Annals of Botany **100**: 1565–1583, 2007
doi:10.1093/aob/mcm257, available online at www.aob.oxfordjournals.org

OXFORD JOURNALS
OXFORD UNIVERSITY PRESS

Towards a Phylogeny for *Coffea* (Rubiaceae): Identifying Well-supported Lineages Based on Nuclear and Plastid DNA Sequences

OLIVIER MAURIN[1,4], AARON P. DAVIS[2,*], MICHAEL CHESTER[1,5], ESTHER F. MVUNGI[3], YASMINA JAUFEERALLY-FAKIM[6] and MICHAEL F. FAY[1]

[1]*Jodrell Laboratory, Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3DS, UK,* [2]*Herbarium, Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3AB, UK,* [3]*Department of Botany, University of Dar es Salaam, PO Box 35060, Dar es Salaam, Tanzania,* [4]*Department of Botany and Plant Biotechnology, University of Johannesburg, PO Box 524, Auckland Park, 2006, Gauteng, South Africa,* [5]*School of Biological and Chemical Sciences, Queen Mary, University of London, Mile End Road, London E1 4NS, UK and* [6]*Faculty of Agriculture, University of Mauritius, Reduit, Mauritius*

Received: 28 June 2007 Returned for revision: 25 July 2007 Accepted: 30 August 2007 Published electronically: 22 October 2007

• *Background and Aims* The phylogenetic relationships between species of *Coffea* and *Psilanthus* remain poorly understood, owing to low levels of sequence variation recovered in previous studies, coupled with relatively limited species sampling. In this study, the relationships between *Coffea* and *Psilanthus* species are assessed based on substantially increased molecular sequence data and greatly improved species sampling.
• *Methods* Phylogenetic relationships are assessed using parsimony, with sequence data from four plastid regions [*trnL–F* intron, *trnL–F* intergenic spacer (IGS), *rpl16* intron and *accD–psa1* IGS], and the internal transcribed spacer (ITS) region of nuclear rDNA (ITS 1/5·8S/ITS 2). Supported lineages in *Coffea* are discussed within the context of geographical correspondence, biogeography, morphology and systematics.
• *Key Results* Several major lineages with geographical coherence, as identified in previous studies based on smaller data sets, are supported. Other lineages with either geographical or ecological correspondence are recognized for the first time. *Coffea* subgenus *Baracoffea* is shown to be monophyletic, but *Coffea* subgenus *Coffea* is paraphyletic. Sequence data do not substantiate the monophyly of either *Coffea* or *Psilanthus*. Low levels of sequence divergence do not allow detailed resolution of relationships within *Coffea*, most notably for species of *Coffea* subgenus *Coffea* occurring in Madagascar. The origin of *C. arabica* by recent hybridization between *C. canephora* and *C. eugenioides* is supported. Phylogenetic separation resulting from the presence of the Dahomey Gap is inferred based on sequence data from *Coffea*.

Key words: Africa, *accD–psa1* IGS, *Coffea*, coffee, Indian Ocean Islands, ITS, Madagascar, molecular phylogeny, *rpl16* intron, Rubiaceae, *trnL–F* intron, *trnL–F* IGS.

INTRODUCTION

The genus *Coffea* L. comprises 103 species (Davis *et al.*, 2006) and occurs naturally in tropical Africa, Madagascar, the Comoros and the Mascarenes (Mauritius and Reunion). *Coffea* species are mostly restricted to humid evergreen forest, although some species are found in seasonally dry deciduous forest and/or bushland. The most recent classifications of *Coffea* (Bridson, 1988*a*, *b*, 2003; Davis *et al.*, 2005, 2006) divide the genus into two subgenera: subgenus *Coffea* (95 spp.) and subgenus *Baracoffea* (J.-F. Leroy) J.-F. Leroy (eight spp.). *Coffea* subgenus *Coffea* occurs throughout the range of the genus, whereas *Coffea* subgenus *Baracoffea* is restricted to the seasonally dry forest and scrubland of western Madagascar (Davis *et al.*, 2005) and, according to Leroy (1982), NE Kenya and SE Somalia. *Coffea* subgenus *Coffea* includes the species used in the production of coffee, i.e. *C. arabica* (Arabica coffee), *C. canephora* (robusta coffee) and *C. liberica* (Liberian coffee). *Coffea arabica* is by far the most important traded species, and provides at least 65 % of commercial production. *Coffea arabica* is the only allotetraploid *Coffea* species ($2n = 4x = 44$; Carvalho, 1952; Grassias and Kammacher, 1975); all other *Coffea* species are diploid ($2n = 2x = 22$). *Coffea arabica* is also self-compatible (Carvalho *et al.*, 1991), thus far only reported in two other species: *C. heterocalyx* (Coulibaly *et al.*, 2002) and *C. anthonyi* ined. (P. Stoffelen, pers. comm.).

It is now well established that *Psilanthus* Hook.f. is the closest relative of *Coffea*. Davis *et al.* (2007) showed that these genera form a well supported monophyletic group within tribe Coffeeae, based on molecular (BP [bootstrap percentage] 100; *b* [Bremer support value/decay value] = 9) and combined molecular–morphological data (BP 100; $b = 13$), for example. *Coffea* and *Psilanthus* share a unique carpel morphology: the endocarp (pyrene shell) is hard (horny/crustaceous), the pyrene (and seed) has a deep ventral invagination (i.e. 'coffee bean' morphology) and the seed coat consists of crushed endotestal cells with ± isolated fibres (Robbrecht and Puff, 1986). *Psilanthus* comprises 22 species and occurs in tropical Africa, southern and SE Asia, and as far east as tropical northern Australia (Davis, 2003). *Psilanthus* is also divided into two subgenera (Bridson, 1988*b*): *Psilanthus*

* For correspondence. E-mail a.davis@kew.org

subgenus *Psilanthus* (two spp.), and *Psilanthus* subgenus *Afrocoffea* (Moens) Bridson (20 spp.). *Psilanthus* subgenus *Psilanthus* is restricted to tropical West and Central Africa, whereas subgenus *Afrocoffea* occurs throughout the range of the genus. The morphological characterization of *Coffea* and *Psilanthus* and their subgenera is reported in detail by Davis *et al.* (2005).

Coffea and *Psilanthus* have been the focus of several recent phylogenetic studies, using systematic data from various sources, including morphology (Stoffelen, 1998; Davis *et al.*, 2005), random amplified polymorphic DNA (RAPD) (Lashermes *et al.*, 1993), sequences from plastid DNA (Cros, 1994; Lashermes *et al.*, 1996; Cros *et al.*, 1998) and internal transcribed spacer (ITS) sequences of nuclear rDNA (Lashermes *et al.*, 1997). At the species level, the studies of Lashermes *et al.* (1997) and Cros *et al.* (1998) provided the most useful data.

On the basis of ITS2 data, Lashermes *et al.* (1997: 953, fig. 4) separated *Coffea* into four main geographical groups, although the bootstrap support for three of these groups was negligible to weak, i.e. Madagascar (BP 53), East Africa (BP 22) and Central Africa (BP 67), and West and Central Africa formed an unresolved group. Cros *et al.* (1998) used sequence data from the *trnL–trnF* intergenic spacer (IGS) region, and separated *Coffea* into five area groupings, four the same as Lashermes *et al.* (1997), but with better support values, i.e. Madagascar (BP 82), East Africa (BP 64), Central Africa (BP 100) and West and Central Africa (BP 41), plus a west Africa clade (BP 100). Cros *et al.* (1998) concluded that there was good agreement between their *trnL–trnF* data analysis and the ITS study of Lashermes *et al.* (1997), including the separation of *Coffea* species into the four major geographical groups as given above, but also noted that there was some conspicuous incongruence between the two data sets. The most notable incongruity was the position of *C. arabica* within a clade of central African taxa [i.e. *C. eugenioides*, and *C.* sp. 'Moloundou' (= *C. anthonyi* ined.); see Davis *et al.*, 2006] in their plastid DNA analysis (BP 100), compared with placement within a 'canephoroid' species group (*C. canephora*, *C. brevipes*, *C. congensis*; BP 53) with ITS2 (Lashermes *et al.*, 1997). Further to this observation (Cros *et al.*, 1998), Raina *et al.* (1998) used genomic *in situ* hybridization (GISH) and fluorescent *in situ* hybridization (FISH) to study the genome organization and evolution of *C. arabica*, and they concluded that *C. congensis* and *C. eugenioides* are the diploid progenitors of *C. arabica*. Using restriction fragment length polymorphism (RFLP) markers in combination with GISH data, Lashermes *et al.* (1999) suggested that *C. arabica* is an amphidiploid formed by hybridization between *C. eugenioides* and *C. canephora*, or ecotypes related to these diploid species. Although Raina *et al.* (1998) did not take into account the diversity of *C. canephora*, *C. congensis* and *C. arabica* (they used only one sample per species), their result is not in any real conflict given that *C. congensis* and *C. canephora* are genetically very similar (Lashermes *et al.*, 1997; Prakesh *et al.*, 2005).

The other topological inconsistencies identified by Cros *et al.* (1998) involved two taxa from West Africa, *C. stenophylla* and *C. humilis*, and *C.* sp. 'X' (= *C. heterocalyx*; Coulibaly *et al.*, 2002, 2003), which were placed in different African groups for ITS and *trnL–trnF*. The plastid analysis of Cros *et al.* (1998) placed *C. stenophylla* as sister to *C. humilis*, and with the Central African, East African and Madagascan species (BP 35). *Coffea* sp. 'X' (= *C. heterocalyx*) was placed with species from West and Central Africa (BP 32; unresolved in relation to two accessions of *C. liberica*). The ITS2 analysis of Lashermes *et al.* (1997) placed *C. stenophylla* with species from West and Central Africa (BP 4), and *C. heterocalyx* with their Central Africa clade (BP = 67; unresolved in relation to *C. eugenioides* and *C.* sp. 'Moloundou' (= *C. anthonyi* ined.). Cros *et al.* (1998) suggested that these could be interpreted as the result of interspecies transfer of plastid DNA mediated by hybridization. However, with generally low levels of support, such discussions are speculative.

Owing to limited sequence divergence between *Coffea* and *Psilanthus* and the nested position of one *Psilanthus* species [*P. travancorensis* (Wight & Arn.) J.-F. Leroy] within *Coffea*, Lashermes *et al.* (1997) concluded that their ITS data did not support the recognition of two genera. On the basis of their *trnL–trnF* data, Cros *et al.* (1998) concurred with Lashermes *et al.* (1997) on this matter, although their tree topology shows an unresolved relationship between the two species of *Psilanthus* sampled (*P. mannii* and *P. ebracteolatus*) and *Coffea*. Cros *et al.* (1998) and Lashermes *et al.* (1997) did not include other representatives of Coffeeae (cf. Davis *et al.*, 2007) as outgroups.

Despite recent advances in *Coffea* systematics, the phylogenetic relationships between species of *Coffea* and *Psilanthus* are still poorly understood. This is mostly due to low levels of sequence variation so far recovered, coupled with a relatively small taxon sample size. Lashermes *et al.* (1997) used 37 samples, including 22 species of *Coffea* (approx. 21 % of known species diversity) and three species of *Psilanthus* (approx. 14 % of species diversity); the number of variable characters (nuclear substitutions and indels) was not given. In the study of Cros *et al.* (1998), 26 samples were used, covering 18 species of *Coffea* (approx. 17 % of known species diversity) and two species of *Psilanthus* (approx. 9 % of species diversity); only 32 variable characters (26 nucleotide substitutions and six indels) were found within their study sample. The studies of Lashermes *et al.* (1997) and Cros *et al.* (1998) did not include any samples of *Coffea* subgenus *Baracoffea* (eight spp.), Mascarene *Coffea* (at least three spp.) or the morphologically and geographically isolated (Bridson, 1979, 1983; Davis *et al.*, 2005) *C. rhamnifolia.* Small sample size is mostly a problem of logistics, as wild *Coffea* and *Psilanthus* species are not well represented in cultivation, are often difficult to find in the wild and DNA suitable for PCR and sequencing is not easily isolated from herbarium specimens. In addition, numerous new species of *Coffea* have been discovered and/or described since 1998, including seven species from west and central tropical Africa (Stoffelen *et al.*, 1996*a*, *b*, 1997*a*, *b*; Cheek *et al.*, 2002; Sonké and Stoffelen,

2004; Sonké *et al.*, 2006*b*), two from East Africa (Davis and Mvungi, 2004) and 13 from Madagascar (Davis and Rakotonasolo, 2000, 2001*a*, *b*, 2003; Davis, 2001). Further new species are in the process of formal description and publication (Davis *et al.*, 2006; Table 1).

In the investigation reported here, plastid sequence data from the *trnL–F* intron, *trnL–F* IGS, *rpl16* intron and *accD–psa1* IGS, and ITS sequences of nuclear rDNA were analysed in an attempt to increase the number of molecular characters available for phylogenetic reconstruction. Better sampling of *Coffea* has been made possible due to recent collecting activities, and it was possible to examine approx. 83 % of total species diversity for the genus, and 32 % of total species diversity for *Psilanthus* (see Materials and Methods). The main objective of the study was to identify well-supported lineages within *Coffea*, and to discuss these within the contexts of geographical coherence, morphology, systematics and biogeography. Secondarily, the aim was to elucidate consistently retrieved lineages within *Coffea* and well-supported lineages within *Psilanthus*, and to assess the relationship between *Coffea* and *Psilanthus*.

MATERIALS AND METHODS

Taxon sampling and plant material

As this study is concerned with assessing relationships above the rank of species, multiple species samples or infraspecific taxa were not used. Only two exceptions were made: a second sample of *C. heterocalyx* (IRD-Montpellier JC 66; voucher K) was included, which it is believed is the same as (or very similar to) that used by Lashermes *et al.* (1997) and Cros *et al.* (1998), and *C. liberica* var. *dewevrei*. *Coffea heterocalyx* is of considerable interest because it is reported to be self-compatible (Coulibaly *et al.*, 2002). N'Diaye *et al.* (2005) report that *C. liberica* var. *liberica* and *C. liberica* var. *dewevrei* have a high genetic differentiation (G_{st} 0·25) with AFLP markers, and the pollen viability of F_1 hybrids between them is low (44·2 %) and similar to interspecific hybrids, indicating that there are marked reproductive barriers between the two varieties. For these reasons, one sample of each variety of *C. liberica* was included.

The samples used in this study, with accepted taxon names, voucher information and GenBank accession numbers for the sequences, are given in Table 1. Most samples are of wild origin, with many collected during recent field expeditions to Madagascar (1997–2004), Tanzania (2001–2003) and Cameroon (in 2002). To complete the sampling, living material held in botanical gardens and coffee research stations, and some herbarium material (leaf samples or single seeds) taken from specimens held at K and BR (abbreviations after Holmgren *et al.*, 1990) were included. In total, 88 samples (86 spp.) of *Coffea* and seven samples (seven spp.) of *Psilanthus* were analysed. The sample includes representatives of all subgenera of *Coffea* and *Psilanthus*. Two species of the genus *Tricalysia* A. Rich ex DC. were used as outgroups, one from Madagascar and the other from Tanzania. *Tricalysia* is a close relative of *Coffea* and a member of Coffeeae (Andreasen and Bremer, 2000; Davis *et al.*, 2007).

The present *Coffea* taxon sampling includes a number of undescribed taxa, although these entities are well documented (Davis *et al.*, 2006) and appear to represent clearly defined species based on morphological data. Inclusion of undescribed species belonging to *Coffea* subgenus *Baracoffea* (Davis *et al.*, 2005) was necessary for testing the monophyly and systematic placement of the subgenus. The floristic study of Mascarene *Coffea* species by Leroy (1989) enumerated three native species of *Coffea*, *C. macrocarpa*, *C. myrtifolia* and *C. mauritiana*. A sample of *C. campaniensis* was also included, which Leroy (1989) placed in the synonymy of *C. mauritiana*.

Taxonomic details for all taxa (below generic rank) mentioned herein (such as author, place and date of publication, synonyms, distribution) follow the *World Rubiaceae Checklist* (www.kew.org/wcsp/rubiaceae). More detailed information for *Coffea* species is given in Davis *et al.* (2006).

DNA extraction, amplification and sequencing

Most of the DNA samples were obtained from silica-dried leaf material (Chase and Hills, 1991). The other DNA samples were extracted from seeds obtained from herbarium specimens; a single seed was used for each sample. DNA extraction was performed from a maximum of 0·3 g of silica-dried leaf material (or from one seed) using the 2 × CTAB method of Doyle and Doyle (1987). The DNA was purified on caesium chloride/ethidium bromide gradients (1·55 g mL^{-1} density) and dialysed before inclusion in the DNA Bank at the Royal Botanic Gardens, Kew (http://www.kew.org/data/dnaBank/homepage.html). To avoid problems of PCR inhibition, all DNA samples were further purified using QIAquick purification columns (QIAgen) following the manufacturer's protocol.

Amplification of the *trnL–F* region (the *trnL* intron and the *trnL–trnF* IGS), the *rpl16* intron, the *accD–psa1* (plastid DNA) IGS and ITS (nuclear encoded internal transcribed spacer) was performed using the primers listed in Table 2. Any ITS trace files showing evidence of heterogeneous ITS copies were cloned to isolate single sequences using the Promega pGem-T Easy Vector kit (catalogue no. A1360). The ITS region was then re-amplified from the transformed bacterial colonies using the M13 primers contained in the kit and a small portion of the colony as the DNA template. Amplification of *trnL–F* was carried out using primers c and f of Taberlet *et al.* (1991). For many taxa, the internal primers d and e also had to be used because of difficulty in amplifying the region as a single piece. Amplification of *rpl16* was carried out using primers 71F and 1661R of Jordan *et al.* (1996). For many taxa, the amplification of DNA using these primers was not satisfactory and so internal primers were designed on the basis of the first sequences in a conserved and GC-rich region suitable for amplification of *rpl16* in two fragments for Rubiaceae. The *accD–psa1* region was amplified using the primers ACCD-769 forward and PSA1-75 reverse from Mendenhall (1994). Two internal

TABLE 1. *Taxon accession data*

Taxon	Voucher	Source	*accD–psa1*	*rpl16*	*trnL–F*	*ITS*
Coffea abbayesii J.-F. Leroy	Davis 2334 (K)	Madagascar	DQ153438	DQ153687	DQ153805	DQ153566
Coffea ambongensis J.-F. Leroy ex A.P. Davis & Rakotonas., ined.	Davis 2509 (K)	Madagascar	DQ153419	DQ153668	DQ153786	DQ153539/ DQ153540/ DQ153541
Coffea andrambovatensis J.-F. Leroy	Davis 2322 (K)	Madagascar	DQ153422	DQ153671	DQ153789	DQ153545
Coffea anthonyi Stoff. & F. Anthony, ined. (*C.* sp. 'Moloundou')	IRD-Montpelier OE 53 (K)	DR Congo	DQ153489	DQ153738	DQ153856	DQ153620
Coffea ankaranensis J.-F. Leroy ex A.P. Davis & Rakotonas.	Davis 2331 (K)	Madagascar	DQ153407	DQ153656	DQ153774	DQ153527
Coffea arabica L.	Jaufeerally-Fakim 29 (K)	Mascarenes (Introduced)	DQ153478	DQ153727	DQ153845	DQ153609
Coffea arenesiana J.-F. Leroy	Davis 2207 (K)	Madagascar	DQ153440	DQ153689	DQ153807	DQ153568
Coffea augagneuri Dubard	Davis 2220 (K)	Madagascar	DQ153433	DQ153682	DQ153800	DQ153561
Coffea bakossii Cheek & Bridson	Lane 361 (K)	Cameroon	DQ153468	DQ153717	DQ153835	DQ153599
Coffea bertrandii A. Chev.	Davis 2348 (K)	Madagascar	DQ153424	DQ153673	DQ153791	DQ153549
Coffea betamponensis Portères & J.-F. Leroy	Davis 2300 (K)	Madagascar	DQ153421	DQ153670	DQ153788	DQ153543/ DQ153544
Coffea boinensis A.P. Davis & Rakotonas., ined.	Davis 2502 (K)	Madagascar	DQ153408	DQ153657	DQ153775	DQ153528
Coffea boiviniana (Baill.) Drake	Davis 2231 (K)	Madagascar	DQ153426	DQ153675	DQ153793	DQ153551/ DQ153552/ DQ153553
Coffea brevipes Heirn	Maurin 8 (K)	Cameroon	DQ153460	DQ153709	DQ153827	DQ153591
Coffea bridsoniae A.P. Davis & Mvungi	Davis 2904 (K)	Tanzania	DQ153455	DQ153704	DQ153822	DQ153584/ DQ153585/ DQ153586
Coffea buxifolia A. Chev.	Rakotonasolo 69 (K, TAN)	Madagascar	DQ153442	DQ153691	DQ153809	DQ153570
Coffea campaniensis J.-F. Leroy	Leroy 55 (K)	Mascarenes (Mauritius)	DQ153470	DQ153719	DQ153837	DQ153601
Coffea canephora Pierre ex Froehn.	Maurin 21 (K)	Cameroon (cultivated)	DQ153462	DQ153711	DQ153829	DQ153593
Coffea commersoniana (Baill.) A. Chev.	Davis 2715 (K)	Madagascar	DQ153432	DQ153681	DQ153799	DQ153560
Coffea congensis A. Froehn.	Harris & Fay 1507 (BR, K, MO)	Cameroon	DQ153467	DQ153716	DQ153834	DQ153598
Coffea costatifructa Bridson	ORSTOM 08 117 (K)	Tanzania	DQ153473	DQ153722	DQ153840	DQ153604
Coffea coursiana J.-F. Leroy	Davis 2278 (K)	Madagascar	DQ153417	DQ153666	DQ153784	DQ153537
Coffea decaryana J.-F. Leroy	Davis 1537 (K)	Madagascar	DQ153429	DQ153678	DQ153796	DQ153556
Coffea dubardii Jum.	Davis 2216 (K)	Madagascar	DQ153435	DQ153684	DQ153802	DQ153563
Coffea eugeniodes S.Moore	Harley 9332 (BR, K)	Tanzania	DQ153457	DQ153706	DQ153824	DQ153588
Coffea fadenii Bridson	Mvungi 9 (DSM, K)	Tanzania	DQ153446	DQ153695	DQ153813	DQ153574
Coffea farafanganensis J.-F. Leroy	Davis 2317 (K)	Madagascar	DQ153405	DQ153654	DQ153772	DQ153525
Coffea grevei Drake ex A. Chev.	Davis 2566 (K)	Madagascar	DQ153414	DQ153663	DQ153781	DQ153534
Coffea heimii J.-F. Leroy	Davis 2241 (K)	Madagascar	DQ153431	DQ153680	DQ153798	DQ153558/ DQ153559
Coffea heterocalyx Stoff.	Maurin 23 (K)	Cameroon	DQ153463	DQ153712	DQ153830	DQ153594
Coffea heterocalyx Stoff.	IRD-Montpelier JC 66 (K)	?Cameroon/ DR Congo)	DQ153492	DQ153741	DQ153859	DQ153623
Coffea homollei J.-F. Leroy	Davis 2305 (K)	Madagascar	DQ153402	DQ153651	DQ153769	DQ153521
Coffea humbertii J.-F. Leroy	Rakotonasolo 50 (K, TAN)	Madagascar	DQ153437	DQ153686	DQ153804	DQ153565
Coffea humblotiana Baill.	Davis 2327 (K)	Madagascar	DQ153411	DQ153660	DQ153778	DQ153531
Coffea humilis A. Chev.	Bamps 1967 (BR)	Ivory Coast	DQ153480	DQ153729	DQ153847	DQ153611
Coffea kapakata (A. Chev.) Bridson	Hepper & Maley 7723 (K)	Angola	DQ153465	DQ153714	DQ153832	DQ153596
Coffea kianjavatensis J.-F. Leroy	Davis 2313 (K)	Madagascar	DQ153482	DQ153731	DQ153849	DQ153613
Coffea kihansiensis A.P. Davis & Mvungi	Mvungi 21 (DSM, K)	Tanzania	DQ153454	DQ153703	DQ153821	DQ153583
Coffea kimbozensis Bridson	Mvungi 6 (DSM, K)	Tanzania	DQ153447	DQ153696	DQ153814	DQ153575
Coffea kivuensis Lebrun	Lebrun 5539 (BR)	Zaire	DQ153481	DQ153730	DQ153848	DQ153612
Coffea pterocarpa A.P. Davis & Rakotonas., ined.	Davis 2519 (K)	Madagascar	DQ153425	DQ153674	DQ153792	DQ153550
Coffea labatii A.P. Davis & Rakotonas., ined.	Davis 3069 (K)	Madagascar	DQ153499	DQ153748	DQ153866	DQ153630
Coffea lancifolia A. Chev.	Davis 2307 (K)	Madagascar	DQ153403	DQ153652	DQ153770	DQ153522
Coffea leroyi A.P. Davis	Davis 2311 (K)	Madagascar	DQ153404	DQ153653	DQ153771	DQ153523/ DQ153524
Coffea liaudii J.-F. Leroy ex A.P. Davis	Rakotonasolo 61 (K, TAN)	Madagascar	DQ153434	DQ153683	DQ153801	DQ153562
Coffea liberica var. *liberica* Bull. ex Hiern	Billiet 19370062 (BR)	DR Congo	DQ153479	DQ153728	DQ153846	DQ153610
Coffea liberica var. *dewerei* (De Wild. & T. Durand) Lebrun	Hepper & Maley 7729 (BR, K, MO)	Central African Republic	DQ153472	DQ153721	DQ153839	DQ153603
Coffea littoralis A.P. Davis & Rakotonas.	Rakotonasolo 261 (K)	Madagascar	DQ153441	DQ153690	DQ153808	DQ153569

Continued

TABLE 1. *Continued*

Taxon	Voucher	Source	*accD–psa1*	*rpl16*	*trnL–F*	*ITS*
Coffea lulandoensis Bridson	Mvungi 2 (DSM, K)	Tanzania	DQ153452	DQ153701	DQ153819	DQ153580
Coffea macrocarpa A. Rich.	Gueho 18555 (K)	Mascarenes (Mauritius)	DQ153471	DQ153720	DQ153838	DQ153602
Coffea mapiana Sonké, Nguembou & A.P. Davis	Sonké 3694 (K, YA)	Cameroon	DQ153509	DQ153758	DQ153876	DQ153640
Coffea mangoroensis Portères	Rakotonasolo 41 (K, TAN)	Madagascar	DQ153503	DQ153752	DQ153870	DQ153634
Coffea manombensis A.P. Davis	Davis 2141 (K)	Madagascar	DQ153445	DQ153694	DQ153812	DQ153573
Coffea mauritiana Lam.	Friedmann 1267 (K)	Mascarenes (Reunion)	DQ153469	DQ153718	DQ153836	DQ153600
Coffea mayombensis A. Chev.	Maurin 16 (K)	Cameroon	DQ153461	DQ153710	DQ153828	DQ153592
Coffea mcphersonii A.P. Davis & Rakotonas.	Davis 2339 (K)	Madagascar	DQ153423	DQ153672	DQ153790	DQ153546/ DQ153547/ DQ153548
Coffea millotii J.-F. Leroy	Davis 2306 (K)	Madagascar	DQ153409	DQ153658	DQ153776	DQ153529
Coffea mongensis Bridson	Mvungi 11 (DSM, K)	Tanzania	DQ153448	DQ153697	DQ153815	DQ153576
Coffea montekupensis Stoff.	Davis 3010 (K)	Cameroon	DQ153459	DQ153708	DQ153826	DQ153590
Coffea montis-sacri A.P. Davis	Davis 2308 (K)	Madagascar	DQ153430	DQ153679	DQ153797	DQ153557
Coffea moratii J.-F. Leroy ex A.P. Davis & Rakotonas.	Davis 2326 (K)	Madagascar	DQ153502	DQ153751	DQ153869	DQ153633
Coffea mufindiensis Hutch. ex Bridson	Mvungi 19 (DSM, K)	Tanzania	DQ153449	DQ153698	DQ153816	DQ153577
Coffea myrtifolia (A. Rich. ex DC.) J.-F. Leroy	Jaufeerally-Fakim 22 (K)	Mascarenes (Mauritius)	DQ153477	DQ153726	DQ153844	DQ153608
Coffea perrieri Drake ex Jum. & H.Perrier	Davis 1174 (K)	Madagascar	DQ153500	DQ153749	DQ153794	DQ153631
Coffea pervillenana (Baill.) Drake	Davis 2328 (K)	Madagascar	DQ153412	DQ153661	DQ153779	DQ153532
Coffea pocsii Bridson	Mvungi 7 (DSM, K)	Tanzania	DQ153453	DQ153702	DQ153820	DQ153581/ DQ153582
Coffea pseudozanguebariae Bridson	Mvungi 16 (DSM, K)	Tanzania	DQ153450	DQ153699	DQ153817	DQ153578
Coffea racemosa Lour.	Hepper & Maley 7717 (BR, K)	Mozambique	DQ153464	DQ153713	DQ153831	DQ153595
Coffea rakotonasoloi A.P. Davis	Davis 2265 (K)	Madagascar	DQ153416	DQ153665	DQ153783	DQ153536
Coffea ratsimamangae J.-F. Leroy ex A.P. Davis & Rakotonas.	Davis 2240 (K)	Madagascar	DQ153444	DQ153693	DQ153811	DQ153572
Coffea resinosa (Hook.f.) Radlk.	Davis 1103 (K)	Madagascar	DQ153428	DQ153677	DQ153795	DQ153555
Coffea rhamnifolia (Chiov.) Bridson	Friis *et al.* 4908 (K, BR, P)	Somalia	DQ153458	DQ153707	DQ153825	DQ153589
Coffea richardii J.-F. Leroy	Davis 2253 (K)	Madagascar	DQ153415	DQ153664	DQ153782	DQ153535
Coffea sahafaryensis J.-F. Leroy	Davis 2345 (K)	Madagascar	DQ153413	DQ153662	DQ153780	DQ153533
Coffea sakarahae J.-F. Leroy	Davis 2167 (K)	Madagascar	DQ153439	DQ153688	DQ153806	DQ153567
Coffea salvatrix Swynn. & Philipson	IRD-Montpelier LA 51 (K)	Mozambique	DQ153491	DQ153740	DQ153858	DQ153622
Coffea sambavensis J.-F. Leroy ex A.P. Davis & Rakotonas.	Davis 2323 (K)	Madagascar	DQ153418	DQ153667	DQ153785	DQ153538
Coffea schliebenii Bridson	Mbago 2256 (DSM)	Tanzania	DQ153456	DQ153705	DQ153823	DQ153587
Coffea sessiliflora Bridson	Mvungi 25 (DSM, K)	Tanzania	DQ153451	DQ153700	DQ153818	DQ153579
Coffea anthonyi Stoff. & F. Anthony, ined.	IRD-Montpelier OE 53 (K)	DR Congo	DQ153489	DQ153738	DQ153856	DQ153620
Coffea sp. 'G' (FTEA)	Mabberley 1417 (K)	Tanzania	DQ153474	DQ153723	DQ153841	DQ153605
Coffea stenophylla G.Don	Hepper & Maley 7723 (K)	Ivory Coast	DQ153466	DQ153715	DQ153833	DQ153597
Coffea tetragona Jum. & H.Perrier	Davis 2318 (K)	Madagascar	DQ153406	DQ153655	DQ153773	DQ153526
Coffea togoensis Jum. & H.Perrier	Hall & Abbins 43367 (K)	Togo	DQ153476	DQ153725	DQ153843	DQ153607
Coffea tsirananae J.-F. Leroy	Davis 2215 (K)	Madagascar	DQ153443	DQ153692	DQ153810	DQ153571
Coffea vianneyi J.-F. Leroy	Davis 2320 (K)	Madagascar	DQ153436	DQ153685	DQ153803	DQ153564
Coffea vatovavyensis J.-F. Leroy	Davis 2316 (K)	Madagascar	DQ153410	DQ153659	DQ153777	DQ153530
Coffea zangueberiae Lour.	Groenendijk 884 (K)	Mozambique	DQ153475	DQ153724	DQ153842	DQ153606
Psilanthus bridsoniae Sivar., Biju & P.Mathew	Biju & Sasi 44800 (K)	India	DQ153397	DQ153646	DQ153764	DQ153516
Psilanthus ebracteolatus Heirn	Davis 3008 (K)	Cameroon	DQ153392	DQ153641	DQ153759	DQ153510
Psilanthus mannii Hook.f.	Maurin 1 (K)	Cameroon	DQ153393	DQ153642	DQ153760	DQ153511
Psilanthus sapinii De Wild.	Sapin s.n. (BR 0856914)	Congo-Kinshasa	DQ153394	DQ153643	DQ153761	DQ153512
Psilanthus semsei Bridson	Kisera 1473 (K)	Tanzania	DQ153395	DQ153644	DQ153762	DQ153513
Psilanthus sp. 'A' (FTEA)	Luke 10197 (K)	Tanzania	DQ153399	DQ153648	DQ153766	DQ153518
Psilanthus travancorensis (Wight & Arn.) J.-F. Leroy	Biju s.n. (K)	India	DQ153398	DQ153647	DQ153765	DQ153517
Tricalysia sp.	Davis 2173 (K)	Madagascar	DQ153400	DQ153649	DQ153767	DQ153519
Tricalysia verdcourtiana Robbr.	Mvungi 43 (DSM, K)	Tanzania	DQ153401	DQ153650	DQ153768	DQ153520

Herbarium abbreviations after Holmgren *et al.* (1990). Where several ITS types were isolated these are listed below with multiple GenBank accesion numbers.

TABLE 2. *Amplification primers for* trnL–F*, ITS,* rpl16 *and* accD–psa1

Locus	Primer	Primer sequence	Reference
trnL intron	Forward (c)	5′-CGAAATCGGTAGACGCTACG-3′	Taberlet *et al.* (1991)
	Reverse (d)	5′-GGGGATAGAGGGACTTGAAC-3′	
trnL–F IGS	Forward (e)	5′-GGTTCAAGTCCCTCTATCCC-3′	Taberlet *et al.* (1991)
	Reverse (f)	5′-ATTTGAACTGGTGACACGAG-3′	
ITS	Forward (17SE or 101)	5′-ACGAATTCATGGTCCGGTGAAGTGTTCG-3′	Sun *et al.* (1994)
	Reverse (26SE or 102)	5′-TAGAATTCCCCGGTTCGCTCGCCGTTAC-3′	
	Internal reverse (ITS 2)	5′-GCTGCGTTCTTCATCGATGC-3′	White (1990)
	Internal forward (ITS 3)	5′-GCATCGATGAAGAACGCAGC-3′	
rpl16	Forward (F71)	5′-GCTATGCTTAGTGTGTGACTCGTTG-3′	Jordan *et al.* (1996)
	Reverse (R1661)	5′-CGTACCCATATTTTTCCACCECGAC-3′	
	Internal forward	5′-GTAAGAAGTGATGGGAACGA-3′	Designed at Kew
	Internal reverse	5′-TCGTTCCCATCACTTCTTAC-3′	
accD–psa1 IGS	Forward (ACCD 769 F)	5′-GGAAGTTTGAGCTTTATGCAAATGG-3′	Mendenhall (1994)
	Reverse (PSA1 75 R)	5′-AGAAGCCATTGCAATTGCCGGAAA-3′	
	Internal forward Coffeeae	5′-GCTAAAAATCTCTSTTGGTTCGG-3′	Designed at Kew
	Internal reverse Coffeeae	5′-CCGAACCAASAGAGATTTTTAGC-3′	

primers were again designed to obtain satisfactory PCR products for the recalcitrant specimens.

The PCR program for *trnL–F* consisted of 2 min at 94 °C followed by 28 cycles of 1 min at 94 °C (denaturation), 1 min at 50 °C (annealing) and 1 min at 72 °C (extension), followed by a final 7 min extension (72 °C). For *rpl16*, the PCR program used was 2 min at 94 °C followed by 30 cycles of 1 min 94 °C, 1 min at 52 °C and 3 min at 72 °C, with a final extension at 72 °C for 7 min. The *accD–psa1* region was amplified using the PCR program: a denaturation phase of 3 min (94 °C), followed by 30 cycles of 1 min at 94 °C, 1 min at 51 °C and 1 min at 72 °C, and a final extension of 72 °C for 5 min. The ITS region was amplified using primer 17SE forward and 26SE reverse from Sun *et al.* (1994). Dimethylsulfoxide (DMSO; 4·5 %) was used to reduce secondary structure problems common with nuclear DNA.

A PCR mastermix containing 2·5 mM $MgCl_2$ (Advanced Biotechnologies, Ltd) was used for *trnL–F* and *rpl16* amplifications. For *accD–psa1*, commercial mastermix did not give good amplifications, and so a pre-mix was prepared using Biotaq DNA polymerase (Bioline, UK), 10× NH_4 reaction buffer (Bioline, UK), 50 mM $MgCl_2$ and dNTPs (Promega, Maddison, WI, USA). Amplified products were purified using QIAquick purification columns (QIAgen) as described in the manufacturer's protocol. Cycle sequencing reactions were carried out using BigDye™ Terminator Mix (Applied Biosystems, Inc., Warrington, Cheshire, UK). The program consisted of 26 cycles of: 10 s denaturation (96 °C), 5 s annealing (50 °C) and 4 min elongation (60 °C). PCR and sequencing reactions were run using a Perkin-Elmer GenAMP™ model 9600 or 9700 PCR system, and sequencing products were run on either an ABI 3100 Genetic Analyzer or an ABI 377 automated sequencer according to the manufacturer's protocols (Applied Biosystems, Inc.). Electropherograms were edited and assembled into contigs using Sequencher version 3·2·2. (Gene Codes Corp., Ann Arbor, MI, USA). The sequences generated were submitted to GenBank using the Sequin Application (version 5·26; available from http://www.ncbi.nlm.nih.gov/Sequin/).

Data matrix composition and parsimony analysis

All sequences were aligned manually in PAUP* (version 4·0b10; Swofford, 2002) without difficulty due to low levels of sequence variation. Areas of ambiguous alignment were excluded from the analysis, as were regions with missing sequences, for example the beginning and end of sequences and around the internal primer-binding sites.

Maximum parsimony was implemented to analyse (*a*) *trnL–F*, (*b*) *rpl16*, (*c*) *accD–psa1*, (*d*) combined plastid data, (*e*) ITS and (*f*) combined sequence data, using PAUP*. In all analyses, gaps were treated as missing data and characters were equally weighted and unordered (Fitch, 1971). All data sets were analysed separately and examined by eye in order to identify topological conflict, i.e. moderate to strong support for placement of a taxon in different clades. Thirteen insertions/deletions were identified and scored for the plastid DNA regions.

Tree searches were conducted using 10 000 replicates of random taxon sequence addition, retaining ten trees at each step, with tree-bisection-reconnection (TBR) branch swapping, delayed transformation (DELTRAN) optimization, MulTrees in effect, and saving a maximum of ten trees per replicate. Support for clades in all analyses was estimated using bootstrap analysis (Felsenstein, 1985), with 10 000 replicates of full heuristic search, simple sequence addition, TBR swapping, with MulTrees in effect and saving a maximum of ten trees per replicate. BPs are described as strongly/well supported (85–100 %), moderate (75–84 %) or low (50–74 %). Support was also estimated by calculating Bremer support values (*b*) (Bremer, 1988, 1994; Källersjö *et al.*, 1992), otherwise known as decay values. These were obtained using PAUP* (Swofford, 2002), in conjunction with AutoDecay 4·0·2 (Eriksson, 1999), with 100 replicates of random addition for each constraint tree.

Map construction

Figure 1 is based on the distribution of individual species as recorded in an African *Coffea* specimen database

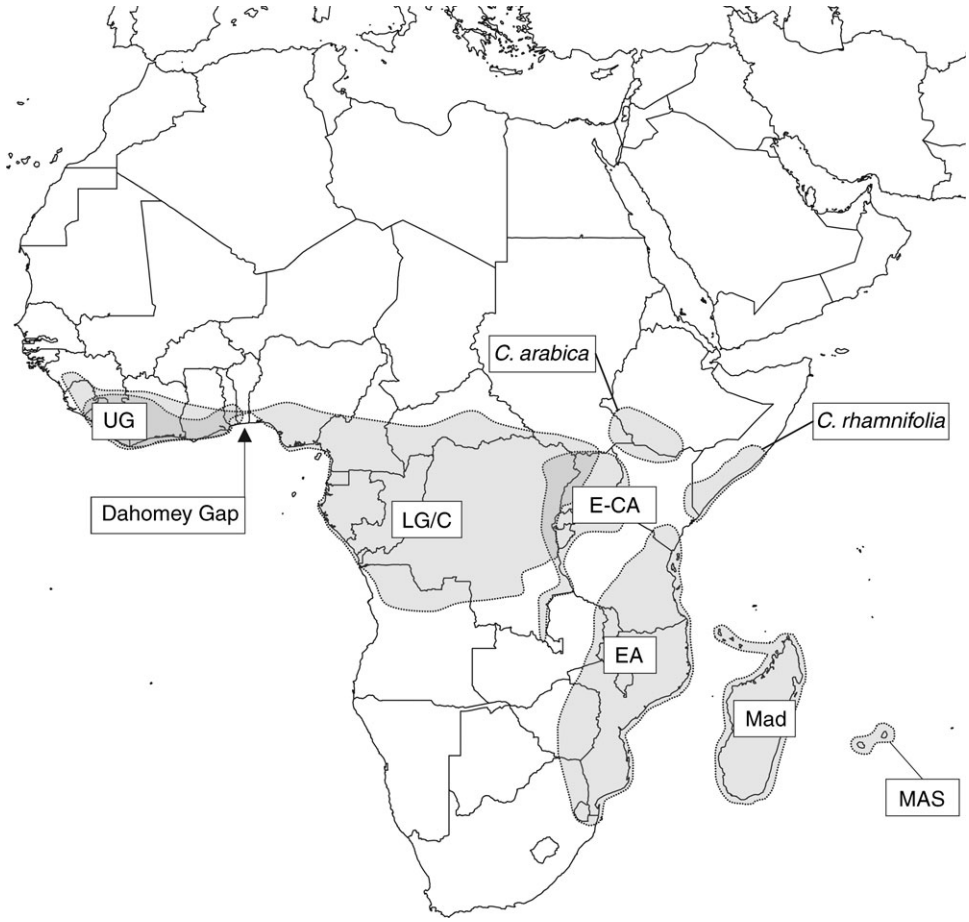

FIG. 1. Distribution map of *Coffea* showing location of clades and groups. UG = Upper Guinea clade; LG/C = Lower Guinea/Congolian clade; E-CA = East-Central Africa clade; EA = East Africa clade; Mad = Madagascan species; MAS = Mascarene clade. The map does not indicate the distribution of the poorly known *C. anthonyi* ined. (see Materials and Methods).

(approx. 2400 records; P. Stoffelen and A. Davis, unpubl. data) and Madagascan/Mascarene *Coffea* specimen database (approx. 1000 records; A. Davis and S. Dawson, unpubl. data). Species distributions maps were plotted and then a generalized map was drawn by hand.

RESULTS

Tree data and statistics for individual and combined analyses using (*a*) *trnL–F*, (*b*) *rpl16*, (*c*) *accD–psa1*, (*d*) combined plastid data, (*e*) ITS and (*f*) combined sequence data are given in Table 3. Individual plastid sequence analyses were topologically consistent (negligible to zero incongruence) and for the purpose of the results and discussion were combined and treated as a single analysis. The combined plastid analysis is largely congruent with the ITS analysis, apart from the placement of *C. arabica*, an accession of *C. heterocalyx* [IRD-Montpellier JC 66 (K); see Table 1], and three species from the Upper Guinea region (*C. humilis*, *C. stenophylla*, *C. togoensis*). *Coffea arabica* and the accession of *C. heterocalyx* were the only strongly supported points of incongruence. It is now generally accepted that *C. arabica* is of hybrid origin, as discussed in the Introduction. On the basis of the present results, it is believed that the accession of *C. heterocalyx* [IRD-Montpellier JC 66 (K)] used here is the same as that sampled by Lashermes *et al.* (1997) and Cros *et al.* (1998). The combined plastid data analysis places *C. heterocalyx* with *C. liberica* var. *dewevrei*, and the ITS analysis with *C. eugenioides*. Examination of the plastid sequences shows that there are only 3 bp differences and one 11 bp deletion (for *C. liberica* var. *dewevrei*) between this accession of *C. heterocalyx* and *C. liberica* var. *dewevrei* [Hepper & Maley 7729 (BR, K, MO)] across the three plastid regions. Cros *et al.* (1988) reported that their *trnL–trnF* sequence of *C. heterocalyx* (no accession data) was identical to one of their samples of *C. liberica*. The present ITS sequence data show that there is only 1 bp difference between this sample of *C. heterocalyx* and *C. eugenioides* [Harley 9332 (BR, K)]. The ITS2 data of Lashermes *et al.* (1997) placed *C. heterocalyx* in an unresolved position within a clade containing four samples of *C. eugenioides* and two of *C.* sp. Moloundou (= *C. anthonyi* ined.). Based on these data, it is believed that this accession of *C. heterocalyx* is a hybrid between *C. eugenioides* and *C. liberica*, resulting from either introgression in the wild or a chance crossing in cultivation. The natural distributions of these taxa overlap in the wild

TABLE 3. *Description of trees for each plastid region, combined plastid region and combined molecular data sets*

Characteristics	*trnL–F*	*rpl16*	*accD–psaI*	Combined plastid data	*ITS*	Combined molecular data
Number of taxa	95	95	95	95	107*	106†
Total number of characters	915	1120	1187	3222	831	3901
Invariable characters	837	995	1052	2884	652	3395
Parsimony uninformative characters	43	73	82	198	72	251
Parsimony informative characters	35	52	53	198	107	255
Tree length	90	106	199	499	415	952
Consistency index (CI)‡	0·808	0·677	0·594	0·563	0·438	0·467
Retention index (RI)	0·941	0·867	0·849	0·809	0·764	0·761
Number of trees	18 753	6860	6290	75 120	7760	11 520

* Including ITS clones.
† Including ITS clones but *C. arabica* removed (see Materials and Methods and Results).
‡ Calculated without uninformative sites.

[Democratic Republic of Congo, Sudan, Uganda (Davis *et al.*, 2006; Fig. 1)], but there do not appear to be field data indicating wild hybrids between *C. liberica* and *C. eugenioides*. The sequences of the sample of *C. heterocalyx* taken directly from the wild [Maurin 23 (K)] in Cameroon do not closely match those of the accession IRD-Montpellier JC 66 (K), as the former is consistently placed with species from West Africa (see Fig. 4). The sample IRD-Montpellier JC 66 (K) was originally held at the IRD coffee breeding station of Divo, Ivory Coast (ex IRD-IFCC station) and then transferred to IRD-Montpellier, France. The accession data imply that it was collected from either Cameroon or the Democratic Republic of Congo during the 1960s; it has been maintained in cultivation for >40 years (F. Anthony, pers. comm.).

Based on the evidence given above, the Montpellier accession of *C. heterocalyx* was excluded from the analyses, and *C. arabica* was removed from the final combined analysis (combined plastid plus ITS). After the deletion of these species, any highly supported incongruence between the combined plastid analysis and the ITS analysis was removed, enabling these two data sets to be combined. The position of three species (*C. humilis*, *C. stenophylla* and *C. togoensis*) from the Upper Guinea region is different for the combined plastid analysis and ITS analysis. In the combined plastid analysis, they form a well-supported clade (the UG clade; BP 100, $b = 5$; Fig. 2), which falls within a clade of species from East and East-Central Africa (BP 70, $b = 1$; Fig. 2). In the ITS analysis, these three species do not form a clade, but are all positioned within a clade of species that contains species from the Lower Guinea/Congo region (the 'canephora alliance'; see below) and *C. arabica* (BP 74, $b = 1$; Fig. 3). Similar results were reported by Cros *et al.* (1998) based on their *trnL–trnF* data for *C. humilis* and *C. stenophylla*, and by comparison with the ITS data of Lashermes *et al.* (1987) for *C. stenophylla* (see Introduction). The combined molecular analysis places the UG clade in approximately the same position as the combined plastid analysis, although this relationship is very weakly supported (BP 23, $b = 1$; Fig. 4). The incongruence identified in the present investigation is not well supported, and the above three species were retained in combined plastid and ITS analysis, where their inclusion does not significantly influence either the topology of the tree or support values. In the combined molecular (plastid–ITS) analysis, the UG clade is in approximately the same position as the combined plastid analysis, although this relationship is only weakly supported (BP 23, $b = 1$; Fig. 4).

Negligible to moderate resolution was found in some parts of analyses, particularly for the Madagascan taxa. This lack of resolution was mostly due to low levels of sequence divergence, as indicated by the values for the consistency index (CI) and retention index (RI), and review of branch lengths (see Fig. 5).

Terminology of clades

Many of the species groups recovered in the present analyses are consistent, or nearly so, with geographical or phytogeographical regions, and geographical abbreviations have been used for clades (cf. Lashermes *et al.*, 1997; Cros *et al.*, 1998). Where the analyses recovered groups congruent with the most recent infrageneric classification of *Coffea* and *Psilanthus* (Davis *et al.*, 2005, 2006), these taxonomic groupings were retained. The following terminology was used for the geographical groupings: Upper Guinea (UG) clade, Lower Guinea/Congolian (LG/C) clade, East Africa-Indian Ocean (EA-IO) clade, East-Central Africa (E-CA) clade, East Africa (EA) clade, and Mascarenes (MAS) clade (see Fig. 1). The EA-IO clade includes the E-CA, EA and MAS clades, and species from Madagascar. The humid Central and West African forests are contained within the Guineo-Congolian Regional Centre of Endemism (White, 1983). Within this major region there are three subcentres of endemism for humid forest species: (1) Upper Guinea; (2) Lower Guinea; and (3) Congolian (White, 1979). For practical purposes, the subcentres (2) and (3) are often put together as the Lower Guinean/Congolian region, and this convention has been followed here. The distribution and systematic positions of *C. arabica* and *C. rhamnifolia* are isolated and are treated independently.

There is considerable agreement between the geographical distribution of species and their placement within clades or assumed species groupings. There is 100 % endemicity

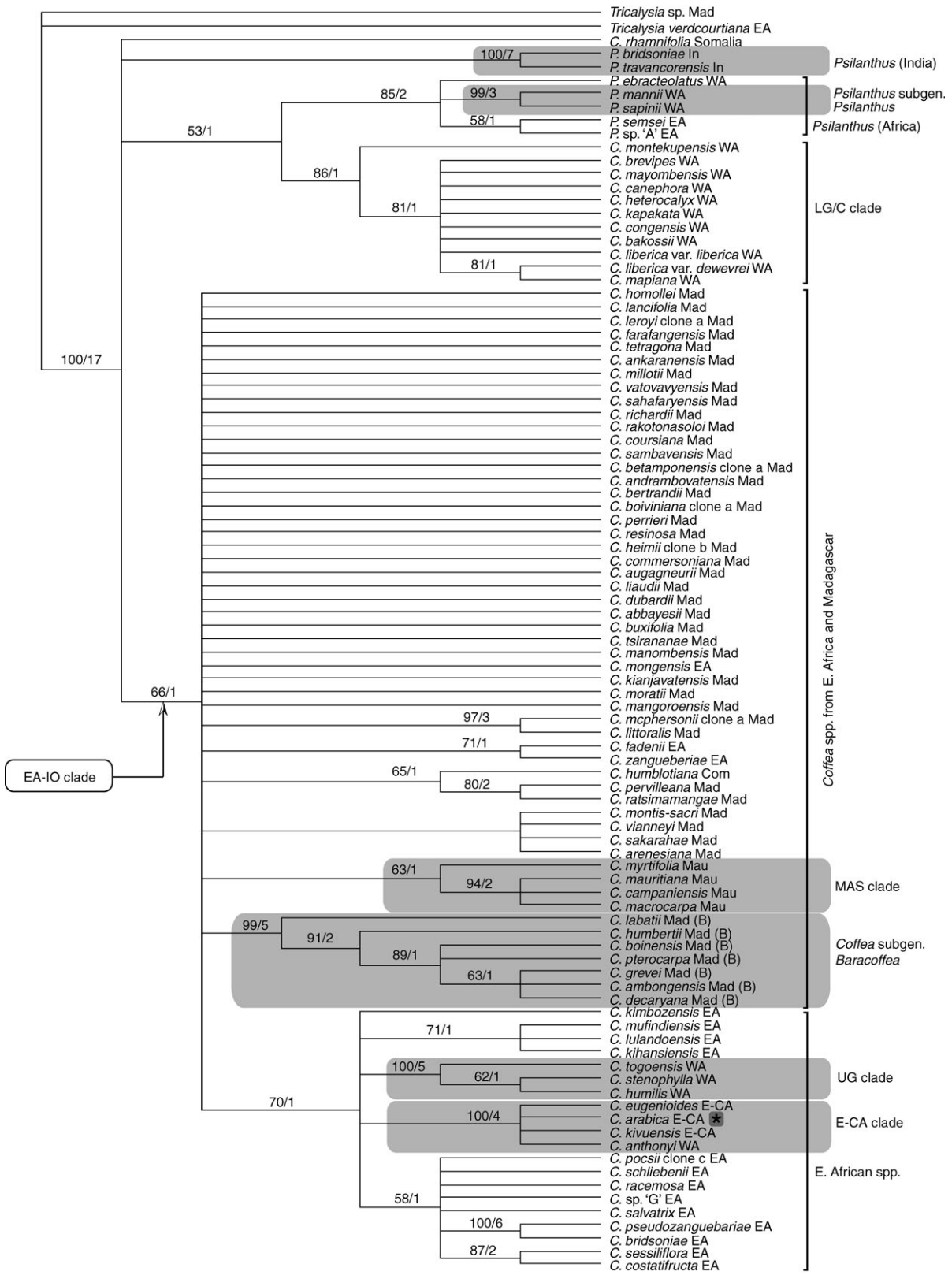

FIG. 2. Strict consensus tree generated from combined plastid analysis. Bootstrap values of >50 % are placed above the branches, followed by Bremer support (decay) values. See Table 1 for species authorities and provenance. EA-IO clade = East Africa-Indian Ocean clade; E-CA = East-Central Africa clade; LG/C clade = Lower Guinea/Congolian clade; UG clade = Upper Guinea clade; MAS clade = Mascarene clade. Regions, given after species names: EA = East Africa; E-CA = East Central Africa; Com = Comoros; In = India; Mad = Madagascar; Mau = Mauritius; WA = West Africa. * = Reference symbol for *C. arabica*. (B) = species belonging to *Coffea* subgenus *Baracoffea*.

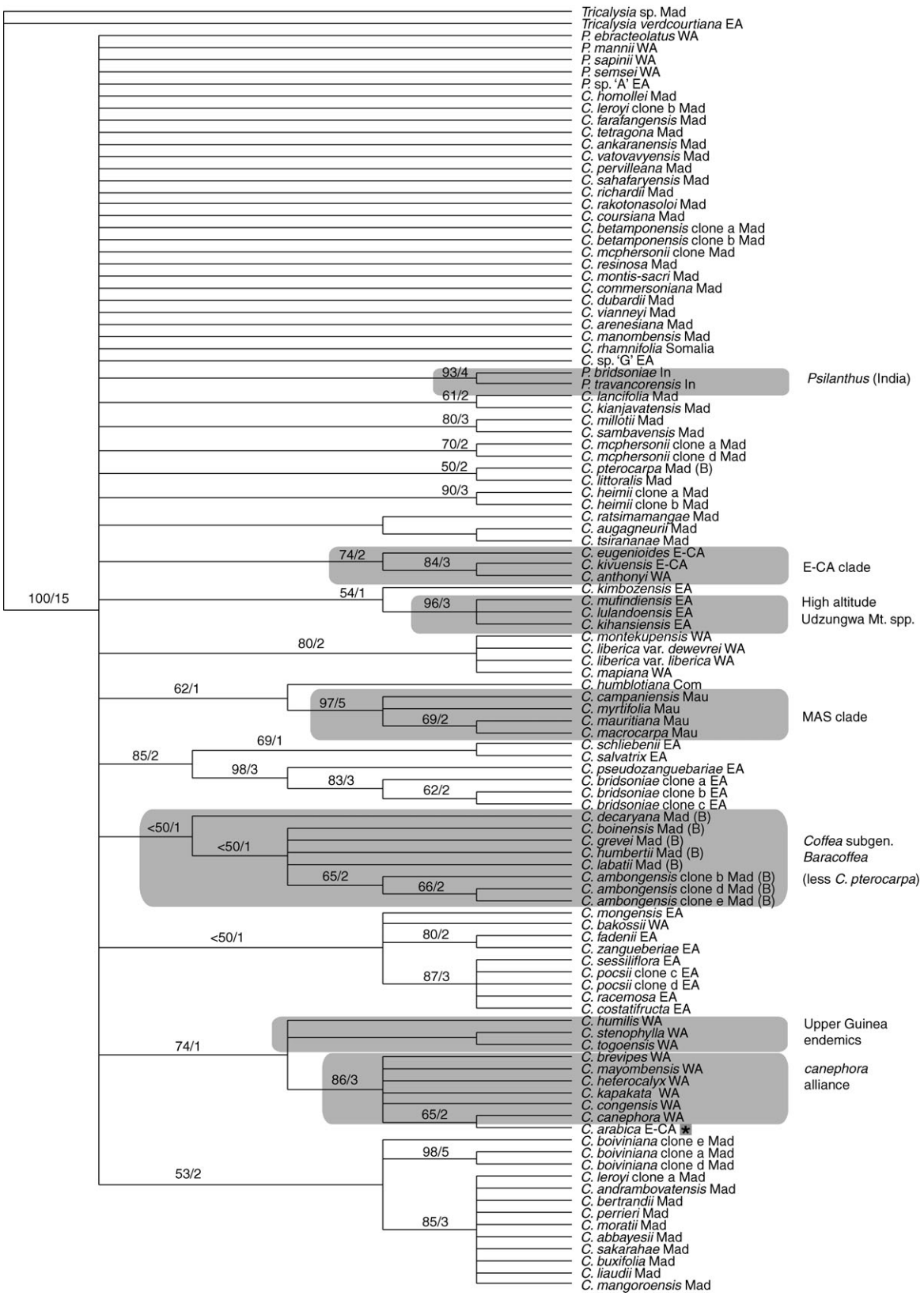

FIG. 3. Strict consensus tree generated from ITS analysis. Bootstrap values of >50 % are placed above the branches, followed by Bremer support (decay) values. See Table 1 for species authorities and provenance. E-CA clade = East-Central Africa clade; MAS clade = Mascarene clade. Regions: EA = East Africa; E-CA = East Central Africa; Com = Comoros; In = India; Mad = Madagascar; Mau = Mauritius; WA = West Africa. * = Reference symbol for *C. arabica.* (B) = species belonging to *Coffea* subgenus *Baracoffea*.

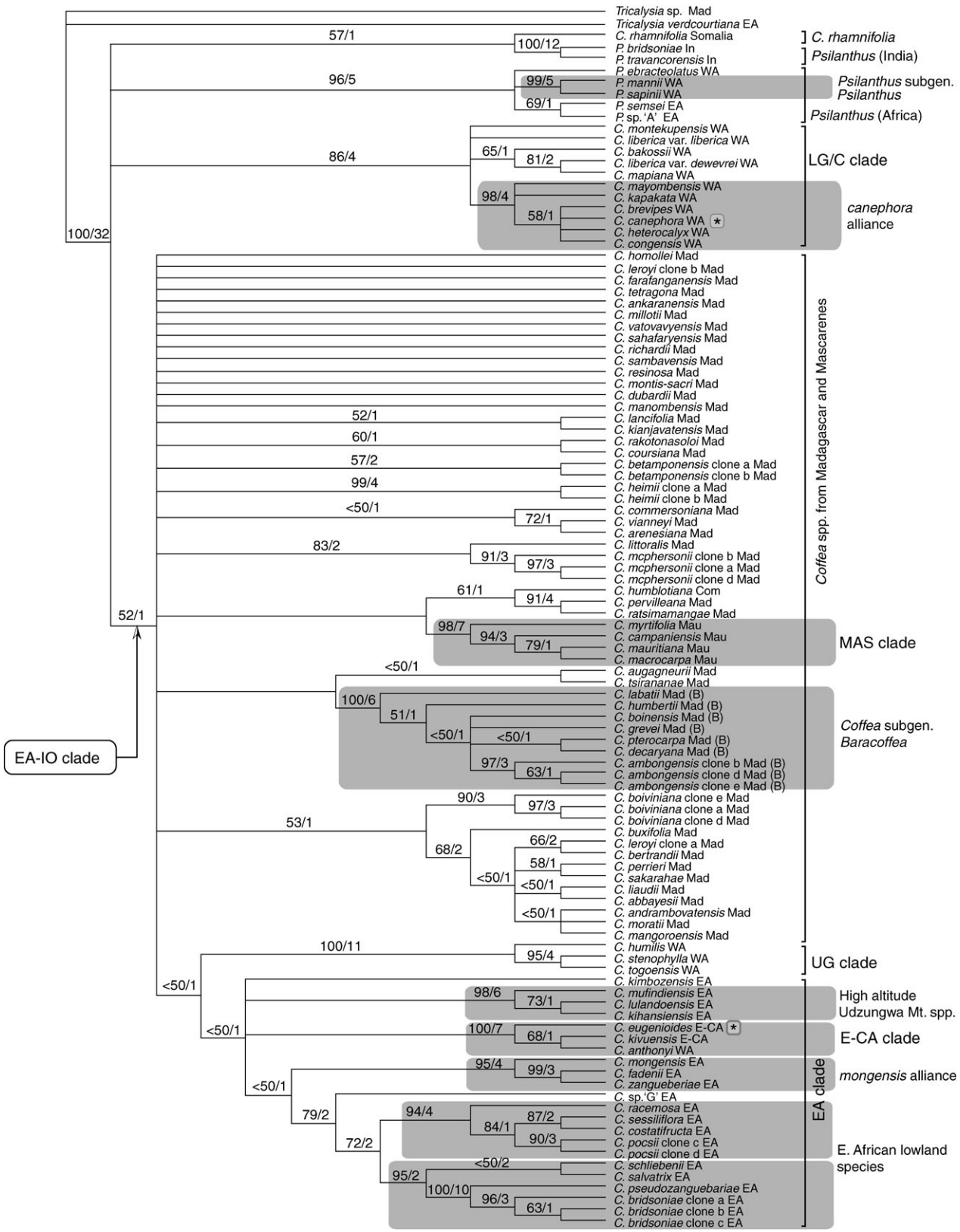

FIG. 4. Strict consensus tree generated from combined molecular (plastid–ITS) analysis. Bootstrap values of >50 % are placed above the branches, followed by Bremer support (decay) values. See Table 1 for species authorities and provenance. EA-IO clade = East Africa-Indian Ocean clade; EA clade = East Africa clade; E-CA clade = East-Central Africa clade; LG/C clade = Lower Guinea/Congolian clade; UG clade = Upper Guinea clade; MAS clade = Mascarene clade. Regions: EA = East Africa; E-CA = East Central Africa; Com = Comoros; In = India; Mad = Madagascar; Mau = Mauritius; WA = West Africa. *Denotes progenitor species for *C. arabica.* (B) = species belonging to *Coffea* subgenus *Baracoffea.*

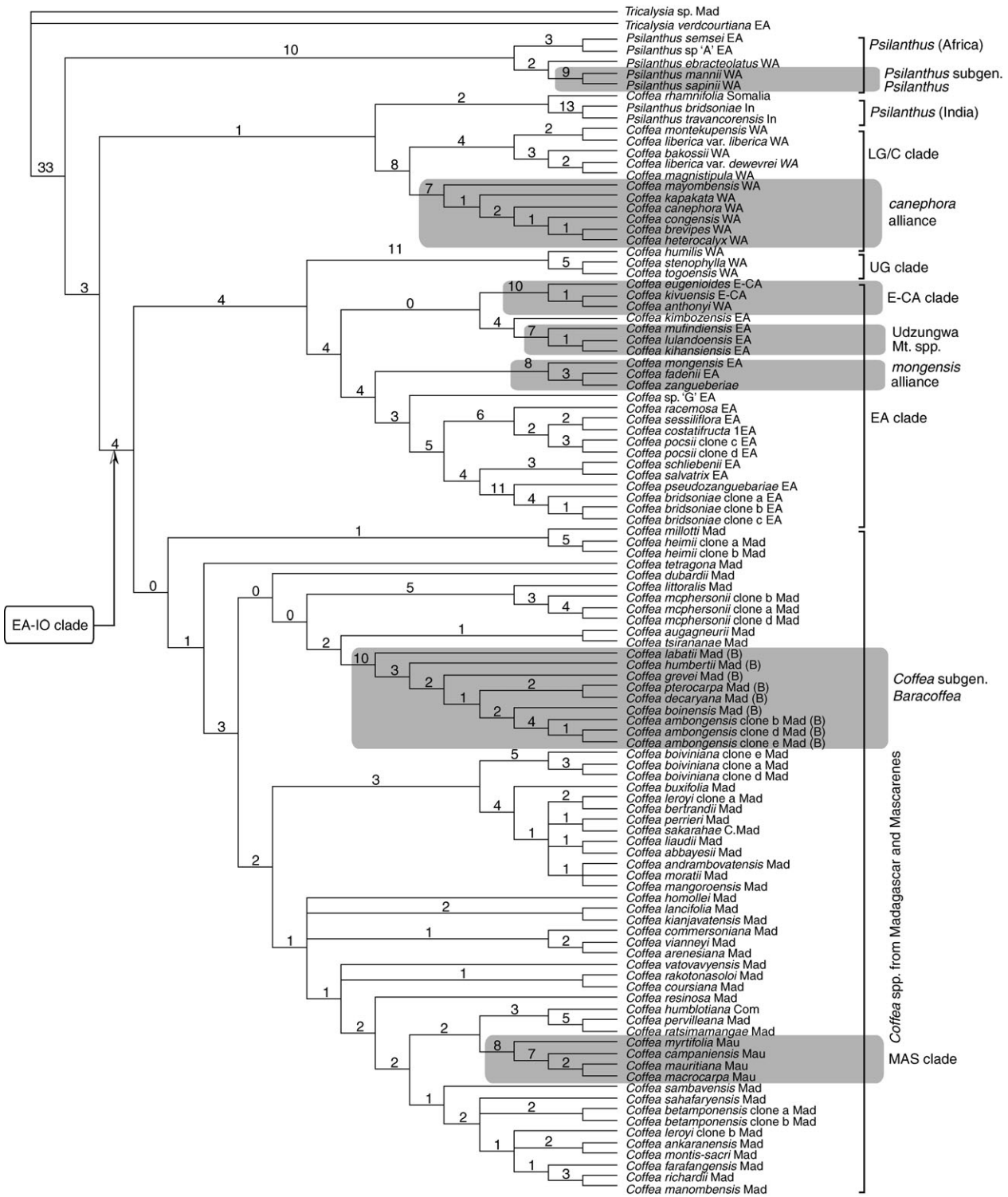

FIG. 5. One tree of 11 520 trees from the combined molecular (plastid–ITS) data analysis, with branch lengths. See Table 1 for species authorities and provenance. EA-IO clade = East Africa-Indian Ocean clade; E-CA clade = East-Central Africa clade; EA clade = East Africa clade; LG/C clade = Lower Guinea/Congolian clade; UG clade = Upper Guinea clade; MAS clade = Mascarene clade. Regions: EA = East Africa; E-CA = East Central Africa; Com = Comoros; In = India; Mad = Madagascar; Mau = Mauritius; WA = West Africa. (B) = species belonging to *Coffea* subgenus *Baracoffea*.

for the MAS, and EA clades. The UG, LG/C and E-CA clades do not have complete endemicity owing to the inclusion of widespread species: *C. canephora* and *C. liberica* occur in both the Upper Guinea (Portères, 1937) and Lower Guinea/Congolian regions and central-east Africa (see Fig. 1); *C. anthonyi* ined. occurs

in the East Central Africa and the Lower Guinea/Congolian regions. However, the natural range of *C. canephora* and *C. liberica* in West Africa has no doubt been obscured by introduction and naturalization, and the distribution of *C. anthonyi* is still poorly known. The Madagascan species do not form a clade, but there is 100 % endemicity for Madagascan *Coffea* species. Owing to the relatively low sample size for *Psilanthus*, and because this study is focused on *Coffea*, abbreviations have not been provided for well-supported clades within *Psilanthus*.

Single plastid analyses

The *accD–psa1* data set yielded the most potentially parsimony informative sites and *trnL–F* the least. Tree statistics for each separate plastid analysis are given in Table 3.

Combined plastid analysis (Fig. 2)

Several well-supported geographical groupings are revealed within *Coffea*, including the UG clade (BP 100; $b = 5$), LG/C clade (BP 86; $b = 1$), E-CA clade (including *C. arabica*) (BP 100; $b = 4$), *Coffea* subgenus *Baracoffea* (BP 99; $b = 5$) and clades within this subgenus. There are also several well-supported *Coffea* species pairs. The MAS clade is weakly supported (BP 63; $b = 1$), although three species within the MAS clade receive better support (BP 94; $b = 2$). The EA-IO clade is weakly supported (BP 66; $b = 1$). *Coffea arabica* is placed within the E-CA clade (BP = 100; $b = 4$), which is consistent with the placement in the Central Africa clade of Cros *et al.* (1998). The relationship of *C. arabica* to the other species in this clade (*C. eugenioides*, *C. kivuensis*, *C. anthonyi* ined.) is unresolved, although the sequences of *C. arabica* and *C. eugenioides* are identical. The monophyly of *Coffea* is not supported: *C. rhamnifolia* is unresolved in relation to other *Coffea* and *Psilanthus* species, and species from the Lower Guinea/Congolian region are sister to African *Psilanthus* (BP 53; $b = 1$). *Psilanthus* is unresolved, although African *Psilanthus* (BP 85; $b = 2$) and Indian *Psilanthus* (BP 100; $b = 7$) are both well-supported.

ITS analysis (Fig. 3)

The cloned ITS sequences from each species showing evidence of heterogeneous copies grouped together, with the exception of the two cloned sequences from *C. leroyi*. The problems associated with direct sequencing in these taxa thus appear to be due to incomplete homogenization of the ITS copies, rather than hybridization or other causes. Groupings within *Coffea* receiving high support include a group of species from the Lower Guinea/Congolian region (*C. brevipes*, *C. mayombensis*, *C. heterocalyx*, *C. congensis* and *C. canephora*), *C. arabica* (originating from Ethiopia) and *C. kapakata* (Angola) (BP 86; $b = 3$). These six species (less *C. arabica*) are here referred to as the canephora alliance. Other well-supported clades include a group of species from the Udzungwa Mountains of East Africa (BP 96; $b = 3$) and the MAS clade (BP 97; $b = 5$). There are also some well-supported species pairs and well-supported groups of ITS clones. The ITS analysis places *C. arabica* within the canephora alliance, sister to *C. canephora* (BP 65; $b = 2$), and with only 2 bp differences between these species. The two species of Indian *Psilanthus* are well supported (BP 93; $b = 4$). The relationship between *Coffea* and *Psilanthus* is unresolved.

Combined plastid and ITS analysis (Fig. 4)

Several well-supported groupings within *Coffea* are recovered, including the UG clade (BP 100; $b = 11$), LG/C clade (BP 86; $b = 4$), E-CA clade (BP 100; $b = 7$), the MAS clade (BP 98; $b = 7$), *Coffea* subgenus *Baracoffea* (BP 100; $b = 6$) and the canephora alliance (BP 98; $b = 4$). The EA-IO clade (BP 52; $b = 1$) and EA clade (BP < 50; $b = 1$) are consistently recovered, but only weakly supported. There are a number of strongly supported lineages of East African species [e.g. Udzungwa Mountains (BP 98; $b = 6$), and the mongensis alliance (BP 95; $b = 4$)], several well-supported species pairs and several strongly supported groupings of ITS clones. Indian *Psilanthus* (BP 100; $b = 12$) and African *Psilanthus* (BP 96; $b = 5$) form well-supported clades. Apart from *Coffea* subgenus *Baracoffea* (BP 100; $b = 6$) and some small groupings of species, the relationships between most Madagascan species are unresolved. In contrast to the well-supported monophyly of *Coffea* subgenus *Baracoffea*, subgenus *Coffea* is paraphyletic. *Psilanthus* subgenus *Psilanthus* (BP 99; $b = 5$) is well supported, but the monophyly of *Psilanthus* subgenus *Afrocoffea* was not substantiated. *Coffea rhamnifolia* is placed with the two species of Indian *Psilanthus*, but this is only weakly supported (BP 57; $b = 1$). As in the combined plastid analysis and ITS analysis, the relationship between *Coffea* and *Psilanthus* is largely unresolved.

DISCUSSION

West African Coffea*: the Upper Guinea (UG) clade*

The UG clade, comprising *C. humilis*, *C. stenophylla* and *C. togoensis*, is one of the most strongly supported groups within the combined analysis (BP 100, $b = 11$; Fig. 4). Cros *et al.* (1998) found good support (BP 100) for *C. stenophylla* and *C. humilis*, which they recognized as the west Africa (W) clade. The convincing phylogenetic support for this clade may well be due to isolation of the Upper Guinea forests, as they are separated from those of the Lower Guinea/Congo region by the Dahomey Gap (Fig. 1), otherwise known as the Dahomey Interval (White, 1979, 1983). The gap is an extension of the woodland savannah of the Sahel to the Gulf of Guinea (Poorter *et al.*, 2004) presently some 250 km wide (White, 1979), which reaches the coast of southern eastern Ghana, Togo and Benin. Booth (1958) suggested that the Dahomey Gap was much wider during periods of glacial aridification, and this is generally supported by more recent studies (e.g. Maley, 1987). The present-day distribution of *Coffea*

species does not show complete separation between the Lower Guinea/Congo region and the Upper Guinea region, as *C. canephora* and *C. liberica* occur in both regions, and *C. togoensis* occurs in isolated humid forest patches in Ghana, Togo and Benin (Sonké *et al.*, 2006*a*). However, the Dahomey Gap has clearly played a role in the evolution of plant (White, 1983) and animal species (e.g. Booth, 1958; Murphy and Collier, 1997) and populations (Sehgal *et al.*, 2005) in the Upper Guinea region. It is proposed that this is first phylogenetic study to support the evolutionary influence of the Dahomey Gap within the flowering plants.

The incongruent position of the UG clade (see Results; Figs 2 and 3) is most probably due to (maternal) plastid genome transfer, which may have pre-dated speciation in this clade. The data imply plastid capture from a species or species lineage progenitor of East African origin or affiliation. Given the geographical location of the UG clade, the movement of a plastid genome from East to West Africa is most probable, perhaps either via long-distance dispersal or through dispersal via a once-continuous forest link between East and West Africa.

West Africa Coffea*: Lower Guinea/Congo (LG/C) clade*

The LG/C clade (BP 86, $b = 1$, Fig. 2; BP 86, $b = 4$, Fig. 4) is a group of ten predominantly lowland rainforest species, largely restricted to the Lower Guinea/Congolian region (Fig. 1). It does, however, include the widespread *C. canephora* and *C. liberica*, which also occur in the Upper Guinea region, and *C. kapakata* from Angola. *Coffea kapakata* occurs in the humid evergreen forests enclaves of the Guinea–Congolian/Zambezia Transition Zone (White, 1979), which is otherwise generally covered by non-forest vegetation. These enclaves are composed almost exclusively of humid forest Guineo-Congolian species (White, 1979), and thus the systematic position of *C. kapakata* within the present analysis is concomitant with humid forest distribution patterns in West Africa (White, 1979, 1983). The LG/C clade supports the findings of Cros *et al.* (1998, fig. 2), who recognized a west and central African (WC) clade (BP 41), which included five species and four provisional/unknown taxa from the Lower Guinea/Congolian region.

Within the Lower Guinea/Congo region, the canephora alliance [*C. brevipes*, *C. mayombensis*, *C. heterocalyx*, *C. congensis*, *C. canephora*, and *C. kapakata* (BP 98, $b = 4$; Fig. 4)] represents an expansion of the 'canephoroid group' (*C. brevipes*, *C. canephora*, and *C. congensis*) as enumerated by Cros *et al.* (1998). Members of this alliance are all very similar morphologically, and (based on single species samples) appear closely related. Recognition of the canephora alliance is important as it provides a well-circumscribed group for further study of the economically important species *C. canephora* and *C. arabica*. Other Lower Guinea/Congo *Coffea* species (*C. carrissoi* A.Chev., *C. dactylifera* Robbr. & Stoff., *C. fotsoana* Stoff. & Sonké and *C. leonimontana* Stoff.) may belong in the canephora alliance; material of these species were not sampled in the present analyses.

Humid forest West African species and species groups have often been considered to represent the earliest diverging lineages within plant genera (Harris *et al.*, 2000; Davis *et al.*, 2002; Plana *et al.*, 2004) or to include species that are phylogenetically isolated (Malcomber, 2002), although systematic studies for African plants are relatively few (Plana *et al.*, 2004). On the basis of the present data, it is not possible to support these assumptions: the combined plastid analysis places the LG/C clade as sister to African *Psilanthus* (BP 53, $b = 1$; Fig. 2) but in the combined molecular analysis (Fig. 4) its position is unresolved. Davis *et al.* (2002) posited that the response of ancient African plant communities to climate change should be detectable through phylogenetic analysis of plants that span both humid and lowland xeric regions of the African continent. This is based on the significant evidence that lowland rainforest dominated much of Africa in the late Cretaceous and was replaced by xeric vegetation as a response to continental uplift (and other Earth events) and consequent widespread aridification beginning in the late Palaeogene. They suggest that if aridification induced a relatively recent period of diversification, then species that inhabit the humid relict forests of West Africa should represent the earliest diverging lineages of these African radiations, whereas species restricted to arid regions of East Africa should be phylogenetically nested (Davis *et al.*, 2002). This pattern is not retrieved in the present analyses, although certain xeric species (*C. costatifructa, C. zanguebariae*, *C. racemosa*, *C. pocsii* and *C. schliebenii*) are convincingly nested within humid-dwelling East African lineages (Fig. 4). *Coffea rhamnifolia*, a species from SE Somalia and NE Kenya, and occurring in a xericenvironment, is intriguing in this respect as it occupies an isolated and equivocal position within the analyses (Figs 2–5).

In the ITS analysis, *C. liberica* var. *dewevrei* is grouped with three other Lower Guinea/Congolian taxa (BP 80, $b = 2$; Fig. 3): *C. liberica* var. *liberica*, *C. montekupensis* and *C. mapiana*, but the relationship among these taxa is unresolved. In the combined plastid analysis (BP 81, $b = 1$; Fig. 2) and combined molecular analysis (BP 81, $b = 2$; Fig. 4), it groups together with the morphologically unusual *C. mapiana* (Sonké *et al.*, 2006*b*). Thus, the results add further evidence in support of the findings of N'Diaye *et al.* (2005), indicating genetic differentiation between the two varieties of *C. liberica*. Further data and sampling are required to assess fully the relationships between the two varieties of *C. liberica* and related taxa in the LG/C clade.

East African–Indian Ocean Coffea*: EA-IO clade*

The EA-IO clade is consistently retrieved in the combined plastid analysis (BP 66, $b = 1$; Fig. 2) and combined molecular analysis (BP 52, $b = 1$; Fig. 4), although the support for this clade is weak. The West African UG clade is placed within the EA-IO clade in these analyses, but this position may be due to (presumably maternal) plastid genome transfer via either dispersal or diffusion (see above). The separation of West African *Coffea* species (the LG/C clades) and East African and Indian Ocean species (EA-IO clade) would be expected given

the geological history of Africa. The formation of the East Albertine African Rift Valley would have provided a climatic and physical obstruction to dispersal, separating a west/east humid forest belt that once spanned the African mainland (Maley, 1987; Davis *et al.*, 2002). In addition, further Neogene aridification would have cyclically separated and fragmented a continuous forest belt or larger forest blocks (Maley, 1987). On the basis of the current analysis, the separation of West Africa and East Africa *Coffea* would require a single vicariance event, such as a climatic incident or long-distance dispersal.

East Africa Coffea*: EA clade*

In the combined molecular analysis, all species from East Africa and East-Central Africa are placed within the EA clade (Fig. 4). There is negligible support for this group (BP 23; $b = 1$), although deleting the UG clade from the combined molecular analysis gives the EA increased support (e.g. BP 52). Lashermes *et al.* (1997) and Cros *et al.* (1998) also recognized an East Africa (E) clade (see Introduction), although their analyses included only four and five taxa, respectively.

Within the EA clade there are other well-supported groups that have geographical/ecological or morphological correspondence. The clade formed by *C. mufindiensis*, *C. lulandoensis* and *C. kihansiensis* (BP 98, $b = 6$; Fig. 4) represents a group of high altitude (800–2300 m) species from the Udzungwa Mountains, one of the mountain groups within the Eastern Arc Mountains (Lovett, 1985, 1988). *Coffea kimbozensis* is also found in the Udzungwa Mountains, but unlike the species above it is restricted to low elevation (300–450 m) on calcareous rocks (Bridson, 1988*a*; A. Davis and E. Mvungi, pers. observ.). The clade formed by *C. mongensis*, *C. fadenii* and *C. zanguebariae* (BP 95, $b = 4$; Fig. 4) does not have a distinct geographical or geological delimitation within East Africa, although they are similar morphologically (Bridson, 1988*a*), and they are labelled here as the 'mongensis group'. The E-CA clade (*C. eugenioides*, *C. kivuensis* and *C. anthonyi* ined.) is placed in an unresolved position at the base of the EA clade; further discussion of the E-CA clade is given below. The largest clade within the EA clade is a group of predominantly lowland species, labelled the 'E. Africa lowland species' (BP 72, $b = 2$; Fig. 4), and includes two well-supported groups (BP 94, $b = 4$; BP 95, $b = 2$). This clade contains mostly species from low elevations (sea level to approx. 500 m, rarely up to 800 m), in seasonally dry forest (some species in xeric woodland). These species mainly occur within the Indian Ocean Coastal belt (White, 1979). One exception is *C. salvatrix*, which normally occurs at altitudes of 850–1650 m (Bridson, 1988*a*) within the 'E. Africa lowland species' clade (Fig. 4).

East-Central Africa Coffea*: E-CA clade*

Coffea eugenioides, *C. kivuensis* and *C. anthonyi* ined. (as *C.* sp. Moloundou, Lashermes *et al.*, 1997; Cros *et al.*, 1998) form the E-CA clade in the combined plastid analysis (BP 100, $b = 4$; Fig. 2) and combined molecular analysis (BP 100, $b = 7$; Fig. 4). These results support the findings of Lashermes *et al.* (1997) and Cros *et al.* (1998), who received weak (BP 67) and strong (BP 100) support (respectively) for a central Africa (C) clade, based on *C. eugenioides* and *C. anthonyi*. The distribution of *C. eugenioides* and *C. kivuensis* falls mostly within the Lake Victoria Regional Mosaic (White, 1983), at elevations normally well above 1000 m, whereas *C. anthonyi* ined. occurs in a few isolated locations at low to mid-elevation [350–650(–900) m] in SE Cameroon and NW Congo. *Ex situ* material of *C. anthonyi* ined. was relied on for the DNA analysis, and further sampling would be desirable to test the close association of this species with *C. eugenioides* and *C. kivuensis*. The placement of the E-CA clade within the EA clade (Fig. 4) is consistent with the geographical proximity and the geological history of Africa (see above). Furthermore, clear associations between the Lake Victoria Regional Mosaic species and the Afromontane species of East Africa have been identified by White (1979); there are numerous species of *Coffea* occurring in the mountains of East Africa (Bridson, 1988*a*; Davis *et al.*, 2006).

Madagascan Coffea *species*

All Madagascan *Coffea* species are placed within a weakly supported EA-IO clade (BP 66, $b = 1$, Fig. 2; BP 52, $b = 1$, Fig. 4). The position of all Madagascan species and species groups is unresolved, due to low levels of sequence divergence (Fig. 4), which is a problem in species-level analysis of Madagascan Rubiaceae, as exemplified by the study of Malcomber (2002). *Coffea* subgenus *Baracoffea* (BP 100, $b = 6$; Fig. 4) is the only well-supported Madagascan clade, apart from *C. pervilleana* and *C. ratsimamangae* (BP 91, $b = 4$; Fig. 4), two closely allied species from northern Madagascar (Davis and Rakotonasolo, 2001*a*). When comparing morphological diversity of *Coffea* species occurring in Madagascar (Davis, 2001; Davis and Rakotonasolo, 2001*a*, *b*, 2003; Davis *et al.*, 2005) with those from Africa (Bridson, 1988*a*, 2003; Stoffelen, 1998), where sequence divergence is higher, the former exhibit far more interspecific differences, even excluding the unusual species of *Coffea* subgenus *Baracoffea* from western Madagascar (Davis *et al.*, 2005). The paucity of sequence divergence in Madagascan *Coffea* subgenus *Coffea* implies either a rapid evolutionary radiation or slow molecular evolution, or perhaps a combination of both.

Suggestions concerning the origin of Madagascan *Coffea* species must be tentative with the present data at hand. A single dispersal event from Africa, followed by insular speciation, has been inferred for *Begonia* L. (Begoniaceae) by Plana *et al.* (2004) and for *Gaertnera* Lam. (Rubiaceae) by Malcomber (2002). Lavin *et al.* (2000) infer that the Madagascan species of the genus *Ormocarpum* P.Beauv. (Fabaceae) are the result of two dispersal events. Investigation of *Streptocarpus* Lindl. (Gesneriaceae) shows multiple colonization events for the Madagascan representatives of the genus (Möller and Cronk, 2001).

A single dispersal event from Africa seems the most likely scenario for Madagascan *Coffea*, and one that would not be in conflict with the present data (Figs 3–5).

In the combined plastid (BP 65, $b = 1$; Fig. 2) and combined molecular (BP 61, $b = 1$; Fig. 4) analyses there is weak support for a sister relationship between *C. humblotiana* (the only *Coffea* species from the Comoros), and two species from Madagascar [*C. pervilleana* and *C. ratsimamangae* (northern Madagascar)], providing an indication for a northern Madagascan origin for the Comorian species.

Mascarene Coffea*: MAS clade*

In all analyses, the four *Coffea* species from the Mascarene Islands (Fig. 1) are placed within the EA-IO clade and form the MAS clade (BP 63, $b = 1$, Fig. 2; BP = 97, $b = 5$, Fig. 3; BP 98, $b = 7$, Fig. 4). Given that these islands are oceanic with a volcanic origin (the oldest approx. 8 million years old), the progenitor species must have arrived on the Mascarenes Islands via a long-distance dispersal event. The position of the MAS clade within the combined plastid and combined molecular analyses (Figs 2 and 5), coupled with the relative proximity of the Mascarenes to Madagascar, infers an 'out of Madagascar' origin for the MAS clade.

Taxonomic groups

The present study clearly shows that *Coffea* subgenus *Baracoffea* is a well-supported group (BP 100, $b = 6$; Fig. 4), restricted to the seasonal drylands (including spiny/xerophytic deciduous forest) of western Madagascar (Davis *et al.*, 2005). *Coffea* subgenus *Baracoffea* is by far the most morphologically distinct group within *Coffea*, having evolved as a response to a seasonally dry environment. Morphological features include congested or shrubby (sympodial growth pattern) habit, indeterminate inflorescences, deciduous leaves and pubescent to densely pubescent leaves and corollas (Davis *et al.*, 2005). In comparison, *Coffea* subgenus *Coffea* are trees (monopodial growth pattern), with determinate inflorescences, evergreen leaves (all except three species, see Davis *et al.*, 2005), glabrous or rarely very sparsely puberulous leaves, and glabrous corollas (Davis *et al.*, 2005). If it can be convincingly demonstrated that *Coffea* subgenus *Baracoffea* has evolved from humid forest ancestors/progenitors, which is possible (Figs 2, 4 and 5), assumptions about the origin of dryland biomes in Madagascar could be posited.

Coffea rhamnifolia cannot be considered a member of *Coffea* subgenus *Baracoffea*, as supposed by Leroy (1982): the combined plastid analysis and ITS analysis (Figs 2 and 3) place *C. rhamnifolia* in an unresolved position at the base of the ingroup, and the combined molecular analysis places it as sister to the two Indian *Psilanthus* (BP 57, $b = 1$; Fig. 4).

Psilanthus subgenus *Psilanthus* (*P. mannii* and *P. sapinii*) is well supported (BP 99, $b = 5$; Fig. 4) and definable by a single morphological synapomorphy: the presence of accrescent calyx lobes. The monophyly of the most species-rich group of *Psilanthus*, subgenus *Afrocoffea*, was not supported in the current analyses (Figs 2–4). Furthermore, the position and strong support for African (BP 85, $b = 2$, Fig. 2; BP 96, $b = 5$, Fig. 4) and Indian (BP 100, $b = 7$, Fig. 2; BP 93, $b = 4$, Fig. 3; BP 100, $b = 12$, Fig. 4) *Psilanthus* implies that the present subgeneric classification of this genus is not consistent with sequence data. Further sequence data and species sampling are required for *Psilanthus*.

The relationship between Coffea *and* Psilanthus

On the basis of ITS and *trnL–trnF* data, Lashermes *et al.* (1997) and Cros *et al.* (1998), respectively, concluded that the division of *Coffea* and *Psilanthus* into two genera was untenable. The present combined molecular analysis does not resolve the issues of monophyly for *Coffea* (including *C. rhamnifolia*) and *Psilanthus*, as the relationship between these genera is largely unresolved, mainly due to a lack of sequence divergence (Fig. 4). There is some support for inferring that *Coffea* and *Psilanthus* are not independent lineages: in the combined plastid analysis there is weak support (BP 53, $b = 1$, Fig. 2) for a sister relationship between African *Psilanthus* and West African *Coffea*, and in the combined molecular analysis *C. rhamnifolia* is weakly supported as the sister group to the two species of Indian *Psilanthus* (BP 57, $b = 1$; Fig. 4).

Three morphological characters separate *Coffea* from *Psilanthus* (Davis *et al.*, 2005): short to long filaments ($\pm$ absent to very short in *Psilanthus*); sub-medifixed anthers (supra-medifixed in *Psilanthus*); and (long) emergent style (short and included in *Psilanthus*). However, *P. melanocarpus* has short filaments and sub-medifixed anthers, as in *Coffea*, but an included style, as in *Psilanthus*. It was not possible to isolate DNA of *P. melanocarpus* from herbarium samples, and it was not included in the present analyses. If *P. melanocarpus* is indeed a species of *Psilanthus*, only one character would separate *Coffea* and *Psilanthus*: short vs. long style. If *P. melanocarpus* nested within *Coffea*, then the two anther characters would separate *Coffea* and *Psilanthus*. The number of pollen apertures (Lobreau-Callen and Leroy, 1980; Chinnappa and Warner, 1981; Stoffelen *et al.*, 1997*c*) has been used as additional evidence to separate *Coffea* and *Psilanthus* (Davis *et al.*, 2005), although considerable polymorphism is evident and there is overlap in the number of apertures between *Coffea* and *Psilanthus* and their subgenera (Stoffelen *et al.*, 1997*c*; Davis *et al.*, 2005). The morphological evidence for the separation of *Coffea* and *Psilanthus* is certainly not convincing, and if *P. melanocarpus* is found to nest within either of these genera there would seem negligible justification for the recognition of two genera on morphological grounds (see above). However, many other genera of Rubiaceae are separated on only one or two morphological characters. *Pagamea* Aubl. and *Gaertnera* are well-supported genera based on molecular data, but are separated by a single synapomorphy (pubescent vs. glabrous corolla lobes), for example (Malcomber, 2002).

The robust morphological (Robbrecht and Puff, 1986; Davis *et al.*, 2005) and molecular support for *Coffea* plus *Psilanthus* (Davis *et al.*, 2007), low sequence diversity between these genera (see above; Fig. 4) and indications of paraphyly (Figs 2 and 4), may be taken as evidence for accepting *Coffea* and *Psilanthus* as a single genus (Lashermes *et al.*, 1997; Cros *et al.*, 1998). However, it is believed that further molecular data are needed to resolve fully the relationship between *Coffea* and *Psilanthus*, and in particular sequence data are required for *P. melanocarpus* and other species of *Psilanthus*.

Origin of Coffea arabica

The present data support a hybrid origin for *C. arabica*, following the findings of Lashermes *et al.* (1997, 1999), Cros *et al.* (1998) and Raina *et al.* (1998). Examination of combined plastid and ITS analyses (Figs 2 and 3), and by pair-wise comparison of sequences (see Results), it is concluded that the progenitor species of *C. arabica* are *C. canephora* and *C. eugenioides*. The results concur with those of Lashermes *et al.* (1999), who found low genetic (RFLP) divergence between the two constituent genomes of *C. arabica* and those of its progenitor species, suggesting perhaps that the speciation of *C. arabica* took place relatively recently. The proposed hybrid origin of *C. arabica* is consistent with the potential recent sympatry of *C. canephora* and *C. eugenioides* based on present-day distribution (Davis *et al.*, 2006).

Conclusions

On the basis of sequences from four plastid regions and ITS, it has been possible to identify several well-supported lineages within *Coffea* that are consistent with major biogeographical regions and geographical areas, including the UG clade (BP 100; $b = 11$), the LG/C clade (BP 86; $b = 4$), the E-CA clade (BP 100; $b = 7$) and the MAS clade (BP 98; $b = 7$), although the distribution of widespread species (*C. canephora* and *C. liberica*) occurs across some of these regions in mainland Africa. The UG, LG/C and E-CA clades are consistent with those either retrieved or identified by Cros *et al.* (1998) and Lashermes *et al.* (1997). Within the LG/C clade, it has been possible to substantiate (Cros *et al.*, 1998) and expand on a group of species related to *C. canephora*, the canephora alliance (BP 98, $b = 4$; Fig. 4). Other groups were consistently retrieved but received weak support, including the EA-IO clade (BP 52; $b = 1$) and the EA clade (BP 23; $b = 1$). Smaller biogeographical groupings were retrieved within the EA clade, including a group of high altitude Udzungwa Mountain species (BP 98; $b = 6$), and the east African lowland species (BP 72; $b = 2$). Groups corresponding to geographical distribution were also recovered in *Psilanthus*, including the two Indian representatives, *P. travancorensis* and *P. bridsoniae* (BP 100; $b = 12$), and species of *Psilanthus* occurring in Africa (BP 96; $b = 5$). Two formal taxonomic groups are well supported: *Coffea* subgenus *Baracoffea* (BP 100; $b = 6$) and *Psilanthus* subgenus *Psilanthus* (BP 99; $b = 5$).

The combined sequence data do not substantiate the monophyly of either *Coffea* or *Psilanthus*, largely due a lack of resolution (Fig. 4), resulting from low levels of sequence divergence (e.g. see Fig. 5). This evidence, together with weak support for some intergeneric (*Coffea*–*Psilanthus*) relationships, and strong molecular and morphological support for a clade comprising *Coffea* plus *Psilanthus* [BP 100; $b = 32$; see also Davis *et al.* (2007)], could be taken as justification for the recognition of a single genus (i.e. *Coffea*), although further data and critical sampling are required to resolve this matter fully. The subgeneric classification of *Coffea* and *Psilanthus* (Bridson 1988*a*, *b*; Davis *et al.*, 2005) is not consistent with phylogenetic data. Most notably, *Coffea* subgenus *Coffea* is paraphyletic, due to the nested position of *Coffea* subgenus *Baracoffea*.

A recent hybrid origin for *C. arabica*, with *C. canephora* and *C. eugenioides* as the likely progenitor species, is supported (Lashermes *et al.*, 1999). This study provides the first phylogenetic evidence for the influence of the Dahomey Gap (West Africa) on plant speciation.

ACKNOWLEDGEMENTS

We would like to thank the following organizations and individuals who provided plant material for this study, including National Centre for Applied Research on Rural Development (FOFIFA), P. De Block (BR), IRD-Montpellier, Q. Luke (EA), F. Mbago (DSM), F. Rakotonasolo (TAN), B. Sonké (Université de Yaoundé) and P. Stoffelen (BR). Permission to sample and study herbarium and/or living collections from DSM, K and BR is gratefully acknowledged. We also thank the DNA bank managers of the Jodrell Laboratory Royal Botanic Gardens, Kew: L. Csiba, E. Kapinos. In Madagascar fieldwork was carried out under collaborative agreements between the Parc Botanique et Zoologique de Tsimbazaza (PBZT), the University of Antananarivo, the Association Nationale de Gestion des Aires Protegées (ANGAP) and the Royal Botanic Gardens, Kew. We would like to thank J. Andriantiana (TAN), H. Ralimanana (K), and especially F. Rakotonasolo (TAN) for help with fieldwork in Madagascar. We would like to acknowledge the Tanzania Commission for Science and Technology (COSTECH) for providing research clearance to undertake fieldwork in Tanzania. In Tanzania, we are grateful to F. Mbago (DSM), S. Hall, A. Ntemi, R. Hind and Z. Bridger. In Cameroon, we thank the staff and Directors at the Université de Yaoundé, National Herbarium of Cameroon (YA) and Limbe Botanical and Zoological Gardens (SCA), and especially B. Sonké (Université de Yaoundé). At the Royal Botanic Gardens, Kew, we would like to thank W. Baker, D. Bridson, S. Dawson, F. Forest, E. Lucas and J. Moat. At IRD-Montpellier we would like to thank F. Anthony, P. Hamon, A. de Kochko, P. Lashermes and M. Noirot. A.D. would like to thank the staff at the Royal Botanic Gardens, Sydney (NSW) and in particular D. Crayn. E.M. gratefully acknowledges the

DANIDA-ENRECA biodiversity project, for their financial support. The funding for this study was largely provided by Kraft Foods (UK), and we are extremely grateful for their financial support for laboratory and herbarium research and the visits of E.M. and Y.J.-F. to Kew.

LITERATURE CITED

Andreasen K, Bremer B. 2000. Combined phylogenetic analysis in the Rubiaceae–Ixoroideae: morphology, nuclear and chloroplast DNA data. *American Journal of Botany* **87**: 1731–1748.

Booth AH. 1958. The Niger, the Volta and the Dahomey Gap as geographical barriers. *Evolution* **12**: 48–62.

Bremer K. 1988. The limits of amino-acid sequence data in angiosperm phylogenetic reconstruction. *Evolution* **42**: 795–803.

Bremer K. 1994. Branch support and tree stability. *Cladistics* **10**: 295–304.

Bridson DM. 1979. The identity of *Paolia* (Rubiaceae). *Kew Bulletin* **34**: 376.

Bridson DM. 1983. The identities of *Plectonia rhamnifolia* Chiov. and *P. sennii* Chiov. (Rubiaceae). *Kew Bulletin* **38**: 320.

Bridson DM. 1988*a*. Coffea. In: Polhill RM, ed. *Flora of Tropical East Africa, Rubiaceae, 2*. Rotterdam/Brookfield: Balkema, 703–723.

Bridson DM. 1988*b*. Classification. In: Wrigley G, *Coffee*. New York, NY: Longmans, 61–75.

Bridson DM. 2003. *Coffea*. In: Pope GV, ed. *Flora Zambesiaca 3*. Surrey: Royal Botanic Gardens, Kew, 452–463.

Carvalho A. 1952. Taxonomia de *Coffea arabica* L. Caracteres morfológicos dos haploides. *Bragantia* **12**: 201–212.

Carvalho A, Medina Filho HP, Fazuoli LC, Guerreiro Filho O, Lima MMA. 1991. Aspectos genéticos do cafeeiro. *Revista Brasileira de Genética* **l4**: 135–183.

Chase MW, Hills HG. 1991. Silica gel: an ideal desiccant for preserving field-collected leaves for use in molecular studies. *Taxon* **41**: 215–220.

Cheek M, Csiba L, Bridson DM. 2002. A new species of *Coffea* (Rubiaceae) from western Cameroon. *Kew Bulletin* **57**: 675–680.

Chinnappa CC, Warner BG. 1981. Pollen morphology in the genus *Coffea* (Rubiaceae) and its taxonomic significance. *Botanical Journal of the Linnean Society* **83**: 221–223.

Coulibaly I, Noirot M, Lorieux M, Charrier A, Hamon S, Louarn J. 2002. Introgression of self-compatibility from *Coffea heterocalyx* to the cultivated species *Coffea canephora*. *Theoretical and Applied Genetics* **105**: 994–999.

Coulibaly I, Louarn J, Lorieux M, Charrier A, Hamon S, Noirot M. 2003. Pollen viability restoration in a *Coffea canephora* P. [*sic*] and *C. heterocalyx* Stoffelen backcross. QTL identification for marker-assisted selection. *Theoretical and Applied Genetics* **106**: 311–316.

Cros J. 1994. *Implications phylogénétiques des variations de l'ADN chloroplastique chez les caféiers (genres* Coffea *L. et* Psilanthus *Hook.f.).* PhD Thesis, Université Montpellier II, France.

Cros J, Combes MC, Trouslot P, Anthony F, Hamon S, Charrier A, Lashermes P. 1998. Phylogenetic analysis of chloroplast DNA variation in *Coffea* L. *Molecular Phylogenetics and Evolution* **9**: 109–117.

Davis AP. 2001. Two new species of *Coffea* L. (Rubiaceae) from eastern Madagascar. *Kew Bulletin* **56**: 479–489.

Davis AP. 2003. A new combination in *Psilanthus* (Rubiaceae) for Australia, and nomenclatural notes on *Paracoffea. Novon* **13**: 182–184.

Davis AP, Mvungi EF. 2004. Two new and endangered species of *Coffea* (Rubiaceae) from the Eastern Arc Mountains (Tanzania) and notes on associated conservation issues. *Botanical Journal of the Linnean Society* **146**: 237–245.

Davis AP, Rakotonasolo F. 2000. Three new species of *Coffea* L. (Rubiaceae) from Madagascar. *Kew Bulletin* **55**: 405–416.

Davis AP, Rakotonasolo F. 2001*a*. Three new species of *Coffea* L. (Rubiaceae) from NE Madagascar. *Adansonia, Sér. 3* **23**: 137–146.

Davis AP, Rakotonasolo F. 2001*b*. Two new species of *Coffea* L. (Rubiaceae) from Northern Madagascar. *Adansonia, Sér. 3*, **23**: 337–345.

Davis AP, Rakotonasolo F. 2003. New species of *Coffea* L. (Rubiaceae) from Madagascar. *Botanical Journal of the Linnean Society* **142**: 111–118.

Davis AP, Bridson DM, Rakotonasolo F. 2005. A reexamination of *Coffea* subgenus *Baracoffea* and comments on the morphology and classification of *Coffea* and *Psilanthus* (Rubiaceae-Coffeeae). In: Keating RC, Hollowell VC, Croat T, eds. *Festschrift for William G. D'Arcy: the legacy of a taxonomist* (Monograph in Systematic Botany 104). Missouri: MBG Press, 398–420.

Davis AP, Govaerts R, Bridson DM, Stoffelen P. 2006. An annotated checklist of the genus *Coffea* L. (Rubiaceae). *Botanical Journal of the Linnean Society* **152**: 465–512.

Davis AP, Chester M, Maurin O, Fay M. 2007. Searching for the relatives of *Coffea* (Rubiaceae, Ixoroideae): the circumscription and phylogeny of Coffeeae based on plastid sequence data and morphology. *American Journal of Botany* **94**: 313–329.

Davis CC, Bell CD, Fritsch PW, Mathews S. 2002. Phylogeny of *Acridocarpus-Brachylophon* (Malpighiaceae): implications for Tertiary tropical floras and Afroasian biogeography. *Evolution* **56**: 2395–2405.

Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemistry Bulletin, Botanical Society of America* **19**: 11–15.

Eriksson T. 1999. *AutoDecay, version 4·0·2*. Computer program distributed by the author. Bergius Foundation, Royal Swedish Academy of Sciences, Stockholm, Sweden.

Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**: 783–791.

Fitch WM. 1971. Toward defining the course of evolution: minimum change for a specific tree topology. *Systematic Zoology* **20**: 406–416.

Grassias M, Kammacher P. 1975. Observations sur le conjugaison chromosomique de *Coffea arabica* L. *Café, Cacao, Thé* **19**: 177–190.

Harris DJ, Poulsen AD, Frimodt-Møller C, Preston J, Cronk QCB. 2000. Rapid radiation in *Aframomum* (Zingiberaceae): evidence from nuclear ribosomal DNA internal transcribed spacer (ITS) sequences. *Edinburgh Journal of Botany* **57**: 377–395.

Holmgren PK, Holmgren NH, Barnett LC. 1990. *Index herbariorum*. In: Part 1: *The herbaria of the world*, 8th edn. *Regnum Vegetabile.* New York, NY: New York Botanical Garden.

Jordan WC, Courtney MW, Neigel JE. 1996. Low levels of intraspecific genetic variation at a rapidly evolving chloroplast DNA locus in North American duckweeds (Lemnaceae). *American Journal of Botany* **83**: 430–439.

Källersjö M, Farris JS, Kluge AG, Bult C. 1992. Skewness and permutation. *Cladistics* **8**: 275–287.

Lashermes P, Cros J, Marmey P, Charrier A. 1993. Use of random amplified DNA markers to analyse genetic variability and relationships of *Coffea* species. *Genetic Resources and Crop Evolution* **40**: 91–99.

Lashermes P, Cros J, Combes M-C, Trouslot P, Anthony F, Hamon S, Charrier A. 1996. Inheritance and restriction fragment length polymorphism of chloroplast DNA in the genus *Coffea* L. *Theoretical Applied Genetics* **93**: 626–632.

Lashermes P, Combes M-C, Trouslot P, Charrier A. 1997. Phylogenetic relationships of coffee-tree species (*Coffea* L.) as inferred from ITS sequences of nuclear ribosomal DNA. *Theoretical Applied Genetics* **94**: 947–955.

Lashermes P, Combes M-C, Robert J, Trouslot P, D'Hont A, Anthony F, Charrier A. 1999. Molecular characterisation and origin of the *Coffea arabica* L. genome. *Molecular and General Genetics* **261**: 259–266.

Lavin M, Thulin M, Labat J-N, Pennington RT. 2000. Africa, the odd man out: molecular biogeography of dalbergioid legumes (Fabaceae) suggests otherwise. *Systematic Botany* **25**: 449–467.

Leroy J-F. 1982. L'origine kenyane du genre *Coffea* L. et la radiation des espèces à Madagascar. *Association Scientifique Internationale du Café (ASIC)* **10e Colloque**: 413–420.

Leroy J-F. 1989. *Coffea* L. In: Bosser J, Cadet T, Guého J, Marais W, eds. *Flore des Mascareignes, La Réunion, Maurice, Rodrigues*. 107. Caprifoliacées à 108bis. Valérianacées. Surrey: Royal Botanic Gardens, Kew, 93–99.

Lobreau-Callen D, Leroy J-F. 1980. Quelques données palynologiques sur le genre *Coffea* et autres genres du cercle des caféiers.

Association Scientifique Internationale du Café (ASIC) **9e Colloque**: 479–506.

Lovett JC. 1985. Moist forests of Tanzania. *Swara* **8**: 8–9.

Lovett JC. 1988. Botanical importance of the Eastern Arc. *Journal of the East African Natural History Society* **87**: 59–74.

Malcomber ST. 2002. Phylogeny of *Gaertnera* Lam. (Rubiaceae) based on multiple DNA markers: evidence of rapid radiation in a widespread morphologically diverse genus. *Evolution* **56**: 42–57.

Maley J. 1987. Fragmentation de la forêt dense humide africaine et extension des biotopes montagnards au Quaternaire recent: nouvelles données polliniques et chronologiques. Implication palaéoclimatiques et biogéographiques. In: Coetzee JA ed. *Palaeoecology of Africa and the surrounding islands*. Rotterdam: A.A. Balkema, 307–334.

Mendenhall M. 1994. *Phylogeny of* Baptisia *and* Thermopsis *(Leguminosae) as inferred from chloroplast DNA and nuclear ribosomal DNA sequences, secondary chemistry, and morphology*. PhD Thesis, University of Texas, Austin.

Möller M, Cronk QCB. 2001. Phylogenetic studies in *Streptocarpus*: reconstruction of biogeographic history and distribution patterns in *Streptocarpus* (Gesneriaceae). *Systematics and Geography of Plants* **71**: 545–555.

Murphy WJ, Collier GE. 1997. A molecular phylogeny for aplocheiloid fishes (Atherinomorpha, Cyprinodontiformes): the role of vicariance and the origins of annualism. *Molecular Biology and Evolution* **14**: 790–799.

N'Diaye AN, Poncet V, Louran J, Hamon S, Noirot M. 2005. Genetic differentiation between *Coffea liberica* var. *liberica* and *C. liberica* var. *dewevrei* and comparison with *C. canephora*. *Plant Systematics and Evolution* **253**: 95–104.

Plana V, Gascoigne A, Forrest LL, Harris D, Pennington RT. 2004. Pleistocene and pre-Pleistocene *Begonia* speciation in Africa. *Molecular Phylogenetics and Evolution* **31**: 449–461.

Poorter L, Bongers F, 'Kouamé FN, Hawthorne WD. 2004. *Biodiversity of west African forests: an ecological atlas of woody plant species*. Wallingford, UK: CABI Publishing.

Portères R. 1937. Étude sur les caféiers spontanés de la section 'des Eucoffea'. *Annales Agricoles de L'Afrique Occidentale* **1**: 68–91.

Prakash NS, Combes M-C, Dussert S, Naveen S, Lashermes P. 2005. Analysis of genetic diversity in Indian robusta coffee genepool (*Coffea canephora*) in comparison with a representative core collection using SSRs and AFLPs. *Genetic Resources and Crop Evolution* **52**: 333–343.

Raina SN, Mukai Y, Yamamoto M. 1998. *In situ* hybridization identifies the diploid progenitor species of *Coffea arabica* (Rubiaceae). *Theoretical and Applied Genetics* **97**: 1204–1209.

Robbrecht E, Puff C. 1986. A survey of the Gardenieae and related tribes (Rubiaceae). *Botanische Jahrbücher für Systematik* **108**: 63–137.

Sehgal RH, Jones HI, Smith TB. 2005. Molecular evidence for host specificity of parasitic nematode microfilariae in some African rainforest birds. *Molecular Ecology* **14**: 3977–3988.

Sonké B, Stoffelen P. 2004. Une nouvelle espèce de *Coffea* L. (Rubiaceae, Coffeeae) du Cameroun et quelques notes sue les affinités avec les espèces voisines. *Adansonia, Sér. 3* **26**: 153–160.

Sonké B, Dessein S, Robbrecht E. 2006*a*. Rubiaceae. In: Akoègninou A, van der Burg WJ, van der Maesen LJG, eds. *Flore analytique du Bénin*. Leiden: Backhuys, 869–910.

Sonké B, Nguembou CK, Davis AP. 2006*b*. A new dwarf *Coffea* (Rubiaceae) from southern Cameroon. *Botanical Journal of the Linnean Society* **151**: 425–430.

Stoffelen P. 1998. Coffea *and* Psilanthus *in tropical Africa: a systematic and palynological study, including a revision of the West and Central African species*. PhD Thesis, Katholieke Universiteit, Leuven, Belgium.

Stoffelen P, Robbrecht E, Smets E. 1996*a*. *Coffea* (Rubiaceae) in Cameroon: a new species and a nomen recognised as species. *Belgian Journal of Botany* **129**: 71–76.

Stoffelen P, Robbrecht E, Smets E. 1996*b*. The '*arabica–congensis* complex': five *Coffea* species with a problematic delimitation. *Opera Botanica Belgica* **7**: 237–248.

Stoffelen P, Cheek M, Bridson DM, Robbrecht E. 1997*a*. A new species of *Coffea* (Rubiaceae) and notes on Mount Kupe (Cameroon). *Kew Bulletin* **42**: 989–994.

Stoffelen P, Robbrecht E, Smets E. 1997*b*. Adapted to the rain forest floor: a remarkable new dwarf *Coffea* (Rubiaceae) from Lower Guinea (tropical Africa). *Taxon* **46**: 37–47.

Stoffelen P, Robbrecht E, Smets E. 1997*c*. Pollen morphology of *Coffea* and *Psilanthus* (Rubiaceae), mainly from Africa. *Grana* **36**: 313–327.

Sun Y, Skinner DZ, Liang GH, Hulbert SH. 1994. Phylogenetic analysis of *Sorghum* and related taxa using internal transcribed spacers of nuclear ribosomal DNA. *Theoretical and Applied Genetics* **89**: 26–32.

Swofford DL. 2002. *PAUP* 4·0b10: Phylogenetic analysis using parsimony (* and other methods, Vers. 4)*. Sunderland, MA: Sinauer Associates.

Taberlet P, Gielly L, Pautou G, Bouvet J. 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology* **17**: 1105–1109.

White F. 1979. The Guineo-Congolian Region and its relationships to other phytochoria. *Bulletin du Jardin Botanique National de Belgique* **49**: 11–55.

White F. 1983. *The vegetation of Africa. A descriptive memoir to accompany the Unesco/AETFAT/UNSO vegetation map of Africa*. Paris: Unesco.