SYSTEMATICS AND PHYLOGENY

# Species delimitation in the genus *Greenwayodendron* based on morphological and genetic markers reveals new species

**Brandet-J. Lissambou,[1,2,7] Thomas L.P. Couvreur,[2,3] Christiane Atteke,[1] Tariq Stévart,[4] Rosalía Piñeiro,[5] Gilles Dauby,[2,6,7] Franck K. Monthe,[7] Davy U. Ikabanga,[1,7] Bonaventure Sonké,[3] Bertrand M'batchi[1] & Olivier J. Hardy[7]**

1 *Département de Biologie, Faculté des Sciences, Université des Sciences et Techniques de Masuku (USTM), BP 943, Franceville, Gabon*
2 *IRD, DIADE, Université Montpellier, Montpellier, France*
3 *Plant Systematics and Ecology Laboratory, Higher Teachers Training College, University of Yaoundé I, BP 047, Yaoundé, Cameroon*
4 *Africa & Madagascar Department, Missouri Botanical Garden, St. Louis, Missouri, U.S.A.*
5 *Physical Geography, University of Exeter, Geoffrey Pop Building, Stocker road, EX44QD, U.K.*
6 *AMAP lab, IRD, CNRS, INRA, Université Montpellier, Montpellier, France*
7 *Evolutionary Biology and Ecology Unit, Faculté des Sciences, Université Libre de Bruxelles, 50 Av. F. Roosevelt, CP 160/12, 1050, Brussels, Belgium*

Address for correspondence: *Brandet-J. Lissambou, lissamboujuniorbrandet@gmail.com*

**DOI** https://doi.org/10.1002/tax.12064

**Abstract** Combining genetic and morphological markers is a powerful approach for species delimitation, much needed in tropical species complexes. *Greenwayodendron* (Annonaceae) is a widespread genus of trees distributed from West to East African rainforests. Two species and four infra-specific taxa are currently recognized. However, preliminary genetic studies and morphological observations suggested the occurrence of additional species, undescribed to date. We tested species delimitation within *Greenwayodendron* by combining morphological and population genetics data. First, a visual inspection of about a thousand specimens suggested the existence of seven morphogroups: four of them occur in Central Africa and overlap in Gabon while three others are allopatric, occurring respectively in West Africa, East Africa, and the islands of São Tomé and Príncipe. Their morphological differentiation was confirmed by analysis of 27 morphological characters coded from 233 herbarium specimens. Second, after genotyping 800 samples at eight nuclear microsatellites, Bayesian clustering analyses (STRUCTURE) identified four genetic clusters corresponding to the well-sampled morphogroups but failed to separate the three remaining morphogroups represented by few samples. However, we show that this is an inherent limit of the STRUCTURE algorithm, whereas factorial correspondence analysis (FCA) and pairwise $F_{ST}$ and $R_{ST}$ measures confirmed the genetic differentiation of all morphogroups. We considered that a clear genetic differentiation occurring between sympatric populations advocates for recognizing distinct species following the biological species concept. Our analyses highlight that the current taxonomic treatment of *Greenwayodendron* underestimates the total number of species. We identified two new species and support separation at the rank of species of two varieties (*G. suaveolens* subsp. *suaveolens* var. *gabonica*, *G. suaveolens* subsp. *suaveolens* var. *suaveolens*) and one subspecies (*G. suaveolens* subsp. *usambaricum*). The taxonomic status of specimens collected in São Tomé and Príncipe remains inconclusive, partly due to the limited fertile material available. Our study highlights the strength of combining morphological and population genetics data for discovering new taxa. Guidelines for using genetic clustering approaches in species delimitation are provided.

**Keywords** Africa; Annonaceae; cryptic species; genetic clustering; *Greenwayodendron*

**Supporting Information** may be found online in the Supporting Information section at the end of the article.

## ■ INTRODUCTION

Until now, the majority of plant species have been delimited by taxonomists based on morphological analyses of character variation (morphological species concept, MSC) (Lewin, 1981; Henderson, 2005; Smith & al., 2018). However, this approach to species delimitation presents many shortcomings either due to intra-specific morphological variation or little morphological difference between closely related species (Pratt & Clark, 2001; Whittall & al., 2004). Today, taxonomy benefits from the growth and development of several disciplines, such as molecular genetics, to decipher species delimitation and identify cryptic species (e.g., Whittall & al., 2004; Ellis & al., 2006; Bickford & al., 2007; Duminil & Di Michele, 2009).

Integrating population genetics methods and morphological data has proven useful to delineate African rainforest tree species complexes such as in *Carapa* Aubl. (Kenfack, 2011;

**Article history:** Received: 4 Jun 2018 | returned for (first) revision: 13 Aug 2018 | (last) revision received: 27 Jan 2019 | accepted: 6 Feb 2019
**Associate Editor:** Michael D. Pirie | 

Duminil & al., 2012), *Ancistrocladus* Wall. (Turini & al., 2014), *Milicia* Sim (Daïnou & al., 2016) and *Santiria* Blume (Ikabanga & al., 2017). Population genetics methods can detect gene flow barriers by identifying well-differentiated genetic clusters (e.g., algorithm of Pritchard & al., 2000 implemented in STRUCTURE software). Hence, when reproductive isolation occurs between sympatric genetic clusters, these are likely to belong to distinct species following the biological species concept (BSC). In this way, population genetics methods can detect distinct species even if they are not yet reciprocally monophyletic at many genes due to incomplete lineage sorting, a common situation between closely related tree species with long generation times (e.g., Duminil & al., 2015). If these methods can be more sensitive than phylogenetic methods based on reciprocal monophyly to detect species, a drawback is that when genetic clusters occur in parapatry or allopatry they may represent different conspecific populations rather than distinct species. It is thus important to associate morphological data and consider the spatial distribution of the detected genetic groups before concluding whether they correspond to distinct species. A constraint of population genetics methods is that they require genotyping tenths of individuals per species at multiple nuclear markers that need to be optimised for each species, such as single nucleotide polymorphisms (SNPs) or microsatellites.

The genus *Greenwayodendron* Verdc. (Annonaceae) is restricted to tropical Africa and occurs mainly in tropical rainforests. Until recently, it has been suggested that *Greenwayodendron* contains only two species: *G. oliveri* (Engl.) Verdc. in West Africa and *G. suaveolens* (Engl. & Diels) Verdc. in Central and East Africa. The species *G. suaveolens* has been divided into two subspecies, subsp. *usambaricum* Verdc. and subsp. *suaveolens*, the latter being further divided into two varieties (var. *gabonica* (Pellegr. ex Le Thomas) Verdc. and var. *suaveolens*) (Le Thomas, 1969; Verdcourt, 1969). Recent observations suggested that the genus *Greenwayodendron* might contain more than two species. In Gabon, *G. suaveolens* subsp. *suaveolens* var. *gabonica* is very distinct morphologically (large hairy leaves) and genetically (unique plastid haplotypes) from the sympatric taxon *G. suaveolens* subsp. *suaveolens* var. *suaveolens* (Dauby & al., 2010). Along the coast of Gabon, specimens originally attributed to *G. oliveri* exhibited small leaves that contrast with typical *G. oliveri* specimens from West Africa (G. Dauby, pers. obs.). In Campo National Park in southern Cameroon, preliminary genetic analyses using nuclear microsatellites identified two distinct genetic groups, one with completely glabrous leaves and one with leaves pubescent at least along the midrib (O.J. Hardy & B.J. Lissambou, pers. obs.). Finally, due to their isolation, it can be questioned whether populations found in East Africa (*G. suaveolens* subsp. *usambaricum*) or on the islands of São Tomé and Príncipe could not correspond to distinct species as well.

The aim of this study is to use an integrative approach to clarify species limits within the genus *Greenwayodendron* using morphological and genetic data, taking advantage of recently developed microsatellite markers (Piñeiro & al., 2016). More specifically, we aim to answer the following questions: (i) Are the previous hypotheses of more than two species confirmed in the light of new population genetics and morphological analyses? (ii) How many species can we propose for the genus *Greenwayodendron*? Within the taxon *G. suaveolens* subsp. *suaveolens* var. *suaveolens*, a recent population study detected four parapatric genetic clusters that were interpreted as the legacy of past forest fragmentation (Piñeiro & al., 2017). We will therefore also test whether genetically differentiated groups should be interpreted as conspecific populations or as distinct species.

## ■ MATERIALS AND METHODS

**Study areas and sampling. —** We used two different types of material. First, we considered *Greenwayodendron* samples collected in the field and georeferenced by some of the authors or their collaborators over the past 12 years. This sampling consisted of leaf material, or sometimes cambium, dried using silica gel to preserve DNA and stored at the Université Libre de Bruxelles ($N = 2115$). For a portion of these samples herbarium vouchers were made and deposited at BR, BRLU, LBV, MO, P and WAG ($N = 80$). Second, we observed approximately a thousand herbarium vouchers from many herbaria (BM, BR, BRLU, K, L, LBV, P, YA) that were visited or that have sent plant material on loan to BRLU. DNA extraction was performed on 98 herbarium samples, but only 50 samples could finally be genotyped.

**Identification of morphological groups. —** To classify herbarium samples into distinct morphological groups, we started with a subjective approach, paying particular attention to the size, shape and pilosity of leaves. The form of connective stamens was also observed in samples with flowers. These characters differentiate species and infraspecific taxa of *Greenwayodendron* following previous taxonomic treatments and Floras (Le Thomas, 1969; Verdcourt, 1969), and recent observations by some of the authors suggested that putative new species can also be differentiated using these traits (Dauby & al., 2010; Dauby, 2012; Piñeiro & al., 2016). This visual inspection of many specimens allowed us to define distinct morphological groups, referred to as "*a priori groups*". For this step, the geographical location of the specimens was also considered, distinguishing samples from West Africa, Central Africa, East Africa and the islands of São Tomé and Príncipe. We used these *a priori* groups for the rest of the study.

To verify whether the resulting *a priori* morphological groups could correspond to distinct taxa, (i) we compared them with genetic groups inferred from microsatellite markers (see below) and (ii) we applied objective multivariate analyses and univariate statistical tests on quantitative morphological traits. For the latter step, we selected 233 herbarium specimens representing the different *a priori* groups, including all flowering specimens available ($N = 41$) and/or fruits

($N = 125$). However, there were no flowering specimens from East Africa (var. *usambaricum*) or from São Tomé. We measured 12 vegetative characters (all samples), 6 fruit characters and 9 floral characters (Tables 1–3). We performed a principal component analysis (PCA) to project and visualize trends in morphological variability across our samples, considering vegetative, floral, and fruit traits separately. We tested whether quantitative traits differed significantly between *a priori* groups using Kruskal-Wallis tests. When a test was significant for a trait, we tested again each pair of *a priori* groups and used letters to identify which groups differed significantly or not (see Tables 1–3). All statistical analyses were carried out using the PAST statistical software v.2.13 (Hammer & al., 2001).

**DNA extraction and genotyping. —** DNA from specimens stored in silica gel was extracted from 15 to 25 mg of leaf material or 30 mg of cambium material using the NucleoSpin 96 Plant II Kit (Macherey-Nagel, Düren, Germany). For herbarium specimens, DNA was extracted following a modified protocol of Doyle & Doyle (1987), where we added two phenol cleaning steps and used the QIAquick column purification kit (Qiagen, Venlo, the Netherlands). Eight microsatellite markers were amplified following Piñeiro & al. (2016). Genotyping was done in a 48-capillary sequencer (3730 DNA Analyzer, Applied Biosystems, Lennik, the Netherlands) using 1 µl of PCR product, 12 µl of HiDi formamide and 0.3 µl of GeneScan-500 LIZ Size Standard (Applied Biosystems, Warrington, U.K.). The resulting chromatograms were interpreted using the Peak Scanner software v.1.0 (Applied Biosystems) to identify the alleles (size of the amplified PCR products). In total, 800 *Greenwayodendron* samples (50 of which from herbarium vouchers) were successfully genotyped for at least five microsatellite loci. Preliminary analyses showed that one genetic group corresponding to the widespread var. *suaveolens* was largely over-represented (ca. 80% of the samples) and sampled at high density in some areas. Therefore, to better balance the sampling among *a priori* morphological groups and geographical regions (see later), we kept all samples associated with a herbarium voucher and no more than one sample per unit area of 1 km² for the best-represented *a priori* groups. This selection left us with 358 specimens for further analyses.

**Identification of genetic groups. —** We identified genetic groups using two different approaches. We found that the under-sampling carried out as explained above facilitated the identification of the least-represented genetic groups (results not shown). First, genetic clusters were identified using the Bayesian clustering algorithm implemented in STRUCTURE v.2.3.4 (Pritchard & al., 2000) without *a priori* grouping (i.e., *a priori* morphological groups were not taken into account). Second, we applied the FCA implemented in GENETIX v.4.0 (Belkhir & al., 2004) to identify genetic groups represented by few samples that may remain undetected by STRUCTURE analysis (Porras-Hurtado & al., 2013; Wang, 2017).

We ran STRUCTURE to delimit the most likely number $K$ of genetic clusters varying from 1 to 10 with 5 replicates per $K$ value. Each run was performed during 100,000 Markov Chain Monte Carlo iterations with a burn-in of 10,000 iterations. We used (i) the admixture model with the independent allele frequencies model (with all options set at their default values), as well as (ii) the alternative setting recommended by Wang (2017) for heterogeneous sample sizes among actual genetic clusters (admixture model with Alpha estimated for each population, initial Alpha set at 0.15, and independent allele frequencies model). To evaluate the number of clusters that best explain the dataset, we considered how the log-likelihood of the data ($\mathrm{Ln}(P)$) varied according to $K$, searching for the minimal $K$ before a plateau was reached, and we also applied the approach proposed by Evanno & al. (2005), using the online tool STRUCTURE HARVESTER v.0.9.94 (Earl & Von Holdt, 2012). We assigned an individual to a cluster $K_i$ ($i$ = 1, 2, 3, …, $K$) when the portion of its genome coming from this cluster was estimated at $q \geq 0.8$. As three *a priori* groups were represented by only 4 to 6 samples, we conducted additional analyses on artificial datasets to assess the power of the STRUCTURE algorithm for detecting genetic groups represented by few samples, as described in supplemental Appendix S1.

To further control the genetic coherence of each genetic cluster and identify individuals that could be genetically differentiated from the genetic clusters they were assigned to, we used the statistical software GENETIX to perform an FCA ordination of the samples and identify outliers. When such an outlier corresponded to one of the *a priori* groups, we considered them as forming a new genetic group. Quantum GIS v.2.8.1 (Quantum GIS Development Team, 2014) was used to display the geographic distribution of each genetic group identified.

**Differentiation between genetic groups. —** We estimated the genetic diversity of each genetic group by the number of alleles ($N_A$) as well as the observed and expected heterozygosities ($H_O$, $H_E$) and the inbreeding coefficient ($F_I$). The differentiation between genetic groups was characterized by computing the fixation index $F_{ST}$ and a similar measure of differentiation considering microsatellite allele sizes, $R_{ST}$. We used SPAGeDi v.1.5 (Hardy & Vekemans, 2002) to compute and compare these indices. When $R_{ST} > F_{ST}$, the differentiation between the genetic groups is probably of ancient origin because mutations, and not only genetic drift, have contributed to their differentiation (Hardy & al., 2013).

**Criteria to delimit species. —** With population genetic data, it is not always straightforward to distinguish distinct species from differentiated conspecific populations. We considered two situations to delimit species. First, if well-differentiated genetic groups display widely overlapping distribution ranges and none or very few admixed individuals occur (hard boundaries between clusters), genetic data provide strong support to consider them as different species following the BSC. If these groups do not show diagnostic morphological traits, they may correspond to cryptic species

**Table 1.** Vegetative quantitative characters of *Greenwayodendron* measured on herbarium specimens ($N = 233$) of the seven groups.

| Vegetative traits | Suaveolens ($N = 116$) | Oliveri ($N = 58$) | Gabonicum ($N = 27$) | Glabrum ($N = 28$) | Littorale ($N = 6$) | Usambaricum ($N = 6$) | São Tomé ($N = 5$) |
|---|---|---|---|---|---|---|---|
| Lped: length of petiole (mm) | 4.31 (2.00–8.00) ±1.36 / b | 3.12 (2.00–3.11) ± 0.856 / a | 6.93 (4.00–11.1) ± 2.293 / c | 4.08 (2.5–6.2) ± 1.025 / b | 2.32 (1.00–3.1) ± 0.786 / a | 5.25 (4.5–6) ± 0.589 / bc | 6.46 (5.2–7.3) ± 0.832 / c |
| Wped: width of petiole (mm) | 1.49 (1.00–2.5) ± 0.385 / b | 1.62 (1.00–2.33) ± 0.335 / c | 2.57 (1.5–3.3) ± 0.510 / d | 1.48 (1.00–2.2) ± 0.358 / bc | 0.98 (0.8–1.2) ± 0.160 / a | 1.83 (1.5–2.1) ± 0.242 / c | 2.1 (2–2.2) ± 0.1 / d |
| Lleaf: length of leaf blade (cm) | 9.59 (5.1–15.6) ± 2.148 / b | 11.10 (4.2–15.4) ± 2.450 / c | 18.27 (11.4–26.3 ± 3.579 / e | 11.78 (6.5–16.1) ± 2.451 / c | 5.78 (4.2–7.8) ± 1.455 / a | 14.16 (11.2–16.5) ± 1.972 / d | 14.92 (10.5–16.8) ± 2.535 / d |
| Wleaf: width of leaf blade (cm) | 3.84 (2–6.7) ± 0.983 / b | 4.68 (2.0–6.1) ± 1.01 / c | 6.83 (4.6–9.6) ± 1.402 / d | 4.33 (2.5–5.8) ± 0.961 / c | 2.90 (2–3.8) ± 0.807 / a | 5.43 (4.1–6.4) ± 0.995 / d | 6.46 (5.2–7.3) ± 0.832 / d |
| Lleaf/Wleaf | 2.55 (1.6–4.00) ± 0.416 / bc | 2.41 (1.5–3.3) ± 0.440 / b | 2.71 (2.1–3.7) ± 0.429 / cd | 2.75 (2.2–3.6) ± 0.349 / d | 2.03 (1.5–2.3) ± 0.266 / a | 2.63 (2.37–3.06) ± 0.245 / cd | 2.30 (2.01–2.46) ± 0.179 / c |
| Lapex: length of apex (mm) | 1.69 (0.6–4.8) ± 0.664 / b | 1.71 (1.00–3.3) ± 0.520 / b | 2.25 (1.3–3.5) ± 0.749 / c | 2.26 (1.4–4.00) ± 0.930 / c | 1.03 (0.5–1.2) ± 0.273 / a | 1.52 (1–2.5) ± 0.621 / b | 1.66 (1.5–2.2) ± 0.304 / b |
| Wapex: width of apex (mm) | 0.44 (0.3–0.6) ± 0.080 / c | 0.51 (0.3–07) ± 0.095 / d | 0.59 (0.5–0.7) ± 0.059 / e | 0.34 (0.2–0.5) ± 0.071 / b | 0.26 (0.2–0.3) ± 0.049 / a | 0.4 (0.3–0.5) ± 0.089 / c | 0.48 (0.3–0.7) ± 0.148 / c |
| Nveins: number of lateral veins | 10.47 (6–12) ± 1.254 / b | 11.14 (6–14) ± 1.849 / c | 13.74 (12–17) ± 1.607 / d | 11.43 (8–14) ± 1.794 / c | 8.17 (6–11) ± 1.941 / a | 15.66 (14–18) ± 1.366 / e | 16 (14–18) ± 1.414 / e |
| Dlveins: distance btw left lateral veins (mm) | 1.53 (0.8–2.2) ± 0.277 / b | 1.70 (1.1–2.6) ± 0.335 / c | 2.46 (1.4–3.5) ± 0.588 / e | 1.86 (1.4–2.5) ± 0.365 / d | 1.22 (1.00–1.4) ± 0.147 / a | 1.88 (1.3–2.8) ± 0.563 / d | 2.36 (1.7–2.8) ± 0.439 / e |
| Drveins: distance btw right lateral veins (mm) | 1.65 (0.8–2.4) ± 0.321 / b | 1.87 (1.3–2.8) ± 0.370 / c | 2.51 (1.8–3.5) ± 0.563 / e | 2.06 (1.6–2.8) ± 0.351 / d | 1.38 (1.1–1.6) ± 0.1726 / a | 1.61 (1.4–2.1) ± 0.248 / abc | 2.1 (1.4–2.5) ± 0.418 / cde |
| Nhair_leafLS: N hairs lower limb side (N/mm²)* | 1.59 (1.0–1.9) ± 0.188 / d | 1.30 (1–1.5) ± 0.145 / c | 1.97 (1.7–2.0) ± 0.082 / e | 0.44 (0–1.1) ± 0.394 / a | 0.89 (0.7–1.1) ± 0.159 / b | 1.43 (1.3–1.6) ± 0.175 / cd | 1.45 (1.3–1.6) ± 0.186/ cd |
| Nhair_midrib: N limb hairs midrib (N/mm²)* | 1.93 (1.3–2.0) ± 0.129 / d | 1.60 (1.3–1.7) ± 0.120 / c | 2.3 (2.3) ± 0 / e | 0.28 (0–0.8) ± 3 / a | 1.25 (1.1–1.3) ± 0.080 / b | 1.56 (1.3–1.7) ± 0.188 / c | 1.67 (1.6–1.7) ± 0.039 / c |

Note: M (X–Y) ± Z represent mean (minimum–maximum) ± standard deviation. Letters "a" to "e" summarize results of Kruskal-Wallis statistical tests where groups that do not differ significantly share a same letter.
*Log-transformed.

**Table 2.** Fruit quantitative traits of *Greenwayodendron* measured on herbarium specimens ($N = 125$) of the seven genetic groups.

| Fruit traits | Suaveolens ($N = 43$) | Oliveri ($N = 35$) | Gabonicum ($N = 22$) | Glabrum ($N = 22$) | Littorale ($N = 3$) | Usambaricum ($N = 2$) | São Tomé ($N = 2$) |
|---|---|---|---|---|---|---|---|
| Lfrped: length of fruit pedicel (mm) | 7.62 (5.5–12.00) ± 1.381 / b | 8.47 (6.00–13.00) ± 1.703 / c | 9.90 (7.00–13.1) ± 1.747 / d | 9.09 (6.00–13.00) ± 1.916/ cd | 5.33 (4.5–6.5) ± 1.041 / a | 12.5 (12–13) ± 0.707 / e | 11.5 (11–12) ± 0.707 / e |
| Wfrped: width of fruit pedicel (mm) | 1.83 (1.5–2.5) ± 0.283 / b | 1.39 (1.00–2.00) ± 0.280 / a | 2.62 (2.00–3.50) ± 0.551 / c | 1.70 (1.3–2.2) ± 0.279 / b | 1.83 (1.5–2.00) ± 0.280 / b | 2.25 (2–2.5) ± 0.353 / c | 1.4 (1.3–1.5) ± 0.141 / a |
| Lstip: length of stipe (mm) | 7.11 (4.5–10.0) ± 1.371 / bc | 6.50 (1.2–8.3) ± 1.452 / b | 7.91 (5.5–12.0) ± 1.503 / c | 7.48 (4.5–10.2) ± 1.442 / c | 3.50 (3.0–4.0) ± 0.500 / a | 6 (5.5–6.5) ± 0.707 / b | 4.25 (4–4.5) ± 0.353 / a |
| Wstip: width of stipe (mm) | 1.79 (1.0–2.5) ± 0.392 / b | 1.25 (1.0–2.2) ± 0.340 / a | 2.23 (1.5–3.1) ± 0.434 / c | 1.95 (1.1–3.2) ± 0.684 / bc | 1.50 (1.0–2.0) ± 0.500 / ab | 1.1 (1–1.2) ± 0.141 / a | 1.1 (1–1.2) ± 0.141 / a |
| Dfruit: diameter of fruit (mm) | 11.18 (7.6–16.4) ± 1.926 / c | 5.56 (3.25–7.9) ± 1.018 / b | 16.06 (12.5–19.25) ± 2.129 / d | 14.95 (11.1–20.95) ± 2.227 / d | 3.67 (2.75–4.15) ± 0.794 / a | 11.9 (11.6–12.2) ± 0.424 / c | 10.15 (10.1–10.2) ± 0.070 / c |
| Dseed: diameter of seed (mm) | 7.34 (4.00–11.1) ± 1.174 / c | 4.58 (2.7–6.25) ± 0.921 / b | 10.61 (7.9–15.35) ± 2.078 / d | 9.98 (6.5–12.5) ± 1.823 / d | 2.97 (1.75–3.9) ± 1.103 / a | 11.07 (11–11.15) ± 0.106 / d | 8.25 (8.6–9.1) ± 0.353 / c |

Note: M (X–Y) ± Z = mean (minimum–maximum) ± standard deviation. Letters "a" to "e" summarize results of Kruskal-Wallis statistical tests where groups that do not differ significantly share a same letter.

**Table 3.** Floral quantitative characters of *Greenwayodendron* measured on herbarium specimens ($N = 41$) of the seven genetic groups.

| Floral traits | Suaveolens ($N = 20$) | Oliveri ($N = 15$) | Gabonicum ($N = 4$) | Glabrum ($N = 1$) | Littorale ($N = 1$) | Usambaricum ($N = 0$) | São Tomé ($N = 0$) |
|---|---|---|---|---|---|---|---|
| Lflped: length of floral pedicel (mm) | 5.09 (3–6.3) ± 1.066 / a | 5.31 (3.8–6.2) ± 0.817 / a | 5.25 (4.5–6.00) ± 0.866 / a | 4 / | 3.1 / | – | – |
| Wflped: width of floral pedicel (mm) | 1.30 (0.8–2.1) ± 0.331 / b | 0.94 (0.5–1.3) ± 0.239 / a | 2.10 (2–2.2) ± 2.2 / c | 1.7 / c | 0.66 / a | – | – |
| Lsep: length of sepal (mm) | 2.90 (1.85–3.55) ± 0.53 / b | 1.71 (1.4–2.35) ± 0.29 / a | 3.78 (3.65–3.95) ± 0.08 / c | 3.05 / b | 1.7 / a | – | – |
| Wsep: width of sepal (mm) | 2.90 (1.9–3.6) ± 0.532 / b | 1.71 (1.4–2.35) ± 0.291 / a | 3.78 (3.7–4.00) ± 0.156 / c | 3.05 / b | 1.7 / a | – | – |
| Lpet: length of petal (mm) | 13.05 (4.7–17) ± 2.718 / b | 11.66 (9.00–17.00) ± 2.115 / a | 23.64 (22.8–24.00) ± 0.567 / c | 12.5 / b | 12 / a | – | – |
| Wpet: width of petal (mm) | 1.87 (1.45–2.33) ± 0.268 / b | 1.89 (1.55–2.15) ± 0.186 / a | 2.78 (2.5–3.00) ± 0.206 / c | 2.25 / c | 1.7 / a | – | – |
| Lstam: length of stamens (mm) | 3.11 (2.00–4.22) ± 0.486 / c | 1.83 (1.5–2.15) ± 0.207 / a | 3.77 (3.50–4.15) ± 0.280 / c | 1.35 / | 1.9 / | – | – |
| Wstam: width of stamens (mm) | 0.57 (0.5–0.73) ± 0.063 / a | 0.59 (0.45–0.83) ± 0.104/ a | 0.54 (0.5–0.58) ± 0.032 / a | 0.35 / | 0.78 / | – | – |
| Dbract: diameter of bract (mm) | 1.95 (1.4–3.1) ± 0.555 / b | 1.33 (1.2–1.5) ± 0.103 / a | 4.98 (4.8–5.2) ± 0.171 / c | 2.2 / b | 0.9 / a | – | – |

Note: M (X–Y) ± Z = mean (minimum–maximum) ± standard deviation; "a" to "c" " summarize results of Kruskal-Wallis statistical tests where groups that do not differ significantly share a same letter.

*sensu stricto*, but if they show diagnostic morphological traits, they would also correspond to distinct species following the MSC. Second, if genetic groups occur in parapatry or allopatry, they might correspond to differentiated conspecific populations, as previously inferred within the taxon *G. suaveolens* subsp. *suaveolens* var. *suaveolens* (Piñeiro & al., 2017), or to different species. The level of genetic and morphological differentiation is then taken into account. More specifically, we consider such groups as distinct species if (i) diagnostic morphological traits occur, *and* (ii) the degree of genetic differentiation ($F_{ST}$, $R_{ST}$) is comparable to, or higher than, that which is observed between other pairs of species from the same genus.

## ■ RESULTS

**Definition of *a priori* groups. —** Seven *a priori* groups were recognized based on their morphological and/or geographical characteristics. The names attributed to these *a priori* groups were derived from one of the four recognized taxon names when they matched one, otherwise we used a characteristic of the group. The large majority of *Greenwayodendron* samples came from Central Africa where different morphogroups could be distinguished, as described hereafter. Nevertheless, other samples came from allopatric areas and led us to recognize three *a priori* groups based on their geographic origin and morphology (Fig. 1). (1) 44 samples from West Africa were called "Oliveri" as they corresponded to the species *G. oliveri*. (2) 7 samples from East Africa were called "Usambaricum" as they corresponded to the subspecies *G. suaveolens* subsp. *usambaricum*. (3) 7 samples from the islands of São Tomé and Príncipe were called "São Tomé". Among the Central African samples, we distinguished four additional *a priori* groups based on leaf traits. (4) 133 samples had medium-sized (ca. 9–18 cm long), slightly pubescent leaves corresponding to the taxon *G. suaveolens* subsp. *suaveolens* var. *suaveolens* and were therefore called "Suaveolens". (5) 98 samples (all from Gabon) presented larger (>18 cm), thick and heavily pubescent leaves corresponding to the taxon *G. suaveolens* subsp. *suaveolens* var. *gabonica* and were therefore called "Gabonicum". (6) 5 samples found along the coast of Gabon had distinctly small leaves (<9 cm), very different from those of *G. oliveri* although these samples were initially determined as *G. oliveri* by their collectors, so that this *a priori* group was called "Littorale". (7) Finally, 69 samples from south-western Cameroon, Equatorial Guinea and western Gabon had completely glabrous leaves (otherwise similar in shape and size to "Suaveolens") and were called "Glabrum".

**Morphological differentiation between *a priori* groups. —** For the 12 vegetative traits measured on the 233 herbarium specimens, axes 1, 2 and 3 of the PCA together accounted for more than 50% of the total variance (Fig. 2A, B). Axis 1 was determined mainly by leaf and petiole size traits (leaf length and width, and petiole length and width; 32.6% of relative contribution), axis 2 by the density of hairs on the lower side of the limb and on the midrib (17.3% of relative contribution), and axis 3 by the number of lateral

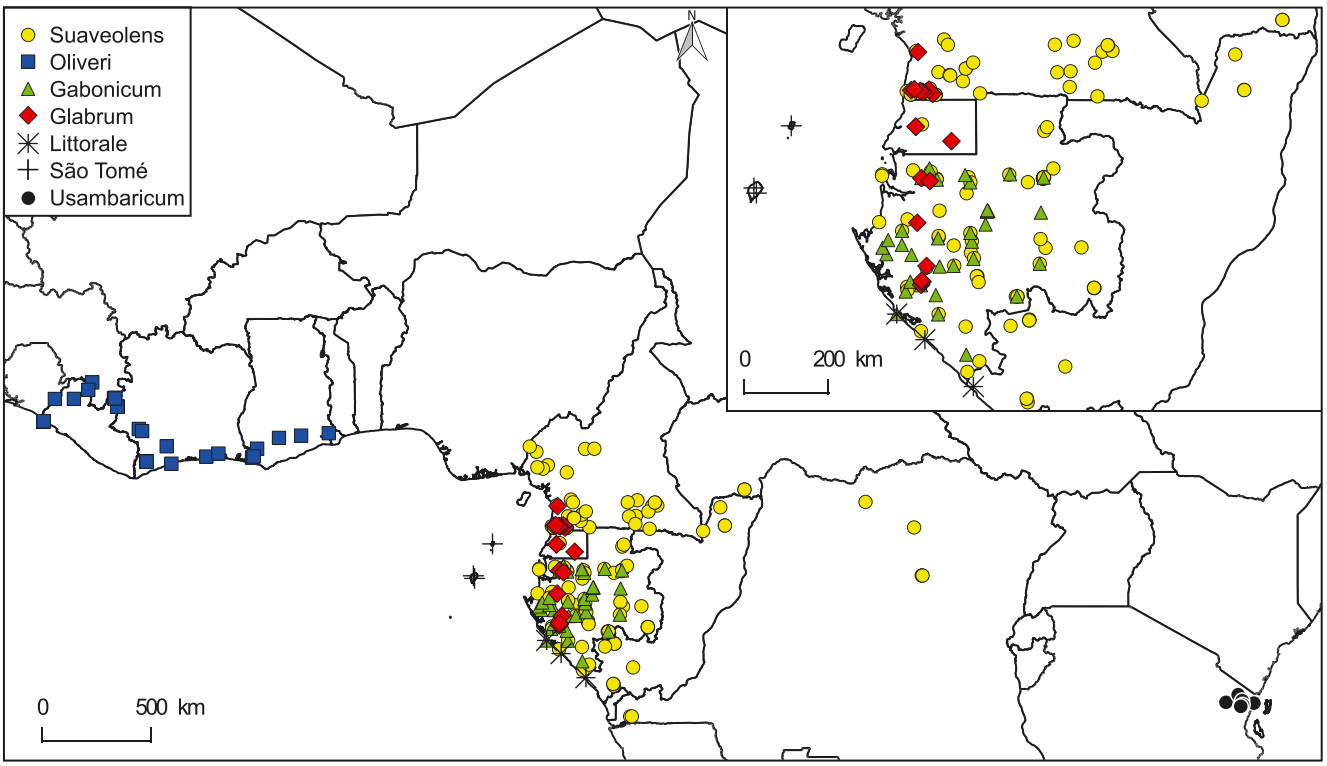


**Fig. 1.** Distribution of the seven *a priori Greenwayodendron* groups based on morphological and geographical criteria.

veins (11.3% of relative contribution). These three axes allowed us to distinguish five main groups of specimens: (1) a group with high scores along axis 1 characterized by large leaves and corresponding to Gabonicum; (2) a group with low scores along axis 1 characterized by small leaves and corresponding to Littorale; (3) a group with low scores along axis 2 characterized by low density of hairs on the main veins and on the underside of the leaves, corresponding to Glabrum; (4) a group with high scores along axis 3 characterized by a high number of secondary veins, corresponding to Usambaricum and São Tomé; and (5) a group consisting of the rest of the specimens with low scores along all three axes, corresponding to Oliveri and Suaveolens.

In the PCA ordination of six quantitative fruit traits based on 125 specimens, axis 1 (53.3% of relative contribution) was determined mainly by traits related to fruit size (diameter of fruit and of seed, width of stipe and of pedicel), while axis 2 (15.6%) was mostly loaded by the length of fruit pedicel and of stipe (Fig. 2C). This ordination did not display clearly isolated groups but *a priori* groups tended to segregate, except Glabrum and Suaveolens.

For the PCA ordination of the nine quantitative floral characters based on 41 specimens, axes 1 and 2 accounted for 65.5% of the total variance (Fig. 2D). Only Oliveri, Suaveolens and Gabonicum were represented by multiple samples and they segregated well along axis 1 which was mainly loaded by the length of petals. Gabonicum showed much larger petals than Oliveri. Suaveolens was intermediate between Oliveri and Gabonicum along axis 1 but was distinguishable from the other two groups when considering the first two axes (Fig. 2D): characters such as length of sepal, length of fruit pedicel and diameter of bract were smaller than in Gabonicum but larger than in Oliveri, as already mentioned by many authors (Le Thomas, 1969; Verdcourt, 1969; Dauby, 2012). The specimens belonging to Gabonicum, Glabrum, Littorale and Suaveolens had a tongue-shaped or lobulated connective while connectives of the stamens in Oliveri were crushed-flattened, as mentioned previously for *G. oliveri* (Aubréville, 1962; Le Thomas, 1969; Verdcourt, 1969). Note that Glabrum and Littorale were each represented by only one flowering specimen while Usambaricum and São Tomé were not included in this analysis because we lacked flowering specimens so that we cannot conclude whether they carry diagnostic floral characters (Fig. 2D).

Comparison of paired medians in the Kruskal-Wallis tests showed that the quantitative variables (vegetative, fruit and floristic) differ significantly among the seven *a priori* groups (Tables 1–3). Gabonicum differs statistically from Littorale, Oliveri and Suaveolens for 11 of the 12 vegetative variables,

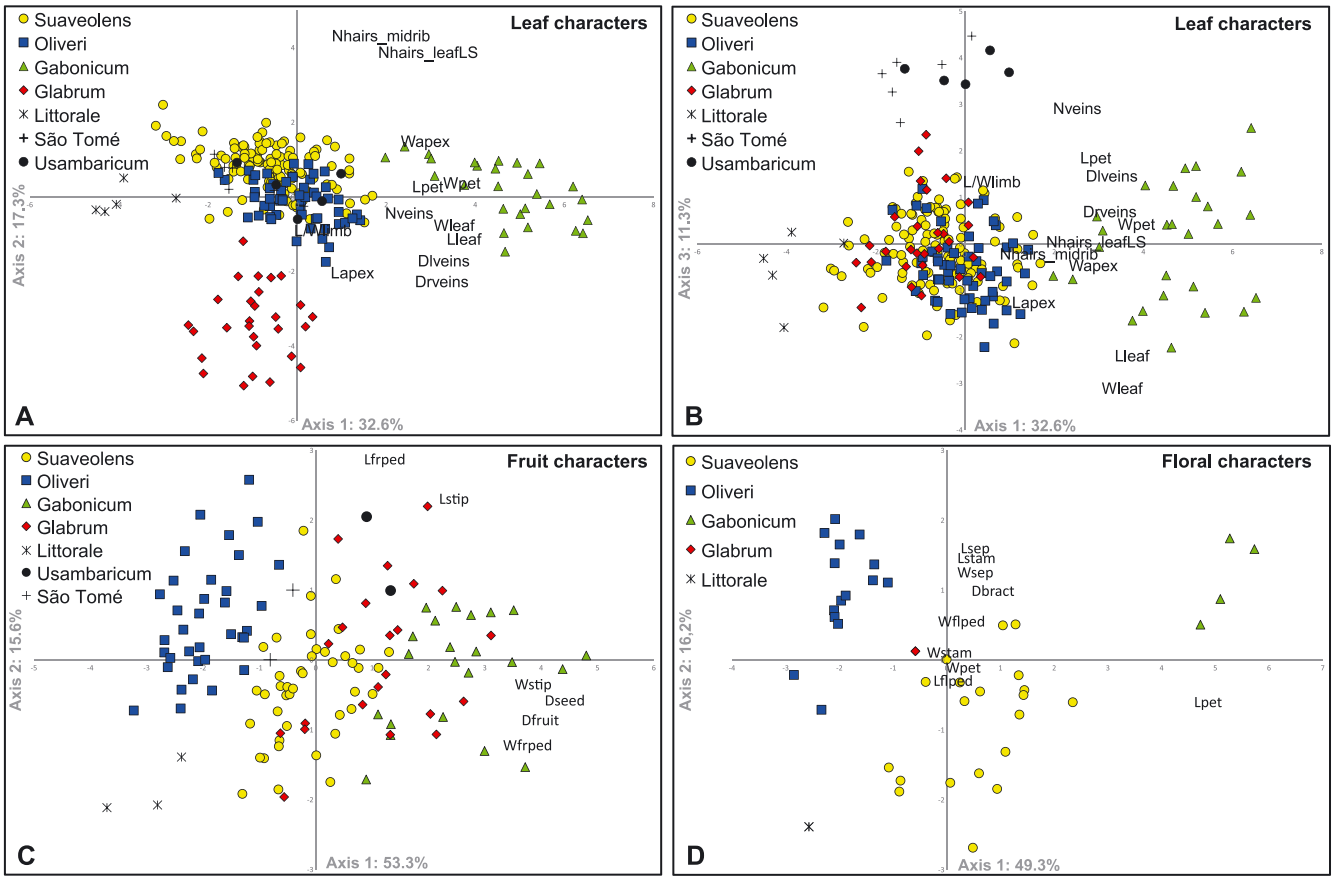


**Fig. 2.** Biplot ordinations of *Greenwayodendron* samples and quantitative variables along the first axes of principal component analyses based on **(A)** 12 leaf variables ($N = 233$, axes 1 and 2), **(B)** 12 leaf variables ($N = 233$, axes 1 and 3), **(C)** 6 fruit variables ($N = 125$, axes 1 and 2), and **(D)** 9 floral variables ($N = 41$, axes 1 and 2).

5 of the 6 fruit traits and 6 of the 9 floral variables. Indeed, Gabonicum has longer leaves and a higher density of hairs (on the midrib and lower side of the leaf) than the other groups. For fruit and floral variables, Gabonicum also has larger fruits and seeds as well as longer sepals, petals and bracts than the other groups. Glabrum differs statistically from all other groups by the absence of hairs on the midrib and lower side of the leaf. The groups Usambaricum and São Tomé present a higher number of lateral veins as well as longer fruit pedicels than the other five groups. Littorale and Oliveri differ significantly from other groups by their smaller fruits and seeds, and Littorale is also characterized by smaller leaves with fewer lateral veins than the other groups.

**Identification of genetic groups. —** The Bayesian structuring analysis using eight nuclear microsatellites performed with STRUCTURE (Pritchard & al., 2000) showed that the likelihood of the data increased substantially with the number of imposed clusters $K$ until $K = 4$, while a plateau was reached for larger $K$ (Fig. 3A). Application of the delta $K$ method of Evanno & al. (2005) also highlighted $K = 4$ as the most likely number of clusters (Fig. 3B). The percentage of samples assigned to a cluster at $q \geq 0.8$ reached 94.4% and was high for all *a priori* groups except for Usambaricum where three of the four samples genotyped were unassigned. We found a very good match between the genetic clusters and the *a priori* groups represented by large sample sizes (Fig. 4). Specifically, cluster K1 matched Suaveolens, K2 Oliveri, K3 Glabrum and K4 Gabonicum. *A priori* groups represented by few samples were assigned to K1 except three unassigned samples of Usambaricum that would be assigned to K2 using a $q \geq 0.75$ threshold. Although the correspondence between genetic clusters and the *a priori* groups was not perfect, 99.1% of the individuals assigned to a cluster were assigned as described above. Running STRUCTURE with the settings recommended by Wang (2017) to deal with unequal sample sizes across species, led

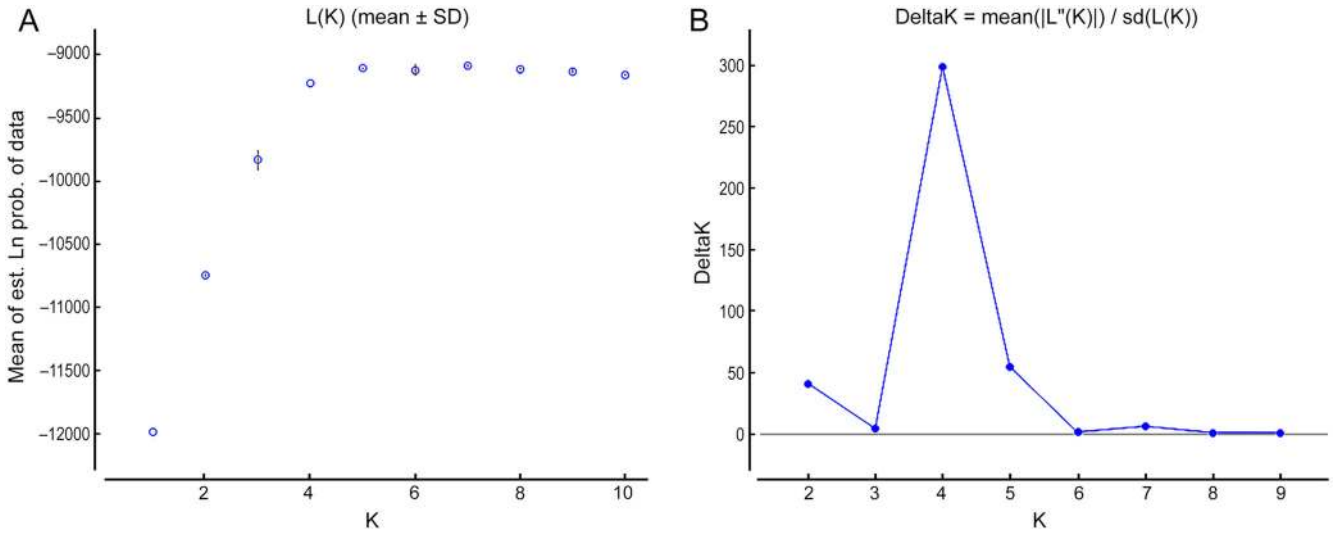


**Fig. 3.** Variation of the likelihood of the data, L(K) **(A)** and of DeltaK **(B)** as a function of the number of genetic groups ($K$) identified in 358 *Greenwayodendron* samples using the software STRUCTURE.

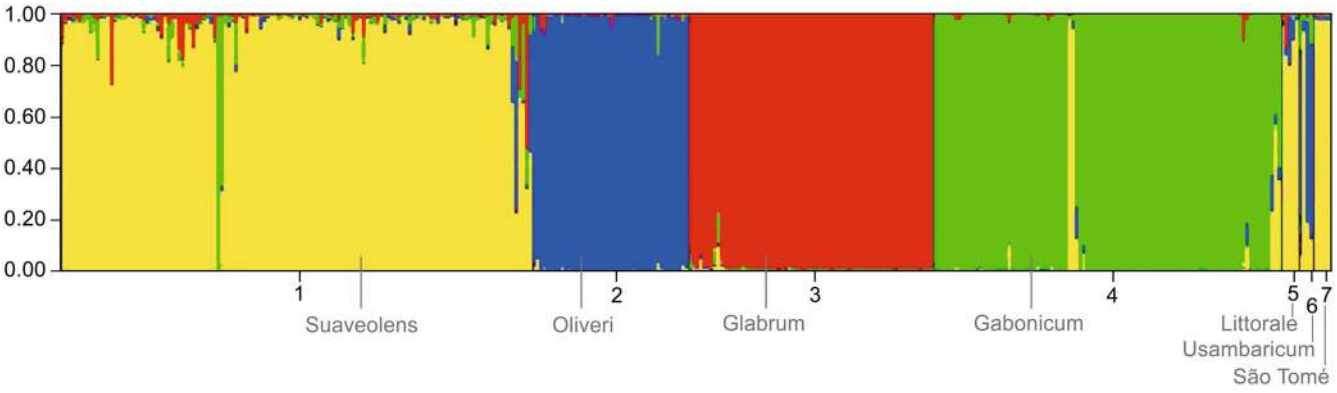


**Fig. 4.** Histogram of genetic assignment of 358 *Greenwayodendron* individuals at $K = 4$. Individuals were ordered along the horizontal axis according to their *a priori* group. The vertical bars indicate how the genome of each individual is partitioned into the four different clusters (K1 yellow, K2 blue, K3 red and K4 green).

to similar results. However, some differences between the two settings appeared when investigating the clustering solutions obtained at higher *K*. For $K=5$, under the default setting but with the independent allele frequency model, the *a priori* group Suaveolens became subdivided into two clusters with many intermediates, corresponding to parapatric clusters described in Piñeiro & al. (2017) for the taxon *G. suaveolens* subsp. *suaveolens* var. *suaveolens*. Under the settings recommended by Wang (2017), five of the ten runs presented the same clustering pattern as above while the other five runs did not subdivide the Suaveolens group but added a cluster made of three samples of Littorale, one sample of Suaveolens and one sample of Gabonicum.

Our tests on data subsets, keeping 130 samples of Suaveolens assigned to K1 and from 5 to 30 samples of each of Oliveri, Glabrum and Gabonicum assigned to K2, K3 and K4, respectively, showed that the least-represented *a priori* groups had to be represented by a minimum of 20 or 30 individuals to be identified as distinct clusters by STRUCTURE (suppl. Appendix S1, suppl. Figs. S1, S2). This confirms that the STRUCTURE algorithm generally fails to identify genetic groups represented by few individuals when one group is well represented (Porras-Hurtado & al., 2013; Wang, 2017).

We further analysed our dataset using FCA to assess whether the *a priori* groups represented by few samples diverge genetically from the other groups. FCA summarizes the genetic diversity found in a sample of genotypes along a limited number of axes. When a sample is made of distinct genetic groups represented by highly unequal sample sizes, the first axes of an FCA may highlight the genetic variation occurring within well-sampled groups rather than the genetic differentiation of groups represented by few individuals. To avoid this, we re-sampled randomly 10 individuals from each of the well-sampled *a priori* groups that were assigned to clusters K1 to K4 and added all the available samples from the three other *a priori* groups, making a total of 55 individuals. The first two axes of the FCA, representing 6.2% and 6.1% of the genetic variation, discriminate fairly well the seven *a priori* groups (Fig. 5). In particular, they show that Usambaricum is well differentiated from Oliveri although STRUCTURE had grouped them in cluster K2 (Fig. 4). Littorale is well separated from Suaveolens while they were assigned to cluster K1, and São Tomé also tends to be differentiated from Suaveolens although they remain close on the first two axes.

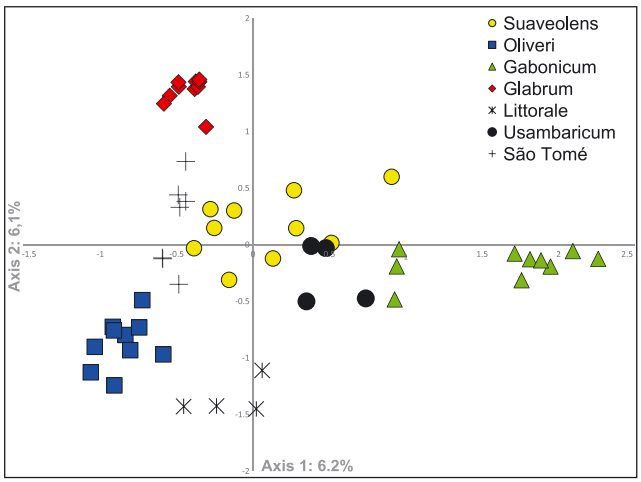


**Fig. 5.** Factorial correspondence analysis (FCA) of microsatellite data from 55 individuals of the genus *Greenwayodendron* representing seven *a priori* groups defined on basis of geographical and/or morphological traits. To ensure that the first axes of this ordination reflected genetic variation occurring between rather than within genetic groups, we used a maximum of 10 randomly chosen samples per *a priori* group.

**Allelic diversity of each genetic group and genetic differentiation. —** According to the STRUCTURE analysis eight samples were apparently misclassified (Fig. 4). A re-examination of the leaf traits of misclassified samples for which a herbarium or at least a leaf was available showed that they had characters intermediate between the group they were initially attributed to and the group deduced from their genotype. They were therefore re-attributed according to the genetic assignation for the following analyses.

The genetic diversity per genetic group varied substantially: Suaveolens displayed by far the highest diversity ($H_E$ = 0.85) and Glabrum the lowest ($H_E$ = 0.34, with only five of the eight loci showing polymorphism), while the five other groups displayed intermediate diversity ($H_E$ ranged from 0.60 to 0.68; Table 4A). Pairwise $F_{ST}$ between genetic groups ranged from 0.138 to 0.571, and $R_{ST}$ ranged from 0.062 to 0.823. Over 21 pairwise comparisons, $R_{ST}$ was significantly larger than $F_{ST}$ in 13 cases, indicating that the accumulation of stepwise microsatellite mutations has contributed to the differentiation of these genetic groups (Table 4B). A relatively low level of differentiation and no phylogeographic signal was, however, observed between Suaveolens and São Tomé ($F_{ST}=0.138$, $R_{ST}=0.062$).

## ■ DISCUSSION

In previous revisions of the genus *Greenwayodendron*, the characters used for identifying different taxa were related to leaf and reproductive organ morphology (Le Thomas, 1969; Verdcourt, 1969). However, previous preliminary genetic studies suggested that this genus might contain several undescribed cryptic species (Dauby & al., 2010; Piñeiro & al., 2016, 2017). We used, for the first time in this genus, multivariate analyses of morphological traits, which have been applied successfully to guide species delimitation in other tropical taxa (e.g., Daïnou & al., 2014, 2016; Ikabanga & al., 2017).

To delineate species, we consider two criteria: (i) the occurrence of diagnostic morphological traits (classical morphological species concept, MSC) and (ii) evidence of reproductive isolation from other species (biological species concept, BSC). Our evidence for BSC is deduced from our population genetics approach: if well-distinct genetic groups occur in sympatry and admixed genotypes are non-existent or very rare in regions where their respective distributions overlap, one can deduce that they are not inter-fertile (except for

occasional hybrids) and belong to distinct species according to the BSC. For genetic groups found in allopatry, the interpretation is not as straightforward because they may potentially correspond to different populations from a same species rather than to different species, and there is no recognized genetic differentiation threshold (e.g., minimal $F_{ST}$ or $R_{ST}$) that can distinguish between the two situations. Nevertheless, the level of genetic differentiation observed between sympatric species of *Greenwayodendron* can help us assess whether genetic entities found in allopatry belong to the same or to distinct species.

**Species delimitation in *Greenwayodendron*. —** Four morphological and genetic groups were identified in Central Africa. One of them, Suaveolens, is by far the most widespread and abundant and, in Gabon, it can be found in sympatry with three other groups: Gabonicum, Glabrum and Littorale (Fig. 1). The marked genetic differentiations between these four groups (Figs. 2–4, Table 4B), even in contact zones, indicate that they correspond to distinct species following the BSC.

Morphologically, the PCA on vegetative, fruit and floral traits revealed a first group of specimens distinct from Suaveolens and attributed to Gabonicum. This taxon shows a clear morphological differentiation by the large dimensions of its leaves, flowers and fruits, but also by its dense pubescence on both sides of the limb (Fig. 2). All these discriminating traits had already been recognized by several authors to distinguish species (Le Thomas, 1969; Verdcourt, 1969; Dauby & al., 2010; Dauby, 2012; Piñeiro & al., 2016) and support the view that the taxon *G. suaveolens* subsp. *suaveolens* var. *gabonica* can be elevated to the rank of species, as already proposed by Dauby & al. (2010).

The PCA of leaf traits separated a second group of samples belonging to the morphogroup Glabrum (Fig. 2A). A preliminary genetic analysis had shown that in the western part of south Cameroon, samples with glabrous leaves and those with slightly pubescent leaves were well differentiated at microsatellite markers (O.J. Hardy & B.J. Lissambou, unpub.). Following this observation, we considered leaf pubescence when checking all herbarium samples and found glabrous specimens in southern Cameroon and in Gabon. The samples of this group are characterized by the absence of hairs on both sides of the limb but also at the level of the midrib and the petiole. Genetically they form a well-differentiated genetic cluster with very low genetic diversity. We, therefore, propose that this morphogroup should be recognized at the rank of species. This is consistent with the fact that foliar characteristics are the most used features for distinguishing tropical tree taxa (Aubréville, 1962; Hawthorne & Jongkind, 2006), and pubescence can be used to distinguish different species in Annonaceae (Le Thomas, 1969).

The PCA ordination circumscribed a third Central African morphogroup, Littorale, present only along coastal forests of Gabon and the Republic of the Congo. This morphogroup is characterized by small trees with relatively small leaves and fruits but also by tongue-shaped short stamen connectives.

**Table 4.** Genetic diversity **(A)** and differentiation **(B)** parameters for the seven genetic groups of the genus *Greenwayodendron* detected using eight microsatellite markers (SSR).

| **A,** Diversity parameters. | | | | | | | |
|---|---|---|---|---|---|---|---|
| Genetic group | Sample size | $N_A$ | $H_E$ | $H_O$ | $F_I$ | | |
| Oliveri | 47 | 8 | 0.602 | 0.353 | 0.428 | | |
| Glabrum | 68 | 5 | 0.337 | 0.228 | 0.337 | | |
| Gabonicum | 94 | 10 | 0.608 | 0.413 | 0.333 | | |
| Suaveolens | 128 | 20 | 0.854 | 0.681 | 0.203 | | |
| Littorale | 5 | 4 | 0.679 | 0.500 | 0.290 | | |
| Usambaricum | 4 | 3 | 0.638 | 0.536 | 0.141 | | |
| São Tomé | 7 | 4 | 0.600 | 0.422 | 0.335 | | |
| **B,** Differentiation parameters. | | | | | | | |
| $F_{ST}\backslash R_{ST}$ | Oliveri | Glabrum | Gabonicum | Suaveolens | Littorale | Usambaricum | São Tomé |
| Oliveri | | 0.823* | 0.821* | 0.493* | 0.626* | 0.587 ns | 0.652 * |
| Glabrum | 0.506 | | 0.466 ns | 0.292 ns | 0.781 ns | 0.507 ns | 0.630 ns |
| Gabonicum | 0.347 | 0.495 | | 0.369* | 0.755* | 0.615* | 0.644* |
| Suaveolens | 0.206 | 0.307 | 0.163 | | 0.189* | 0.311* | 0.062 ns |
| Littorale | 0.291 | 0.571 | 0.302 | 0.145 | | 0.576* | 0.173 ns |
| Usambaricum | 0.258 | 0.508 | 0.234 | 0.151 | 0.242 | | 0.397* |
| São Tomé | 0.341 | 0.508 | 0.346 | 0.138 | 0.282 | 0.187 | |

$N_A$: mean number of alleles observed per locus; $H_E$: expected heterozygosity (unbiased estimator); $H_O$: observed heterozygosity; $F_I$: inbreeding coefficient. $F_{ST}$ and $R_{ST}$ are classical measures of genetic differentiation based on SSR allele identity and allele sizes, respectively. The asterisks (*) following the $R_{ST}$ values show cases where $R_{ST}$ is significantly larger than $F_{ST}$ ($P < 0.05$), indicating that SSR mutations contributed to the genetic differentiation of the compared groups.

This group was generally identified as *G. oliveri* by collectors of samples, but it appears morphologically and genetically very distinct from our Oliveri group. The quantitative and qualitative diagnostic characteristics of floral and fruit traits, along with the observed genetic differentiation of the four genotyped individuals, lead us to propose that the Littorale morphogroup be considered as a species.

Thus, in Central Africa, the recognition of four species is well supported by both the occurrence of diagnostic morphological traits and clear-cut discontinuities of genetic variation in areas of sympatry. We now consider the three remaining *a priori* groups that are allopatric to all the other ones.

The PCA of fruit and floral traits distinguished a group of individuals assigned to Oliveri, present only in West Africa, and displaying small fruits, though not as small as in Littorale. In addition to these quantitative traits, Oliveri presents connectives of stamens with a crushed-flattened shape. This diagnostic trait had already been identified to distinguish this species from *G. suaveolens* (Aubréville, 1962; Le Thomas, 1969). Given that this group was already recognized as a distinct species for which we can corroborate its morphological differentiation and that it is well differentiated genetically from all other groups, we thus confirm the species status of *G. oliveri*, which is endemic to Upper Guinea.

Two other *a priori* groups, Usambaricum and São Tomé, well isolated geographically from all other groups, differentiated from them along axis 3 of the PCA of vegetative traits, which was essentially determined by the number of veins (Fig. 2B). Univariate tests (Tables 1, 2) indicate that these morphogroups differ significantly from Suaveolens by at least nine vegetative and fruit characters (number of lateral veins, width of petiole, length of leaf blade, width of leaf blade, distance between lateral veins, number of hairs on midrib, length and width of fruit pedicel, width of stipe). However, the only quantitative variable that does not show overlap is the number of lateral veins (≤12 veins in Suaveolens and ≥14 in São Tomé and Usambaricum), a type of trait which is nevertheless diagnostic to differentiate for example *Milicia excelsa* (Welw.) C.C.Berg from *M. regia* (A.Chev.) C.C.Berg (Daïnou & al., 2014). Unfortunately, no floral traits could be examined for these groups and fruits are rather similar. Genetic data are harder to interpret in terms of species delimitation when they concern allopatric entities because population differentiation within species can sometimes reach high values. Nevertheless, both Usambaricum and São Tomé are genetically closest to Suaveolens so that we can compare their level of differentiation with Suaveolens to the differentiation between the latter and the other species recognized so far. According to $R_{ST}$, Usambaricum is more divergent from Suaveolens than Glabrum or Littorale are from Suaveolens. Hence, there is genetic support to consider Usambaricum as a distinct species, and thus recognize *G. suaveolens* subsp. *usambaricum* at the rank of species. This is in fact also supported by plastid and nuclear DNA sequences indicating that Usambaricum does not fall within the clade including all Suaveolens (Couvreur & al., 2019; Migliore & al., 2019).

The status of the São Tomé group is less straightforward. This group is least genetically differentiated from Suaveolens compared to all other groups, so that we cannot exclude that it is simply a divergent population from that species. Although it differs from Suaveolens in vegetative and fruit traits as much as Usambaricum differs from Suaveolens, we consider here that additional data from herbarium fertile specimens will be needed to conclude on the taxonomic status of this group.

**The identification of genetic groups when sample sizes are heterogeneous. —** The Bayesian clustering of *Greenwayodendron* samples using eight microsatellite loci led to the recognition of four genetic clusters corresponding to four well-sampled and widespread morphogroups (Fig. 2). However, the STRUCTURE algorithm did not form separate genetic clusters for individuals of Littorale, São Tomé and Usambaricum, despite further analyses (FCA ordination; $F_{ST}$ and $R_{ST}$ differentiation metrics), confirming that these groups are genetically differentiated. This is most likely due to their low sample sizes ($N$ = 4 to 7) while the other groups had much higher sample sizes, a problem already reported for heterogeneous sampling (e.g., Porras-Hurtado & al., 2013; Wang, 2017). This probably explains why Littorale and São Tomé were assigned to the same cluster as Suaveolens, and Usambaricum was clustered with Oliveri. Wang (2017) identified the origin of the problem on theoretical grounds and proposed alternative settings when running the STRUCTURE algorithm. Tested on our dataset and artificial datasets (suppl. Appendix S1), these alternative settings indeed improved the detection of clusters represented by fewer samples, but at least 20 samples per actual genetic group were still needed with our microsatellites (suppl. Figs. S1, S2). This might be caused by the fact that allele frequencies cannot be estimated correctly by the STRUCTURE algorithm using smaller sample sizes given the large number of alleles per locus. It is thus possible that SNP data at many loci could be more powerful than microsatellites because a genetic group represented by few samples might still be identified as a distinct genetic cluster if diagnostic alleles are fixed at a few loci.

Despite the inherent limits of the clustering algorithm used, multivariate ordination of our genetic data using FCA allowed us to discriminate all our *a priori* groups according to the first two axes when the most sampled groups had been subsampled to achieve fairly balanced sample sizes. Hence, combining a Bayesian clustering method with an ordination is useful, especially when other data (here morphology) provide prior hypotheses about species delineation that can be tested.

**Variation of genetic diversity among genetic groups. —** Intra-population polymorphism indices varied substantially among groups, with high diversity in Suaveolens and low diversity in Glabrum, suggesting that the latter corresponds to a recently formed new species (founder event) or that it suffered a strong demographic bottleneck in the recent past. Within each genetic group, the mean observed heterozygosity

($H_O$) is lower than the mean expected heterozygosity ($H_E$) under Hardy-Weinberg conditions, leading to positive inbreeding coefficients ($F_I$), which probably results from genetic substructure, as observed by Piñeiro & al. (2017) in Suaveolens, but also potentially from the impact of null alleles.

Genetic differentiation between each pair of genetic groups was evaluated by $F_{ST}$ which varied from 0.138 to 0.571. The minimum $F_{ST}$ is observed between the genetic groups Suaveolens and São Tomé while the maximum $F_{ST}$ value was reached between the genetic groups Glabrum and Littorale (Table 4B). $R_{ST}$ was higher than $F_{ST}$, indicating ancient divergence, in most pairwise comparisons involving Oliveri, Gabonicum and Littorale, which provides additional support to recognize them as distinct species. The results of these tests were, however, not significant in most pairwise comparisons involving Glabrum, which may result from limited testing power when allelic polymorphism is very low, as observed in Glabrum. The test results were also non-significant between São Tomé and Suaveolens, consistent with the limited morphological differentiation between these groups; but significant between Usambaricum and both São Tomé and Suaveolens, highlighting the ancient divergence of this isolated East African group.

## ■ CONCLUSION

Our morphometric and genetic analyses have shown that *Greenwayodendron* includes a greater number of species than previously thought. Indeed, based on our integration of morphological and genetic data, we have identified two new species (Glabrum and Littorale) and showed that *G. suaveolens* subsp. *suaveolens* var. *gabonica*, *G. suaveolens* subsp. *suaveolens* var. *suaveolens* and *G. suaveolens* subsp. *usambaricum* also deserve to be considered as distinct species. A formal taxonomic description of these species is handled in a separate publication (Lissambou & al., 2018). The status of a group of specimens, all gathered in São Tomé and Príncipe, remains inconclusive. This is partly related to the little fertile material we have to date. Future collections on these islands will better classify this group. Beyond the case of *Greenwayodendron*, our study illustrates the strength of combining morphological and population genetics data, as it shows that new taxa can be discovered using these approaches. It also adds to mounting evidence that the current taxonomic delineation of African tree species tends to underestimate the actual number of species (e.g., Daïnou & al., 2016; Bouka, 2017; Ikabanga & al., 2017).

## ■ AUTHOR CONTRIBUTIONS

BJL, TLPC, CA, BS, BM and OJH designed the research. BJL, TLPC, TS, GD, FKM, DUI, BS and OJH contributed to the sampling and morphological analyses. BJL and RP performed the genotyping. BJL and OJH performed the genetic data analyses. BJL, TLPC and OJH wrote the first draft of the manuscript. All authors contributed to the interpretation of the results and the final writing of the manuscript. — TLPC, https://orcid.org/0000-0002-8509-6587; RP, https://orcid.org/0000-0002-4679-7305; GD, https://orcid.org/0000-0002-9498-413X; FKM, https://orcid.org/0000-0003-4664-658X; OJH, https://orcid.org/0000-0003-2052-1527

## ■ ACKNOWLEDGEMENTS

We thank the Agence Universitaire de la Francophonie (AUF), the Agence Nationale des Bourses du Gabon (ANBG) and the AFRODYN project funded by the Agence Nationale de la Recherche (grant number ANR-15-CE02-0002-01) and the Fonds de la Recherche Scientifique (F.R.S.-FNRS) through project T.0163.13, and by the Belgian Science Policy (Belspo) through project AFRIFORD for funding field work, lab work, and herbarium visits. We would like to express our gratitude to IRD-Cameroun, to the Institut de Pharmacopée et de Médecine Traditionnelle (IPHAMETRA), to the Institut de Recherche en Ecologie Tropicale (IRET) of the Centre National de la Recherche Scientifique et Technologique (CENAREST) and to the Agence Nationale des Parcs Nationaux du Gabon (ANPN) for the assistance and logistics provided during the various sampling missions. We would also like to thank the curators and staff members of all herbariums (BR, BRLU, COI, K, LBV, MB, MO, P, WAG, YA) for their help during the various visits and for the availability of herbarium specimens.

## ■ LITERATURE CITED

**Aubréville, A.** 1962. *Flore du Gabon*, vol. 3, *Irvingiaceae, Simaroubaceae, Burseraceae*. Paris: Muséum national d'Histoire naturelle.

**Belkhir, K., Borsa, P., Chikhi, L., Raufaste, N. & Bonhomme, F.** 2004. GENETIX, version 4.05. Windows software for population genetics. Laboratoire génome, populations, interactions, CNRS UMR 5000. Université de Montpellier II, Montpellier (France). https://kimura.univ-montp2.fr/genetix/

**Bickford, D., Lohman, D.J., Sodhi, N.S., Ng, P.K., Meier, R., Winker, K., Ingram, K.K. & Das, I.** 2007. Cryptic species as a window on diversity and conservation. *Trends. Ecol. Evol.* 22: 148–155. https://doi.org/10.1016/j.tree.2006.11.004

**Bouka, D.U.G.** 2017. *Structuration de la biodiversité des forêts africaines et changements climatiques: Une étude à travers le genre Khaya (Meliaceae)*. Ph.D. thesis, University of Montpellier, France.

**Couvreur, T.L.P., Helmstetter, A.J., Koenen, E.J.M., Bethune, K., Brandão, R.D., Little, S.A., Sauquet, H. & Erkens, R.H.J.** 2019. Phylogenomics of the Major Tropical Plant Family Annonaceae Using Targeted Enrichment of Nuclear Genes. *Frontiers Pl. Sci.* 9. https://doi.org/10.3389/fpls.2018.01941

**Daïnou, K., Mahy G., Duminil, J., Dick, C., Doucet, J.-L., Donkpégan, A., Pluijgers, M., Sinsin, B., Lejeune, P. & Hardy, O.J.** 2014. Speciation slowing down in widespread and long-living tree taxa: Insights from the tropical timber tree genus *Milicia* (Moraceae). *Heredity* 113: 74–85. https://doi.org/10.1038/hdy.2014.5

**Daïnou, K., Blanc-Jolivet, C.C., Degen, B., Kimani, P.P., Ndiade-Bourobou, D., Donkpegan, A.S.L., Tosso, F., Kaymak, E., Bourland, N., Doucet, J.-L. & Hardy, O.J.** 2016. Revealing hidden species diversity in closely related species using nuclear SNPs, SSRs and DNA sequences – A case study in the tree genus *Milicia*. *B. M. C. Evol. Biol.* 16: 259. https://doi.org/10.1186/s12862-016-0831-9

**Dauby, G.** 2012. *Structure spatiale de la diversité intra- et interspécifique en Afrique centrale: Le cas des forêts gabonaises*. Ph. D. thesis, Université Libre de Bruxelles, Belgium.

**Dauby, G., Duminil, J., Heuertz, M. & Hardy, O.J.** 2010. Chloroplast DNA polymorphism and phylogeography of a central African tree species widespread in mature rainforests: *Greenwayodendron suaveolens* (Annonaceae). *Trop. Pl. Biol.* 3: 4–13. https://doi.org/10.1007/s12042-010-9041-6

**Doyle, J.J. & Doyle, J.L.** 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull. Bot. Soc. Amer.* 19: 11–15.

**Duminil, J. & Di Michele, M.** 2009. Plant species delimitation: A comparison of morphological and molecular markers. *Pl. Biosyst.* 143: 528–542. https://doi.org/10.1080/11263500902722964

**Duminil, J., Kenfack, D., Viscosi, V., Grumiau, L. & Hardy, O.J.** 2012. Testing species delimitation in sympatric species complexes: The case of an African tropical tree, *Carapa* spp. (Meliaceae). *Molec. Phylogen. Evol.* 62: 275–285. https://doi.org/10.1016/j.ympev.2011.09.020

**Duminil, J., Mona, S., Mardulyn, P., Doumenge, C., Walmacq, F., Doucet, J.-L. & Hardy, O.J.** 2015. Late Pleistocene molecular dating of past population fragmentation and demographic changes in African rain forest tree species supports the forest refuge hypothesis. *J. Biogeogr.* 42: 1443–1454. https://doi.org/10.1111/jbi.12510

**Earl, D.A & Von Holdt, B.M.** 2012. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genet. Resources* 4: 359–361. https://doi.org/10.1007/s12686-011-9548-7; http://taylor0.biology.ucla.edu/structureHarvester/

**Ellis, J., Knight, M.E., Carvell C. & Goulson D.** 2006. Cryptic species identification: A simple diagnostic tool for discriminating between two problematic bumblebee species. *Molec. Ecol. Notes* 6: 540–542. https://doi.org/10.1111/j.1471-8286.2006.01231.x

**Evanno, G., Regnaut, S. & Goudet, J.** 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molec. Ecol.* 14: 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

**Hammer, Ø., Harper, D.A.T. & Ryan P.D.** 2001. PAST: Paleontological statistics software package for education and data analysis. *Palaeontol. Electronica* 4: 9 pp. https://palaeo-electronica.org/2001_1/past/issue1_01.htm; http://folk.uio.no/ohammer/past

**Hardy, O.J. & Vekemans, X.** 2002. SPAGeDi: A versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molec. Ecol. Notes* 2: 618–620. https://doi.org/10.1046/j.1471-8286.2002.00305.x

**Hardy, O.J., Born, C., Budde, K., Daïnou, K., Dauby, G., Duminil, J., Ewédjé, E.-E.B., Gomez C., Heuertz M. & Koffi G.K.** 2013. Comparative phylogeography of African rain forest trees: A review of genetic signatures of vegetation history in the Guineo-Congolian region. *Compt. Rend. Geosci.* 345: 284–296. https://doi.org/10.1016/j.crte.2013.05.001

**Hawthorne, W.D. & Jongkind, C.C.** 2006. *Woody plants of Western African forests: A guide to the forest trees, shrubs and lianes from Senegal to Ghana*. Richmond: Royal Botanic Gardens, Kew.

**Henderson, A.** 2005. The methods of herbarium taxonomy. *Syst. Bot.* 30: 456–459. https://doi.org/10.1600/0363644054223701

**Ikabanga, D.U., Stévart, T., Koffi, K.G., Monthé, F.K., Doubindou, E.C.N., Dauby, G., Souza, A., M'batchi, B. & Hardy, O.J.** 2017. Combining morphology and population genetic analysis uncover species delimitation in the widespread African tree genus *Santiria* (Burseraceae). *Phytotaxa* 321: 166–180. https://doi.org/10.11646/phytotaxa.321.2.2

**Kenfack, D.** 2011. Resurrection in *Carapa* (Meliaceae): A reassessment of morphological variation and species boundaries using multivariate methods in a phylogenetic context. *Bot. J. Linn. Soc.* 165: 186–221. https://doi.org/10.1111/j.1095-8339.2010.01104.x

**Le Thomas, A.** 1969. *Flore du Gabon*, vol. 16, *Annonaceae*. Paris: Muséum national d'Histoire naturelle.

**Lewin, R.** 1981. Three species concepts. *Taxon* 30: 609–613. https://doi.org/10.2307/1219942

**Lissambou, B.-J., Hardy, O.J., Atteke, C., Stevart, T., Dauby, G., M'batchi, B., Sonke, B. & Couvreur, T.L.P.** 2018. Taxonomic revision of the African genus *Greenwayodendron* (Annonaceae). *PhytoKeys* 114: 55–93. https://doi.org/10.3897/phytokeys.114.27395

**Migliore, J., Kaymak, E., Mariac, E., Couvreur, T.L.P., Lissambou, B.-J., Piñeiro, R. & Hardy, O.J.** 2019. Pre-Pleistocene origin of phylogeographical breaks in African rain forest trees: New insights from *Greenwayodendron* (Annonaceae) phylogenomics. *J. Biogeogr.* 46: 212–223. http://doi.org/10.1111/jbi.13476

**Piñeiro, R., Micheneau, C., Dauby, G. & Hardy, O.J.** 2016. Isolation, characterisation and cross-species amplification of nuclear microsatellites in the African tree genus *Greenwayodendron* (Annonaceae). *J. Trop. Forest Sci.* 28: 121–131. http://www.jstor.org/stable/43799215

**Piñeiro, R., Dauby, G., Kaymak, E. & Hardy, O.J.** 2017. Pleistocene population expansions of shade-tolerant trees indicate fragmentation of the African rainforest during the Ice Ages. *Proc. Roy. Soc. London, Ser. B, Biol. Sci.* 284: 20171800. https://doi.org/10.1098/rspb.2017.1800

**Porras-Hurtado, L., Ruiz, Y., Santos, C., Phillips, C., Carracedo, Á. & Lareu, M.V.** 2013. An overview of STRUCTURE: Applications, parameter settings, and supporting software. *Frontiers Genet.* 4: 98. https://doi.org/10.3389/fgene.2013.00098

**Pratt, D.B. & Clark, L.G.** 2001. *Amaranthus rudis* and *A. tuberculatus*, one species or two? *J. Torrey Bot. Soc.* 282–296. https://doi.org/10.2307/3088718

**Pritchard, J.K., Stephens, M. & Donnelly, P.** 2000. Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.

**Quantum GIS Development Team** 2014. QGIS Geographic Information System. Open Source Geospatial Foundation Project. http://qgis.org/fr/site

**Smith, J.F., Mansfield, D.H., Stevens, M., Sosa, E., Feist, M.A., Downie, S.R., Plunkett, G.M. & Darrach, M.** 2018. Try Tri again? Resolving species boundaries in the *Lomatium triternatum* (Apiaceae) complex. *J. Syst. Evol.* 56: 218–230. https://doi.org/10.1111/jse.12418

**Turini, F.G., Steinert, C., Heubl, G., Bringmann, G., Lombe, B.K., Mudogo, V. & Meimberg, H.** 2014. Microsatellites facilitate species delimitation in Congolese *Ancistrocladus* (Ancistrocladaceae), a genus with pharmacologically potent naphthylisoquinoline alkaloids. *Taxon* 63: 329–341. https://doi.org/10.12705/632.36

**Verdcourt, B.** 1969. The status of the genus *Polyalthia* Blume (Annonaceae) in Africa. *Adansonia*, ser. 2, 9(1): 87–94.

**Wang, J.** 2017. The computer program STRUCTURE for assigning individuals to populations: Easy to use but easier to misuse. *Molec. Ecol. Notes* 17: 981–990. https://doi.org/10.1111/1755-0998.12650

**Whittall, J.B., Hellquist, C.B., Schneider, E.L. & Hodges, S.A.** 2004. Cryptic species in an endangered pondweed community (*Potamogeton*, Potamogetonaceae) revealed by AFLP markers. *Amer. J. Bot.* 91: 2022–2029. https://doi.org/10.3732/ajb.91.12.2022