

Species delimitation in the African tree genus *Lophira* (Ochnaceae) reveals cryptic genetic variation.

Eben-Ezer Baba Kayode Ewédjè^{1,5,£}, Simon Jansen¹, Guillaume Kouame Koffi^{1,3}, Adrien Staquet¹, Rosalia Piñeiro, Rodolphe Abessole Essaba¹, Nestor Laurier Engone Obiang^{4,6}, Kasso Dainou², Achille Bernard Biwolé², Jean-Louis Doucet², Olivier J. Hardy¹

* corresponding author: Simon Jansen; phone: +43 699 10 41 12 91;
E-mail: simon.jansen@posteo.de

£ these authors contributed equally to this work

Authors' affiliation:

1. Evolutionary Biology and Ecology Unit, Faculté des Sciences, Université Libre de Bruxelles, Avenue F. D. Roosevelt 50, CP 160/12, B-1050, Belgium
2. Gembloux Agro-bio Tech, Université de Liège, Passage des Déportés 2, B-5030 Gembloux, Belgium
3. Laboratoire de Génomique Fonctionnelle et Amélioration Génétique, Université Nangui Abrogoua, BP 801, Abidjan 02, Côte d'Ivoire
4. Institut de Recherche Ecologie Tropicale (IRET), BP 13 354, Libreville, Gabon.
5. Laboratoire de Biologie, Ecologie végétale appliquée et de génétique forestière; Faculté des Sciences et Techniques de Dassa Zoumè; Université Nationale des Sciences, Technologies, Ingénierie et Mathématiques (UNSTIM-Abomey), Bénin
6. Herbier National du Gabon, BP 1156 Libreville

Keywords: Central and West Africa, cryptic species, genetic structure, introgression, microsatellites, speciation, ecotone, hybridization.

This is the Postprint of the article published in Conservation Genetics:

Ewédjè, E.B.K., Jansen, S., Koffi, G.K. et al. Species delimitation in the African tree genus *Lophira* (Ochnaceae) reveals cryptic genetic variation. *Conserv Genet* 21, 501–514 (2020).
<https://doi.org/10.1007/s10592-020-01265-7>

ABSTRACT

Species delimitation remains a crucial issue for widespread plants occurring across forest-savanna ecotone such as *Lophira* (Ochnaceae). Most taxonomists recognize two parapatric African tree species, widely distributed and morphologically similar but occurring in contrasted habitats: *L. lanceolata* in the Sudanian dry forests and savannahs and *L. alata* in the dense Guineo-Congolian forests. Both species co-occur along a c. 3000 km long forest-savanna mosaic belt, constituting ideal models for investigating hybridization patterns and the impact of past glacial periods on the genetic structures in two types of ecosystems. We genotyped 10 nuclear microsatellites for 803 individuals sampled across the distribution range of *Lophira*. Both species exhibit similar levels of genetic diversity ($H_e = 0.52$ (*L. alata*); 0.44 (*L. lanceolata*)) and are well differentiated, consistent with taxonomic delimitation ($F_{ST} = 0.36$; $R_{ST} = 0.49$), refuting the hypothesis that they might constitute ecotypes rather than distinct species. Furthermore, *L. alata* displayed two deeply differentiated clusters ($F_{ST} = 0.37$; $R_{ST} = 0.53$) distributed in parapatry, one endemic to Western Gabon while another cluster extended over the remaining species range, suggests that *L. alata* is made of two cryptic species. We showed that rare hybrids occur in some contact zones between these three species, leaving a weak signal of introgression between *L. lanceolata* and the northern cluster of *L. alata*. At the intra-specific level, the latter species also show weak genetic structuring between Upper and Lower Guinea and the intensity did not differ strikingly between rainforest and savanna ecosystems. The discovery of a new species of *Lophira* with a narrow distribution in West Gabon where it is intensively exploited for its timber requires to evaluate its conservation status.

INTRODUCTION

A species is a group of potentially interbreeding and evolving populations that are reproductively isolated from other species (the key criterion of the Biological Species Concept, BSC; Schemske 2000). Species delimitation is an essential step for the conservation and management of forest ecosystems as inaccurate taxonomical identification can bias scientific conclusions on biodiversity patterns and population fitness (Reed and Frankham 2003). The high plant biodiversity found in tropical biomes provides serious challenges for researchers dealing with species delimitation issues, and the fact that many samples of tropical trees in herbarium collections lack reproductive organs complicates the work of taxonomists. Fortunately, population genetics can help to disentangle species boundaries in sister taxa (e.g., Dauby et al. 2010; Ikabanga et al. 2017; Lissambou et al. 2018) or to identify new cryptic species (Dainou et al. 2016) by identifying reproductively isolated entities that occur in sympatry.

Ecotones (areas at the junction of two different biomes) are often where the distribution boundaries of two closely related species meet, hence increasing chance of hybridization (e.g., Dainou et al. 2017). Whereas this is well documented for plant and animal species in several continents (e.g., Rieseberg and Carney 1998; Morgan-Richards et al. 2010 and references therein), hybrid zones have been less documented in African biomes (but see animal examples in Mondol et al. 2015). Hybrid zones are qualified as unimodal or bimodal, according to the shape of the abundance distribution of genotypic classes (Harrison and Bogdanowicz 1997; Jiggins and Mallet 2000). Unimodal distribution (predominance of hybrids – including backcross genotypes) in the heart of the zone questions the degree of speciation between the studied taxa whereas bimodal distribution (rare hybrids and predominance of each species in the hybrid zone) tends to confirm the high differentiation between the target taxa. Jiggins and Mallet (2000) argue that hybrid-index plots are highly informative to understand the underlying biological processes but they remain too seldom used in the concerned literature. For African tree taxa, while many population genetic studies focused on either forest taxa (e.g., Duminil et al. 2010; Lowe et al. 2010; Born et al. 2011; Hardy et al. 2013) or savannah taxa (e.g., Fontaine et al. 2004; Allal et al. 2011; Odee et al. 2012), we lack comparative studies on sister taxa that preferably inhabit adjacent biomes.

Genetic differentiation between populations can be due to long or repetitive geographical isolation (reduced gene flow leading to genetic drift), sometimes combined with adaptation to changing

environments (Smith et al. 2001). Gene flow among such populations can be limited by a reinforcement of assortative mating through diverse mechanisms (e.g., delay of flowering phenology, Lowry et al. 2008; Sobel et al. 2009; Daïnou et al. 2017, de Lafontaine et al. 2018) and ultimately lead to speciation. Within a given vegetation type, habitat fragmentation is reported as one of the most important cause of genetic differentiation. African forest taxa have presumably been strongly impacted by past glacial periods associated with an expansion of savannahs, leading to range contraction, fragmentation and subsequent genetic differentiation (Maley 1996; Dupont et al. 2000; Hardy et al. 2013). The Guineo-Congolian rainforest is made of two major forest blocks: the West African forest block (Upper Guinea) and the Central African forest block which is divided into the Lower-Guinean (on the West) and the Congolian (on the East) phytogeographic domains (White 1979). These two blocks are separated by a ca. 200 km wide savannah corridor stretching from East-Ghana to Benin and named 'Dahomey Gap'. This corridor is considered as an important ecogeographical barrier, as it coincides either with the distribution limits of numerous species (e.g., Fayolle et al. 2014) or with the limits between intra-specific lineages or genetic clusters of many widespread species (e.g., Demenou et al. 2016, 2017). Past glacial-interglacial fluctuations probably had less impact on the dry woodlands and savannahs situated north of the Guineo-Congolian region because we can hypothesize that this biome changed in latitudinal range and extent but without fragmentation, which may explain the limited phylogeographic differentiation observed in this area for some species (e.g., Odee et al. 2012; Sexton et al. 2015).

The African context offers the opportunity to investigate the relative role of ecotones and past population fragmentation on the genetic diversity of parapatric sister species as (1) evergreen rainforests (from Upper and Lower Guinea) and savannas (from the Sudanian zone) meet along an ecotone on at least 5000 km of longitude from West to Central Africa, and (2) some widespread plant genera are distributed on each side of the forest-savanna contact zone. The genus *Lophira* Banks ex C. F. Gaertn. (Ochnaceae) is interesting in this respect because it includes two widespread species found in well-differentiated habitats: the Guineo-Congolian rainforest tree *L. alata* Banks ex C. F. Gaertn., which requires high precipitation, and the Guineo-sudanian woody savannah and dry forest tree *L. lanceolata* Tiegh. ex Keay, which occurs in regions with a marked dry season (Aubreville 1959; Hutchinson and Dalziel 1954; Bamps 1970; Biwole et al. 2012). These species differ in their general stature (*L. alata* reaches much larger sizes than *L. lanceolata*) but they are otherwise very similar

morphologically (Aubreville 1959) so that their origin and relationships are unclear (Persinos and Quimby 1968). They have sometimes been considered as synonyms (Keay 1953), assumed to be distinct morphotypes resulting from ecological adaptations (Letouzey 1968). Letouzey (1957) suggested that *L. alata* would result from an ecological adaptation of *L. lanceolata* to Guineo-Congolian rainforests, and that these species might be ecotypes rather than truly distinct species. According to Maley (1996), *L. alata* is among the plant species that have spread out of the African forests during the Holocene humid period. In transitional vegetation types between rainforest and savannah, the two species co-occur.

We can hypothesize that if *L. alata* and *L. lanceolata* are distinct species we should find (i) bimodal hybrid zones with no or few hybrids and (ii) higher population genetic differentiation across the forest-savanna boundary than across past biogeographic barriers (e.g. the Dahomey Gap). On the contrary, if *L. alata* and *L. lanceolata* are ecotypes with regular gene flow in contact zones, we should find (i) unimodal hybrid zones and (ii) higher population genetic differentiation across biogeographic barriers (e.g. between Upper and Lower Guinea) than across the forest-savanna boundary. Using molecular markers, we aimed to decipher species boundaries in *Lophira* and compare the genetic structures of populations found in rainforest versus savannah biomes to address the following questions: (1) Do *L. alata* and *L. lanceolata* constitute distinct genetic entities as expected in reproductively isolated species or are they ecotypes from a same species showing evidence of hybridization and/or introgression in the forest-savanna transition zone? (2) Are there large-scale genetic subdivisions within each species that would suggest past population fragmentation and are they less marked in the savanna biome than the rainforest biome? Finally, as our results indicate that a cryptic species with a narrow distribution range likely occurs within *L. alata*, we will also address conservation issues.

MATERIALS and METHODS

Study species, habitat and uses

Lophira alata is a large long-living tree reaching 60 m height and a trunk diameter up to 1.8 m, distributed from Sierra Leone to Congo (Letouzey 1957; Bamps 1970; Biwolé et al. 2012) throughout the evergreen to semi-deciduous dense forests of UG and LG and in some parts of the DG. It is an economically important timber tree species known under the trade name Azobé or Ekki, used for

outdoor constructions such as harbors, hydraulic infrastructure, railways, and bridges in Central Africa (Biwolé et al. 2012). Due to its current heavy rate of logging it is classified as a vulnerable species in the red list of the International Union for Conservation of Nature (IUCN 2014).

Lophira lanceolata (Méné) is a light-demanding and pioneer tree species reaching 24 m height and a trunk diameter of 70 cm. It is widespread in Sudano-Guinean savannahs and forest-savannah mosaics (Bamps 1970; Biwolé et al. 2012) and distributed from Sierra Leone to Cameroon. Its seeds are processed in an edible oil widely used in food, cosmetics, and medicinal preparations in West Africa.

Both *Lophira* species possess the same number of chromosomes ($2n = 28$, Satabié 1982) and similar morphology of pollen, flower, fruit. Apart from the tree size, the main distinguishing morphological traits concern leaf branch texture (smooth in *L. alata* vs. scaly-corky in *L. lanceolata*) and petiole size which are shorter in *L. alata* than in *L. lanceolata* (von Thieme 1929; see also Persinos and Quimby 1968). *Lophira* flowers are large, whitish, hermaphrodite and fragrant, producing numerous stamens with yellow to orange anthers (Meunier et al. 2015), and bee-pollinated (Mapongmetsem 2007; Oriana Bhasin, pers. Comm.). Their fruits are conical achenes (up to 13 cm long) supplied with two unequal outer sepals extremely accrescent and forming wings. They are mainly disseminated by wind (~200 m) and are consumed by ruminants and rodents (Anon 1954).

Sampling and microsatellite genotyping

We sampled silica gel-dried leaves or cambium from trees during multiple field expeditions in nine countries, covering most of the natural distributions of the two *Lophira* species (Fig. 1A). One of these expeditions focused on the forest-savanna mosaic area of Cameroon where the two species occur in close vicinity. Another expedition, in Gabon, sampled trees near the town of Ndjolé where a contact zone between two genetic groups of *L. alata* was identified during preliminary data analyses. Total genomic DNA from 433 *L. alata* and 370 *L. lanceolata* individuals was isolated using the NucleoSpin 96 plant kit (Macherey-Nagel, Düren, Germany) according to the manufacturer's protocol. We amplified 13 polymorphic nuclear microsatellite loci (namely P12, P18, P24, P31, P34, P36, P40, P44, P47, P51, P53, P62, P66) isolated from *L. alata* and optimized in two multiplex reactions according to Piñeiro et al. (2015). Three loci (P24, P36, P44) that suffered from weak amplification were discarded. When an individual presented no trace of PCR amplification product for a given locus while alleles

were clearly visible at the other loci amplified in the same PCR mix, it was considered homozygote for a null allele.

Data analyses

Genetic structure

In order to group individuals by their genetic affinity and assess whether *L. alata* and *L. lanceolata* constitute distinct genetic groups, we used (a) the individual-based Bayesian genetic assignment method implemented in STRUCTURE 2.3.4, which attempts to delimit genetic clusters within which Hardy-Weinberg and linkage equilibria hold (Pritchard et al. 2000) and (b) conducted a principal component analysis (PCA) using the R package “adeigenet” (Jombart 2008).

To account for the potential presence of hierarchical genetic structure, the STRUCTURE analysis was performed by a two-step approach using different settings. The first analyses were performed on the complete dataset running the Monte-Carlo Markov Chain (MCMC) 10 times for each number of clusters K ranging from 1 to 8 using 20 000 generations with a burn-in length of 10 000. Following Pritchard et al. (2010), we performed these initial runs under different models considering admixture or no admixture, and independent or correlated allele frequencies. Subsequently, we performed further STRUCTURE analyses using only individuals that were assigned to a specific genetic cluster in the previous step. For these runs we assumed K = 1-5, using the ancestry model with admixture and correlated allele frequencies. All STRUCTURE results were processed through the online tool STRUCTURE SELECTOR (Li and Liu 2018) to determine the optimal K. This tool calculates the widely used ΔK method (Evanno et al. 2005) and four alternative statistics (MEDMEDK, MEDMEAK, MAXMEDK and MAXMEAK) recently introduced by Puechmaille (2016). While ΔK tends to identify the uppermost hierarchy of population structures, the remaining statistics appear to be more accurate for subtle population patterns and robust towards uneven sampling data. The program CLUMPAK (Kopelman et al. 2015) was used to summarize the 10 simulations for the best number of clusters found.

SSR diversity and inbreeding

We characterized the genetic diversity at two scales: (a) for each genetic cluster identified, and (b) at a finer scale according to broad biogeographic regions, distinguishing Upper Guinea, Dahomey Gap, Lower Guinea and Western Gabon (called populations hereafter). We computed the following statistics used SPAGeDi version 1.5 (Hardy and Vekemans 2002): mean number of alleles per locus, effective number of alleles, allelic richness after rarefaction, expected and observed heterozygosity (Nei 1978), and deviation from Hardy-Weinberg equilibrium (HWE). To check whether departure from HWE at a given locus might be explained by the presence of null alleles, null allele frequencies and the mean inbreeding coefficient were estimated using the software INEst version 2.1 under the population inbreeding model (Chybicki and Burczyk 2009). This Bayesian model was run using the following number of MCMC iterations, burning-period and thinning: 200 000, 1000 and 20 000, respectively.

Fixation indices and test of phylogeographic structure

After defining the global pattern of population structure with the above methods, we computed pairwise fixation statistics (F_{ST} , R_{ST}) between inferred clusters and geographical regions using the software SPAGeDi (Hardy and Vekemans 2002). F_{ST} can be interpreted as measuring “nearness to fixation” while R_{ST} is a measure of the homogeneity of SSR allele sizes within populations, although these indices have often been interpreted as measuring genetic differentiation, which is not accurate (Jost et al. 2018). We did not compute true measures of allelic differentiation, such as JostD (Jost et al. 2018), because their high dependency on the mutation rate makes them poor descriptors of the demographic history of populations (Whitlock 2011). R_{ST} and F_{ST} have the same theoretical expectations when population genetic structuring results solely from genetic drift, while if stepwise mutations contribute to genetic differentiation (mutation rate non negligible compared to the migration rate or the reciprocal of the number of generations since population isolation) we expect $R_{ST} > F_{ST}$ (Hardy et al. 2003). The latter situation implies that mean allele sizes vary between populations for at least some loci, leading to a phylogeographic signal. The latter was tested using the allele sizes permutation test (Hardy et al. 2003) implemented in SPAGeDi. Furthermore, to account for a potential positive bias in F_{ST} values due to null alleles, the ENA method implemented in FReeNA was used to recalculate pairwise F_{ST} values (Chapuis and Estoup, 2007).

Hybridization

We used the STRUCTURE analysis (admixture model) to identify potential hybrids between genetic clusters. Individuals were considered as admixed when their probability of assignment to the most likely genetic cluster had a Q-value lower than 0.8. For the contact zones between genetic clusters (see results) we produced hybrid-index plots, and mapped the few admixed individuals. Furthermore, we used the program NewHybrids (Anderson and Thompson 2001) under the “Jeffreys prior” settings to characterize putative admixed individuals into six frequency categories: purebred (A/B), F1 hybrid, F2 hybrid, backcrossed F1 to purebred (A/B). We also checked the genotypes of the putative hybrids to assess whether they correspond to F1 hybrids according to Mendelian inheritance.

Spatial genetic structure and introgression between species

Hybridization does not necessarily lead to introgression - defined here as gene flow between species due to hybridization plus repeated backcrossing - if hybrids or their progeny are sterile or strongly counter-selected. Therefore, we attempt to detect introgression by testing for a spatial covariation in the genetic structures of hybridizing species (Hardy and Vekemans 2001). To this end, first we compared fixation indices between parapatric versus allopatric populations of distinct species, expecting lower values between parapatric populations if they are subject to inter-specific gene flow. Second, we examined the decay of kinship coefficients with geographic distance within and between species (Hardy and Vekemans 2001). The unweighted mean allele frequencies over the two species were used as reference to estimate pairwise kinship coefficients between individuals, F_{ij} (estimator of Loiselle et al. 1995, implemented in SPAGeDi). F_{ij} values were then averaged among i-j pairs separated by predefined spatial distance intervals, distinguishing i-j *L. alata* pairs, i-j *L. lanceolata* pairs and heterospecific i-j pairs. If the two species introgress recurrently in the contact zones, we expect that heterospecific pairs of individuals are on average more genetically similar if there are spatially close than if they are distant (Hardy and Vekemans 2001). We thus examined the slope of the interspecific kinship-ln(distance) curve, using a jack-knife procedure over loci to obtain an approximate test of the significance of the regression slope. This analysis was conducted at two scales: (i) a large scale including all 665 samples from Ivory Coast to Cameroon (excluding samples from Gabon where *L. lanceolata* is absent and a peculiar genetic cluster occurs in *L. alata*), and (ii) a fine scale including 166 samples around the forest-savannah contact zone of Cameroon where the two species can

potentially interbreed. At the two scales, two hybrid individuals from the contact zone in Cameroon were excluded from these analyses, as they could not be genetically assigned to a given species.

RESULTS

Genetic structure and species delimitation

For all STRUCTURE runs using the complete dataset the likelihood of the data increased with K, especially from K = 1 to K = 3. Accordingly, the K estimate following the approach of Evanno et al. (2005) corresponds to K = 2, which matched to the two taxonomic species recognized and suggested that *L. alata* and *L. lanceolata* were well separated genetically (Fig. 2B). However, the clustering solution at K = 3 appeared more satisfactory because all but 8 samples were assigned to one of the three clusters with an ancestry coefficient $Q > 0.8$, a result robust to model specifications (i.e., with or without admixture, with correlated or independent allele frequencies; see also Supplementary Online 1). Under K = 3, which will be referred to as 'major clusters', in addition to the *L. lanceolata* cluster, two clusters occurred among *L. alata* samples (Fig. 2B and C, see Fig. 1A for their spatial distribution): one endemic to Western Gabon (hereafter *L. alata* 1) and the second spread from West Africa to Central Africa and essentially distributed in the Northern hemisphere except in Eastern Gabon (hereafter *L. alata* 2). The methods proposed by Puechmaille (2016) - MEDMEDK, MEDMEAK, MAXMEDK and MAXMEAK - consistently suggested K = 5, subdividing in two both the *L. lanceolata* and the *L. alata* 2 clusters into western (Upper Guinea) and eastern (Lower Guinea) sub-clusters. However, many samples appeared as admixed between these western and eastern sub-clusters. The subsequent Structure analysis performed on *L. alata* samples only recovered the *L. alata* 1 and *L. alata* 2 clusters at K = 2. Additional clustering analyses within *L. lanceolata* and within *L. alata* 2 again showed that at K = 2, samples from Upper and Lower Guinea tended to be assigned to different sub-clusters but with many intermediate samples (Fig. 2C).

The recognition of three major clusters was confirmed by the PCA on the genetic data which exhibited three well-delineated groups along axes 1 and 2, explaining respectively 31.8 % and 7.6 % of the total variance (Fig. 2A). The sub-clusters detected in *L. lanceolata* and *L. alata* 2 were, by contrast, highly overlapping (Fig. 2C).

Microsatellite diversity and inbreeding

Over the ten loci, the number of alleles (A) per locus range from 8 to 18 (mean $A = 9.7$; see also Supplementary Online 2) and the expected heterozygosity (H_E) per locus considering all samples ranged from 0.38 (P34) to 0.73 (P66) (mean $H_E = 0.62$). Diversity measured within each of the three major genetic clusters show that the cluster *L. alata* 1 (West-Gabon) presents the highest genetic diversity ($H_E = 0.58$), while cluster *L. alata* 2 (North) and *L. lanceolata* show similar levels of genetic diversity ($H_E = 0.45$ and 0.44 , respectively).

Further, nSSR genetic diversity was estimated within each of the seven populations defined on the basis of biogeographic regions and major genetic clusters (Table 1). In *L. alata*, allelic richness (A_r) ranged from 2.24 (Dahomey Gap) to 2.69 and 3.18 (Lower Guinea and West-Gabon respectively) while H_E varied between 0.33 and 0.46 for all populations apart from West-Gabon for which it reached 0.58. In *L. lanceolata*, A_r and H_E ranged in 2.74-2.86 and 0.40-0.47, respectively. Globally, *L. alata* populations from Central Africa displayed higher level of genetic diversity than those from Upper Guinea. Deficit of heterozygotes was observed in all populations (F_{IS} ranging from 0.09 to 0.36) but after accounting for the bias due to null alleles with INEst, populations did not deviate significantly from HWE, suggesting that outbreeding predominates.

Fixation indices between major clusters and populations

Pairwise fixation indices were high between *L. lanceolata* and *L. alata*, consistently with taxonomic delimitation ($F_{ST} = 0.36$ and $R_{ST} = 0.49$), but also between cluster *L. alata* 1 from West-Gabon and both *L. lanceolata* ($F_{ST} = 0.38$, $R_{ST} = 0.40$) and *L. alata* 2 from the northern hemisphere ($F_{ST} = 0.37$, $R_{ST} = 0.53$). Pairwise F_{ST} values corrected for null alleles using the ENA method were 0.03 to 0.05 units lower than uncorrected F_{ST} values (Table 2A). A phylogeographic signal (R_{ST} significantly larger than F_{ST}) was only observed between the two *L. alata* clusters ($p = 0.04$). By contrast, low fixation indices occurred among populations of *L. alata* 2 ($F_{ST} = 0.03$, $R_{ST} = 0.02$ among Upper Guinea, Dahomey Gap and Lower Guinea) or among *L. lanceolata* populations ($F_{ST} = 0.06$, $R_{ST} = 0.06$ among Upper Guinea, Dahomey Gap, Cameroon; see also Table 2B). These results highlight a strong genetic discontinuity between the genetic clusters detected in *L. alata*, similar to the difference between species, confirming the pattern obtained with STRUCTURE.

Hybridization

A closer look at the location of admixed individuals inferred from STRUCTURE at $K = 3$ allowed us to identify four potential hybrid zones where the hybrid index was systematically bimodal (Fig 1 E-G). In Upper Guinea, one *L. lanceolata* sample located close to the former forest-savannah boundary of Ivory Coast (nowadays substantially deforested) appeared admixed between clusters *L. lanceolata* (64%) and *L. alata 2* (36%) (Fig. 1B and E). In southern Dahomey Gap (southern Benin), three individuals of *L. lanceolata* have $Q = 30\%$ of their genomes assigned to the *L. alata 2* cluster, while they are located c. 50 km from a population of pure *L. alata 2* (Fig. 1B). In the forest-savannah mosaic of Cameroon where we sampled many *L. alata* and *L. lanceolata*, three individuals identified as *L. alata* appeared admixed between *L. lanceolata* ($Q = 29\%, 40\%, 43\%$) and *L. alata 2* (Fig. 1C and F). At three loci (P12, P40, P66) containing diagnostic alleles of each species, the genotypes of these admixed samples were systematically heterozygote with one allele of *L. lanceolata* and the second from cluster *L. alata 2* (Supplementary Online 2). Finally, in Gabon, just at the contact zone between clusters *L. alata 1* and *L. alata 2*, there was one admixed individual between these clusters ($Q = 59\%$) (Fig. 1D and G).

Further analyses using NewHybrids identified the same admixed individuals and classified them into four different hybrid categories (see 1B - D). While two F1 hybrids and one unassigned admixed individual were detected in the forest-savannah mosaic in Cameroon, the remaining admixed individuals between *L. alata 2* and *L. lanceolata* were categorized as F2 hybrids. NewHybrids failed to assign the admixed individual between *L. alata 1* and 2 into a particular category.

Detection of interspecific gene flow

Fixation indices between populations from distinct species occurring in the same contact zone ($F_{ST} = 0.38$ in LG and $F_{ST} = 0.49$ in Upper Guinea) were not systematically lower than between allopatric populations from West and Central Africa ($F_{ST} = 0.45$ between Upper Guinea *L. alata* and Cameroonian *L. lanceolata*, $F_{ST} = 0.43$ between Lower Guinea *L. alata* and Upper Guinea *L. lanceolata*). Hence, there is no evidence of interspecific introgression at this scale based on fixation indices. Considering the spatial genetic structures, at large scale we observed a significant decay of kinship coefficients between individuals with geographic distance within each species, indicating

isolation by distance: regression slopes with Ln(distance) equal -0.010 (SE = 0.002) and -0.020 (SE = 0.005) for *L. alata* and *L. lanceolata*, respectively (Fig. 3A). When the inter-specific regression slope was computed over the full distance range (whole set of heterospecific pairs of individuals), the regression slope equaled 0.003 with a SE = 0.009 and is thus not significantly different from 0. However, when the regression was limited to heterospecific pairs of individuals separated by less than 200 km, or when it was computed over samples from the Cameroonian contact zone (Fig. 3B), the regression slope was -0.020 with SE = 0.010. This result suggests that some introgression between species occurred but it affected their genetic structures only at a local scale in the Cameroonian contact zone.

DISCUSSION

Species delimitation in *Lophira*

The principal component analysis (PCA) and the Bayesian clustering algorithm detected three main genetic clusters within *Lophira* spp. (Fig. 2A - C). It confirmed that *L. alata* and *L. lanceolata* are well separated genetically, consistently with taxonomical species delimitation, and congruently with the bimodal distributions of the hybridization index in all contact zones (Fig. 1E, 1F). Our results thus support the view that the two taxa cannot be considered as ecotypes of a same species. It is not possible with our data to demonstrate that a model of sympatric speciation would apply because the current parapatric distribution of the species may result from a secondary contact after allopatric speciation. Molecular dating based on DNA sequences should help to check whether their divergence predates the Quaternary as assumed for most congeneric African plant species (Plana et al. 2004). Unsuspectedly, our results also highlighted two parapatric genetic clusters within *L. alata* which are as much genetically separated from each other as they are from *L. lanceolata*, and which also form a bimodal hybrid zone, suggesting they could represent cryptic species. Overall, our data support the recognition of three species of *Lophira*.

Hybridization between *L. alata* and *L. lanceolata*

Although there is no evidence of inter-specific introgression at large scale according to inter-specific F_{ST} values and kinship-distance curve, the decay of the inter-specific kinship-distance curve at a local scale in the Cameroonian contact zone suggests that occasional gene flow occurs, creating a slight

spatial correlation between the allele frequencies of the two species. The detection of hybrid individuals in this region (F1), as well as one in Ivory Coast and the Dahomey Gap (F2), confirms that hybridization is possible, although it must remain rare to explain the maintenance of high fixation indices. Although, NewHybrids identified F1 and F2 hybrids, the classification should be interpreted with caution as using 10 microsatellite markers limits the resolution. The low inbreeding coefficient estimated in each population after accounting for null alleles indicated that both species are essentially outcrossing. Both species have similar pollen morphology and structure (Satabié 1982). Their reproductive isolation in the contact zone could result from postzygotic barrier (Rieseberg and Carney 1998) or from floral phenological shift because flowering is reported mainly in November-January for *L. alata* and January-July for *L. lanceolata* (Biwolé et al. 2012, Aubreville 1959, Akoegninou et al. 2006, herbaria data). Further studies characterizing hybridization pattern (proportion and genetic composition of hybrid individuals in the contact zone, genotyping seed lots in different contact zones) must be undertaken to detect eventual pre- and/or post-zygotic isolation processes and assess the fertility of hybrids.

Genetic clusters within *L. alata*: evidence of cryptic species?

A striking and unexpected result of this study is that besides the clear genetic discontinuity between *L. alata* and *L. lanceolata*, is that *L. alata* is itself made of two parapatric genetic clusters forming a sharp genetic discontinuity with a bimodal hybrid zone in their contact area (Fig. 1G). The complementary sampling conducted in the surroundings of the Gabonese town Ndjolé, which was found to be close to the limit between these clusters during our initial investigations, allowed us to detect only one hybrid individual (see Fig 1D and G). However, the NewHybrids algorithm failed to identify its hybridization class (F1, F2, Backcross). As all analyses confirm the pronounced differentiation between these two clusters (the Bayesian clustering algorithm, the PCA ordination and the F_{ST}), either they represent distinct cryptic species between which a barrier to gene flow is well established, or they are deeply differentiated conspecific populations that formed a secondary contact zone so recently that admixed individuals are still uncommon.

Intra-specific genetic clusters have been repeatedly reported in African forest tree species (e.g., Hardy et al. 2013) and are generally interpreted as evidence of past population fragmentation (forest refuge hypothesis), possibly driven by Pleistocene climate change or even older events (e.g. Migliore et al.

2018). The West-Gabon *L. alata* cluster displays relatively high genetic diversity and allelic endemism, and may therefore be a legacy of ancient forest fragmentation. However, the fixation indices among *L. alata* clusters ($F_{ST} = 0.37$, $R_{ST} = 0.53$) are well above the values typically reported between parapatric genetic clusters in other Guineo-Congolian tree species (e.g., $F_{ST} = 0.07-0.28$; Duminil et al. 2013; Demenou et al. 2016, 2017). Moreover, the transition between parapatric clusters in *L. alata* is extremely sharp, occurring over a width of only c. 1 km near Ndjolé, while in other tree species it typically occurs over tens of kilometers, sometimes hundreds (e.g. Pineiro et al 2017; Demenou et al 2016, 2017; Dainou et al 2016; Duminil et al. 2013). The bimodal hybrid zone displayed by *L. alata* seems more similar to the contact zone between two *Erythrophleum* sister tree species in Central Africa: *E. ivorense* occurs in the wettest evergreen forest along the coast up to c. 200-300 km inland, while *E. suaveolens* occurs in slightly drier semi-evergreen forest occurring more inland (Duminil et al. 2013). Analogous to the *L. alata* clusters, the distribution ranges of these *Erythrophleum* species are parapatric and seem to form a sharp boundary that does not correspond to any obvious biogeographical barrier. This configuration might be explained if the *Erythrophleum* species or the *L. alata* clusters were excluding each other. Local exclusion of a minority species/cluster may be driven by heterospecific pollen transfer, whereby crosses between species/clusters eventually lead to the loss of ovules due to pre or post-zygotic processes, and reduce much more the reproductive success of the locally least abundant species/cluster (Arceo-Gómez and Ashman 2013; Morales and Traveset 2008).

Overall, our data rather support the hypothesis that the two *L. alata* clusters represent cryptic species. To further test it, we should record phenological data, genotype seeds from each cluster in the contact zone, and obtain herbarium samples with reproductive organs to check if diagnostic traits occur. A first comparison of eight observations (flowering time) provided by the herbarium of Leiden (NL; D-Y Ouedraogo, pers. comm.) supports the assumption of a flowering delay in Gabon (West-Gabon: October, November; North-Gabon: December, January). Preliminary analyses of herbarium material that should belong to the two *L. alata* genetic clusters did not reveal any distinctive morphological trait but as only one fertile specimen was available from West Gabon in the different herbaria consulted (BS4657 from BRLU and BR), new herbarium material with reproductive organs need to be collected. *Lophira alata* being an important commercial timber tree, a comparison of wood anatomical features

and technological properties of the two clusters would also be interesting, in particular for the forestry sector.

Very low genetic structuring between West and Central Africa in *Lophira*

We found a weak genetic structure over a large area in *L. lanceolata* and in the cluster *L. alata* 2. If we exclude the West Gabonese *L. alata* which seems to form a separate species, there is thus no support for our hypothesis that genetic differentiation at large scale is stronger in the rainforest ecosystem than in the savannah ecosystem because the former was more prone to fragmentation during past climate changes. The weak genetic structure between Upper Guinea and Lower Guinea in *L. alata* 2 contrasts with the clear differentiation reported at this scale in most of the other forest tree species studied so far (Gomez et al. 2009; Koffi et al. 2011; Daïnou et al. 2016; Demenou et al. 2016, 2017). For savanna species extending across the Sudanian region, different patterns have been reported: a clear East-West genetic clustering in *Vitellaria paradoxa* (Allal et al. 2011) but a weak genetic structuring in *Acacia senegal* (at least within the Sudanian area, Odee et al. 2012) and *Khaya senegalense* (Sexton et al. 2015), similar to what we observed in *L. lanceolata*. While glacial-interglacial cycles of the Pleistocene have probably resulted in repeated fragmentation of the Guineo-Congolian forest cover, the savanna area corresponding to the current Sudanian vegetation belt has probably fluctuated in area and in latitudinal range but it has not necessarily been fragmented. This might explain why we find generally less genetic structuring in savanna trees than in forest trees, at least in West and Central Africa. The large distribution of the *L. alata* 2 cluster (spread from West Africa to Central Africa) is thus surprising. It may result from the maintenance of substantial gene flow during the glacial periods, due to either extensive gene dispersal or weakly fragmented populations, or to limited genetic drift if large effective population sizes were maintained. As *L. alata* is considered as a pioneer species it may have developed mechanisms to survive unsuitable conditions. However, according to species ecological requirements (evergreen forests) and demographic dynamics (7 mm/year in diameter increment and 55-60 years to become reproductive; for review see Biwole et al. 2012) a recent and rapid expansion from a single forest refuge is doubtful. As genetic diversity in *L. alata* clusters is highest in Lower Guinea and West-Gabon, one may thus hypothesize that *L. alata* mainly survived in two main refuge areas that remained isolated for a very long time, and that the Lower Guinea

population has expanded rapidly more recently to cover most of the Northern hemisphere of West and Central Africa. However, further data will be needed to test this hypothesis in the future, and we cannot exclude that the polymorphism of our microsatellite markers was insufficient to identify discontinuities in the spatial pattern of genetic variation if actual intra-specific clusters were still weakly differentiated.

Conservation issues

Despite its wide distribution, *L. alata* is considered as Vulnerable in the red list of the IUCN due to its heavy rate of logging. In Gabon, production of roundwood and sawnwood has steadily increased from c. 3.5×10^6 m³ in 2010 to 4.2×10^6 m³ in 2016 (database of the International Tropical Timber Organization, www.itto.int). However, we have no indication on the relative origin (West-Gabon *L. alata* 1 versus *L. alata* 2) of these volumes, nor on the remaining stocks and growth rates in natural forests for each putative species. As *L. alata* 1 ranges over a relatively small area (c. 300 km latitude by c. 180 km longitude), most of which apparently lying outside of protected areas, an evaluation of the impact of logging on its conservation status seems urgent, as well as inventories and genetic analyses of samples from protected areas. The focus on *L. alata* as a target species in international projects (Large-Scale, TimTrace) opposing illegal timber trade by genetic barcoding are a promising sign for the species conservation, but also emphasizes the importance and urgency of clarifying species delimitation.

CONCLUSION

This study investigated the diversification of the genus *Lophira* and genetic structuring patterns that occurred within rainforests and savannas of tropical Africa. It confirmed that *L. alata* and *L. lanceolata* formed well distinct species genetically, although hybridization can still sometimes occur in contact zones and possibly lead to weak introgression at a local scale. It revealed that *L. alata* probably contains two cryptic species, adding to the mounting genetic evidences of cryptic species in African trees (e.g., Ikabanga et al. 2017; Lissambou et al. 2018). Additional investigations on molecular phylogeny and morphological, anatomical and phenological traits are needed to characterize the population of western Gabon, especially as it is potentially threatened by logging activities and may require particular conservation measures.

ACKNOWLEDGEMENT

Financial support was provided by the Belgian Federal Science Policy Office (Belspo) through a postdoctoral grant (EEBK) within the project AFRIFORD, and by the F.R.S.-FNRS (grant n° J.0292.17). Furthermore, this project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 801505. We would like to thank Dakis-Yaoba Ouedraogo for providing herbarium data for *L. alata* individuals from Gabon including phenological traits, and Marc Sosef for having compared the few herbarium vouchers from western Gabon and eastern/northern Gabon available at the Meise Botanic Garden (BR). We thank CENAREST (Gabon) for the sampling permission (AR0012/18). Moreover, we thank the three anonymous reviewers whose comments have greatly improved this manuscript.

REFERENCES

- Akoegninou A, van der Burg WJ, van der Maesen (2006) Flore Analytique du Bénin. 1064 p
- Allal F, Sanou H, Millet L, Vaillant A, Camus-Kulandaivelu L, Logossa ZA et al. (2011) Past climate changes explain the phylogeography of *Vitellaria paradoxa* over Africa. *Heredity* **107**:601
DOI:10.1038/hdy.2011.112
- Anon (1954) Monographie de Azobé, *Lophira procera* A. Chev. Nogent-sur-Marne, France: Centre Technique Forestier Tropical
- Assi-Kaudjhis C, Digbehi BZ, Roche Eand Lezine A-M (2010) Synthesis on the palaeoenvironments' evolution of Atlantic West Africa since the end of the Last Glacial Maximum. Climatic and human influences. *Geo Eco Trop* **34**:1-28
- Aubreville A (1959) La flore forestière de la Côte d'Ivoire. 2nd éd. Nogent-sur-Marne, France: Centre Technique Forestier Tropical, pp.958
- Bamps P (1970) Répartition géographique du genre *Lophira* Banks ex Gaertn. (Ochnaceae). *Bull Jard Bot Belg* **40**:291-294
- Biwole A, Bourland N, Daïnou K, Doucet J-L (2012) Définition du profil écologique de l'azobé, *Lophira alata*, une espèce ligneuse africaine de grande importance : synthèse bibliographique et perspectives pour des recherches futures. *Biotechnol Agron Soc Environ* **16**:217-228
- Born C, Alvarez N, McKey D, Ossari S, Wickings EJ, Hossaert-McKey M et al. (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
- Channan S, Collins K, Emanuel WR (2014) Global mosaics of the standard MODIS land cover type data. University of Maryland and the Pacific Northwest National Laboratory, College Park, Maryland, USA.
- Chapuis M-P, Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. *Mol Biol Evol* **24**:621–631

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

Dauby G, Zaiss R, Blach-Overgaard A, Catarino L, Damen T, Deblauwe V et al. (2016) Rainbio: a megadatabase of tropical Africa vascular plants distribution. *Phyto Keys* **74**:1-18

Dainou K, Flot J-F, Degen B, Blanc-Jolivet C (2017) DNA taxonomy in the timber genus *Milica*: evidence of unidirectional introgression in the West African contact zone. *Tree Genet Genomes* **13** DOI:10.1007/s11295-017-1174-4

Dainou K, Blanc-Jolivet C, Degen B, Kimani P, Ndiade-Bourobou D, Donkpedan ASL et al. (2016) Revealing hidden species diversity in closely related species using nuclear SNPs, SSRs and DNA sequences – a case study in the tree genus *Milica*. *Evol Biol* **16**:259 DOI:10.1186/st12862-016-0831-9

de Lafontaine G, Napier JD, Petit RJ, Hu FS (2018) Invoking adaptation to decipher the genetic legacy of past climate change. *Ecology* **99**: 1530-1546

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

Demenou B, Pineiro R, Hardy OJ (2016) Origin and history of the Dahomey Gap separating West and Central African rainforests: insights from the phylogeography of the legume tree *Distemonanthus benthamianus*. *J Biogeogr* **43**:1020-1031

Duminil J, Brown RP, Ewedje EBK, Mardulyn P, Doucet, J-L, Hardy OJ (2013) Large-scale pattern of genetic differentiation within African rainforest trees: insights on the roles of ecological gradients and past climate changes on the evolution of *Erythrophleum* spp. (*Fabaceae*). *BMC Evol Biol* **13**:1-13

Duminil J, Heuertz M, Doucet J-L, Bourland N, Cruaud C, Gavory F et al. (2010) Cp-DNA based species identification and phylogeography: application to African tropical tree species. *Mol Ecol* **19**:5469-5483 DOI:10.1111/j.1365-294X.2010.04917.x

Dupont LM, Jahns S, Marret F, Ning S (2000) Vegetation change in equatorial West Africa: time-slices for the last 150 ka. *Palaeogeogr Palaeoclimatol Palaeoecol* **155**:95-122 DOI:10.1016/S0031-0182(99)00095-4

Earl DA, von Holdt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour* **4**:359-361

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 DOI:10.1111/j.1365-294X.2005.02553.x

Fayolle A, Swaine M D, Bastin, J-F, Bourland N, Comiskey J, Dauby G, Doucet J-L, Gillet J-F, Gourlet-Fleury S, Hardy O, Kirunda B, Kouamé FN, Plumptre A J (2014). Patterns of tree species composition across tropical African forests. *J Biogeogr.* **41**. 10.1111/jbi.12382.

Fontaine C, Lovett PN, Sanou H, Maley J., Bouvet J-M (2004) Genetic diversity of the shea tree (*Vitellaria paradoxa* C.F. Gaertn), detected by RAPD and chloroplast microsatellite markers. *Heredity* **93**:639–648 DOI: 10.1038/sj.hdy.6800591

Friedl MA, Sulla-Menashe D, Tan B, Schneider A, Ramankutty N, Sibley A, Huang X (2010) MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets, 2001-2012, Collection 5.1 IGBP Land Cover, Boston University, Boston, MA, USA.

Gomez C, Dussert S, Hamon P, Hamon S, Kochko A, Poncet V (2009) Current genetic differentiation of *Coffeacaneophora* Pierre ex A. Froehn in the Guineo-Congolian African zone: cumulative impact of ancient climatic changes and recent human activities. *BMC Evol Biol* **9**:167 DOI:10.1186/1471-2148-9-167

Hardy OJ, Born C, Budde K, Daïnou K, Dauby G, Duminil J et al. (2013) Comparative phylogeography of African rainforest trees: a review of genetic signatures of vegetation history in the Guineo-Congolian region. *C R Geosci* **345**:284-296 DOI:10.1016/j.crte.2013.05.001

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, Vekemans X (2001) Patterns of allozyme variation in diploid and tetraploid *Centaurea jacea* at different spatial scales. *Evolution* **55**: 943-954

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 DOI:10.1046/j.1471-8286.2002.00305.x

Harrison RG, Bogdanowicz (1997) Patterns of variation and linkage disequilibrium in a field cricket hybrid zone. *Evolution* **51**:493 DOI:10.2307/2411122

Hutchinson J, Dalziel JM (1954) *Flora of West Tropical Africa*. 2nd edn. Crown Agents for Overseas Governments Administrations, London

Ikabanga DU, Stévant T, Koffi GK, Monthe FK, Nzigou Doubindou EC, Dauby G et al. (2017) Combining morphology and population genetic analysis uncover species delimitation in the widespread African tree genus *Santiria* (Burseraceae). *Phytotaxa* **321**:166–180 DOI:10.11646/phytotaxa.321.2.2

Jiggins CD, Mallet J (2000) Bi-modal hybrid zones and speciation. *Trends Ecol Evol* **15**:250-255

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**:1403-1405 DOI:10.1093/bioinformatics/btn129

Keay RWJ (1953) Revision of the “Flora of West Tropical Africa. - Ochnaceae. *Kew Bulletin* **8**:487-492

Koffi KG, Hardy OJ, Doumenge C, Cruaud C, Heuertz M (2011) Diversity gradients and phylogeographic patterns in a widespread African tree typical of mature rainforests, *Santiria trimera* (Burseraceae). *Am J Bot* **98**:254-264 DOI:10.3732/ajb.1000220

Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I (2015) CLUMPAK : a program for identifying clustering modes and packaging population structure inferences across K. *Mol Ecol Res* **15**:1179-1191 DOI :10.1111/1755-0998.12387

Letouzey R (1968) Étude phytogéographique du Cameroun. Paris, pp 511

Letouzey R (1957) La forêt à *Lophira alata* de la zone littorale camerounaise. *Bois For Trop* **53**:9-20

Li YL, Liu JX (2018) StructureSelector: a web based software to select and visualize the optimal number of clusters using multiple methods. *Mol Ecol Res* **18**:176–177

- Lissamou B-J, Couvreur TLP, Atteke C, Stevart T, Piñeiro R, Dauby G et al. (2018) Species delimitation in the genus *Greenwayodendron* based on morphological and genetic markers reveal new species. *Taxon* (in press)
- Loiselle BA., Sork VL, Nason J, Graham C (1995) Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). *Am J Bot* **82**:1420-1425
- Lowe AJ, Harris D, Dormontt E, Dawson IK. (2010). Testing putative African tropical forest refugia using chloroplast and nuclear DNA phylogeography. *Trop Plant Biol* **3**:50–58 DOI:10.1007/s12042-010-9045-2
- Lowry DB, Rockwood R C and Willis JH (2008) Ecological reproductive isolation of coast and inland races of *Mimulus guttatus*. *Evolution* **62**:2196-2214 DOI:10.1111/j.1558-5646.2008.00457.x
- Maley J (1996) The African rainforest - main characteristics of changes in vegetation and climate from the upper cretaceous to quaternary. *Proceedings of the Royal Society of Edinburgh: Essays on the ecology of the Guineo-Congo rainforest*, Huddersfield, UK, **104B**:31-73.
- Mapongmetsem P-M (2007) *Lophira lanceolata* Tiegh. ex Keay. [Internet] Record from PROTA4U. van der Vossen HAM. and Mkamilo GS. (Editors). PROTA (Plant Resources of Tropical Africa / Ressources végétales de l'Afrique tropicale), Wageningen, Netherlands. <<http://www.prota4u.org/search.asp>>. Accessed 23 December 2019.
- Meunier Q, Moubogou C, Doucet J-L (2015) *Les Arbres Utiles du Gabon*. pp. 246
- Mondol S, Moltke I, Hart J, Keigwin M, Brown L, Stephans et al. (2015) New evidence for hybrid zones of forest and savanna elephants in Central and West Africa. *Mol Ecol* **24**:6134-6147 DOI:10.1111/mec.13472
- Morales CL, Traveset A (2008) Interspecific pollen transfer: magnitude, prevalence and consequences for plant fitness. *Crit Rev Plant Sci* **27**:221–238 DOI:10.1080/07352680802205631
- Morgan-Richards M, Trewick SA, Stringer IA (2010) Geographic parthenogenesis and the common tea-tree stick insect of New Zealand. *Mol Ecol* **19**:1227–1238 DOI:10.1111/j.1365-294X.2010.04542.x.

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89**:583-590

Odee DW, Telford A, Wilson J, Gaye A, Cavers S (2012) Plio-Pleistocene history and phylogeography of *Acacia senegal* in dry woodlands and savannahs of sub-Saharan tropical Africa: evidence of early colonisation and recent range expansion. *Heredity* **109**:372-382 DOI:10.1038/hdy.2012.52

Persinos GJ, Quimby MW (1968) Studies on Nigerian plants V. comparative anatomy of *Lophira lanceolata* and *Lophira alata*. *Econ Bot* **22**:206-220

Piñeiro R, Staquet A, Hardy OJ (2015) Isolation of Nuclear Microsatellites in the African Timber Tree *Lophira alata* (*Ochnaceae*) and Cross-Amplification in *L. lanceolata*. *Appl Plant Sci* **3**:1500-1056 DOI:10.3732/apps.1500056

Plana V, Gascoigne A, Forrest LL, Harris D, Pennington RT (2004) Pleistocene and pre-pleistocene *Begonia* speciation in Africa. *Mol Phylogenetics Evol* **31**:449-461 DOI:10.1016/j.ympev.2003.08.023

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

Puechmaille SJ (2016) The program STRUCTURE does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem, *Mol Ecol Res* **16**:608-627 DOI:10.1111/1755-0998.12512

Reed DH, Frankham R (2003) Correlation between fitness and genetic diversity. *Conserv Biol* **17**:230-237 DOI:10.1046/j.1523-1739.2003.01236.x

Rieseberg LH, Carney SE (1998) Plant hybridization. *New Phytol* **140**:599-624

Salzmann U, Hoelzmann P (2005) The Dahomey Gap: an abrupt climatically induced rainforest fragmentation in West Africa during the late Holocene. *Holocene* **15**:1-10
DOI:10.1191/0959683605hl799rp

Satabie B (1982) Le phénomène de vicariance chez deux espèces ecophylétiques au Cameroun: *Lophira alata* Banks ex Gaertn. f. et *Lophira lanceolata* Van Tiegh. ex Keay (*Ochnaceae*), PhD thesis, Université de Yaounde, Cameroon

Schemske D W (2000) Understanding the origin of species. *Evolution* **54**:1069–1073

Sexton GJ, Fere CH, Kalinganire A, Uwamariya A, Lowe AJ, Godwin ID et al. (2015) Influence of putative forest refugia and biogeographic barriers on the level and distribution of genetic variation in an African savannah tree, *Khaya senegalensis* (Desr.) A. *Tree Genet Genomes* **11**:103
DOI:10.1007/s11295-015-0933-3

Smith TB, Schneider CJ, Holder K (2001) Refugial isolation versus ecological gradients. *Genetica* **112/113**:383-398

Sobel JM, Chen GF, Watt LR and Schemske DW (2009) The biology of speciation. *Evolution* **64**:295-315
DOI:10.1111/j.1558-5646.2009.00877.x

IUCN (2014) The IUCN Red List of Threatened Species. Version 2014.3. Website www.iucnredlist.org
[accessed 25 September 2015]

Vincens A, Williamson D, Thevenon F, Taieb M, Buchet G, Decobert M et al. (2003) Pollen-based vegetation changes in southern Tanzania during the last 4200 years: climate change and/ or human impact. *Palaeogeogr Palaeoclimatol Palaeoecol* **198**:321-334

Von Thieme HW (1929) Das Bongossiholz und seine Abstammung. *Botanisches Archiv*. Band 26. Leipzig

White F (1979) The Guineo-Congolian region and its relationships to other phytochoria. *Bull Jard Bot Belg / Bull van de National Plantentuin van België* **49**:11-55

Whitlock MC (2011) G_{st} and D do not replace F_{st} . *Mol Ecol* **20**:1083–1091

Tables

Table 1: Population-level and cluster-level genetic diversity: N : sample size, Ar : allelic richness (for a subsample of 12 allele copies), H_E : expected heterozygosity, F_{IS} : inbreeding coefficient based on ratio of observed over expected heterozygosities, F_{IS}' : inbreeding coefficient accounting for null alleles following the method implemented in INEst and its posterior distribution (HDP)

Species	Population	N	$Ar(12)$	H_E	F_{IS}	F_{IS}'	HPD
<i>L. lanceolata</i>	Upper Guinea (UG)	223	2.74	0.42	0.26	0.04	0.00-0.07
	Dahomey Gap (DG)	47	2.86	0.47	0.21	0.04	0.00-0.11
	Cameroon (CM)	100	2.8	0.4	0.14	0.07	0.06-0.14
<i>L. alata</i>	Upper Guinea (UG)	86	2.54	0.37	0.09	0.03	0.00-0.08
	Dahomey Gap (DG)	7	2.24	0.33	0.36	0.31	0.00-0.51
	Lower Guinea (LG)	280	2.69	0.46	0.25	0.05	0.00-0.09
	West Gabon (WG)	60	3.18	0.58	0.18	0.06	0.00-0.14
<u>Genetic clusters</u>							
<i>L. alata 1</i>	West Gabon	60	3.18	0.58	0.18		
<i>L. alata 2</i>	Northern range	373	2.77	0.45	0.24		
<i>L. lanceolata</i>	complete range	370	2.91	0.44	0.25		

Table 2: Pairwise fixation indices between **(A)** the three major genetic clusters inferred in the genus *Lophira* and **(B)** the biogeographic regions using F_{ST} (below diagonal) or its analogue index accounting for allele sizes, R_{ST} (above diagonal). In parenthesis: Estimated Fst of Weir (1996) for each pair of clusters using the ENA correction described in Chapuis and Estoup (2007). Allele sizes permutation tests indicate whether R_{ST} is significantly larger than F_{ST} (* P-value < 0.05).

A)

$F_{ST} \setminus R_{ST}$	<i>L. alata</i> 1	<i>L. alata</i> 2	<i>L. lanceolata</i>
<i>L. alata</i> 1		0.53*	0.4
<i>L. alata</i> 2	0.37 (0.32)		0.56
<i>L. lanceolata</i>	0.38 (0.35)	0.4 (0.35)	

B)

Species :	<i>Lophira alata</i>				<i>Lophira lanceolata</i>		
Major cluster :	<i>L. alata</i> 2			<i>L. alata</i> 1	<i>L. lanceolata</i>		
Population :	Upper Guinea	Dahomey Gap	Lower Guinea	West Gabon	Upper Guinea	Dahomey Gap	Cameroon
Upper Guinea		0	0.05	0.48	0.62	0.57*	0.49
Dahomey Gap	0 (0)		0.01	0.5	0.64*	0.61*	0.47
Lower Guinea	0.06 (0.06)	0.03 (0.04)		0.52*	0.61*	0.56*	0.52*
West Gabon	0.42 (0.42)	0.39 (0.40)	0.35 (0.33)		0.49	0.36	0.28
Upper Guinea	0.49 (0.47)	0.48 (0.46)	0.43 (0.41)	0.41 (0.39)		0.05	0.09
Dahomey Gap	0.43 (0.41)	0.39 (0.38)	0.35 (0.33)	0.33 (0.33)	0.05 (0.05)		0.04
Cameroon	0.45 (0.43)	0.44 (0.42)	0.38 (0.36)	0.38 (0.38)	0.11 (0.11)	0.03 (0.03)	

Figures

Fig. 1: **(A)** Top: Location of the 433 samples of *Lophira alata* (circles) and 370 *L. lanceolata* individuals (triangles) and geographical distribution of the three genetic clusters. Dotted lines separate geographical regions of population assignment (Upper Guinea [UG], Dahomey Gap [DG], Cameroon [CM], Lower Guinea [LG], and West-Gabon [WG]). Gray and green shaded backgrounds distinguish four land cover types derived from MODIS images (Friedl et al. 2010; Channan et al. 2014). Middle: Detailed maps of contact zones between **(B)** *L. alata* and *L. lanceolata* [UG], **(C)** *L. alata* and *L. lanceolata* [CM], and **(D)** between genetic clusters of *L. alata*. F1 = F1 Hybrid. F2 = F2 Hybrid, F* = no hybrid classification possible by NewHybrids. Bottom: Hybridization index within contact zones **(E)** UG, **(F)** CM, and **(G)** between genetic clusters of *L. alata*. Filled Asterisks indicate detected hybrid individuals.

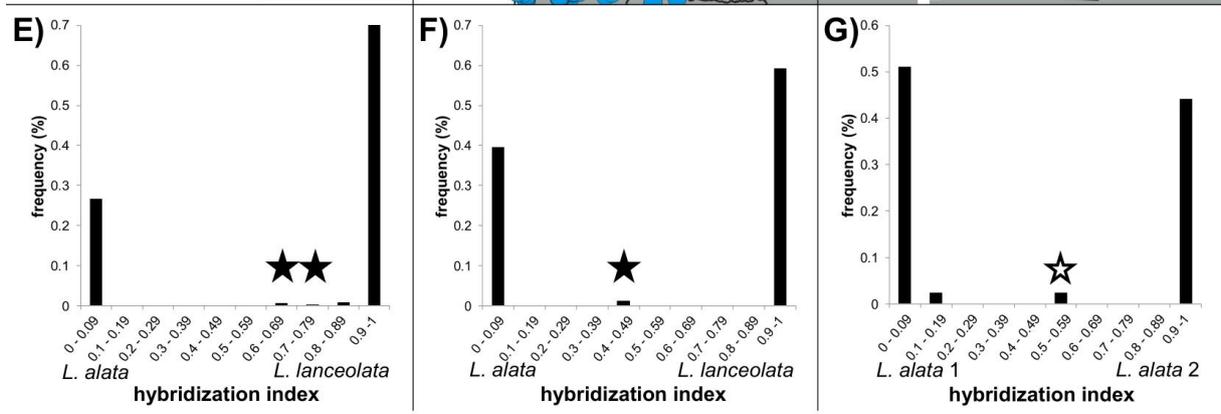
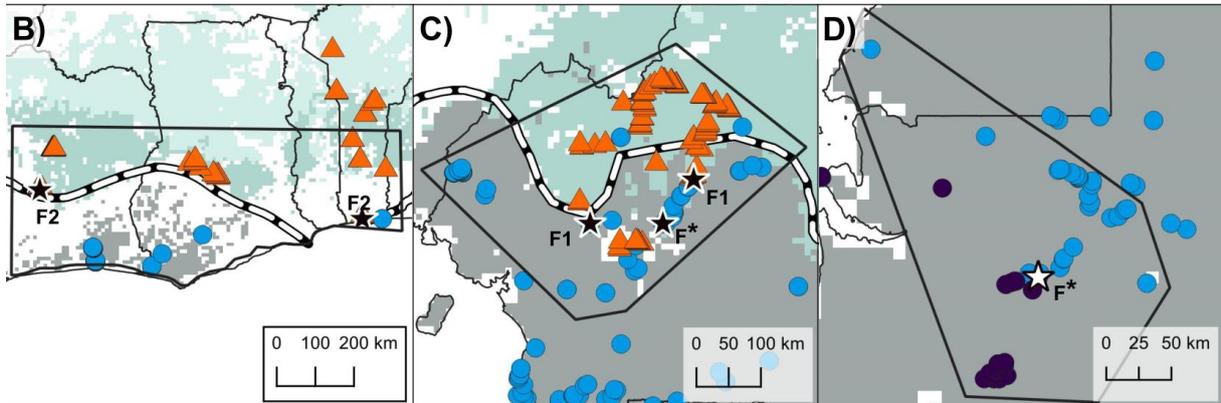
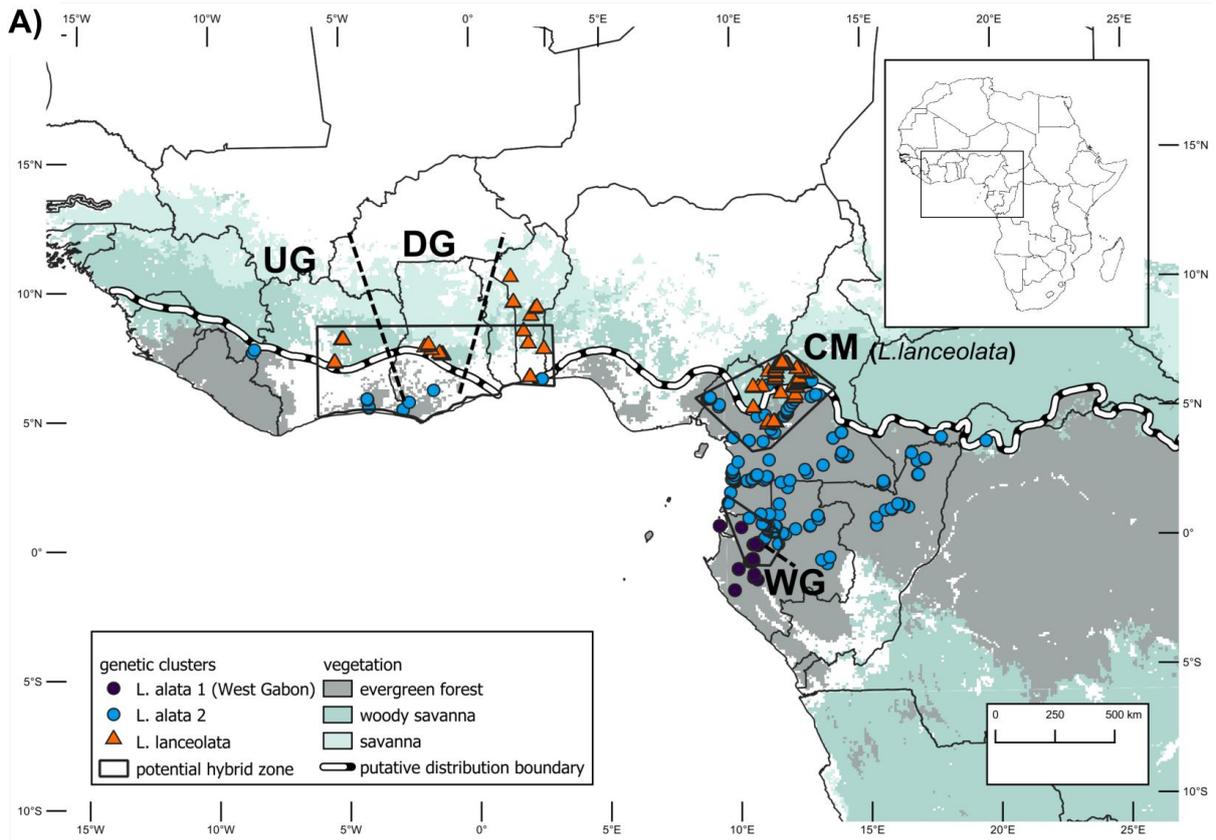


Fig 2: (A) PCA ordination of the 803 *Lophira* spp. individuals from 7 geographical regions according to their genetic structure. Axis 1 and axis 2 represent respectively 31.8 and 7.6 % of the total variance (Code: *L.alata*: [UG] Upper Guinea, [DG] Dahomey Gap, [LG] Lower Guinea, [WG] West-Gabon; *L.lanceolata*: [UG] Upper Guinea, [DG] Dahomey Gap, [CA] Cameroon)

(B) Histograms of all individuals assignment to clusters for K = 2 (top) to K = 5 (bottom) clusters (using admixture and correlated allele frequencies). **(C)** Histograms of hierarchical structure analysis using only *L.alata* cluster 2 individuals (left) and *L.lanceolata* individuals (right).

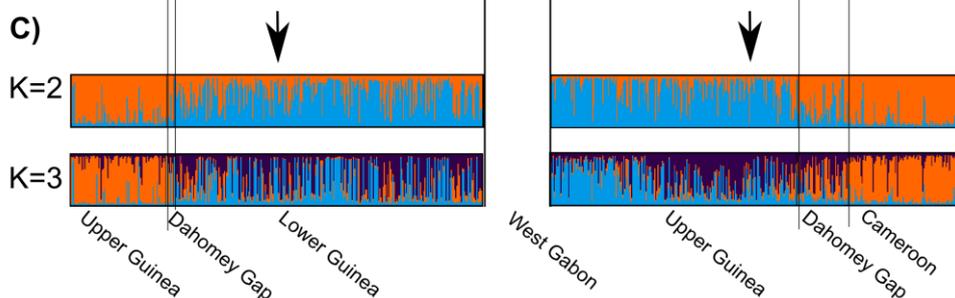
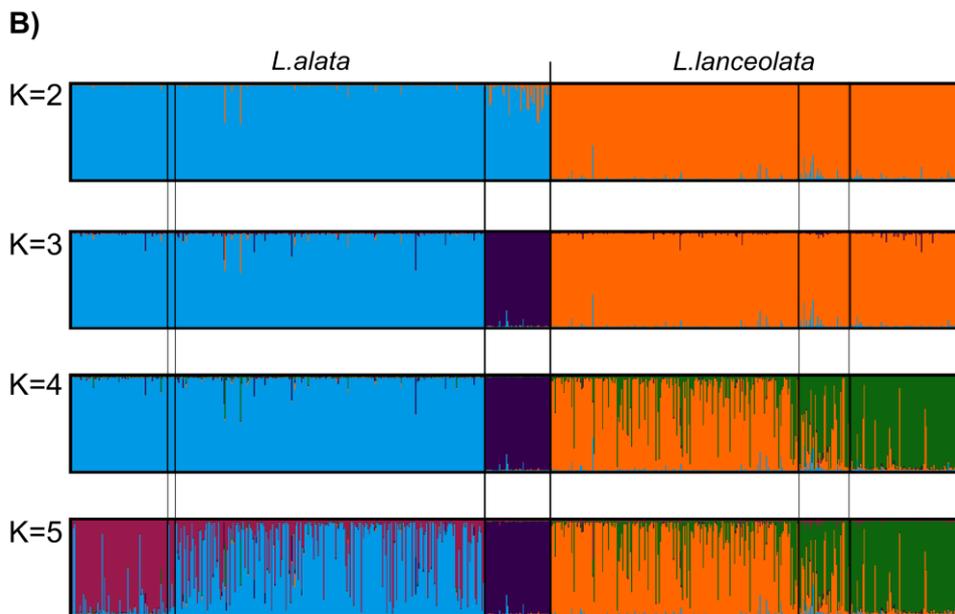
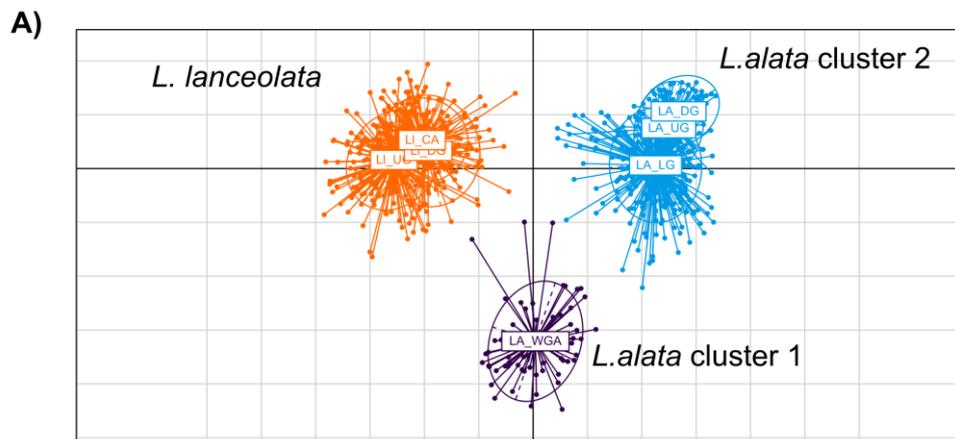
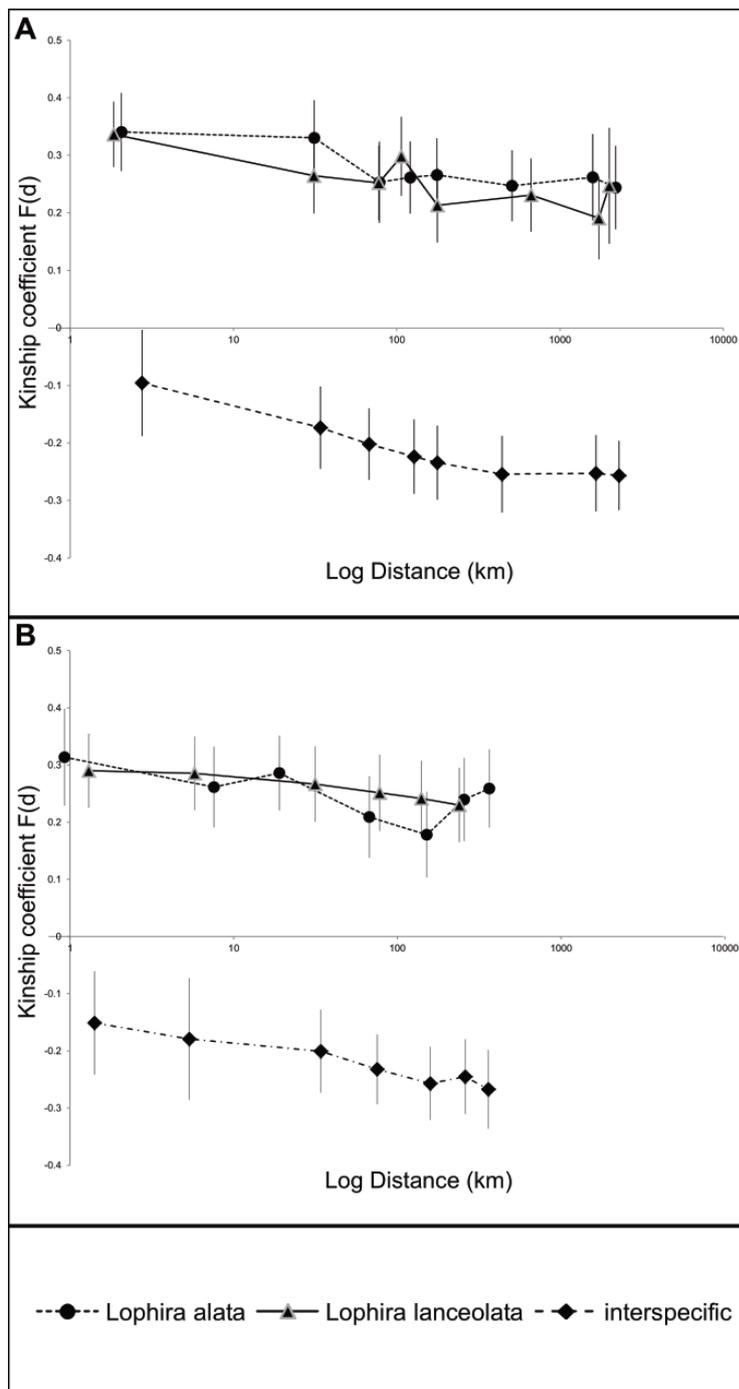
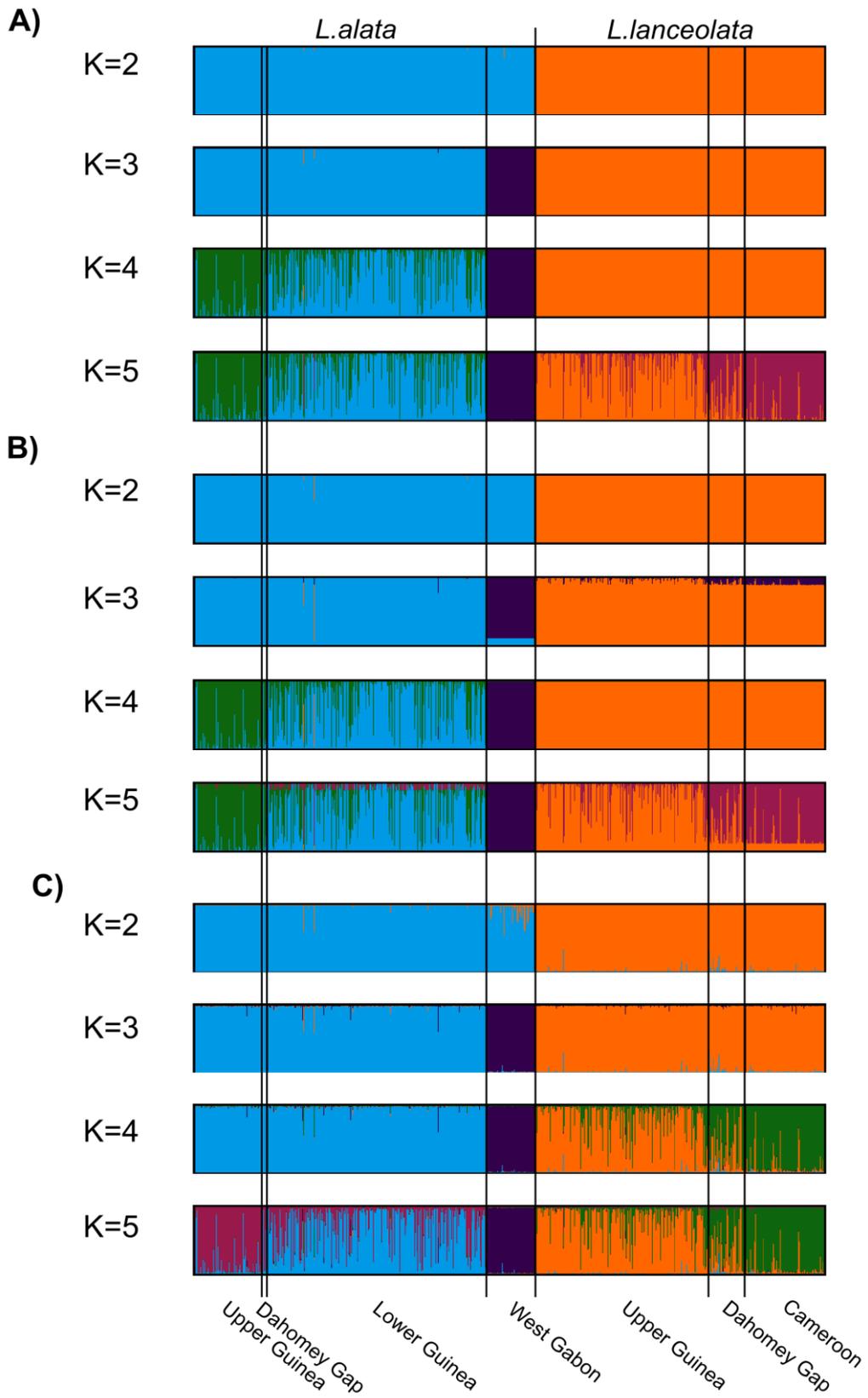


Fig. 3: Mean kinship coefficients between *Lophira alata* and *L. lanceolata* individuals according to geographical distance at **(A)** large scale (Ivory-Coast to Cameroon) and **(B)** fine scale around the forest-savannah contact zone in Cameroon. We distinguish kinship-distance curve within each species and between species. These analyses excluded two putative hybrids from the Cameroonian contact zone between species as well as samples from Gabon where *L. lanceolata* is absent.



Supplementary Material

Supplementary Online 1: Histograms of all individuals assignment to clusters for $K = 2-5$ using **(A)** no admixture and independent allele frequencies, **(B)** no admixture and correlated allele frequencies, and **(C)** admixture and independent allele frequencies.



Supplementary Online 2: List of alleles of frequency >0.05 within each genetic cluster of *Lophira*, with cluster-diagnostic alleles shown in bold, at **(A)** large scale (complete data set) and **(B)** the interspecific contact zone in Cameroon, and **(C)** between the two clusters in *L. alata*. Total number of alleles per locus.

Supplementary Online 2

Genetic cluster	Marker									
	P53	P31	P66	P12	P47	P51	P18	P34	P40	P62
A.										
<i>L. alata</i> 1	201	158	271	111	181	189	130	166	163	257
	203	161	273	117	183	191	133	169	166	259
			275		189	193	143	172	171	261
			277		191			181	176	
					193					
<i>L. alata</i> 2	203	158	275	105	181	191	133	163	176	257
	205	161	277	111	183	193	136	166	179	259
		167	279	120	185	195			181	261
		176								263
<i>L. lanceolata</i>	203	152	269	114	181	191	130	160	171	255
	205	158	271		183	193	133	163	176	257
	207		273		185		136			259
	209									
	211									
B.										
<i>L. alata</i> 2	203	158	275	105	181	191	133	163	176	257
	205	161	277	111	183	193	136	166	179	259
		167	279	120	185	195			181	261
		176								263
<i>L. lanceolata</i>	203	152	269	114	181	191	130	160	171	255
	205	158	271		183	193	133	163	176	257
	207		273		185		136			259
	209									
	211									
C.										
<i>L. alata</i> 2	203	158	275	105	181	191	133	163	176	257
	205	161	277	111	183	193	136	166	179	259
		167	279	120	185	195			181	261
		176								263
<i>L. alata</i> 1	201	158	271	111	181	189	130	166	163	257
	203	161	273	117	183	191	133	169	166	259
			275		189	193	143	172	171	261
			277		191			181	176	
					193					
total number of alleles per locus	10	9	9	8	18	10	9	8	8	8