 capacity suggests that the old signals of fragmentation may be more easily erased in these  
321 species than in species with limited dispersal like *S. zenkeri*. The ability of long-distance  
322 dispersal may have made a difference over subsequent cycles of forest fragmentation and  
323 recolonisation that took place during the Pleistocene. Our results also show a decline of  
324 genetic diversity from forest refugia in long-distance-dispersed species only. While *D.*  
325 *benthamianus* likely survived in coastal refugia, *P. elata* probably did so in inland refugia in  
326 the Congo Basin. The fact that we were able to trace the genetic significant declines of genetic  
327 diversity outside refugia in the long-distance dispersed species suggests that we may be  
328 detecting the signal of a recent dispersal for those species.

329

### 330 **Conclusions**

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

342

## 343 MATERIALS AND METHODS

### 344 **Study species: dispersal capacity, light tolerance, reproductive system**

345

346 *Pericopsis elata* is a light-demanding, pioneer or non-pioneer and wind-dispersed. Seeds  
347 disperse on average 214 m with a very flat tail. Most seeds disperse less than 100m, but a  
348 significant amount of seeds experience long distance dispersal (>1 km) (Olivier Hardy, personal  
349 observation). *D. benthamianus* is a pioneer light-demanding and wind-dispersed species  
350 (28). Individual trees are not aggregated in the field. It is an indicator of disturbed  
351 rainforests. Although most of the *D. benthamianus* seeds disperse over short distances (c. 70  
352 m), 30% of seed immigration was detected (15). This indicates that the distribution of  
353 dispersal distances is very flat tailed, indicative of long-distance dispersal events. The two  
354 *Erythrophleum* species are light-demanding (28), the coastal congener *E. ivorensis* is a



355 pioneer while *E. suaveolens* is a non-pioneer (Anais Gorel, personal communication). The  
356 fruits exhibit primary ballistic dispersal and secondary dispersal by primates, *Cephalophus*  
357 species and rodents has been reported (29). The distribution is non gregarious, and they are  
358 indicators of secondary rainforests. Based on genetic markers seed dispersal distances of 210  
359 m were detected in *E. suaveolens*, but long distance dispersal events seem rare (30).  
360 *Scorodophloeus zenkeri* is a shade tolerant species that exhibits ballistic dispersal through  
361 the explosion of the pods (R Piñeiro 2017, personal observation). Trees are locally aggregated  
362 in the field. It requires high environmental humidity (31) and is an indicator of undisturbed  
363 rainforests.

364

### 365 **DNA extraction and Genotyping by Sequencing**

366 Leaf and cambium material were collected in the rainforests of West and Central Africa  
367 between 2005 and 2014 (S8). The samples were immediately dried with silica-gel in order to  
368 preserve the DNA quality. Between 18 and 46 samples of each species were selected in order  
369 to represent their distribution in Central Africa, with special emphasis in the biogeographic  
370 region of Lower Guinea. A few samples from the less accessible rainforest of the Congo Basin  
371 were included. Outgroup taxa (S8) were selected based on available legume phylogenies (32,  
372 33).

373

374 Overall 362 GBS libraries, from 182 individuals, were initially sequenced on four Illumina  
375 lanes (HiSeq2000 San Diego, CA, USA), using 100-bp Single Read chemistry. For each  
376 library two DNA extractions were performed using the DNeasy Plant Minikit columns  
377 (Qiagen), and pooled in order to generate sufficient DNA for the GBS protocol. One blank  
378 per plate was included. DNA quality was checked on a 1.5% agarose gel and DNA quantity  
379 was measured with Qbit HS (Life technologies, Grand Island, NY). The DNA was purified  
380 with a ZR-96 DNA Clean up kit (Zymo Research Corp.). Subsequently, genotyping by  
381 Sequencing (GBS) was performed at the Genomic Diversity and Computational Biology  
382 Service Unit at Cornell University (Ithaca, NY) according to a published protocol (34). One  
383 microgram of DNA of each species was initially used in order to optimise the GBS protocol,  
384 in particular aiding the choice of the most appropriate restriction enzyme. Specifically, three  
385 libraries were built for each species using three different enzymes: ApeKI (4.5-base cutter),  
386 EcoT22I and PstI (both 6-base cutters) and checked for appropriate fragment sizes (<500bp)  
387 and distribution on an Experion automatic electrophoresis system (Bio-Rad laboratories,  
388 USA). Given these results, we elected to use the enzyme EcoT22I for subsequent data  
389 generation, as it yielded appropriate fragment sizes (<500bp) and distributions for all study  
390 species.

391

### 392 **Genotype Calling and Site Frequency Spectrum**

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

397

#### 398 *Genotype calls without outgroups*

399 For those analyses that require single nucleotide polymorphism (SNP) and genotype calls,  
400 we used the Universal Network-Enabled Analysis Kit (UNEAK) pipeline (35) within the  
401 software TASSEL 3.0 (36), suitable for analysis of GBS data at the intraspecific level in the  
402 absence of a reference genome. Reads were trimmed to 64 bps (to avoid sequencing errors at  
403 the ends of reads) and identical reads collapsed into tags (i.e. alleles). Tag pairs having a  
404 single base pair mismatch were identified as candidate loci and used for SNP calling. Tag  
405 pairs forming complicated networks, likely to result from repeats, paralogs and sequencing

406 error, were filtered out. In order to filter out false-positive SNPs: (i) the minimum number of  
407 reads per GBS tag (i.e. allele) was set to five (ii), an error tolerance rate (ETR) of 0.05 was  
408 established, and (iii) SNPs with a genotype missing rate >50% were removed.

#### 409 *Genotype likelihood framework for analyses with outgroups*

410 Building a “reference” catalog: As the first step in estimating genotype likelihoods at variable  
411 sites across the sequenced loci, we constructed a reference catalog against which we could  
412 align the GBS reads. We used the tags built with TASSEL to obtain 64 bp long reads.  
413 Subsequently, the query read sequences from TASSEL were concatenated with a spacer of  
414 200 Ns to obtain the reference.

415 Genotype likelihood computation: For many of the downstream analyses, a genotype calling  
416 approach from NGS data may bias the population genetic estimates (37). Therefore we  
417 compute the genotype likelihoods at all the variable sites using the software Analysis of Next-  
418 Generation Sequencing Data, ANGSD v0.914 (38). First, we mapped the GBS reads from  
419 each sample to the constructed reference catalog using the PALEOMIX pipeline (39). As part  
420 of this pipeline, AdapterRemoval v2 (40) was used to trim adapter sequences, merge the  
421 paired reads and discard reads shorter than 30 bp. BWA v0.7.15 was then used to map the  
422 processed reads to the reference (41). Minimum mapping quality was set at 15 while  
423 minimum base quality was set to five. Subsequently, genotype likelihoods were calculated in  
424 ANGSD using the following filters, (i) the baq option was used to recompute base alignment  
425 qualities, allowing us to reduce false SNPs due to misalignment, (ii) a Hardy-Weinberg  
426 Equilibrium (HWE) p-value greater than 0.001, and (iii) the total coverage at any site cannot  
427 exceed 55 times the number of samples. The depth distribution was plotted for each species  
428 and a cut-off was chosen in order to exclude outlier sites at extremely high coverage. Given  
429 the diversity in our reads, the -C filter, downgrading mapping quality for reads containing  
430 excessive mismatches, was not used.

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

440

#### 441 **Inference of genetic clusters**

442 The software Admixture v.13 was used to calculate the probability of assignment of  
443 individuals to genetic clusters (42) from the genotype calls datasets without outgroups.  
444 Individuals exhibiting admixture between clusters were detected. For each species, K = 1 to  
445 12 genetic clusters were tested with 20 randomly seeded replicate runs at each K and five-  
446 fold cross-validation (CV). Barplots showing the probability of assignment of individuals to  
447 the genetic clusters were generated in R using the package ggplot2.

448 The number of K that best fits the data was estimated using the lowest five-fold CV. Since  
449 these values did not always work appropriately due to low sample sizes in some of the  
450 clusters (43), we used (i) the congruence of the genetic clusters with the geography, and (ii)  
451 the congruence of the genetic clusters with the phylogenies as additional criteria to select the  
452 optimal K.

453

#### 454 **Phylogenetic relationships**

455 Maximum likelihood phylogenies were built for each species using RAxML 8.0 (44) from the  
456 Genotype Calls datasets with outgroups. Rooting with outgroups allowed us to assess the  
457 direction of evolution. All SNPs were concatenated into a single alignment, with missing data  
458 (Ns) and heterozygous positions entered as needed. Bootstrap support was calculated from  
459 100 replicate searches with random starting trees using the GTR+ gamma nucleotide  
460 substitution model.

461

#### 462 **Demographic inference with $\delta\alpha\delta\iota$**

463 Population demographic models were estimated using  $\delta\alpha\delta\iota$  (45), a method that uses the SFS  
464 of populations to infer their demographic history. Based on the admixture and phylogeny  
465 results, for the three species widespread in Upper and Lower Guinea -*D. benthamianus*, *E.*  
466 *suaveolens*, *E. ivorensis*-, we fitted a three-population model with the tree topology (Upper  
467 Guinea-UG (Northern Lower Guinea-LG-North, Southern Lower Guinea-LG-South) (S4 A).  
468 In addition to a model with symmetric migration among the three clusters, we fitted a model  
469 with no migration allowed between UG and LG-South after the split. For *S. zenkeri*, absent in  
470 Upper Guinea, we estimated split times based on the tree topology (LG-North, (LG-South1  
471 and LG-South2)). The observed SFS was compared to the expected SFS under the Isolation  
472 with Migration (IM) model with symmetric gene flow (S4 B-E). Using Maximum Likelihood,  
473 split time parameters values were estimated in generations. We assumed a mutation rate of  $\mu$   
474 =  $2.5 \times 10^{-9}$  ( $1.7 \times 10^{-9}$  to  $3.5 \times 10^{-9}$ ) per site per year, estimated for *Populus* (46, 47) and  
475 a generation time of 100 years (12, 48). To avoid biasing the demographic inferences due to  
476 uneven depth of coverage, which is typical of GBS data, we excluded all singletons, i.e. alleles  
477 found only once among the populations, while estimating the demographic parameters.  
478 Further, while estimating SFS, admixed individuals with less than 70% genetic ancestry to a  
479 single group were excluded.

480 For each such demographic model, we ran 100 replicates, and selected the parameter  
481 estimates from the best fitting model, i.e. the model with the highest log-likelihood, with one  
482 important exception. Model fits in  $\delta\alpha\delta\iota$ , which masked more than 5% of the SFS were  
483 excluded. Finally, we performed 1000 replicates for *D. benthamianus*, as this species yielded  
484 fewer SNPs and larger sample sizes, in order to get a stable estimate of the demography.

485

#### 486 **Gradients of genetic diversity over space and climatic niche modelling**

487 In order to visualise the patterns of genetic diversity over space in Lower Guinea, for each  
488 species we calculated the genetic diversity of each of the genotyped individuals, and plotted it  
489 against the geographical distance from the closest Last Glacial Maxima refugia (LGM). A  
490 mixed effect model was performed using the genetic diversity as the independent variable,  
491 the distance to LGM refugia as explanatory variable (minimal distance between the sample  
492 and the limit of a postulated refugia) and the genetic cluster as a random variable in order to  
493 account for genetic diversity differences across genetic lineages

494 We calculated the observed heterozygosity ( $H_o$ ) of each individual as a proxy for genetic  
495 diversity using GenALEx (49). Admixed individuals with less than 70% genetic ancestry to a  
496 single group were excluded. Three different hypotheses of LGM forest refugia were  
497 considered. The forest refugia postulated by Maley and Anhuf, based on palaeoclimatic and  
498 palynological data (Figure 2): (i) LGM-Maley (14), and (ii) LGM-Anhuf, (27). In addition,

499 specific forest refugia based on the LGM niche models of each species were tested (Figure 2,  
500 S7): (iii) LGM-species niche models (see below).

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

505

## 506 **Species distribution models for the LGM**

507 We collected occurrence records for each species from RAINBIO database (50) and our own  
508 records. We visually inspected records to eliminate outliers and deduplicated records per  
509 species to obtain one per pixel according to the climate layers resolution (see below). After  
510 this, we had 1,029 occurrences distributed among species as follows: *D. benthamianus*  
511 (n=427), *E. ivorensis* (n=74), *E. suaveolens* (n = 180), *P. elata* (n=118), *S. zenkeri* (n=230)  
512 (S7).

513

514 We used the 19 bioclim layers from the Worldclim dataset v1.4 at 2.5 arc-min (51)  
515 representing the climate between 1960-1990 and calculated the total precipitation of the  
516 Austral Summer (December, January and February) using the monthly layers. After  
517 performing correlation analysis and Principal component analysis (PCA) we finally selected 5  
518 variables that were minimally correlated to run the models: annual mean temperature (bio1),  
519 temperature annual range (bio7), total annual precipitation (bio12), precipitation of the  
520 driest month (bio14), and precipitation of December, January and February (pp\_djf). We  
521 also used the selected variables from simulations with the CCSM4 global climate model for  
522 the LGM from Worldclim to project species distributions for that time.

523

524 We modelled the distribution of the five species with the biomod2 package in R (52). We  
525 randomly selected 5000 pseudoabsences. We used three different algorithms (generalized  
526 linear models, GLM; random forest, RF and Maxent) and five repetitions, obtaining a total of  
527 15 models for each species. Models were calibrated using 80% of the occurrences and  
528 evaluated using the remaining 20% of occurrences and the TSS and ROC statistics (53). We  
529 used a consensus approach to produce an ensemble model using the weighted mean of all  
530 models which had at least TSS values above 0.7 and ROC values above 0.8. Binary maps  
531 (presence/absence maps) were produced using the TSS threshold (S7).

532

## 533 **ACKNOWLEDGEMENTS**

534 We would like to thank Boris Demenou, Armel Donkpegan, Myriam Heuertz, Théophile  
535 Ayolle, Charlotte Hansen, Esra Kaymak, Jean-Louis Doucet, Jérôme Duminil, Jérôme Chave,  
536 and Hermann Daniel. This work received financial support from the Marie Curie FP7-  
537 PEOPLE-2012-IEF program (project AGORA) awarded to Rosalía Piñeiro.

538

539

## 540 **REFERENCES**

- 541 1. Salzmann U & Hoelzmann P (2005) The Dahomey Gap: an abrupt climatically induced  
542 rain forest fragmentation in West Africa during the late Holocene. *The Holocene*  
543 15(2):190-199.
- 544 2. Miller CS & Gosling WD (2014) Quaternary forest associations in lowland tropical West  
545 Africa. *Quaternary Science Reviews* 84:7-25.

- 546 3. Hey J (2009) The divergence of chimpanzee species and subspecies as revealed in  
547 multipopulation isolation-with-migration analyses. *Molecular biology and evolution*  
548 27(4):921-933.
- 549 4. Telfer PT, *et al.* (2003) Molecular evidence for deep phylogenetic divergence in  
550 *Mandrillus sphinx*. *Molecular Ecology* 12(7):2019-2024.
- 551 5. Anthony NM, *et al.* (2007) The role of Pleistocene refugia and rivers in shaping gorilla  
552 genetic diversity in central Africa. *Proceedings of the National Academy of Sciences*  
553 104(51):20432-20436.
- 554 6. Hardy OJ, *et al.* (2013) Comparative phylogeography of African rain forest trees: a  
555 review of genetic signatures of vegetation history in the Guineo-Congolian region.  
556 *Comptes Rendus Geoscience* 345(7-8):284-296.
- 557 7. Heuertz M, Duminil J, Dauby G, Savolainen V, & Hardy OJ (2014) Comparative  
558 phylogeography in rainforest trees from Lower Guinea, Africa. *PloS one* 9(1):e84307.
- 559 8. Droissart V, *et al.* (2018) Beyond trees: biogeographical regionalization of tropical  
560 Africa. *Journal of biogeography* 45(5):1153-1167.
- 561 9. Gibbard P, Ehlers J, & Hughes P (2016) Quaternary Glaciations. *International*  
562 *Encyclopedia of Geography: People, the Earth, Environment and Technology: People, the*  
563 *Earth, Environment and Technology*:1-10.
- 564 10. Duminil J, *et al.* (2015) Late Pleistocene molecular dating of past population  
565 fragmentation and demographic changes in African rain forest tree species supports the  
566 forest refuge hypothesis. *Journal of biogeography* 42(8):1443-1454.
- 567 11. Faye A, *et al.* (2016) Phylogenetics and diversification history of African rattans  
568 (Calamoideae, Ancistrophyllinae). *Botanical journal of the Linnean Society* 182(2):256-  
569 271.
- 570 12. Piñeiro R, Dauby G, Kaymak E, & Hardy OJ (2017) Pleistocene population expansions of  
571 shade-tolerant trees indicate fragmentation of the African rainforest during the Ice Ages.  
572 *Proc. R. Soc. B* 284(1866):20171800.
- 573 13. Migliore J, *et al.* (2019) Pre-Pleistocene origin of phylogeographical breaks in African  
574 rain forest trees: New insights from Greenwayodendron (Annonaceae) phylogenomics.  
575 *Journal of biogeography* 46(1):212-223.
- 576 14. Maley J (1996) The African rain forest—main characteristics of changes in vegetation and  
577 climate from the Upper Cretaceous to the Quaternary. *Proceedings of the Royal Society of*  
578 *Edinburgh, Section B: Biological Sciences* 104:31-73.
- 579 15. Hardy OJ, *et al.* (2019) Seed and pollen dispersal distances in two African legume timber  
580 trees and their reproductive potential under selective logging. *Molecular ecology*  
581 28:3119-3134.
- 582 16. White F (1983) The vegetation of Africa: a descriptive memoir to accompany the  
583 UNESCO/AETFAT/UNSO vegetation map of Africa by F White. *Natural Resources*  
584 *Research Report XX, UNESCO, Paris, France.*
- 585 17. Fayolle A, *et al.* (2014) Patterns of tree species composition across tropical African  
586 forests. *Journal of Biogeography* 41(12):2320-2331.
- 587 18. Linder HP, *et al.* (2012) The partitioning of Africa: statistically defined biogeographical  
588 regions in sub-Saharan Africa. *Journal of Biogeography* 39(7):1189-1205.
- 589 19. Altschul SF, *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein  
590 database search programs. *Nucleic acids research* 25(17):3389-3402.
- 591 20. Gagneux P, Gonder MK, Goldberg TL, & Morin PA (2001) Gene flow in wild chimpanzee  
592 populations: what genetic data tell us about chimpanzee movement over space and time.  
593 *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*  
594 356(1410):889-897.
- 595 21. Gonder MK, *et al.* (1997) A new west African chimpanzee subspecies? *Nature*  
596 388(6640):337.
- 597 22. Fuchs J & Bowie RC (2015) Concordant genetic structure in two species of woodpecker  
598 distributed across the primary West African biogeographic barriers. *Molecular*  
599 *phylogenetics and evolution* 88:64-74.

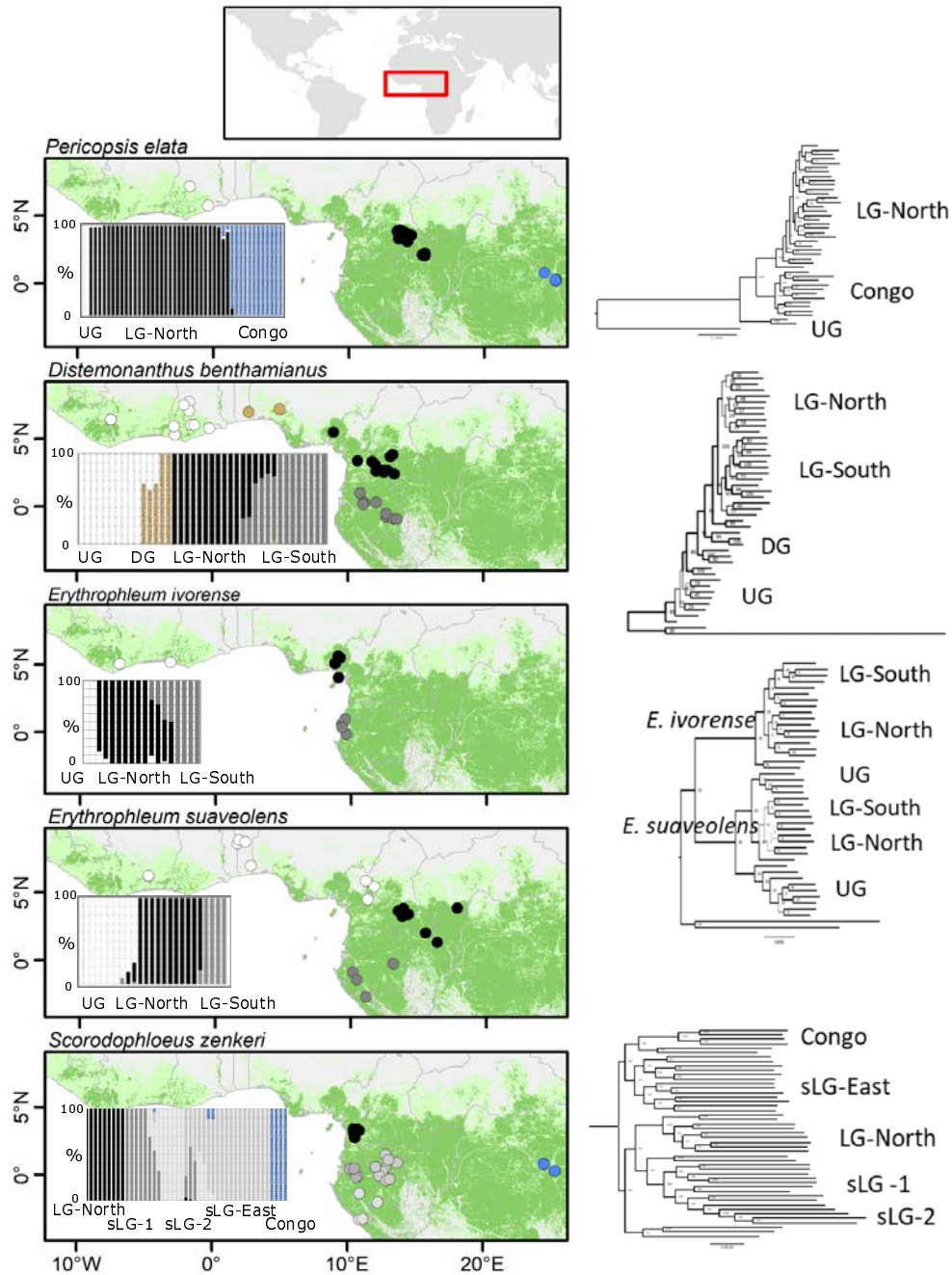
- 600 23. Beresford P & Cracraft J (1999) Speciation in African forest robins (*Stiphronis*): species  
601 limits, phylogenetic relationships, and molecular biology. *American Museum novitates*;  
602 no. 3270.
- 603 24. Schmidt BK, Foster JT, Angehr GR, Durrant KL, & Fleischer RC (2008) A new species of  
604 African forest robin from Gabon (Passeriformes: Muscicapidae: *Stiphronis*).
- 605 25. Marks BD (2010) Are lowland rainforests really evolutionary museums?  
606 Phylogeography of the green hylia (*Hylia prasina*) in the Afrotropics. *Molecular*  
607 *Phylogenetics and Evolution* 55(1):178-184.
- 608 26. Nicolas V, *et al.* (2011) The roles of rivers and Pleistocene refugia in shaping genetic  
609 diversity in *Praomys misonnei* in tropical Africa. *Journal of Biogeography* 38(1):191-207.
- 610 27. Anhof D, *et al.* (2006) Paleo-environmental change in Amazonian and African rainforest  
611 during the LGM. *Palaeogeography, Palaeoclimatology, Palaeoecology* 239(3-4):510-527.
- 612 28. Meunier Q, Moumbogou C, & Doucet J-L (2015) *Les arbres utiles du Gabon* (Presses  
613 agronomiques de Gembloux).
- 614 29. Gorel A-P, Fayolle A, & Doucet J-L (2015) Ecology and management of the multipurpose  
615 *Erythrophleum* species (Fabaceae-Caesalpinioideae) in Africa. A review. *Biotechnologie,*  
616 *Agronomie, Société et Environnement* 19(4):415-429.
- 617 30. Duminil J, *et al.* (2010) CpDNA-based species identification and phylogeography:  
618 application to African tropical tree species. *Molecular ecology* 19(24):5469-5483.
- 619 31. Abanda G (2011) La Surexploitation du *Scorodophloeus zenkeri* Harms (Arbre aAil):  
620 Atténuer l Impact de la Gestion Non Durable de l Arbre aAil dans le Massif Forestier de  
621 Ngovayang du Sud Cameroun. *Editions Universitaires Europeennes*.
- 622 32. Bruneau A, Mercure M, Lewis GP, & Herendeen PS (2008) Phylogenetic patterns and  
623 diversification in the caesalpinoid legumes. *Botany* 86(7):697-718.
- 624 33. Azani N, *et al.* (2017) A new subfamily classification of the Leguminosae based on a  
625 taxonomically comprehensive phylogeny The Legume Phylogeny Working Group  
626 (LPWG). *Taxon* 66(1):44-77.
- 627 34. Elshire RJ, *et al.* (2011) A robust, simple genotyping-by-sequencing (GBS) approach for  
628 high diversity species. *PLoS one* 6(5):e19379.
- 629 35. Lu F, *et al.* (2013) Switchgrass genomic diversity, ploidy, and evolution: novel insights  
630 from a network-based SNP discovery protocol. *PLoS genetics* 9(1):e1003215.
- 631 36. Bradbury PJ, *et al.* (2007) TASSEL: software for association mapping of complex traits in  
632 diverse samples. *Bioinformatics* 23(19):2633-2635.
- 633 37. Nielsen R, Paul JS, Albrechtsen A, & Song YS (2011) Genotype and SNP calling from next-  
634 generation sequencing data. *Nature Reviews Genetics* 12(6):443.
- 635 38. Korneliussen TS, Albrechtsen A, & Nielsen R (2014) ANGSD: analysis of next generation  
636 sequencing data. *BMC bioinformatics* 15(1):356.
- 637 39. Schubert M, *et al.* (2014) Characterization of ancient and modern genomes by SNP  
638 detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nature*  
639 *protocols* 9(5):1056.
- 640 40. Lindgreen S (2012) AdapterRemoval: easy cleaning of next-generation sequencing  
641 reads. *BMC research notes* 5(1):337.
- 642 41. Li H & Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler  
643 transform. *bioinformatics* 25(14):1754-1760.
- 644 42. Alexander DH, Novembre J, & Lange K (2009) Fast model-based estimation of ancestry  
645 in unrelated individuals. *Genome research*.
- 646 43. Wang J (2017) The computer program STRUCTURE for assigning individuals to  
647 populations: easy to use but easier to misuse. *Molecular ecology resources* 17(5):981-  
648 990.
- 649 44. Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis  
650 of large phylogenies. *Bioinformatics* 30(9):1312-1313.
- 651 45. Gutenkunst RN, Hernandez RD, Williamson SH, & Bustamante CD (2009) Inferring the  
652 joint demographic history of multiple populations from multidimensional SNP  
653 frequency data. *PLoS genetics* 5(10):e1000695.

- 654 46. Tuskan GA, *et al.* (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. &  
655 Gray). *science* 313(5793):1596-1604.
- 656 47. Ingvarsson PK (2008) Multilocus patterns of nucleotide polymorphism and the  
657 demographic history of *Populus tremula*. *Genetics*.
- 658 48. Baker TR, *et al.* (2014) Fast demographic traits promote high diversification rates of  
659 Amazonian trees. *Ecology Letters* 17(5):527-536.
- 660 49. Peakall R & Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic  
661 software for teaching and research. *Molecular ecology notes* 6(1):288-295.
- 662 50. Dauby G, *et al.* (2016) RAINBIO: a mega-database of tropical African vascular plants  
663 distributions. *PhytoKeys* (74):1.
- 664 51. Hijmans RJ, Cameron SE, Parra JL, Jones PG, & Jarvis A (2005) Very high resolution  
665 interpolated climate surfaces for global land areas. *International Journal of Climatology*:  
666 *A Journal of the Royal Meteorological Society* 25(15):1965-1978.
- 667 52. Thuiller W, Lafourcade B, Engler R, & Araújo MB (2009) BIOMOD—a platform for  
668 ensemble forecasting of species distributions. *Ecography* 32(3):369-373.
- 669 53. Zou KH, Liu A, Bandos AI, Ohno-Machado L, & Rockette HE (2011) *Statistical evaluation*  
670 *of diagnostic performance: topics in ROC analysis* (CRC Press).
- 671 54. Mayaux P, Bartholomé E, Fritz S, & Belward A (2004) A new land-cover map of Africa for  
672 the year 2000. *Journal of Biogeography* 31(6):861-877.

673

674 FIGURES

675 Figure 1. Genetic ancestry and phylogenetic reconstructions for *Pericopsis elata*,  
 676 *Distemonanthus benthamianus*, *Erythrophleum ivorense*, *E. suaveolens*, and  
 677 *Scorodophloeus zenkeri* based on Genotyping by Sequencing (GBS). Left: ADMIXTURE  
 678 analyses. Barplots show the probability of assignment of individuals to genetic clusters.  
 679 Centre: geographic distribution of individual trees according to the genetic cluster they  
 680 belong to (admixed individuals with <70% assignment probability to a single genetic cluster  
 681 have been removed). Right: RAxML phylogenetic analyses using the GBS genotype calls with  
 682 outgroups. Branch width is proportional to bootstrap supports. Green areas in map  
 683 represent present-day rainforest cover (54) (see also S3).





685 Figure 2. Decline of genetic diversity, based on Genotyping by Sequencing, with distance  
 686 from refugia in Central Africa for *Pericopsis elata*, *Distemonanthus benthamianus*,  
 687 *Erythrophleum ivorense*, *E. suaveolens*, and *Scorodophloeus zenkeri*. Linear regression of  
 688 the genetic diversity -observed heterozygosity,  $H_o$ - of each genotyped individual tree against  
 689 the geographical distance from the closest Last Glacial Maxima refugia ca. 20,000 yr. BP,  
 690 and the genetic cluster as a random variable in order to account for genetic diversity  
 691 differences across genetic lineages. Admixed individuals with less than 70% genetic ancestry  
 692 to a single group were excluded. Three different hypotheses of LGM forest refugia were  
 693 considered. The forest refugia postulated by Maley and Anhuf, based on palaeoclimatic and  
 694 palynological data: (i) LGM-Maley (14), and (ii) LGM-Anhuf, (27). In addition, specific forest  
 695 refugia based on the LGM niche models of each species were tested (S7): (iii) LGM- species  
 696 niche models (see methods).

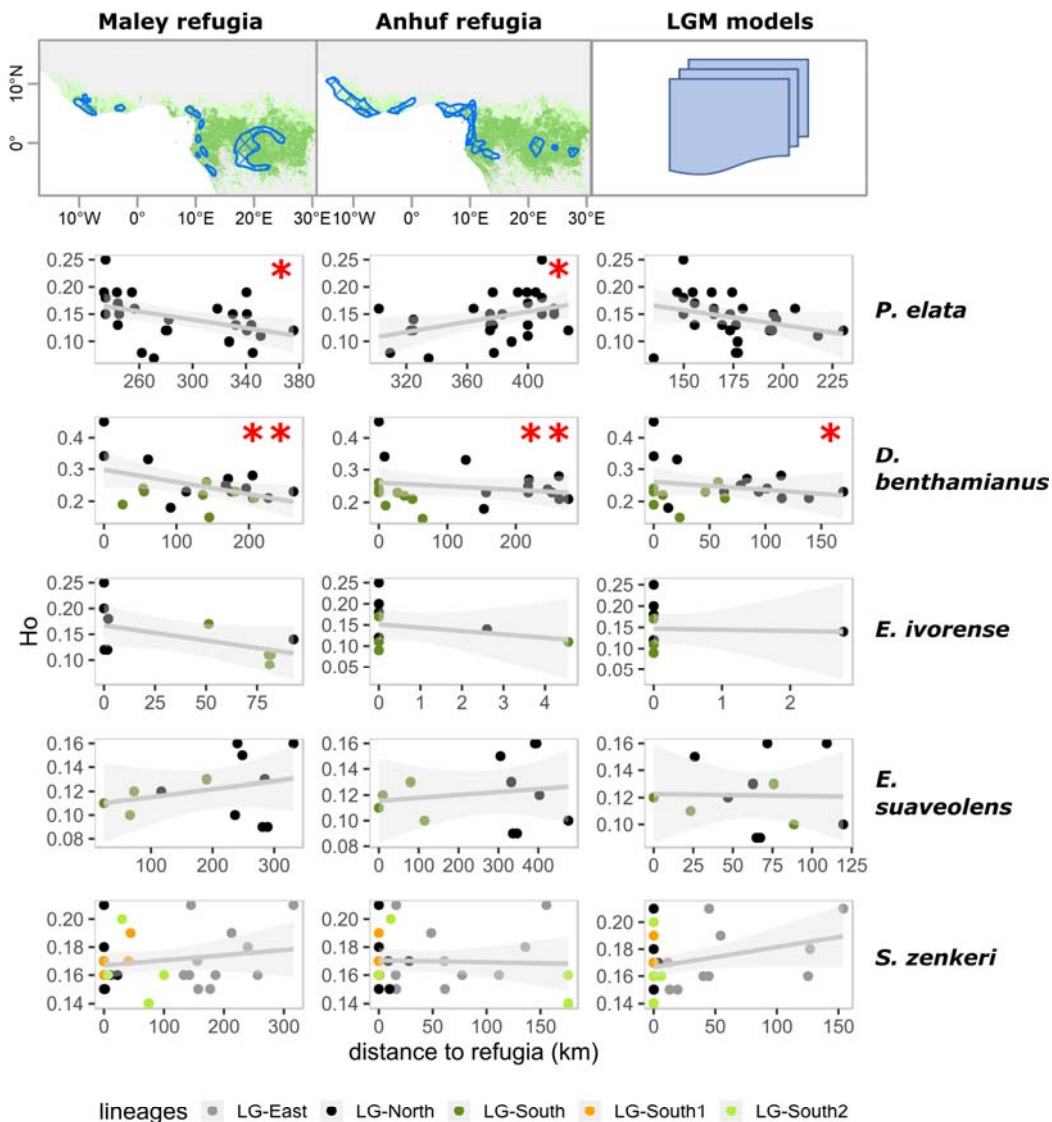


Table 1. Demographic history of *Distemonanthus benthamianus*, *Erythrophleum ivorense*, *E. suaveolens*, and *Scorodophloeus zenkeri* using  $\partial a\partial I$ .

	Dataset FILTERING 1				Dataset FILTERING 2			
	<i>D. benthamianus</i>	<i>E. ivorense</i>	<i>E. suaveolens</i>	<i>S. zenkeri</i>	<i>D. benthamianus</i>	<i>E. ivorense</i>	<i>E. suaveolens</i>	<i>S. zenkeri</i>
<b>Divergence time (Kyr. ago)</b>								
nLG – sLG	28.98	103.06	359.28	3436.99	7.70	76.90	101.78	3085.94
UG – (nLG, sLG)	1224.79	1151.31	3493.44		115.53	1034.19	1053.05	
sLG1, sLG2				1462.52				174.03
<b>Model details</b>								
Log-likelihood	-15764.60	-394.29	-25466.29	-17122.81	-11098.82	-637.26	-14168.11	-13590.11
# of SNPs	17921.38	29513.13	61667.59	70271.16	11323.50	22151.64	35042.61	54156.53

Divergence time in 1000 years are shown for the best-fitting model under Isolation with Migration. Based on the admixture and phylogeny results, for the three species widespread in Upper and Lower Guinea -*D. benthamianus*, *E. suaveolens*, *E. ivorense*-, we fitted a three-population model with the tree topology (Upper Guinea-UG (Northern Lower Guinea-nLG, Southern Lower Guinea-sLG). In addition to a model with symmetric migration among the three populations, we fitted a model with no migration between UG and sLG after the split. Note that the best fitting-model for the three species does not include a migration component between UG and sLG. For *S. zenkeri*, absent in Upper Guinea, we estimated split times based on the tree topology (Northern Lower Guinea-nLG, (Southern Lower Guinea 1-sLG1 and Southern Lower Guinea2-sLG2)). 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 (see full results in S5).

Table 2. Decline of genetic diversity with distance from refugia in Central Africa (Lower Guinea) for *Pericopsis elata*, *Distemonanthus benthamianus*, *Erythrophleum ivorense*, *E. suaveolens*, and *Scorodophloeus zenkeri*.

	LGM- Maley					LGM-Anhuf					LGM-spp				
	Value	Std.Error	DF	t-value	p-value	Value	Std.Error	DF	t-value	p-value	Value	Std.Error	DF	t-value	p-value
<i>P. elata</i>	<b>-0.00000038</b>	0.00000015	25	-2.44	<b>0.02</b>	0.00000047	0.00000021	25	2.27	0.03	-0.00000055	0.00000034	25	-1.61	0.12
<i>D. benthamianus</i>	<b>-0.00000044</b>	0.00000015	18	-2.90	<b>0.01</b>	<b>-0.00000059</b>	0.00000012	18	-4.79	<b>0.00</b>	<b>-0.00000059</b>	0.00000027	18	-2.21	<b>0.04</b>
<i>E. ivorense</i>	-0.00000058	0.000000338	8	-1.72	0.12	-0.00000692	0.000010134	8	-0.68	0.51	-0.00000595	0.000018381	8		0.75
<i>E. suaveolens</i>	0.00000007	0.000000074	9	0.89	0.40	0.00000002	0.000000047	9	0.49	0.63	-0.00000002	0.000000228	9	-0.07	0.95
<i>S. zenkeri</i>	0.00000004	0.000000038	24	0.97	0.34	-0.00000001	0.000000064	24	-0.22	0.83	0.00000015	0.000000082	24	1.82	0.08

Linear regression of the genetic diversity -observed heterozygosity,  $H_o$ - of each genotyped individual tree against the geographical distance from the closest Last Glacial Maxima refugia (LGM). Negative, significant correlations are highlighted in bold. Three different hypotheses of LGM forest refugia were considered. The forest refugia postulated by Maley and Anhuf, based on palynological and palaeoclimatic data (Figure 2): (i) LGM-Maley (14), and (ii) LGM-Anhuf (27). In addition, specific forest refugia based on the LGM niche models of each species were tested (Figure 2 and S7): (iii) LGM-spp.

