This is the postprint of the article published in *Tree Genetics & Genomes*:

Monthe, F.K., Duminil, J., Kasongo Yakusu, E., Beeckman, H., Bourland, N., Doucet, J.-L., Sosef, M.S.M., Hardy, O.J., 2018. The African timber tree *Entandrophragma congoense* (Pierre ex De Wild.) A.Chev. is morphologically and genetically distinct from *Entandrophragma angolense* (Welw.) C.DC. *Tree Genetics & Genomes* 14, 66. https://doi.org/10.1007/s11295-018-1277-6

The African timber tree *Entandrophragma congoense* (Pierre ex De Wild.) A.Chev. is morphologically and genetically distinct from *Entandrophragma angolense* (Welw.) C.DC.

Franck K. Monthe[1$], Jérôme Duminil[1,2,3], Emmanuel Kasongo Yakusu[4,5,6], Hans Beeckman[4], Nils Bourland[4,7,8], Jean-Louis Doucet[9], Marc S.M. Sosef[10] and Olivier J. Hardy[1]

[1] Université Libre de Bruxelles, Faculté des Sciences, Service Evolution Biologique et Ecologie, CP 160/12, 50 Av. F. Roosevelt, 1050 Bruxelles, Belgium.
[2] DIADE, IRD, Univ Montpellier, 911 Avenue Agropolis, BP 64501, 34394 Montpellier, France.
[3] Bioversity International, Forest Genetic Resources and Restoration Programme, Sub-Regional Office for Central Africa, P.O. Box 2008 Messa, Yaoundé, Cameroon.
[4] Wood Biology Service, Royal Museum for Central Africa (RMCA), Leuvensesteenweg 13, 3080 Tervuren, Belgium.
[5]Ghent University, Laboratory of Wood Technology, Coupure Links 653, 9000 Ghent, Belgium
[6] Université de Kisangani, Faculté de Gestion des Ressources Naturelles Renouvelables, B.P. : 2012, Avenue Kitima, 3, Kisangani, République Démocratique du Congo.
[7] Center For International Forestry Research, Situ Gede, Bogor Barat 16115, Indonesia.
[8] Resources and Synergies Development Pte Ltd, Raffles Quay 33-03, Singapore 048581, Singapore
[9] TERRA Teaching and Research Centre, Central African Forests, BIOSE Department, Gembloux Agro-Bio Tech, Université de Liège, 2 Passage des Déportés, 5030 Gembloux, Belgium.
[10]Botanic Garden Meise, Nieuwelaan 38, 1860 Meise, Belgium.

[$] Corresponding author. Université Libre de Bruxelles, Faculté des Sciences, Service Evolution Biologique et Ecologie, CP 160/12, 50 Av. F. Roosevelt, 1050 Bruxelles, Belgium. fmonthek@ulb.ac.be Tel. +32 (0)2 650 45 11. Fax. +32 (0)2 650 24 45.

ORCID: 0000-0003-4664-658X

Abstract

Interpreting morphological variability in terms of species delimitation can be challenging. However, correcting species delineation can have strong implications for the sustainable management of exploited species. Up to now, species delimitation between two putative timber species from African forests, *Entandrophragma congoense* and *E. angolense*, remained unclear. To investigate their differences, we applied an integrated approach which combines morphological traits and genetic markers. We characterized 13 morphological characters from 81 herbarium specimens and developed 15 new polymorphic microsatellite markers to genotype 305 samples (herbarium samples and specimens collected in the field across the species distribution ranges). Principal Component Analysis (PCA) of morphological data and Bayesian clustering analyses of genetic data were used to assess differentiation between putative species. These analyses support two well-differentiated groups (F_{ST} = 0.30) occurring locally in sympatry. Moreover, these two groups present distinct morphological characters at the level of the trunk, leaflets and seeds. Our genetic markers identified few individuals (4%) that seem to be hybrids, though there is no evidence of genetic introgression from geographic patterns of genetic variation. Hence, our results provide clear support to recognize *E. congoense* as a species distinct from *E. angolense*, with a much lower genetic diversity than the latter, and that should be managed accordingly. This work highlights the power of microsatellite markers in resolving species boundaries.

Keywords: Species delimitation, microsatellite markers, Bayesian assignment, Africa rainforest, tiama, *Entandrophragma*, Meliaceae

1. Introduction

Identification and delimitation of species is of particular importance for species conservation, notably in the light of global change (Dayrat 2005; Schlick-Steiner et al. 2010). Defining species has long been contentious, leading to the development of many methods and concepts (e.g. Le Guyader 2002, De Queiroz 2007). Two species concepts focus the attention of most plant taxonomists (Le Guyader 2002). First, the 'typological species concept' defines species as a group of individuals whose members share common characteristics that differ from other species (Mayr 1992, Le Guyader 2002, De Queiroz 2007). Unfortunately, classifying individuals sharing similar morphological traits is not always obvious, especially when those individuals represent cryptic species (Janzen et al. 2017) or come from contrasted habitats (Tarasjev et al. 2009). In these cases, it may not be easy to separate species-specific traits from individual phenotypic variability, leading to an excessive splitting or lumping of species. Second, the 'biological species concept' defines species based on the interfecondity of individuals, thus the absence of reproductive isolation mechanisms (Mayr 1942). This concept can nowadays be easily investigated with the help of genetic markers able to identify interbreeding individuals using population genetics principles (Duminil and Di Michele 2009).

Tropical African rainforests exhibit a high richness of tree species (Slik et al. 2015) but comprising still many groups with a weak taxonomic framework (Sosef et al. 2017). In this context, species delimitation based on morphological characters might be difficult. Main factors as environment, phenology, growth stage can affect phenotypic variability among species (Poorter 1999; Tarasjev et al. 2009). Using molecular methods can help but they have their own drawbacks, for example hybridization can blur the delineation of species boundaries (Duminil and Di Michele 2009; Ley and Hardy 2017; Weber et al. 2017). Accordingly, applying an approach integrating morphological and genetic data is generally necessary to unravel species delimitation. Such an approach has been successful in the resolution of several plant species complexes of African forests, in some cases resulting in the description of new species (e.g., Ley and Hardy 2010; Duminil et al. 2012; Dainou et al. 2016; Ikabanga et al. 2017).

Correct species delimitation is a fundamental issue for the sustainable management of timber tree species populations (e.g. Tosso et al. 2015). The genus *Entandrophragma* (Meliaceae), described in 1894 by Casimir De Candolle, includes economically important timber species, both from rain and dry forests. The genus has undergone important taxonomic revisions, which resulted in more than 44 taxonomic names published, but today only 10 or 11 species are recognized depending on the database which recognizes, or not, *E. congoense* (Pierre ex De Wild.) A.Chev. as a synonym of *E. angolense* (Welw.) C.DC (African Plant Database 2018; The Plant List 2013). This uncertainty of the taxonomic status of an important timber species, possibly due to wide intra-specific phenotypic variability, also relates to market issues (Kasongo-Yakusu et al. 2018).
Entandrophragma angolense (Welw.) C.DC. was first described as *Swietenia angolensis* Welw. but after new observations, de Candolle (1894) declared it was not a *Swietenia* and transferred it to his new genus *Entandrophragma*. Taxonomic revisions conducted within *Entandrophragma* consider many taxa as synonyms of *E. angolense* (Chevalier 1909; Staner 1941). Up to now, more than 14 taxa, including *E. congoense*, have been assigned to this species (Kasongo-Yakusu et al. 2018). *Entandrophragma congoense* was firstly described as *Leioptyx congoensis* Pierre ex De Wild in 1908 (Sprague 1910). One year later, the species was transferred to *Entandrophragma* by Chevalier (1909) and subdivided into two distinct species : *Entandrophragma pierrei* A.Chev. and *Entandrophragma congoense* A.Chev. Staner (1941) considered these two species as synonyms of *E. angolense*. Later, Liben and Dechamps (1966) and Liben (1970) recognized *E. congoense* again as different from *E. angolense*, based on morphological characters such as the absence of developed buttresses at the base, the presence of scaly rhytidome, generally glabrous ribs, acute-apiculate leaflet apex, much smaller capsules (18 cm long, 2 cm wide and

about 3.5 cm thick) and seeds with truncated base and narrower than the wings. More recently, in a revision of the Meliaceae family in the Flora of Gabon series, de Wilde (2015) considered *E. congoense* as a distinct species and described new diagnostic floral characters. Furthermore, in logging concessions, foresters use differences in trunk aspect to distinguish individuals belonging to each species (white tiama for *E. angolense* and black tiama for *E. congoense*) (Meunier et al. 2015; J-L Doucet, pers. comm). It is worth noting that *E. congoense* is exclusively distributed in the Congo basin region, while *E. angolense* is more widely distributed throughout the African rain forest (Liben 1970; de Wilde 2015; Meunier et al. 2015). Nevertheless, doubt still persists on the status of *E. congoense* and many authors are still considering *E. congoense* as a synonym of *E. angolense* (for example Klopper et al. 2006).

In the present study, we combined morphological data and molecular markers to assess the taxonomic status of *E. congoense* and *E. angolense*. More specifically, we address the following questions. (i) Do they form distinct morphological and genetic entities suggesting the presence of two distinct species? (ii) If yes, do they hybridize and/or is there some evidence of genetic introgression that could explain the difficulty to delimit them in previous taxonomical works?

2. Material and methods

2.1. Sampling

For morphological analyses, we used 81 herbarium vouchers from different herbaria (BR and WAG—now at L—) attributed to *E. angolense* or *E. congoense* (Table 1). Samples were visually separated in two morphogroups (morphogroup A for *"E. angolense"* and morphogroup C for *"E. congoense"*) based on leaflet and seed characters following Liben and Dechamps (1966) and Liben (1970) (Appendix 1).
For genetic analyses, we used a piece of leaflet from each successfully amplified herbarium voucher. We also collected leaf or cambium (dehydrated with silica gel) from 261 adult individuals sampled in the field across the Guineo-Congolian forest area (Fig. 1A), among which 88 specimens sampled within the Forest Stewardship Council (FSC)-certified logging concession granted to 'Pallisco' in Eastern Cameroon (mean coordinates: 13.37°E, 3.29°N; Fig. 1B) where both taxa would occur (de Wilde 2015; J-L Doucet, comm. pers.). These specimens were also separated in two morphogroups based on trunk aspect, slash characters and leaflet shape (Appendix 1). Genetic analyses were performed at two spatial scales: the whole Guineo-Congolian forest (all 261 individuals, maximal distance between samples c. 4000 km) and within the Pallisco forest concession (88 individuals, maximal distance between samples c. 5 km).

2.2. Morphometric traits and analyses

To confirm objectively the morphological differentiation between morphogroups, for each of the 81 herbarium samples we observed and measured 13 morphological traits indicated by previous authors as being diagnostic. First, three qualitative traits: apex (acute-apiculate or obtuse and exceptionally retuse and mucronate), median vein (glabrous or pubescent) and domatia (thick tuft or generaly absent). Then, three quantitative variables related to leaflets size and numbers. Finally, seven traits associated to fruits, but which were available for only five samples of each species (Table 2, Appendix 1). We performed a Hill-Smith ordination, an extended principal component analysis (PCA) for datasets containing both qualitative and quantitative variables, on the vegetative traits of the 81 samples applying the function "dudi.hillsmith" of the Ade4 package available in R 3.1.2 (Chessel et al. 2004; Dray et al. 2007). For all quantitative leaf and fruit variables, we tested the difference between morphogroups by Welch Two Sample t-tests using the package MASS available in R 3.1.2.

2.3. Molecular genetic analyses

2.3.1. DNA extraction, microsatellites markers development and genotyping

Total genomic DNA was extracted using the cetyltrimethylammonium bromide (CTAB) protocol (Doyle and Doyle 1987) for herbarium specimens and the NucleoSpin 96 Plant II kit (Macherey-Nagel, Duren, Germany) for dry plant material collected in the field. Fifteen microsatellites loci were developed from a genomic library of a sample of *E. angolense* (GEM10; Table 1) using the protocol described in Monthe et al. (2017) developed for two other *Entandrophragma* species. The microsatellite loci were amplified in three multiplexes developed following the protocol of Micheneau et al. (2011). These multiplexes named "Mix 1", "Mix 2" and "Mix 3" were, respectively, composed of six (EnA-ssrEnA-ssr7, EnA-ssr2, EnA-ssr35, EnA-ssr23, EnA-ssr14, EnA-ssr48), five (EnA-ssr5, EnA-ssr34, EnA-ssr21, EnA-ssr42, EnA-ssr36) and four (EnA-ssr3, EnA-ssr29, EnA-ssr15 and EnA-ssr44) microsatellite markers (Table 3). PCR amplification was performed in a total volume of 15 µL containing 0.3 µL of the reverse (0.2 µM form 100 µM initial concentration) and 0.1 µL of the forward (0.07 µM form 100 µM initial concentration) primers with a Q1–Q4 universal sequence at the 5′ end (see Table 3), 0.15 µL of Q1–Q4 labelled primer (0.2 µM each), 7.5 µL of Type-it Microsatellite PCR Kit (QIAGEN), 15µL of H_2O, and 1.5 µL of DNA extract. PCR program conditions were: 95°C for 3 min; 30 PCR cycles of 95°C for 30 s / 57°C for 90 s / 72°C for 1 min; and 60°C for 30 min. Using 1 µL of PCR product, 12 µL of Hi-Di Formamide (Life Technologies, Carlsbad, California, USA), and 0.3 µL of MapMarker 500 labelled with DY-632 (Eurogentec, Seraing, Belgium). Loci successfully cross-amplified between morphogroups. Among the 81 herbarium samples, only 44 successfully amplified, so that a total of 305 individuals were genotyped using an ABI3730 sequencer (Applied Biosystems, Lennik, The Netherlands; ULB-EBE platform). The data generated for each individual were scored using the microsatellite plugin in Geneious 9.1.6 (Kearse et al. 2012). The first screening revealed that all samples were diploid as no more than two alleles per individual and per locus were found.

2.3.2. Population genetic analyses

The genetic structure was investigated through (i) a Bayesian clustering algorithm implemented in STRUCTURE v.2.3.4 (Pritchard et al. 2000), (ii) a principal coordinate analysis (PCoA) on pairwise genetic distances between samples, performed using GenAlEx v.6.5 (Peakall and Smouse 2012).

Considering all 305 samples, we applied Bayesian clustering using the admixture and the correlated allele frequency model, declaring the presence of null alleles for all loci, without any location or population priors. We tested K=1 to 10 genetic clusters with runs of 500,000 MCMC generations (burnin period of 100,000 generations) and 10 runs for each K value. The online application STRUCTURE HARVESTER (http://taylor0.biology.ucla.edu/structureHarvester/) was used to compute and plot the deltaK statistic against the range of K values (Evanno et al. 2005; Earl et al. 2012). The optimum number of genetic clusters (K = 2, see results) was identified considering the important gain in likelihood as K increases. Each individual was assigned to a genetic cluster when its probability of assignment to the most likely cluster, *q*, was higher than 0.9, while the remaining individuals were considered as unassigned.

For each identified cluster (hereafter called A and C given their correspondence with the A and C morphogroups), the following genetic diversity indices were computed for each locus using all samples (global scale): the number of alleles (*A*), the observed heterozygosity (H_o), the expected heterozygosity (H_e), the inbreeding coefficient (*F*), using SPAGeDi 1.5 (Hardy and Vekemans 2002). Null allele frequencies (*r*) were estimated for each genetic cluster in STRUCTURE. We used INEST 2.0 under a Population Inbreeding Model to estimate $F_{(null)}$, an unbiased estimator of inbreeding coefficient robust to the presence of null alleles (Chybicki and Burczyk 2009). These analyses were repeated at the local scale (Pallisco) to factor out the potential impact of a phylogeographic structure occurring within each genetic cluster. We also assessed the differentiation between genetic clusters through the estimation of fixation

indices (F_{ST} and R_{ST}), and tested whether stepwise mutations contributed to genetic differentiation (test if $R_{ST} > F_{ST}$; Hardy et al. 2003), using SPAGeDi 1.5 (Hardy and Vekemans 2002).
To evaluate whether unassigned individuals could represent hybrids, we simulated new genotypes under random mating based on the allele frequencies inferred by the STRUCTURE algorithm under K=2. To this end, we generated 151 genotypes from cluster A and 77 genotypes from cluster C (sampling randomly two alleles per locus following the allele frequencies of cluster A or C, respectively), to respect the same proportions as in the real data set, and 50 hybrid genotypes (sampling randomly one allele from cluster A and one from cluster C following the respective allele frequencies). The 278 simulated genotypes were then analysed in STRUCTURE under K=2 using the same parameters as described above to assess the distribution of *q* values for each category of genotypes. To further verify the occurrence of hybrids, this time only at a local scale (88 individuals from Pallisco), we applied the NewHybrids method (Anderson and Thompson 2002) under the 'Jeffrey priors' settings assuming six genotype frequency categories: purebred Cluster A (A-A), purebred Cluster C (C-C), F1 hybrids (F1), F2 hybrids (F2), backcrossed F1 to purebred Cluster A (BckA-A) and backcrossed F1 to purebred Cluster C (BckC-C).
To test introgression between genetic clusters, we computed pairwise kinship coefficients between 208 individuals sampled in Central Africa (where the two clusters are sympatric), keeping only 17 random samples attributed to cluster C in Pallisco to better balance samples sizes, and excluding unassigned samples. To this end, we used SPAGeDi 1.5 (Hardy and Vekemans 2002; estimator of J. Nason) to describe patterns of isolation by distance from the kinship-distance curves computed (i) within each genetic cluster and (ii) between the two clusters, using the mean allele frequencies observed between the two clusters as reference to estimate kinship coefficients.

3. Results

3.1. Morphometric differentiation between species

Considering the quantitative and qualitative traits observed on the 81 herbarium specimens, the first two axes of the Hill-Smith ordination summarized 53% of the total variance and allowed to distinguish two groups of samples segregating along the first axis (Fig. 2). All but three samples of morphogroup A showed negative scores along axis 1, while all but two samples of morphogroup C showed positive scores along axis 1. The Welch Two Sample t-test revealed significant differences between morphogroups for two quantitative leaf traits (number of leaflets per leaf and the length/width ration of leaflets were higher in samples attributed to *E. congoense*) and two fruits traits (the length and the width of capsules were higher in samples attributed to *E. angolense*) (Table 2). We also observed an important difference in seed base and wing (straight in morphogroup C and more curved in morphogroup A). Concerning leaflet characteristics, the specimens of morphogroup A are characterized by pubescent veins, a generally rounded apex, obovate leaflets and an absence of pilosity between the main and secondary veins, and they showed high contribution on the first component (Table S1). Specimens of morphogroup C exhibited glabrous veins, an acute apex, long leaflets, carrying thick tufts of hairs (domatia) between the main and secondary veins.

3.2. Development of microsatellite markers

In *E. angolense*, 15 highly polymorphic microsatellites markers were successfully developed. We observed 10 to 21 alleles per locus, with *A* ranging from 10 to 21 alleles and H_E from 0.72 to 0.91 (mean H_E = 0.85) among samples attributed to *E. angolense* (Table 4). These markers globally amplified on individuals attributed to *E. congoense* but four of them were monomorphic and genetic diversity was globally much lower, with *A* ranging from 1 to 12 alleles and H_E from 0 to 0.87 (mean H_E = 0.31; Table 4). Substantial heterozygote deficit was observed at most loci in both taxa and were at least partially explained by the presence of null alleles (Table 4). However, an analysis performed at the local scale

(Pallisco) showed that the inbreeding coefficient was null in each taxon after factoring out the impact of null alleles (Table S2).

3.3. Inferred genetic clusters and correspondence with morphogroups

The Bayesian clustering analysis indicated that the likelihood of the data increased substantially from K=1 to K=2, and moderately at higher K, hence we conclude that the most likely number of clusters is two, consistent with the maximum deltaK statistic at K=2 (Fig. S1). There was a 96% correspondence between a priori identification of morphogroups and genetic clusters. At a threshold of $q \geq 0.9$ we observed that 15 individuals, representing 4% of the total sample, were not assigned to a genetic cluster (Fig. 3). At a local scale in S-E Cameroon ('Pallisco'), where the two genetic clusters sampled are distributed in sympatry (Fig. 1), seven of the 88 (8%) individuals were unassigned.

The STRUCTURE analysis applied on the simulated genotypes at K=2 showed that the thresholds of q values for pure parental species were $q > 0.80$ and $q < 0.20$ and that all F1 hybrids showed $0.20 < q < 0.80$ (Fig. 3B). The Newhybrids approach applied in the contact area ('Pallisco') correctly identified the two groups at $q > 0.90$, 94% of individuals belonging to category (A-A) were all assigned to the expected group except for one individual ($q > 0.72$). In the (C-C) category, all individuals were correctly assigned at $q > 0.90$. The seven putative hybrid individuals according to STRUCTURE analysis were identified as F2 hybrids, whereas no F1 and no backcrossed individuals were identified by NewHybrids.

Results from PCoA analysis performed with GenAlEx were consistent with the Bayesian clustering analysis: the two main genetic clusters segregated along axis 1 while 15 unassigned samples ($0.1 < q < 0.90$) had intermediate scores (Fig. 4).

The genetic differentiation between clusters was high ($F_{ST} = 0.30$), and the corresponding index accounting for microsatellite allele sizes was even higher ($R_{ST} = 0.48$). Allele size permutation tests (Hardy et al. 2003) revealed that four loci (EnA-ssr3, EnA-ssr23, EnA-ssr5, EnA-ssr42) showed significant shift in allele sizes between the two clusters (single-locus R_{ST} significantly larger than single-locus F_{ST}). Four other loci (EnA-ssr14, EnA-ssr48, EnA-ssr36, EnA-ssr15) were polymorphic in Cluster A but mostly monomorphic in Cluster C (see Fig. S2). These main differences are supported by the much higher genetic diversity and allelic richness in Cluster A compared to cluster C (see Fig. S2, Table 4). On the other hand, identification of diagnostic alleles (frequency ≥ 0.30 in one cluster and below 0.10 in the other cluster) reveals that Cluster C has more diagnostic alleles (14) than Cluster A (4).

The kinship-distance curves for pairs of samples of the same genetic cluster (A-A or C-C pairs) decay with distance, indicating isolation by distance within each cluster (Fig. 5). However, the curve for pairs of samples from different clusters (A-C pairs) shows negative pairwise kinship coefficients without any trend with spatial distance (Fig. 5). Regression slopes of pairwise kinship coefficients on the logarithm of spatial distance equal $b \pm SE = -0.009 \pm 0.005$ for C-C pairs, -0.018 ± 0.002 for A-A pairs and 0.001 ± 0.003 for A-C pairs, where A and C indicate the genetic clusters of pairs of samples compared.

4. Discussion

4.1. Morphological differentiation

Our analysis of morphological traits collected on herbarium specimens assigned to *E. angolense* or *E. congoense* confirmed that they can be morphologically differentiated based on characters of leaves and fruits (flowers were not observed in our study) (Fig. 2, Table 2). According to Liben and Dechamps (1966), other diagnostic traits concern the maximal dimensions of the tree (up to 50 m high and 200 cm DBH, diameter at breast height, in *E. angolense*, compared to ≤45 m high and ≤90 cm in DBH in *E. congoense*), and the trunk base which is smooth to scaly in *E. angolense*, and generally cracked leading to rectangular elongated scales in *E. congoense*. The seeds are described as larger in *E. angolense*, although we did not observe significant differences, probably due to our low sample sizes. However, the

wing shape appears to be quite different (straight in *E. congoense* and more curved in *E. angolense*; Fig. S3). Moreover, in a recent revision for "Flore du Gabon", de Wilde (2015) mentions small differences in the staminal tube length (3-4 mm in *E. angolense*, 2-3 mm in *E. congoense*). Despite these observations, the presence of intermediate individuals (Fig. 2) can be explained by the limiting discriminating power of the variables used and/or the presence of hybrids. Unfortunately, none of the herbarium samples showing intermediate scores on axis one of the Hill-Smith ordination (Fig. 2) was successfully genotyped so that we cannot confirm if they corresponded to genetic hybrids.

4.3. Population genetics-based species delimitation

The present study confirms the validity of morphological characters described in Liben and Dechamps (1966) to distinguish the two taxa. However, in many cases morphological characteristics have showed their limits to confirm the distinction of a taxon at species level (Edwards and Knowles 2014).

We developed 15 nuclear microsatellite markers to test the differentiation between the putative species. Interestingly, the markers developed from a sample attributed to *E. angolense* amplified on *E. congoense* samples, while cross-amplification of microsatellite markers between other *Entandrophragma* species often failed (Monthe et al. 2017), indicating the comparatively close phylogenetic relationship between *E. angolense* and *E. congoense*. However, microsatellites markers show heterozygote deficit due to null alleles (Table S2), and probably also due to Wahlund effect considering the wide distribution range of our samples.

The Bayesian clustering and PCoA analysis support two well-differentiated genetic clusters corresponding to the two putative species (Figs. 3 and 4). Cluster A (*E. angolense*) displayed more alleles and much higher genetic diversity indices (H_E = 0.85) than Cluster C (*E. congoense*; H_E = 0.31; Table 4). This difference in levels of polymorphism probably reflects a difference of genetic diversity between species but could also result from an ascertainment bias given that microsatellites were developed from an *E. angolense* individual. However, while allele sizes were on average smaller in Cluster C than in Cluster A at five loci, the reverse pattern occurred at four other loci where polymorphism was generally lower in Cluster C (Fig. S2). Hence, a significant ascertainment bias (i.e. a selection of longer and more variable microsatellite loci in *E. angolense*) seems unlikely. The origin of the relatively low genetic diversity of *E. congoense* should probably be searched in its demographic history and would justify further research.

We observe a high genetic differentiation between the two clusters (F_{ST} = 0.30). Moreover, the genetic differentiation is also well marked at a local scale as we can distinguish the two clusters distributed in sympatry in a forest concession (Fig. 4). While most loci displayed several alleles shared between species, with the notable exception of locus EnA-ssr21 which was fixed in *E. congoense* for an allele not found in *E. angolense*, many loci also showed alleles at high frequency in one cluster and (near) absent in the other (Fig. S2). Some loci also displayed a global shift of allele size ranges between species (e.g. EnA-ssr 3, 5, 14, 21, 23, 42; Fig. S2), resulting in a signal whereby R_{ST} is significantly larger than F_{ST}, which implies long-term differentiation due to the accumulation of stepwise mutations (Hardy et al. 2003). Overall, morphological and genetic analyses give strong support for the recognition of two species: *E. congoense* and *E. angolense*.

4.4. Evidence of hybridization but not of introgression between species

Despite the strong genetic differentiation between the clusters, our genetic clustering analyses indicate the presence of putative hybrid individuals that were found only in regions where the two species co-occur (Fig. 1). The presence of hybrids was observed in STRUCTURE analysis ($0.11 < q < 0.88$) and confirmed by applying the NewHybrids method at a local scale, where around 8% of individuals appear to represent a hybrid (Fig. 3). This is relatively low compared to hybridization rates reported for other contact zones of congeneric African plant species: 13 - 40% for *Haumania* (Ley and Hardy 2017), 12%

for *Milicia* (Daïnou et al. 2017). Unfortunately, we were not able to assess the morphological characteristics of genetic hybrids, as the latter were individuals collected in the field without herbarium vouchers. Additional investigations on hybrid specimens are needed to find out if they are morphologically intermediate.

Hybrids observed here may results from occasional hybridization events between the clusters, a phenomenon frequently observed between other closely related plant species with overlapping distribution ranges (Haselhorst and Buerkle 2011; Duminil et al. 2012; Dainou et al. 2016; Ikabanga et al. 2017). The presence of hybrids does not necessarily imply gene flow between clusters (i.e. introgression) because hybrids may be sterile or unable to back-cross with either parental cluster. Surprisingly, according to Newhybrids analyses, the seven hybrids detected appeared to be second-generation hybrids (F2 hybrids). The absence of F1 hybrids may be due to our limited sample size, however, the absence of back-crosses with a parental cluster is consistent with the absence of evidence of genetic introgression. Indeed, if gene flow occurred regularly between the clusters in contact zones, we would expect pairs of individuals from different clusters to be on average more genetically related when sampled in the same contact zone than when sampled far apart (Hardy et al. 2001), which is not the case (Fig. 5). Hence, further research is needed to understand the underlying mechanisms (e.g. pre-zygotic isolation due to phenological shift, selection against introgressed genotypes) explaining such observation.

5. Conclusion

Our combined morphological and genetic approach confirmed that morphogroups A and C constitute distinct taxa that can be identified using the morphological characters described by Liben and Dechamps (1966). Our work therefore also confirms the correctness of the differentiation made between *E. angolense* and *E. congoens*e in the recent revision of Meliaceae in Flore du Gabon (de Wilde 2015). Although occasional hybridization events do occur, these do not cause significant genetic introgression. Hence, because of the fair number of morphological differences and the strong genetic signal, we conclude a distinction at species level is most appropriate. The recognition of *E. congoense* as a distinct species implies that its populations must be managed separately from those of *E. angolense*. This can be easily implemented as field technicians in forestry concessions are already used to distinguish these species. The much lower genetic diversity found in *E. congoense* may also have management implications, but the origin of this feature must still be understood.

Acknowledgments

We are grateful to Meise Botanic Garden (Belgium), the Herbarium and African Botanical Library (BRLU) of Université Libre de Bruxelles (ULB) and Naturalis Biodiversity Center (Leiden, the Netherlands), for making the herbarium specimens available. We thank the forest company PALLISCO for its support to field sampling, the Fonds pour la Formation à la Recherche dans l'Industrie et l'Agriculture (FRIA-FNRS, Belgium) through a PhD grant to the first author, the Fonds de la Recherche Scientifique (FRS-FNRS) through grant X.3040.17, and the Belgian Science Policy (Belspo) for funding the project AFRIFORD. We are also grateful to the DynAfFor and P3FAC projects, funded by the Fonds Français pour l'Environnement Mondial, which facilitated access to the field, and to two anonymous referees for their comments.

Data archiving statement

The microsatellite markers are submitted to GenBank (Table 3).

References

African Plant Database - version 3.4.0 (2018). Conservatoire et Jardin botaniques de la Ville de Genève and South African National Biodiversity Institute, Pretoria. Retrieved June 2018, from <http://www.ville-ge.ch/musinfo/bd/cjb/africa/>.

Anderson EC, Thompson EA (2002) A model-based method for identifying species hybrids using multilocus genetic data. Genetics 160:1217–29

http://www.ville-ge.ch/musinfo/bd/cjb/africa

Chessel D, Dufour AB, Thioulouse J, others (2004) The ade4 package-I-One-table methods. R news 4:5–10

Chevalier A (1909) Les végétaux utiles de l'Afrique tropicale française: études scientifiques et agronomiques. Dépot des publications

Chybicki IJ, Burczyk J (2009) Simultaneous estimation of null alleles and inbreeding coefficients. J Hered 100:106–113. doi: 10.1093/jhered/esn088Culley TM, Weller SG, Sakai AK, Putnam KA (2008) Characterization of microsatellite loci in the Hawaiian endemic shrub *Schiedea adamantis* (Caryophyllaceae) and amplification in related species and genera. Mol Ecol Resour 8:1081–1084

Dainou K, Blanc-Jolivet C, Degen B, Kimani P, Ndiade-Bourobou D, Donkpegan AS, Tosso F, Kaymak E, Bourland N, Doucet JL, 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. doi: 10.1186/s12862-016-0831-9

Daïnou K, Flot J-F, Degen B, Blanc-Jolivet C, Doucet J-L, Lassois L, Hardy OJ (2017) DNA taxonomy in the timber genus *Milicia*: evidence of unidirectional introgression in the West African contact zone. Tree Genet Genomes 13:90. doi: 10.1007/s11295-017-1174-4

Dayrat B (2005) Towards integrative taxonomy. Biol J Linn Soc 85:407–415

de Candolle A (1813) Théorie élémentaire de la botanique: ou, Exposition des principes de la classification naturelle et de l'art de décrire et d'étudier les végétaux. Déterville, Paris.

De Queiroz K (2007) Species concepts and species delimitation. Syst Biol 56:879–886

de Wilde JJFE (2015) Meliaceae. In: Sosef MSM, Florence J, Ngok BL, Bourobou BHP (eds) Flore du Gabon 47:5-74. Margraf Publishers, Wekersheim

Doyle JJ, Doyle JL (1987) CTAB DNA extraction in plants. Phytochem Bull 19:11–15

Dray S, Dufour A-B, others (2007) The ade4 package: implementing the duality diagram for ecologists. J Stat Softw 22:1–20

Duminil J, Di Michele M (2009) Plant species delimitation: a comparison of morphological and molecular markers. Plant Biosyst 143:528–542

Duminil J, Kenfack D, Viscosi V, Grumiau L, Hardy OJ (2012) Testing species delimitation in sympatric species complexes: the case of an African tropical tree, *Carapa* spp.(Meliaceae). Mol Phylogenet Evol 62:275–285

Earl DA, VonHoldt BM, von Reumont B (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4:359–361. doi: 10.1007/s12686-011-9548-7

Edwards DL, Knowles LL (2014) Species detection and individual assignment in species delimitation: can integrative data increase efficacy? Proceedings Biol Sci 281:20132765. doi: 10.1098/rspb.2013.2765

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

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, Charbonnel N, Fréville H, Heuertz M (2003) Microsatellite allele sizes: a simple test to assess their significance on genetic differentiation. Genetics 163:1467–82

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

Haselhorst M, Buerkle C (2011) Detection of Hybrids in Natural Populations of *Picea glauca* and *Picea* e*nglemannii*. Univ Wyoming Natl Park Serv Res Cent Annu Rep 33

Ikabanga DUDU, Stévart T, Koffi KGG, Monthé FK, Doubindou ECN, Dauby G, Souza A, M'BATCHI B, Hardy O. (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

Janzen DH, Burns JM, Qian Cong, Hallwachs W, Dapkey T, Manjunath R, Hajibabaei M, Hebert PDN, Grishin NV (2017) Nuclear genomes distinguish cryptic species suggested by their DNA barcodes and ecology. Proc Natl Acad Sci U S A 114:8313–8. doi: 10.1073/pnas.1621504114

Kasongo-Yakusu E, Monthe FK, Nils B, Hardy OJ, Louppe D, Lokanda FBM, Hubau W, Kahindo J-MM, Van Den Bulcke J, Van Acker J, Beeckman H (2018) Le genre *Entandrophragma* (Meliaceae) : taxonomie et écologie d'arbres africains d'intérêt économique (synthèse bibliographique). Biotechnol Agron société Environ 22:1–15

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A (2012) Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647–1649. doi: 10.1093/bioinformatics/bts199

Klopper RR, Chatelain C, Habashi C, Gautier L, Spichiger R-E (2006) Checklist of the flowering plants of sub-Saharan Africa: An index of accepted names and synonyms. Sabonet

Le Guyader H (2002) Doit-on abandonner le concept d'espèce? Le Courr l'environnement l'INRA 51–64

Ley A, Hardy O (2010) Species delimitation in the Central African herbs *Haumania* (Marantaceae) using georeferenced nuclear and chloroplastic DNA sequences. Mol Phylogenet Evol. doi: 10.1016/j.ympev.2010.08.027

Ley AC, Hardy OJ (2017) Hybridization and Asymmetric Introgression after Secondary Contact in Two Tropical African Climber Species, *Haumania danckelmaniana* and *Haumania liebrechtsiana* (Marantaceae). Int J Plant Sci 178:421–430. doi: 10.1086/691628

Liben L (1970) La répartition géographique d'*Entandrophragma congoënse* (De Wild.) A. Chev.(Meliaceae). Bull du Jard Bot Natl Belgique/Bulletin van Natl Plantentuin van Belgie 299–300

Liben L, Dechamps R (1966) *Entandrophragma congoense* (De Wild.) A. Chev. espèce méconnue du Congo. Bull du Jard Bot l'Etat, Bruxelles/Bulletin van den Rijksplantentuin, Brussel 415–424

Martinez-Meier A, Sanchez L, Dalla-Salda G, Gallo L, Pastorino M, Rozenberg P (2009) Ring density record of phenotypic plasticity and adaptation to drought in Douglas-fir. For Ecol Manage 258:860–867. doi: 10.1016/J.FORECO.2009.03.021

Mayr E (1992) Species concepts and their application. Units Evol MIT 15–25

Mayr E (1942) Systematics and the origin of species, from the viewpoint of a zoologist. Harvard University Press

Meunier Q, Moumbogou C, Doucet J-L (2015) Les arbres utiles du Gabon. Presses Agronomiques de Gembloux

Micheneau C, Dauby G, Bourland N, Doucet JL, Hardy OJ (2011) Development and characterization of microsatellite loci in *Pericopsis elata* (Fabaceae) using a cost-efficient approach. Am J Bot 98:e268-70. doi: 10.3732/ajb.1100070

Monthe F, Duminil J, Tosso F, Migliore J, Hardy OJ (2017) Characterization of microsatellite markers in two exploited African trees, *Entandrophragma candollei* and *E. utile* (Meliaceae). Appl Plant Sci 5:1600130. doi: 10.3732/apps.1600130

Peakall R, Smouse PE (2012) GenALEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics 28:2537–2539. doi: 10.1093/bioinformatics/bts460

Poorter L (1999) Growth responses of 15 rain-forest tree species to a light gradient: the relative importance of morphological and physiological traits. Funct Ecol 13:396–410. doi: 10.1046/j.1365-2435.1999.00332.x

Pritchard JK, Stephens M, Donnelly P (2000) Inference of Population Structure Using Multilocus Genotype Data. Genetics 155:

Schlick-Steiner BC, Steiner FM, Seifert B, Stauffer C, Christian E, Crozier RH (2010) Integrative taxonomy: a multisource approach to exploring biodiversity. Annu Rev Entomol 55:421–438

Slik JWF, Arroyo-Rodríguez V, Aiba S-I, Alvarez-Loayza P, Alves LF, Ashton P, Balvanera P, Bastian ML, Bellingham PJ, van den Berg E, Bernacci L, da Conceição Bispo P, Blanc L, Böhning-Gaese K, Boeckx P, Bongers F, Boyle B, Bradford M, Brearley FQ, et al (2015) An estimate of the number of tropical tree species. Proc Natl Acad Sci U S A 112:7472–7. doi: 10.1073/pnas.1423147112

Sosef MSM, Dauby G, Blach-Overgaard A, van der Burgt X, Catarino L, Damen T, Deblauwe V, Dessein S, Dransfield J, Droissart V, Duarte MC, Engledow H, Fadeur G, Figueira R, Gereau RE, Hardy OJ, Harris DJ, de Heij J, Janssens S, Klomberg Y, Ley AC, Mackinder BA, Meerts P, van de Poel JL, Sonké B, Stévart T, Stoffelen P, Svenning J-C, Sepulchre P, Zaiss R, Wieringa JJ, Couvreur TLP (2017) Exploring the floristic diversity of tropical Africa. BMC Biology 15:15. doi: 10.1186/s12915-017-0356-8

Sprague TA (1910) *Entandrophragma*, *Leioptyx* and *Pseudocedrela*. Bull Misc Inf (Royal Gard Kew) 1910:177–182

Staner P (1941) Les Méliacées du Congo Belge. Bull du Jard Bot l'Etat, Bruxelles/Bulletin van den Rijksplantentuin, Brussel 16:109–251

Tarasjev A, Barisić Klisarić N, Stojković B, Avramov S (2009) Phenotypic plasticity and between population differentiation in Iris pumila transplants between native open and anthropogenic shade habitats. Genetika 45:1078–86

The Plant List (2013). Version 1.1. Retrieved June 2018, from http://www.theplantlist.org/

Tosso F, Dainou K, Hardy OJ, Sinsin B, Doucet J-L (2015) Le genre *Guibourtia* Benn., un taxon à haute valeur commerciale et sociétale (synthèse bibliographique). Biotechnol Agron société Environ 19:71–88

Weber AA-T, Stöhr S, Chenuil A (2017) Species delimitation in the presence of strong incomplete lineage sorting and hybridization. bioRxiv 240218. doi: 10.1101/240218

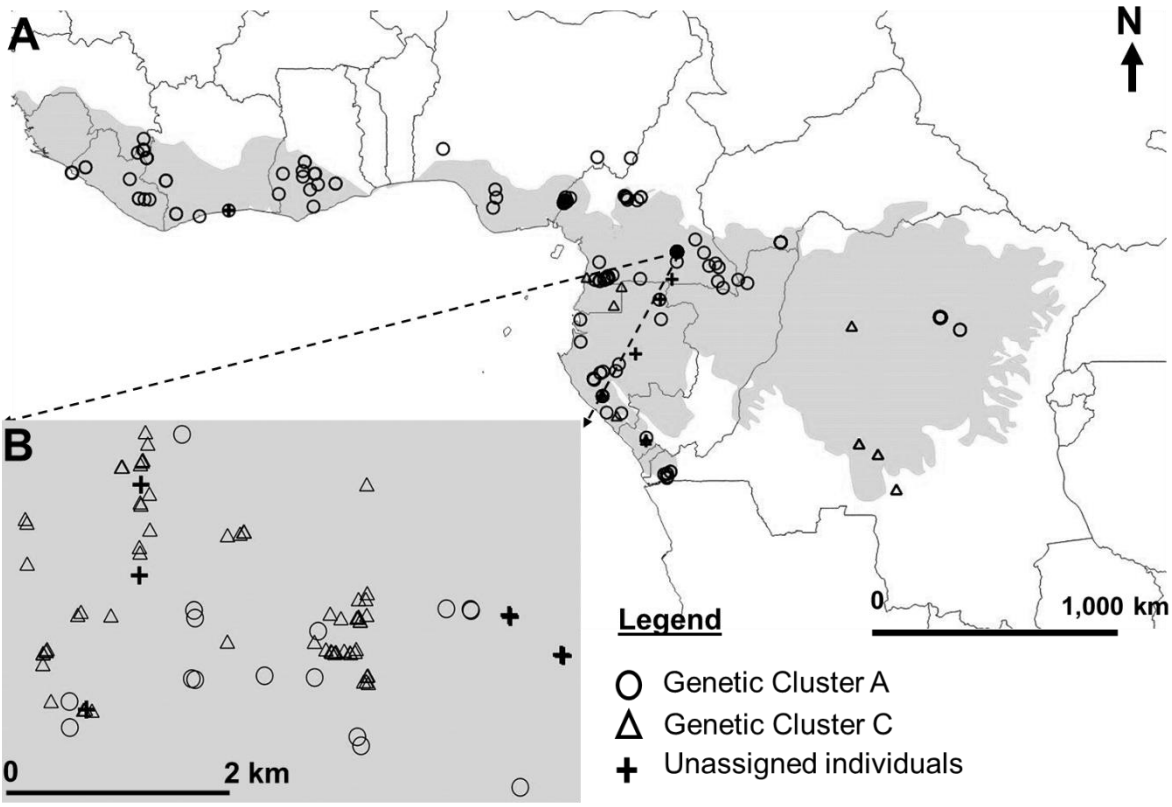


Fig. 1. Spatial distribution of genotyped samples of *Entandrophragma angolense* / *E. congoense* (A) across the Guineo-Congolian forest (grey area), and (B) in a forest from Eastern Cameroon. The symbols represent the output of the clustering algorithm (STRUCTURE) which assigned each sample to one of two genetic clusters (A or C) or left them unassigned when both clusters contributed to >10% of the genome.

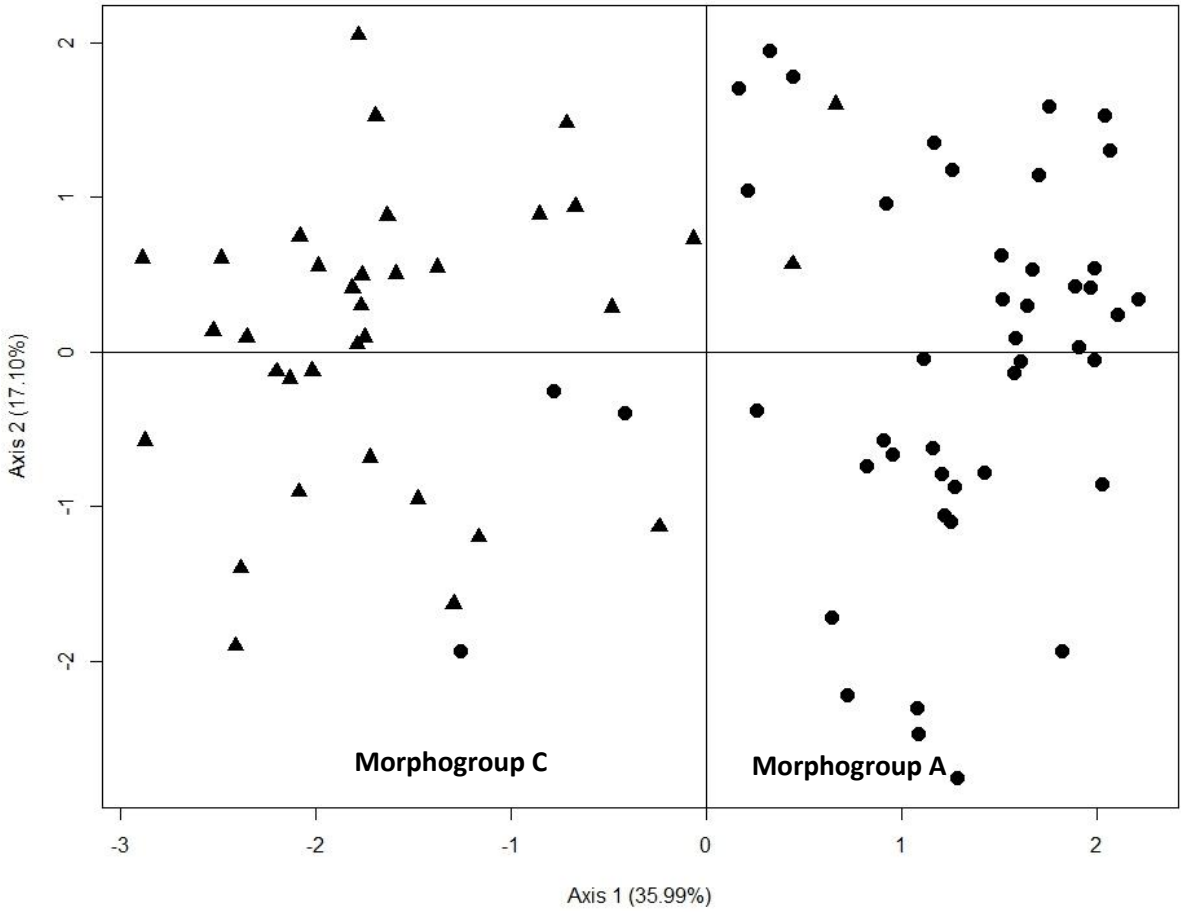


Fig. 2. Hill-Smith ordination of 81 *Entandrophragma angolense* / *E. congoense* specimens of morphogroups A (circles) and C (triangles) for six quantitative and qualitative leaf traits, using the two first axes (53% of variance explained).

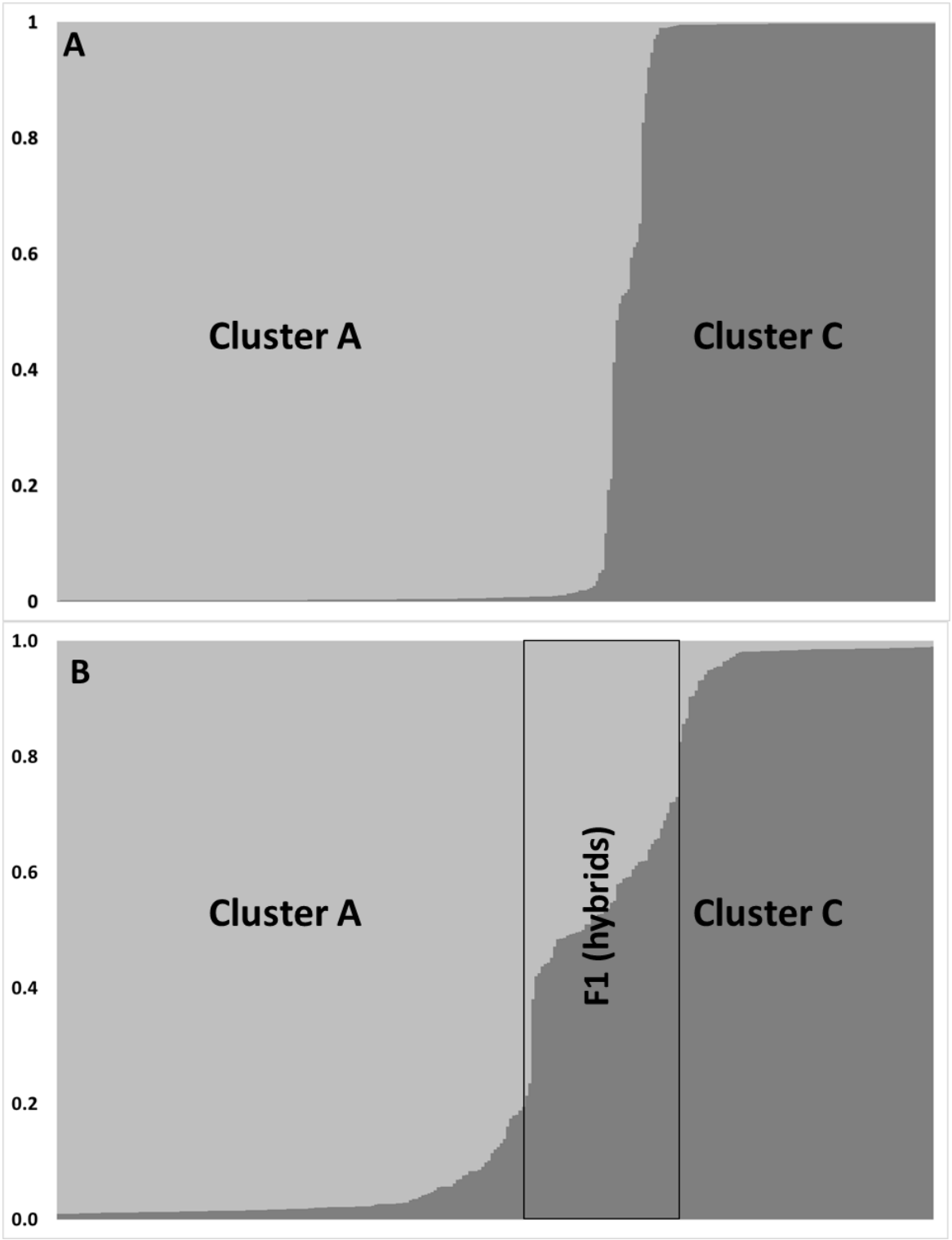


Fig. 3. (A) Histogram of genetic assignment of 305 individuals genotyped at 15 microsatellites loci at K=2 genetic clusters, according to a Bayesian clustering analysis. Each bar indicates the proportion of the genome (*q*) of an individual being assigned to each genetic cluster. (B) Identical analysis performed on 278 simulated genotypes (151 pure cluster A, 77 pure cluster C, 50 F1 hybrids) to identify the *q* value thresholds corresponding to hybrids between the two genetic clusters (0.2<*q*<0.8).

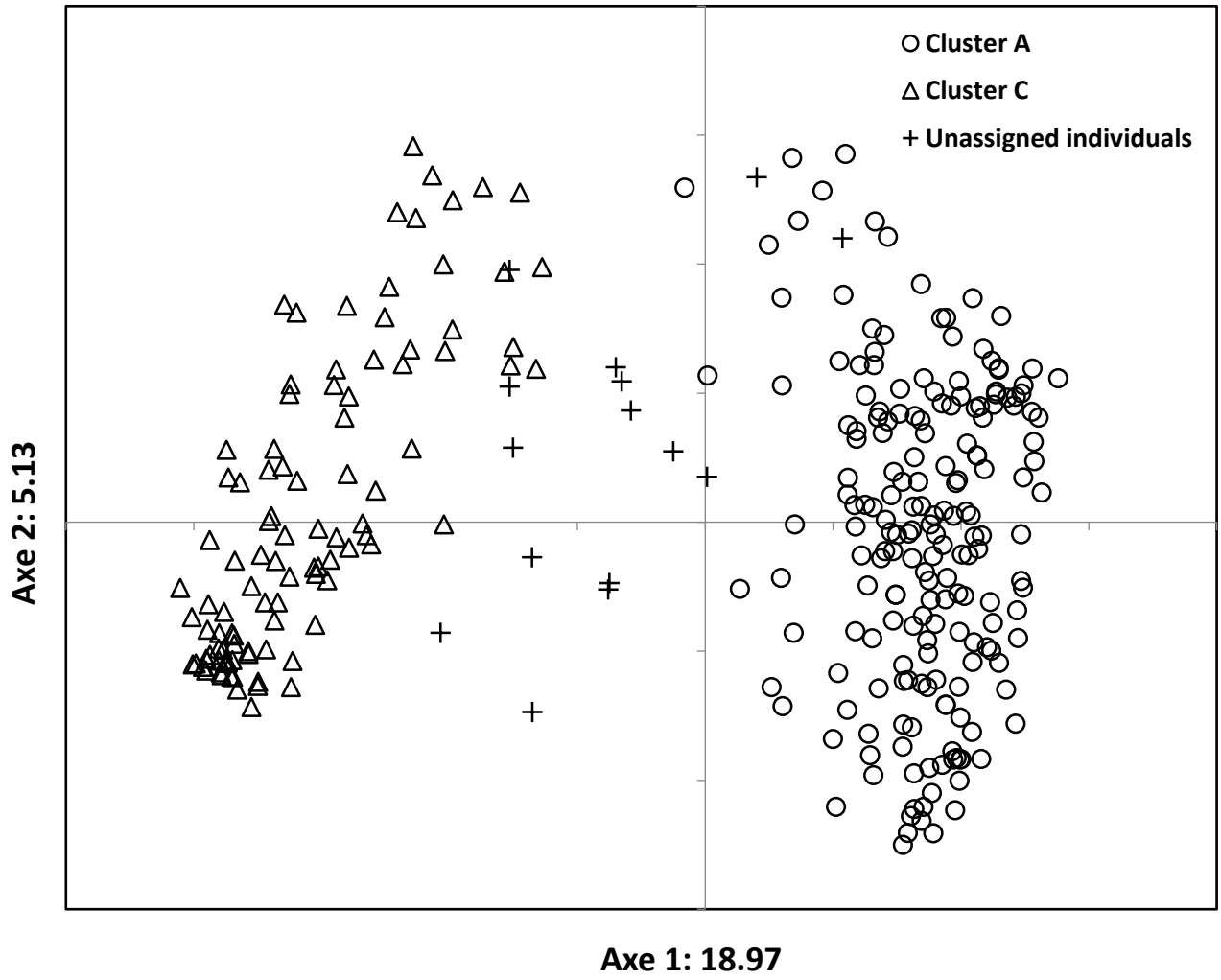


Fig. 4. Principal coordinate analysis (PCoA) of pairwise genetic distances between 305 individuals, with cluster assignments according to a Bayesian clustering analysis.

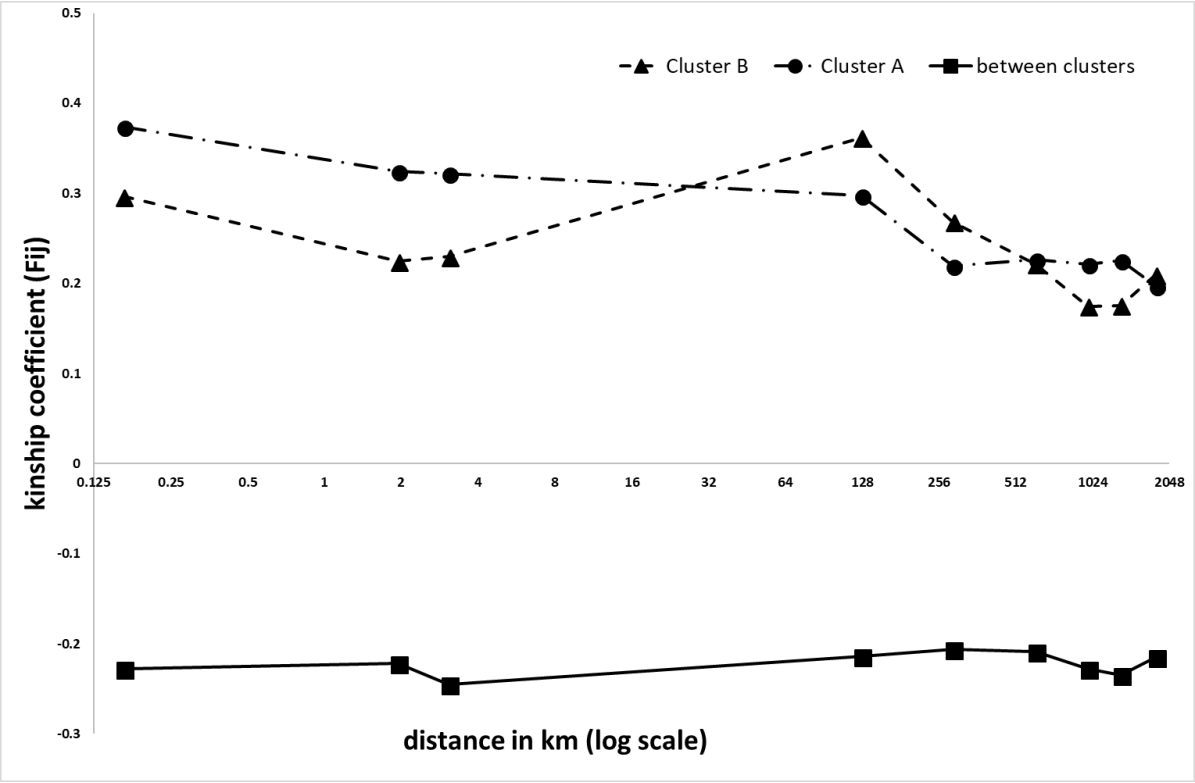


Fig. 5. Average kinship coefficients (F_{ij}) between pairs of individuals plotted against the logarithm of geographical distance for different pairwise comparisons between genetic clusters (A or C): A-A (dashed line), C-C (broken line), and A-C (solid line).

Table 1. Samples used for the morphometric analyses. The letters in voucher identification tags indicate the herbarium in which specimens were collected: L for Leiden, BR for Meise, WAG for Wageningen; and GEM for Gembloux greenhouse.

Taxon name	Geographic origin	Voucher
Entandrophragma angolense	Central African Republic	WAG1096893, BR000005862887, BR0000013596002
	Cameroon	GEM10[a]
	Democratic Republic of the Congo	L2158046, WAG1096868, WAG1096871, BR0000013596576, BR0000013266974, BR0000013946036, BR0000013266981, BR0000013946081, BR0000013646104, BR0000013946128, BR0000013946180, BR0000013946277, BR0000013946364, BR0000013946371, BR0000013946401, BR0000013946418, BR0000013596422, BR0000013946050
	Gabon	WAG1096873, WAG1096895
	Ghana	WAG1096874, BR0000013592738
	Guinea-Bissau	WAG1096864, WAG1096865, WAG1488141
	Ivory Coast	L2158048, WAG1096837, WAG1096844, WAG1096878, WAG1096879, BR0000013596095
	Liberia	WAG1096851, WAG1096854, WAG1096858, WAG1096861, BR0000013592714
	Nigeria	WAG1096834, WAG1096843, WAG1096862, WAG1097368
	Sierra Leone	L2158047, L2158049, WAG1096857
Entandrophragma congoense	Central African Republic	WAG1096927
	Democratic Republic of the Congo	L2158016, L2158017, L2158018, L2158019, L2158020, L2158022, WAG1096928, WAG1096930, WAG1096931, WAG1096932, WAG1096935, WAG1096937, WAG1096938, WAG1096940, WAG1096942, WAG1096944, WAG1096945, WAG1096946, WAG1096947, WAG1096949, WAG1096950, WAG1096951, WAG1096953, BR0000013977382, BR0000013977504, BR0000013977511, BR00000672842, BR0000013977566, BR0000013977733, BR0000013977764
	Gabon	WAG1096956, BR00000586403, BR00000586280
	Nigeria	WAG1096838

[a] Sample used for genomic libraries available from Dr. Olivier Hardy collection (ULB, EBE team).

Table 2. Comparison between morphogroups A and B for nine quantitative variables measured from leaves from 81 samples and from seeds of 10 samples. Mean, standard deviation and [minimum-maximum] values are reported, as well as P values of the Welch two sample t-tests.

	Characters	**Morphogroup A** (n=46 for leaves, n= 5 for seeds)	**Morphogroup C** (n=35 for leaves, n= 5 for seeds)	**P**
Leaves	NF (number of leaflets per leaf)	7.64 ± 2.23 [3-13]	9.37 ±2.97 [5-18]	0.01
	LF (leaflet length)	12.43 ± 3.93 [5.2-25.5]	11.93 ± 1.75 [7.5-16.5]	0.43
	wF (leaflet width)	4.77 ± 1.49 [2.9 - 9]	3.66 ± 0.91 [2.3 - 6.4]	<0.01
	RLF (ratio of LF/wF)	2.62 ± 0.41 [1.48-4.25]	3.38 ± 0.69 [1.87-4.8]	<0.01
Seeds and capsules	Lcp (capsule length)	16.32 ± 1.4 [14.6-17.5]	12.9 ± 1.60 [10-14]	0.01
	wcp (capsule width)	2.78 ± 0.7 [2-3.5]	2 ± 0.00 [2-2]	0.06
	Rcp (ratio of Lcp/wcp)	6.13 ± 1.75 [4.17-8.5]	6.75 ± 0.82 [5-7]	0.77
	Lsd (seed length)	8.74 ± 1.52 [6.5-10.2]	7.58 ± 0.75 [6.4-8.4]	0.18
	wsd (seed width)	1.62 ± 0.32 [1.2-2.1]	1.3 ± 0.12 [1.2-1.5]	0.09
	Rsd (ratio of Lsd/wsd)	5.42 ± 0.38 [4.85-5.88]	5.86 ± 0.68 [4.92-6.46]	0.25

Table 3. Characterization of 15 polymorphic nuclear microsatellite loci isolated from *Entandrophragma angolense*.

Locus[a]	Primer sequences (5′–3′)	Fluorescent label[b]	Repeat motif	Allele size range (bp)	GenBank accession no.
EnA-ssr2	F: TGTGGAGAAACTGAGGGACC R: CGAATTGCAGATTGAGAGCTT	Q1-6-FAM	$(AG)_{16}$	212-253	MH382769
EnA-ssr3	F: CCCACCAATCCCTCTCAAA R: CCCTGCAGATGAAACCCTAA	Q1-6-FAM	$(AG)_{18}$	184-226	MH382770
EnA-ssr5	F: CTAGTGGGCGAACACAAACA R: CAAATTCAAGTCTGCTTTCGG	Q1-6-FAM	$(AT)_{15}$	152-186	MH382771
EnA-ssr7	F: GCCACGACATTATTTCCACC R: CAGTTGTTGCGGTCACAATC	Q1-6-FAM	$(AG)_{14}$	141-173	MH382772
EnA-ssr14	F: AACTCTGACACGTGCGGTTA R: GCTGCCAGCATTGATAGTGA	Q2-NED	$(AG)_{17}$	184-226	MH382773
EnA-ssr15	F: CCATGGGTAAGCTCTCAACAA R: GGAGTTTGGCCTCTCACCTT	Q2-NED	$(AG)_{15}$	159-219	MH382774
EnA-ssr21	F: TTGAGCATGGTTTATGTATCCG R: AACGTGAAGGTACAGGTTGTATCA	Q2-NED	$(AT)_{14}$	122-156	MH382775
EnA-ssr23	F: TGCTAACATCTGGTTGCATCA R: AAGTGCCTACCAGCCTTACTTT	Q2-NED	$(AC)_{12}$	115-182	MH382776
EnA-ssr29	F: AGATGGGCGACTAAAGCTGA R: ACAGGCACAGTACACCTGGA	Q3-VIC	$(AG)_{15}$	135-201	MH382777
EnA-ssr34	F: CATAGAGATTTGGGACATGGG R: ATGGCATACAGATGCAACGA	Q3-VIC	$(AC)_{12}$	157-190	MH382778
EnA-ssr35	F: CAGCATTTGAGTGTATGTTCCC R: TAATAGGGCAGACGGCTTGT	Q3-VIC	$(AG)_{11}$	121-158	MH382779
EnA-ssr36	F: TCTTTCCCACCAATTCAAGG R: TGAGGGTCTGAAACAAAGTGAA	Q4-PET	$(AAG)_{12}$	216-302	MH382780
EnA-ssr42	F: ACGGAAACCATTACCACACC R: TTTCATCGGGAAGAAGGC	Q4-PET	$(AC)_{16}$	145-186	MH382781
EnA-ssr44	F: AGAAGAATAAACAACACCACCC R: CTGTTCTTATGATGTCCATGGTG	Q4-PET	$(AG)_{18}$	119-157	MH382782
EnA-ssr48	F: TTGTTGTTCTGCAAGGATGG R: GGCCGAAGTCCCTTCTAATC	Q4-PET	$(AG)_{11}$	136-166	MH382783

[a] Optimal annealing temperature was 57°C for all loci.
[b] Q1 = TGTAAAACGACGGCCAGT; Q2 = TAGGAGTGCAGCAAGCAT;
Q3 = CACTGCTTAGAGCGATGC; Q4 = CTAGTTATTGCTCAGCGGT (Q1 after Schuelke, 2000; Q2–Q4 after Culley et al., 2008).

Table 4. Genetic characterization of 15 polymorphic microsatellite loci for the two genetic clusters (individuals assigned at $q \geq 0.90$) at the scale of the Guineo-Congolian forests.

	Cluster A (n = 190)					Cluster C (n = 100)				
Locus[a]	*A*	H_e	*Ho*	*F*	*r*	*A*	H_e	*Ho*	*F*	*r*
EnA-ssr2	21	0.89	0.48	0.45	0.18	6	0.63	0.55	0.13	0.01
EnA-ssr3	15	0.89	0.70	0.21	0.05	2	0.03	0.01	0.66	0.09
EnA-ssr5	10	0.81	0.22	0.73	0.32	3	0.50	0.56	-0.11	0.01
EnA-ssr7	17	0.91	0.47	0.48	0.21	3	0.26	0.06	0.77	0.19
EnA-ssr14	18	0.90	0.42	0.53	0.24	1	0.00	0.00	-	0.03
EnA-ssr15	16	0.80	0.71	0.16	0.06	1	0.00	0.00	-	0.29
EnA-ssr21	12	0.77	0.32	0.58	0.25	2	0.01	0.01	0.00	0.02
EnA-ssr23	23	0.88	0.77	0.12	0.05	12	0.87	-	0.15	0.07
EnA-ssr29	21	0.91	0.44	0.51	0.24	2	0.01	0.01	0.00	0.15
EnA-ssr34	15	0.72	0.25	0.64	0.27	4	0.51	0.23	0.53	0.17
EnA-ssr35	14	0.72	0.40	0.44	0.18	4	0.55	0.58	-0.05	0.00
EnA-ssr36	13	0.87	0.43	0.49	0.22	1	0.00	0.00	-	0.15
EnA-ssr42	18	0.90	0.45	0.5	0.00	6	0.55	0.58	-0.05	0.22
EnA-ssr44	18	0.91	0.51	0.44	0.18	7	0.76	0.43	0.43	0.20
EnA-ssr48	17	0.91	0.36	0.60	0.27	1	0.00	0.00	-	0.29
Multilocus average	16.53	0.85	0.46	0.45		3.67	0.31	0.25	0.19	

Note: *A* = number of alleles; *F* = fixation index; *H*e = expected heterozygosity; *Ho* = observed heterozygosity; *n* = number of individuals, *r* = null allele frequency according to Bayesian structure analysis (STRUCTURE).

Appendix 1: Morphological characters differentiating *Entandrophragma* species as reported by Liben and Dechamps (1966) and by de Wilde (2015). [$] indicates qualitative variables used for morphological analyses.

	E. congoense	***E. angolense***
Trunk		
Tree dimensions	Up to 45 m high and 90 cm diameter.	Up to 50 m high and 200 cm diameter
base	Straight more or less thickened, sometimes sub-winged, but without developed buttresses	Sometimes simply thickened, but usually has well-developed aliform buttresses.
Rhytidome	Suberous peeling off into regular plaques (square or rectangular)	Non-suberous, peeling off by irregular plaques
Leaflet		
Form	Elliptical-oblong, rarely ovate	Generally obovate, rarely elliptical or oblong
Base	Generally acute, rarely obtuse, asymmetrical	Generally acute, rarely obtuse or round
Apex[$]	Acute-apiculate, more rarely bluntly rounded and apiculate	Obtuse to rounded, with long deciduous acumen, rarely acute, exceptionally retuse and mucronate
Midrib[$]	Glabrous, exceptionally pubescent when young	Pubescent, rarely glabrous below
Domatia[$]	Present, a thick tuft of hairs	Generally absent
Capsules		
Valves	(9.5-)11-16(-18) cm long (1-)1.5-2 cm wide (1-)1.5-3(-3.5) mm thick	(12-)15-19(-22) cm long 2-2.7(-3) cm wide 4-6(-7) mm thick
Columella	Almost straight and subcircular section	Curved at the extremity and with a long elliptical section
Seeds		
Base	Straight, narrower than the wing	Wavy, as wide as the wing
External teguments	Violet blue	Brown

Electronic Supplementary Material

The African timber tree *Entandrophragma congoense* (Pierre ex De Wild.) A.Chev. is morphologically and genetically distinct from *Entandrophragma angolense* (Welw.) C.DC.

Franck K. Monthe[1], Jérôme Duminil[1,2,3], Emmanuel Kasongo Yakusu[4,5,6], Hans Beeckman[4], Nils Bourland[4,7,8], Jean-Louis Doucet[9], Marc S.M. Sosef[10] and Olivier J. Hardy[1]

[1] Université Libre de Bruxelles, Faculté des Sciences, Service Evolution Biologique et Ecologie, CP 160/12, 50 Av. F. Roosevelt, 1050 Bruxelles, Belgium.
[2] DIADE, IRD, Univ Montpellier, 911 Avenue Agropolis, BP 64501, 34394 Montpellier, France.
[3] Bioversity International, Forest Genetic Resources and Restoration Programme, Sub-Regional Office for Central Africa, P.O. Box 2008 Messa, Yaoundé, Cameroon.
[4] Wood Biology Service, Royal Museum for Central Africa (RMCA), Leuvensesteenweg 13, 3080 Tervuren, Belgium.
[5]Ghent University, Laboratory of Wood Technology, Coupure Links 653, 9000 Ghent, Belgium
[6] Université de Kisangani, Faculté de Gestion des Ressources Naturelles Renouvelables, B.P. : 2012, Avenue Kitima, 3, Kisangani, République Démocratique du Congo.
[7] Center For International Forestry Research, Situ Gede, Bogor Barat 16115, Indonesia.
[8] Resources and Synergies Development Pte Ltd, Raffles Quay 33-03, Singapore 048581, Singapore
[9] TERRA Teaching and Research Centre, Central African Forests, BIOSE Department, Gembloux Agro-Bio Tech, Université de Liège, 2 Passage des Déportés, 5030 Gembloux, Belgium.
[10]Botanic Garden Meise, Nieuwelaan 38, 1860 Meise, Belgium.

[$] Corresponding author. Université Libre de Bruxelles, Faculté des Sciences, Service Evolution Biologique et Ecologie, CP 160/12, 50 Av. F. Roosevelt, 1050 Bruxelles, Belgium. fmonthek@ulb.ac.be Tel. +32 (0)2 650 45 11. Fax. +32 (0)2 650 24 45.

ORCID: 0000-0003-4664-658X

Table S1: Contribution (Eigenvector) of morphological variables to the three first principal components: NF (number of leaflets per leaf), LF (leaflet length), RLF (ratio of LF/wF), NM (midrib pilosity), Do (domatia), Acu (acumen shape), FA (apex form).

Variables	Comp1	Comp2	Comp3
NF	-0.17	0.57	-0.41
LF	0.02	0.39	0.85
RLF	-0.77	-0.01	0.24
NM	0.83	-0.15	-0.08
Do	-0.67	-0.40	-0.22
Acu	-0.08	-0.71	0.26
FA	0.86	-0.14	0.04

Table S2. Genetic diversity at the 15 microsatellite loci of two genetic clusters (individuals assigned at $q \geq 0.90$) detected at fine scale (Pallisco).

	Cluster A (n = 17)					**Cluster C (n = 64)**				
Loci	A	H_e	*Ho*	*F*	*r*	A	H_e	*Ho*	*F*	*r*
EnA-ssr2	-	-	-	-	-	5	0.63	0.62	0.02	0
EnA-ssr3	9	0.81	0.64	0.20	0.01±0.12	2	0.03	0	1	0
EnA-ssr5	3	0.7	0	1	0	3	0.50	0.71	-0.41	0
EnA-ssr7	8	0.86	0.53	0.38	0.14±0.16	2	0.14	0.05	0.64	0.30±0.10
EnA-ssr14	9	0.81	0.56	0.31	0.08±0.17	1	0	0		
EnA-ssr15	7	0.78	0.70	0.10	-	-	-	-	-	-
EnA-ssr21	6	0.75	0.17	0.77	0.26±0.09	2	0.01	0.01	0	
EnA-ssr23	13	0.85	0.70	0.17	0	9	0.84	0.77	0.08	0
EnA-ssr29	9	0.87	0.35	0.6	0.21±0.13	1	0	0		
EnA-ssr34	4	0.6	0.18	0.69	0.28±0.14	2	0.48	0.24	0.50	0.30±0.10
EnA-ssr35	7	0.57	0.23	0.59	0.14±0.12	3	0.63	0.75	-0.18	0
EnA-ssr36	6	0.79	0.70	0.10	0	1	0	0		
EnA-ssr42	10	0.86	0.35	0.59	0.20±0.13	4	0.58	0.61	-0.05	0
EnA-ssr44	7	0.86	0.35	0.59	0.35±0.19	6	0.74	0.47	0.36	0.20±0.1
EnA-ssr48	8	0.84	0.14	0.83	0.46±0.12	1	-	-	-	-
Multilocus		0.78	0.40	0.49			0.38	0.35	0.08	
$F_{(null)}$				**0**					**0.136±003**	

Note: *A* = number of alleles; *F* = Inbreeding coefficient (potentially biased by null alleles); *H*e = expected heterozygosity; *Ho* = observed heterozygosity; *n* = number of individuals, *r* = null allele frequencies; $F_{(null)}$, corrected multilocus inbreeding coefficient under a population inbreeding model (unbiased by null alleles); *r* and $F_{(null)}$ are estimated by INEst.

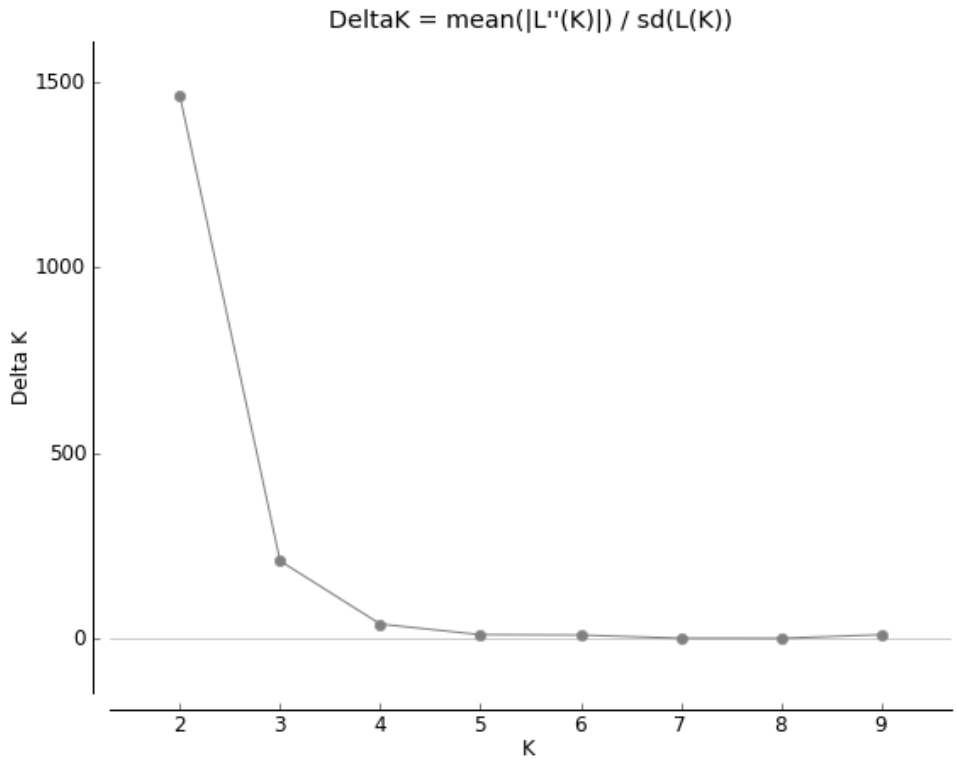
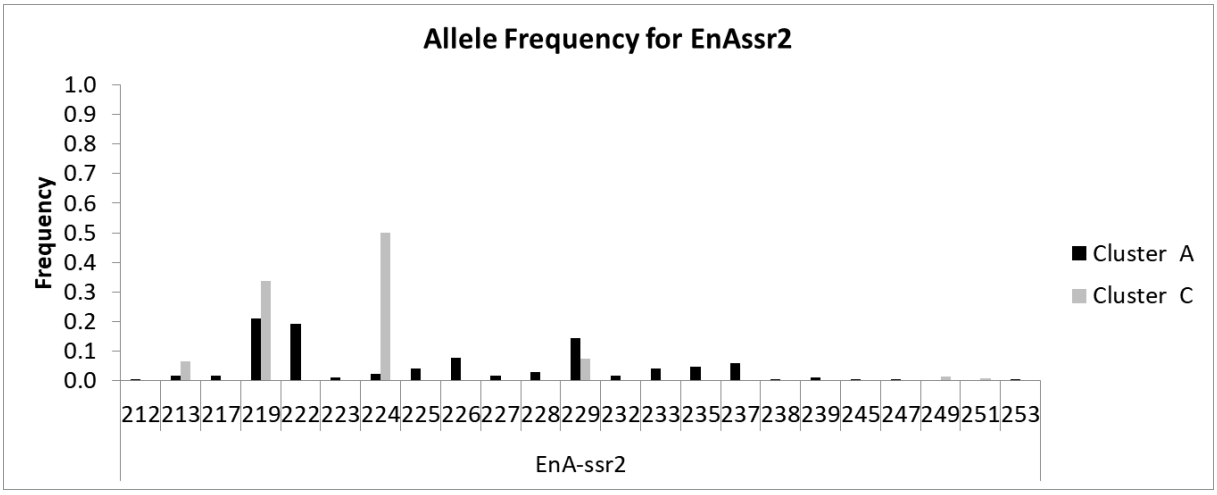
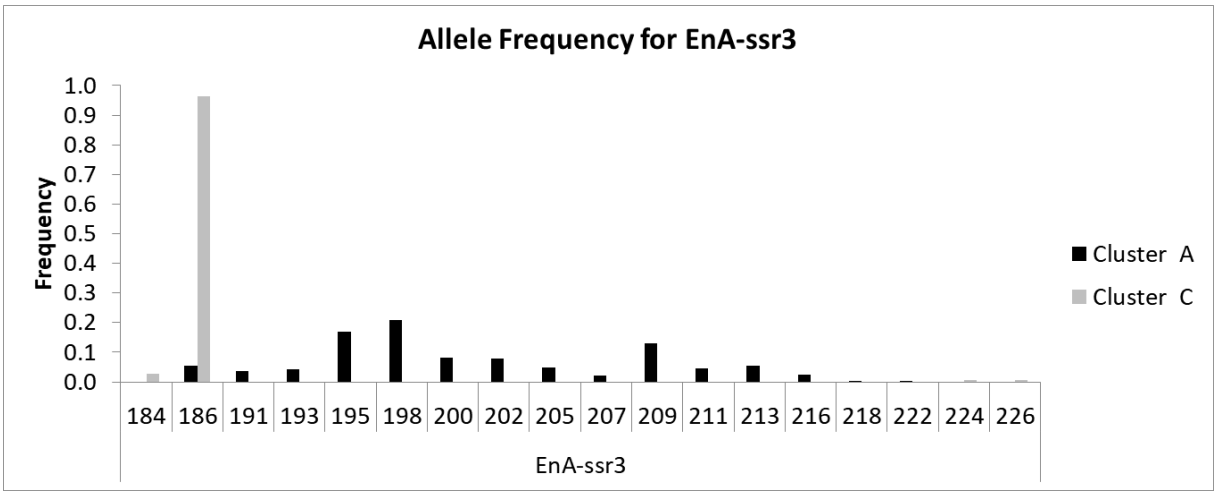
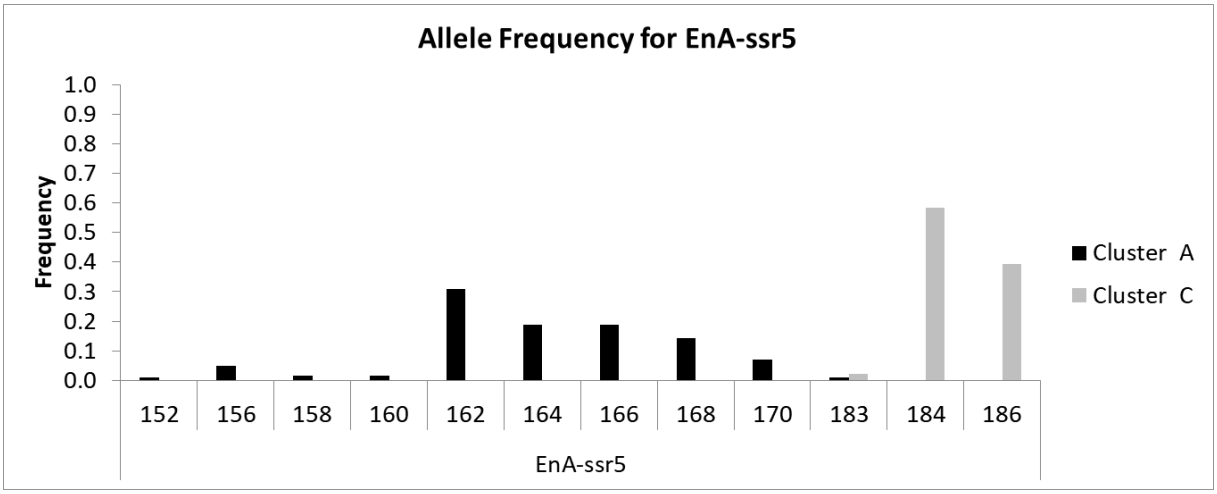
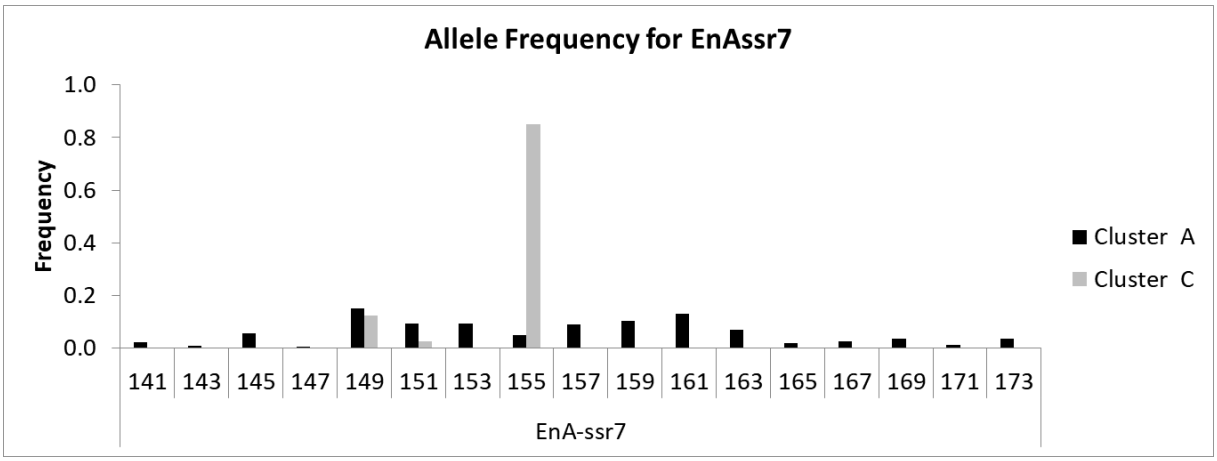


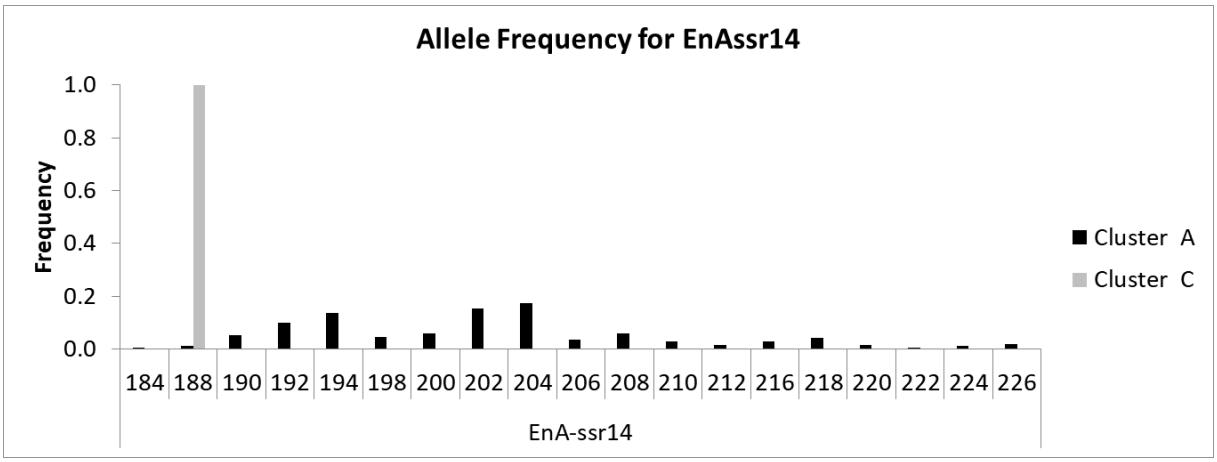
Fig. S1. Variation of deltaK values as a function of the number of hypothetical genetic clusters (K) according to the clustering algorithm implemented in STRCTURE. We considered K = 2 for further analyses.

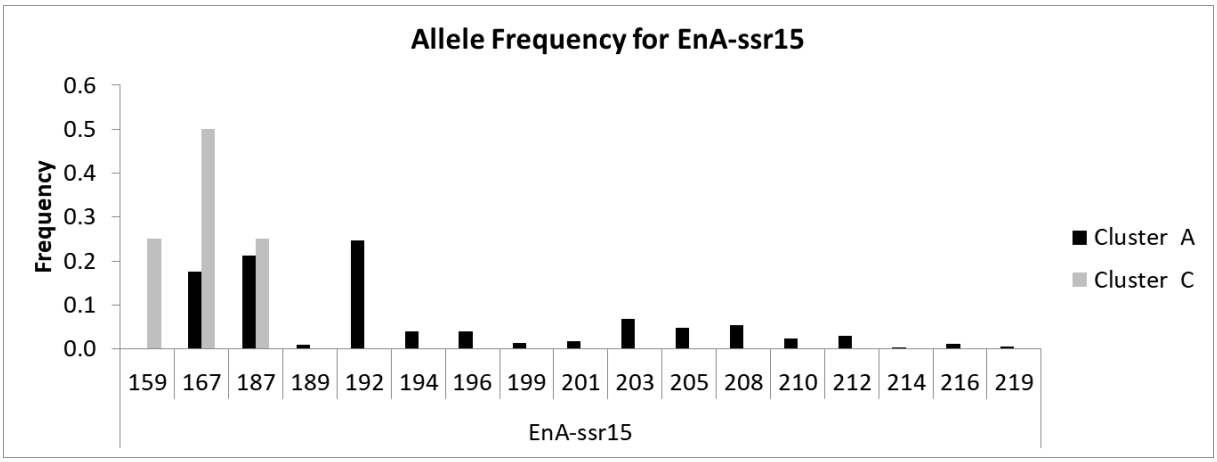

Allele Frequency for EnAssr2
Frequency
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
212 213 217 219 222 223 224 225 226 227 228 229 232 233 235 237 238 239 245 247 249 251 253
EnA-ssr2
Cluster A
Cluster C

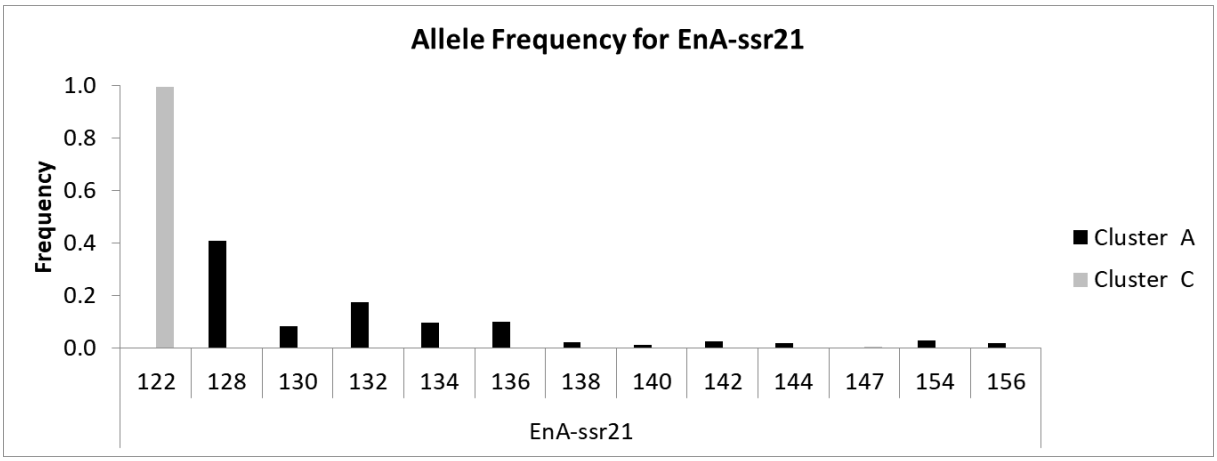

Allele Frequency for EnA-ssr3
Frequency
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
184 186 191 193 195 198 200 202 205 207 209 211 213 216 218 222 224 226
EnA-ssr3
Cluster A
Cluster C

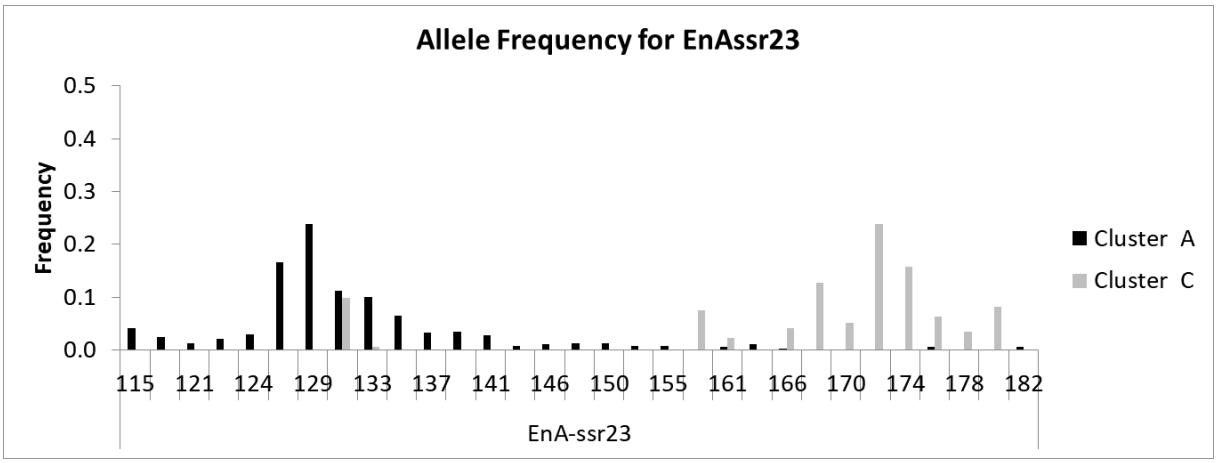

Allele Frequency for EnA-ssr5
Frequency
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
152 156 158 160 162 164 166 168 170 183 184 186
EnA-ssr5
Cluster A
Cluster C

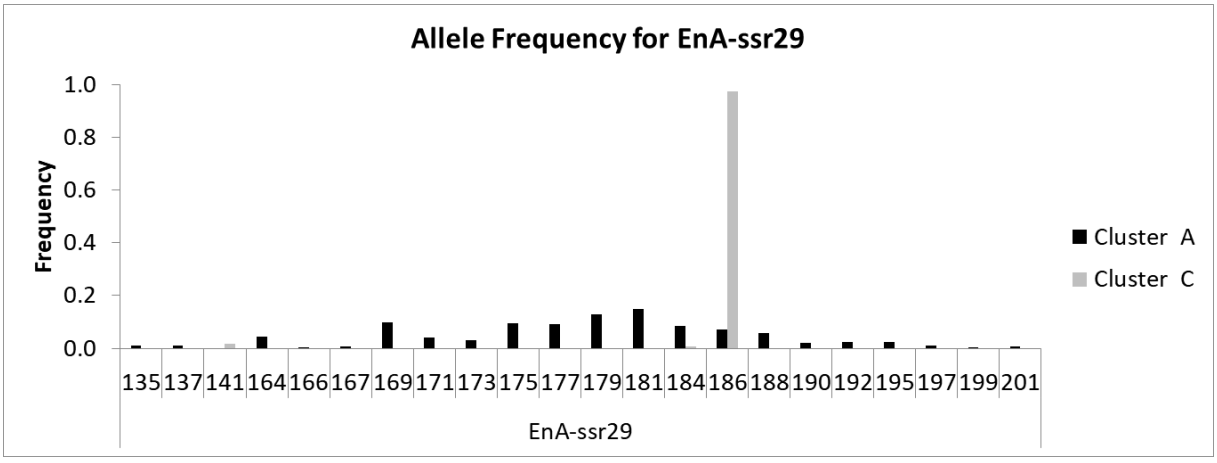

Allele Frequency for EnAssr7
Frequency
1.0
0.8
0.6
0.4
0.2
0.0
141 143 145 147 149 151 153 155 157 159 161 163 165 167 169 171 173
EnA-ssr7
Cluster A
Cluster C

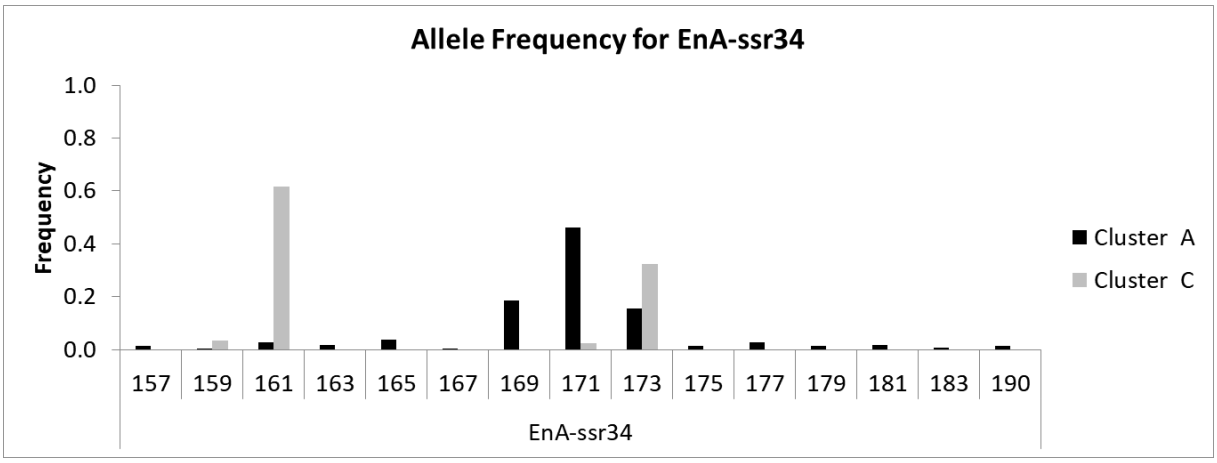

Allele Frequency for EnAssr14
Frequency
1.0
0.8
0.6
0.4
0.2
0.0
184 188 190 192 194 198 200 202 204 206 208 210 212 216 218 220 222 224 226
EnA-ssr14
Cluster A
Cluster C

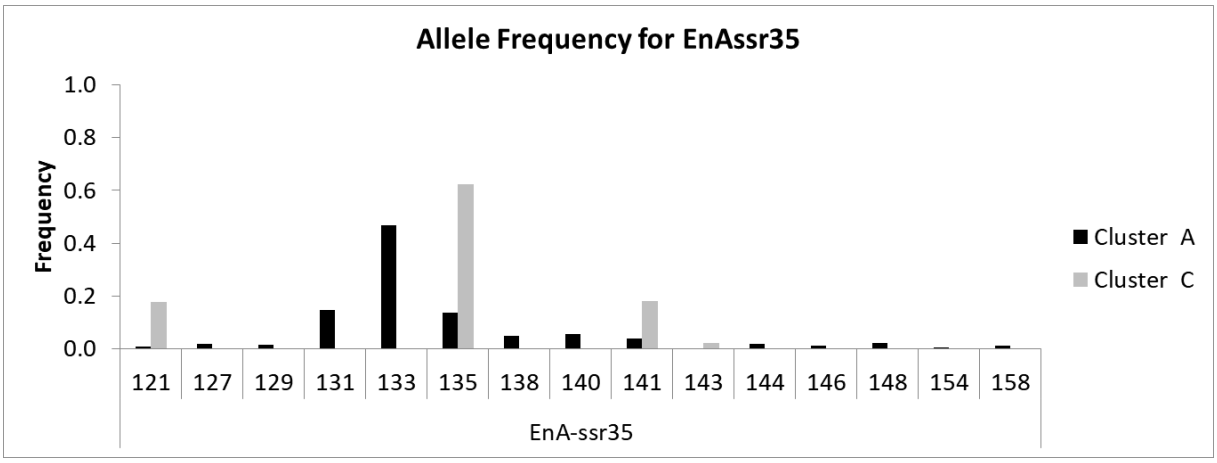

Allele Frequency for EnA-ssr15
Frequency
0.6
0.5
0.4
0.3
0.2
0.1
0.0
159 167 187 189 192 194 196 199 201 203 205 208 210 212 214 216 219
EnA-ssr15
Cluster A
Cluster C

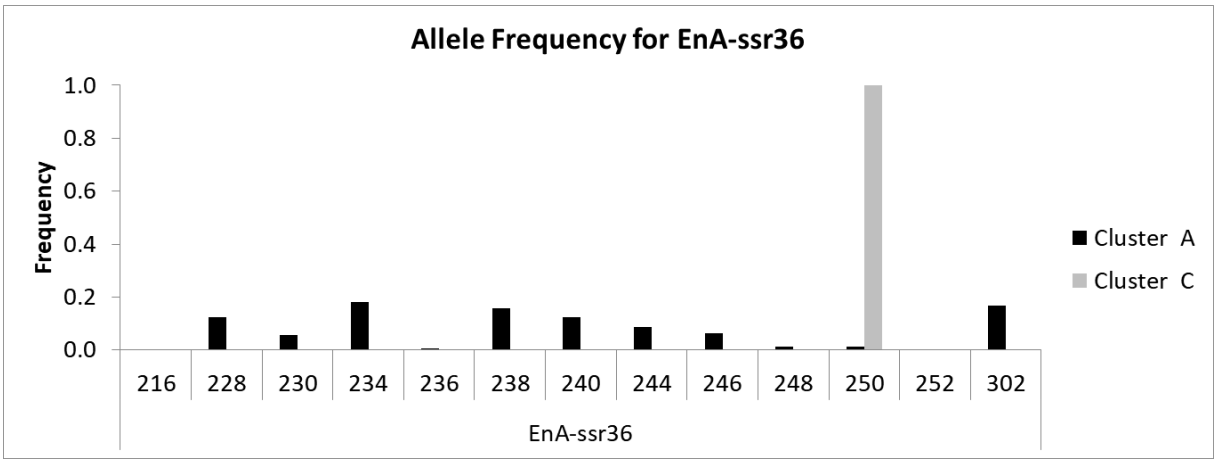

Allele Frequency for EnA-ssr21
Frequency
1.0
0.8
0.6
0.4
0.2
0.0
122 128 130 132 134 136 138 140 142 144 147 154 156
EnA-ssr21
Cluster A
Cluster C


Allele Frequency for EnAssr23
Frequency
0.5
0.4
0.3
0.2
0.1
0.0
115 121 124 129 133 137 141 146 150 155 161 166 170 174 178 182
EnA-ssr23
Cluster A
Cluster C


Allele Frequency for EnA-ssr29
Frequency
1.0
0.8
0.6
0.4
0.2
0.0
135 137 141 164 166 167 169 171 173 175 177 179 181 184 186 188 190 192 195 197 199 201
EnA-ssr29
Cluster A
Cluster C


Allele Frequency for EnA-ssr34
Frequency
1.0
0.8
0.6
0.4
0.2
0.0
157 159 161 163 165 167 169 171 173 175 177 179 181 183 190
EnA-ssr34
Cluster A
Cluster C


Allele Frequency for EnAssr35
Frequency
1.0
0.8
0.6
0.4
0.2
0.0
121 127 129 131 133 135 138 140 141 143 144 146 148 154 158
EnA-ssr35
Cluster A
Cluster C


Allele Frequency for EnA-ssr36
Frequency
1.0
0.8
0.6
0.4
0.2
0.0
216 228 230 234 236 238 240 244 246 248 250 252 302
EnA-ssr36
Cluster A
Cluster C

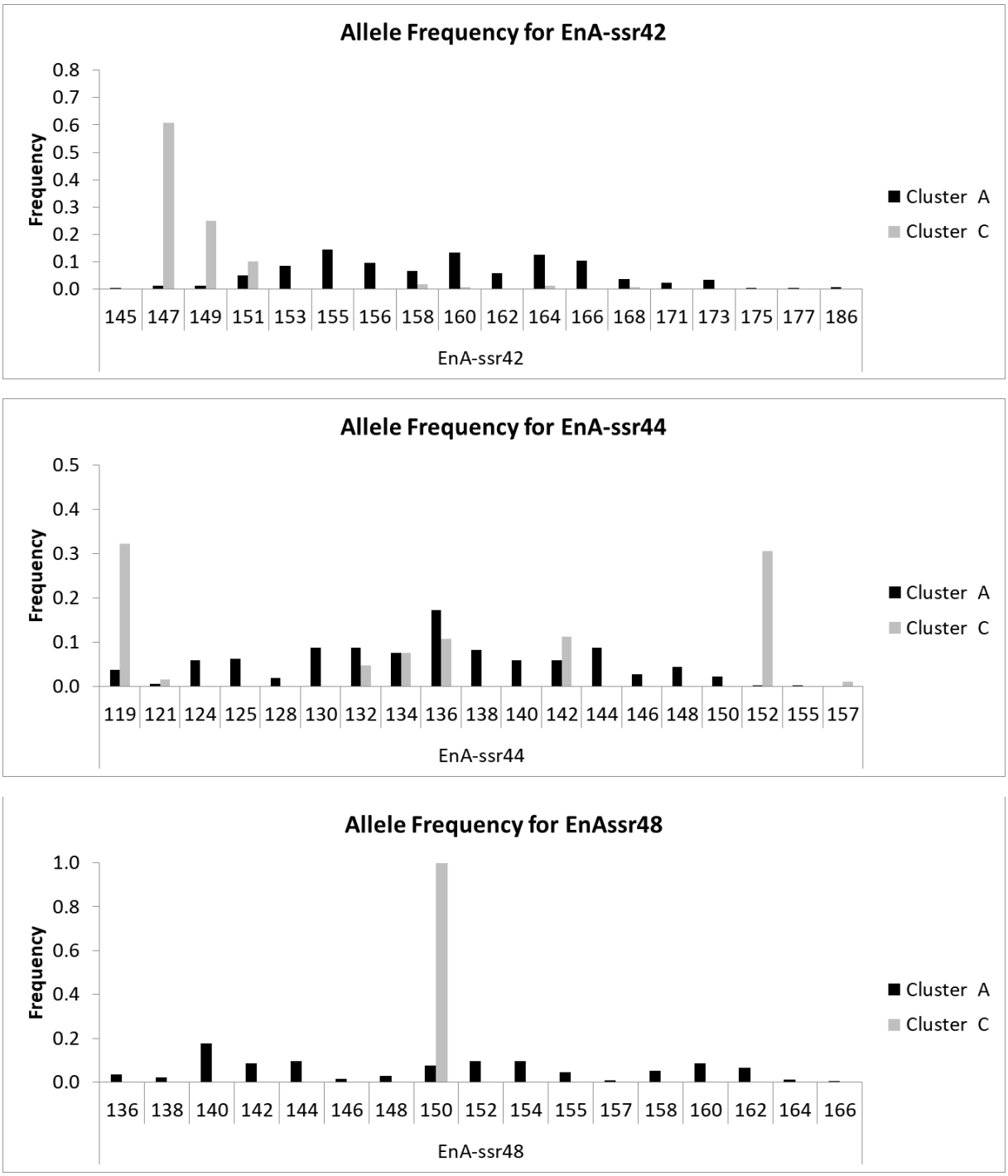


Fig. S2. Comparison of allele frequencies for each genetic cluster (A corresponds to *Entandrophragma angolense*, C to *E. congoense*).

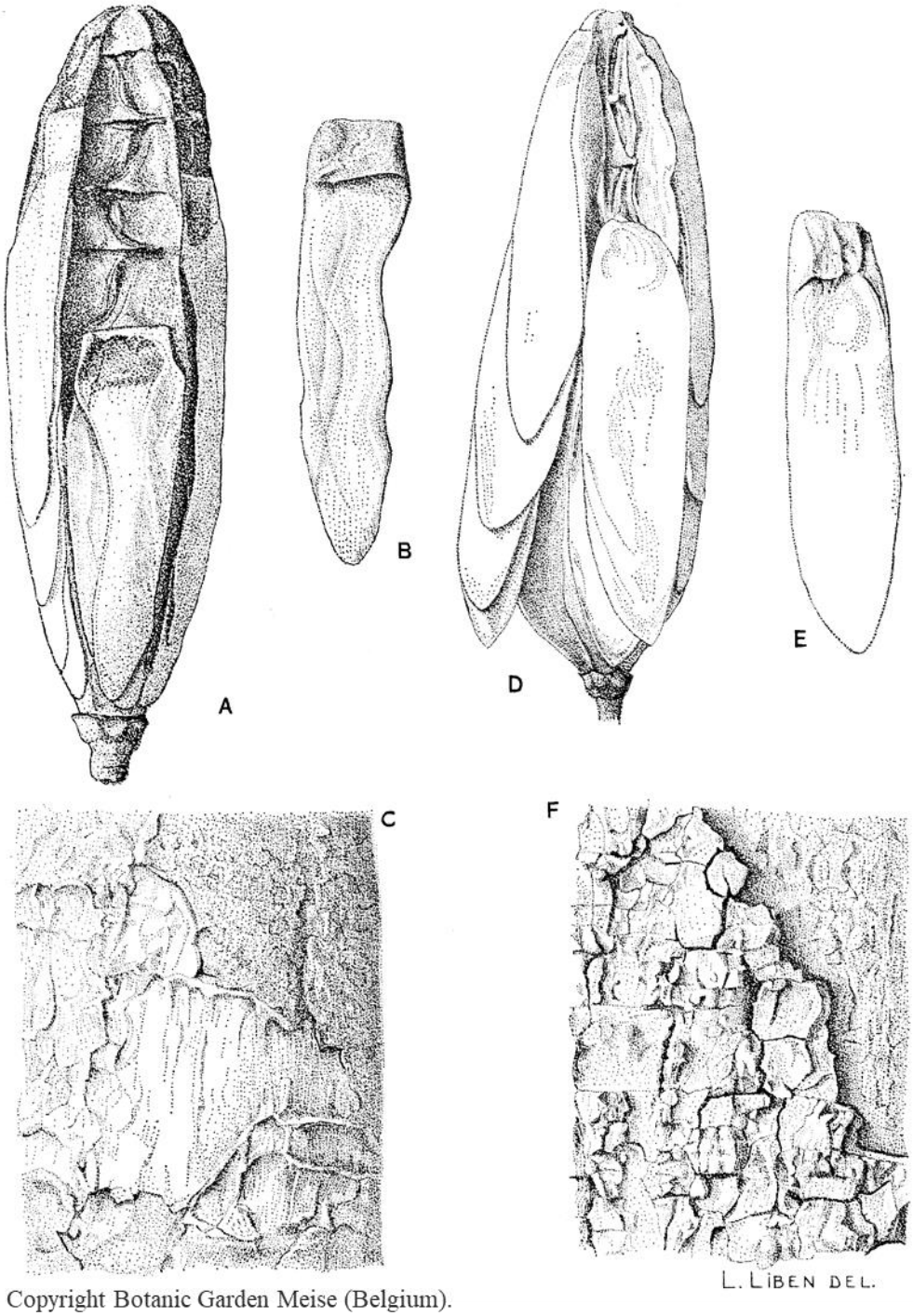


Fig. S3: Some morphological differences between *E. angolense* and *E. congoense* as reported by Liben and Dechamps (1966). Figures A, B and C represent respectively: columella, seed internal face and rhytidome of *E. angolense*. Figures D, E and F represent: columella, internal seed surface, rhytidome of *E. congoense.*