Tree Genetics & Genomes (2020) 16:15
https://doi.org/10.1007/s11295-019-1408-8

ORIGINAL ARTICLE

Revisiting the North-South genetic discontinuity in Central African tree populations: the case of the low-density tree species *Baillonella toxisperma*

Dyana Ndiade-Bourobou[1,2] · Kasso Daïnou[3,4,5] · Olivier J. Hardy[3] · Charles Doumenge[6] · Félicien Tosso[5] · Jean-Marc Bouvet[2]

Received: 21 September 2019 / Revised: 4 December 2019 / Accepted: 10 December 2019

Abstract

How the Central African rain forests have been affected by climatic fluctuations of the Quaternary remains debated. Phylogeographical studies have shown that tree species from western Central Africa often display spatially congruent genetic discontinuities, supporting the hypothesis that the forest was previously fragmented. Extensive seed dispersal is expected to accelerate the admixture between gene pools but most of the species studied so far have presumably limited seed dispersal abilities. Here, we genotyped 15 nuclear and three plastid microsatellite markers in a low-density Central African tree species with long-distance seed dispersal: *Baillonella toxisperma* (Sapotaceae). While plastid markers revealed a weak structure in Cameroon, nuclear markers highlighted three genetic clusters: two distributed in Cameroon and separating Atlantic coastal forests from the inland forests, and one cluster occurring in Gabon. Substantial genetic differentiation with a phylogeographical signal was detected only between Cameroonian and Gabonese populations, suggesting two major genetic clusters located approximately North and South of the equatorial climatic hinge. Genetic differentiation was very low between the clusters within Cameroon. This pattern could be partially explained by the climate niche distribution modelling applied on the Last Glacial Maximum (LGM) which predicts a unique remnant population per country. The deep North-South differentiation in a species with long-distance seed dispersal supports the hypothesis that Central African rain forests have been fragmented at the height of the equator during a substantial part of the Quaternary.

Keywords *Baillonella toxisperma* · Spatial genetic structure · Equator genetic cline · Nuclear microsatellites · Chloroplast microsatellites · Niche modelling

Communicated by J. Beaulieu

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11295-019-1408-8) contains supplementary material, which is available to authorized users.

✉ Kasso Daïnou
kdainou@uliege.be

[1] CENAREST, Institut de Recherches Agronomiques et Forestières (I.R.A.F), BP: 842 Gros-Bouquet, Libreville, Gabon

[2] CIRAD, Research Unit "Amélioration génétique et adaptation des plantes méditerranéennes et tropicales", Université de Montpellier, Campus International de Baillarguet TA A-108/C, 34398 Montpellier Cedex 5, France

[3] Service Evolution Biologique et Ecologie, CP160/12, Faculté des Sciences, Université Libre de Bruxelles, 50 av. F. Roosevelt, 1050 Brussels, Belgium

[4] Ecole de Foresterie et d'Ingénierie du Bois, Université Nationale d'Agriculture. BP 43, Kétou, Benin

[5] Nature+ asbl / TERRA Forest Is Life, Gembloux Agro-Bio Tech, Université de Liège, Passage des déportés 2, 5030 Gembloux, Belgium

[6] CIRAD – Research Unit "Forêts & Sociétés", Université de Montpellier, Campus International de Baillarguet TA C-105/D, F-34398 Montpellier Cedex 5, France

Published online: 02 January 2020

Introduction

Spatial genetic discontinuities have been documented in many plant species. When environmental barriers (mountain chains, large rivers, etc.) are not involved, this pattern is generally associated to historical events that have isolated populations for long periods of time before they merged again, forming a secondary contact zone (Duminil et al. 2006; Knowles and Alvarado-Serrano 2010). Recent phylogeographical studies of African rain forest plant species support the hypothesis that the forest cover fragmented under past climate changes, although additional investigations are needed to better understand when and how this happened (e.g. Lowe et al. 2000; Daïnou et al. 2014; Heuertz et al. 2014; Blatrix et al. 2017; Piñeiro et al. 2017). In particular, debates are still ongoing about the location of persistent forest fragments during glacial periods, especially in the Guineo-Congolian region where the Last Glacial Maximum is thought to have been more influential than in the other tropical regions (Anhuf et al. 2006).

A recent review on the phylogeographical structure of Guineo-Congolian forest tree species by Hardy et al. (2013) exhibited a clear North-South genetic cline—observed in eight out of nine taxa—in Lower Guinea (i.e. western part of the Central African forest block) at the height of the equatorial climatic hinge. The climatic hinge is close to the border between Cameroon and Gabon or Equatorial Guinea and represents the latitudinal area that separates the boreal and austral climatic regimes (Gonmadje et al. 2012). The North-South genetic cline may be explained by old forest fragmentations along the climatic hinge but current models that reconstruct past climates do not highlight a period of lasting dryness in that area. Alternatively, this North-South genetic divide may be explained by reduced pollen-mediated gene flow among populations located on both sides of the climatic hinge. Indeed, inversion of climatic seasons between the northern and southern hemispheres can result in a 6-month delay in flowering phenology. A shortcoming of that assumption is that the phenological inversion is generally not abrupt but gradual thus allowing a certain gene flow especially in widespread plant species. Moreover, species where gene flow is largely mediated by extensive seed dispersal are less likely to be isolated by a phenological delay. The North-South genetic divide could hide more complex patterns. Dauby et al. (2014) showed that tree populations of northern localities displayed congruent patterns of differentiation at plastid genes, while those of southern sites presented species-specific divergence patterns, suggesting a more complex vegetation history in the southern hemisphere. Blatrix et al. (2017) interpreted the abrupt North-South genetic divide found in taxa of a plant-ant symbiosis as a tension zone between populations that evolved in allopatry and became partially incompatible (incipient speciation). To complement findings from phylogeographical approaches, the influence of glacial periods on the locations of ancestral populations can be discussed using species distribution modelling tools. Basically, it consists in pairing species-specific divergence patterns with paleo and current distribution in order to identify genetic lineages that differ in ecological response to environmental variation and change (Castellanos-Morales et al. 2016; Gutiérrez-Tapia and Palma 2016; Ikeda et al. 2017). This approach has rarely been applied on African tree species for phylogeographical issues (but see Allal et al. 2011; Born et al. 2011; Daïnou et al. 2016; Faye et al. 2016; Blatrix et al. 2017).

Baillonella toxisperma (Sapotaceae; commercial name: Moabi) is a large canopy tree species widely distributed in Lower Guinea, from Nigeria to western Democratic Republic of the Congo. A fine-scale genetic structure analysis of *B. toxisperma* revealed extensive gene flow and long-distance seed dispersal (probably over a few km per generation), and the absence of genetic divide at the scale of a Gabonese forest block of about 3000 km^2 (Ndiade-Bourobou et al. 2010). Although widespread in Lower Guinean forests, *B. toxisperma* occurs at low population density (Letouzey 1985; Debroux et al. 1998). Phenological data suggest a clear inversion of flowering and fruiting periods between populations from Cameroon and Gabon (Plenderleith and Brown 2000). Hence, *B. toxisperma* is an interesting model to further investigate phylogeographical patterns of Central African trees because extensive gene flow is expected to erase more rapidly the signatures of past population fragmentation, and long-distance seed dispersal should limit the potential impact of phenological inversion on genetic differentiation across the climatic hinge. Using a set of 15 nuclear and three plastid microsatellite markers, we genotyped a few hundred individuals of *B. toxisperma* distributed in Lower Guinea (Cameroon and Gabon) to address the following questions: (1) Does the known long gene dispersal distance in *B. toxisperma* mitigate any strong genetic structure, in particular across the climatic hinge? (2) Assuming that different genetic clusters are detected, is there evidence of phylogeographical signal that would highlight long-term or repeated geographical isolation? (3) Could the analysis of current and paleo distribution modelling explain the observed spatial genetic structure?

Materials and methods

Study species and sites

Baillonella toxisperma Pierre (Sapotaceae) is a diploid large tree up to 50 m height and 300 cm of trunk diameter. It is endemic to African rain forests along the Gulf of Guinea (White 1986) and distributed from the southern part of Nigeria to the mouth of Congo river in the Democratic

Republic of the Congo (DRC). Hence, the heart of its range covers Gabon and Cameroon. Forest inventory data reported only 5–10 trees/km^2 (individuals with dbh ≥ 10 cm) in the study area (Letouzey 1985; Debroux et al. 1998). *Baillonella toxisperma* harbours hermaphrodite flowers (Debroux 1998) which are probably pollinated by insects and bats (Ndiade-Bourobou et al. 2009). Regular fruiting diameter is estimated around 70 cm dbh (Debroux 1998). Elephants and humans are probably the main seed dispersers (Gautier-Hion et al. 1985; Debroux 1998; Ndiade-Bourobou et al. 2010). Adult trees of *B. toxisperma* are exploited for the great economic value of the wood whereas local human populations use the fruits and seeds for food purposes. *B. toxisperma* is classified as a vulnerable species by IUCN (UICN 1996) and its logging is forbidden in Gabon since 2009.

We collected leaves or cambium pieces of 517 georeferenced trees in Cameroon and Gabon in order to cover the majority of the distribution range of *B. toxisperma* in these countries (Fig. 1; Table S1 in Online Resource). Some of the sampling sites are located within the postulated Quaternary forest refugia of Lower Guinea sensu Maley (1996). After collection, the sampled material was directly stored in bags containing silica gel crystals.

Genotyping of nuclear and chloroplast microsatellites and sequencing attempts

DNA was extracted following Bousquet et al. (1990) and Dolezel et al. (1989). A total of 15 nuclear microsatellite loci (nSSRs) were genotyped following Ndiade-Bourobou et al. (2010). These loci were characterized in Ndiade-Bourobou et al. (2009). Three polymorphic chloroplast microsatellite loci, *ccmp1*, *ccmp3* and *ccmp4* (cpSSRs; Weising and Gardner 1999) were genotyped on the same individuals using the protocol in Ndiade-Bourobou et al. (2010). From the initial sample of 517 trees, 303 individuals were retained for further analyses at nSSRs. Individuals were removed either due to a high rate of missing genotypes (23 individuals) or to avoid very dense sampling at particular geographic locations (191 individuals excluded through a systematic approach keeping at most one individual per km^2 in order to get a more balanced overall sample). For the cpSSRs, 232 individuals were successfully genotyped at the three loci and used for the analyses.

Polymorphism was also screened in 13 plastid non-coding introns and intergenic regions, namely, *trnS-trnG-trnG*, *rpl32-trnL*, *BD*, *B1B2*, *DT*, *SFM*, *QS*, *SR*, *trnL*, *nad1*, *trnK*, *ndhA* intron and 3'*rps16*-5'*trnK* (Demesure et al. 1995; Wakasugi et al. 1998; Grivet et al. 2001; Shaw et al. 2007). Unfortunately, no variation was observed on the tested sequences so that no further analysis was conducted.

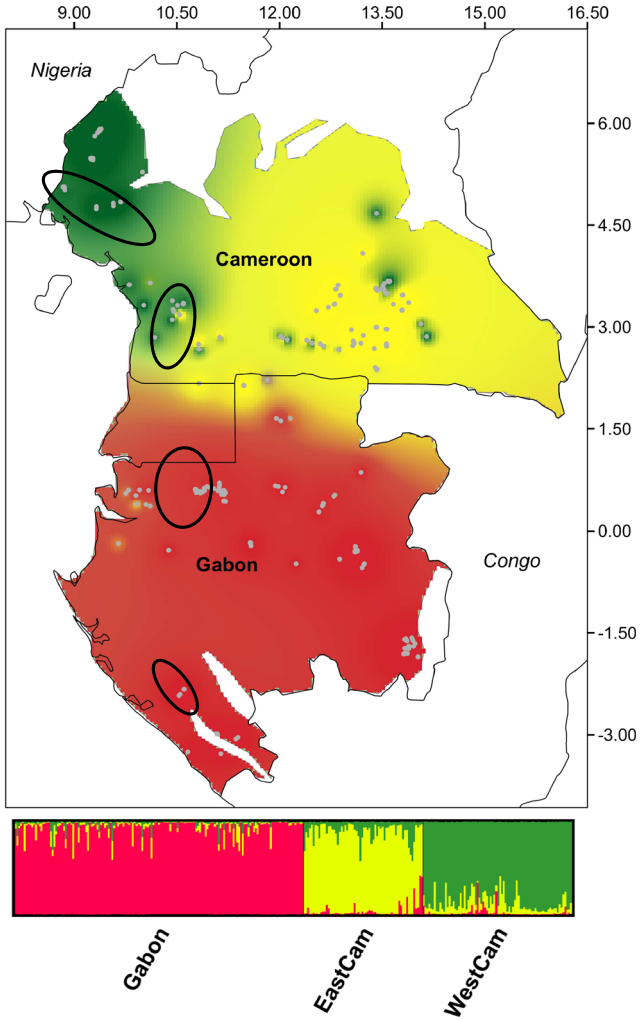

Fig. 1 Spatial range of the three genetic clusters inferred from Cameroonian and Gabonese samples of *Baillonella toxisperma* (grey points), using STRUCTURE which was highly congruent with TESS (93.7%). The figure is the interpolated map—restricted to the rain forest zone within each country—from the ancestry coefficients of the samples: for each of the three colours, the gradient increases with the value of the individual ancestry coefficient. Green: cluster WestCAM; yellow: cluster EastCAM; red: cluster GABON. Grey dots represent sampled individuals and dashed curves indicate some postulated glacial refuges according to Maley (1996). The graph below the spatial distribution represents each individual ancestry proportions for the three clusters

Detection of genetic clusters at nuclear SSR using Bayesian methods

First, we used the Bayesian approach implemented in STRUCTURE (Pritchard et al. 2000) to estimate the number K of different genetic clusters and the associated membership coefficient *q* for each individual. We used the admixture model assuming either correlated or independent allele frequencies. Ten runs were performed for each value of K_{max} which was fixed from 1 to 10. Each run consisted in 100,000 sweeps preceded by a 10,000-burn-in. The most likely number of

genetic clusters according to STRUCTURE, K_{Str}, was estimated by visualizing the evolution of the log-likelihood of the data as a function of $K_{max,}$ as recommended by the program manual. With our data, this approach clearly tended to provide more reliable estimates of the number of clusters than those from Evanno et al. (2005) and Verity and Nichols (2016) (Fig. S1 in Online Resource; see also Janes et al. 2017). Second, because the spatial proximity among individuals may explain population genetic structure, we also used the algorithm implemented in TESS (Chen et al. 2007) where spatial autocorrelation among individuals can be controlled through the interaction parameter ψ. We assumed $2 \leq K_{max} \leq 10$ under the admixture model with a burn-in period of 10,000 sweeps and a total of 100,000 sweeps. Ten runs were conducted for each value of K_{max} and for each of the following values of ψ: 0.0 (no consideration of the spatial location of individuals) and 0.6 (more chance that two neighbourhood samples belong to a same cluster). The optimal number of genetic clusters from TESS analyses, K_{Tess}, was determined through the curve of the deviance information criterion (DIC) plotted against K_{max} as proposed by Chen et al. (2007). For both approaches, we used CLUMPP (Jakobsson and Rosenberg 2007) to average the individuals' assignment probabilities to each cluster from the 50% best runs (identified through the average log-likelihood). Each individual was assigned to a cluster on the basis of a minimum membership probability $q \geq 50\%$. To construct a continuous map of the range of each inferred genetic cluster, we spatially modelled q values using the interpolation tool implemented in Quantum GIS 2.18 (QGIS Development Team 2018) using the inverse square distance-weighting method (Lam 1983) with a distance coefficient $P = 2$. The ancestry proportions of each individual for the detected genetic clusters were illustrated using DISTRUCT 1.1 (Rosenberg 2004).

Detection of genetic structure at plastid SSR

Each combination of the three alleles belonging to the three microsatellite loci was considered as a haplotype. Genetic distance was estimated using the median-joining method (Saitou and Nei 1987). A haplotype network was constructed using NETWORK 4.6 and based on all the shortest trees: this procedure deletes superfluous links and is recommended when homoplasy is important (Bandelt et al. 1999). Because there was a weak spatial structure at chloroplast SSR (see results), the genetic clusters inferred through nuclear SSR were also utilized as sampling units to assess diversity and divergence patterns at chloroplast SSR.

Diversity and differentiation parameters

We computed for each cluster and for each genome (nuclear and plastid DNA) the effective number of alleles *NAe*, the allelic richness *AR* (number of alleles in a random subsample of constant size for the different populations) and the expected heterozygosity H_E. Genetic differentiation based on allele identity at nuclear SSR was estimated through the statistic F_{ST} whereas G_{ST} was computed to estimate genetic divergence at chloroplast SSR. To verify whether long-term isolation (phylogeographical signal) and accumulation of stepwise mutations may explain the observed divergence between clusters, we also computed R_{ST} for nuclear SSR based on allele size and N_{ST} for chloroplast SSR based on the minimum number of mutations separating haplotypes (Pons and Petit 1996). R_{ST} and N_{ST} are expected to be larger than their respective allele identity-based relatives (F_{ST} and G_{ST}) when a phylogeographical signal occurs (Hardy et al. 2003). The latter was tested using permutation tests: R_{ST} and N_{ST} were compared to the distributions of $R_{STpermuted}$ ($\approx F_{ST}$) and $N_{STpermuted}$ ($\approx G_{ST}$) after 10,000 permutations of allele size among allelic states or of genetic distances among haplotypes, respectively. The program SPAGEDI 1.5 (Hardy and Vekemans 2002) was used for all these computations and tests.

In order to evaluate the inbreeding coefficient in the detected clusters (299 assigned individuals in total; Table 1) and take into account possible occurrence of null alleles at some nuclear loci (Ndiade-Bourobou et al. 2009), we employed the program INEST (Chybicki and Burczyk 2009; see also Campagne et al. 2012 who compared it to another well-known similar program). INEST provides estimates of inbreeding coefficients across loci by considering the potential occurrence of both null alleles and genotyping failures. For each genetic cluster, the mean inbreeding coefficient F_{is} and 95% highest posterior density interval, 95%HPD, were provided. In case of significant inbreeding coefficient, we were able to assess the relative importance of null alleles in the data set by comparing the deviance information criterion (DIC) from three models: (i) presence of both null alleles and inbreeding, (ii) presence of only inbreeding and (iii) presence of only null alleles. We set the number of cycles to 300,000 and the burning period to 30,000.

SPAGEDI 1.5 was also used to characterize isolation-by-distance (IBD) within each cluster using the nuclear dataset. Pairwise kinship coefficients F_{ij} between individuals (i and j) were averaged over a set of distance intervals and regressed against the spatial distance, providing the regression slope b_{ld} (from F_{ij} on the logarithm of spatial distance); this allowed the computation of the S_p statistic which quantifies spatial genetic structure (Vekemans and Hardy 2004).

Bioclimatic data and ecological niche modelling

To model the potential distribution of *Baillonella toxisperma*, we used (i) the geographical coordinates of the 303 individuals used for genotyping with nSSRs and (ii) occurrence data drawn from RAINBIO database, a

Table 1 Estimates of genetic diversity and inbreeding coefficient from the three genetic clusters inferred from the nuclear SSR dataset

nSSR-based clusters	Nuclear microsatellites					Chloroplast microsatellites			
	Genetic diversity parameters and inbreeding coefficients								
	N	*NAe*	*AR*	H_E	F_{is} (95%HPD)	*N*	*NAe*	AR_{cp}	*Hs*
WestCAM	80	2.94	7.32	0.521	0.041 (0.001–0.079)	52	4.13	7.00	0.759
EastCAM	64	3.01	7.10	0.551	0.024 (0.000–0.057)	53	4.19	6.96	0.762
GABON	155	2.68	7.77	0.557	0.038 (0.011–0.069)	123	4.28	6.87	0.767
	Genetic differentiation parameters: F_{ST} *(above the diagonal) and* R_{ST} *for nuclear SSR data, and* G_{ST} *(above the diagonal) and* N_{ST} *for chloroplast SSR data*								
$R_{ST}F_{ST}$	WestCAM	EastCAM	GABON			$N_{ST}\backslash G_{ST}$	WestCAM	EastCAM	GABON
WestCAM		0.078	0.185					0.112	0.063
EastCAM	0.004 ns		0.142				0.019 ns		0.167
GABON	0.381***	0.275***					0.051 ns	0.019 ns	

N, number of individuals; *NAe*, effective number of alleles; *AR*, allelic richness; H_E, gene diversity corrected for sample size; H_S, gene diversity with unordered alleles; F_{is} (95%HPD), population inbreeding coefficient and 95% highest posterior density interval

The genetic differentiation parameters F_{ST} (above the diagonal) and R_{ST} were computed for the nuclear SSR data set, whereas G_{ST} (above the diagonal) and N_{ST} were computed for the chloroplast SSR data set. The significance of permutation tests ($R_{ST} > R_{STpermuted}$ and $N_{ST} > N_{STpermuted}$) was indicated as follows: ns = not significant, * = $P < 0.05$, ** = $P < 0.01$ and *** = $P < 0.001$

standardized database of tropical African vascular plants occurrences (Dauby et al. 2016). In total, 608 occurrence records from Cameroon, Equatorial Guinea, Gabon and Republic of Congo were obtained. This dataset was cleaned according to Vinceti et al. (2013) and using the "systematic sampling" method as in Fourcade et al. (2014), selecting one occurrence per area of 20 km by 20 km, leaving a total of 132 occurrence records. To model the potential distribution of *B. toxisperma* in the present, these 132 occurrence records were used with 19 bioclimatic variables from the WorldClim 1.4 dataset (Hijmans et al. 2005) at 2.5 min (~ 4.5 km at the equator) spatial resolution (http://www.worldclim.org/current). To avoid problems of multicollinearity that may violate statistical assumptions and alter model predictions (Heikkinen and Luoto 2006), we removed some correlated variables as follows. Within groups of highly cross-related variables (absolute value of the correlation coefficient > 0.8), variables were ranked according to the number of correlated variables (Fourcade et al. 2014) and we retained only the variable with the highest rank. If two climatic variables had the same rank, the sum of the absolute correlation values was used to choose the most representative variable. We finally retained the 10 following climatic variables: annual temperature range (BIO7), mean temperature of driest quarter (BIO9), max temperature of warmest month (BIO5), mean diurnal range (BIO2; mean of monthly (max temp - min temp)), mean temperature of wettest quarter (BIO8), precipitation of wettest month (BIO13), precipitation of driest quarter (BIO17), precipitation of driest month (BIO14), precipitation of coldest quarter (BIO19) and precipitation of warmest quarter (BIO18). The R package *mopa* (Iturbide et al. 2015) was used to generate 500 pseudo-absence at random, keeping a 0.415° (~ 46 km) exclusion buffer around known presence localities. We followed the approach of Araújo and New (2007) and Riul et al. (2013) based on an ensemble of predictions derived from multiple modelling algorithms because it reduces the uncertainties resulting from a single-modelling method (Thuiller et al. 2009; Singer et al. 2016). Four algorithms were used using the R package *sdm* (Naimi and Araújo 2016): (i) Generalized Linear Model (GLM), (ii) Generalized Boosted Regression Modelling (GBM), (iii) Mixture Discriminant Analysis (MDA) and (iv) Flexible Discriminant Analysis (FDA). For each, we performed 10 runs using 70% of the data for training and 30% for testing. The evaluation of the model performance was based on threshold-independent statistics (AUC and COR; Fielding and Bell 1997), and threshold-dependent statistics (true skill statistic or TSS; Allouche et al. 2006). We created the ensemble models by considering a weighted averaging consisting in using only high-quality models based on the TSS and AUC evaluation metrics thresholds (TSS > 0.65; AUC > 0.8).

In order to model the potential distribution of *Baillonella toxisperma* in the LGM, the results of ensemble models were projected using the median values of the three downscaled General Circulation Models for the LGM (about 22,000 years ago), available from the WorldClim website (http://www.worldclim.org/paleo-climate1) at 2.5 min spatial resolution.

The R package *biomod2* (Thuiller et al. 2014) allowed to convert the raster layers resulting from the current and paleo (LGM) distribution modelling into binary rasters of suitable vs. unsuitable habitat by using the threshold at which the true skill statistic was maximized.

Results

Detection of spatial genetic cline

Nuclear DNA-based structuring First, the two clustering programs were highly congruent at $K_{max} = 2$ clusters, distinguishing a northern (Cameroon) and a southern (Gabon) cluster with a congruence rate > 97%, whatever the allele frequency model in STRUCTURE and the value of the interaction parameter in TESS. On the basis of the admixture model implemented in STRUCTURE, the independent and correlated allele frequency models suggested three and four genetic clusters, respectively (Fig. S2 in Online Resource). The program TESS consistently suggested three genetic clusters (Fig. S2 in Online Resource) based on the admixture model and whatever the value of the interaction parameter (values of DIC at $K_{max} = 3$ were statistically similar for $\psi =$ 0.0 and 0.6; Mann-Whitney test: $W = 40$, $p = 0.4813$). The 3-cluster scenario, which subdivides the northern cluster found at $K_{max} = 2$ into a western and an eastern cluster, was supported by additional observations: (i) for $K_{max} = 3$, TESS and STRUCTURE were highly congruent provided that the independent allele frequency model was chosen in STRUCTURE: 93.7% (vs. 51.2% of congruence with STRUCTURE's model of correlated allele frequencies); (ii) most samples were assigned to a genetic cluster at $q > 0.7$, although there was evidence of substantial gene flow between clusters (see the graph of individual ancestry proportion in Fig. 1); (iii) for $K_{max} \geq 4$, no individual was assigned by TESS into a fourth cluster even at $q \geq 0.50$ (spurious cluster in the sense of Guillot et al. 2005 and Puechmaille 2016).

There was a clear North-South divide close to the boundary between Cameroon and Gabon. The clusters will be called hereafter WestCAM, EastCAM and GABON, to refer to their respective distributions, as highlighted by the interpolations of cluster membership values (Fig. 1).

Plastid DNA-based structuring The most diverse chloroplast microsatellite marker was *ccmp1* which exhibited four alleles, whereas *ccmp3* and *ccmp4* showed each two alleles. Overall, the three chloroplast markers defined 10 haplotypes over 232 individuals in the study zone. Three haplotypes exhibited extremely low frequencies (found in 1 to 2 individuals: *h2*, *h3* and *h4*; Fig. S3 in Online Resource). Excluding these three haplotypes resulted in a certain spatial pattern in Cameroon with the western zone dominated by haplotype *h7*, while the eastern part of the country was characterized by four major haplotypes: *h5*, *h6*, *h8* and *h10*. It is remarkable that the location of the geographic border between the eastern and the western genetic groups found in Cameroon from the plastid genome was consistent with the one inferred from the nuclear microsatellites (Figs. 1 and 2). No clear geographic substructure was observed in Gabon (Fig. 2).

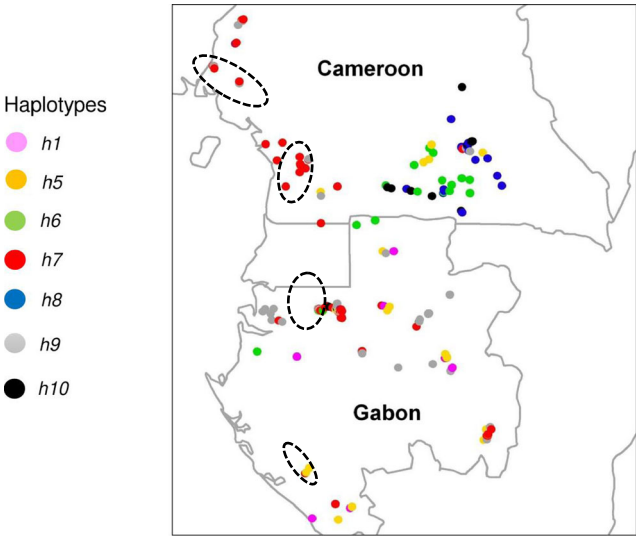

Fig. 2 The geographic distribution of haplotypes derived from three chloroplast microsatellite markers, *ccmp1*, *ccmp3* and *ccmp4*, in samples of *Baillonella toxisperma*. The dashed curves indicate some postulated glacial refuges according to Maley (1996)

Diversity, inbreeding coefficient and differentiation patterns at cluster level

For nSSRs, genetic diversity parameters did not substantially vary among clusters (Table 1). Pairwise F_{ST} was lowest between WestCAM and EastCAM ($F_{ST} = 0.078$) and highest between the two countries ($F_{ST} = 0.142$ to 0.185), and R_{ST} was significantly higher than $R_{STpermuted}$ ($\approx F_{ST}$) only for these between-country pairs of clusters ($R_{ST} = 0.27$ to 0.38; Table 1) indicating a phylogeographical signal. With the chloroplast microsatellites, genetic diversity did not substantially vary among clusters but the between-country differentiation was here globally low ($G_{ST} = 0.063$ to 0.167) and no phylogeographic signal was detected ($N_{ST} < G_{ST}$; Table 1).

F_{is} was significantly different from zero in the genetic clusters WestCam and Gabon although the 95% highest posterior density interval never exceeded 0.08 (Table 1). For these two genetic clusters, the values of the DIC was always the lowest for the INEST model that assumes both null alleles and inbreeding (Table S2 in Online Resource). DIC was also lower for the model assuming just null alleles than the one assuming just inbreeding, indicating that null alleles are the likely cause of $F_{is} > 0$ in these clusters (Table S2 in Online Resource).

Isolation-by-distance within clusters

Isolation-by-distance (IBD) was significant in each genetic cluster. S_p ranged in 0.005 to 0.018 with the highest values obtained from the Cameroonian clusters: 0.012 in WestCAM, 0.018 in EastCAM and 0.005 for the unique Gabonese cluster.

Modelling historical and current ecological niches

Niche modelling was carried out to detect putative changes in the spatial range of *B. toxisperma* since the LGM (Fig. 3). The ensemble model displayed good model performance (mean AUC = 0.89, TSS = 0.69) and the present potential distribution coincides approximately with the humid forest zone of Lower Guinea, although some occurrence points in eastern Cameroon and south-eastern Gabon occur outside the predicted range (Fig. 3). The potential distribution of *B. toxisperma* is estimated to have expanded by about 74.1% from the LGM to the present (Fig. 3). The LGM reconstruction predicted a suitable area located in central East-Gabon and a small suitable area at the border between Cameroon and the Republic of Congo, at the margin of the current distribution range of the species (Fig. 3).

Discussion

We hypothesized that extensive gene dispersal distance that has been previously demonstrated in the low-density tree species *Baillonella toxisperma* (Ndiade-Bourobou et al. 2010) could ensure sufficient gene flow to dilute any genetic structure generated during Pleistocene forest fragmentations. At most, we were expecting a very weak spatial genetic structure. However, nuclear microsatellite markers did not support our hypothesis since a clear genetic discontinuity was detected between the northern and southern part of the species distribution range, across 1.5° to 2° N latitude. In addition, divergence time was long enough for stepwise mutations to accumulate and generate a phylogeographical signal across this North-South cline. Each of the nuclear markers used showed a F_{ST} significantly different from zero (5000 individuals

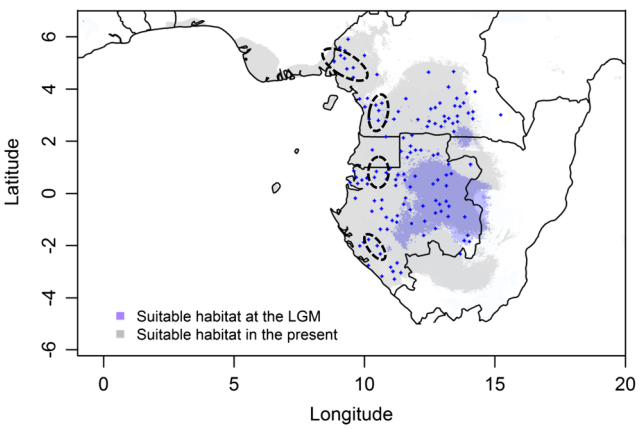

Fig. 3 Mixed plot of the suitable habitat of *B. toxisperma* under the current and LGM climate conditions. Medium purple area represents suitable habitat during the LGM and the grey one indicates the current potential suitable habitat. The blue dots represent the occurrence records used for the current and paleo distribution modelling; the dashed curves indicate some postulated glacial refuges according to Maley (1996)

permutation) when we considered the 2-cluster scenario (results not shown). Whereas that latitudinal divide was well confirmed, sub-clusters were confirmed only in Cameroon from both genomes. The existence of sub-clusters in Gabon as suggested by the correlated allele frequency model in STRUCTURE remains questionable: (i) STRUCTURE and TESS were not congruent in identifying two Gabonese clusters; (ii) the plastid microsatellites did not exhibit any clear spatial structuring in Gabon. The unclear genetic structure in Gabon could be explained by a pattern of isolation-by-distance in these populations, that complicates clustering inference (François and Durand 2010).

Possible origins of the North-South genetic divide in Lower Guinea

Our niche modelling of *Baillonella toxisperma* populations exhibited two populations located in the studied countries during the LGM. As this just partially fitted the current genetic clustering scenario and despite evidence of reduction of forest cover during the LGM due to significant drops in temperature and rainfalls (Maley 1997; Anhuf et al. 2006), the role of the LGM in shaping the genetic structure of *B. toxisperma* populations seems limited. The historical geographic isolation of the fragmented populations may be much more ancient than hypothesized before. The first major divide separating northern and southern populations of Lower Guinea is congruent with recent findings from other Central African tree species (e.g. Daïnou et al. 2010; Dauby et al. 2010; Duminil et al. 2010; Debout et al. 2011; Daïnou et al. 2014; Blatrix et al. 2017; Piñeiro et al. 2017) although the limit between northern and southern genetic clusters can range from 0.5° and 2.5° North according to the species (Hardy et al. 2013). This outcome suggests that surviving populations of *B. toxisperma* have been isolated long enough in each zone during the Pleistocene, and the current distribution results from their expansion since the end of the LGM.

In the particular case of *B. toxisperma*, the genetic divergence between the northern and southern clusters might have started much before the last glacial period. Indeed, detecting a phylogeographical signal with microsatellites ($R_{ST} > F_{ST}$) indicates that stepwise mutations have contributed to genetic differentiation in at least some of the microsatellite loci. This requires that (i) the (historical) migration rate is smaller or similar to the mutation rate ($m \leq \mu$), and (ii) the number of generations since divergence is larger than or similar to the reciprocal of the mutation rate ($t \geq 1/\mu$; Hardy et al. 2003). Thus, considering a mutation rate of $\mu = 10^{-3}$ per generation for nuclear microsatellites (Estoup and Angers 1998) and a mean generation time assumed to equal the average *B. toxisperma* flowering age (diameter at breast height, or $dbh = 85$ cm equivalent to ≈280 years; Lee White, unpublished), the Cameroonian genetic clusters would have

diverged from the Gabonese cluster since at least 280,000 BP. Much uncertainty remains on the true divergence date but it seems to predate largely the previous glacial period (c. 12,000–70,000 BP).

If northern and southern *B. toxisperma* populations were isolated during the LGM as suggested by our species distribution modelling, we expect that these populations were connected again since the Holocene when the rain forest presumably reached its maximal extension (Anhuf et al. 2006). Given the long generation time of the species, we could assume that about 35–40 generations have elapsed since the beginning of the Holocene, which may well be too short for an efficient admixture of the populations. Indeed, even if we assume seed dispersal capacities of a few kilometres (Ndiade-Bourobou et al. 2010), each population might have moved by no more than 100–200 km towards each other, while they were presumably separated by c. 300 km during the LGM (Fig. 3). These are just estimates but they highlight that interglacial periods such as the Holocene, which usually lasted c. 20 ky during the late Pleistocene (glacial periods were cyclical leading to regular phases of forest expansion and fragmentation), were probably too short for allowing a good genetic homogenization of isolated long-living tree populations, even if they present good dispersal abilities. Low admixture is supported by a recent molecular dating in *Greenwayodendron suaveolens* showing that adjacent genetic clusters of a characteristic tree of mature African forests have sometimes diverged for millions of years (Migliore et al. 2018). Hence, the North-South genetic divide might result from the accumulation of genetic differentiation among forest fragments during glacial periods, differentiation that was not substantially counterbalanced by the species-specific extensive gene flow during interglacial periods.

Nevertheless, the reason gene flow may have been broken or limited between populations located apart from the climatic hinge is controversial because neither pollen records nor past climate modelling indicates a particular forest fragmentation nearby the equator (Dupont et al. 2000; Levinsky et al. 2013). An alternative explanation for the occurrence of the often found North-South genetic discontinuity in Lower Guinea is related to flowering phenology, because this discontinuity matches approximately the position of the equatorial climatic hinge that causes a 6-month inversion in the precipitation regime on both sides of the equator (Hardy et al. 2013). In contrast to some tree species that display a gradual change of phenology across this climatic hinge (e.g. *Milicia excelsa*; K. Daïnou pers. obs.), *B. toxisperma* displays a quite sharp change in phenology when moving from Cameroon to Gabon: 6 months separate the fruiting times in these regions (Plenderleith and Brown 2000). This is susceptible to limit pollination across the climatic hinge. However, this should not impact seed dispersal which is known to be very extensive in *B. toxisperma* (Ndiade-Bourobou et al. 2010), so that we think the phenological shift hypothesis poorly explains the substantial North-South genetic divide in these populations. However, we cannot exclude that the sharp phenological delay across the climatic hinge might result from the restricted genetic introgression between northern and southern *B. toxisperma* genetic clusters, possibly because these had not time to intermingle enough since they came into contact.

We did not observe a clear North-South genetic divide from the plastid microsatellite markers although some haplotypes are more represented on one side of the equator and despite the fact that the markers used exhibited substantial diversity. In *Baillonella toxisperma*, chloroplasts are maternally inherited and dispersed with the seeds mainly through movements of elephants and humans (Debroux 1998; Ndiade-Bourobou et al. 2010). Because its main seed dispersers can move long distances and in various directions, this may be an explanation of the absence of the latitudinal genetic cline from the plastid markers.

The East-West genetic gradient confirms the low impact of long-distance seed dispersal

Both types of markers (nuclear and plastid microsatellites) suggested the same location of the genetic divide in Cameroon. The discontinuity proposed between eastern and western populations was also reported at various intensities in other tree species such as *Milicia excelsa* (Daïnou et al. 2010), *Distemonanthus benthamianus* (Debout et al. 2011; Demenou et al. 2016), *Greenwayodendron suaveolens* and *Scorodophleus zenkeri* (Piñeiro et al. 2017), though not always at the same longitude (Hardy et al. 2013). The Cameroonian East-West genetic structure formed a gradient rather than a sharp divide. Under the hypothesis that this genetic structure also results from past population fragmentation, our results suggested that the East-West gene flow barrier was not as old and/or was partial compared to the North-South divide.

Such a finding in Cameroon is not surprising and confirms hypotheses of Dauby et al. (2014) who argued that populations of plant species at the northern side of the equator may have been more affected by past environment changes than their relatives located southwards (see also Dupont et al. 2000 who suggested that dense humid forests in Gabon may have experienced less changes than in Eastern Cameroon). In other words, phylogeographical signals would be more evident in areas at the north of the climatic hinge, where forests in the hollow of the Gulf of Guinea have experienced less environmental changes than the inland forests. Overall, the detection of two clear genetic clusters in Cameroon along with the existence of isolation-by-distance pattern tends to prove that seed dispersal at long distance in *B. toxisperma* may have less impact in obscuring long-lasting effects of population fragmentation.

Conclusive remarks

The present study points out that several factors such as forest fragmentation in relation to historical climate oscillations, mating and efficiency of the pollen and seeds vectors may strongly affect the trajectory of diversity in space and time in tree species populations. Whereas it confirmed the existence of a North-South genetic cline in Central Africa, which seems of ancient origin, the causes of a limited gene flow at both sides of the equator are still difficult to explain. The detection of genetic clusters in *B. toxisperma* range is also a criterion to take into account when delimiting regions of provenances for sustainable forest management and conservation purposes, such as sub-regional seed bank collections, forest plantations or species reintroduction programs.

Acknowledgements We are grateful to persons and institutions that facilitated the leaf collection: Christelle Gonmadje, Yves Issembe, Edouard Nzengue, Hugues Moussavou, Raoul Niangadouma, Myriam Heuertz, Gembloux Agro-Bio Tech (University of Liege), Nature+ asbl, Institut des Recherches Agronomiques et Forestières (IRAF/CENAREST, Gabon) and the logging company Rougier Gabon.

Data archiving statement Nuclear and plastid microsatellites data are submitted into DRYAD repository.

References

Allal F, Sanou H, Millet L, Vaillant A, Camus-Kulandaivelu L, Logossa ZA, Lefèvre F, Bouvet JM (2011) Past climate changes explain the phylogeography of *Vitellaria paradoxa* over Africa. Heredity 107: 174–186

Allouche O, Tsoar A, Kadmon R (2006) Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). J Appl Ecol 43:1223–1232

Anhuf D, Ledru MP, Behling H, Da Cruz FW, Cordeiro RC, Van der Hammen T, Karmann I, Marengo JA, De Oliveira PE, Pessenda L, Siffedine A (2006) Paleo-environmental change in Amazonian and African rainforest during the LGM. Palaeogeogr Palaeocl 239:510–527

Araújo MB, New M (2007) Ensemble forecasting of species distributions. Trends Ecol Evol 22:42–47

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

Blatrix R, Peccoud J, Born C, Piatscheck F, Benoit L, Sauve M, Djiéto-Lordon C, Atteke C, Wieringa JJ, Harris DJ, McKey D (2017) Comparative analysis of spatial genetic structure in an ant–plant symbiosis reveals a tension zone and highlights speciation processes in tropical Africa. J Biogeogr 44:1856–1868

Born C, Alvarez N, McKey D, Ossari S, Wickings EJ, Hossaert-McKey M, Chevallier M-H (2011) Insights into the biogeographical history of the Lower Guinea Forest domain: evidence for the role of refugia in the intraspecific differentiation of *Aucoumea klaineana*. Mol Ecol 20:131–142

Bousquet J, Simon L, Lalonde M (1990) DNA amplification from vegetative and sexual tissues of trees using polymerase chain reaction. Can J For Res 20:254–257

Campagne P, Smouse PE, Varouchas G, Silvain JF, Leru B (2012) Comparing the van Oosterhout and Chybicki-Burczyk methods of estimating null allele frequencies for inbred populations. Mol Ecol Resour 12:975–982

Castellanos-Morales G, Gutiérrez-Guerrero YT, Gámez N, Eguiarte LE (2016) Use of molecular and environmental analyses for integrated in situ and ex situ conservation: the case of the Mexican prairie dog. Biol Conserv 204:284–295

Chen C, Durand E, Forbes F, François O (2007) Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. Mol Ecol Notes 7:747–756

Chybicki IJ, Burczyk J (2009) Simultaneous estimation of null alleles and inbreeding coefficients. J Heredity 100:106–113

Daïnou K, Bizoux JP, Doucet J-L, Mahy G, Hardy OJ, Heuertz M (2010) Forest refugia revisited: nSSRs and cpDNA sequences support historical isolation in a wide-spread African tree with high colonization capacity, *Milicia excelsa* (Moraceae). Mol Ecol 19:4462–4477

Daïnou K, Mahy G, Duminil J, Dick C, Doucet J-L, Donkpégan ASL, Pluijgers M, Sinsin B, Lejeune P, Hardy OJ (2014) Speciation slowing down in widespread and long-living tree taxa: insights from the tropical timber tree genus *Milicia* (Moraceae). Heredity 113:74–85

Daïnou K, Blanc-Jolivet C, Degen B, Kimani P, Ndiade-Bourobou D, Donkpegan ASL, Tosso F, Kaymak E, Bourland N, Doucet J-L, Hardy OJ (2016) Revealing hidden species diversity in closely related species using nuclear SNPs, SSRs and DNA sequences–a case study in the tree genus *Milicia*. BMC Evol Biol 16:259

Dauby G, Duminil J, Heuertz M, Hardy OJ (2010) Chloroplast DNA polymorphism and phylogeography of a Central African tree species widespread in mature rainforests: *Greenwayodendron suaveolens* (Annonaceae). Trop Plant Biol 3:4–13

Dauby G, Duminil J, Heuertz M, Koffi G, Stevart T, Hardy OJ (2014) Congruent phylogeographical patterns of eight tree species in Atlantic Central Africa provide insights into the past dynamics of forest cover. Mol Ecol 23:2299–2312

Dauby G, Zaiss R, Blach-Overgaard A, Catarino L, Damen T, Deblauwe V, Dessein S, Dransfield J, Droissart V, Duarte MC, Engledow H (2016) RAINBIO: a mega-database of tropical African vascular plants distributions. PhytoKeys 74:1–18

Debout GD, Doucet JL, Hardy OJ (2011) Population history and gene dispersal inferred from spatial genetic structure of a Central African timber tree, *Distemonanthus benthamianus* (Caesalpinioideae). Heredity 106:88–99

Debroux L (1998) L'aménagement des forêts tropicales fondé sur la gestion des populations d'arbres:l'exemple du moabi (*Baillonella toxisperma* Pierre) dans la forêt du Dja, Cameroun. Thèse de doctorat. Faculté universitaire des sciences agronomiques de Gembloux

Debroux L, Delvingt W, Mbolo M, Amougou A (1998) Régénération du moabi et du mukulungu au Cameroun: perspectives pour l'aménagement. Bois For Trop 255:5–17

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

Demesure B, Sodzi N, Petit R (1995) A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. Mol Ecol 4:129–134

Dolezel J, Binarova P, Lucretti S (1989) Flow cytometry was used to analyse the DNA content of nuclei isolated from plant tissues and from callus and cell suspension cultures in vitro. Biol Plantarum 62: 113–120

Duminil J, Caron H, Scotti I, Cazal SO, Petit RJ (2006) Blind population genetics survey of tropical rainforest trees. Mol Ecol 15:3505–3513

Duminil J, Heuertz M, Doucet JL, Bourland N, Cruaud C, Gavory F, Doumenge C, Navascués M, Hardy OJ (2010) CpDNA-based species identification and phylogeography: application to African tropical tree species. Mol Ecol 19:5469–5483

Dupont LM, Jahns S, Marret F, Ning S (2000) Vegetation change in equatorial West Africa: time-slices for the last 150 ka. Palaeogeogr Palaeocl 155:95–122

Estoup A, Angers B (1998) Microsatellites and minisatellites for molecular ecology: theoretical and empirical considerations. In: Carvalho GR (ed) Advances in molecular ecology. IOS Press, Amsterdam

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620

Faye A, Deblauwe V, Mariac C, Richard D, Sonké B, Vigouroux Y, Couvreur TLP (2016) Phylogeography of the genus *Podococcus* (Palmae/Arecaceae) in Central African rain forests: climate stability predicts unique genetic diversity. Mol Phylogenet Evol 105:126–138

Fielding AH, Bell JF (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. Environ Conserv 24:38–49

Fourcade Y, Engler JO, Rödder D, Secondi J (2014) Mapping species distributions with MAXENT using a geographically biased sample of presence data: a performance assessment of methods for correcting sampling bias. PLoS One 9:e97122

François O, Durand E (2010) Spatially explicit Bayesian clustering models in population genetics. Mol Ecol Resour 10:773–784

Gautier-Hion A, Duplantier J-M, Quris R, Feer F, Sourd C, Decoux J-P, Dubost G, Emmons L, Erard C, Hecketsweiler P, Moungazi A (1985) Fruit characters as a basis of fruit choice and seed dispersal in a tropical forest vertebrate community. Oecologia 65:324–337

Guillot G, Estoup A, Mortier F, Cosson JF (2005) A Spatial Statistical Model for Landscape Genetics. Genetics 170 (3):1261–1280

Gonmadje CF, Doumenge C, Sunderland TCH, Balinga MPB, Sonké B (2012) Analyse phytogéographiques des forêts d'Afrique Centrale: le cas du massif de Ngovayang (Cameroun). Plant Ecol Evol 145: 152–164

Grivet D, Heinze B, Vendramin G, Petit R (2001) Genome walking with consensus primers: application to the large single copy region of chloroplast DNA. Mol Ecol Notes 1:345–349

Gutiérrez-Tapia P, Palma RE (2016) Integrating phylogeography and species distribution models:cryptic distributional responses to past climate change in an endemic rodent from the Central Chile hotspot. Divers Distrib 22:638–650

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

Hardy OJ, Charbonnel N, Fréville H, Heuertz M (2003) Microsatellite allele sizes: a simple test to assess their significance on genetic differentiation. Genetics 163:1467–1482

Hardy OJ, Born C, Budde K, Daïnou K, Dauby G, Duminil J, Ewédjé EEB, Gomez C, Heuertz M, Koffi GK, Lowe AJ, Micheneau C, Ndiade-Bourobou D, Piñeiro R, Poncet V (2013) Comparative phylogeography of African rain forest trees: a review of genetic signatures of vegetation history in the Guineo-Congolian region. Compt Rendus Geosci 345:284–296

Heikkinen RK, Luoto M (2006) Methods and uncertainties in bioclimatic envelope modelling under climate change. Prog Phys Geogr 6:751–777

Heuertz M, Duminil J, Dauby G, Savolainen V, Hardy OJ (2014) Comparative phylogeography in rainforest trees from Lower Guinea, Africa. PLoS One 9:e84307

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. Int J Climatol 25:1965–1978

Ikeda DH, Max TL, Allan GJ, Lau MK, Shuster SM, Whitham TG (2017) Genetically informed ecological niche models improve climate change predictions. Glob Chang Biol 23:164–176

Iturbide M, Bedia J, Herrera S, del Hierro PO, Gutiérrez JM (2015) A framework for species distribution modelling with improved pseudo-absence generation. Ecol Model 312:166–174

Jakobsson M, Rosenberg NA (2007) Clumpp: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23:1801–1806

Janes JK, Miller JM, Dupuis JR, Malenfant RM, Gorrell JC, Cullingham CI, Andrew RL (2017) The K = 2 conundrum. Mol Ecol 26:3594–3602

Knowles L, Alvarado-Serrano DF (2010) Exploring the population genetic consequences of the colonization process with spatio-temporally explicit models: insights from coupled ecological, demographic and genetic models in montane grasshoppers. Mol Ecol 19: 3727–3745

Lam NSN (1983) Spatial interpolation methods: a review. Am Cartographer 10:129–150

Letouzey R (1985) Notice de la carte phytogeographique du Cameroun au 1:500,000. Institut de la carte internationale de la végétation. Toulouse, France

Levinsky I, Araújo MB, Nogués-Bravo D, Haywood AM, Valdes PJ, Rahbek C (2013) Climate envelope models suggest spatio-temporal co-occurrence of refugia of African birds and mammals. Glob Ecol Biogeogr 22:351–363

Lowe A, Gillies A, Wilson J, Dawson I (2000) Conservation genetics of bush mango from Central/West Africa: implications from random amplified polymorphic DNA analysis. Mol Ecol 9:831–841

Maley J (1996) The African rain forest - main characteristics of changes in vegetation and climate from the Upper Cretaceous to the Quaternary. In: Alexander IJ, Swaine MD, Watling R (eds) Essays on the ecology of the Guinea-Congo rain forest. The Royal Society of Edinburgh, Edinburgh, pp 31–73

Maley J (1997) Middle to Late Holocene changes in tropical Africa and other continents: paleomonsoon and sea surface temperature variations. In: Dalfes HN, Kukla G, Weiss H (eds) Third millennium BC climate change and Old World collapse. Springer, Berlin, pp 611–640

Migliore J, Kaymak E, Mariac C, Couvreur TL, Lissambou BJ, Piñeiro R, Hardy OJ (2018) Pre-Pleistocene origin of phylogeographical breaks in African rain forest trees: new insights from *Greenwayodendron* (Annonaceae) phylogenomics. J Biogeogr. https://doi.org/10.1111/jbi.13476

Naimi B, Araújo MB (2016) Sdm: a reproducible and extensible R platform for species distribution modelling. Ecography 39:368–375

Ndiade-Bourobou D, Vaillant A, Favreau B, Gayrin E, Bouvet J-M (2009) Isolation and characterization of 15 nuclear microsatellite markers for *Baillonella toxisperma* Pierre (Sapotaceae), a low-density tree species of Central Africa. Mol Ecol Resour 9:1135–1138

Ndiade-Bourobou D, Hardy OJ, Favreau B, Moussavou H, Nzengue E, Mignot A, Bouvet J-M (2010) Long-distance seed and pollen dispersal inferred from spatial genetic structure in the very low-density rainforest tree, *Baillonella toxisperma* Pierre, in Central Africa. Mol Ecol 19:4949–4962

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

Plenderleith K, Brown N (2000) *Baillonella toxisperma*: a state of knowledge study. Oxford Forestry Institute (Oxford University), Oxford

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

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Puechmaille SJ (2016) The program does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem. Molecular Ecology Resources 16 (3):608–627

QGIS Development Team (2018) QGIS geographic information system. Open Source Geospatial Foundation Project. http://qgis.osgeo.org

Riul P, Targino CH, Júnior LAC, Creed JC, Horta PA, Costa GC (2013) Invasive potential of the coral *Tubastraea coccinea* in the Southwest Atlantic. Mar Ecol Prog Ser 480:73–81

Rosenberg NA (2004) Distruct: a program for the graphical display of population structure. Mol Ecol Notes 4:137–138
Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425
Shaw J, Lickey EB, Schilling EE, Small RL (2007) Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. Am J Bot 94:275–288
Singer A, Schückel U, Beck M, Bleich O, Brumsack HJ, Freund H, Geimecke C, Lettmann KA, Millat G, Staneva J, Vanselow A (2016) Small-scale benthos distribution modelling in a North Sea tidal basin in response to climatic and environmental changes (1970s-2009). Mar Ecol Prog Ser 551:13–30
Thuiller W, Lafourcade B, Engler R, Araújo MB (2009) BIOMOD - a platform for ensemble forecasting of species distributions. Ecography 32:369–373
Thuiller W, Georges D, Engler R (2014) Biomod2: ensemble platform for species distribution modeling. http://CRAN.Rproject.org/package=biomod2
Vekemans X, Hardy OJ (2004) New insights from fine-scale spatial genetic structure analyses in plant populations. Mol Ecol 13:921–935
Verity R, Nichols RA (2016) Estimating the number of subpopulations (K) in structured populations. Genetics 203:1827–1839
Vinceti B, Loo J, Gaisberger H, van Zonneveld MJ, Schueler S, Konrad H, Kadu CA, Geburek T (2013) Conservation priorities for *Prunus africana* defined with the aid of spatial analysis of genetic data and climatic variables. PLoS One 8:e59987
Wakasugi T, Sugita M, Tsudzuki T, Sugiura M (1998) Updated gene map of tobacco chloroplast DNA. Plant Mol Biol Rep 16:231–241
Weising K, Gardner RC (1999) A set of conserved PCR primers for the analysis of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms. Genome 42:9–19
White F (1986) La Vegetation de L'Afrique (the vegetation of Africa), Vol 20. IRD Editions (France)

This study is a part of Dyana Ndiade-Bourobou's doctoral thesis which was funded by the French Ministry of Foreign Affairs and the French Embassy in Libreville (Gabon).