Molecular Phylogenetics and Evolution 62 (2012) 275–285

Contents lists available at SciVerse ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Testing species delimitation in sympatric species complexes: The case of an African tropical tree, *Carapa* spp. (Meliaceae)

Jérôme Duminil [a,b,*], David Kenfack [c], Vincenzo Viscosi [d], Laurent Grumiau [a], Olivier J. Hardy [a]

[a] Service Evolution Biologique et Ecologie, CP160/12, Faculté des Sciences, Université Libre de Bruxelles, 50 Av. F. Roosevelt, 1050 Brussels, Belgium
[b] Laboratoire de Foresterie des Régions Tropicales et Subtropicales, Unité de Gestion des Ressources Forestières et des Milieux Naturels, Gembloux Agro-Bio Tech., Université de Liège, Passage des Déportés, 2, 5030 Gembloux, Belgium
[c] Center for Tropical Forest Science, Arnold Arboretum of the Harvard University, 1300 Centre Street, Boston, MA 02131, USA
[d] Museo Erbario del Molise (MEM), Dipartimento S.T.A.T., Università Degli Studi del Molise, C. da Fonte Lappone, IT-86090 Pesche (IS), Italy

ARTICLE INFO

Article history:
Received 5 April 2011
Revised 20 September 2011
Accepted 28 September 2011
Available online 14 October 2011

Keywords:
Species boundary
Hybridization
Bayesian assignment
Rainforest
Africa
Speciation

ABSTRACT

Plant species delimitation within tropical ecosystems is often difficult because of the lack of diagnostic morphological characters that are clearly visible. The development of an integrated approach, which utilizes several different types of markers (both morphological and molecular), would be extremely useful in this context. Here we have addressed species delimitation of sympatric tropical tree species that belong to *Carapa* spp. (Meliaceae) in Central Africa. We adopted a population genetics approach, sampling numerous individuals from three locations where sympatric *Carapa* species are known to exist. Comparisons between morphological markers (the presence or absence of characters, leaf-shape traits) and molecular markers (chloroplast sequences, ribosomal internal transcribed spacer region (ITS) sequences, and nuclear microsatellites) demonstrated the following: (i) a strong correlation between morphological and nuclear markers; (ii) despite substantial polymorphism, the inability of chloroplast DNA to discriminate between species, suggesting that cytoplasmic markers represent ineffective DNA barcodes; (iii) lineage sorting effects when using ITS sequences; and (iv) a complex evolutionary history within the genus *Carapa*, which includes frequent inter-specific gene flow. Our results support the use of a population genetics approach, based on ultra-polymorphic markers, to address species delimitation within complex taxonomic groups.

1. Introduction

Species identification and delimitation are fundamentally important within the fields of biology, biogeography, ecology, and conservation. In the tropics, species identification can, however, be problematic because of the lack of diagnostic morphological characters that are clearly visible. Moreover, for the vast majority of tropical organisms, there are no DNA data. As such, it is difficult to test species boundaries using molecular polymorphism information in a statistically rigorous framework. Addressing species delimitation within tropical ecosystems is nonetheless of great practical importance for the conservation of these biodiversity hotspots (Agapow et al., 2004).

The causes of species delimitation difficulties in some tropical taxa groups are poorly known, and many evolutionary processes can be invoked to explain this problem, including recent phylogenetic divergence, introgression, high phenotypic plasticity, ongoing differentiation of varieties, and partial barriers to gene flow between ecotypes. These processes are usually considered when attempting to define a species complex that do present clear morphological differences or not (cryptic species complexes, which often result from incomplete reproductive isolation). Disentangling the roles that these evolutionary processes play in confounding species delimitation is particularly challenging. It is generally agreed that one should examine a set of morphological and molecular markers and analyze these data using complementary statistical methods. In this way, an objective and rigorous overview of a species' evolutionary history can be achieved, while avoiding the inherent limitations associated with using a single locus or character and a single method of analysis. By accumulating a diverse set of morphological and molecular data and applying a multidisciplinary approach, we have addressed species delimitation of a difficult tree genus from the mahogany family, *Carapa*, in tropical Africa.

Morphological markers are the main characters used for plant species delimitation, as phenotypic similarity has been the criterion used historically by taxonomists to group individuals into species. Many morphological characters can be used to this end (Duminil and Di Michele, 2009). These types of markers can, however, fail to discriminate between morphologically similar species (e.g., cryptic

* Corresponding author at: Service Evolution Biologique et Ecologie, CP160/12, Faculté des Sciences, Université Libre de Bruxelles, 50 Av. F. Roosevelt, 1050 Brussels, Belgium. Fax: +32 2 650 24 45.
E-mail address: jduminil@ulb.ac.be (J. Duminil).

1055-7903/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2011.09.020

species, which can be distinguished based only on molecular markers; Bond et al., 2001; Whittall et al., 2004). Morphological analyses can also improperly subdivide species through inaccurate interpretation of natural phenotypic diversity over their distribution area (e.g., ecotypes, and varieties; Duminil and Di Michele, 2009). Alternatives to traditional morphological characters are geometric morphometrics, which allow for quantitative analyses of shape as "size-free" properties of a biological structure (Neto et al., 2006). Thus, the analysis of the geometric properties of landmark points can be used to study the variation of a biological organ (i.e., landmark-based methods imply true homology). This type of approach would be particularly helpful in cases of conflict among morphological characters, biased sampling of morphological traits, or cryptic ecological speciation (Lexer et al., 2009). Interestingly, this method has also been applied to cluster intermediate individuals recognized as hybrids. This method, therefore, can provide information on potential inter-species gene flow (Viscosi et al., 2009).

Molecular markers provide insights into the genealogical descent of lineages and represent alternative and powerful tools for species delimitation. Ideally, each species would be identified by a single, unique DNA-tag (a DNA barcode), defined as "short segments of DNA that can be used to uniquely identify an unknown specimen to species" (Sass et al., 2007). Species delimitation could then be assessed by the presence of a DNA barcoding gap between intra- and inter-specific genetic variation (Lahaye et al., 2008) or by the use of phylogenetic methods, where monophyly and branch support represent the diagnostic criteria for species delimitation (Fazekas et al., 2008). Although DNA barcoding represents a powerful approach (Hollingsworth et al., 2009; Lahaye et al., 2008), it is ineffective under a range of circumstances. These include the lack of DNA barcode divergence among closely related species (Muellner et al., 2011; Seberg and Petersen, 2009), inadequate taxonomy, incomplete lineage sorting (i.e., a gene tree does not always reflect the species tree) (Devos et al., 2003), hybridization (Dobes et al., 2004), and cytoplasmic capture (Tsitrone et al., 2003).

Relying on DNA-based assignment using multilocus genotype data in a population genetics framework appears relevant for species delimitation (e.g., using Bayesian clustering methods; Duminil et al., 2006; Lepais et al., 2009). Ultra-polymorphic markers, such as nuclear microsatellites, are particularly useful when studying closely related species because (i) they are more informative than traditional candidate barcodes, as they present a higher genetic diversity, and (ii) they are distributed across multiple loci, thereby providing a more accurate representation of the entire genome and limiting the risk of incomplete lineage sorting, which can confound single-locus analyses. Under a population genetics framework, species boundaries can be identified using model-based clustering methods for assigning sympatric individuals to gene pools according to genotype data (e.g., Pritchard et al., 2000). These algorithms delimit gene pools in such a way that each pool roughly approximates a panmictic population (e.g., Hardy–Weinberg and linkage equilibria). Distinct gene pools must also be genetically differentiated, reflecting barriers to gene flow. Interestingly, these non-tree based approaches can reveal inter-specific gene flow. Importantly, the detection of distinct gene pools that have allopatric or parapatric patterns of distribution does not necessarily indicate distinct species. These gene pools may instead represent well-differentiated populations of the same species. However, when distinct gene pools are found in sympatry (Koffi et al., 2010), it is a strong argument for the presence of distinct species, following the Biological Species Concept (BSC; i.e., separate reproductive entities).

In Africa, the genus *Carapa* (Meliaceae) has historically been represented by a single species (*C. procera* D.C.), despite substantial morphological variation. Attempts to delineate species of African *Carapa* clearly depend on author bias and illustrate the opposition between 'lumpers' and 'splitters' (Candolle, 1878; Harms, 1917; Noamesi, 1958; Staner, 1941; Styles and White, 1991). According to recent revisions, which integrate morphological and molecular markers (ITS sequences), the genus *Carapa* contains ~27 species, 16 of which are from tropical Africa (Kenfack, 2008, 2011).

We have investigated species boundaries in African *Carapa* using samples from three sites where distinct morphospecies are found in sympatry. A morphospecies is defined in this study as a taxonomic entity with particular diagnostic morphological characters (that enable its identification and classification), but for which reproductive isolation has not been demonstrated. Combining different types of markers (morphology, nuclear sequences, chloroplast sequences, and nuclear microsatellites) and analytic methods from different disciplines (geometric morphometrics, phylogenetics, and population genetics), we have assessed the degree of inter-individual morphological and genetic polymorphism within localities to address a number of questions. (i) Do the various markers and methods identify the same species boundaries? (ii) Are the species delimitations presented in Kenfack (2011), which are based on morphological and ITS markers, supported by our integrated analysis? (iii) Do *Carapa* morphospecies correspond to a unique species that is characterized by high morphological polymorphism or to differentiated species that match the BSC? (iv) In what way does our combined approach help clarify the evolutionary history of this taxa?

2. Materials and methods

2.1. Biological model description

The genus *Carapa* (from the mahogany family Meliaceae) is composed of both small and large trees, up to 35 m tall. The genus was first described in the Neotropics (Aublet, 1775). *Carapa* species are monoecious, with seeds generally dispersed by animals but also potentially by water (Forget, 1996; Forget et al., 1999). Flowering and fruiting occur throughout the year; however, the majority of species flower from January to May and produce fruit from June to September.

2.2. Sampling strategy

Samples from 130 individual trees were collected from three locations within the Lower Guinea forests of Southern Cameroon: (i) the Dja Forest Reserve (Dja; 3.336°N–12.735°E); (ii) the Ngovayang forest within the Bipindi-Akom region (Bipindi; 3.179°N–10.512°E); and (iii) the Korup National Park (Korup; 5.053°N–8.864°E; Fig. 1). These locations were chosen because they contain different morphospecies of *Carapa*, according to local botanists or our own observations. These sites are in evergreen forests with either a four-season equatorial climate (Dja and Bipindi) or a two-season pseudo-equatorial climate (Korup). Annual precipitation levels are 1650 mm (Dja), 2000 mm (Bipindi), and 5000 mm (Korup). In Dja, samples were collected from 40 individual trees, mainly found within ~500 ha of lowland forest (400–700 m of elevation; Fig. 1). In Bipindi, samples were collected from 33 individuals within ~200 ha (300–900 m of elevation; Fig. 1). In Korup, samples were collected from 57 trees within an ~600-ha region (Fig. 1). Thirty-four of the trees within Korup came from a 50-ha permanent plot established by the Centre for Tropical Forest Science (Kenfack et al., 2007; Thomas et al., 2003). For each of the 130 individuals included in the study, cambium or leaf tissue samples were preserved in silica gel for molecular studies. In addition, voucher-herbarium specimens were collected from 92 trees for further morphological identification. For the majority of trees, no flowers or fruits could be collected (Table 1). Total DNA was isolated using the NucleoSpin® plant kit from Macherey–Nagel.

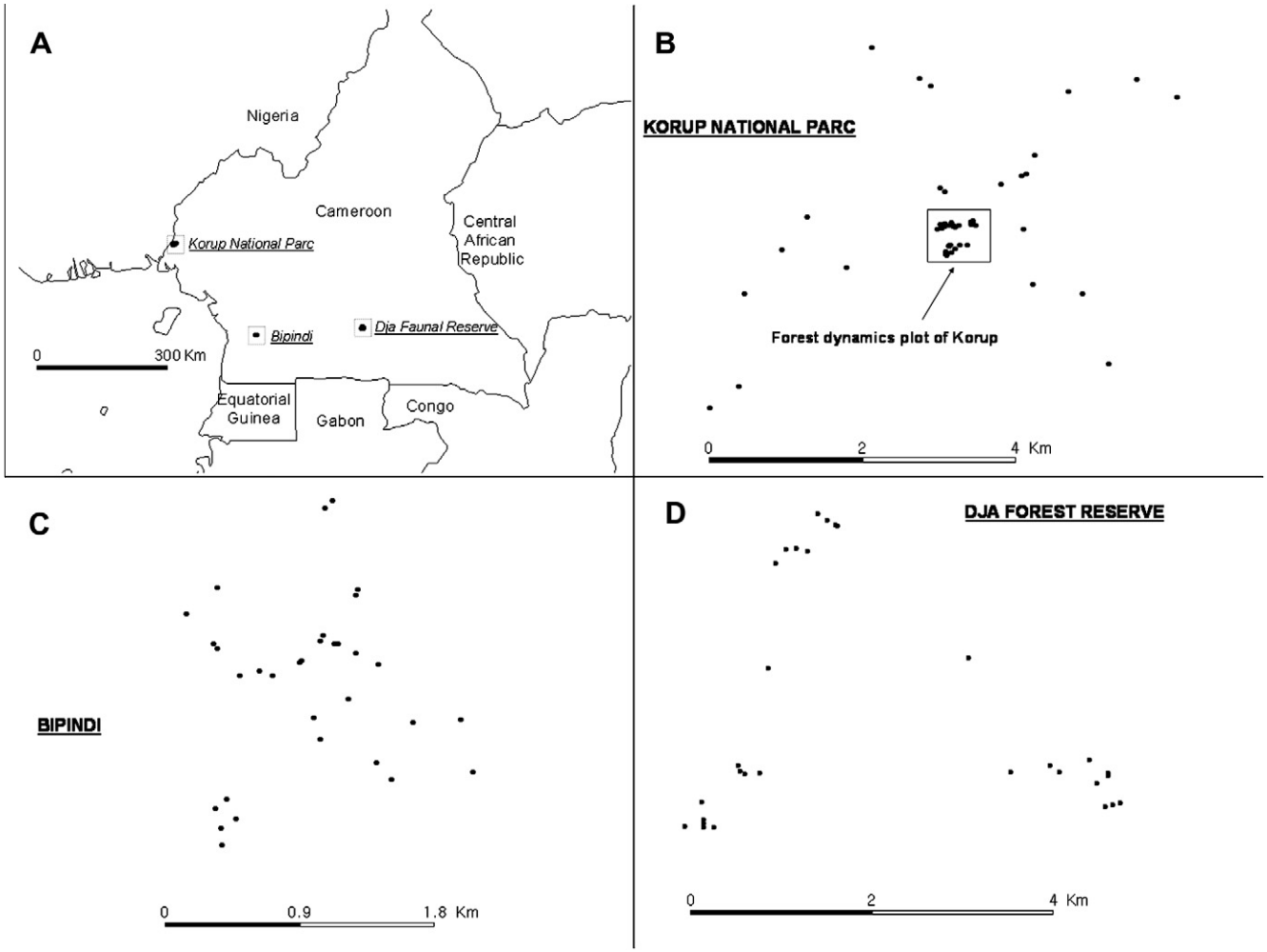

Fig. 1. Localization and repartition of the individuals in Cameroon: (A) The three localities where *Carapa* sp. individuals were sampled. (B) Spatial distribution of the individuals in Korup National Park. (C) Spatial distribution of the individuals in Bipindi. (D) Spatial distribution of the individuals in Dja Forest Reserve.

2.3. Taxonomic identification based on morphological traits

Voucher identification followed a recent description of *Carapa* species (Kenfack, 2008). Hereafter, to avoid confusion among different species concepts, the species recognized in Kenfack (2008) will be referred to as morphospecies. Because *Carapa* trees are large, most of the vouchers consisted of immature leaves, often collected from juveniles or low shoots. As a result, these samples did not contain flowers or fruits, which exacerbated the identification difficulties. Four main vegetative morphological characters were used to separate the specimens into morphospecies: (i) the number of pairs of leaflets, (ii) the relative length of the petiole, (iii) the shape of the leaflets, and (iv) the presence or absence of indumentum on the midrib, the latter of which is important to separate *C. dinklagei* from the other taxa (Kenfack, 2011). When voucher collection was not possible, tentative identification was performed in the field using dry leaves found on the ground. In these cases, the identification was less reliable, and posterior re-identification was almost impossible.

2.4. Geometric morphometrics: semi-landmark method

The semi-landmark method was tested on individuals from Korup, which was the only location where vouchers could be assigned to a particular morphospecies with high confidence. The test was done using 30 leaflets from nine trees identified as *C. dinklagei* and 93 leaflets from 31 trees identified as *C. parviflora* (Supplementary material 1). The position of the leaflets on the respective leaves was not taken into account because of the small sample size. Closely spaced points were digitized around the periphery of the leaflet blades by manual curve tracing using the TpsDig program (Rohlf, 2001). Each trace began at the intersection between the leaflet blade and the petiolule and ended at the analogous point on the trailing edge of the leaflet. A set of 100 equally-spaced points was used to represent the information contained in the homologous curves traced around the periphery of each leaf blade. Points obtained in this way were subsequently analyzed using the semi-landmark method (Navarro et al., 2004; Sheets et al., 2006). Semi-landmarks were first submitted to a generalized Procrustes analysis to remove all differences in location, orientation, and scale. Then, semi-landmark alignment was computed by perpendicular projection, which aligned semi-landmarks along the curves using the SemiLand6 program (Sheets et al., 2004).

Using a size-free approach, we performed two types of statistical analyses. The first analysis (*K*-means) grouped individuals according to leaflet morphology differences that were not visible by eye. This analysis did not take into account each sample's morphospecies assignment. The second analysis (Mulitvariate Analysis of variance, MANOVA) incorporated the morphospecies assignment data and tested whether shape differences could be detected between morphospecies (as supported by Bayesian clustering).

Table 1
List of the individuals and their characteristics.

Individual id	Loc[a]	Morphotype[b]	Nr[c]	Cp[d]	Longitude[e]	Latitude[e]	Herbarium[f]
OH1245*	B	*C. macrantha*	nr-A(1)	H8	3.1755	10.5394	Yes
OH1277*	B	*C. macrantha*	nr-A(1)	H4	3.1749	10.5455	Yes
OH1283	B	*C.* sp	nr-A(1)	H4	3.1780	10.5417	Yes
OH1302	B	*C.* sp	nr-A(1)	H4	3.1770	10.5358	Yes
OH1304	B	*C.* sp	nr-A(1)	H4	3.1732	10.5299	Yes
OH1311*	B	*C. macrantha*	nr-A(1)	H4	3.1861	10.5381	Yes
OH1312	B	*C. macrantha*	nr-A(1)	H4	3.1819	10.5346	Yes
OH1315*	B	*C. parviflora*	nr-A(2)	H4	3.1830	10.5291	Yes
OH1316	B	*C.* sp	nr-A(2)	H4	3.1866	10.5293	Yes
OH1317	B	*C.* sp	nr-A(2)	H4	3.1849	10.5273	Yes
OH1322*	B	*C.* sp	nr-A(2)	H4	3.1921	10.5366	No
OH1323*	B	*C. parviflora*	nr-A(2)	H4	3.1916	10.5361	Yes
OH1324	B	*C.* sp	nr-A(1)	H4	3.1818	10.5345	Yes
OH1325*	B	*C. macrantha*	nr-A(1)	H4	3.1810	10.5328	Yes
OH1334	B	*C.* sp	nr-A(1)	H4	3.1795	10.5376	Yes
OH1335	B	*C.* sp	nr-A(1)	H4	3.1726	10.5292	Yes
OH1336	B	*C.* sp	nr-A(1)	H4	3.1713	10.5295	Yes
OH1337	B	*C.* sp	nr-A(1)	H4	3.1703	10.5296	Yes
OH1338*	B	*C. cf. hygrophila*	nr-A(1)	H4	3.1719	10.5305	Yes
OH1339	B	*C.* sp	nr-A(1)	H4	3.1824	10.5381	Yes
OH1342*	B	*C. macrantha*	nr-A(1)	H4	3.1835	10.536	Yes
OH1343	B	*C.* sp	nr-A(1)	H4	3.1817	10.5395	Yes
OH1246*	B	*C. dinklagei*	nr-C	H10	3.1744	10.5403	Yes
OH1278	B	*C. dinklagei*	nr-C	H10	3.1782	10.5447	Yes
OH1307	B	*C. dinklagei*	nr-C	H10	3.1832	10.5358	Yes
OH1313*	B	*C. dinklagei*	nr-C	H1	3.1813	10.532	Yes
OH1326*	B	*C. dinklagei*	nr-C	H11	3.1810	10.5307	Yes
OH1333	B	*C. dinklagei*	nr-C	H1	3.1783	10.5354	Yes
OH1341*	B	*C. dinklagei*	nr-C	H1	3.1830	10.5367	Yes
OH1306	B	*C. dinklagei*	na	H10	3.1830	10.5369	Yes
OH1314*	B	*C. cf. hygrophila*	na	H4	3.1827	10.5293	Yes
OH1318*	B	*C. macrantha*	na	H7	3.1832	10.5358	Yes
OH1319*	B	*C. cf. hygrophila*	na	H5	3.1864	10.5382	Yes
LD0003	D	*C.* sp	nr-A	H2	3.3258	12.7149	No
LD0007	D	*C.* sp	nr-A	H2	3.3262	12.7149	Yes
LD0012	D	*C.* sp	nr-A	H2	3.3262	12.7149	No
LD0023*	D	*C. cf. macrantha*	nr-A	H2	3.3262	12.7149	Yes
LD0026*	D	*C. cf. parviflora*	nr-A	H2	3.3258	12.7160	Yes
LD0039*	D	*C.* sp	nr-A	H2	3.3284	12.7147	Yes
LD0060	D	*C.* sp	nr-A	H2	3.3313	12.7192	No
LD0069	D	*C.* sp	nr-A	H2	3.3317	12.7187	Yes
LD0070	D	*C.* sp	nr-A	H2	3.3322	12.7185	No
LD0074	D	*C.* sp	nr-A	H2	3.3315	12.7208	No
LD0090	D	*C.* sp	nr-A	H2	3.3424	12.7216	No
LD0098	D	*C.* sp	nr-A	H2	3.3424	12.7216	No
LD0113*	D	*C.* sp	nr-A	H3	3.3534	12.7224	No
LD0117*	D	*C. cf. palustris*	nr-A	H9	3.3548	12.7235	Yes
LD0119*	D	*C. cf. parviflora*	nr-A	H2	3.3549	12.7246	Yes
LD0120*	D	*C. cf. palustris*	nr-A	H2	3.3547	12.7258	Yes
LD0124	D	*C. cf. palustris*	nr-A	H2	3.3573	12.7289	Yes
LD0129*	D	*C. cf. palustris*	nr-A	H2	3.3579	12.7278	Yes
LD0136	D	*C.* sp	nr-A	H2	3.3586	12.7268	No
LD0188	D	*C.* sp	nr-A	H2	3.3282	12.7576	No
LD0215	D	*C.* sp	nr-A	H2	3.3304	12.7560	No
LD0221	D	*C.* sp	nr-A	H2	3.3304	12.7560	No
LD0233*	D	*C. cf. parviflora*	nr-A	H2	3.3329	12.7552	Yes
LD0235*	D	*C. cf. palustris*	nr-A	H2	3.3329	12.7552	Yes
LD0240	D	*C.* sp	nr-A	H2	3.3316	12.7470	No
LD0242*	D	*C. cf. palustris*	nr-A	H2	3.3316	12.7520	Yes
LD0243	D	*C.* sp	nr-A	H2	3.3315	12.7571	No
LD0245*	D	*C. cf. palustris*	nr-A	H2	3.3312	12.7571	Yes
LD0024*	D	*C. cf. palustris*	nr-A	H2	3.3266	12.7149	Yes
LD0154	D	*C.* sp	nr-A	H2	3.3435	12.7426	No
LD0163	D	*C.* sp	nr-A	H2	3.3435	12.7426	No
LD0203	D	*C.* sp	nr-A	H2	3.3280	12.7568	No
LD0121*	D	*C. cf. palustris*	nr-A	H2	3.3574	12.7287	Yes
LD0168	D	*C.* sp	nr-A	?	3.3283	12.7584	No
LD0194	D	*C.* sp	nr-A	H2	3.3282	12.7576	No
LD0192*	D	*C. cf. angustifolia*	nr-B	H1	3.3282	12.7576	Yes
LD0238*	D	*C.* sp	nr-B	?	3.3322	12.7511	No
LD0028*	D	*C.* sp	na	H2	3.3259	12.7130	No
LD0111	D	*C. cf. parviflora*	na	H1	3.3534	12.7224	Yes
LD0202*	D	*C.* sp	na	H2	3.3280	12.7568	No
32–1040	K	*C. parviflora*	nr-B	H6	5.0619	8.8546	No

Table 1 (*continued*)

Individual id	Loc[a]	Morphotype[b]	Nr[c]	Cp[d]	Longitude[e]	Latitude[e]	Herbarium[f]
JD0015*	K	*C. parviflora*	nr-B	H6	5.0614	8.8541	Yes
JD0016*	K	*C. parviflora*	nr-B	H6	5.0615	8.8535	Yes
JD0018*	K	*C. parviflora*	nr-B	H6	5.0612	8.8535	Yes
JD0019*	K	*C. parviflora*	nr-B	H6	5.0611	8.8536	Yes
JD0025*	K	*C. parviflora*	nr-B	H6	5.0650	8.8564	Yes
JD0027*	K	*C. parviflora*	nr-B	H6	5.0647	8.8530	Yes
JD0028	K	*C. parviflora*	nr-B	H6	5.0647	8.8528	Yes
JD0029*	K	*C. parviflora*	nr-B	H6	5.0646	8.8537	Yes
JD0031*	K	*C. parviflora*	nr-B	H6	5.0646	8.8539	Yes
JD0033*	K	*C. parviflora*	nr-B	H6	5.0647	8.8564	Yes
JD0037	K	*C. parviflora*	nr-B	H6	5.0644	8.8533	Yes
JD0038	K	*C. parviflora*	nr-B	H6	5.0646	8.8538	Yes
JD0039*	K	*C. parviflora*	nr-B	H6	5.0642	8.8530	Yes
JD0040*	K	*C. parviflora*	nr-B	H6	5.0646	8.8569	Yes
JD0042BIS	K	*C. parviflora*	nr-B	H6	5.0643	8.8546	Yes
JD0051*	K	*C. parviflora*	nr-B	H6	5.0641	8.8524	Yes
JD0054*	K	*C. parviflora*	nr-B	H6	5.0596	8.8418	Yes
JD0070	K	*C. parviflora*	nr-B	H6	5.0566	8.8299	Yes
JD0075*	K	*C. parviflora*	nr-B	H6	5.0457	8.8292	Yes
JD0078*	K	*C. parviflora*	nr-B	H6	5.0432	8.8258	Yes
JD0090	K	*C. parviflora*	nr-B	H6	5.0623	8.856	Yes
JD0091	K	*C. parviflora*	nr-B	H6	5.0623	8.8551	Yes
JD0095	K	*C. parviflora*	nr-B	H6	5.0622	8.8539	Yes
JD0096	K	*C. parviflora*	nr-B	H6	5.0623	8.8540	Yes
JD0099	K	*C. parviflora*	nr-B	H6	5.0685	8.8533	Yes
JD0101	K	*C. parviflora*	nr-B	H6	5.0689	8.8528	Yes
JD0123	K	*C. parviflora*	nr-B	H6	5.0706	8.8629	Yes
JD0124*	K	*C. parviflora*	nr-B	H6	5.0728	8.8639	Yes
JD0130*	K	*C. parviflora*	nr-B	H6	5.0803	8.8678	Yes
JD0134	K	*C. parviflora*	nr-B	H6	5.0817	8.8758	Yes
JD0135	K	*C. parviflora*	nr-B	H6	5.0796	8.8805	Yes
OH741	K	*C. parviflora*	nr-B	H6	5.0641	8.8626	No
OH742	K	*C. parviflora*	nr-B	H6	5.0809	8.8517	No
OH743	K	*C. parviflora*	nr-B	H6	5.0656	8.8372	No
OH744	K	*C. parviflora*	nr-B	H6	5.0854	8.8448	No
OH745	K	*C. parviflora*	nr-B	H6	5.0818	8.8504	No
OH746	K	*C. parviflora*	nr-B	H6	5.0566	8.8695	No
OH747	K	*C. parviflora*	nr-B	H6	5.0484	8.8725	No
OH814	K	*C. parviflora*	nr-B	H6	5.0694	8.8600	Yes
16–2256	K	*C. dinklagei*	nr-C	H6	5.0646	8.8544	No
32–1169	K	*C. dinklagei*	nr-C	H6	5.0619	8.8546	No
JD0007*	K	*C. dinklagei*	nr-C	H6	5.0651	8.8566	Yes
JD0026	K	*C. dinklagei*	nr-C	H6	5.0649	8.8541	Yes
JD0032	K	*C. dinklagei*	nr-C	H6	5.0647	8.8567	Yes
JD0034	K	*C. dinklagei*	nr-C	H6	5.0648	8.8564	Yes
JD0036	K	*C. dinklagei*	nr-C	H6	5.0646	8.8550	Yes
JD0077	K	*C. dinklagei*	nr-C	H6	5.0432	8.8258	Yes
JD0092	K	*C. dinklagei*	nr-C	H6	5.0623	8.8539	Yes
JD0093	K	*C. dinklagei*	nr-C	H6	5.0623	8.8540	Yes
JD0094	K	*C. dinklagei*	nr-C	H6	5.0622	8.8538	Yes
JD0122	K	*C. dinklagei*	nr-C	H6	5.0704	8.8623	No
JD0010	K	*C. parviflora*	nr-C	H6	5.0645	8.8543	No
JD0013*	K	*C. parviflora*	nr-C	H6	5.0615	8.8535	No
JD0006*	K	*C.* sp	nr-C	H6	5.0577	8.8637	Yes
JD0026Bis	K	*C.* sp	nr-C	H6	5.0648	8.8533	No
OH0684*	K	*C.* sp	nr-C	H6	5.0618	8.8342	No

[a] Locality of origin (B: Bipindi, D: Dja, K: Korup).
[b] Assignation of the individuals to species using morphological characters.
[c] Results of the assignation of the species to a gene pool nr-A, nr-B or nr-C using STRUCTURE software (For Bipindi, the number in brackets indicates individual assignation to cluster nr-A1 or nr-A2 with analyses done at the intra-locality level).
[d] Chloroplastic haplotype based on the polymorphism observed at the *trn*C-*pet*N1R and the *trn*H–*trn*K intergenic spacers.
[e] Latitude and longitude in decimal degrees.
[f] Availability of a herbarium specimen for taxonomic identification.
* Individuals used for ITS analyses.

The dataset of aligned shape coordinates was subjected to *K*-means clustering using three numbers of clusters ($K = 2$, 3, or 4), which generated specimen groups based on their leaf shape traits. We then tested by MANOVA the differences between three types of specimen groupings: (i) *K*-mean groups, (ii) morphospecies, and (iii) groups based on genetic markers (see below). Subsequently, using information derived from individual assignment to species using genetic markers, the semi-landmark approach was used to test significant differences between morphospecies by one-way MANOVA, with *F*-tests at 999 permutations. The ManovaBoard program was used for this analysis (Sheets et al., 2004).

2.5. Chloroplast DNA polymorphism

For all samples, two intergenic regions (*trn*C-*pet*N1R and *trn*H–*trn*K) from chloroplast DNA (cpDNA)were amplified and sequenced

using universal primers for the full set of individuals (Demesure et al., 1995; Lee and Wen, 2004). For the *trn*C-*pet*N1R region, PCR reactions included 2 μL of template DNA (10–100 ng), 0.1 μL *Taq* polymerase (Qiagen), 2 μL PCR buffer, 1 μL $MgCl_2$ (25 mM), 0.5 μL dNTPs (10 μM), 0.25 μL of each primer (10 μM), and 18.9 μL H_2O (total volume, 25 μL). The PCR cycling profile for the *trn*C-*pet*N1R region included 3 min at 94 °C, followed by 35 cycles of 30 s at 94 °C, 30 s at 56 °C, and 1 min at 72 °C. Reactions were then incubated for 5 min at 72 °C. For the *trn*H–*trn*K region, PCR reactions included 2 μL of template DNA (10–100 ng), 0.1 μL *Taq* polymerase (Phusion high-fidelity DNA polymerase from Finnzymes), 5 μL 5× Phusion HF buffer, 0.5 μL dNTP**[s]** (10 μM), 0.5 μL of each primer (10 μM), and 17.8 μL H_2O (total volume, 25 μL). The PCR cycling profile for the *trn*H–*trn*K region included 30 s at 98 °C, followed by 35 cycles of 10 s at 98 °C, 30 s at 60 °C, and 1 min at 72 °C. Reactions were then incubated for 5 min at 72 °C. Sequences were resolved on a 3100 Genetic Analyzer (Applied Biosystems).

2.6. ITS ribosomal DNA polymorphism

The 18S rDNA-26S rDNA (which comprises ITS1, 5.8S rDNA, and ITS2) was amplified using the primers ITSi and ITS4 (Urbatschl et al., 2000; White et al., 1990). PCR reactions included 1.5 μL of template DNA (10–100 ng), 0.25 μL *Taq* polymerase (Qiagen), 2.5 μL PCR buffer, 2.5 μL $MgCl_2$ (25 mM), 2.5 μL Q solution, 0.4 μL dNTPs (10 μM), 0.3 μL of each primer (10 μM), and 14.75 μL H_2O (total volume, 25 μL). The PCR cycling profile for the 18S rDNA-26S rDNA fragment included 3 min at 94 °C, followed by 40 cycles of 30 s at 94 °C, 40 s at 54 °C (−0.1 °C for the T_m at each subsequent PCR cycle), and 50 s at 72 °C. Reactions were then incubated for 10 min at 72 °C. A subset of 55 individuals was used for these analyses, which included at least one representative from each nucleocytoplasmic association (see Table 1).

2.7. Phylogenetic reconstruction

A median joining network (Bandelt et al., 1999) of cpDNA haplotypes was reconstructed using NETWORK 4.1.5.6 (Fluxus Technology Ltd.). ITS sequences of 24 *Carapa* species were acquired from GenBank, corresponding to the work of Kenfack (2011). ITS sequences from three species of *Khaya* (a sister species of *Carapa*) were used as the outgroup for the phylogenetic analyses. A phylogenetic tree based on parsimony analyses was constructed using PAUP* (Swofford, 2002). Gaps were treated as missing characters. A heuristic search was done using the tree bisection and reconnection (TBR) method of branch swapping, with retention of a maximum of 10,000 trees. Branch support values were evaluated by bootstrap analysis, which used a heuristic search with TBR branch swapping. Moreover, to compare with the parsimony method, Bayesian analyses were performed from random starting trees using Mr. BAYES 3.1.2 software (Ronquist and Huelsenbeck, 2003). The (GTR + I + G) model of nucleotide substitution was used, as it was determined to be the best model according to the Akaike Information Criterion in the jModelTest 0.1.1 software (Posada, 2008). Gaps were treated as missing characters. The data were run for 10,000,000 generations, with a tree sampled every 1000 generations. Chain stationarity was checked using both the average standard deviation of split frequencies (which should be <0.01 for a sufficient number of generations) and the Potential Scale Reduction Factor (which should approach 1 as runs converge) (Gelman and Rubin, 1992). Chain stationarity was achieved after 20,000 generations (burn-in). A 50% majority rule consensus tree was constructed from the remaining trees.

2.8. Microsatellite markers

Six microsatellite markers were amplified from the entire set of samples: Cg1I, Cg7, Cg16, Cg17, Cg06, and Cg6. These markers were originally designed for the South American species *Carapa guianensis* (Dayanandan et al., 1999; Vinson et al., 2005) but are also useful when analyzing Central African species of *Carapa*. Genotypes were resolved on a 3100 Genetic Analyzer. SSR fragment lengths were determined by comparison with the GeneScan 350 ROX size standard (Applied Biosystems), using the Genemapper v3.0 software (Applied Biosystems).

2.9. Population genetics analyses

To identify distinct gene pools with the microsatellite data set, we used a Bayesian clustering algorithm to create *K* groups of individuals in such a way that Hardy–Weinberg equilibrium (as well as linkage equilibrium) was maintained for as long as possible within each group. To this end, we ran the software structure (Pritchard et al., 2000) 10 times for $K = 1$ to 10 on the entire sample set with the following parameters sample acquisition site. We used the admixture model, in which the fraction of ancestry from each gene pool was estimated for each individual. The parameter of individual admixture, alpha, was chosen to be the same for all gene pools, and it was given a uniform prior. Allele frequencies were kept independent among gene pools, to avoid overestimating the number of gene pools (Falush et al., 2003). The approach of Evanno et al. (2005) was used to determine the optimal number of clusters (*K*) in our sample of individuals. The software structure was also run for each sampling site separately to test the robustness of the results. Under the admixture model, structure partitions each individual among the different gene pools. As such, we considered an individual assigned to a particular gene pool when at least 80% of its genome came from that gene pool. Otherwise, the individual was left unassigned. The SPAGeDi software (Hardy and Vekemans, 2002) was used to calculate Nei's standard genetic distance Ds. between the gene pools (Nei, 1978).

3. Results

3.1. Morphological identification

Individuals that could not be assigned with confidence to a particular morphospecies were denoted *Carapa* sp. (Table 1). The morphospecies present within the 50-ha plot of Korup had been identified previously (Thomas et al., 2003), whereas this type of information was not available for Bipindi or Dja. As such, individuals from Korup were assigned accurately to either *C. dinklagei* or *C. parviflora* (the two morphospecies found in that region) based on the presence or absence of indumentum on the midrib. In Bipindi, where more than two morphospecies are known to reside (Kenfack, 2011), leaf pubescence was used to separate *C. dinklagei* from the rest of the morphospecies. Based on leaflet shape and the tertiary venation pattern, remaining individuals (all glabrous) were assigned with confidence to either *C. macrantha* or *C. parviflora* and with less confidence to *C. hygrophila* or *C. palustris* (individuals identified with low confidence were denoted "*C. cf.*"; Table 1). In Dja, all individuals were glabrous and were assigned to one of four morphospecies, but with limited confidence at times (Table 1). Based on our morphological observations, at least five morphospecies were present within the three sampled regions. Overall, 41 individuals could not be identified based on morphology, as fresh leaves were not accessible (these were adult trees from which only cambium was recovered). Seventy-one individuals were identified with high confidence, and 18 were identified with low confidence.

3.2. Geometric morphometrics

3.2.1. K-mean clustering

According to MANOVA, the most significant grouping was obtained for $K = 2$ ($F = 54.6797$; df1 = 196, df2 = 23,716; $p < 0.001$), but these clusters were not coherent with respect to morphological identification or genetic assignment. These two clusters consisted of 59 and 64 leaflets. The first cluster grouped 19 leaflets of *C. dinklagei* and 40 leaflets of *C. parviflora*, whereas the second cluster contained 11 *C. dinklagei* and 53 *C. parviflora* leaflets. In addition, leaflets from the same individual were not always assigned to the same group. This method, therefore, did not accurately identify species.

3.2.2. One-way MANOVA

MANOVA indicated a significant difference in leaflet shape between *C. dinklagei* and *C. parviflora* ($F = 8.9299$; df1 = 196, df2 = 23,716; $p < 0.001$). Concerning total variance, however, the percentage of variability that was due to leaf shape differences between the two species was small (6.87%). The main differences between these two species, when compared with the consensus configuration, were found within basal and apical leaf regions. *C. dinklagei* presented a lanceolate leaf, with a more acute basal and apical shape than did *C. parvifolia* (Supplementary material 2). These shapes, however, are not maintained in adult leaves.

3.3. Chloroplast DNA polymorphism

Based on sequence data from the *trn*C-*pet*N1R and *trn*H–*trn*K intergenic spacers, 11 chloroplastic haplotypes were identified. A total of 16 polymorphic nucleotides were found (seven within *trn*C-*pet*N1R and nine within *trn*H–*trn*K). Sequences of the *trn*C-*pet*N1R and *trn*H–*trn*K intergenic regions were deposited in GenBank (accession numbers JN564653–JN564665). For two individuals from the Dja site (LD0168 and LD0238), one of the cpDNA fragments could not be amplified. Only haplotype H1 was shared among sample sites (Fig. 2). Bipindi had the highest amount of polymorphism (seven haplotypes), followed by Dja (four haplotypes). In contrast, the Korup group was entirely composed of a single haplotype. Haplotypes H2, H4, and H6 were frequent (17–44%), whereas all others were rare (Supplementary material 3). Interestingly, haplotypes H1 and H6 were associated with the nr-B and nr-C gene pools (Fig. 2; see below for gene pool definitions).

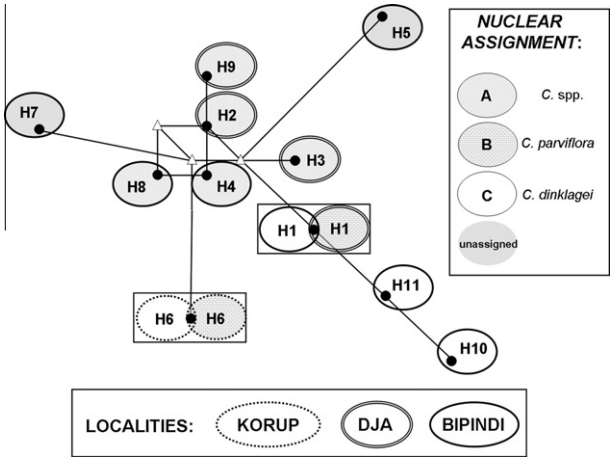

Fig. 2. CpDNA phylogenetic network and nucleocytoplasmic association. Background circle style indicates the nuclear gene pools as obtained with STRUCTURE (gene pool nr-A, nr-B, nr-C or unassigned individuals). In front of nuclear gene pools is indicated the corresponding taxonomic identification of the individuals. Circle contour style indicates the locality of origin of the haplotype. Only H1 is common to two localities: Bipindi and Dja. Both in Korup and in Bipindi cpDNA haplotypes (H6 for Korup and H1 for Bipindi) are shared between individuals assigned to gene pools nr-B and nr-C.

3.4. ITS polymorphisms

The ITS1-5.8S-ITS2 fragment was 640–642 bp long in the 55 sequences obtained. A total of 63 polymorphic sites were identified, of which 46 were parsimony informative. Sequences were deposited in GenBank (accession numbers JN564666–JN564678). All individuals from the nr-C gene pool (see below for gene pool definitions) grouped together with individuals defined as *C. dinklagei* (sequences obtained from GenBank) to form a monophyletic group (Fig. 3). All but three individuals from gene pool nr-B formed a monophyletic group. Surprisingly, none of the sequences obtained from GenBank were included in this group. Most individuals from the nr-A gene pool grouped with individuals defined as *C. grandiflora*, *C. macrantha*, or *C. batesii* (sequences obtained from GenBank). Individuals that were not assigned to a particular gene pool were grouped with individuals from gene pool nr-A or nr-B.

3.5. Assignment of individuals to gene pools and genetic distances among gene pools

Including all individuals, the highest likelihood was obtained for $K = 3$ gene pools (Supplementary material 4). These three gene pools were named nr-A, nr-B, and nr-C. Seven individuals were not assigned to any gene pool, using a probability threshold of 0.8 (Table 1). In the Bipindi group, seven individuals were assigned to nr-C, 22 to nr-A, and 4 were unassigned. In the Dja group, two individuals were assigned to nr-B, 35 were assigned to nr-A, and 3 were unassigned. In the Korup group, 17 individuals were assigned to nr-C, and 40 were assigned to nr-B. Therefore, gene pool nr-A was present in Bipindi and Dja, nr-B was present in Dja and Korup, and nr-C was present in Bipindi and Korup. Nei's genetic distances between gene pools demonstrated that nr-A and nr-B were the most closely related ($Ds_{\text{nr-A/nr-B}} = 0.78$, $Ds_{\text{nr-A/nr-C}} = 1.79$, $Ds_{\text{nr-B/nr-C}} = 1.81$). This result agreed with the topology of the phylogeny obtained using ITS sequences (Fig. 3).

Cluster analyses were also performed on individuals from each sample site (intra-site analyses). These results essentially agreed with global analyses described above, which included individuals from every site. Only one gene pool was obtained from the Dja group, whereas two gene pools that fully corresponded to those defined in the global analysis were obtained from the Korup group. Three gene pools rather than two were, however, obtained from the Bipindi group. The main difference in the Bipindi group was that gene pool nr-A was divided into two gene pools (nr-A1 and nr-A2).

3.6. Correlations between morphological taxonomy (morphospecies), nuclear gene pools, and chloroplast haplotypes

In Bipindi, there was partial agreement between morphospecies and the nuclear gene pools. *C. dinklagei* matched nr-C, whereas *C. parviflora*, *C. macrantha,* and *C. cf. hygrophila* matched nr-A. Four individuals from different morphospecies were, however, not assigned a gene pool. This correlation may have been improved when gene pools were defined based on their locality. In this case, the six *C. macrantha* individuals were assigned to nr-A1, and the two *C. parviflora* individuals were assigned to nr-A2 (the two *C. cf. hygrophila* individuals were assigned to nr-A1 or nr-A2). There was a strict association between nuclear groups and chloroplast haplotypes, as individuals from gene pool nr-A were either H4 or H8 (only one individual), and individuals from gene pool nr-C were H1, H10, or H11. Unassigned individuals were H4, H5, H7, or H10.

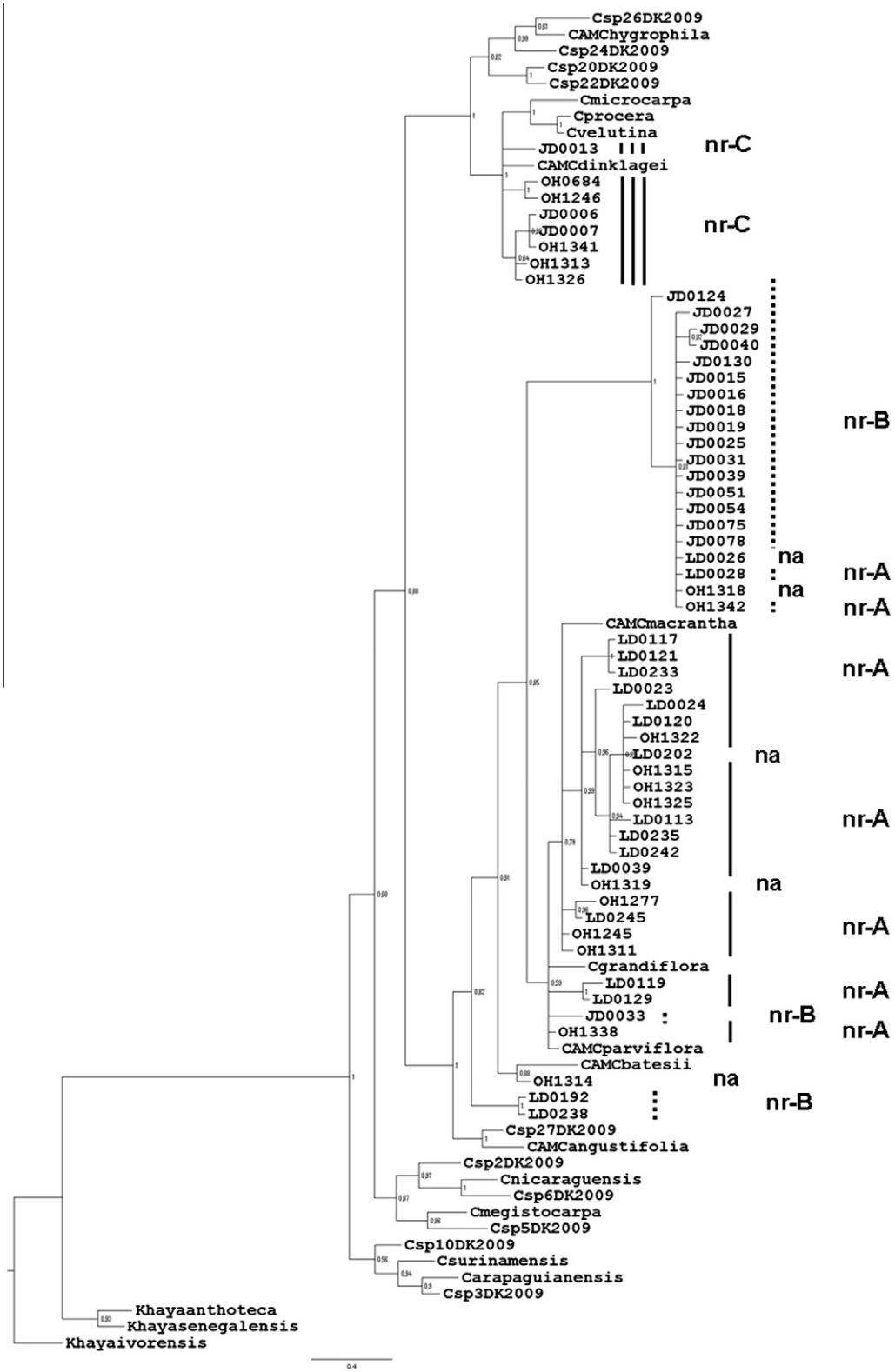

Fig. 3. ITS 50% majority rule consensus tree based on the ITS sequences. Individual assignment to nuclear gene pools nr-A, nr-B or nr-C is indicated on the right ("na" indicates "non-assigned" individuals).

In Dja, only one morphospecies was described, as very few herbarium specimens were available from this location. All but two individuals were assigned to a single gene pool, nr-A. Most individuals from this site were chloroplast haplotype H2, but rare examples of H1, H3, and H9 haplotypes were also identified.

In Korup, all individuals identified as *C. dinklagei* were assigned to gene pool nr-C, whereas most *C. parviflora* individuals were assigned to gene pool nr-B. Two individuals described as *C. parviflora* were, however, assigned to nr-C. As we had no herbarium specimens for these two individuals, their morphological identification

could not be confirmed. The three individuals that were not identified morphologically were also assigned to gene pool nr-C. Regardless of the morphospecies or the gene pool, all individuals from Korup carried the H6 chloroplast haplotype.

4. Discussion

For species identification and delimitation, the molecular markers analyzed in this study were not equally relevant. We found strong associations between morphological taxonomic identification and nuclear ribosomal DNA (nrDNA) markers, although the strength of this association varied between sample sites. In contrast, cpDNA data were often incongruent with taxonomic identifications and nrDNA markers.

CpDNA markers did not reliably distinguish between species. For example, in the Korup group two morphospecies were clearly identified, *C. dinklagei* and *C. parviflora*. Every Korup individual, however, shared the same chloroplast haplotype. Such an observation demonstrates the pitfalls associated with using chloroplast molecular markers (and single DNA tags in general) as a DNA barcode (Duminil et al., 2006; Jakob and Blattner, 2006; Seberg and Petersen, 2009; Tsitrone et al., 2003). Similar observations have been made in animal species regarding the use of mitochondrial DNA sequences (Rubinoff and Holland, 2005). In general, the use of cpDNA sequences should be confined to species with strong barriers to inter-specific gene flow that have diverged for a sufficient period of time so as to reach monophyly at most characters (Duminil et al., 2010). These limitations highlight the need for nuclear-DNA data when identifying species, at least when closely related species are involved.

ITS sequences, the other unique DNA tag used in this study (but of nuclear origin), were more useful than were cpDNA sequences for identifying *Carapa* species. Using an in-depth coverage of taxonomic units, ITS sequences represent a more efficient DNA barcode than cpDNA (Chen et al., 2010; Muellner et al., 2011; Sass et al., 2007; Yao et al., 2010). The morphospecies identified in our study were, however, not always monophyletic in the ITS phylogenetic tree. It has been proposed that strongly supported monophyly should represent an important criterion for successful species resolution (Fazekas et al., 2008). Here, monophyly was strictly demonstrated for one morphospecies, *C. dinklagei*. Partial monophyly was demonstrated for a second morphospecies (*C. parviflora*), but monophyly was absent from the remaining morphospecies (*C. hygrophylla*, *C. palustris*, and *C. macrantha*). Indeed, ITS sequences were highly diverse within these morphospecies, and no associations between morphological identification and ITS polymorphisms could be found. These results can be interpreted as (i) a lack of power for ITS markers (for example, when paralogs exist); (ii) evidence that the *C. hygrophylla*, *C. palustris,* and *C. macrantha* morphospecies are not true species (separate reproductive entities); or (iii) erroneous taxonomic identifications based on sterile herbarium vouchers. Under the assumption that these morphospecies represent true species, these results illustrate the analytic limits associated with ITS sequences. These limits are primarily due to problems of unconcerted evolution (Calonje et al., 2009), a common phenomenon in recent species radiations (Shaffer and Thomson, 2007). In addition, these data illustrate the general problem associated with using a single DNA tag. A single tag provides only one gene tree estimate, which may or may not represent the true species genealogy (Degnan and Rosenberg, 2006; Maddison, 1997; Rosenberg, 2002). To avoid these limitations, one should sequence several unlinked DNA regions to generate multiple independent representations of the genealogical history of a lineage (Chase et al., 2005; Fazekas et al., 2008; Newmaster et al., 2006; Seberg and Petersen, 2009). Alternatively, the use of ultra-polymorphic markers (such as microsatellites) provides an efficient approach for species identification (e.g., Caetano et al., 2008; Gugerli et al., 2008; Lepais et al., 2009).

For delimiting closely related species, population genetics approaches represent promising alternatives to DNA barcoding. In these approaches, nuclear microsatellites allow one to study associations between morphological and molecular polymorphisms, overcoming problems associated with the use of cpDNA, ITS, and unique DNA tags in general. These markers allow for the grouping of individuals (gene pools) that seem to behave like a panmictic population. Gene pools thus represent separate reproductive entities, which result from strong barriers to gene flow. When distinct gene pools are found in sympatry, this strongly supports the presence of separate species, according to the BSC. Allopatric gene pools can, however, also occur within a species because of population differentiation. In such cases, one cannot assess whether distinct species exist. The efficient use of microsatellite data to identify species has already been demonstrated for *Carapa* species from tropical America (Duminil et al., 2006). Here we have confirmed that two *Carapa* sympatric morphospecies within the Korup site (*C. dinklagei* and *C. parviflora*) form distinct gene pools and do not seem to generate hybrids. Similarly, *C. dinklagei* was well separated from other *Carapa* morphospecies within the Bipindi region. Furthermore, Bipindi individuals identified as *C. macrantha* and *C. parviflora* may also form separate gene pools, although the small sample size does not allow for a firm conclusion. As for ITS, however, associations between gene pools and morphospecies were not perfect. When all individuals were subjected to Bayesian analyses, three gene pools were identified, despite the identification of at least five morphospecies. Interestingly, associations between morphology and genetics agreed with previous observations based on ITS. As such, the validity of morphological identifications, or of taxonomic species delimitations, must be questioned. Moreover, among individuals identified as *C. parviflora*, those from Korup formed a gene pool that was entirely distinct from Bipindi individuals. Although gene pool nr-A contained several morphospecies, we cannot exclude the possibility that some morphospecies were insufficiently sampled, thereby preventing the algorithm from unambiguously separating gene pools. Additional sampling of these morphospecies is thus necessary to conclude whether they are separate species, according to the BSC.

In tropical plants, difficulties are frequently encountered when trying to define reliable morphological diagnostic characters. Despite our efforts to collect as many herbarium vouchers as possible, species identification using sterile vouchers was often difficult and in some cases was impossible. The availability of reliable morphological markers for species diagnosis often requires deep taxonomic investigations, a condition rarely met for tropical species (Koffi et al., 2010; Muellner et al., 2009, 2005). In the case of *Carapa*, the identification of species-level vouchers is still problematic, despite recent revisions to the genus (Kenfack, 2008). Vouchers collected for this study had limited utility because they were often comprised of leaves collected from juvenile trees or from young shoots. As such, these samples lacked essential morphological characters that are necessary to distinguish between species. Leaves from mature trees are more difficult to obtain but contain important discriminative characters such as the number of leaflet pairs, leaflet shape, and the relative length of the petiole.

Limitations associated with traditional qualitative markers can often be overcome using alternative approaches that asses shape information from morphological structures (Andres-Sanchez et al., 2009; Ekrt et al., 2010; Hearn, 2009). We applied a morphometric approach to the Korup group but failed to discriminate between species when unidentified individuals were analyzed (i.e., the raw data). This method failed, in part, because we did not take into consideration variations in leaflet shape and size within a single leaf. Used without *a priori* information concerning morphospecies assign-

ments, this approach was less efficient than the use of qualitative traits to diagnose species. This can be explained by the presence of both juvenile and adult leaves in our dataset, as these leaves present shape differences according to their maturity. When morphospecies assignment data were included, however, this approach detected significant differences in leaf shape between *C. parviflora* and *C. dinklagei* from Korup. Although we agree that morphometric analyses have limited utility in the field, we would like to emphasize the potential that these types of approaches have to define diagnostic morphological characters retrospectively. This top-down approach is valuable both for genetic studies (with morphological markers aiding the identification of individuals) and for taxonomic studies (with molecular data providing additional important information).

Another factor that makes species identification and delimitation more difficult is related to the evolutionary history of the genus itself. Comparisons between *Carapa* nrDNA and cpDNA sequences support the occurrence of hybridization events among *Carapa* species in the past. For example, *C. dinklagei* and *C. parviflora* share identical chloroplast haplotypes (H1 and H6), although they belong to two distinct nuclear gene pools. Given the substantial genetic distance between these two haplotypes, shared ancestral polymorphism is an unlikely explanation. In addition, shared ancestral polymorphism cannot explain the fact that shared haplotypes are found within the same geographic area as observed here (H6 is found only in Korup). The substantial genetic distance at a nuclear gene (ITS) also supports introgression rather than shared ancestral polymorphism. Finally, the absence of intermediates in the Bayesian clustering method suggests that inter-specific gene flow between these two species is relatively ancient. In tropical America, traces of inter-specific gene flow were found between *C. procera* and *C. guyanensis,* demonstrating the permeability of species barriers within this genus (Duminil et al., 2006).

Our results have demonstrated the presence of species complexes in both Bipindi and Dja. A species complex was attested by the assignment of different morphospecies to a unique nuclear gene pool and by their repartition on the ITS phylogenetic tree. Results indicate that these morphospecies may have diverged recently and that a complete barrier to gene flow between these species has not yet been established. Alternatively, these morphospecies may belong to the same species but display substantial morphological variation. A third alternative is that these morphospecies were represented by too few individuals (i.e., insufficient sampling), which prevented the identification of separate gene pools. These results illustrate the difficulties encountered when assessing species status as well as the problems associated with underlying definitions of species concepts.

5. Conclusions

Despite the use of multiple diagnostic characters, *Carapa* species identification and delimitation is complex in the Lower Guinean region. Overall, a population genetics approach, which relied on a set of ultra-polymorphic markers, more reliably delimited *Carapa* species than approaches based on a single DNA locus. Despite the limited utility of cpDNA sequences as a barcode in the genus *Carapa*, using them in association with microsatellite markers provided support for historical inter-specific gene flow. From a taxonomical point of view, our analyses support the designation of *C. dinklagei* as a distinct species, following the BSC. A similar conclusion can be drawn for *C. parviflora*, although with less confidence, as all but two individuals formed another well-supported gene pool. Other described morphospecies (*C. macrantha*, *C. hygrophila*, *C. palustris*, *C. angustifolia*) were not easily distinguished using these genetic markers and might represent a complex of species with frequent gene flow.

Acknowledgments

We are grateful to two anonymous referees for their critical comments on a previous version of the manuscript. This work was funded by the Gembloux Agricultural University (FUSAGx, Belgium) via project PPR 10.000, the Fonds de la Recherche Scientifique (FNRS, Belgium; Grants FRFC 2.4.576.07.F and MIS F.4.519.10.F) and by the IFORA (Iles Forestières Africaines) project financed by the French ANR (Agence Nationale de la Recherche) under the ANR-BIODIV program.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2011.09.020.

References

Agapow, P.M., Bininda-Emonds, O.R., Crandall, K.A., Gittleman, J.L., Mace, G.M., Marshall, J.C., Purvis, A., 2004. The impact of species concept on biodiversity studies. Q. Rev. Biol. 79, 161–179.

Andres-Sanchez, S., Rico, E., Herrero, A., Santos-Vicente, M., Martinez-Ortega, M.M., 2009. Combining traditional morphometrics and molecular markers in cryptic taxa: towards an updated integrative taxonomic treatment for *Veronica* subgenus *Pentasepalae* (Plantaginaceae sensu APG II) in the western Mediterranean. Bot. J. Linn. Soc. 159, 68–87.

Aublet, J.B.C., 1775. Histoire des plantes de la Guiane Françoise rangées suivant la méthode sexuelle (P.-F. Didot jeune, Londres et Paris, quatre volumes).

Bandelt, H.-J., Forster, P., Röhl, A., 1999. Median-joining networks for inferring intraspecific phylogenies. Mol. Biol. Evol. 16, 37–487.

Bond, J.E., Hedin, M.C., Ramirez, M.G., Opell, B.D., 2001. Deep molecular divergence in the absence of morphological and ecological change in the Californian coastal dune endemic trapdoor spider *Aptostichus simus*. Mol. Ecol. 10, 899–910.

Caetano, S., Nusbaumer, L., Naciri, Y., 2008. Chloroplast and microsatellite markers in *Astronium urundeuva* (Allemao) Engl. and close species of Anacardiaceae: toward the definition of a species complex? Candollea 63, 115–130.

Calonje, M., Martin-Bravo, S., Dobes, C., Gong, W., Jordon-Thaden, I., Kiefer, C., Kiefer, M., Paule, J., Schmickl, R., Koch, M.A., 2009. Non-coding nuclear DNA markers in phylogenetic reconstruction. Plant Syst. Evol. 282, 257–280.

Candolle, C.D., 1878. Meliaceae. In: De Candolle, A.C. (Ed.), Monographiae Phanerogamarum, vol. 1, pp. 419–758.

Chase, M.W., Salamin, N., Wilkinson, M., Dunwell, J.M., Kesanakurthi, R.P., Haidar, N., Savolainen, V., 2005. Land plants and DNA barcodes: short-term and long-term goals. Philos. T. Roy. Soc. B 360, 1889–1895.

Chen, S.L., Yao, H., Han, J.P., Liu, C., Song, J.Y., Shi, L.C., Zhu, Y.J., Ma, X.Y., Gao, T., Pang, X.H., Luo, K., Li, Y., Li, X.W., Jia, X.C., Lin, Y.L., Leon, C., 2010. Validation of the ITS2 Region as a Novel DNA Barcode for Identifying Medicinal Plant Species. Plos One 5.

Dayanandan, S., Dole, J., Bawa, K., Kesseli, R., 1999. Population structure delineated with microsatellite markers in fragmented populations of a tropical tree, *Carapa guianensis* (Meliaceae). Mol. Ecol. 8, 1585–1592.

Degnan, J.H., Rosenberg, N.A., 2006. Discordance of species trees with their most likely gene trees. PLoS Genet. 2, 762–768.

Demesure, B., Sodzi, N., Petit, R.J., 1995. A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. Mol. Ecol. 4, 129–131.

Devos, N., Tyteca, D., Raspe, O., Wesselingh, R.A., Jacquemart, A.L., 2003. Patterns of chloroplast diversity among western European *Dactylorhiza* species (Orchidaceae). Plant Syst. Evol. 243, 85–97.

Dobes, C.H., Mitchell-Olds, T., Koch, M.A., 2004. Extensive chloroplast haplotype variation indicates Pleistocene hybridization and radiation of North American *Arabis drummondii*, *A. x divaricarpa*, and *A-holboellii* (Brassicaceae). Mol. Ecol. 13, 349–370.

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

Duminil, J., Caron, H., Scotti, I., Cazal, S.-O., Petit, R.J., 2006. Blind population genetics survey of tropical rainforest trees. Mol. Ecol. 15, 3505–3513.

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

Ekrt, L., Holubova, R., Travnicek, P., Suda, J., 2010. Species boundaries and frequency of hybridization in the Dryopteris Carthusiana (Dryopteridaceae) complex: a taxonomic puzzle resolved using genome size data. Am. J. Bot. 97, 1208–1219.

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

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.

Fazekas, A.J., Burgess, K.S., Kesanakurti, P.R., Graham, S.W., Newmaster, S.G., Husband, B.C., Percy, D.M., Hajibabaei, M., Barrett, S.C.H., 2008. Multiple Multilocus DNA Barcodes from the Plastid Genome Discriminate Plant Species Equally Well. Plos One 3.

Forget, P.M., 1996. Removal of seeds of *Carapa procera* (Meliaceae) by rodents and their fate in rainforest in French Guiana. J. Trop. Ecol. 12, 751–761.

Forget, P.M., Mercier, F., Collinet, F., 1999. Spatial patterns of two rodent-dispersed rain forest trees *Carapa procera* (Meliaceae) and *Vouacapoua americana* (Caesalpiniaceae) at Paracou, French Guiana. J. Trop. Ecol. 15, 301–313.

Gelman, A., Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. Statist. Sci. 7, 211–457.

Gugerli, F., Brodbeck, S., Holderegger, R., 2008. Utility of multilocus genotypes for taxon assignment in stands of closely related European white oaks from Switzerland. Ann. Bot. 102, 855–863.

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

Harms, H., 1917. Meliaceae africanae. Notizblatt des Königl. botanischen Gartens und Museums zu Berlin 65, 223–232.

Hearn, D.J., 2009. Shape analysis for the automated identification of plants from images of leaves. Taxon 58, 934–954.

Hollingsworth, P.M., Forrest, L.L., Spouge, J.L., Hajibabaei, M., Ratnasingham, S., van der Bank, M., Chase, M.W., Cowan, R.S., Erickson, D.L., Fazekas, A.J., Graham, S.W., James, K.E., Kim, K.J., Kress, W.J., Schneider, H., van AlphenStahl, J., Barrett, S.C.H., van den Berg, C., Bogarin, D., Burgess, K.S., Cameron, K.M., Carine, M., Chacon, J., Clark, A., Clarkson, J.J., Conrad, F., Devey, D.S., Ford, C.S., Hedderson, T.A.J., Hollingsworth, M.L., Husband, B.C., Kelly, L.J., Kesanakurti, P.R., Kim, J.S., Kim, Y.D., Lahaye, R., Lee, H.L., Long, D.G., Madrinan, S., Maurin, O., Meusnier, I., Newmaster, S.G., Park, C.W., Percy, D.M., Petersen, G., Richardson, J.E., Salazar, G.A., Savolainen, V., Seberg, O., Wilkinson, M.J., Yi, D.K., Little, D.P., Grp, C.P.W., 2009. A DNA barcode for land plants. Proc. Natl. Acad. Sci., U.S.A. 106, 12794–12797.

Jakob, S.S., Blattner, F.R., 2006. A chloroplast genealogy of Hordeum (Poaceae): long-term persisting haplotypes, incomplete lineage sorting, regional extinction, and the consequences for phylogenetic inference. Mol. Biol. Evol. 23, 1602–1612.

Kenfack, D., 2008. Systematic studies in *Carapa* (Meliaceae–Swietenioideae). Ph.D. Thesis, University of Missouri-St. Louis, 265p.

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.

Kenfack, D., Thomas, D.W., Chuyong, G., Condit, R., 2007. Rarity and abundance in a diverse African forest. Biodivers. Conserv. 16, 2045–2074.

Koffi, K.G., Heuertz, M., Doumenge, C., Onana, J.-M., Gavory, F., Hardy, O.J., 2010. A combined analysis of morphological traits, chloroplast and nuclear DNA sequences with *Santiria trimera* (Burseraceae) suggest several species following the Biological Species Concept. Plant Ecology and Evolution 143, 160–169.

Lahaye, R., Van der Bank, M., Bogarin, D., Warner, J., Pupulin, F., Gigot, G., Maurin, O., Duthoit, S., Barraclough, T.G., Savolainen, V., 2008. DNA barcoding the floras of biodiversity hotspots. Proc. Natl. Acad. Sci., U.S.A. 105, 2923–2928.

Lee, C., Wen, J., 2004. Phylogeny of *Panax* using chloroplast *trn* C-*trn* D intergenic region and the utility of *trn* C-*trn* D in interspecific studies of plants. Mol. Phylogenet. Evol. 31, 894–903.

Lepais, O., Petit, R.J., Guichoux, E., Lavabre, J.E., Alberto, F., Kremer, A., Gerber, S., 2009. Species relative abundance and direction of introgression in oaks. Mol. Ecol. 18, 2228–2242.

Lexer, C., Joseph, J., van Loo, M., Prenner, G., Heinze, B., Chase, M.W., Kirkup, D., 2009. The use of digital image-based morphometrics to study the phenotypic mosaic in taxa with porous genomes. Taxon 58, 349–364.

Maddison, W.P., 1997. Gene trees in species trees. Syst. Biol. 46, 523–536.

Muellner, A.N., Samuel, R., Chase, M.W., Pannell, C.M., Greger, H., 2005. Aglaia (Meliaceae): an evaluation of taxonomic concepts based on DNA data and secondary metabolites. Am. J. Bot. 92, 534–543.

Muellner, A.N., Pennington, T.D., Chase, M.W., 2009. Molecular phylogenetics of Neotropical Cedreleae (mahogany family, Meliaceae) based on nuclear and plastid DNA sequences reveal multiple origins of "*Cedrela odorata*". Mol. Phylogenet. Evol. 52, 461–469.

Muellner, A.N., Schaefer, H., Lahaye, R., 2011. Evaluation of candidate DNA barcoding loci for economically important timber species of the mahogany family (Meliaceae). Mol. Ecol. Res. 11, 450–460.

Navarro, N., Zatarain, X., Montuire, S., 2004. Effects of morphometric descriptor changes on statistical classification and morphospaces. Biol. J. Linn. Soc. 83, 243–260.

Nei, M., 1978. Estimation of average heterozygosity and genetic distance for small number of individuals. Genetics 89, 583–590.

Neto, J.C., Meyer, G.E., Jones, D.D., Samal, A.K., 2006. Plant species identification using Elliptic Fourier leaf shape analysis. Comput. Electron. Agric. 50, 121–134.

Newmaster, S.G., Fazekas, A.J., Ragupathy, S., 2006. DNA barcoding in land plants: evaluation of rbcL in a multigene tiered approach. Can. J. Bot. 84, 335–341.

Noamesi, G.K., 1958. A revision of Xylocarpae (Meliceae). Ph.D. Thesis, University of Wisconsin.

Posada, J., 2008. jModelTest: phylogenetic model averaging. Mol. Biol. Evol. 25, 1253–1256.

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

Rohlf, F.J., 2001. Comparative methods for the analysis of continuous variables: geometric interpretations. Evolution 55, 2143–2160.

Ronquist, F., Huelsenbeck, J.P., 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.

Rosenberg, N.A., 2002. The probability of topological concordance of gene trees and species trees. Theor. Popul. Biol. 61, 225–247.

Rubinoff, D., Holland, B.S., 2005. Between two extremes: mitochondrial DNA is neither the panacea nor the nemesis of phylogenetic and taxonomic inference. Syst. Biol. 54, 952–961.

Sass, C., Little, D.P., Stevenson, D.W., Specht, C.D., 2007. DNA Barcoding in the Cycadales: Testing the Potential of Proposed Barcoding Markers for Species Identification of Cycads. Plos One 2.

Seberg, O., Petersen, G., 2009. How Many Loci Does it Take to DNA Barcode a Crocus? Plos One 4.

Shaffer, H.B., Thomson, R.C., 2007. Delimiting species in recent radiations. Syst. Biol. 56, 896–906.

Sheets, H.D., Kim, K., Mitchell, C.E., 2004. A combined landmark and outline-based approach to ontogenetic shape change in the Ordovician trilobite *Triarthrus becki*. In: Morphometrics: Applications in Biology and Paleontology Edited by Elewa AMT. Springer, Berlin, pp. 67–82.

Sheets, H.D., Covino, K.M., Panasiewicz, J.M., Morris, S.R., 2006. Comparison of geometric morphometric outline methods in the discrimination of age-related differences in feather shape. Front. Zool. 3, 15.

Staner, P., 1941. Les Méliacées du Congo Belge. Bulletin du Jardin botanique de l'Etat 16, 114–125.

Styles, B.T., White, F., 1991. Meliaceae. In: Polhil, R.M. (Ed.), Flora of Tropical East Africa. A.A. Balkema, Rotterdam.

Swofford, D.L., 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland.

Thomas, D.W., Kenfack, D., Chuyong, G.B., Moses, S.N., Losos, E.C., Condit, R.S., Songwe, N.C., 2003. Tree Species of Southwestern Cameroon: tree Distribution Maps, Diameter Tables, and Species Documentation of the 50-Hectare Korup Forest Dynamics Plot. Tree Species of Southwestern Cameroon: tree Distribution Maps, Diameter Tables, and Species Documentation of the 50-Hectare Korup Forest Dynamics Plot, xviii, + 254pp.

Tsitrone, A., Kirkpatrick, M., Levin, D.A., 2003. A model for chloroplast capture. Evolution 57, 1776–1782.

Urbatschl, E., Baldwin, B.L., Donoghue, M.J., 2000. Phylogeny of the coneflowers and relatives (Heliantheae: Asteraceae) based on nuclear rDNA internal transcribed spacer (ITS) sequences and chloroplast DNA restriction site data. Syst. Bot. 25, 539–565.

Vinson, C.C., Azevedo, V.C.R., Sampaio, I., Ciampi, A.Y., 2005. Development of microsatellite markers for *Carapa guianensis* (Aublet), a tree species from the Amazon forest. Mol. Ecol. Notes 5, 33–34.

Viscosi, V., Lepais, O., Gerber, S., Fortini, P., 2009. Leaf morphological analyses in four European oak species (*Quercus*) and their hybrids: a comparison of traditional and geometric morphometric methods. Plant Biosyst. 143, 564–574.

White, T.J., Birns, T., Lee, S., Taylor, J., 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis, M., Gelfand, D., Sninsky, J., White, T. (Eds.), PCR Protocols: A Guide to Methods and Applications. Academic Press, San Diego, California, USA, pp. 315–322.

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. Am. J. Bot. 91, 2022–2029.

Yao, H., Song, J.Y., Liu, C., Luo, K., Han, J.P., Li, Y., Pang, X.H., Xu, H.X., Zhu, Y.J., Xiao, P.G., Chen, S.L., 2010. Use of ITS2 Region as the Universal DNA Barcode for Plants and Animals. Plos One 5.