



HAL
open science

Khaya revisited: genetic markers and morphological analysis reveal six species in the widespread taxon *K. anthotheca*

Gaël U.D. Bouka, Charles Doumenge, Marius R.M. Ekué, Kasso Daïnou, Jacques Florence, Bernd Degen, Jean Joël Loumeto, Doyle Mckey, Olivier J. Hardy

► **To cite this version:**

Gaël U.D. Bouka, Charles Doumenge, Marius R.M. Ekué, Kasso Daïnou, Jacques Florence, et al.. Khaya revisited: genetic markers and morphological analysis reveal six species in the widespread taxon *K. anthotheca*. *Taxon*, the journal of the International Association for Plant Taxonomy, 2022, 71 (4), pp.814-832. <10.1002/tax.12720>. <hal-04957682>

HAL Id: hal-04957682

<https://hal.science/hal-04957682v1>

Submitted on 19 Feb 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-NC 4.0 - Attribution - Non-commercial use - International License

RESEARCH ARTICLE

Khaya revisited: Genetic markers and morphological analysis reveal six species in the widespread taxon *K. anthotheca*

Gaël U.D. Bouka,^{1,2,3} Charles Doumenge,² Marius R.M. Ekué,⁴ Kasso Daïnou,^{5,6} Jacques Florence,⁷ Bernd Degen,³ Jean Joël Loumeto,¹ Doyle McKey⁸ & Olivier J. Hardy⁹

1 Université Marien NGOUABI, Faculté des Sciences et Techniques, Laboratoire de Biodiversité de Gestion des écosystèmes et de l'Environnement, BP 69, Brazzaville, Congo

2 CIRAD, Forêts et Sociétés, Université de Montpellier, Campus International de Baillarguet TA C-105/D, 34398 Montpellier cedex 5, France

3 Thünen Institute of Forest Genetics, Sieker Landstraße 2, 22927 Großhansdorf, Germany

4 Bioversity International, s/c IITA, BP 2008, Yaoundé, Cameroon

5 Laboratory of Applied Ecology, Faculty of Agronomic Science, University of Abomey-Calavi, Cotonou, Benin

6 Laboratory of Tropical and Subtropical Forestry, Unit of Forest and Nature Management, Gembloux Agro-Bio Tech, University of Liege, Gembloux 5030, Belgium

7 Muséum national d'histoire naturelle, Herbarium national, Case postale 39, 57 rue Cuvier, 75231 Paris cedex 05, France

8 CEFÉ, UMR 5175, CNRS, University of Montpellier, University Paul Valéry, EPHE, SupAgro Montpellier, INRA, IRD Montpellier, France

9 Evolutionary Biology and Ecology Unit, Faculté des Sciences, Université Libre de Bruxelles, Avenue F. D. Roosevelt 50, CP 160/12, Brussels 1050, Belgium

Address for correspondence: Gaël U.D. Bouka, ulrichbouka@yahoo.fr

DOI <https://doi.org/10.1002/tax.12720>

Abstract *Khaya* (Meliaceae) is a widespread genus of trees distributed from the tropical forests of Africa to Madagascar and the Comoros. *Khaya* species are very close morphologically and some contain considerable phenotypic and ecological diversity, raising the question of their delimitation: this is the case of *Khaya anthotheca* s.l. (sensu lato, including *K. nyasica*), which is often lumped with *K. grandifoliola* and *K. ivorensis*. We tested species delimitation within *K. anthotheca* by combining population-genetic and morphological data. First, a visual inspection of over a thousand specimens suggested the existence of five morphogroups. Second, after genotyping 498 samples at 71 nuclear SNPs, Bayesian clustering analyses (STRUCTURE) first identified five genetic clusters (K1–K5), while further analyses subdivided K5 into two clusters. The six genetic clusters obtained are distributed in parapatry or allopatry from West Africa to East Africa, except for cluster K4, which is in sympatry with K3 in Central Africa. The partitioning of genetic clusters is also largely congruent with the morphogroups identified. The morphological differentiation of these clusters was confirmed by statistical analyses of 27 leaf characters. The clear-cut genetic differentiation between clusters, with few admixed genotypes even in contact zones, argues for the recognition of distinct species following the biological species concept. Our analyses highlight that the current taxonomic treatment of *Khaya* underestimates by one-half the total number of species. We identify two new species, circumscribe *K. anthotheca* and support the rehabilitation at the rank of species of three taxa considered to be synonyms of *K. anthotheca* (*K. agboensis*, *K. euryphylla*, *K. nyasica*). These results illustrate the strength of combining population-genetic and morphological data to delineate an African rainforest tree species complex.

Keywords African mahogany; cryptic species; genetic clustering; Meliaceae; population-genetics

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

■ INTRODUCTION

Integrating population-genetic methods and morphological data has proven useful to delineate African rainforest tree and vine species complexes such as in *Carapa* Aubl. (Kenfack, 2011; Duminil & al., 2012), *Ancistrocladus* Wall. (Turini & al., 2014), *Milicia* Sim (Daïnou & al., 2016), *Santiria* Blume (Ikabanga & al., 2017) and *Greenwayodendron* Verdc. (Lissambou & al., 2019). This combination of methods has proven particularly successful in detecting hidden species (Bickford & al., 2007).

The delimitation of cryptic species (“two or more distinct species that are erroneously classified [and hidden] under one

species name”: Bickford & al., 2007: 148) has implications for the interpretation of phylogenies and evolutionary processes but also for management and biodiversity conservation (Liu & al., 2015). Species delimitation has to be confronted with different species concepts, in particular morphological and biological ones (De Queiroz, 2007; Birky & al., 2010).

Morphological analysis remains essential in species delimitation, but in some cases, the grouping of individuals sharing common morphological traits into species is not obvious because speciation is not always accompanied by clear morphological differentiation (Bickford & al., 2007). On the other hand, the biological species concept, which emphasizes the

Article history: Received: 31 Dec 2020 | returned for (first) revision: 22 Feb 2021 | (last) revision received: 25 Nov 2021 | accepted: 29 Nov 2021

Associate Editor: Michael D. Pirie | © 2022 International Association for Plant Taxonomy.

existence of reproductive barriers that can be tested using genetic markers, is often called into question because many plant species hybridize and produce fertile offspring (Stace & al., 2015; Dainou & al., 2016). Thus, species delimitation should be based on several pieces of evidence, among them morphological, geographic and molecular characters (Dayrat, 2005).

Nowadays, following the rapid development of new DNA-based technologies, phylogenetic and population-genetic approaches are increasingly used to delimit species (e.g., Whittall & al., 2004; Bickford & al., 2007; Duminil & Di Michele, 2009; Monthe & al., 2018). According to the phylogenetic approach, species can be identified if they separate into different lineages (monophyletic species concept). However, the identification of these lineages can be problematic because of biological processes such as hybridization, introgression or incomplete lineage sorting (Whitney & al., 2010; Liu & al., 2015; Naciri & Linder, 2015), which generate paraphyletic or polyphyletic groups (Reeves & Richards, 2007; Naciri & Linder, 2015). In addition, local adaptation and genetic drift can lead to divergent populations within the same species, which makes it difficult to distinguish intraspecific from interspecific variation (Harrison, 2012). As these forces operate at the population level, a population-genetic approach, which examines the genetic structure of populations, can assist in species delimitation, particularly in closely related species where the process of speciation may not be complete.

Population-genetic methods can detect barriers to gene flow by identifying well-differentiated genetic clusters (e.g., algorithm of Pritchard & al., 2000, implemented in the STRUCTURE software). Hence, when well-differentiated genetic clusters occur in sympatry without showing sign of introgression, these are likely to belong to distinct species according to the biological species concept because they are reproductively isolated. Population-genetic 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; Dainou & al., 2016). Although 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. Thus, it is important to associate morphological data with spatial distribution of the detected genetic groups before concluding whether they correspond to distinct species.

The genus *Khaya* A.Juss. contains large trees naturally distributed in tropical Africa, Madagascar and the Comoros. *Khaya* species provide one of the most-prized woods on the African continent, sold under the name of African mahogany (Bouka & al., 2019). Currently, five species are accepted (Aubréville, 1959; CTFT, 1988; African Plant Database, 2020). Four of them are present on the Afrotropical realm: *Khaya anthotheca* (Welw.) C.DC., *K. grandifoliola* C.DC., *K. ivorensis* A.Chev. and *K. senegalensis* (Desr.) A.Juss. The fifth species, *K. madagascariensis* Jum. & H.Perrier, is endemic to Madagascar and the Comoros (Keay & Hepper, 1963; CTFT, 1988). Due

to considerable morphological variability within each species and morphological proximity between them, several taxa were described and published, then put into synonymy of one or the other of these five species (Aubréville, 1959; Voorhoeve, 1965; Styles & White, 1991; Bouka & al., 2017). Although *K. grandifoliola*, *K. ivorensis*, *K. senegalensis* and to a certain extent *K. madagascariensis* clearly represent good species, the taxonomic status of *K. anthotheca* remains contested (Danquah & al., 2011; Bouka, 2017; Bouka & al., 2019; Pakull & al., 2019). Indeed, *K. anthotheca* s.l. (sensu lato, including all current synonyms such as *K. nyasica* s.l.) is the taxon that has the highest number of synonyms and an unclear delimitation within *Khaya* (Bouka & al., 2019). For instance, *K. nyasica* Stapf ex Baker f., a taxon distributed in East and South-East Africa, is considered synonymous with *K. anthotheca* s.l. (Styles & White, 1991; African Plant Database, 2020) or a different species (Staner & Gilbert, 1958). In addition, *K. anthotheca* s.l. is often lumped with *K. grandifoliola* and *K. ivorensis* (Aubréville, 1959; Hawthorne & Gyakari, 2006; De Wilde, 2015). Currently listed as vulnerable by the IUCN (Hawthorne, 1998), *K. anthotheca* s.l. has a large ecological amplitude and geographic range, occurring from Guineo-Congolian semi-deciduous forests in West and Central Africa to dry forests in East Africa.

Despite the uncertain taxonomic delimitation of *Khaya anthotheca* s.l., no taxonomic studies have been performed to resolve this problem. However, a phytochemical study carried out in the genus *Khaya* identified, in addition to *K. nyasica*, two chemical groups within *K. anthotheca*, which also differ by the leaves and fruits morphology (Adesida & al., 1971). A more recent study based on molecular markers carried out by Pakull & al. (2019) on the whole genus *Khaya* could not clearly distinguish *K. anthotheca* from *K. nyasica* and *K. madagascariensis* but 4 of the 15 single nucleotide polymorphism nuclear markers (nSNPs; “golden markers”) in this study separated *K. anthotheca* from *K. nyasica*. Finally, a recent phylogenetic study within the genera *Entandrophragma* C.DC. and *Khaya* has shown that for *Khaya*, plastome and ribosomal DNA phylogenies have very divergent topologies, possibly due to a recent diversification involving incomplete lineage sorting and/or recurrent hybridization events (Monthe & al., 2019).

In the present study, we used an integrative taxonomic approach to clarify the limits of the taxon *K. anthotheca* s.l. using morphological and genetic data, taking advantage of the recently developed nSNP markers (Pakull & al., 2016). More specifically, we address the following questions: (i) Do genetic data identify multiple genetic groups? (ii) Are there several morphological groups? (iii) Is there a correlation between genetic and morphological data, allowing delimitation of species?

■ MATERIALS AND METHODS

Source of genetic data and sampling. — A random sampling of *Khaya anthotheca* s.l. individuals was carried out in 15 African countries (Angola, Cameroon, Central African

Republic, Democratic Republic of the Congo, Ghana, Ivory Coast, Liberia, Malawi, Mozambique, Republic of the Congo, Sierra Leone, Tanzania, Uganda, Zambia, Zimbabwe) with a minimum distance of at least 50 m between two trees in the same site. For each tree, herbarium vouchers were collected with a unique collection number. That number was also applied to all materials from the same tree used for genetic analyses. The material used for genetic analyses consisted of pieces of leaflets and/or cambium dehydrated using silica gel. We collected 442 samples from live trees, 9 of which are attached to herbarium specimens. In addition to these samples, 73 other samples were taken from herbarium vouchers available at Paris (P) and Meise (BR) herbaria (see Appendix 1). A total of 498 georeferenced samples were collected across the 15 countries. The majority of the samples taken from the herbaria originate from the Democratic Republic of the Congo (D.R.C.) and from East African countries. All of these samples are stored at Thünen Institute of Forest Genetics (Hamburg-Großhansdorf, Germany).

Source of morphological data and preliminary herbarium studies. — A total of around 1300 herbarium samples attributed to *Khaya anthotheca* s.l. were examined. Of these samples, 11 originated from our own field collections between 2013 and 2015 from three countries (Cameroon, Republic of the Congo, D.R.C.). The remaining specimens were obtained through loans from the following herbaria: BM, BR, FHO, G, HBG, K, L, P, PRE and WAG (now in L). Since the herbaria loans did not cover the entire distribution range of the species, we visited and consulted BM, BR, BRLU, FHO, HBG, K, P (including P-JU and P-LA) and YA.

Up to now, identification of specimens belonging to *Khaya* has essentially been based on the number and the shape

of leaflets because they remain roughly constant in adulthood within each species. On the other hand, the flowers do not vary among species, while the fruits vary in size among conspecific individuals (Staner & Gilbert, 1958; Aubréville, 1959). A preliminary morphological study of herbarium specimens of *K. anthotheca* s.l. was therefore carried out based on distinctive characteristics of leaflets (size, number, shape, nature of the apex and base of the lamina, aspect of the veinlets) mentioned in the descriptions of this taxon. This first work allowed us to evaluate the variation of these characters and to constitute five groups of specimens on the basis of the morphological resemblance of all the samples consulted (Kenfack, 2011). These morphogroups were working hypotheses for the remainder of the study and were named G1 to G5 (see Fig. 1).

DNA extraction, genotyping and sequencing. — DNA isolation from 498 individuals was carried out according to Dumolin & al. (1995). A set of 67 nSNPs developed by Pakull & al. (2016) via the restriction site associated DNA sequencing method (RADseq) (Miller & al., 2007) and four other nuclear markers developed by the same authors (unpub. work), i.e., 71 nSNPs, were used for genotyping these 498 individuals through a single-base extension and mass spectrometry-based MassARRAY iPLEX method (Agena Biosciences, Hamburg, Germany). Of 498 sequenced individuals, 481 that were successfully genotyped for at least half of the 71 nSNPs were used for further analyses.

Identification of genetic groups. — We identified genetic groups through a Bayesian grouping algorithm using the Markov Chain Monte Carlo method (MCMC) in two steps. The first analysis was performed on all 481 individuals (see data used for genetic analyses in suppl. Appendix S1). A second analysis was carried out on each genetic group

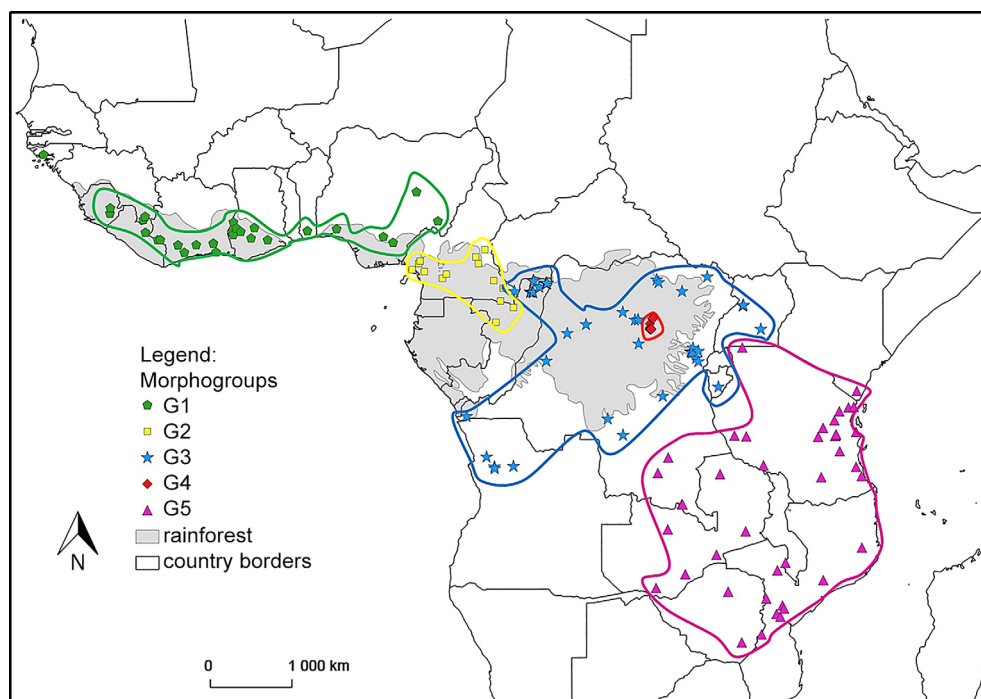


Fig. 1. Distribution of a priori morphogroups G1–G5 in the complex *Khaya anthotheca* s.l.

obtained at the end of the first analysis in order to identify genetic groups represented by few samples. Indeed, genetic groups represented by small sample sizes may remain undetected when analysed together with well-represented groups (Porrás-Hurtado & al., 2013; Wang, 2017).

For both steps, we identified genetic clusters using STRUCTURE v.2.3.4 (Pritchard & al., 2000; Falush & al., 2003) without a priori grouping (i.e., morphogroups were not taken into account). We ran STRUCTURE to delimit the most likely number K of genetic clusters varying from 1 to 10 with 10 replicates per K value. Each run was performed during 1,000,000 MCMC iterations with a burn-in of 100,000 (10%) iterations. We used the admixture model with the independent allele frequencies model (with all options set at their default values). To evaluate the number of clusters that best explains the dataset. We considered how the log-likelihood of the data ($\ln(P)$) 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).

The mean proportion of the individual genomes (q) was finally determined via the CLUMPP v.1.1.1 program (Jakobsson & Rosenberg, 2007), and individuals were assigned to a given genetic cluster K_i ($i = 1, 2, 3, \dots$) when $q > 0.5$. Finally, the results were graphically illustrated with the DISTRUCT v.1.1 program (Rosenberg, 2004).

Diversity and differentiation parameters between genetic groups. — We estimated the genetic diversity of each genetic group by calculating the effective number of alleles (N_A), as well as observed and expected heterozygosities (H_O , H_E), and inbreeding coefficient (F_I). The differentiation between genetic groups was characterized by computing the fixation index or genetic differentiation index (F_{ST}). All these calculations were done with SPAGeDi v.1.5 (Hardy & Veke­mans, 2002).

Morphometric data matrix construction. — In order to quantify the morphological differences that support the morphological groups defined a priori and to carry out a more detailed characterization, the morphological characters (leaf characters) were defined on the basis of a bibliography of the genus *Khaya* (Danquah & al., 2011; Bouka & al., 2019). We selected a total of 27 leaf characters, including 18 quantitative and 9 qualitative traits (Table 1). The 9 qualitative characters were divided into 44 different states (Table 2), following the terminology of Stearn (1992). In addition to these leaf characters, we also measured and took information on the nature of the reproductive organs: ovary 4- or 5-locular, type of inflorescence, size of corolla, number of ovules per loculus and diameter of the fruit (reproductive organ characters).

Out of the 1300 samples examined, we removed all the uncertain and deteriorated samples, for example those whose leaflets were not present or were cut. We also discarded young individuals to avoid introducing ontogenetic variation because the leaves of *Khaya anthotheca* s.l. are very broad in young individuals (Aubréville, 1959). Mature individuals were identified on the herbarium samples by tree measurements and/or

when the samples had flowers or were labelled “fruit in carpothèque”. The 71 leaf variables (18 quantitative characters and 44 qualitative character states) selected were thus measured or observed on 178 samples or individuals, representing all the morphogroups: G1 ($N = 25$), G2 ($N = 52$), G3 ($N = 34$), G4 ($N = 20$) and G5 ($N = 47$). We measured the pair of leaflets preceding the terminal pair and the pair of leaflets preceding the two pairs of terminal leaflets as the median pairs respectively, for leaves having 4 and 6 pairs of leaflets. Leaves with two pairs of leaflets present in a single morphogroup were not measured but used as a qualitative character for that morphogroup. Because of the rarity of fertile herbarium specimens, information on reproductive organs was collected and

Table 1. List of leaf characters to assess morphological variation.

Code	Quantitative character
NLP	Number of leaflet pairs
PL	Length of the petiole
RL	Length of the rachis
PLBL	Length of the basal-leaflet petiolule
LLBL	Length of the basal-leaflet limb
LWBL	Width of the basal-leaflet limb
NLVBL	Number of lateral veins of the basal leaflet
ALBL	Apex length of basal leaflet
PLML	Length of the middle-leaflet petiolule
LLML	Length of the middle-leaflet limb
LWML	Width of middle-leaflet limb
NLVML	Number of lateral veins of the middle leaflet
ALML	Apex length of middle leaflet
PLTL	Length of the terminal-leaflet petiolule
LLTL	Length of the terminal-leaflet limb
LWTL	Width of the terminal-leaflet limb
NLVTL	Number of lateral veins of the terminal leaflet
ALTL	Apex length of terminal leaflet
Code	Qualitative character
FBL	Form of the basal leaflet
FML	Form of the middle leaflet
FTL	Form of the terminal leaflet
BBL	Base of the basal leaflet
BML	Base of the middle leaflet
BTL	Base of the terminal leaflet
TNVLS	Tertiary network of the veins on the lower surface
TNVUS	Tertiary network of the veins on the upper surface
SL	Apex leaflet

The terminology follows Stearn (1992). The codes represent characters used in the statistical analyses.

measured only from 23 individuals from all the morphogroups G1 ($N=3$), G2 ($N=3$), G3 ($N=5$), G4 ($N=4$) and G5 ($N=8$).

Morphological data analyses. — To confirm the morphological differentiation between morphogroups constituted a priori, we performed statistical tests through the software R v.3.5.2 (R Core Team, 2020) using (i) hierarchical cluster analysis (HCA) on all variables (Kaufman & Rousseeuw, 1990) and (ii) Kruskal-Wallis tests on quantitative variables as well as (iii) the Hill-Smith ordination (Hill & Smith, 1976), an extended principal component analysis (PCA) for datasets containing both qualitative and quantitative variables (see data used for statistical analyses in suppl. Appendix S2).

Regarding the HCA, the daisy function (Struyf & al., 1997) of the R-package cluster was used to compute a distance matrix between samples using the Gower index (Gower, 1971), while the hierarchical classification was performed with the hclust function of the stat package using the Ward.D2 aggregation criterion of Ward's method (Ward, 1963). A confusion matrix was built after HCA to check if individuals included in each a priori morphogroup were classified accordingly. When individuals of a morphogroup were divided into several different classes of HCA, we re-examined the herbarium samples of this morphogroup to assess if characters that escaped our first assessment could explain this difference and distinguish other morphological groups.

We displayed the morphological variation of the different morphogroups constituted a priori and those obtained following the HCA using box-and-whiskers plots for quantitative characters (Tukey, 1977) (see suppl. Figs. S1–S5) and stacked bar diagrams for qualitative characters (see suppl. Figs. S6–S8). Then, we checked whether quantitative characters significantly differed between those groups using the Kruskal-Wallis test, and using Wilcoxon-Mann-Whitney tests to compare each pair of groups. We attributed letters to identify groups differing significantly or not at each character.

The Hill-Smith ordination was realized via the dudi.hill-smith function of the ade4 package (Dray & Dufour, 2007) to identify the main variables that discriminate morphogroups and also to independently verify whether the a priori morphogroups are well supported.

The number of individuals per morphogroup being very low for the reproductive organ characters, we used for all the above analyses only the leaf characters. The reproductive organ characters were taken into account only in the species characterization.

Comparison of morphogroups and genetic clusters. —

Given that each morphogroup was represented by at least eight genotyped herbarium samples, we examined (i) the congruence between the classifications of individuals into morphogroups and genetic clusters using a confusion matrix, and

Table 2. List of the states of nine qualitative leaf characters (see Table 1 for character codes).

Characters			
FBL; FML; FTL	BBL; BML; BTL	TNVLS; TNVUS	SL
Elliptic (E)	Rounded (AR)	Marked (MQ)	Acuminate (ACM)
Elliptic-oblong (EO)	Rounded-symmetric (AS)	Less-marked (MMQ)	Apiculate (APL)
Narrowly-elliptic (EE)	Rounded-asymmetric (AA)	Very marked (TM)	Caudate (CAD)
Narrowly-elliptic-oblong (EEO)	Rounded-asymmetric-truncate (AAT)	–	Emarginate (EMG)
Narrowly-elliptic-oval (EEOV)	Rounded-asymmetric-truncate-obtuse (AATO)	–	Retuse (RTU)
Narrowly-oboval (EOBV)	Rounded-obtuse-truncate (AOT)	–	Truncate (TRQ)
Narrowly-oblong (EOBL)	Cuneate (CU)	–	–
Narrowly-oval (EO)	Cuneiform-rounded (CUA)	–	–
Narrowly-ovato-oblong (EEO)	Cuneiform-rounded-asymmetric (CAA)	–	–
Widely-elliptic (LE)	Cuneiform-rounded-truncate (CAT)	–	–
Widely-elliptic-oblong (LEO)	Cuneiform-asymmetric (CA)	–	–
Widely-oblong (LO)	Cuneiform-asymmetric-acute-obtuse (CAAO)	–	–
Widely-oboval (LOBV)	Cuneiform-asymmetric-rounded-truncate (CAAT)	–	–
Widely-oval (LOV)	Cuneiform-obtuse (CO)	–	–
Oblong (OB)	Obtuse-rounded-asymmetric (OAA)	–	–
Oboval (OBV)	Obtuse-rounded (OA)	–	–
Oval (OV)	–	–	–
Ovato-oblong (OO)	–	–	–
Ovato-elliptic (OE)	–	–	–

The letters in parentheses represent character state codes used in the statistical analyses.

(ii) the coherence of the geographical distributions of each morphogroup and the corresponding genetic cluster using all samples. However, only a few herbarium samples were jointly measured and genotyped due to sampling restrictions on herbarium vouchers established by some herbaria. Nevertheless, the congruence between the 83 samples of genotyped herbaria (of which 52 were genotyped and measured) and the morphogroups were checked.

Criteria to delimit species. — The primary goal of reassessing patterns of morphological variation in *Khaya anthotheca* s.l. was to look for morphological discontinuities that could highlight the presence of cryptic species and to validate them according to the biological species concept using population-genetic tools. Firstly, we considered as putative species the homogeneous morphogroups highlighted by the statistical analyses. Secondly, we checked whether the well-differentiated genetic groups corresponded to the morphogroups and whether there were no or very few admixed individuals. We considered such groups as distinct species if (i) diagnostic morphological traits occurred, (ii) the degree of genetic differentiation (F_{ST}) was high, and (iii) admixed individuals were absent or rare. Ultimately, our species delimitation respects both the morphological and biological species concepts.

Comparison between different species and nomenclatural types. — After delimiting species, all individuals of each species were compared to the nomenclatural types (Bouka & al., 2017) of currently accepted or *Khaya* names or synonyms and to the diagnoses of all species described in this genus. Each specimen was therefore formally identified, and the names of the existing *Khaya* taxa were assigned to each species. Specimens that did not correspond to any diagnosis and nomenclatural type were given a temporary name. A key for discriminating these species was then constructed from the diagnostic characters obtained from PCA,

a re-examination of the herbarium samples and reproductive characters.

RESULTS

Identification of genetic groups. — The first Bayesian clustering analysis using 71 nSNPs on 481 specimens showed that the likelihood of the data increased rapidly with the number of imposed clusters K until $K = 5$, while it increased very slowly for larger K (Fig. 2A). Application of the delta K method of Evanno & al. (2005) highlighted $K = 2$ and $K = 5$ as the most likely numbers of clusters (Fig. 2B).

The five genetic clusters obtained were spatially structured by longitude, so that numbers were reattributed from west (K1) to east (K5). Most clusters were distributed in parapatry or allopatry, except cluster K4, which has a narrow distribution range included within the range of K3 (Fig. 3A). More than 91% of individuals were assigned to a cluster at a $q > 80\%$. The few seemingly admixed individuals were usually intermediate between the clusters K2, K3 or K4 (Fig. 3B). Overall, 11 individuals could not be assigned to a cluster (i.e., $q < 0.5$ towards each cluster). Eight of these unassigned individuals were found in the north of the Republic of the Congo on the border between the K2 and K3 clusters. For the other three, after re-examination of the herbarium samples, one (BR00000 16995048) was found not to be a specimen of the genus *Khaya*, another (BR0000016995116) was found to have been planted in the Luki reserve in D.R.C. and the last was intermediate between K4 and K5.

Further Bayesian clustering analyses performed on each of the five genetic clusters did not highlight further genetic structuring in clusters K1 to K4 (when imposing $K = 2$, the majority of individuals appeared admixed, results not shown). However, cluster K5 revealed two genetic clusters (under

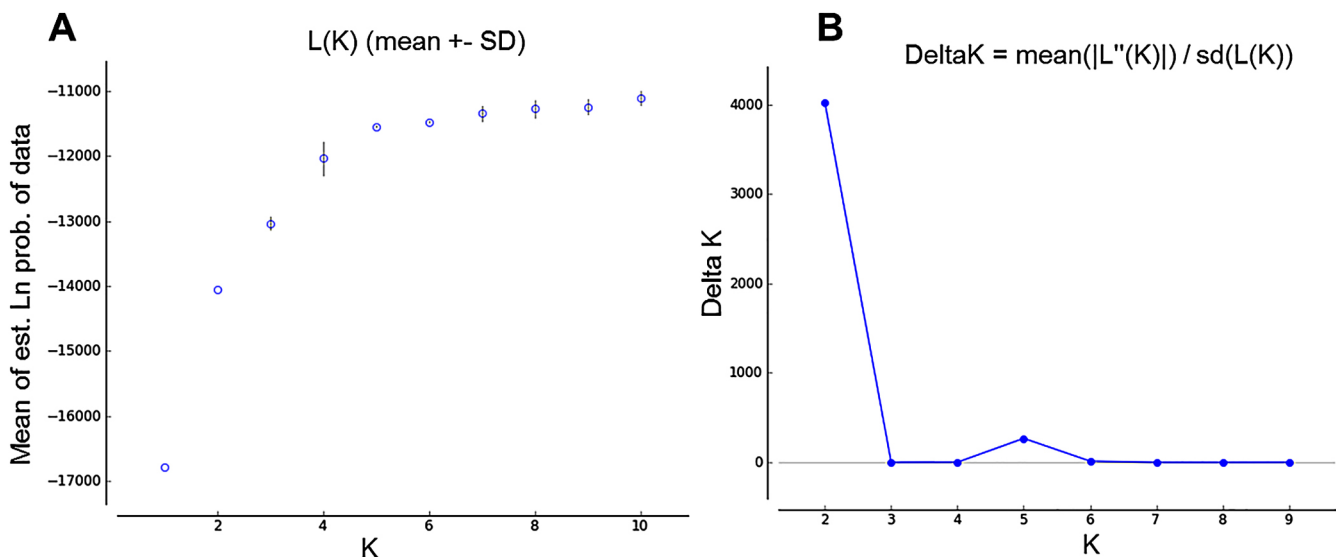


Fig. 2. Variation of the likelihood of the data $L(K)$ (A) and of DeltaK (B) as a function of the number of genetic groups (K) identified in 481 *Khaya anthotheca* s.l. samples using the software STRUCTURE.

$K = 2$) according to the increase in Ln P(D) method (see suppl. Fig. S9A) and the method of Evanno & al. (2005) (see suppl. Fig. S9B), with a complete absence of admixed individuals. These two clusters, named K5a and K5b (Fig. 3A,C), are mostly distributed in parapatry: K5a is more westerly, ranging from the southern D.R.C. to northern Zambia and Malawi, while K5b is more eastern, ranging from Tanzania to Malawi, Mozambique and Zimbabwe, with a few samples in Zambia (Fig. 3A). Hereafter, we will thus consider a total of six genetic groups (K1, K2, K3, K4, K5a, K5b).

Diversity and differentiation parameters between genetic groups. — The genetic diversity per cluster varied substantially (Table 3). Clusters K1 to K4 displayed by far the highest diversity with H_E ranging from 0.091 to 0.144, while K5a and K5b displayed very low genetic diversity values ($H_E = 0.038$ for K5a and 0.060 for K5b). Overall, the mean observed heterozygosities (H_O) were similar or slightly lower than expected heterozygosities (H_E), leading to inbreeding coefficients (F_I) ranging from -0.02 to 0.18.

The genetic differentiation index (F_{ST}) between pairs of genetic clusters ranged from 0.24 to 0.66 (Table 4),

the lowest values obtained between K4 and K3 ($F_{ST} = 0.24$) and between K4 and K2 ($F_{ST} = 0.26$), while the maximal values were obtained for all pairs involving K5a and/or K5b ($F_{ST} > 0.48$). All these values were significantly larger than zero.

Comparison of morphogroups. — We distinguished five morphogroups (G1–G5) based on our subjective visual

Table 3. Parameters of genetic diversity within genetic clusters.

Genetic clusters	N	N_A	H_E	H_O	F_I
K1	76	1.19	0.115	0.118	-0.022
K2	146	1.24	0.144	0.140	0.030
K3	157	1.15	0.091	0.077	0.154
K4	49	1.17	0.102	0.085	0.182
K5a	17	1.07	0.038	0.035	0.097
K5b	25	1.10	0.060	0.052	0.150

N = Sample size; N_A = effective number of alleles; H_E = expected heterozygosity; H_O = observed heterozygosity; F_I = inbreeding coefficient.

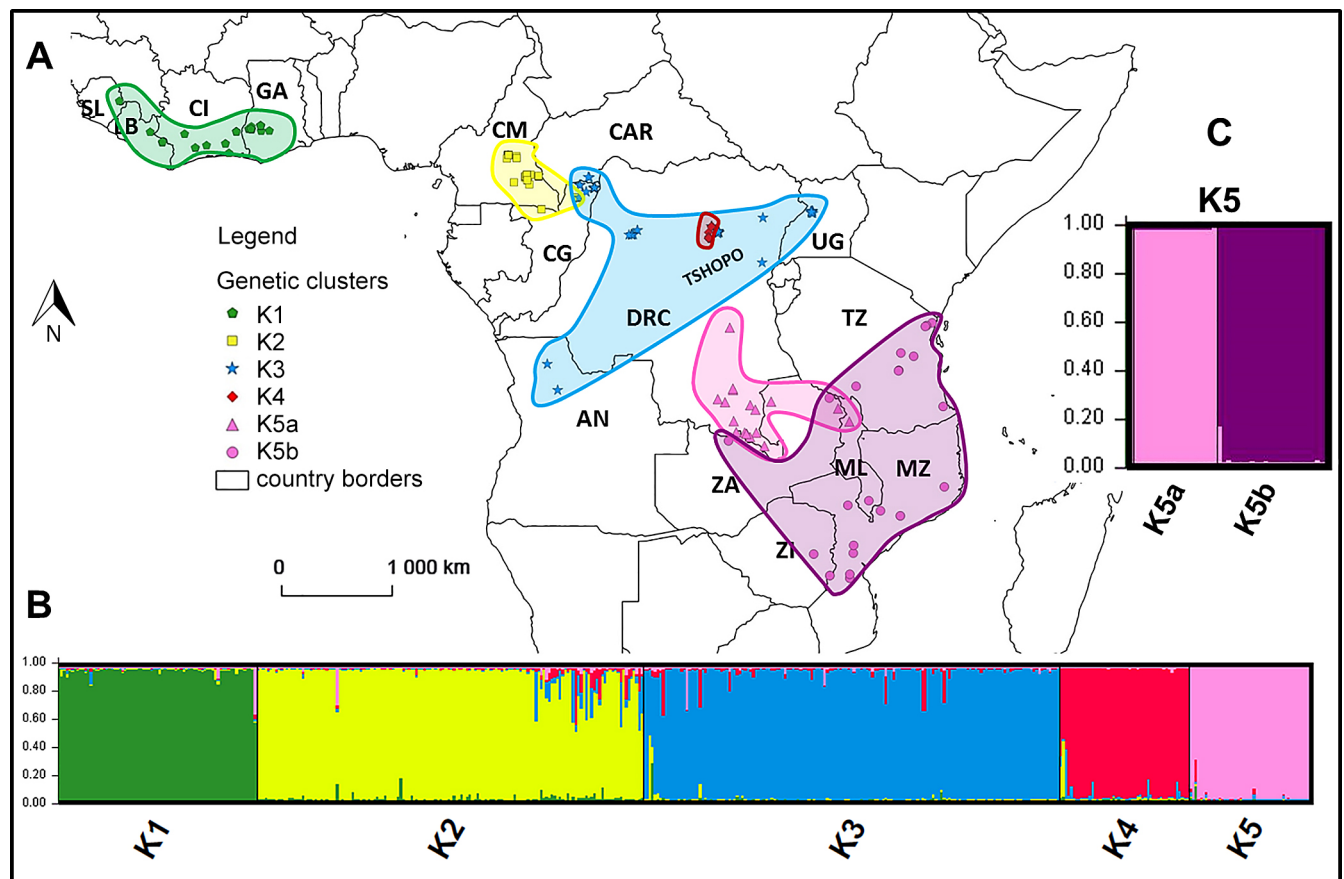


Fig. 3. Geographical distribution (A) of the six genetic clusters (K1–K5b) of *Khaya anthotheca* s.l. identified by STRUCTURE in two steps (identification of K1 to K5 followed by subdivision of K5 into K5a and K5b). Histogram of genetic assignment of (B) 481 *K. anthotheca* s.l. individuals at $K = 5$ and (C) of 42 K5 individuals at $K = 2$. Individuals were ordered along the horizontal axis according to their geographic affiliation. The vertical bars indicate how the genome of each individual is partitioned into the different clusters. Country code: AN, Angola; CAR, Central African Republic; CG, Congo; CI, Ivory Coast; CM, Cameroon; DRC, Democratic Republic of the Congo; GA, Ghana; LB, Liberia; ML, Malawi; MZ, Mozambique; SL, Sierra Leone; TZ, Tanzania; UG, Uganda; ZA, Zambia; ZI, Zimbabwe.

examination of all herbarium materials. Morphogroup numbers were again chosen to reflect the geographic position of most samples, from west (G1) to east (G5). Using an objective method based on 27 leaf characters (quantitative and qualitative) assessed on 178 individuals, the HCA also distinguished five classes (Fig. 4; suppl. Fig. S10). The confusion matrix shows a good one-to-one correspondence between most morphogroups and HCA classes (Table 5). However, while the majority of individuals in morphogroups G1 to G4 were mainly found in a single HCA class (with a few exceptions in G1 and G2), those categorized in morphogroup G5 were distributed between classes 3 (28 samples) and 5 (19 samples). A re-examination of the herbarium samples representing

incongruent individuals of morphogroups G1 and G2 compared to expected HCA classes revealed that these are extreme forms of these morphogroups. We therefore kept them in these initial morphogroups. However, the re-examination of the individuals constituting morphogroup G5 confirmed that this morphogroup can be further subdivided into two distinct morphogroups based on the following diagnostic characters:

Table 4. Index of genetic differentiation, F_{ST} , between pairs of genetic clusters.

Genetic clusters	K1	K2	K3	K4	K5a
K2	0.307				
K3	0.506	0.382			
K4	0.402	0.256	0.243		
K5a	0.658	0.481	0.617	0.561	
K5b	0.571	0.486	0.521	0.510	0.655

Table 5. Confusion matrix showing the congruence between two morphological classifications of 178 *Khaya anthotheca* herbarium samples based on (i) a visual inspection of more than 1000 herbarium samples allowing distinction of five morphogroups, and (ii) an objective statistical analysis (HCA) based on 27 leaf characters allowing distinction of five classes.

Morpho-groups	Classes				
	1	2	3	4	5
G1	2	1	0	0	22
G2	0	47	0	0	5
G3	34	0	0	0	0
G4	0	0	0	20	0
G5	0	0	28	0	19

In bold, the majority of individuals of each morphogroup included in each class.

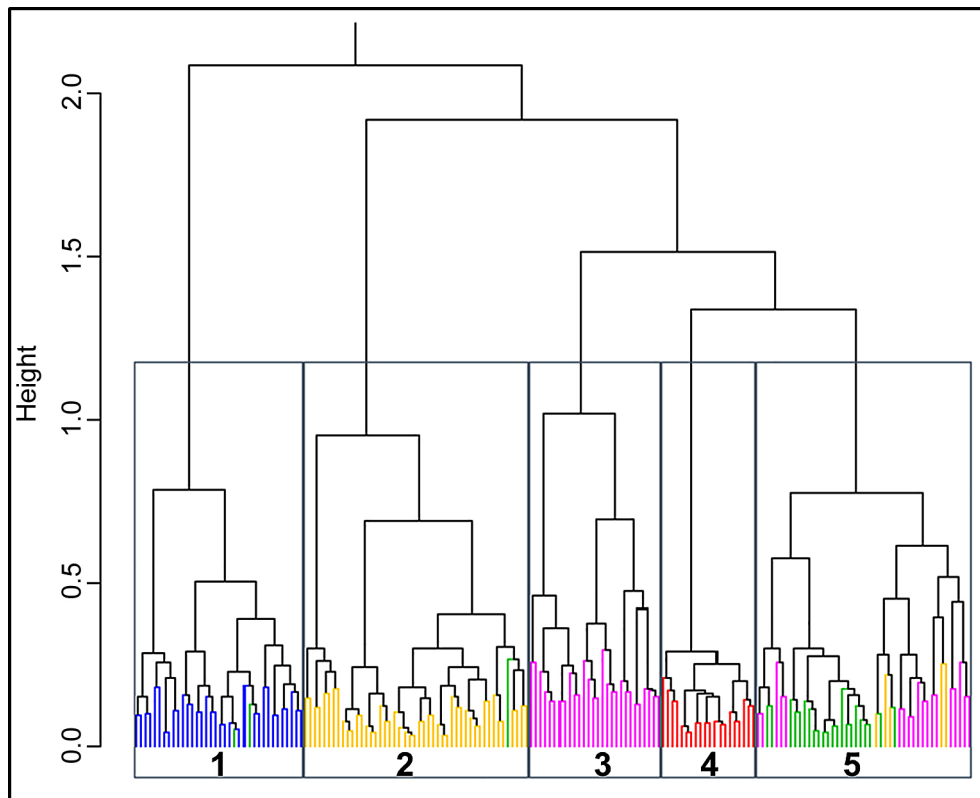


Fig. 4. Dendrogram of the hierarchical cluster analysis (HCA) applied on the morphological traits of 178 herbarium samples. The scale on the left corresponds to the inertia translating the distance of similarity between each class. The colours associated to tips follow the colour code of Fig. 1 to indicate the five morphogroups identified initially from a visual inspection of all available herbarium material (see Fig. 1). More detailed figures including a code for each sample can be found in suppl. Fig. S10.

tree habit (drooping or upright), flower (pentamerous or tetramerous), presence and absence of furrows on fruit, density and number of secondary veins (very dense and less dense). These two morphogroups were named G5a and G5b, finally giving six morphogroups when integrating the two approaches to categorize herbarium samples.

A Kruskal-Wallis test performed on all morphogroups (G1–G4, G5a, G5b) confirmed that all the quantitative variables measured showed significant differences among morphogroups ($P < 0.001$ for all variables). Wilcoxon-Mann-Whitney tests showed that most morphogroups differ from each other by 8 (G4 and G5b) to 17 (G2 and G5b) of the 18 quantitative characters (Table 6). However, G1 and G4 on the one hand and G5a and G5b on the other hand differ by only four variables: in particular, petiolule length of basal leaflet (PLBL), limb width of middle and terminal leaflets (LWML; LWTL), and apex length of middle leaflet (ALML) differ between G1 and G2, while limb length of middle and terminal leaflets (LLML; LLTL) and apex length of middle and terminal leaflets (ALML; ALTL) differ between G5a and G5b (Table 6).

Ordination of morphogroups. — The first four principal coordinates of the Hill and Smith ordination explain 33.1% of the total variance (12.4%, 9.4%, 6.7% and 4.6% for each axis). In general, the first two axes (1 and 2) mostly express the variation of quantitative variables, and the next two axes (3 and 4) are mostly influenced by qualitative variables (Fig. 5, Table 7). Nevertheless, the first qualitative characters that have a great contribution on axes 1 and 2 are the shapes of the base (BTL-BBL-BML.CAAO) and the nature of the apex leaflet (SL.EMG). The first axis mainly depends on the size of leaflets, petiole and petiolules (LWBL, LWML, LLBL, LLTL, LLML, LWTL, PLML, PLTL). The second axis is mostly related to the number of leaflet pairs and the number of lateral veins of leaflets, length of the rachis and the apex length of basal, middle, terminal leaflet (NLVTL, NLP, NLVML, RL, ALML, NLVBL, ALBL, ALTL), as well as to the length of the rachis and of the apex of leaflets (RL, ALML). The base of the leaflets (BBL, BTL, BML), the nature of apex leaflet (when acuminate) and the apex length of basal, middle and terminal leaflets (ALTL, ALBL, ALML) largely to contribute to axis 3 (when rounded-symmetric, AS) and 4 (when rounded, AR, or cuneiform-rounded, CUA).

The six morphogroups tend to be well separated on the first four axes of the Hill and Smith ordination (Fig. 6A,B): morphogroups G1 and G3 separate on axis 1, G2 and G5b on axis 2, G4 on axis 3 and G5a on axis 4 (the contribution of each morphogroup to the four axes is given in suppl. Table S1). In Fig. 6, the morphogroups G5a and G5b appear superimposed but not in the same plane. This could be due to the great variation of quantitative variables on all axes. These two morphogroups cannot therefore be separated on the sole basis of these quantitative variables.

Congruence between morphogroups, genetic clusters and their spatial distribution. — Six morphogroups and six genetic groups were eventually identified. Among the

52 individuals that could be both genotyped and measured, we found a perfect one-to-one correspondence between these morphogroups and genetic groups (Table 8). In addition, considering all examined individuals, the geographic distribution of these morphogroups coincides well with that of the corresponding genetic clusters (Figs. 1, 3A).

Comparison between different species and nomenclatural types. — The comparison of the final six morphogroups with diagnoses and nomenclatural types (see Bouka & al., 2019) allows clarifying the identification of taxa. No diagnosis of any named taxon, not even that of *Khaya anthotheca*, combined the characteristics of all the morphogroups. The morphogroups were identified as follows: G1 corresponds to the diagnosis of *Khaya agboensis* A.Chev.; G2 to that of *K. euryphylla* Harms; G3 to that of *K. anthotheca* (Welw.) C.DC. (sensu stricto, s.str.); and G5b to that of *K. nyasica* Stapf ex Baker f. s.str. No diagnosis or nomenclatural type corresponded to morphogroups G4 and G5a. In accordance with our species delimitation criteria, these morphogroups have been identified as putative new taxa and have been temporarily named *Khaya* sp. 1 for G4 and *Khaya* sp. 2 for G5a. The key identifies all of these species (the correspondence between the herbarium samples, the genetic clusters and the species to which we attribute them, including the country of origin, is given in suppl. Table S2).

Identification key for the *Khaya anthotheca* species complex

1. Leaves with 4 or 5 (6) pairs of often slightly falcate leaflets; flowers and fruits always 4-merous; corolla less than 4 mm long; ovules 4 to 10 (to 12) per locule; capsules 3–4.5(–5) cm diameter; southern Zambia and Malawi, northwest and southern Tanzania, Mozambique and Zimbabwe.....*K. nyasica*
1. Leaves with (1) 2 to 6 pairs of leaflets, ± similar and regular or the terminal pair unequal and ± falcate or different in shape; flowers and fruits 4- or 5-merous; corolla 3.2–5.7 mm long; ovules 12 to 24 per locule; capsules 4–7 cm diameter.....2
2. Leaves with (3) 4 to 6 pairs of leaflets, petiolules brownish-red to orangish-brown when dry; acumen caudate, generally plicate when dry; flowers 5-merous; capsules not exceeding 4.5 cm diameter; Katanga Province in D.R.C., northern Zambia and Malawi, northeastern Tanzania.....*K. sp. 2*
2. Leaves with (1) 3 to 5 pairs of leaflets; petiolules never brownish-red to orangish-brown; acumen neither caudate or plicate when dry; capsules always greater than 5 cm diameter.....3
3. Terminal leaflets always ± falcate (10.5–) 12.5–18.5 (–23) cm long, base always asymmetric; obtuse on one side, rounded on the other (Fig. 7C,D); southeastern D.R.C., southern Sudan, northeastern Republic of the Congo, and northern and central D.R.C. and Angola.....*K. anthotheca*
3. Terminal leaflets never falcate, (3–) 5–10 (–17) cm long, base generally symmetric (Fig. 7E).....4

Table 6. Quantitative leaf characters of the six morphogroups of *Khaya anthotheca* s.l. measured on 178 herbarium specimens at 18 traits.

Trait	Morphogroup					
	G1	G2	G3	G4	G5a	G5b
NLP	3.48 (3–4) ± 0.51 / b	3.19 (3–4) ± 0.4 / a	3.88 (3–5) ± 0.59 / c	3.5 (3–4) ± 0.51 / b	4.37 (3–6) ± 0.9 / d	4.71 (3–6) ± 0.94 / d
PL	6.7 (4.1–10.5) ± 1.69 / a	7.99 (5.7–11.7) ± 1.25 / b	10.39 (6.4–15.6) ± 2.22 / c	6.83 (2.9–10.7) ± 1.6 / a	8.26 (4.4–11.5) ± 1.9 / b	8.67 (5.6–12.2) ± 2.14 / b
RL	9.26 (5.9–16.9) ± 3.04 / ab	10.41 (6.8–18.3) ± 2.83 / b	14.95 (7.4–28.1) ± 5.53 / c	8.6 (5.1–13) ± 2.43 / a	12.77 (6.6–20) ± 3.69 / c	14.79 (7–23.2) ± 4.25 / c
PLBL	1 (0.5–1.7) ± 0.3 / b	1.01 (0.7–1.5) ± 0.17 / b	1.46 (0.8–2.7) ± 0.5 / c	0.76 (0.5–1.1) ± 0.19 / a	1.09 (0.3–5) ± 1.01 / ab	0.79 (0.4–1.5) ± 0.33 / a
LLBL	8.74 (5.8–13.7) ± 2.06 / a	10.8 (6–14.8) ± 1.62 / b	12.9 (8.6–18.7) ± 2.95 / c	8.13 (5.9–12) ± 1.37 / a	7.92 (5.5–11.2) ± 1.74 / a	8.3 (6.5–12.2) ± 1.48 / a
LWBL	5.25 (3.8–7.9) ± 1.09 / b	6.37 (4–8.8) ± 1.13 / c	6.74 (4.3–10) ± 1.73 / c	4.62 (3.3–6.1) ± 0.86 / ab	4.13 (2.9–5.8) ± 0.86 / a	4.26 (3–5.7) ± 0.78 / a
NLVBL	6.08 (5–8) ± 0.81 / a	6.17 (4–9) ± 1 / a	7.5 (6–10) ± 1.08 / b	6 (5–8) ± 0.92 / a	8.16 (5–11) ± 1.64 / bc	8.46 (6–13) ± 1.73 / c
ALBL	0.37 (0.2–0.7) ± 0.12 / bc	0.19 (0–0.6) ± 0.16 / a	0.4 (0.2–0.8) ± 0.14 / bc	0.43 (0.3–0.6) ± 0.09 / c	0.29 (0–0.6) ± 0.17 / b	0.37 (0–0.8) ± 0.17 / bc
PLML	0.88 (0.4–1.6) ± 0.31 / ab	0.98 (0.7–1.5) ± 0.18 / b	1.34 (0.7–2) ± 0.42 / c	0.72 (0.4–1.1) ± 0.19 / a	0.85 (0.3–1.4) ± 0.32 / ab	0.78 (0.3–1.3) ± 0.3 / a
LLML	11 (8.3–19.3) ± 2.63 / a	13.07 (1–16.2) ± 2.26 / c	15.9 (10.3–22.3) ± 3.56 / d	9.99 (7.5–13.2) ± 1.62 / a	10.99 (9.1–14.4) ± 1.57 / a	12.31 (7.8–16.7) ± 2.31 / b
LWML	5.3 (4.3–8.8) ± 1.15 / b	7.05 (4.8–9.6) ± 1.19 / c	7 (3.9–10.2) ± 1.68 / d	4.6 (3.6–6) ± 0.71 / a	4.4 (3.4–4.95) ± 0.8 / a	4.9 (3–5.9) ± 0.78 / a
NLVML	5.52 (6–12) ± 1.41 / a	6.92 (5–10) ± 1.19 / a	7.03 (6–11) ± 1.43 / b	4.72 (5–9) ± 1.01 / a	4.46 (8–13) ± 1.44 / c	4.79 (7–17) ± 2.57 / c
ALML	0.36 (0.2–0.7) ± 0.14 / bc	0.18 (0–0.5) ± 0.15 / a	0.39 (0.2–0.7) ± 0.13 / cd	0.44 (0.3–0.6) ± 0.09 / d	0.3 (0.1–0.6) ± 0.16 / b	0.42 (0–0.8) ± 0.17 / cd
PLTL	0.76 (0.3–1.7) ± 0.34 / ab	0.82 (1–1.2) ± 0.25 / b	1.16 (0.5–1.7) ± 0.32 / c	0.59 (0.3–1) ± 0.18 / a	0.71 (0.3–1.2) ± 0.3 / ab	0.65 (0.3–1.2) ± 0.3 / a
LLTL	10.48 (7–17.2) ± 2.69 / ab	13.22 (9.6–16.7) ± 1.74 / c	15.36 (9.9–22) ± 3.46 / d	9.51 (6.9–15.2) ± 1.95 / a	10.17 (7.3–13) ± 1.68 / a	11.51 (8.4–17.1) ± 2.29 / b
LWTL	5.11 (3.2–7.8) ± 1.15 / c	6.54 (3.6–9.1) ± 1.39 / d	5.91 (3.9–7.7) ± 1.27 / d	4.22 (3.2–6) ± 0.75 / b	3.75 (2.9–5) ± 0.62 / a	4.16 (2.6–6.1) ± 0.88 / ab
NLVTL	7.56 (6–12) ± 1.36 / a	7.06 (5–11) ± 1.23 / a	9.21 (7–13) ± 1.57 / b	7 (5–8) ± 0.92 / a	11.58 (8–15) ± 2.17 / c	11.93 (8–17) ± 2.29 / c
ALTL	0.36 (0.2–0.6) ± 0.12 / b	0.18 (0–0.5) ± 0.14 / a	0.38 (0.2–0.6) ± 0.11 / b	0.38 (0.2–0.6) ± 0.1 / b	0.24 (0.1–0.4) ± 0.12 / a	0.35 (0–0.7) ± 0.18 / b

Data given as: mean (minimum–maximum) ± standard deviation. Letters “a” to “d” summarize results of Wilcoxon–Mann–Whitney statistical tests where groups that do not differ significantly share a same letter. See Table 1 for the code of traits.

4. Leaves with (1) 2 or 3 (4) pairs of leaflets, lamina of the terminal pair rectangular to obovate-oblong, if 2 pairs present, the terminal pair always obovate-oblong and longer than the other one, apex emarginate, rounded to truncate, mucate or abruptly apiculate to shortly acuminate (Fig. 7A,B), the acumen often at the base of a wide, shallow sinus; Cameroon, southwestern D.R.C. and northern Republic of the Congo (Likouala and Sangha Provinces)*K. euryphylla*
4. Leaves with 3 or 4 pairs of leaflets, lamina of the terminal pair usually different in shape, rarely rectangular or obovate-oblong never with an acumen borne at the base of a sinus.....5
5. Lower surface of leaflets with a dense, discolorous network of weakly raised, distinct tertiary veins; inflorescence a dense, ± corymbiform panicle with straight flowering axes; Nigeria to Sierra Leone.....*K. agboensis*
5. Lower surface of leaflets with a low, slightly impressed and weakly visible network of tertiary veins: inflorescence an open, pyramidal panicle with arched flowering axes; D.R.C. (Tshopo Province), usually on hydromorphic soils.....*K. sp. 1*

■ DISCUSSION

How many species in the widespread taxon *Khaya anthotheca*? — *Khaya anthotheca* s.l. is the taxonomically most unstable species of the genus *Khaya* (Styles & White, 1991; Hawthorne & Gyakari, 2006; De Wilde, 2015; Bouka & al., 2019). Due to its large morphological variability, several

taxa were described and published, and later put in synonymy (Aubréville, 1959; Voorhoeve, 1965; Styles & White, 1991; Danquah & al., 2011). Here, combining statistical analyses of morphological traits with population-genetic data on large sample sizes (1407 herbarium vouchers examined, 481 samples genotyped) allowed us to better delineate species in *K. anthotheca* s.l., as was the case in other tropical tree taxa (e.g., Daïnou & al., 2014, 2016; Ikabanga & al., 2017; Lissambou & al., 2019; Ewédjè & al., 2020).

We identified six putative species following the criteria used by Lissambou & al. (2019): 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 biological species concept. This is the situation of the sympatric K4 and K3, and to some extent the parapatric K3 and K2, with a narrow area of range overlap in northern Republic of the Congo (Fig. 3). For genetic groups occurring in allopatry or parapatry, the interpretation is not as straightforward because they may potentially correspond to different populations of the same species rather than to different species, and there is no recognized genetic differentiation threshold (e.g., minimal F_{ST}) that can distinguish between the two situations. Nevertheless, the level of genetic differentiation observed between sympatric genetic clusters of *K. anthotheca* s.l. can help us assess whether genetic entities found in allopatry and parapatry belong to the same or to distinct species. Here, we found that the F_{ST} between clusters in allopatry or parapatry are higher than the F_{ST} between the sympatric K4 and K3 clusters (Table 4), supporting the

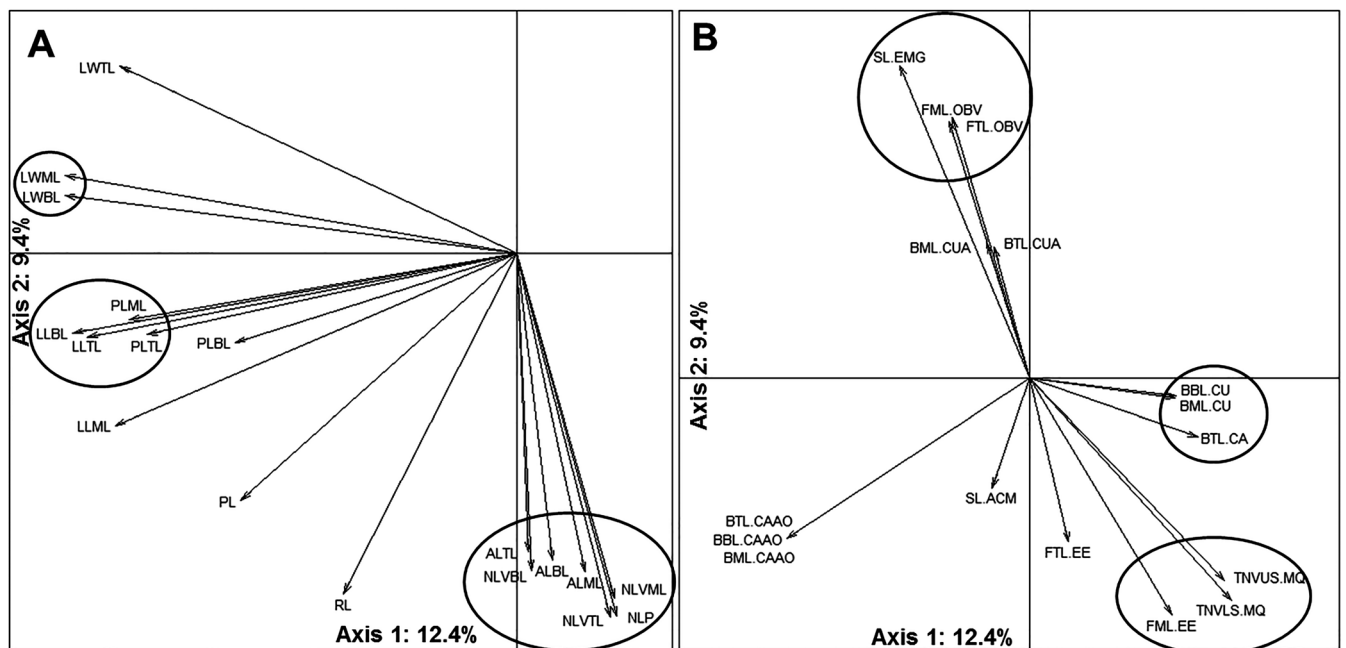


Fig. 5. Distribution of morphological variables in the plane of axes 1 and 2. A, Distribution of quantitative variables; B, Distribution of qualitative variables.

view that they also represent distinct species. Finally, we found a very good correspondence between these clusters and morphogroups, so that the morphological and biological species concepts are congruent. We thus identified a total of six genetic and morphological groups with allopatric (K1 and K2), parapatric (K2 and K3, K4 and K5a, K5a and K5b) and sympatric (K3 and K4) distributions. The striking correspondence between the geographic distributions of these six morphological and genetic groups, which are based on only partially overlapping samples, definitely confirms that six taxa should be recognized within *K. anthotheca* s.l.

It is worth noting that the initial classification of herbarium specimens by morphological affinity recognized only five groups, corresponding to five genetic clusters initially identified when STRUCTURE was run on all genotyped samples. Particular population-genetic features of K5 (strong deficit

in heterozygotes and spatial genetic structure, results not shown) drew our attention to the possibility that this cluster was possibly strongly sub-structured, as we confirmed by re-applying the STRUCTURE algorithm on these samples. This led us to re-examine herbarium samples from eastern and southern Africa, corresponding to *Khaya nyasica*, and apply a clustering algorithm on measured leaf characters. This analysis finally confirmed the existence of two morphogroups, which also differ in traits of reproductive organs, which are often lacking in herbarium specimens. Similarly, prior knowledge of the existence of genetic clusters greatly helped the recognition of corresponding morphogroups in the genera *Milicia* (Dainou & al., 2016), *Santiria* (Ikabanga & al., 2017) and *Greenwayodendron* (Lissambou & al., 2019). Nevertheless, genetic clustering, such as that obtained using STRUCTURE, is not necessarily a panacea for pre-identifying putative species and must be

Table 7. Morphological variables having greater-than-average contribution for the four first axes of the Hill and Smith analysis.

Axis 1		Axis 2		Axis 3		Axis 4	
Variable	(%)	Variable	(%)	Variable	(%)	Variable	(%)
LWBL	8.55	NLVTL	7.16	BBL.AS	10.11	BBL.AR	11.04
LWML	8.53	NLP	7.16	BTL.AS	10.11	BML.AR	11.04
LLBL	8.22	NLVML	6.46	BML.AS	10.11	BTL.AR	9.04
LLTL	7.69	RL	6.35	SL.ACM	4.88	BBL.CUA	8.76
LLML	6.70	ALML	5.53	ALTL	4.61	BML.CUA	7.35
LWTL	6.59	NLVBL	5.46	ALBL	4.16	BTL.CUA	7.12
PLML	6.20	ALBL	5.17	ALML	4.13	FBL.LE	3.64
PLTL	5.63	ALTL	4.91	NLVTL	3.34	BML.AOT	3.08
BTL.CAAO	3.78	SL.EMG	3.68	NLVML	3.19	BTL.AOT	3.08
BBL.CAAO	3.78	PL	3.35	TNVLS.MQ	2.93	BBL.AOT	3.08
BML.CAAO	3.78	BTL.CAAO	2.36	TNVUS.MQ	2.79	BTL.CU	3.03
PLBL	3.31	BBL.CAAO	2.36	NLVBL	2.27	BML.CU	2.19
PL	3.15	BML.CAAO	2.36	FML.EE	2.22	BBL.CU	2.12
BTL.CA	2.04	TNVLS.MQ	2.19	BBL.CUA	1.93	FBL.OB	1.93
BBL.CU	1.74	BTL.CUA	2.18	SL.EMG	1.62	FML.EE	1.66
BML.CU	1.69	FML.EE	2.12	SL.APL	1.52	BML.CA	1.48
TNVLS.MQ	1.45	BML.CUA	2.10	BML.CUA	1.41	FTL.EO	1.45
TNVUS.MQ	1.42	SL.ACM	1.95	BTL.CUA	1.37	TNVUS.MQ	1.43
RL	1.21	FML.OBV	1.91	BTL.CU	1.31	TNVLS.MQ	1.29
		TNVUS.MQ	1.91	BBL.CAAO	1.30		
		FTL.OBV	1.88	BML.CAAO	1.30		
		LWTL	1.87	BTL.CAAO	1.30		
		FTL.EE	1.67	FTL.OBV	1.28		
		LLML	1.65	FTL.EE	1.26		
				BML.CA	1.24		

Qualitative variables are represented by their character state. For example BML.AS: base of the middle leaflet (BML).rounded-symmetric (AS). The meaning of each variable can be found in Tables 1 and 2.

applied with caution. Firstly, genetic clusters may correspond to isolated conspecific populations and, alone, cannot be considered as evidence of distinct species if they do not occur in sympatry. Secondly, marker polymorphism, and especially the sample sizes of actual genetic clusters, can be critical for identifying them using a clustering algorithm. In our case, using the whole genetic dataset, we first detected only five clusters. It was only after looking further for genetic structuring within each cluster that we identified a sixth cluster, dividing the East African samples into two. This can probably be explained by the smaller sample sizes of the two East African clusters thus identified ($N = 17$ and 25 , compared to at least $N = 49$ for the other clusters, Table 3). Indeed, numerical simulations using a microsatellite dataset showed that an actual genetic cluster may remain undetected if its sample size is below 20 or 30 individuals (Lissambou & al., 2019).

A new circumscription of species in *Khaya anthotheca* s.l. and their morphological traits. — Our work gives a clearer picture of the species complex *Khaya anthotheca* s.l. than previously exposed in Pakull & al. (2019) and Monthe & al. (2019). We confirm the first conclusions of Pakull & al. (2019) that *K. nyasica* is a different taxon than *K. anthotheca*, but found, moreover, that *K. nyasica* itself comprises two different taxa. The comparison between morphogroups, diagnoses and nomenclatural types showed that four of the six morphogroups have already been described in the past and named as different species, before being put in synonymy with *K. anthotheca* (Bouka & al., 2017). The other two morphogroups appear to be “cryptic species” in the sense that they have never been identified as distinct taxa (Bickford & al., 2007).

The PCA of Hill and Smith showed that the most-discriminating variables that divide *K. anthotheca* s.l. morphogroups are: leaflet base shape, number of veins, length and width of leaflets, number of pairs of leaflets, petiole length, rachis length, leaflet shape, leaflet apex and tertiary network of veins on the lower surface (Table 6), as well as the nature of the inflorescences, flowers and fruits, especially ovary 4- or 5-locular, inflorescence type, size of corolla, number of ovules per loculus and diameter of the fruit. The specific characters used to differentiate each species can be found in the identification key. The most discriminating characters include qualitative characteristics, and floral characters reinforce the robustness of the morphological delineation, because these characters are often less phenotypically plastic than others (Kenfack, 2011).

Table 8. Confusion matrix of 52 individuals that could be categorized based both on their genotypes (cluster) and their morphological traits (morphogroup).

Morphogroups	Cluster					
	K1	K2	K3	K4	K5a	K5b
G1	8	–	–	–	–	–
G2	–	8	–	–	–	–
G3	–	–	9	–	–	–
G4	–	–	–	11	–	–
G5a	–	–	–	–	8	–
G5b	–	–	–	–	–	8

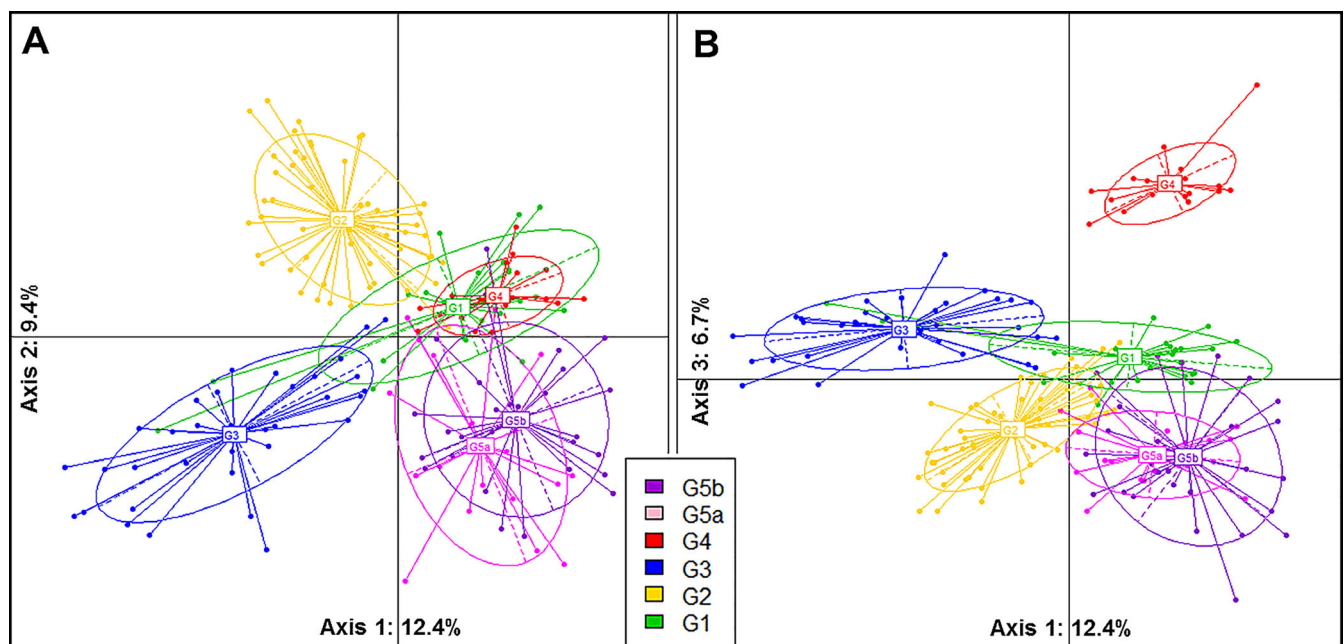


Fig. 6. Projection of individuals on the first three axes of the principal coordinates analysis of Hill and Smith. A, Projection on axes 1 and 2; B, Projection on axes 1 and 3. The coloured points group together the individuals assigned to the same morphogroup according to our a priori grouping.

We propose to recognize the following taxa:

- Morphogroup G1 and genetic cluster K1 correspond to *Khaya agboensis*, which is characterized by its oval leaflets, 3 to 4 pairs, with a rounded apex and ending abruptly with a very short acumen, and the lower surface of leaflets with a dense, discolorous network of weakly raised, distinct tertiary veins.

- Morphogroup G2 and genetic cluster K2 correspond to *Khaya euryphylla*, having the following diagnostic characteristics: leaves with (1) 2 or 3 (4) pairs of leaflets, blade of the terminal pair rectangular to obovate-oblong, if two pairs present, the terminal pair always obovate-oblong and longer than the other one, apex rounded, weakly emarginate or often narrowed abruptly into a short to wide acumen, with a shallow sinus.

- Morphogroup G3 and genetic cluster K3 correspond to *Khaya anthotheca*, which is characterized by its terminal pair of leaflets always falcate, base always irregular and asymmetric, obtuse on one side, rounded on the other on the other, or acute on both sides at different angles, sometimes uneven.

- Morphogroup G5b and gene cluster K5b correspond to *Khaya nyasica*, which is distinct by having leaves with 4 or 5 (6) pairs of often slightly falcate leaflets, and flowers and fruits always 4-merous.

Morphogroup/clusters G4/K4 (*Khaya* sp. 1) and G5a/K5a (*Khaya* sp. 2) correspond to two undescribed species (cryptic species within the meaning of Bickford & al., 2007), distinct from the other ones respectively by the following characters:

- *Khaya* sp. 1 shows a lower surface of leaflets with a low, slightly impressed and weakly visible network of tertiary veins, and its inflorescence is an open, pyramidal panicle with arched flowering axes.

- *Khaya* sp. 2 has leaves with (3) 4 to 6 pairs of leaflets, the terminal pair with (9 to) 13 to 15 (to 17) pairs of secondary veins; its petiolules are strongly discolorous, brownish-red to orangish-brown when dry; the acumen is caudate, often laterally pinched when dry; and flowers are always 5-merous.

The six morphogroups constituting the taxon *Khaya anthotheca* s.l. can be separated on the basis of leaf characters,

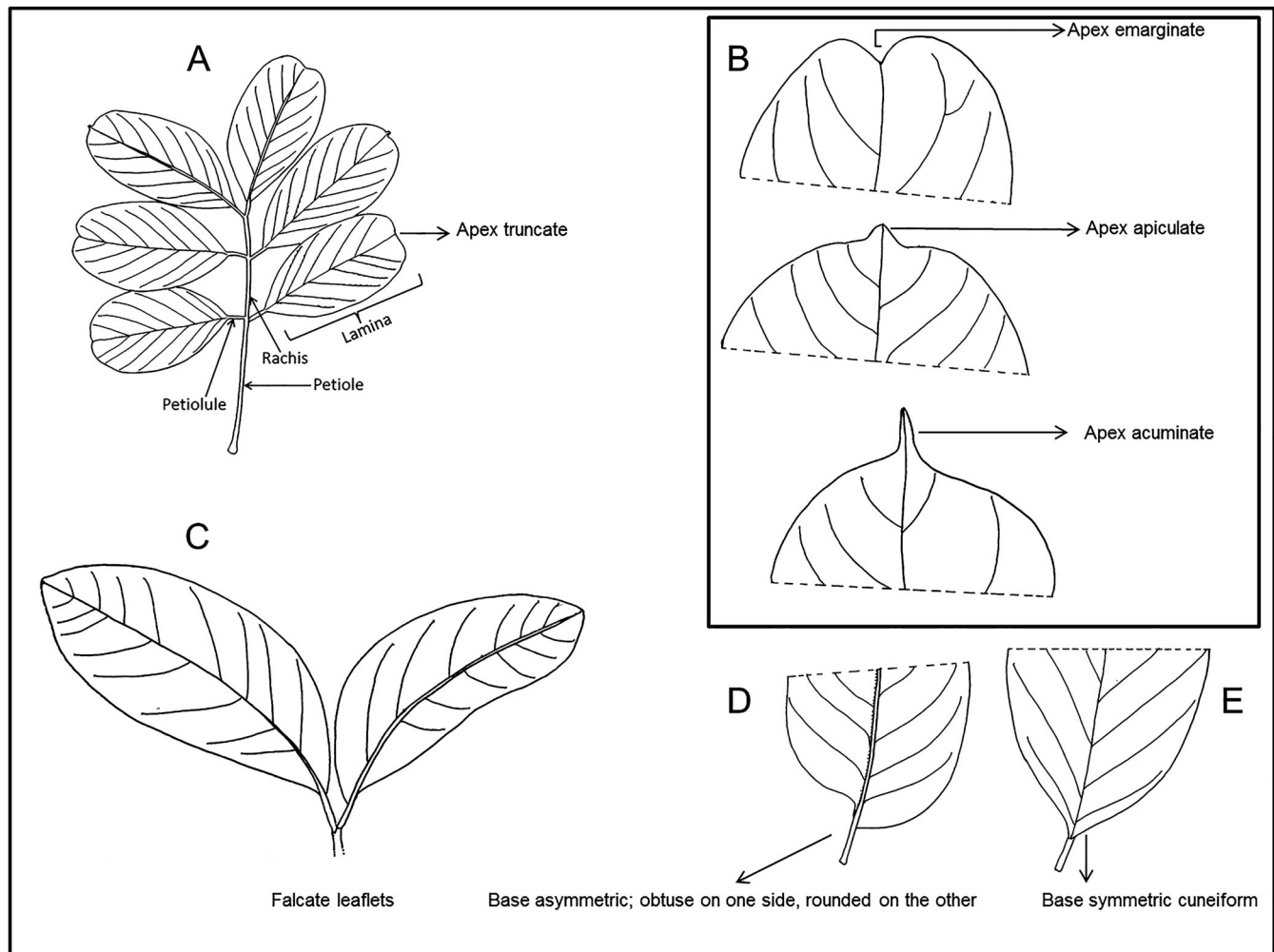


Fig. 7. Some characters subjected to morphological analysis and in the identification keys. **A**, Different parts of a compound leaf; **B**, Different types of apex; **C**, Falcate leaflets; **D**, Base asymmetric; obtuse on one side, rounded on the other; **E**, Base symmetric cuneiform.

which seems surprising for a genus described as showing very high intra-species variability in leaf traits (Aubréville, 1959; Styles & White, 1991). Because of this variability, we would have expected a very heterogeneous dataset with a large overlap of characters. In addition to these leaf traits and despite the paucity of fertile herbarium specimens, our study also shows that certain floral traits contribute to the recognition of morphologically distinct species (see especially *K. nyasica*, split by our study into two species). This is in contradiction with Staner & Gilbert (1958) and Aubréville (1959), who argued that floral characters do not separate the species of the genus *Khaya*. This contradiction might be explained by the large range and ecological amplitude of *K. anthotheca* s.l., so that the floral samples these authors examined may not have covered the taxon's whole geographical distribution.

Ecological distribution of the newly recognized species. — All genetic groups from K1 to K4 are typical of lowland dense, humid, semi-deciduous forests of West and Central Africa, and genetic groups K5a and K5b are more characteristic of drier woodlands and gallery forests on plateaux (White, 1983; Bouka & al., 2019). However, data on herbarium labels or collected from various people in Tshopo Province of Kisangani (unpub. field observations) suggest that *Khaya* sp. 1 may be largely restricted to forests growing on hydromorphic soils. More data are necessary to assess these ecological preferences but, if proven, they might explain the sympatric distribution of *Khaya* sp. 1 (K4) and *K. anthotheca* s.str. (K3). Interestingly, the terra firme lowland species *K. agboensis*, *K. euryphylla* and *K. anthotheca* (see respectively morphogroups G1, G2 and G3 on Fig. 1 and genetic groups K1, K2 and K3 on Fig. 3A) display a geographical range mainly at the periphery of dense, humid forest refugia of past dry geological times, with the contact zone between the first two species being materialised by the Cameroon volcanic line and the contact zone between the last two species by the Sangha River Interval (Maley, 1996; Maley & al., 2018). Genetic group K5a (*Khaya* sp. 2) seems to occupy high plateaux of the interior west of the African Rift, while group K5b (*K. nyasica*) is widespread in areas more subject to Indian Ocean climatic influences. These species might have also undergone different climatic histories during past geological times, but more thorough phyto-geographic and genetic studies are necessary in order to disentangle their evolutionary history.

Genetic diversity and divergence of genetic groups. — The polymorphism indices between the clusters varied substantially, with high diversity in clusters K1 to K4 and low diversity in K5a and K5b, suggesting that these last two may be recently shaped new species (founder events) or that they suffered a strong demographic bottleneck in the recent past. However, the low diversity of K5a and K5b could also be explained by the low representation of individuals from these clusters in the samples used to identify SNPs (a similar problem was reported by Dainou & al., 2016, for *Milicia regia* in Central Africa). Within each genetic group, the observed mean heterozygosity (H_O) is generally slightly lower than the expected mean heterozygosity (H_E) under Hardy-Weinberg

conditions, leading to positive inbreeding coefficients (F_I), which probably result from genetic substructure, but also potentially from the impact of null alleles (Garnier-Géré & Chikhi, 2013).

The genetic differentiation (F_{ST}) values between different genetic clusters highlight considerable differentiation (Wright, 1978). The minimum F_{ST} is observed between the genetic groups in sympatry (K3 and K4), while the maximum F_{ST} value was reached between the genetic groups in parapatry K5a and K5b. These F_{ST} values suggest that the six clusters of *Khaya anthotheca* s.l. can be considered to belong to distinct species. A dated phylogeny based on plastid and ribosomal DNA indicates that *Khaya* species appear to have diverged during the Pliocene to Pleistocene, more recently than in another genus of Meliaceae, *Entandrophragma*, in which species diverged mainly during the Miocene (Monthe & al., 2019). Unfortunately, the resolution was limited within the genus *Khaya*, and the topology was incongruent between ribosomal and plastid DNA, so that we cannot yet infer how the genus diversified.

Past and present subdivision of *Khaya anthotheca* s.l. — A phytochemical study by Adesida & al. (1971) on the woods and seeds of *Khaya* identified three different chemotypes within *K. anthotheca* s.l. (Bouka & al., 2019). These three chemical types corresponded to: (1) individuals from West Africa (Ghana, Nigeria, Sierra Leone), (2) individuals collected in Central and East Africa (Republic of the Congo, Uganda, Angola) and (3) individuals identified as *K. nyasica* s.l. coming from East Africa. In comparison with our morphogroups, the first chemical type refers to individuals of *K. agboensis*, the second to individuals of *K. anthotheca* s.str. and the third to those of *K. nyasica* s.str.

These authors also noted morphological differences in the individuals constituting these three chemical groups, in particular the leaf and fruit traits. They even reported that a form of *Khaya anthotheca* s.l. common in Budongo Forest, in Uganda, differs from the West African material by the morphology of the leaves and the fruits, while the specimens of Republic of the Congo and Angola appeared intermediate. However, despite the existence of these morphological differences, Adesida & al. (1971) concluded that these three chemical groups of *K. anthotheca* s.l. could not support separation into different species because the distribution of *K. anthotheca* s.l. is discontinuous. In addition to the morphological differences, our results show that these three forms are also genetically different and correspond to *K. agboensis*, *K. anthotheca* s.str. and *K. nyasica* s.str. These three species have been described as such in the past, and our results justify their maintenance as different species.

Apart from these three chemical groups, Adesida & al. (1971) also identified other chemotypes, at least one of which was specific to each of the four other *Khaya* species that are currently accepted in this genus and which have no taxonomic problems (Bouka, 2017; Bouka & al., 2017, 2019). In view of our results and in comparison with those of Adesida & al. (1971), each species of the genus *Khaya* appears to contain

at least one specific chemotype. As hypothesized in Bouka & al. (2019), chemical composition in *Khaya* seems to show taxonomic value and should be further investigated.

Another more recent study, carried out by Pakull & al. (2019) on the whole genus *Khaya*, could not clearly distinguish *K. anthotheca* s.l. from *K. nyasica* s.l. In this study, these two species and *K. madagascariensis* were grouped in a single genetic group. However, 4 of the 15 nSNPs “golden markers” in this study separated *K. anthotheca* from *K. nyasica*. Nevertheless, this non-separation of these two species could be due to the small number of samples of *K. nyasica* in their study compared to other species (53 samples of *K. nyasica* out of a total of 2222). Indeed, genetic groups represented by a small number of samples may not be detected when they are analysed with well-represented groups (Porrás-Hurtado & al., 2013; Wang, 2017).

■ CONCLUSION

The combination of genetic and morphological methods have made it possible to show that *Khaya anthotheca* s.l. includes several taxa and that these can be considered distinct species according to the biological and morphological species concepts. Indeed, our results reveal that *K. nyasica* s.l. contains two species: *K. nyasica* s.str. and *Khaya* sp. 2 (a putative new species still to be fully described). In addition, within *K. anthotheca* s.l., apart from *K. anthotheca* s.str., which must be circumscribed, two other species must be rehabilitated: *K. agboensis* and *K. euryphylla*. Finally, another potential new species (*Khaya* sp. 1), which seems to be a narrow endemic of the center of the Democratic Republic of the Congo, and in particular in and around Tshopo Province, has also been identified. Detailed studies of the flower and comparison of these species with other species of the genus *Khaya* are, however, necessary to assess the degree of their differentiation and to carry out a full taxonomic revision of the genus. This work is in progress and will be published separately. As underlined by Bouka & al. (2019), a better delimitation of tree species brings also management issues, as rare hidden species of *Khaya* are exploited under the name of more widespread ones, potentially leading to their disappearance. As such, a revision of the conservation status of all the *Khaya* species is required. Beyond clarifying the taxonomic limits of *K. anthotheca* s.l., our study illustrates the strength of combining morphological and population-genetic data, as it shows that new taxa can be discovered using these approaches, underlining also that the hidden species richness of African trees is still greatly underestimated (e.g., Ikabanga & al., 2017; Lissambou & al., 2019).

■ AUTHOR CONTRIBUTIONS

GUDB, CD, DM, JLL and JF designed the research. GUDB, CD, BD, OJH and MRME contributed to the sampling. GUDB,

CD, JLL and JF conducted morphological analyses. GUDB, MRME, KD, CD, BD, and OJH performed the genetic data analyses. GUDB, CD 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. — GUDB, <https://orcid.org/0000-0001-8850-522X>; MRME, <https://orcid.org/0000-0002-5829-6321>; DM, <https://orcid.org/0000-0002-7271-901X>; OJH, <https://orcid.org/0000-0003-2052-1527>

■ ACKNOWLEDGMENTS

We thank the Service de coopération et d’action culturelle de l’Ambassade de France au Congo (SCAC), the International Foundation for Science (IFS), the Conservation Action Research Network Congo Basin Program Grant (CARN-CBGP), the Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD), the German Federal Ministry of Food and Agriculture (BMEL) in frame of the “Large scale project on genetic timber verification”, the International Tropical Timber Organization (ITTO), the Prince Albert II of Monaco Foundation, the CGIAR Research Program on Forests, Trees and Agroforestry (FTA), and DoubleHelix Tracking Technologies for funding field work, lab work, and herbarium visits. MassARRAY MultiPLEX genotyping was performed at the Genome and Transcriptome Facility of Bordeaux and funded by grants from the Conseil Régional d’Aquitaine (n° 20030304002FA and n° 20040305003FA), from the European Union (FEDER n° 2003227) and from Investissements d’Avenir (Convention attributive d’aide n° ANR-10-EQPX-16-01). We would like to express our gratitude to the forestry companies Thanry, Mokabi and Likouala Timber for the assistance and logistical supports provided during the various sampling missions. We would also like to thank the curators and staff members of all herbaria (BM, BR, BRLU, FHO, G, HBG, K, L, P, PRE, WAG, YA) for their help during the various visits, for making herbarium specimens available, and for the authorisation to take DNA samples from them. Our special thanks to Raymond Baudoin, Vivian Kuhlenkamp and Sydney T. Ndolo Ebika for their excellent assistance in statistical analyses, DNA extraction and for making illustrations, respectively.

■ LITERATURE CITED

- Adesida, G.A., Adesogan, E.K., Okorie, D.A., Taylor, D.A.H. & Styles, B.T. 1971. The limonoid chemistry of the genus *Khaya* (Meliaceae). *Phytochemistry* 10: 1845–1853. [https://doi.org/10.1016/S0031-9422\(00\)86448-1](https://doi.org/10.1016/S0031-9422(00)86448-1)
- African Plant Database 2020. African Plant Database, version 3.4.0. Conservatoire et Jardin botaniques de la Ville de Genève and South African National Biodiversity Institute, Pretoria. <http://africanplantdatabase.ch> (accessed 24 Apr 2020).
- Aubréville, A. 1959. *La flore forestière de la Côte d’Ivoire*, 2nd ed., vol. 2. Nogent-sur-Marne: Centre Technique Forestier Tropical.
- Bickford, D., Lohman, D.J., Sodhi, N.S., Ng, P.K.L., 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>
- Birky, C.W., Jr., Adams, J., Gemmel, M. & Perry, J. 2010. Using population genetic theory and DNA sequences for species detection and identification in asexual organisms. *PLoS ONE* 5(5): e10609. <https://doi.org/10.1371/journal.pone.0010609>
- 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.
- Bouka, D.U.G., Florence, J., Doumenge, C., Loumeto, J.J. & McKey, D. 2017. *Khayae* (Meliaceae) species nomenclator. *Adansonia* 39: 15–30. <https://doi.org/10.5252/a2017n1a2>

- Bouka, D.U.G., Doumenge, C., Loumeto, J.J., Florence, J. & McKey, D. 2019. Des confusions entre espèces préjudiciables à la gestion durable des essences forestières: l'exemple des acajous d'Afrique (*Khaya*, Meliaceae). *Bois Forêts Trop.* 339: 17–32. <https://doi.org/10.19182/bft2019.339.a31714>
- CTFT [Centre technique forestier tropical] 1988. *Khaya senegalensis* (Desr.) A. Juss. *Bois Forêts Trop.* 218: 43–56.
- Dainou, 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>
- Dainou, 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>
- Danquah, J.A., Appia, M. & Ari, P. 2011. Eco-geographic variation in leaf morphology of African Mahogany (*Khaya anthotheca* and *Khaya ivorensis*) provenances in Ghana. *Eur. J. Sci. Res.* 51: 18–28.
- Dayrat, B. 2005. Towards integrative taxonomy. *Biol. J. Linn. Soc.* 85: 407–415. <https://doi.org/10.1111/j.1095-8312.2005.00503.x>
- De Queiroz, K. 2007. Species concepts and species delimitation. *Syst. Biol.* 56: 879–886. <https://doi.org/10.1080/10635150701701083>
- De Wilde, H.J.F.E. 2015. Meliaceae. Pp. 5–74. In: Sosef, M.S.M., Florence, J., Ngok, B.L. & Bourobou, B.H.P. (eds.), *Flore du Gabon*, vol. 47. Weikersheim: Margraf Publishers.
- Dray, S. & Dufour, A.B. 2007. The ade4 package: Implementing the duality diagram for ecologists. *J. Statist. Softw.* 22(4): 1–20. <https://doi.org/10.18637/jss.v022.i04>
- 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., Walmaçq, 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>
- Dumolin, S., Demesure, B. & Petit, R. 1995. Inheritance of chloroplast and mitochondrial genomes in pedunculate oak investigated with an efficient PCR method. *Theor. Appl. Genet.* 91: 1253–1256. <https://doi.org/10.1007/BF00220937>
- 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/>
- 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>
- Ewédjè, E.E.B.K., Jansen, S., Koffi, G.K., Staquet, A., Piñeiro, P., Essaba, R.A., Obiang, N.L.E., Dainou, K., Biwolé, A.B., Doucet, J.L. & Hardy, O.J. 2020. Species delimitation in the African tree genus *Lophira* (Ochnaceae) reveals cryptic genetic variation. *Conservation Genet.* 21: 501–514. <https://doi.org/10.1007/s10592-020-01265-7>
- Falush, D., Stephens, M. & Pritchard, J.K. 2003. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* 164: 1567–1587. <https://www.genetics.org/content/164/4/156>
- Garnier-Géré, P. & Chikhi, L. 2013. Population subdivision, Hardy-Weinberg equilibrium and the Wahlund effect. In: *eLS*. Chichester: John Wiley & Sons. <https://doi.org/10.1002/9780470015902.a0005446.pub3>, (accessed 25 Apr 2020).
- Gower, J.C. 1971. A general coefficient of similarity and some of its properties. *Biometrics* 27: 857–871. <https://www.jstor.org/stable/2528823>
- 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>
- Harrison, R.G. 2012. The language of speciation. *Evolution* 66 (12): 3643–3657. <https://doi.org/10.1111/j.1558-5646.2012.01785.x>
- Hawthorne, W. 1998. *Khaya anthotheca*. The IUCN Red List of Threatened Species 1998: e.T32235A9690061. <https://doi.org/10.2305/IUCN.UK.1998.RLTS.T32235A9690061.en> (accessed 16 May 2020).
- Hawthorne, W.D. & Gyakari, N. 2006. *Photoguide for the forest trees of Ghana: A tree-spotter's field guide for identifying the largest trees*. Oxford: Oxford Forestry Institute, Department of Plant Sciences.
- Hill, M.O. & Smith, A.J.E. 1976. Principal component analysis of taxonomic data with multi-state discrete characters. *Taxon* 25: 249–255. <https://www.jstor.org/stable/1219449>
- Ikabanga, D.U., Stévant, T., Koffi, K.G., Monthe, 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>
- Jakobsson, M. & Rosenberg, N.A. 2007. CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23: 1801–1806. <https://doi.org/10.1093/bioinformatics/btm233>
- Kaufman, L. & Rousseeuw, P.J. 1990. *Partitioning around medoids (program PAM): Finding groups in data; An introduction to cluster analysis*. New York: Wiley.
- Keay, R.W.J. & Hepper, F.N. 1963. *Flora of West Tropical Africa*, vol. 2. Millbank London: Crown Agents for Oversea Governments and Administrations.
- 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>
- Liu, J., Zhou, W. & Gong, X. 2015. Species delimitation, genetic diversity and population historical dynamics of *Cycas diannanensis* (Cycadaceae) occurring sympatrically in the Red River region of China. *Frontiers Pl. Sci. (Online journal)* 6: 696. <https://doi.org/10.3389/fpls.2015.00696>
- Lissambou, B.J., Couvreur, T.L.P., Atteke, C., Stévant, T., Piñeiro, R., Dauby, G., Monthe, F.K., Ikabanga, D.U., Sonké, B., M'batchi, B. & Hardy, O.J. 2019. Species delimitation in the genus *Greenwayodendron* based on morphological and genetic markers reveals new species. *Taxon* 68: 442–454. <https://doi.org/10.1002/tax.12064>
- Maley, J. 1996. The African rain forest – Main characteristics of changes in vegetation and climate from the Upper Cretaceous to the Quaternary. *Proc. Roy. Soc. Edinburgh, B, Biol. Sci.* 104: 31–73. <https://doi.org/10.1017/S0269727000006114>
- Maley, J., Doumenge, C., Giresse, P., Mahé, G., Philippon, N., Hubau, W., Lokonda, M.O., Tshibamba, J.M. & Chepstow-Lusty, A. 2018. Late Holocene forest contraction and

- fragmentation in central Africa. *Quatern. Res.* 89: 43–59. <https://doi.org/10.1017/qua.2017.97>
- Miller, M.R., Dunham, J.P., Amores, A., Cresko, W.A. & Johnson, E.A. 2007. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Res.* 17: 240–248. <https://doi.org/10.1101/gr.5681207>
- Monthe, F.K., Duminil, J., Kasongo, Y.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.D.C. *Tree Genet. Genomes* 14: 66. <https://doi.org/10.1007/s11295-018-1277-6>
- Monthe, F.K., Migliore, J., Duminil, J., Bouka, G., Demenou, B.B., Doumenge, C., Blanc-Jolivet, C., Ekué, M.R.M. & Hardy, O.J. 2019. Phylogenetic relationships in two African Cedreloideae tree genera (Meliaceae) reveal multiple rain/dry forest transitions. *Perspect. Pl. Ecol.* 37: 1–10. <https://doi.org/10.1016/j.ppees.2019.01.002>
- Naciri, Y. & Linder, P. 2015. Species delimitation and relationships: The dance of the seven veils. *Taxon* 64: 3–16. <https://doi.org/10.12705/641.24>
- Pakull, B., Mader, M., Kersten, B., Ekué, M.R.M., Bouka, D.U.G., Paulini, M., Bouda, Z.H.N. & Degen, B. 2016. Development of nuclear, chloroplast and mitochondrial SNP markers for *Khaya* sp. *Conservation Genet. Resources* 8: 283–297. <https://doi.org/10.1007/s12686-016-0557-4>
- Pakull, B., Ekué, M.R.M., Bouka, D.U.G., Doumenge, C., McKey, D.B., Loumeto, J.J., Opuni-Frimpong, E., Yorou, S.N., Nacoulma, B.M.Y., Guelly, K.A., Ramamonjisoa, L., Thomas, D., Guichoux, E., Loo, J. & Degen, B. 2019. Genetic differentiation among different *Khaya* species based on a large SNP array. *Conservation Genet.* 20: 1035–1044. <https://doi.org/10.1007/s10592-019-01191-3>
- 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>
- Pritchard, J.K., Stephens, M. & Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959. <https://www.genetics.org/content/155/2/945>
- R Core Team 2020. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Available from <http://www.R-project.org/> (accessed 25 Apr 2020).
- Reeves, P.A. & Richards, C.M. 2007. Distinguishing terminal monophyletic groups from reticulate taxa: Performance of phenetic, tree-based, and network procedures. *Syst. Biol.* 56: 302–320. <https://doi.org/10.1080/10635150701324225>
- Rosenberg, N.A. 2004. DISTRUCT: A program for the graphical display of population structure. *Molec. Ecol. Notes* 4: 137–138. <https://doi.org/10.1046/j.1471-8286.2003.00566.x>
- Stace, C.A., Preston, C.D. & Pearman, D.A. 2015. *Hybrid flora of the British Isles*. Durham: Botanical Society of Britain & Ireland.
- Staner, P. & Gilbert, G. 1958. Meliaceae. Pp. 174–180 in: Robyns, W., Staner, P., Demaret, F., Germain, R., Gilbert, G., Hauman, L., Homès, M., Jurion, F., Lebrun, J., Vaaden Abeel, M. & Boutique R. (eds.), *Flore du Congo belge et du Ruanda-Urundi: Spermatophytes*, vol. 7. Brussels: Institut National pour l'Étude Agronomique du Congo belge.
- Stearn, W.T. 1992. *Botanical Latin*, 4th ed. Trowbridge: Timber Press.
- Struyf, A., Hubert, M. & Rousseeuw, P.J. 1997. Integrating robust clustering techniques in S-PLUS. *Computat. Statist. Data Analysis* 26: 17–37. [https://doi.org/10.1016/S0167-9473\(97\)00020-0](https://doi.org/10.1016/S0167-9473(97)00020-0)
- Styles, B.T. & White, F. 1991. *Flora of tropical East Africa: Meliaceae*. Rotterdam: A.A. Balkema.
- Tukey, J. 1977. *Exploratory data analysis*. Reading: Addison-Wesley.
- 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>
- Voorhoeve, A.G. 1965. *Liberian high forest trees: A systematic botanical study of the 75 most important or frequent high forest trees, with reference to numerous related species*. Wageningen: Centrum voor Landbouwpublikaties en Landbouwdocumentatie. <https://edepot.wur.nl/187074>
- 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>
- Ward, J.H., 1963. Hierarchical grouping to optimize an objective function. *J. Amer. Statist. Assoc.* 58: 236–244. <https://doi.org/10.1080/01621459.1963.10500845>
- White, F. 1983. *The vegetation of Africa: A descriptive memoir to accompany the UNESCO/AETFAT/UNSO vegetation map of Africa*. Paris: Unesco.
- Whitney, K.D., Ahern, J.R., Campbell, L.G., Albert, L.P. & King, M.S. 2010. Patterns of hybridization in plants. *Perspect. Pl. Ecol.* 12: 175–182. <https://doi.org/10.1016/j.ppees.2010.02.002>
- 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>
- Wright, S. 1978. *Evolution and the genetics of population*, vol. 4, *Variability within and among natural populations*. Chicago: University of Chicago Press.

Appendix 1. Herbarium vouchers from which leaflet fragments used for the genetic analysis were taken.

Khaya agboensis A.Chev., Ivory Coast, *A. Aubréville 10* (P, P00453216); *Khaya agboensis* A.Chev., Ivory Coast, *A.J.B. Chevalier 57B 22342* (P, P00453217); *Khaya agboensis* A.Chev., Ivory Coast, *A.J.M. Leeuwenberg 3744* (P, P00453220); *Khaya agboensis* A.Chev., Sierra Leone, *H.D. Jordan 2081* (P, P00453250); *Khaya agboensis* A.Chev., Liberia, *C. Versteegh & J.W. Jansen 797* (BR, BR0000016994836); *Khaya agboensis* A.Chev., Liberia, *C. Versteegh & J.W. Jansen 819* (BR, BR0000016994836); *Khaya agboensis* A.Chev., Liberia, A.G. Voorhoeve 1129 (BR, BR0000016994812); *Khaya agboensis* A.Chev., Ivory Coast, *A. Aubréville 10* (BR, BR0000016994850); *Khaya agboensis* A.Chev., Ivory Coast, *A.J.M. Leeuwenberg 3744* (BR, BR0000016994843); *Khaya agboensis* A.Chev., Ivory Coast, *A.J.B. Chevalier 22342* (P, P00453217); *Khaya euryphylla* Harms, Republic of Congo, *G.U.D. Bouka 1* (IEC); *Khaya euryphylla* Harms, Republic of Congo, *G.U.D. Bouka 2* (IEC); *Khaya euryphylla* Harms, Cameroon, *G.U.D. Bouka 6* (IEC); *Khaya euryphylla* Harms, Cameroon, *G.U.D. Bouka 11* (IEC); *Khaya euryphylla* Harms, Cameroon, *G.U.D. Bouka 7* (IEC); *Khaya euryphylla* Harms, Cameroon, *G.U.D. Bouka 8* (IEC); *Khaya euryphylla* Harms, Cameroon, *G.U.D. Bouka 9* (IEC); *Khaya euryphylla* Harms, Cameroon, *G.U.D. Bouka 10* (IEC); *Khaya euryphylla* Harms, Cameroon, *G.U.D. Bouka 12* (IEC); *Khaya anthotheca* (Welw.) C.D.C., Angola, *R. Dechamps, F. Murta & M. Da Silva 1492* (BR, BR0000016995055); *Khaya anthotheca* (Welw.) C.D.C., RDC, *R. Pierlot 2334* (BR, BR0000016996014); *Khaya anthotheca* (Welw.) C.D.C., RDC, *Terese B. Hart 370* (BR, BR0000016995673); *Khaya anthotheca* (Welw.) C.D.C., Uganda, *B. Styles 125* (BR, BR0000016995000); *Khaya anthotheca* (Welw.) C.D.C., Uganda, *B. Styles 124* (BR, BR0000016994997); *Khaya anthotheca* (Welw.) C.D.C., Uganda, *B. Styles 338* (BR, BR0000016994980); *Khaya anthotheca* (Welw.) C.D.C., Uganda, *B. Styles 339* (BR, BR0000016995017); *Khaya anthotheca* (Welw.) C.D.C., Uganda, *B. Styles 105* (BR, BR0000016994966); *Khaya anthotheca* (Welw.)

Appendix 1. Continued.

C.D.C., Uganda, *B. Styles* 67 (BR, BR0000016994935); *Khaya anthotheca* (Welw.) C.D.C., Uganda, *B. Styles* 98 (BR, BR0000016994911); *Khaya anthotheca* (Welw.) C.D.C., Central African Republic, *C. Tisserant* 2008 (BR, BR0000016994898); *Khaya anthotheca* (Welw.) C.D.C., Uganda, *B. Styles* 337 (BR, BR0000016995024); *Khaya anthotheca* (Welw.) C.D.C., *D. Kombi* 2/A (P, P02286982); *Khaya nyasica* Stapf ex Baker f., Tanzania, *Semsei* 1076 (BR, BR0000016996984); *Khaya nyasica* Stapf ex Baker f., Tanzania, *Riva Saw Mills* 3044 (BR, BR0000016997011); *Khaya nyasica* Stapf ex Baker f., Tanzania, *Maughan Brown* 2883 (BR, BR0000016997004); *Khaya nyasica* Stapf ex Baker f., Tanzania, *F. Haerdi* 175/0 (BR, BR0000016997035); *Khaya nyasica* Stapf ex Baker f., Tanzania, *F. Haerdi* 175/0 (BR, BR0000016997042); *Khaya nyasica* Stapf ex Baker f., Tanzania, *A. Stolz* 1626 (BR, BR0000016996922); *Khaya nyasica* Stapf ex Baker f., Zambia, *Brenan J.P.M. & Greenway P.J.* 7816 (BR, BR0000016997080); *Khaya nyasica* Stapf ex Baker f., Tanzania, *H.J. Schlieben* 5342 (BR, BR0000016996991); *Khaya nyasica* Stapf ex Baker f., Mozambique, *F.A. Mendonça* 2566 (BR, BR0000016997240); *Khaya nyasica* Stapf ex Baker f., Malawi, *J. Burt Davy* 21948 (BR, BR0000016997158); *Khaya nyasica* Stapf ex Baker f., Mozambique, *F.A. Mendonça* 4126 (BR, BR0000016997233); *Khaya nyasica* Stapf ex Baker f., Mozambique, *M.P. Correia & A. Marques* 2447 (BR, BR0000016997219); *Khaya nyasica* Stapf ex Baker f., Mozambique, *H. Faulkner* 469 (BR, BR0000016997127); *Khaya nyasica* Stapf ex Baker f., Mozambique, *F.A. Mendonça* 2368 (BR, BR0000016997264); *Khaya nyasica* Stapf ex Baker f., Mozambique, *F.A. Mendonça* 2680 (BR, BR0000016997257); *Khaya nyasica* Stapf ex Baker f., Mozambique, *M.F. De Carvalho* 859 (BR, BR0000016997202); *Khaya nyasica* Stapf ex Baker f., Mozambique, *M.F. De Carvalho* 750 (BR, BR0000016997226); *Khaya nyasica* Stapf ex Baker f., Zimbabwe, *B. Goldsmith* 47/65 (BR, BR0000016996953); *Khaya nyasica* Stapf ex Baker f., Zimbabwe, *H. Wild* 2211 (BR, BR0000016996946); *Khaya nyasica* Stapf ex Baker f., Zambia, *J. Burt-Davy* 20886 (BR, BR0000016997103); *Khaya nyasica* Stapf ex Baker f., Zimbabwe, *B. Goldsmith* 47/65 (BR, BR0000016996953); *Khaya nyasica* Stapf ex Baker f., Mozambique, *H. Faulkner* 17 (BR, BR0000016997110); *Khaya sp1*, D.R.C., *J. Louis* 9379 (BR, BR0000016995796); *Khaya sp1*, D.R.C., *E. Maudoux* 1281 (BR, BR0000016995901); *Khaya sp1*, D.R.C., *E. Maudoux* 1282 (BR, BR0000016995918); *Khaya sp1*, D.R.C., *E. Maudoux* 3241 (BR, BR0000016995949); *Khaya sp1*, D.R.C., *F. Szafranski* 1592 (BR, BR0000016996076); *Khaya sp1*, D.R.C., *E. Maudoux* 1279 (BR, BR0000016995895); *Khaya sp1*, D.R.C., *J. Louis* 9210 (P, P00453249); *Khaya sp1*, D.R.C., *G.C.C. Gilbert* 194 (P, P00453223bis); *Khaya sp1*, D.R.C., *D. Kombi* 03/A (P, P02286983); *Khaya sp1*, D.R.C., *D. Kombi* 01/A (P, P00790069); *Khaya sp2*, D.R.C., *A. Thiebaud* 567 (BR, BR0000016997363); *Khaya sp2*, D.R.C., *H. Heine* U44 (BR, BR0000016997370); *Khaya sp2*, D.R.C., *G.F. De Witte* 4104 (BR, BR0000016997387); *Khaya sp2*, D.R.C., *De Troyer* 142 (BR, BR0000008388933); *Khaya sp2*, D.R.C., *W. Robyns* 3595 (BR, BR0000016997493); *Khaya sp2*, Zambia, *D.B. Fanshawe* 3632 (BR, BR0000016997073); *Khaya sp2*, D.R.C., *F. Malaisse* 4569 (BR, BR0000016997394); *Khaya sp2*, Zambia, *L.T. Wigg* 470 (BR, BR0000016996960); *Khaya sp2*, D.R.C., *R. Wechuysen* 770 (BR, BR0000016998117); *Khaya sp2*, D.R.C., *M. Strel* 187 (BR, BR0000016997523); *Khaya sp2*, D.R.C., *F. Ritschard* 1452 (BR, BR0000016997431); *Khaya sp2*, D.R.C., *F. Ritschard* 1733 (BR, BR0000016997462); *Khaya sp2*, D.R.C., *Jean-Jacques Symoens* 14100 (BR, BR0000016997547); *Khaya sp2*, D.R.C., *Symoens J.* 8803 (BR, BR0000016997530); *Khaya sp2*, D.R.C., *F. Malaisse* 13246 (BR, BR0000016997417); *Khaya sp2*, D.R.C., *F. Ritschard* 1734 (BR, BR0000016997455); *Khaya sp2*, D.R.C., *N. Thomas* 1247 (BR, BR0000016997561); *Khaya sp2*, Malawi, *J. Pawek* 13252 (BR, BR0000016997165).